ИНСТИТУТ ЗА ФИЗИКУ

ПРИМЉЕНО: 25. 02. 2025			
Рад.јед.	б р о ј	Арх.шифра	Прилог
0801	336/1		

Научном већу Института за физику у Београду

Београд, 25.2.2025. године

Предмет: ***Молба за покретање поступка за избор у звање научни саветник***

Молим Научно веће Института за физику у Београду да, у складу с Правилником о поступку и начину вредновања и квантитативном исказивању научно-истраживачких резултата истраживача, покрене поступак за мој избор у звање научни саветник.

У прилогу достављам:

1. Мишљење руководиоца пројекта са предлогом чланова комисије за избор у звање
2. Стручну биографију
3. Преглед научне активности
4. Елементе за квалитативну оцену научног доприноса
5. Елементе за квантитативну оцену научног доприноса
6. Списак објављених радова и других публикација
7. Копије радова и других публикација
8. Податке о цитираности радова
9. Фотокопију решења о избору у претходно звање
10. Доказе о испуњености наведених квалитативних услова

С поштовањем,

Андреја Стојић

др Андреја Стојић

виши научни сарадник

Институт за физику у Београду

Научном већу Института за физику у Београду

Београд, 25.2.2025. године

Предмет: Мишљење руководиоца лабораторије о избору др Андреје Стојића у звање научни саветник

Др Андреја Стојић је запослен у Лабораторији за физику животне средине Института за физику у Београду. Његов рад обухвата теме из области опште и интердисциплинарне физике, хемије животне средине и науке о подацима с посебним фокусом на истраживање утицаја атмосферског загађења на животну средину и здравље људи. С обзиром на то да постигнути резултати задовољавају услове прописане Правилником о поступку, начину вредновања и квантитативном исказивању научноистраживачких резултата истраживача, сагласан сам са покретањем поступка за избор др Андреје Стојића у звање научни саветник.

За чланове комисије за избор др Андреје Стојића у звање научни саветник предлажем:

1. др Владимир Удовичић, научни саветник, Институт за физику у Београду
2. др Марија Врањеш Милосављевић, научни саветник, Институт за физику у Београду
3. проф. др Горан Попарић, редовни професор, Физички факултет, Универзитет у Београду

Руководилац Лабораторије за физику
животне средине

Зоран Мијић

др Зоран Мијић
виши научни сарадник
Институт за физику у Београду

2. БИОГРАФСКИ ПОДАЦИ О КАНДИДАТУ

Др Андреја Стојић је рођен 3. јануара 1976. године у Јагодини где је завршио основну школу и гимназију. Дипломирао је 2007. године на Физичком факултету Универзитета у Београду на смеру Примењена физика и информатика са просечном оценом 9,32. Дипломски рад под називом *Испитивање електричних и спектроскопских карактеристика коаксијалног диелектричног баријерног пражњења* је одбранио под менторством проф. др Братислава Обрадовића и проф. др Милорада Кураице.

Од јула 2007. године кандидат је запослен у Институту за физику у Београду као истраживач-приправник. Докторске студије на смеру Физика атома и молекула на Физичком факултету Универзитета у Београду завршио је са просечном оценом 10. Докторску дисертацију под називом *Анализа расподела и динамике испарљивих органских једињења и аеросола у тропосфери: лидар и масена спектрометрија* (*Spatio-temporal Distribution of Volatile Organic Compounds and Aerosols in Troposphere: Lidar and Mass Spectrometry*) одбранио је 2015. године под менторством др Зорана Мијића, вишег научног сарадника Института за физику у Београду.

Током докторских студија, бавио се истраживањем утицаја атмосферског загађења на животну средину и здравље људи. Примарни фокус био је утврђивање порекла атмосферских аеросола и испарљивих органских једињења (ИОЈ), њихове динамике и просторне расподеле, као и изучавање феномена и међусобних спрега који их дефинишу.

Кандидат је руководио (назначено), учествовао или учествује на следећим пројектима:

Међународни пројекти

1. 2019-2022. NI4OS-Europe: *National Initiatives for Open Science in Europe*; European Commission, Horizon 2020, Implementing the European Open Science Cloud
2. 2018-2022. *Persistent organochlorine compounds in breast milk and their effect on the level of primary DNA damage in human cells,* bilateral cooperation between the Republic of Serbia and Croatia
3. 2017-2021. *International network to encourage the use of monitoring and forecasting dust products,* COST Action CA16202, European Cooperation in Science and Technology
4. 2016-2018. GEO-CRADLE – *Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East, and Balkans and Developing Links with GEO related initiatives towards GEOSS*, Horizon 2020 (H2020) research and innovation programme under grant agreement No 690133
5. 2015-2019. No 654109: ACTRIS-2 (Aerosols, Clouds, and Trace gases Research Infrastructure) Project supported by the European Commission Horizon 2020 Research and Innovation Framework Programme
6. 2014-2017. *Atmospheric pressure plasma jet for neutralisation of CBW (chemical biological weapons)* – financed by NATO (SfP 984555)
7. 2006-2009. *Reinforcing Experimental Centre for Non-equilibrium Studies with Application in Nano-technologies, Etching of Integrated Circuits and Environmental Research* (IPB-CNP-026328), FP6

Национални пројекти

1. 2024-2027. crAIRsis, *Characterizing crises-caused air pollution alternations using an artificial intelligence-based framework*, бр. 7373, Фонд за науку Републике Србије, Програм ПРИЗМА – руководилац тима из Института за физику у Београду,

2. 2020-2022. ATLAS, *Artificial Intelligence Theoretical Foundations for Advanced Spatio-Temporal Modelling of Data and Processes*, бр. 6524105, Фонд за науку Републике Србије, Програм за развој пројеката у области вештачке интелигенције – руководилац тима из Института за физику у Београду,
3. 2018. *Мапирање извора токсичних, мутагених и канцерогених испарљивих органских једињења на територији Града Београда*, Зелени фонд, Министарство заштите животне средине Републике Србије – руководилац пројекта,
4. 2018. *Студија изводљивости имплементације националне мреже за континуално и аутоматизовано праћење значајних параметара из домена заштите животне средине*, Зелени фонд, Министарство заштите животне средине Републике Србије,
5. 2018. *Временске варијације и просторне карактеристике присуства испарљивих органских једињења и атмосферских честица у широј зони Београда – Реализација кампање фиксног и мобилног прикупљања података током грејне сезоне са аналитичким инструментима минутне резолуције*, Зелени фонд, Министарство заштите животне средине Републике Србије,
6. 2011-2019. *Истраживање климатских промена и њиховог утицаја на животну средину – праћење утицаја, адаптација и ублажавање* – III 43007, Министарство просвете, науке и технолошког развоја Републике Србије,
7. 2011-2019. *Примене нискотемпературних плазми у биомедицини, заштити човекове околине и нанотехнологијама* – III 41011, Министарство просвете, науке и технолошког развоја Републике Србије,
8. 2006-2010. *Емисија и трансмисија полутаната у атмосфери урбане средине* – OI 141012, Министарство просвете, науке и технолошког развоја Републике Србије,
9. 2008-2010. *Примена плазма игле у медицинским и биолошким истраживањима и брза и поуздана детекција волатилних супстанци хуманог и биљног порекла* – TR 23106, Министарство просвете, науке и технолошког развоја Републике Србије,
10. 2008-2009. *Развој и примена савремених археометријских-недеструктивних метода у анализи артефаката културног наслеђа* – TR 19046, Министарство просвете, науке и технолошког развоја Републике Србије.

Апликативни пројекти

1. 2024. *План квалитета ваздуха у агломерацији Нови Пазар*, Секретаријат за заштиту животне средине Града Новог Пазара,
2. 2023. *План квалитета ваздуха у агломерацији Зрењанин*, Секретаријат за заштиту животне средине Града Зрењанина,
3. 2022. *План квалитета ваздуха у агломерацији Панчево*, Секретаријат за заштиту животне средине Града Панчева,
4. 2021. *План квалитета ваздуха у агломерацији Нови Сад*, Секретаријат за заштиту животне средине Града Новог Сада,
5. 2020. *План квалитета ваздуха у агломерацији Београд*, Секретаријат за заштиту животне средине Града Београда – руководилац тима из Института за физику у Београду,
6. 2016. *План квалитета ваздуха у агломерацији Београд*, Секретаријат за заштиту животне средине Града Београда.

Такође, кандидат је током 2016-2017. године руководио фазама и активностима Националног центра изузетних вредности за примену плазме у нанотехнологијама, биомедицини и екологији Института за физику у Београду.

Истраживачки рад кандидата обухвата области опште и интердисциплинарне физике, хемије животне средине и науке о подацима. Активности се могу поделити у три сегмента:

- прикупљање података, које укључује експериментална мерења (концентрације загађујућих материја у амбијенталном ваздуху у реалним и симулираним мултифазним системима животне средине) и коришћење јавно доступних база података (загађујуће материје – EEA и US EPA, метеоролошки параметри – NOAA и ARL, морталитет – надлежне институције у Републици Србији, мобилност становништва – *Apple* и *Google*, мере државних институција – OxCGRT, пандемијска статистика – *Worldometer* и економски показатељи);
- моделирање података применом великог броја статистичких метода, метода машинског учења, метода оптимизације хиперпараметара алгоритама машинског учења и метода за објашњење модела машинског учења и
- развој експерименталних и/или статистичких метода за истраживање мултифазних система животне средине, транспорта загађења ваздуха и утицаја фактора животне средине на здравље људи и морталитет.

Теме истраживања су фокусиране на (1) анализу утицаја фактора животне средине који одређују еволуцију концентрација загађујућих материја у атмосфери у времену и простору, (2) симулацију и анализу мултифазних система животне средине, (3) анализу транспорта загађења ваздуха, (4) анализу утицаја фактора животне средине на живи свет, здравље људи и морталитет, као и (5) развој концепта за анализу података из области животне средине базираног на вештачкој интелигенцији.

Досадашњи рад кандидата укључује 28 радова категорија М20, као и 15 поглавља у међународним монографијама категорија М10. Од 28 радова, 5 је објављено у часописима изузетних вредности категорије М21а, 12 у врхунским међународним часописима категорије М21. Кандидат има развијену међународну научну сарадњу са истраживачком групом у Републици Хрватској (Институт за медицинска истраживања и медицину рада) и започету сарадњу са истраживачком групом у Републици Литванији (*Kaunas University of Technology*). Члан је научног одбора међународне конференције SINTEZA, односно организационог одбора међународне конференције CNN *Tech*.

Кандидат је учествовао је у израдама 2 докторске дисертације, 3 мастер рада и 8 дипломских радова, током школске 2016/2017. године водио је пројекат студентске праксе за студенте Физичког факултета Универзитета у Београду, а током 2019. године био ментор матурског рада ученика Математичке гимназије у Београду. Ментор је 1 докторске дисертације на Универзитету Сингидунум чија је одбрана планирана 2025. године.

Током 2019. године кандидат је учествовао у акредитацији, а потом је ангажован и као предавач на студијском програму *Животна средина и одрживи развој* Универзитета Сингидунум у Београду, на основним (3 предмета), мастер (2 предмета) и докторским студијама (1 предмет). Коаутор је уџбеника за практичну и теоријску наставу из уже научне области Наука о заштити животне средине Универзитета Сингидунум. Од 2024. године кандидат је у наставном звању ванредни професор.

3. ПРЕГЛЕД НАУЧНЕ АКТИВНОСТИ

Др Андреја Стојић бави се изучавањем утицаја фактора и процеса животне средине на концентрације загађујућих материја у атмосфери и њихову еволуцију у простору и времену који укључују природне и антропогене изворе емисије, физичко-хемијске процесе, метеоролошке и биотичке факторе, топографске карактеристике и др. Савремени трендови, као што су пораст броја становника урбаних средина, економски развој, потребе за енергијом, урбанизација и транспорт, питање загађења стављају у средиште пажње савременог друштва, првенствено због штетних ефеката на јавно здравље, животну средину и климатски систем. Идентификација и карактеризација појединачних извора загађења, локалне и макроскопске динамике (регионалне и глобалне), просторних расподела, доприноса укупном загађењу или режима животне средине који обликују загађење и механизама трансформације, представљају темељ за дубљи увид у њихов утицај на природне и антропогене екосистеме, укључујући њихову отпорност. Резултати ових истраживања пружају научну основу за формирање стратегија усмерених на побољшање квалитета животне средине, унапређење здравља људи и ублажавање ефеката климатских промена.

Кандидат у свом раду користи експерименталне и теоријске концепте и методе из различитих области опште и интердисциплинарне физике, атмосферске хемије и науке о подацима. Његова досадашња истраживања обухватају увођење напредних, унапређење постојећих и развој нових метода за мерење и анализу података који контекстуализују загађење ваздуха у отвореном и затвореном простору, мултифазне системе животне средине, транспорт загађења ваздуха и утицај фактора животне средине на биосферу, здравље људи и морталитет. У новијим истраживањима, примарни фокус кандидата усмерен је на анализу глобалне еволуције загађења ваздуха засноване на јавно доступним подацима и примени напредних метода ML, метахеуристика и XAI.

Рад кандидата се може поделити на следеће теме:

- Физика и хемија животне средине,
- Мултифазни системи животне средине,
- Транспорт загађења ваздуха,
- Утицај фактора животне средине на живи свет, здравље људи и морталитет,
- Развој концепта за анализу података из области животне средине базираног на вештачкој интелигенцији.

Напомена: радови објављени након претходног избора у звање су означени звездицом (*).

3.1 Физика и хемија животне средине

Циљ истраживања у оквиру ове теме, коју је кандидат покренуо током докторских студија, је испитивање просторно-временских расподела загађења ваздуха насталих појединачним и комбинованим утицајима фактора животне средине. Ови фактори обухватају изворе емисије, дисперзију, транспорт загађења, сезоналност, механизме трансформације, топографију, итд. Истраживања су базирана на анализи ИОЈ,[1] аеросола

[1] Прве научне активности кандидата су биле везане за увођење методе масене спектрометрије са трансфером протона (*proton transfer reaction mass spectrometry* – PTR-MS) и мерење концентрација ИОЈ у реалном времену

и њиховог хемијског састава (елементни и органски угљеник, тешки метали, јони, полициклични ароматични угљоводоници), неорганских гасова (угљен моноксид, тропосферски озон, оксиди азота, сумпор диоксид и радон) и чађи. Примена великог броја статистичких метода (методе анализе временских серија, рецепторски модели, мултифрактал и инверзна мултифрактал анализа, хибридни рецепторски модели, различите врсте поларних зависности од компоненти ветра, итд.), нумеричких модела (дисперзија), метода машинског учења и њихова хибридизација, омогућила је значајно побољшање квалитета и општег нивоа закључака у вези са процесима који одређују порекло и еволуцију загађујућих материја у ваздуху. Коришћена методологија обезбедила је прецизнију идентификацију и квантификацију локалних, регионалних и удаљених извора загађења, као и фактора животне средине који одређују нивое, промене, флуктуације и сингуларитете концентрација загађујућих материја у тропосфери.

У оквиру ове теме испитане су могућности прогнозе концентрација PM_{10} и доприноса извора ИОЈ базиране на примени машинског учења (TMVA, ROOT). Применом рецепторских модела (позитивна факторизација матрица – PMF и *Unmix*) на концентрације ИОЈ, суспендованих честица (PM_{10}) и неорганских гасова (CO, NO_x, NO, NO_2 и SO_2), анализирани су доприноси који потичу из саобраћаја и индустрије.

Наведени резултати приказани су у следећим радовима:

- * *The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources.*
  S. Stanišić, M. Perišić, G. Jovanović, T. Milićević, S.H. Romanić, A. Jovanović, A., Šoštarić, V. Udovičić, **A. Stojić**
  Environ. Res. 193, p.110520 (2021) (ИФ: 8,431)
- *Comprehensive analysis of PM_{10} in Belgrade urban area on the basis of long-term measurements*
  A. Stojić, S.S. Stojić, I. Reljin, M. Čabarkapa, A. Šoštarić, M. Perišić, Z. Mijić
  Environ. Sci. Pollut. R. 23(11), 10722-10732 (2016) (ИФ: 2,741)
- *Assessment of PM_{10} pollution level and required source emission reduction in Belgrade area*
  M. Todorović, M. Perišić, M. Kuzmanoski, **A. Stojić**, A. Šoštarić, Z. Mijić, S. Rajšić
  J. Environ. Sci. Heal. A 50(13), 1351-1359 (2015) (ИФ: 1,276)
- *Forecasting hourly particulate matter concentrations based on the advanced multivariate methods*
  M. Perišić, D. Maletić., S.S. Stojić, S. Rajšić, **A. Stojić**
  Int. J. Environ. Sci. Te. 14(5), 1047-1054 (2017) (ИФ: 2,037)
- *Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods*
  A. Stojić, D. Maletić, S.S. Stojić, Z. Mijić, A. Šoštarić
  Sci. Total Environ. 521, 19-26 (2015) (ИФ: 3,976)
- *Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling*
  A. Stojić, S. S. Stojić, Z. Mijić, A. Šoštarić, S. Rajšić
  Atmos. Environ. 106, 71-79 (2015) (ИФ: 3,459)
- *Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling*
  A. Stojić, S.S. Stojić, A. Šoštarić, L. Ilić, Z. Mijić, S. Rajšić
  Environ. Sci. Pollut. R. 22(17), 13137-13152 (2015) (ИФ: 2,76)

- *Estimation of required PM_{10} emission source reduction on the basis of a 10-year period data*
  M. Perišić, M., **A. Stojić,** S.S. Stojić, A. Šoštarić, Z. Mijić, S. Rajšić
  Air Qual. Atmos. Hlth. 8(4), 379-389 (2014) (ИФ: 1,804)
- *Receptor modeling studies for the characterization of PM_{10} pollution sources in Belgrade*
  Z. Mijić, **A. Stojić**, M. Perišić, S. Rajšić, M. Tasić
  Chemical Industry and Chemical Engineering Quarterly, 18(4-2), 623-634 (2012) (ИФ: 0,533)
- *Seasonal variability and source apportionment of metals in the atmospheric deposition in Belgrade*
  Z. Mijić, **A. Stojić**, M. Perišić, S. Rajšić, M. Tasić, M. Radenković, J. Joksić
  Atmos. Environ. 44(30), 3630-3637 (2010) (ИФ: 3,226)

3.2 Мултифазни системи животне средине

Моноароматични угљоводоници бензен, толуен, етилбензен и изомери ксилена (*benzene, toluene, ethylbenzene, xylene* – BTEX) сматрају се главним носиоцима загађења пореклом из антропогених извора. Повишене концентрације ових токсичних, мутагених и канцерогених једињења у урбаним срединама представљају значајан научни и практични проблем, због њиховог штетног утицаја на животну средину и здравље људи. Циљ истраживања спроведених у оквиру ове теме био је испитивање механизама уклањања BTEX из атмосфере путем процеса мокре депозиције. Ова истраживања спроведена су у симулираним (лабораторијским) и реалним условима, како би се обезбедио шири увид у њихово понашање у атмосфери.

Развој и примена експерименталне методологије за анализу мултифазних система животне средине

За потребе симулације интеракција које прате мокру депозицију ИОЈ и експерименталног одређивања коефицијента расподеле ових једињења између течне и гасне фазе, развијен је оригинални динамички аналитички систем. Овај систем састоји се из динамичке реакционе коморе, у којој се симулирају процеси мокре депозиције, и PTR-MS, помоћу кога се мере промене концентрација ИОЈ у реалном времену. За обраду резултата мерења развијен је низ статистичких процедура за квантитативно одређивање релевантних параметара. Процедуре обухватају примену параметарских функција за фитовање сигнала и испитивање квалитета фита, одређивање карактеристика еквилибријума као што су време успостављања и концентрација, одређивање количина аналита у течној и гасној фази и друге.

Истраживања су показала да је фактор обогаћења течне фазе једињењима BTEX, дефинисан као однос коефицијента расподеле између фаза и Хенријеве константе, знатно већи од вредности коју предвиђа Хенријев закон. Иако су ароматична једињења на макроскопском нивоу хидрофобна, бензенов прстен може имати улогу акцептора водоничне везе. Афинитет BTEX према формирању водоничне везе одређен је њиховом тенденцијом да отпуштају електроне, израженим као јонизациони потенцијал. Показано је да због изразито негативне линеарне везе између фактора обогаћења и јонизационог потенцијала, водоничне везе нису механизам који доводи до обогаћења течне фазе. Анализе су показале да је адсорпција на граници фаза доминантан механизам обогаћења течне фазе. Ово је потврђено позитивном линеарном везом између фактора обогаћења и

хидрофобности, дефинисане логаритмом коефицијента расподеле између октанола и воде. Такође, утврђено је да обогаћење површине на граници фаза расте са величином молекула, као и да је обогаћење обрнуто пропорционално запремини течне фазе и молским уделима BTEX у гасној фази, што такође иде у прилог овом механизму. На крају, позитивна линеарна зависност између фактора обогаћења и ван дер Валсове површине указује на значајан утицај ван дер Валсових интеракција на адсорпцију (*физисорпцију*) и расподелу BTEX једињења у мултифазним системима.

Испитивање капацитета кише за уклањање ИОЈ из атмосфере

Поред адсорпције на граници фаза, у истраживањима спроведеним у реалним условима разматрани су утицаји молских удела BTEX у амбијенталном ваздуху, метеоролошких параметара и физичко-хемијских карактеристика и састава кише (утицај матрице). Утицај промене температуре на Хенријеву константу и фактор обогаћења размотрен је анализом вертикалних профила температуре између површине тла и висине базе облака. На овај начин добијен је вертикални профил фактора обогаћења, из кога се може закључити да се услови који одређују интеракције које доводе до расподеле BTEX између течне и гасне фазе не мењају значајно са повећањем висине.

Применом различитих статистичких метода и метода машинског учења идентификована су четири типа извора који одређују састав кише: извор гасовитих органских једињења, извор чврсте фракције, извор који карактерише висок удео елемената пореклом из Земљине коре и извор који представља фракцију аеросола. Резултати су показали да молски удели BTEX у амбијенталном ваздуху и поједине физичко-хемијске карактеристике имају значајнији утицај на расподелу ових једињења између фаза у односу на метеоролошке параметре. Такође, утврђено је да је фактор обогаћења узорака кише већи него у симулираним условима што се може приписати присуству молекула BTEX адсорбованих на површини аеросола.

Напредак у машинском учењу довео је до примене сложених алгоритама за предикцију, као и до развоја метода за интерпретацију добијених модела. У оквиру ове теме, применом машинског учења (*Extreme gradient boosting* – XGBoost) први пут су моделиране везе између концентрација TEX у кишници и фактора обогаћења, с једне стране, и различитих фактора животне средине, с друге стране (концентрације TEX у амбијенталном ваздуху, физичко-хемијски параметри кишнице и метеоролошки параметри).

Дубљи увид у физичко-хемијске процесе који управљају депозицијом TEX постигнут је интерпретацијом добијених модела коришћењем напредне методе за интерпретацију модела машинског учења *Shapley additive explanations* (SHAP). На овај начин су по први пут утврђене расподеле утицаја фактора животне средине на концентрације TEX у кишници и факторе обогаћења кишнице овим једињењима. Резултати су показали да су концентрације TEX у амбијенталном ваздуху, као и температуре кишнице и ваздуха, доминантни фактори који обликују расподеле ових једињења у кишници. Знатно мањи утицај приписан је брзини ветра, атмосферском притиску, замућености кишнице и садржају укупног органског угљеника, NO_3^-, Cl^- и K^+, док су се утицаји осталих фактора показали занемарљивим.

Наведени резултати су приказани у следећим радовима:

- *Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition*

A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić, L. Lazić
Sci. Total Environ. 653, 140–147 (2019) (ИФ: 5,589)

- *Rainwater capacities for BTEX scavenging from ambient air*
A. Šoštarić, S.S. Stojić, G. Vuković, Z. Mijić, **A. Stojić**, I. Gržetić
Atmos. Environ. 168, 46-54 (2017) (ИФ: 3,708)
- *Quantification and mechanisms of BTEX distribution between aqueous and gaseous phase in a dynamic system*
A. Šoštarić, **A. Stojić**, S.S. Stojić, I. Gržetić
Chemosphere, 144, 721-727 (2016) (ИФ: 4,208)

3.3 Транспорт загађења ваздуха

Идентификација и карактеризација удаљених извора емисије загађења ваздуха и њиховог доприноса измереним концентрацијама загађујућих материја на месту рецептора (мерно место) могу се извршити применом хибридних рецепторских модела (*hybrid receptor models* – HRM). Ови модели заснивају се на повезивању концентрација измерених на месту рецептора са трајекторијама транспорта ваздуха из удаљених области.

Међутим, постоје четири основна недостатка стандардних HRM:

- **Недовољно укључивање релевантних фактора за транспорт загађења**. Стандардни модели не узимају у обзир кључне факторе који утичу на транспорт загађења са места рецептора. То доводи до прецењивања утицаја удаљених извора и нетачне идентификације географских области које представљају извор загађења.
- **Дводимензионални приступ**. Стандардни модели користе дводимензионални приступ, што онемогућава моделирање вертикалних расподела загађења. Овај недостатак је значајан, јер вертикални профили играју важну улогу у разумевању циркулације ваздуха и процени изложености људи и животне средине.
- **Ниска резолуција крајњих тачака трајекторија**. Резолуција крајњих тачака у стандардним моделима је веома ниска, што онемогућава прецизну идентификацију области релевантних за анализу транспорта. Ово ограничење такође спречава примену HRM за анализу локалних извора загађења, који се налазе у релативној близини мерног места.
- **Немогућност карактеризације типова извора загађења**. Стандардни модели не могу се користити за карактеризацију загађења и диференцирање различитих врста извора емисије.

Најважнији резултат у оквиру истраживања у овој теми је развој тродимензионалног HRM – гранични слој отежињен концентрацијама (*concentration weighted boundary layer* – CWBL), који представља једини HRM који обезбеђује анализу континуалних вертикалних расподела загађења ваздуха дуж путања транспорта посматраних са места рецептора. По први пут су у истраживање укључени ефекти флуктуације планетарног граничног слоја и њихов утицај на транспорт загађења, као и на измерене концентрације на месту рецептора.

Поред наведених, значајни резултати огледају се и у развоју тродимензионалних варијанти постојећих HRM, попут тродимензионалне функције потенцијалних доприноса (3D *potential source contribution function* – 3D PSCF) и тродимензионалних трајекторија отежињених концентрацијама (3D *concentration weighted trajectory* – 3D CWT). Ови модели омогућавају анализу дискретних вертикалних расподела загађења, чиме пружају дубљи увид у транспорт загађујућих материја и њихово ширење унутар планетарног граничног слоја.

Веома значајан резултат истраживања представља и унапређење приступа анализи транспорта загађења применом HRM који обухвата прецизну формулацију релевантних података и апроксимација унутар модела. Ово унапређење се састоји из два кључна аспекта:

- **Издвајање удела транспортованог загађења у измереним концентрацијама загађујућих материја**. Овај приступ омогућава диференцијацију апсолутног удела позадинског нивоа загађења, локалних извора и процеса транспорта у концентрацијама измереним на месту рецептора. Укључивањем само удела концентрација који одговарају транспортованом загађењу, значајно се смањује проблем прецењивања утицаја удаљених извора емисије који је карактеристичан за стандардне моделе.
- **Идентификација репрезентативних крајњих тачака трајекторија кретања ваздуха**. Унапређење укључује постављање критеријума за избор крајњих тачака трајекторија кретања ваздуха на основу висине планетарног граничног слоја. На овај начин из анализе транспорта искључују се крајње тачке које се не могу довести у везу са измереним концентрацијама на месту рецептора. Ово значајно побољшава прецизност идентификације географских области које представљају извор загађења.

Такође, као резултат пројекта *Мапирање извора токсичних, мутагених и канцерогених испарљивих органских једињења на територији Града Београда*, који је финансирао Зелени фонд Министарства заштите животне средине Републике Србије којим је кандидат руководио, развијена је иновативна методологија која је остварила и практичну примену. Методологија је заснована на новим, локалним рецепторски оријентисаним моделима и алгоритмима вештачке интелигенције имплементираним кроз методе машинског учења и XAI. Она омогућава мапирање и карактеризацију извора, као и просторно-временску прогнозу концентрација загађујућих материја у ваздуху. Такође, доприноси решавању два значајна проблема стандардних HRM, проблема репрезентативности области релевантних за анализу транспорта и ограничења које онемогућава анализу локалних извора загађења и карактеризацију извора емисије.

Истраживања у оквиру ове теме показала су да развијене методе издвајања удела концетрација и репрезентативних крајњих тачака трајекторија, као и нови модели, омогућавају знатно прецизнију идентификацију и карактеризацију локалних и удаљених извора загађења ваздуха. Поред испитивања расподела појединих загађујућих материја, истраживања обухватају и анализу расподела извора загађења, укључујући саобраћај и индустрију. Овим унапређењима HRM обезбеђују тачније и поузданије резултате у

анализи транспорта загађења ваздуха, омогућавајући дубље разумевање порекла и утицаја како блиских, тако и удаљених извора емисије.

Наведени резултати приказани су у следећим радовима:

- *The innovative concept of three-dimensional hybrid receptor modeling*
  A. Stojić and S.S. Stojić
  Atmos. Environ. 164, 216-223 (2017) (ИФ: 3,708)
- *Levels of PM_{10} bound species in Belgrade, Serbia: spatio-temporal distributions and related human health risk estimation*
  M. Perišić, S. Rajšić, A. Šoštarić, Z. Mijić, **A. Stojić**
  Air Qual. Atmos. Hlth. 10(1), 93-103 (2017) (ИФ: 2,662)
- *Comprehensive analysis of PM_{10} in Belgrade urban area on the basis of long-term measurements*
  A. Stojić, S.S. Stojić, I. Reljin, M. Čabarkapa, A. Šoštarić, M. Perišić, Z. Mijić Environ. Sci. Pollut. R. 23(11), 10722-10732 (2016) (ИФ: 2.741)
- *Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling*
  A. Stojić, S. S. Stojić, Z. Mijić, A. Šoštarić, S. Rajšić
  Atmos. Environ. 106, 71-79 (2015) (ИФ: 3,459)

3.4 Утицај фактора животне средине на живи свет, здравље људи и морталитет

Током последњих деценија загађење ваздуха препознато је као једна од највећих глобалних претњи по здравље људи. Процењује се да је преко 8 милиона смртних случајева годишње, узрокованих кардиоваскуларним, малигним и хроничним респираторним обољењима, директно повезано са загађењем ваздуха.

У оквиру ове теме истраживани су канцерогени и неканцерогени утицаји честичног загађења у Београду, са посебним фокусом на његов хемијски састав, који укључује тешке метале и бензо[а]пирен. Резултати су показали да хром и бензо[а]пирен значајно повећавају ризик од настанка канцера, док су ефекти арсена и никла изразито токсични, нарочито на урбаним локацијама под утицајем саобраћаја. Поред тога, утврђено је да транспортовано загађење у одређеним периодима може значајно допринети укупном загађењу, са уделом од чак 36%.

Поред канцерогених и неканцерогених здравствених ризика, истраживани су и утицаји краткорочне и дугорочне изложености загађујућим материјама (PM_{10}, SO_2, NO_2 и чађ) на морталитет изазван кардио-васкуларним и респираторним обољењима. Нелинеарна веза између изложености загађењу, ризика од смртности и одложених ефеката услед варијација температуре моделирана је применом *distributed lag nonlinear models* – DLNM. Резултати су показали да краткорочна изложеност повишеним концентрацијама загађујућих материја не повећава значајно ризик од смртности. Међутим, јасна повезаност утврђена је у случају изложености у трајању од 90 дана. Хронична изложеност загађењу показала је израженији утицај на смртност од респираторних обољења у поређењу са кардио-васкуларним, посебно у мушкој популацији млађој од 65 година.

У оквиру ове теме развијена је нова метода за укључивање кумулативних средњорочних ефеката загађења ваздуха у Поасонов регресиони модел, са циљем процене ризика смртности од кардио-васкуларних и респираторних обољења под утицајем климатских

фактора. Реалистичније процене ризика повезаног са екстремним климатским условима постају све важније за креирање будућих стратегија и мера прилагођавања актуелним климатским променама.

Истраживања су показала да су кумулативни средњорочни ефекти загађења ваздуха значајнији од одложених (*lag-specific*), који су традиционално укључивани у регресионе моделе. Такође, утврђено је постојање оптималног температурног опсега у коме се не очекује повећање стопе смртности узроковане променама температуре, што се разликује од налаза претходних студија. С друге стране, резултати указују да ефекти загађења ваздуха пружају боље објашњење ризика од смртности током хладнијег времена, који је раније био приписиван искључиво утицају температуре. Поред тога, утврђено је да је релативни значај честичног загађења мањи у односу на сумпор-диоксид, азот-диоксид и чађ. Ово указује да укључивање искључиво података о концентрацији аеросола не представља најефикаснији приступ за процену утицаја загађења ваздуха на здравље људи.

У оквиру ове теме разматрана је акумулација перзистентних органских полутаната (*persistent organic pollutants* – POPs) у мајчином млеку и њихова повезаност са годинама мајке и бројем рођене деце. Применом великог броја статистичких метода и ML истражена је улога конституентних дескриптора конгенера у акумулацији органохлорних пестицида (OCPs) и полихлорованих бифенила (PCBs). Посебно су значајним показани број и положај атома хлора на фенил прстену, попут орто-положаја. Резултати су показали да нивои PCBs не зависе од броја рођене деце. С друге стране, утврђена је значајна међусобна веза између PCB конгенера -153, -180, -170, -118, -156, -105 и -138, што се приписује њиховој хемијској структури и метаболичким процесима у организму мајке.

Такође, у оквиру ове теме испитивана је акумулација масних киселина, POPs и елемената у траговима у малим пелагичним рибама (сардина, инћун, округла сардинела, скуша и белаци) у источном Медитерану. Резултати су показали да су неорганска једињења знатно заступљенија у ткиву риба у поређењу са органохлорним ксенобиотицима, попут OCPs и PCBs. Анализа профила повезаности ових једињења идентификовала је пет извора, укључујући исхрану морским организмима из нижих трофичких ниша и доприносе из околних компоненти екосистема. Поређење са токсиколошким параметрима показало је да су испитиване врсте риба безбедне за исхрану људи, док садржај масних киселина указује на њихову вредност као нутритивно корисне хране. Истраживање није открило значајну корелацију између 18 масних киселина и липофилних OCPs. Ови резултати допринели су бољем разумевању односа органских и неорганских загађујућих материја и нутритивних параметара у пелагичним рибама , чиме се омогућава прецизнија процена квалитета јестивих врста риба и стања акватичног екосистема источног Медитерана. Истраживање наглашава важност ових врста као извора нутритивно вредне хране, али и као потенцијалних извора опасних органохлорних једињења која могу негативно утицати на здравље људи. Резултати представљају основу за будуће процене ризика, као и за еколошке анализе и оцене стања животне средине у овом риболовном подручју.

Наведени резултати су приказани у следећим радовима:

- * *Fatty acids, persistent organic pollutants, and trace elements in small pelagic fish from the eastern Mediterranean Sea,*
  S. Herceg Romanić, G. Jovanović, B. Mustać, J. Stojanović-Đinović, **A. Stojić,** T. Čadež,

A. Popović
Mar. Pollut. Bull, 170, p. 112654 (2021) (ИФ: 7,001)

- *Introducing of modeling techniques in the research of POPs in breast milk – A pilot study*
  G. Jovanović, S. Herceg Romanić, **A. Stojić**, D. Klinčić, M. Matek Sarić, J. Grzunov Letinić, A. Popović
  Ecotox. Environ. Safe., **172**, 341-347 (2019) (ИФ: 4,527)
- *Levels of PM_{10} bound species in Belgrade, Serbia: spatio-temporal distributions and related human health risk estimation*
  M. Perišić, S. Rajšić, A. Šoštarić, Z. Mijić, **A. Stojić**
  Air Qual. Atmos. Hlth. **10(1)**, 93-103 (2017) (ИФ: 2,662)
- *Temperature-related mortality estimates after accounting for the cumulative effects of air pollution in an urban area*
  S.S. Stojić, N. Stanišić, **A. Stojić**
  Environ. Health, **15(1)**, 73 (2016) (ИФ: 3,816)
- *Single and combined effects of air pollutants on circulatory and respiratory system-related mortality in Belgrade, Serbia*
  S.S. Stojić, N. Stanišić, **A. Stojić**, A. Šoštarić
  J. Toxicol. Env. Heal. A **79(1)**, 17-27 (2016) (ИФ: 2,731)
- *Seasonal mortality variations of cardiovascular, respiratory and malignant diseases in the City of Belgrade*
  S.S. Stojić, N. Stanišić, **A. Stojić**, V. Džamić
  Stanovništvo, 54(1), 83-104 (2016)
- *Heavy metal accumulation in wheat and barley: The effects of soil presence and liquid manure amendment*
  S.S. Stojić, L. Ignjatović, S. Popov, S. Škrivanj, A. Đorđević, **A. Stojić**
  Plant Biosyst. 150(1), 104-110 (2016) (ИФ: 1,39)
- *Essential oils of two Nepeta species inhibit growth and induce oxidative stress in ragweed (Ambrosia artemisiifolia L.) shoots in vitro*
  S. Dmitrović, M. Perišić, **A. Stojić**, S. Živković, J. Boljević, J.N. Živković, D. Mišić
  Acta Physiol. Plant. 37(3)**,** 1-15 (2015) (ИФ: 1,563)

3.5 Развој концепта за анализу података из области животне средине базираног на вештачкој интелигенцији

Наука о животној средини суочава се са бројним проблемима у разумевању сложености утицаја различитих фактора на нивое и динамику загађења. Проблеми укључују изразиту комплексност, мултифазност и нелинеарност феномена унутар екосистема, недовољну доступност података високог квалитета, што онемогућава адекватну контекстуализацију феномена, као и изразиту хетерогеност у квалитету примењених научних концепата и метода.

Концентрације загађујућих материја одређују бројни чиниоци, као што су карактеристике медијума животне средине, физичко-хемијске трансформације, климатски и метеоролошки услови, топографија, биотички ефекти, утицај друштвених активности и многи други. Ови фактори, који чине контекст који одређује стање животне средине, интерагују на сложене начине, изазивајући значајне промене у просторној и временској расподели загађења.

У оквиру пројеката Фонда за науку Републике Србије у којима је кандидат учествовао, односно учествује (ATLAS и crAIRsis), развијен је напредни концепт заснован на вештачкој интелигенцији (AI) ради превазилажења ограничења традиционалних приступа. Централно место представља идентификација и детаљна карактеризација режима животне средине (*environmental settings*), концепта који је кандидат увео са својим истраживачким тимом. Режими животне средине интегришу интеракције природних и антропогених фактора одговорних за понашање загађујућих материја и стање животне средине.

Примарни ниво концепта подразумева контекстуализацију феномена применом алгоритама машинског учења оптимизованих метахеуристикама. Имплеметиране су напредне методе ансамбала стабала одлучивања и дубоког учења, као и многобројне метахеуристике. Овај приступ омогућава повезивање варијабли са феноменом који се истражује у комплексан модел. Добијени модел се интерпретира применом многих метода XAI попут SHAP, nSHAP, gSHAP, SAGE, iSAGE и Banzhaf-ов индекс. Идентификација и карактеризација режима животне средине врши се кластеризацијом локалних утицаја помоћу метода попут HDBSCAN уз претходну редукцију димензионалности методама попут UMAP, Trimap и PaCMAP, као и детаљну статичку анализу добијених кластера конвенционалним методама анализе података из области животне средине. Средњи ниво концепта подразумева симулације путем виртуелних експеримената који омогућавају развој сценарија и евентуалну процену ефеката предложених мера за редукцију загађења. Финални ниво представља аутоматизовану интерпретацију добијених резултата и њихову интеграцију са језичким моделима који омогућавају додатно истраживање резултата применом сентимент анализе или XAI. Цео концепт подразумева и анализу пропагација несигурности, од грешака мерења, до грешака сваког нивоа моделирања.

Овај концепт примењен је за анализу динамике бензена током ванредног стања у Београду током пандемије Ковид-19. Почетак пандемије обележен је минималним активностима људи, попут транспорта и индустрије, док су активности попут грејања и кућних активности достигле максимуме. Коришћени су подаци који обухватају концентрације ИОЈ, метеоролошке параметре, бројеве заражених и умрлих, мере примењене у циљу сузбијања пандемије, мобилност и друштвено-економске показатеље са више од 400 варијабли. Испитани су и одложени (*lag*) ефекти појединих променљивих. Идентификовано је девет режима животне средине одговорних за варијације концентрација бензена у различитим фазама ванредног стања (увођење ванредног стања, увођење полицијског часа и релаксација мера). Извршена је карактеризација утицаја индустријске производње хемијских једињења, сагоревања нафтних деривата, емисија које не укључују сагоревање, ноћне хемије, метеоролошких прилика, локалних индустријских и других процеса.

Такође, истраживања у оквиру ове теме обухватила су и анализу утицаја режима животне средине на динамику полицикличних ароматичних угљоводоника у ваздуху отвореног и затвореног простора. Идентификована су четири различита режима и одговарајући подрежими одговорни за концентрације бензо[a]пирена који укључују различите облике непотпуног и високо-температурног сагоревања, депозицију, међусобну интеракцију ваздуха спољашњег и унутрашњег простора кроз инфилтрацију и друге механизме.

У оквиру ове теме започето је истраживање стања квалитета ваздуха у Европи и Сједињеним Америчким Државама сагледано кроз 70 загађујућих материја (EEA и US EPA) контекстуализовано уз помоћ више од 100 додатних варијабли (метеоролошки параметри – NOAA, ARL, мобилност становништва – *Apple* и *Google*, мере државних

институција – OxCGRT, пандемијски – *Worldometer* и економски показатељи). Истраживање обухвата појединачну и збирну анализу података са преко 6000 мерних места.

Наведени резултати приказани су у следећим радовима:

- ** The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere*
  N. Bačanin, M. Perišić, G. Jovanović, R. Damaševičius, S. Stanišić, V. Simić, M. Zivković, **A. Stojić**
  Sci. Total Environ. 929, 172195 (2024) (ИФ: 9.8)
- ** Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing PAHs Environmental Fate*
  G. Jovanović, M. Perišić, N. Bačanin, M. Živković, S. Stanišić, I. Štrumberger, F. Alimpić, **A. Stojić**
  Toxics 11(4):394 (2023) (ИФ: 4.6)
- ** The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels*
  A. Stojić, G. Jovanović, S. Stanišić, S.H. Romanić, A. Šoštarić, V. Udovičić, M. Perišić, T. Milićević
  Chemosphere, 289, p.133154 (2022) (ИФ: 8.943)
- ** The explainable potential of coupling metaheuristics-optimized-xgboost and shap in revealing VOCs' environmental fate*
  L. Jovanović, G. Jovanović, M. Perišić, F. Alimpić, S. Stanišić, N. Bačanin, M. Živković, **A. Stojić**
  Atmosphere, 14(1), p.109 (2023) (ИФ: 3.11)
- ** The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part III: Role of environmental settings in elevating indoor concentrations of benzo(a)pyrene*
  G. Jovanović, M. Perišić, T. Bezdan, S. Stanišić, K. Radusin, A. Popović, **A. Stojić**, Atmosphere 15(12), 1520 (2024) (ИФ: 2.9)
- ** An AI-based framework for characterizing the atmospheric fate of air pollutants within diverse environmental settings*
  N. Radić, M. Perišić, G. Jovanović, T. Bezdan, S. Stanišić, N. Stanić, **A. Stojić**
  Atmosphere, 16(2), 231 (2025) (ИФ: 2.9)

4. ЕЛЕМЕНТИ ЗА КВАЛИТАТИВНУ ОЦЕНУ НАУЧНОГ ДОПРИНОСА КАНДИДАТА

4.1 Квалитет научних резултата

4.1.1 Научни ниво и значај резултата, утицај научних радова

Др Андреја Стојић је у свом досадашњем раду дао кључни допринос у истраживању на укупно 28 радова објављених у категорији М20, као и 15 поглавља у књизи категорије М10, од којих је 10 објављено у истакнутим монографијама међународног значаја. Од 28 радова, 5 је објављено у часописима категорије М21а (међународни часописи изузетних вредности), 12 у часописима категорије М21 (врхунски међународни часописи), 7 у часописима категорије М22 (истакнути међународни часописи), 2 у часописима категорије М23 (међународни часописи), док је 1 објављен у категорији М24 (национални часописи међународног значаја).

У периоду након доношења одлуке Научног већа Института за физику о предлогу за стицање претходног научног звања, кандидат је објавио 8 радова у часописима са ISI листе. Од тога су 3 рада објављена у часопису категорије М21а (међународни часописи изузетних вредности), 2 у часописима категорије М21 (врхунски међународни часописи) и 3 у часописима категорије М22 (истакнути међународни часописи). Такође, кандидат је у том периоду објавио 2 поглавља у истакнутим монографијама међународног значаја категорије М13, 3 поглавља у монографијама међународног значаја категорије М14 и одржао 1 предавање по позиву на међународном скупу.

Као пет најзначајнијих радова кандидата издвајају се:

1. *The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources*
   S. Stanišić, M. Perišić, G. Jovanović, T. Milićević, S.H. Romanić, A. Jovanović, A., Šoštarić, V. Udovičić, **A. Stojić**
   Environ. Res. 193, p.110520 (2021) (ИФ: 8,431) цитиран 14 пута.
2. *Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition*
   A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić and L. Lazić
   Sci. Total Environ. 653, 140–147 (2019), M21 (ИФ: 5,589), цитиран 61 пут.
3. *The innovative concept of three-dimensional hybrid receptor modeling*
   A. Stojić and S. Stanišić Stojić
   Atmos. Environ. 164, 216 223 (2017), M21 (ИФ: 3,708), цитиран 18 пута.
4. *Temperature-related mortality estimates after accounting for the cumulative effects of air pollution in an urban area*
   S. Stanišić Stojić, N. S. Stanišić and **A. Stojić**
   J. Environ. Health, 15(1), 73 (2016), M21a (ИФ: 3,816), цитиран 12 пута.
5. *The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere*
   N. Bačanin, M. Perišić, G. Jovanović, R. Damaševičius, S. Stanišić, V. Simić, M.

Zivković and **A. Stojić**
Sci. Total Environ. 929, 172195 (2024), M21a (ИФ: 9,8), цитиран 1 пут.

У свих 5 радова кандидат је дао кључни научни допринос и може сматрати основним/најважнијим аутором. Први и пети рад објављени су у периоду након избора у претходно звање.

У првом раду приказана је широка анализа полицикличних ароматичних угљоводоника у ваздуху отвореног и затвореног простора која обухвата велики број измерених параметара и метода обраде података. Резултати студије омогућили су разумевање интеракција између метеоролошких фактора и загађујућих материја у сложеном урбаном окружењу. Мерења су обухватала концентрације O_3, CO, SO_2, NOx, радона, $PM_{2,5}$ и њиховог састава (16 приоритетних US EPA PAH, елементи As, Cd, Cr, Mn, Ni и Pb, као и јони Cl^-, Na^+, Mg^{2+}, Ca^{2+}, K^+, NO_3^-, SO_4^{2-} и NH_4^+), као и више од 30 метеоролошких параметара. Методе анализе података обухватале су традиционалне приступе које укључују анализе временских серија, густина расподела података, рецепторске моделе и многе друге. Највећи утицај на квалитет ваздуха приписан је изворима попут сагоревања угља и сродних пирогених процеса. Уочене су значајне корелације између PAH са пет и шест прстенова, док, изузев CO, није идентификована значајна линеарна зависност са другим истраженим варијаблама. Регресиона анализа концентрација PAH у затвореном и отвореном простору извршене су применом алгоритма машинског учења XGBoost, који је показао висок степен прилагођавања непараметарским расподелама података и толеранцију на шум. Овај приступ омогућио је значајно побољшање квалитета и општости нивоа закључака у вези са процесима који одређују порекло односа између метеоролошких параметара и концентрација O_3, CO, SO_2, NOx, радона, $PM_{2,5}$ и елемената у траговима.

У другом раду истражени су фактори животне средине који утичу на уклањање толуена, етилбензена и ксилена из амбијенталног ваздуха путем влажне депозиције у оквиру биогеохемијског циклуса ИОЈ. Анализиране су расподеле TEX између течне и гасне фазе, као и одговарајући фактори обогаћења кишнице. Показано је да су концентрације ових једињења у амбијенталном ваздуху, као и температуре кишнице и ваздуха, доминантни фактори који утичу на расподелу TEX у кишници. Далеко мање важни утицаји могу се приписати брзини ветра, атмосферском притиску, замућености кишнице и садржају укупног органског угљеника, NO_3^-, Cl^- и K^+, док су утицаји осталих фактора занемарљиви. По први пут, односи између концентрација TEX у кишници, фактора обогаћења и бројних фактора животне средине (концентрације TEX у амбијенталном ваздуху, физичко-хемијски параметри кишнице и метеоролошки параметри) моделирани су применом алгоритма машинског учења XGBoost. Физичко-хемијски процеси који управљају депозицијом TEX анализирани су применом методе SHAP. Овај приступ омогућио је утврђивање расподеле утицаја фактора животне средине на концентрације TEX у кишници и детаљну анализу фактора који утичу на обогаћење кишнице овим једињењима.

У трећем раду приказани су резултати унапређеног приступа анализи транспорта загађења ваздуха базираног на хибридним рецепторским моделима. Основни недостаци традиционалних модела огледали су се у недовољном укључивању фактора релевантних за транспорт загађења посматраног са места рецептора, као и ограничавање на анализу у две димензије (географска ширина и географска дужина). Ови недостаци онемогућавали су испитивање вертикалних расподела загађења, које су од суштинског значаја за разумевање циркулације ваздуха и процену утицаја на изложеност људи и стање животне средине. Унапређење приказано у раду укључује развој три кључна сегмента који

омогућавају анализу вертикалних расподела загађења ваздуха дуж путања транспорта. Први сегмент односи се на издвајање удела транспортованог загађења из измерених концентрација загађујућих материја, чиме се постиже диференцијација између позадинских нивоа, локалних извора и утицаја транспорта. На овај начин моделирају се само удели концентрација који су директно повезани са транспортованим загађењем. Други сегмент односи се на идентификацију крајњих тачака трајекторија кретања ваздуха које су репрезентативне за транспорт загађења, при чему се критеријуми заснивају на висини планетарног граничног слоја. Овај приступ елиминише тачке које се не могу повезати са измереним концентрацијама на месту рецептора, повећавајући прецизност анализе. Трећи сегмент подразумева развој првих тродимензионалних хибридних рецепторских модела, укључујући 3D функцију потенцијалних доприноса (3D PSCF), 3D трајекторије отежињене концентрацијама (3D CWT) и гранични слој отежињен концентрацијама (CWBL). Ови напредни модели омогућавају детаљну анализу вертикалних расподела и сложених процеса који утичу на транспорт загађења, значајно унапређујући разумевање циркулације ваздуха и утицаја на животну средину.

У четвртом раду представљена је нова метода за интеграцију кумулативних средњорочних ефеката загађења ваздуха у Поасонов регресиони модел, са циљем процене ризика од смртности од кардио-васкуларних и респираторних обољења под утицајем климатских фактора. Овај приступ пружа реалистичније процене ризика повезаних са екстремним климатским условима, што је од кључног значаја за планирање будућих стратегија и мера прилагођавања климатским променама. Резултати су показали да су кумулативни средњорочни ефекти загађења ваздуха значајнији од одложених, који су традиционално укључивани у регресионе моделе. Такође, показано је и да постоји оптимални температурни опсег унутар кога се не очекује повећање стопе смртности услед промене температуре, што се разликује од претходних студија. Истовремено, анализе су показале да ефекти загађења ваздуха могу прецизније објаснити ризик од смртности током хладнијих периода, који се раније приписивао искључиво ниским температурама. На тај начин утврђено је да глобално не важи правило по којем је смртност узрокована екстремно хладним временом један ред величине нижа од смртности повезане са топлотним таласима. На крају, анализа је показала да је релативни значај честичног загађења мањи у односу на сумпор-диоксид, азот-диоксид и чађ, што указује да коришћење података искључиво о концентрацији аеросола није довољно за ефикасну процену утицаја загађења ваздуха на здравље људи.

У петом раду доказана је валидност развијеног концепта за анализу података из области животне средине базираног на вештачкој интелигенцији који истиче важност интеракција између загађујућих материја и метеоролошких услова у моделирању квалитета ваздуха. Такође, представљена је унапређена верзија алгоритма метахеуристике *Reptile search*, коришћена за оптимизацију хиперпараметара методе XGBoost са циљем истраживања атмосферских нивоа толуена и његових интеракција са другим загађујућим материјама у различитим режимима животне средине. Толуен, као неуротоксични ароматични угљоводоник и један од главних представника ИОЈ, познат је по својој распрострањености, негативним утицајима на здравље и улози у формирању других загађујућих материја попут озона. Истраживање је базирано на двогодишњем скупу података сатне резолуције, који укључује концентрације неорганских гасовитих загађујућих материја (NO, NO_2, NOx и O_3), фракције суспендованих честица (PM_1, $PM_{2,5}$ и PM_{10}), бензена, м,п-ксилена, толуена, укупних неметанских угљоводоника и метеоролошке параметре. Експериментални резултати валидирани су у поређењу са XGBoost моделима оптимизованим применом седам савремених метахеуристика, а најбољи модел интерпретиран је применом методе SHAP. Установљено је да је бензен

најзначајнији предиктор за динамику толуена, а интеракције између ове две загађујуће материје значајно варирају у зависности од режима животне средине. Идентификована су три режима која одређују повезаност толуена и бензена, као и њихову интеракцију са другим загађујућим материјама. Предложени хибридни алгоритам *Reptile search* показао се као супериорна метода за оптимизацију методе XGBoost, са одличним перформансама у анализи динамике толуена у комплексном урбаном окружењу.

4.1.2 Позитивна цитираност научних радова кандидата

Према бази *ISI Web of Science*, радови кандидата укупно су цитирани 485 пута, док је број цитата без аутоцитата 408. Према бази *Scopus*, укупан број цитата је 623, док је број цитата без аутоцитата 506. Према подацима из обе базе, Хиршов индекс радова кандидата је 13, односно 15.

Прилог: подаци о цитираности радова из интернет база ISI Web of Science и Scopus до 25.2.2025. године

4.1.3 Параметри квалитета часописа

Као елемент за процену квалитета научних радова користи се и импакт-фактор часописа у којима су радови објављени. Кандидат је објављивао радове у часописима категорија M21a, M21, M22, M23 и M24, при чему су подвучени импакт-фактори часописа у којима су публиковани радови након одлуке Научног већа Института за физику у Београду о предлогу за стицање претходног научног звања:

- 3 рада у *Science of the Total Environment* (ИФ 9,8 за 1 рад, ИФ 5,589 за 1 рад и ИФ 3,816 за 1 рад)
- 1 рад у *Marine Pollution Bulletin* (ИФ 7,001)
- 1 рад у *Environmental Research* (ИФ 8,431)
- 1 рад у *Toxics* (ИФ 4,6)
- 2 рада у *Chemosphere* (ИФ 8,943 за 1 рад и ИФ 4,208 за 1 рад)
- 3 рада у *Atmosphere* (ИФ 3,110 за 1 рад, ИФ 2,9 за 1 рад и ИФ 2,9 за 1 рад)
- 1 рад у *Ecotoxicology and Environmental Safety* (ИФ 4,527)
- 1 рад у *Environmental Health: A Global Access Science Source* (ИФ 3,816)
- 4 рада у *Atmospheric Environment* (ИФ 3,708 за 2 рада и ИФ 3,459 за 1 рад и ИФ 3,226 за 1 рад) 123
- 2 рада у *Environmental Science and Pollution Research* (ИФ 2,76 за 1 рад и ИФ 2,741 за 1 рад)
- 1 рад у *Journal of Toxicology and Environmental Health*, Part A (ИФ 2,731)
- 2 рада у *Air Quality, Atmosphere and Health* (ИФ 2,662 за 1 рад и ИФ 1,804 за 1 рад) 1
- 1 рад у *International Journal of Environmental Science and Technology* (ИФ 2,037)
- 1 рад у *Plant Biosystems* (ИФ 1,39)
- 1 рад у *Acta Physiologiae Plantarum* (ИФ 1,563)
- 1 рад у *Journal of Environmental Science and Health, Part A* (ИФ 1,276)
- 1 рад у *Chemical Industry and Chemical Engineering Quarterly* (ИФ 0,533)

Укупан импакт-фактор радова кандидата износи 103,239, а у периоду након одлуке Научног већа Института за физику у Београду о предлогу за стицање претходног научног звања радова сумарни импакт фактор је 47,685. Часописи у којима објављује су цењени по свом угледу и водећи у његовим областима рада. Међу поменутим часописима посебно се истичу *Science of the Total Environment, Environmental Health, Environmental Research, Toxics, Ecotoxicology and Environmental Safety, Atmospheric Environment* и *Chemosphere*.

Додатни библиометријски показатељи према Упутству о начину писања извештаја о изборима у звања које је усвојио Матични научни одбор за физику приказани су у следећој табели:

	ИФ	**М**	**СНИП**
Укупно	47,685	61	9,85
Усредњено по чланку	5,961	7,625	1,231
Усредњено по аутору	6,072	7,843	1,264

4.1.4 Степен самосталности и степен учешћа у реализацији радова у научним центрима у земљи и иностранству

Од 28 објављених радова, кандидат је први аутор на 7 радова, други наведени аутор на 4 рада, трећи аутор на 3 рада и последњи аутор на 11 радова. На радовима који су објављени у периоду након одлуке Научног већа Института за физику о предлогу за стицање претходног научног звања, кандидат је први аутор на 1 раду и последњи аутор на 6 радова. Од 15 поглавља у монографијама од међународног значаја, кандидат је први аутор на 3, други наведени аутор на 4, трећи аутор на 2 и последњи аутор на 2. На поглављима објављеним у периоду након одлуке Научног већа Института за физику о предлогу за стицање претходног научног звања, кандидат је први аутор на 1 и последњи аутор на 4.

При изради поменутих публикација, кандидат је учествовао у осмишљавању и формулацији проблема, конструкцији релевантних прикупљању података, развоју теоријских метода за анализу добијених резултата, моделирању, обради података, визуелизацији и писању.

Током докторских студија кандидат је истраживао утицај атмосферског загађења на животну средину и здравље људи. Имплементирао је методу масене спектрометрије са трансфером протона за мерење концентрација широког спектра испарљивих органских једињења у амбијенталном ваздуху и у контролисаним, лабораторијским условима. Истраживање је било усмерено на идентификацију извора атмосферских аеросола и ИОЈ, анализу њихове динамике и просторне расподеле, као и на разумевање процеса и међусобних интеракција који утичу на њихово формирање, трансформацију и динамику у атмосфери. Поред тога, кандидат је активно учествовао у развоју новог концепта за прогнозу динамике доприноса извора загађујућих материја базираног на примени напредних алгоритама машинског учења.

Након завршетка докторских студија, кандидат је наставио своје истраживачке активности са фокусом на проучавање улоге загађења ваздуха у циклусу од извора емисије, преко атмосферских процеса и феномена који утичу на његову трансформацију и транспорт, до његовог утицаја на живи свет, здравље људи и стање животне средине. Ове активности могу се сврстати у три области:

- прикупљање података, које укључује експериментална мерења (концентрације загађујућих материја у амбијенталном ваздуху у реалним и симулираним мултифазним системима животне средине) и коришћење јавно доступних база података (загађујуће материје – EEA и US EPA, метеоролошки параметри – NOAA и ARL, морталитет – надлежне институције у Републици Србији, мобилност становништва – *Apple* и *Google*, мере државних институција – OxCGRT, пандемијска статистика – *Worldometer* и економски показатељи);
- моделирање података применом великог броја статистичких метода, метода машинског учења, метода оптимизације хиперпараметара алгоритама машинског учења и метода за објашњење модела машинског учења и
- развој експерименталних и/или статистичких метода за истраживање мултифазних система животне средине, транспорта загађења ваздуха и утицаја фактора животне средине на здравље људи и морталитет.

4.1.5 Награде

Сертификати о завршеним тренинзима на 3rd, 4th и 7th *Hands on PTR-MS* (2009, 2011 и 2019. године, Аустрија).

4.2 Ангажованост у формирању научних кадрова

Др Андреја Стојић је ментор:

1. 1 докторске дисертације у изради:
   1.1. Наташа Букумирић, *Процена утицаја измењених активности људи током пандемије Ковид-19 на расподеле испарљивих органских једињења у Београду*, Животна средина и одрживи развој, Универзитет Сингидунум (очекивана одбрана током 2025. године)
2. 1 мастер рада:
   2.1. Филип Алимпић, 2021. *Промене у концентрацијама испарљивих органских једињења пореклом из индустрије под утицајем пандемије Ковид-19 у Београду,* Животна средина и одрживи развој Универзитет Сингидунум.

Такође, учествовао још и у изради:

1. 2 докторске дисертације:
   1.1. Мирјана Перишић, 2016. година, *Примена хибридних рецепторских модела у анализи квалитета ваздуха и транспорта загађујућих материја у Београду,* Физички факултет Универзитета у Београду;
   1.2. Андреј Шоштарић, 2017. година, *Механизми уклањања лако испарљивих моноароматичних угљоводоника (BTEX) из амбијенталног ваздуха мокром депозицијом*, Хемијски факултет Универзитета у Београду),
2. 2 мастер рада:
   2.1. Ружица Шебек, 2017. година, *Сезонске варијације концентрација PM_{10} за Београд рачунате дисперзионим моделом*, Физички факултет Универзитета у Београду;
   2.2. Наташа Станојковић, 2019. година, *Климатске промене: могући утицај на здравље и морталитет у Новом Саду*, Животна средина и одрживи развој, Универзитет Сингидунум
3. 8 дипломских радова

3.1. Никола Петровић, 2008. Мониторинг испарљивих органских једињења у ваздуху, Физички факултет Универзитета у Београду.
3.2. Драгослав Ристић, 2010. Мерење испарљивих органских једињења спектрометром са трансфером протона, Физички факултет Универзитета у Београду,
3.3. Марија Тодоровић, 2012. Мерење испарљивих органских једињења уређајем PTR-MS – проблеми мерења у зависности од услова у реакционој комори, Факултет за физичку хемију Универзитета у Београду,
3.4. Вера Цукуђан, 2020. Утицај промена у активностима људи током пандемије Ковид-19 на квалитет ваздуха у Београду, Универзитет Сингидунум, Београд,
3.5. Биљана Стојковић, 2021. Утицај метана на климатски систем, Универзитет Сингидунум, Београд,
3.6. Елена Петровска, 2022. Угљенично-неутрална Србија до 2050: ризици и потенцијали кроз анализу тренутног стања и трендова у енергетском сектору, Универзитет Сингидунум, Београд,
3.7. Ђорђе Тодоровић, 2023. Утицај Ковид-19 шока на честично загађење у Републици Србији, Универзитет Сингидунум, Београд,
3.8. Павле Јовановић, 2023. Утицај Ковид-19 шока на стање квалитета ваздуха у Републици Србији, Универзитет Сингидунум, Београд.

У сарадњи са Регионалним центром за таленте Земун, кандидат је током 2012. године радио на изради експерименталних радова са ученицима који су учествовали на Републичком такмичењу младих талената за основне школе.

Током 2019. године кандидат је био ментор матурског рада *Примена метода машинског учења у физици животне средине* Лазара Златића, Математичка гимназија у Београду.

Кандидат је током школске 2016-2017. године водио пројекат студентске праксе *Истраживање квалитета ваздуха*, на коме су учествовала два студената треће године Физичког факултета у Београду.

Током 2019. године кандидат је учествовао у акредитацији, а потом је ангажован и као предавач на студијском програму *Животна средина и одрживи развој* Универзитета Сингидунум у Београду, на основним (3 предмета: Загађење и заштита ваздуха, Климатске промене и међународни односи и *Analytical methods of environmental quality*), мастер (2 предмета: Методологија научног истраживања и Савремене методе обраде података из животне средине) и докторским студијама (1 предмет: Напредне методе обраде података из животне средине).

Прилог:

- *Потврда Универзитета Сингидунум о менторствима и избору у наставно звање*

4.3 Нормирање броја коауторских радова, патената и техничких решења

Сви радови кандидата објављени након одлуке Научног већа Института за физику у Београду о предлогу за стицање претходног научног звања укључују резултате комплексних мерења и анализе података. Три од осам радова имају седам коаутора, тако да се рачунају са пуним бројем бодова у односу на број коаутора, док је за остале радове извршено нормирање.

4.4 Руковођење пројектима, потпројектима и пројектним задацима

Кандидат је руководио, односно руководи тимом из Института за физику у Београду на пројектима Фонда за науку Републике Србије ATLAS (2020-2022, бр. 6524105) и crAIRsis (2024-2027, бр. 7373).

Током 2020. године кандидат је руководио научном студијом спроведеном у оквиру апликативног пројекта *План квалитета ваздуха у агломерацији Београд*, које је финансирао Секретаријат за заштиту животне средине Града Београда.

У оквиру националног пројекта интердисциплинарних истраживања ИИИ 43007, под називом *Истраживање климатских промена и њиховог утицаја на животну средину – праћење утицаја, адаптација и ублажавање* руководи фазама истраживања које се односе на мерења и анализе ИОЈ и аеросола.

У периоду од 2019. до 2021. године, кандидат учествује на пројекту билатералне сарадње између Републике Србије и Републике Хрватске *Дуготрајна органохлорна једињења у мајчином млеку и њихов утицај на примарна оштећења ДНК у људским ћелијама*, на коме руководи активностима у вези са обрадом података.

Кандидат је током 2016/2017. године руководио фазама и активностима *Националног центра изузетних вредности за примену плазме у нанотехнологијама, биомедицини и екологији*, Института за физику у Београду.

Током 2018. године кандидат је био руководилац пројекта *Мапирање извора токсичних, мутагених и канцерогених испарљивих органских једињења на територији Града Београда*, финансираног од стране Зеленог фонда, Министарства заштите животне средине Републике Србије.

Прилог:

- *Потврда о пројектима crAIRsis* и *ATLAS*

4.5 Активност у научним и научно-стручним друштвима

Кандидат је рецензент за часописе (рецензије након избора у претходно звање):

- *Atmospheric Pollution Research* (2020. и 2024. година),
- *Environmental Pollution* (2022. и 2024. година),
- *Environmental Research* (2024. година),
- *Future Generation Computer Systems* (2021. година),
- *Science of the Total Environment* (2 рада 2021, 1 рад 2022. и 1 рад 2024. година),
- *Tropical Medicine and Infectious Disease* (2023. година),
- *Applied Sciences* (2023. година),
- *Sustainability* (2 рада 2022. године),
- *Atmosphere* (1 рад 2020, 1 рад 2021. и 2 рада 2022. године),
- *Electronics* (2022. година),
- *Remote Sensing* (4 рада 2021. и 1 рад 2022. године).

Кандидат је члан Асоцијације италијанских и српских научника и истраживача (AIS3), као и *topical advisory panel member* часописа *Atmosphere*.

Током 2022. године кандидат је био гостујући едитор у часопису *Atmosphere* (*Guest editor of special issue – Is our future up in the air? Odorous volatile organic compounds (VOCs) and greenhouse gas emissions*).

Кандидат је био члан научног одбора конференције *International scientific conference on information technology, computer science, and data science (Sinteza)* 2021, 2022, 2023. и 2024. године.

Кандидат је био члан организационог одбора међународне конференције *International conference of experimental and numerical investigations and new technologies* (CNN *Tech*) 2021. године.

Кандидат је учествовао у консултацијама у оквиру израде препоруке о етици вештачке интелигенције UNESCO 2020. године.

Прилог:

- *Сертификати издавача о рецензији радова*
- *Сертификат часописа Atmosphere – topical advisory panel member*
- *Сертификат часописа Atmosphere – гостујући едитор*
- *Синтеза – научни одбор*
- *CNN Tech – организациони одбор*
- *Позив за учешће у* консултацијама у оквиру израде препоруке о етици вештачке интелигенције UNESCO

4.6 Утицајност научних резултата

Утицај научних резултата кандидата је приказан у секцији 5.1 овог документа. Поред тога, списак свих публикација и цитата је дат у прилогу, на основу чега се такође може закључити да су радови кандидата јасно препознати у оквиру области опште и интердисциплинарне физике.

4.7 Конкретан допринос кандидата у реализацији радова у земљи и иностранству

Кандидат је значајно допринео сваком раду у чијој припреми је учествовао. Од 8 радова објављених у периоду након одлуке Научног већа Института за физику у Београду о предлогу за стицање претходног научног звања, сви радови су урађени у сарадњи с колегама из земље и иностранства. Кандидат је у овим радовима имао кључни допринос: на 1 раду је први аутор, док је на 6 радова последњи аутор. Током израде ових публикација, он је осмислио тему истраживања и радио на развоју одговарајућих мерних поставки и симулација, прикупљању и анализи релевантних података, развоју теоријских модела, метода и техника анализа проблема, писању радова, а такође је био у комуникацији с уредницима часописа при слању радова за објављивање.

У Институту за физику у Београду кандидат је увео иновативне методе у проучавање порекла, еволуције и утицаја загађујућих материја у атмосфери на животну средину и здравље људи. Методе су засноване на мерењима у реалном времену у реалним и симулираним (лабораторијским) условима, као и примени напредних статистичких метода и алгоритама вештачке интелигенције које укључују методе машинског учења (*machine learning* – ML), методе оптимизације хиперпараметара алгоритама машинског

учења (метахеуристике) и методе за објашњење модела машинског учења (*explainable artificial intelligence* – XAI) за обраду података. Знања и искуства које је стекао у теоријском моделирању и аналитичким методама у области опште и интердисциплинарне физике, хемије животне средине и науке о подацима успешно преноси млађим сарадницима у Лабораторији за физику животне средине и студентима кроз 5 предмета студијског програма *Животна средина и одрживи развој*, као и 1 предмета на свим техничким факултетима Универзитета Сингидунум, Београд.

4.8 Уводна предавања на конференцијама и друга предавања

Након претходног избора у звање, кандидат је одржао следеће уводно предавање по позиву на конференцији:

- **A. Stojić**
  Understanding pollution – contextual environmental data analysis
  Sinteza 2020, International scientific conference on information technology and data related research, Belgrade, October 17th, M32

Пре претходног избора у звање одржао је следеће уводно предавање по позиву на конференцији:

- **A. Stojić**
  Modeling particulate matter in urban areas: Experiences of the Institute of Physics Belgrade
  The 7th International WeBIOPATR, 1-3 October, 2019, Belgrade, Serbia, M32

Поред тога, одржао је и следећа предавања на међународним конференцијама:

- **A. Stojić**, M. Perišić, G. Jovanović, S. Stanišić, N. Stanić and T. Milićević
  Parsing environmental factors which shape particulate matter pollution using explainable artificial intelligence
  The 7th International WeBIOPATR, 1-3 October, 2019, Belgrade, Serbia, M34

- **A. Stojić** and S.S. Stojić
  Concentration weighted boundary layer hybrid receptor model for analyzing particulate matter altitude distribution
  The 6th International WeBIOPATR, 6-8 September, 2017, Belgrade, Serbia, M33

Прилог: Сертификат о одржаном предавању и програм конференције

5. ЕЛЕМЕНТИ ЗА КВАНТИТАТИВНУ ОЦЕНУ НАУЧНОГ ДОПРИНОСА КАНДИДАТА

Остварени резултати у периоду након одлуке Научног већа Института за физику о предлогу за стицање претходног научног звања дати су у табели. Према бази ISI *Web of Science*, радови кандидата укупно су цитирани 485 пута, док је број цитата без аутоцитата 408. Према бази *Scopus*, укупан број цитата је 623, док је број цитата без аутоцитата 506. Према подацима обе базе, Хиршов индекс радова кандидата је 13, односно 15.

Категорија	М бодова по раду	Број радова	Укупно М бодова	Нормирани број М бодова
M13	7	1	7	5
M14	4	3	12	12
M21a	10	3	30	25,476
M21	8	2	16	13,333
M22	5	3	15	14,167
M32	1,5	1	1,5	1,5
M33	1	10	10	10
M34	0,5	8	4	4
M42	5	1	5	5

Поређење оствареног броја М-бодова с минималним квантитативним условима за избор у звање научни саветник:

Минималан број М бодова		Услов	Остварено (нормирано)
Укупно		70	103,5 (90,476)
M10+M20+M31+M32+M33 +M41+M42		50	99,5 (86,476)
M11+M12+M21+M22+M23		35	61 (52,976)

6. СПИСАК РАДОВА КАНДИДАТА

6.1 Монографска студија/поглавље у књизи М11 или рад у тематском зборнику водећег међународног значаја (М13)

Радови објављени након претходног избора у звање

1. Stanišić, S., Perišić, M., Jovanović, G., Maletić, D., Vudragović, D., Vranić, A., **Stojić, A.**, 2021. What information on volatile organic compounds can be obtained from the data of a single measurement site through the use of artificial intelligence? In *Artificial Intelligence: Theory and Applications*, pp. 207-225. Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-030-72711-6_12

2. **Stojić, A.**, Mustać, B., Jovanović, G., Đinović Stojanović, J., Stanišić, S., Herceg Romanić, S., 2021. Patterns of PCB-138 bioaccumulation in small pelagic fish from the Eastern Mediterranean Sea using explainable machine learning prediction. In Artificial Intelligence: Theory and Applications, pp. 175-189. Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-030-72711-6_10

Радови објављени пре претходног избора у звање

3. Stanišić, S., **Stojić, A.**, 2020, Urban Air Pollution and Environmental Health. In: Leal Filho W., Azul A., Brandli L., Özuyar P., Wall T. (eds) Sustainable Cities and Communities. Encyclopedia of the UN Sustainable Development Goals. Springer, Cham, ISBN: 978-3-319-71061-7 https://doi.org/10.1007/978-3-319-71061-7 https://link.springer.com/referenceworkentry/10.1007/978-3-319-71061-7_120-1

4. **Stojić, A.**, Vuković, G., Perišić, M., Stanišić, S., Šoštarić, A., 2018. Urban air pollution: an insight into its complex aspects. In: A Closer Look at Urban Areas, Editor: Sahar Romero, Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0, pp. 69-123. http://www.novapublishers.org/catalog/product_info.php?products_id=65599

5. Stanišić, S., **Stojić, A.**, Prodanović, M., 2018. Health aspects of urban life. In: A Closer Look at Urban Areas, Editor: Sahar Romero, Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0, pp. 49-64. http://www.novapublishers.org/catalog/product_info.php?products_id=65599

6. Stanišić, S., **Stojić, A.**, Prodanović, M., 2018. Environmental concerns in Serbia – with specific regard to air-pollution and its effects on human health, Editor: Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0, pp. 215-229. https://novapublishers.com/shop/serbia-current-issues-and-challenges-in-the-areas-of-natural-resources-agriculture-and-environment/

7. Stanišić Stojić, S., Stanišić, N., **Stojić, A.**, 2016. Short- and long-term effects of urban air pollution on cardiopulmonary and malignant death rates. In: Air Pollution: Management Strategies, Environmental Impact and Health Risks, Editor: Gerald L. Burns, Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0. pp. 41-68. http://www.novapublishers.org/catalog/product_info.php?products_id=58708

8. **Stojić, A.**, Stanišić Stojić, S., Mijić, Z., Ilić, L., Tomašević, M., Todorović, M., Perišić, M., 2015. Comprehensive analysis of VOC emission sources in Belgrade urban area. In: Urban and Built Environments: Sustainable Development, Health Implications and

Challenges, Editor: Alexis Cohen, Nova Science Publishers, NY, USA, ISBN: 978-1-62417-735-4, pp. 55-88. https://novapublishers.com/shop/urban-and-built-environments-sustainable-development-health-implications-and-challenges/

9. Tomašević, M., Mijić, Z., Aničić, M., **Stojić, A.**, Perišić, M., Kuzmanoski, M., Todorović, M., Rajšić, S., 2013. Air quality study in Belgrade: particulate matter and volatile organic compounds as threats to human health. In: Air Pollution: Sources, Prevention and Health Effects, Editor: Rajat Sethi, Nova Science Publishers, NY, USA, ISBN: 978-1-62417-735-4, pp. 315-346. https://novapublishers.com/shop/environmental-and-agricultural-research-summaries-volume-1/

10. Aničić M., Mijić, Z., Kuzmanoski, M., **Stojić, A.**, Tomašević, M., Rajšić, S., Tasić, M., 2012. A study of airborne trace elements in Belgrade urban area: instrumental and active biomonitoring approach. In: Trace Elements: Environmental Sources, Geochemistry and Human Health, Editors: Diego Alejandro De Leon and Paloma Raquel Aragon, Nova Science Publishers, NY, USA, ISBN: 978-1-62081-401-7, pp.1-30. https://novapublishers.com/shop/trace-elements-environmental-sources-geochemistry-and-human-health/

6.2 Монографска студија/поглавље у књизи M12 или рад у тематском зборнику међународног значаја (M14)

Радови објављени након претходног избора у звање

1. Jovanović, G., Herceg Romanić, S., **Stojić, A.**, 2020, Advanced modeling of persistent organic pollutants (POPs) patterns in biomatrices. In: Leal Filho W., Azul A., Brandli L., Özuyar P., Wall T. (eds) Sustainable Cities and Communities. Encyclopedia of the UN Sustainable Development Goals. Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0, pp. 105-145. https://novapublishers.com/shop/advances-in-environmental-research-volume-77/

2. Jovanović, G., Perišić, M., Stanišić, S., and **Stojić, A.**, 2022. Explaining xylene wet deposition using artificial intelligence. In Horizons in World Physics, Volume 307. Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0, pp. 1-30. https://novapublishers.com/shop/horizons-in-world-physics-volume-307/

3. Stanišić, S., Jovanović, G. Perišić, M., Herceg Romanić, S., Milićević, T. and **Stojić, A.**, 2022. Explaining the environmental fate of PAHs in indoor and outdoor environments by the use of artificial intelligence. In Polycyclic Aromatic Hydrocarbons: Sources, Exposure and Health Effects. Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0, pp. 1-36. https://novapublishers.com/shop/polycyclic-aromatic-hydrocarbons-sources-exposure-and-health-effects/

Радови објављени пре претходног избора у звање

4. Mijić, Z., **Stojić, A.**, Perišić, M., Rajšić, S., Tasić M., 2012. Statistical character and transport pathways of atmospheric aerosols in Belgrade. In: Air Quality – New Perspective, Edited by Gustavo Lopez Badilla, Benjamin Valdez and Michael Schorr,

Published by InTech, ISBN: 978-953-51-0674-6, pp. 199-226.
https://www.intechopen.com/books/air-quality-new-perspective/statistical-character-and-transport-pathways-of-atmospheric-aerosols-in-belgrade

5. Mijić, Z., Rajšić, S., Žekić, A., Perišić, M., **Stojić, A.**, Tasić M., 2010. Characteristics and application of receptor models to the atmospheric aerosols research, Book chapter in Air quality edited by Ashok Kumar, ISBN 978-953-307-131-2, pp. 143-167.
https://www.intechopen.com/books/air-quality/characteristics-and-application-of-receptor-models-to-the-atmospheric-aerosols-research

6.3 Радови у међународним часописима изузетних вредности (M21a)

Радови објављени након претходног избора у звање

1. Bacanin, N., Perisic, M., Jovanovic, G., Damaševičius, R., Stanisic, S., Simic, V., Zivkovic, M. and **Stojić, A.**, 2024. The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere. *Science of The Total Environment*, *929*, p.172195.
https://doi.org/10.1016/j.scitotenv.2024.172195
(ИФ: 9,8)

2. Herceg Romanić, S, Jovanović, G., Mustać, B., Stojanović-Đinović, J., **Stojić, A.**, Čadež, T., Popović, A., 2021. Fatty acids, persistent organic pollutants, and trace elements in small pelagic fish from the eastern Mediterranean Sea, Marine Pollution Bulletin, 170, p. 112654.
https://doi.org/10.1016/j.marpolbul.2021.112654
(ИФ: 5,553)

3. Stanišić, S., Perišić, M., Jovanović, G., Milićević, T., Romanić, S.H., Jovanović, A., Šoštarić, A., Udovičić, V., **Stojić, A.**, 2021. The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources. Environmental Research, 193, p.110520.
https://doi.org/10.1016/j.envres.2020.110520
(ИФ: 6,498)

Радови објављени пре претходног избора у звање

4. Stanišić Stojić, S., Stanišić, N., **Stojić, A.**, 2016. Temperature-related mortality estimates after accounting for the cumulative effects of air pollution in an urban area. Environmental Health. 15(1), 73.
https://doi.org/10.1186/s12940-016-0164-6
(ИФ: 3,816)
5. **Stojić, A.**, Maletić, D., Stojić, S. S., Mijić, Z., Šoštarić, A., 2015. Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. Science of the Total Environment, 521, 19-26.
https://doi.org/10.1016/j.scitotenv.2015.03.098
(ИФ: 3,976)

6.4 Радови у врхунским међународним часописима (M21)

Радови објављени након претходног избора у звање

1. Jovanović, G., Perišić, M., Bačanin, N., Živković, M., Stanišić, S., Štrumberger, I., Alimpić, F., **Stojić, A.**, 2023. Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing PAHs Environmental Fate. *Toxics*. 11(4):394. https://doi.org/10.3390/toxics11040394
   (ИФ: 4.472)

2. **Stojić, A.**, Jovanović, G., Stanišić, S., Romanić, S.H., Šoštarić, A., Udovičić, V., Perišić, M. and Milićević, T., 2022. The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels. Chemosphere, 289, p.133154. https://doi.org/10.1016/j.chemosphere.2021.133154
   (ИФ: 8.943)

Радови објављени пре претходног избора у звање

3. **Stojić, A.**, Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Science of The Total Environment, 653, 140–147. https://doi.org/10.1016/j.scitotenv.2018.10.368
   (ИФ: 5,589)

4. Jovanović, G., Herceg Romanić, S., **Stojić, A.**, Klinčić, D., Matek Sarić, M., Grzunov Letinić, J., Popović, A., 2019. Introducing of modeling techniques in the research of POPs in breast milk – A pilot study, Ecotoxicology and Environmental Safety, 172, 341-347. https://doi.org/10.1016/j.ecoenv.2019.01.087
   (ИФ: 4,527)

5. **Stojić, A.**, Stanišić Stojić, S., 2017. The innovative concept of three-dimensional hybrid receptor modeling. Atmospheric Environment, 164, 216-223. https://doi.org/10.1016/j.atmosenv.2017.06.009
   (ИФ: 3,708)

6. Šoštarić, A., Stojić, S. S., Vuković, G., Mijić, Z., **Stojić, A.**, Gržetić, I., 2017. Rainwater capacities for BTEX scavenging from ambient air. Atmospheric Environment, 168, 46-54. https://doi.org/10.1016/j.atmosenv.2017.08.045
   (ИФ: 3,708)

7. Perišić, M., Rajšić, S., Šoštarić, A., Mijić, Z., **Stojić, A.**, 2017. Levels of PM_{10} bound species in Belgrade, Serbia: spatio-temporal distributions and related human health risk estimation. Air Quality, Atmosphere and Health, 10(1), 93-103. https://doi.org/10.1007/s11869-016-0411-6
   (ИФ: 2,662)

8. Stanišić Stojić, S., Stanišić, N., **Stojić, A.**, Šoštarić, A., 2016. Single and combined effects of air pollutants on circulatory and respiratory system-related mortality in Belgrade, Serbia. Journal of Toxicology and Environmental Health, Part A, 79(1), 17-27.

https://doi.org/10.1080/15287394.2015.1101407
(ИФ: 2,731)

9. **Stojić, A.**, Stanišić Stojić, S., Reljin, I., Čabarkapa, M., Šoštarić, A., Perišić, M., Mijić, Z., 2016. Comprehensive analysis of PM_{10} in Belgrade urban area on the basis of long-term measurements. Environmental Science and Pollution Research, 23(11), 10722-10732.
https://doi.org/10.1007/s11356-016-6266-4
(ИФ: 2,741)

10. Šoštarić, A., **Stojić, A.**, Stojić, S. S., Gržetić, I., 2016. Quantification and mechanisms of BTEX distribution between aqueous and gaseous phase in a dynamic system. Chemosphere, 144, 721-727.
https://doi.org/10.1016/j.chemosphere.2015.09.042
(ИФ: 4,208)**Stojić, A.**, Stojić, S. S., Mijić, Z., Šoštarić, A., Rajšić, S., 2015. Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling. Atmospheric Environment, 106, 71-79.
https://doi.org/10.1016/j.atmosenv.2015.01.071
(ИФ: 3,459)

11. **Stojić, A.**, Stanišić Stojić, S., Šoštarić, A., Ilić, L., Mijić Z., Rajšić S., 2015. Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling. Environmental Science and Pollution Research, 22(17), 13137-13152.
https://doi.org/10.1007/s11356-015-4540-5
(ИФ: 2,76)

12. Mijić, Z., **Stojić, A.**, Perišić, M., Rajšić, S., Tasić, M., Radenković, M., Joksić, J., 2010. Seasonal variability and source apportionment of metals in the atmospheric deposition in Belgrade. Atmospheric Environment, 44(30), 3630-3637.
https://doi.org/10.1016/j.atmosenv.2010.06.045
(ИФ: 3,226)

6.5 Радови у истакнутим међународним часописима (М22)

Радови објављени након претходног избора у звање

1. N. Radić, M. Perišić, G. Jovanović, T. Bezdan, S. Stanišić, N. Stanić, **A. Stojić**, 2025. An AI-based framework for characterizing the atmospheric fate of air pollutants within diverse environmental settings Atmosphere, 16(2), 231
https://doi.org/10.3390/atmos16020231
(ИФ: 2,900)
2. Jovanović, G., Perišić, M., Bezdan, T., Stanišić, S., Radusin, K., Popović, A. and **Stojić, A.**, 2024. The PM 2.5-Bound Polycyclic Aromatic Hydrocarbon Behavior in Indoor and Outdoor Environments, Part III: Role of Environmental Settings in Elevating Indoor Concentrations of Benzo (a) pyrene. Atmosphere, 15(12).
https://doi.org/10.3390/atmos15121520
(ИФ: 2,900)
3. Jovanović, L., Jovanović, G., Perišić, M., Alimpić, F., Stanišić, S., Bačanin, N., Živković, M., **Stojić, A.**, 2023. The explainable potential of coupling metaheuristics-optimized-XGBoost and SHAP in revealing VOC's environmental fate, *Atmosphere*,

14(1), p.109.
https://doi.org/10.3390/atmos14010109
(ИФ: 3,110)

Радови објављени пре претходног избора у звање

4. Perišić M., Maletić D., Stanišić Stojić S., Rajšić S., **Stojić A.**, 2017. Forecasting hourly particulate matter concentrations based on the advanced multivariate methods. International Journal of Environmental Science and Technology, 14(5), 1047-1054. https://doi.org/10.1007/s13762-016-1208-8
(ИФ: 2,037)

5. Stanišić Stojić, S., Ignjatović, L., Popov, S., Škrivanj, S., Đorđević, A., **Stojić, A.**, 2016. Heavy metal accumulation in wheat and barley: The effects of soil presence and liquid manure amendment. Plant Biosystems, 150(1), 104-110.
https://doi.org/10.1080/11263504.2014.976288
(ИФ: 1,39)

6. Dmitrović, S., Perišić, M., **Stojić, A.**, Živković, S., Boljević, J., Živković, J. N., Mišić, D., 2015. Essential oils of two Nepeta species inhibit growth and induce oxidative stress in ragweed (Ambrosia artemisiifolia L.) shoots in vitro. Acta Physiologiae Plantarum, 37(3), 1-15.
https://doi.org/10.1007/s11738-015-1810-2
(ИФ: 1,563)

7. Perišić, M., **Stojić, A.**, Stojić, S. S., Šoštarić, A., Mijić, Z., Rajšić, S., 2014. Estimation of required PM_{10} emission source reduction on the basis of a 10-year period data. Air Quality, Atmosphere and Health, 8(4), 379-389.
https://doi.org/10.1007/s11869-014-0292-5
(ИФ: 1,804)

6.6 Радови у међународним часописима (М23)

Радови објављени пре претходног избора у звање

1. Todorović, M., Perišić, M., Kuzmanoski, M., **Stojić, A.**, Šoštarić, A., Mijić, Z., Rajšić, S., 2015. Assessment of PM_{10} pollution level and required source emission reduction in Belgrade area. Journal of Environmental Science and Health, Part A, 50(13), 1351-1359.
https://doi.org/10.1080/10934529.2015.1059110
(ИФ: 1,276)

2. Mijić, Z., **Stojić, A.**, Perišić, M., Rajšić, S., Tasić, M., 2012. Receptor modeling studies for the characterization of PM_{10} pollution sources in Belgrade. Chemical Industry and Chemical Engineering Quarterly, 18(4-2), 623-634. doi: 10.2298/CICEQ120104108M
http://www.ache.org.rs/CICEQ/2012/no04-II.html
(ИФ: 0,533)

6.7 Радови у националним часописима међународног значаја (М24)

Радови објављени пре претходног избора у звање

1. Stanišić Stojić, S., Stanišić, N., **Stojić, A.**, Džamić, V., 2016. Seasonal mortality variations of cardiovascular, respiratory and malignant diseases in the City of Belgrade. Stanovništvo, 54(1), 83-104. https://www.idn.org.rs/ojs3/stanovnistvo/index.php/STNV/article/download/74/65/

6.8 Предавања по позиву с међународних скупова штампана у изводу (М32)

Радови објављени након претходног избора у звање

1. **Stojić, A.**, 2020. Understanding pollution – contextual environmental data analysis, Sinteza 2020, International scientific conference on information technology and data related research, Belgrade, October 17th.

Радови објављени пре претходног избора у звање

2. **Stojić**, **A.**, 2019, Modeling particulate matter in urban areas: Experiences of the Institute of Physics Belgrade, The 7th International WeBIOPATR, 1-3 October, Belgrade, Serbia, pp. 67.

6.9 Саопштења с међународних скупова штампана у целини (М33)

Радови објављени након претходног избора у звање

1. Bezdan, T., Perišić, M., Jovanović, G., Bačanin-Džakula, N., **Stojić, A.**, 2024. Artificial intelligence-based framework for analyzing crises-caused air pollution, Sinteza 2024 - International Scientific Conference on Information Technology and Data Related Research, Singidunum University, 16th May, pp. 281-280.
2. Jovanović, G., Perišić, M., Stanišić, S., Bačanin-Džakula, N., **Stojić, A.**, 2023. Revealing toluene behaviour in the atmosphere based on coupling of metaheuristics, XGBoost, and SHAP, Sinteza 2023 - International Scientific Conference on Information Technology and Data Related Research, Singidunum University, 27th May, pp. 17-22.
3. Alimpić, F., Perišić, M., Stanišić, S., Jovanović, G., **Stojić, A.**, 2021. Evolution of industry-related volatile organic compound levels affected by COVID-19 lockdown in Belgrade, Sinteza 2021 - International Scientific Conference on Information Technology and Data Related Research, Singidunum University, 25th Jun, pp. 1-4.
4. Bukumirić, N., Perišić, M., Stanišić, S., Jovanović, G., **Stojić, A.**, 2021. The influence of COVID-19 lockdown on BTEX distributions in Belgrade, Sinteza 2021 - International Scientific Conference on Information Technology and Data Related Research, Singidunum University, 25th Jun, pp. 72-76.
5. Jovanović, G., Stanišić, S., Perišić, M., **Stojić, A.**, 2021. Structural characteristics of particulate matter time series observed in an urban environment, Sinteza 2021 - International Scientific Conference on Information Technology and Data Related Research, Singidunum University, 25th Jun, pp. 98-101.

6. Stanišić, S., Perišić, M., Jovanović, G., **Stojić, A.**, 2021. Receptor oriented modeling for revealing air pollution emission sources affecting an urban area, Sinteza 2021 - International Scientific Conference on Information Technology and Data Related Research, Singidunum University, 25th Jun, pp. 94-97.

7. Perišić, M., Jovanović, G., Stanišić, S., Šoštarić, A., **Stojić, A.**, 2021. Meteorological factors governing particulate matter distribution in an urban environment, Sinteza 2021 - International Scientific Conference on Information Technology and Data Related Research, Singidunum University, 25th Jun, pp. 89-93.

8. **Stojić, A.**, Mustać, B., Jovanović G., 2020. Explainable machine learning prediction of PCB-138 patterns in edible fish from Croatian Adriatic, Sinteza 2020, International scientific conference on information technology and data related research, Belgrade, October 17th, pp. 23-28.

9. **Stojić, A**., Matek Sarić, M., Herceg Romanić, S., 2020. Shapley additive explanations of indicator PCB-138 distribution in breast milk, Sinteza 2020, International scientific conference on information technology and data related research, Belgrade, October 17th, pp. 35-40.

10. Stanišić, S., Perišić, M., **Stojić, A.**, 2020. The use of innovative methodology for the characterization of benzene, toluene, ethylbenzene and xylene sources in the Belgrade area, Sinteza 2020, International scientific conference on information technology and data related research, Belgrade, October 17th, pp. 41-45.

Радови пре претходног избора у звање

11. **Stojić, A.**, Vuković, G., Stanišić, S., Udovičić, V., Stanić, N, Šoštarić, A. 2019. Explainable machine learning prediction of VOC in an university building microenvironment, 8th International PTR-MS Conference, February 3-8, Innsbruck, Austria, pp. 267-271.

12. **Stojić, A.**, Vuković, G., Stanišić, S., Ćućuz, V., Trifunović, D., Udovičić, V., Šoštarić, A. 2019. Multifractality of isoprene temporal dynamics in outdoor and indoor university environment, 8th International PTR-MS Conference, February 3-8, Innsbruck, Austria, pp. 271-275.

13. Mijić, Z., Perišić, I., Ilić, L., **Stojić, A.**, Kuzmanoski, M. 2017. Air mass transport over Balkan region identified by atmospheric modeling and aerosol lidar technique, 49th International October Conference on Mining and Metallurgy, October 18-21, Bor Lake, Serbia, pp. 69-72.

14. **Stojić, A.**, Stanišić Stojić, S., 2017. Concentration weighted boundary layer hybrid receptor model for analyzing particulate matter altitude distribution. 6th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, September 6-8, Belgrade, Serbia, pp. 163-166.

15. **Stojić, A.**, Stanišić Stojić, S., Perišić, M., Mijić, Z., 2017. Multiscale multifractal analysis of nonlinearity in particulate matter time series. 6th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, September 6-8, Belgrade, Serbia, pp. 114-117.

16. Perišić, M., Vuković, G., Mijić, Z., Šoštarić, A., **Stojić, A.**, 2017. Relative importance of gaseous pollutants and aerosol constituents for identification of PM_{10} sources of

variability. 6th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, September 6-8, Belgrade, Serbia, pp. 109-112.

17. Stanišić Stojić, S., **Stojić, A.**, Perišić, M., 2016. Relationship between isoprene, related gaseous pollutants and meteorological factors in an urban area. 13th International Conference on Fundamental and Applied Aspects of Physical Chemistry, September 26-30, Belgrade, Serbia, Vol. II, pp. 711-714.

18. Perišić, M., **Stojić, A.**, Stanišić Stojić, S., 2016. Impact of remote sources on chromium concentrations in Belgrade and the related health risk. 13th International Conference on Fundamental and Applied Aspects of Physical Chemistry, September 26-30, Belgrade, Serbia, Proceedings Vol. II, pp. 735-738.Stanišić Stojić, S., Stanišić, N., Šoštarić, A., **Stojić, A.**, Mladenović, S., 2015. The association between short term exposure to PM_{10} and soot and circulatory system related mortality in Belgrade area. 5th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, October 14-16, Belgrade, Serbia, pp. 211-216.

19. Stanišić Stojić, S., Stanišić, N., Šoštarić, A., **Stojić, A.**, Mladenović, S., 2015. The association between short term PM_{10} exposure and mortality caused by respiratory system diseases in Belgrade area. 5th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, October 14-16, Belgrade, Serbia, pp. 191-195.

20. Perišić, M., **Stojić, A.**, Todorović, M., Mijić, Z., Šoštarić, A., 2015. Transport contribution to $PM_{2.5}$ mass concentrations in Belgrade sub urban area. 5th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, October 14-16, Belgrade, Serbia, pp. 99-102.

21. Mijić, Z., Perišić, M., **Stojić, A.**, Kuzmanoski, M., Ilić, L., 2015. Estimation of atmospheric aerosol transport by ground based remote sensing and modeling. XIX International Eco-Conference 2015, September 23 25, Novi Sad, Serbia, pp. 375-382.

22. **Stojić, A.**, Stanišić Stojić, S., Šoštarić, A., Mijić, Z., Perišić M., Rajšić, S., 2014. The contribution of chemical industry to ambient VOC levels in Belgrade. 12th International Conference on Fundamental and Applied Aspects of Physical Chemistry, September 22-26, 2014, Belgrade, Serbia, pp. 949-952.

23. Šoštarić, A., Perišić, M., **Stojić, A.**, Mijić, Z., Rajšić, S., 2014. Dynamics of gaseous pollutants in Belgrade urban area. 12th International Conference on Fundamental and Applied Aspects of Physical Chemistry, September 22-26, Belgrade, Serbia, Vol. I, pp. 953-956.

24. Todorović, M., Perišić, M., **Stojić, A.**, Rajšić, S., 2014. Source apportionment study in Belgrade urban area. 12th International Conference on Fundamental and Applied Aspects of Physical Chemistry, September 22-26, Belgrade, Serbia, Vol. I, pp. 929-932.

25. Šoštarić, A., **Stojić, A.**, Stanišić Stojić, S., Mijić, Z., 2014. Traffic-related VOC dynamics in Belgrade urban area. 12th International Conference on Fundamental and Applied Aspects of Physical Chemistry, September 22-26, Belgrade, Serbia, Vol. III, pp. 953-956.

26. Perišić, M., **Stojić, A.**, Mijić, Z., Todorović, M., Rajšić, S., 2013. Source apportionment of ambient VOCs in Belgrade semi-urban area. 6th International Conference on Proton Transfer Reaction Mass Spectrometry and Its Application, February 3-8, Innsbruck, Austria, pp. 204-208.

27. Perišić, M., Mijić, Z., **Stojić, A.**, 2013. Frequency analysis of PM_{10} time series and assessing source reduction for air quality compliance in Serbia. 4th WeBIOPATR Workshop Conference, October 2-6, Belgrade, Serbia, pp. 64 68.

28. Šoštarić, A., Perišić, M., **Stojić, A.**, Mijić, Z., Rajšić, S., Tasić, M., 2013. The influence of air mass origin and potential source contributions on PM_{10} in Belgrade. 4th WeBIOPATR Workshop Conference, October 2-6, Belgrade, Serbia, pp.39-43.

29. **Stojić, A.**, Perišić, M., Mijić, Z., Rajšić, S., 2011. Ambient VOCs measurements in winter: Belgrade semi-urban area. 5th International Conference on Proton Transfer Reaction Mass Spectrometry and Its Application, January 26-February 2, Innsbruck, Austria, pp. 248-251.

30. Perišić, M., **Stojić, A.**, Rajšić, S., Mijić, Z., 2010. Assessment of VOCs concentrations in Belgrade semi-urban area. 10th International Conference of Fundamental and Applied aspects of Physical Chemistry, September 21-24, Belgrade, Serbia, pp. 579-581.

31. **Stojić, A.**, Rajšić, S., Perišić, M., Mijić, Z., Tasić, M., 2009. Assessment of ambient VOCs levels in Belgrade semiurban area, 4th International Conference on Proton Transfer Reaction Mass Spectrometry and its Applications, February 16-21, Obergurgl, Austria, pp. 289-293.

32. Nestorović, J., Mišić, D., **Stojić, A.**, Perišić, M., Živković, S., Šiler, B., Aničić, M., Malović, G., Grubišić, D., 2009. *In vitro* selection of nepetalactone-rich genotypes of *Nepeta rtanjensis* by using HPCL and PTR-MS. 4th International Conference on Proton Transfer Reaction Mass Spectrometry and its Applications, February 16-21, Obergurgl, Austria, pp. 263-267.

33. Tasić, M., Mijić, Z., Rajšić, S., **Stojić, A.**, Radenković, M., Joksić, J., 2009. Source apportionment of atmospheric bulk deposition in the Belgrade urban area using positive matrix factorization. In Journal of Physics: Conference Series, IOP Publishing, April, Vol. 162, No. 1, pp. 012-018.

6.10 Саопштења с међународних скупова штампана у изводу (M34)

Радови објављени након претходног избора у звање

1. Perišić, M., **Stojić, A.**, Jovanović, G., 2023. The impact of the covid-19 pandemic on particulate air pollution in the republic of Serbia, WeBIOPATR WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, 28th November to 1st December, pp96.
2. **Stojić, A.**, Perišić, M., Jovanović, G., Stanišić, S., 2022. Artificial intelligence in revealing air pollution related processe. SICAA: 1st Serbian International Conference on Applied Artificial Intelligence, May 19-20, Kragujevac, Serbia.

3. Jovanović, G., Stanišić, S., Perišić M., Šoštarić, A. and **Stojić, A.**, 2021. Key factors governing particulate matter environmental fate in an urban environment, 8th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management, Decembar 1st, Belgrade, Serbia.

4. Perišić, M., **Stojić, A.**, Jovanović, G., Šoštarić, A., Maletić, D., Vudragović, D., Stanišić, S., 2021. The potential for forecasting the particulate matter levels in complex urban environment, CNN TECH Conference 2021, International conference of

experimental and numerical investigations of new technologies, 30th Jun-2nd July, Zlatibor, Serbia, pp 39.

5. **Stojić, A.**, Jovanović, G., Stanišić, S., Šoštarić, A., Vranić, A., Mitrović Dankulov, M., Perišić, M., 2021. The impact of humidity and temperature on particulate matter environmental fate, CNN TECH Conference 2021, International conference of experimental and numerical investigations of new technologies, 30th Jun-2nd July, Zlatibor, Serbia, pp 40.
6. Stanišić, S., Perišić, M., **Stojić, A.**, Šoštarić, A., Vudragović, D., Maletić, D., Perišić, M., 2021. The impact of gaseous pollutants on particulate matter distribution, CNN TECH Conference 2021, International conference of experimental and numerical investigations of new technologies, 30th Jun-2nd July, Zlatibor, Serbia, pp 41.
7. Stupar, N., Vranic, A., **Stojić, A.**, Vukovic, G., Vudragovic, D., Maletic, D., Mitrovic Dankulov, M., , 2021. Spatio-temporal analysis of mobility patterns in the city of Belgrade, The impact of gaseous pollutants on particulate matter distribution, CNN TECH Conference 2021, International conference of experimental and numerical investigations of new technologies, 30th Jun-2nd July, Zlatibor, Serbia, pp 42.
8. Jovanović, G., Stanišić, S., Perišić, M., Šoštarić, A., Mitrović Dankulov, M., Vranić, A., **Stojić, A.**, 2021. Environmental factors governing particulate matter distribution in an urban environment, CNN TECH Conference 2021, International conference of experimental and numerical investigations of new technologies, 30th Jun-2nd July, Zlatibor, Serbia, pp 43.

Радови објављени пре претходног избора у звање

1. **Stojić, A.**, Perišić, M., Jovanović, G., Stanišić, S., Stanić, N., Milićević, T., 2019, Parsing environmental factors which shape particulate matter pollution using explainable artificial intelligence, The 7th International WeBIOPATR 1-3 October, 2019, Belgrade, Serbia, pp. 34.
2. Perišić, M., **Stojić, A.**, Jovanović, G., Stanišić, S., 2019, Receptor oriented modeling of urban particulate air pollution: source characterization and spatial distribution, The 7th International WeBIOPATR 1-3 October, 2019, Belgrade, Serbia, pp. 75.
3. Jovanović, G., **Stojić, A.**, Perišić, M., Stanišić, S., Stanić, N., Milićević, T., 2019, Explainable relations of particulate matter and environmental factors in an urban area, The 7th International WeBIOPATR 1-3 October, 2019, Belgrade, Serbia, pp. 94.
4. Dmitrović, S., Perišić, M., **Stojić, A.**, Živković, S., Boljević, J., Nestorović Živković, J., Aničić, N., Ristić, M., Mišić, D., 2015. The oxidative stress in Ambrosia artemisiifolia L. shoots grown in vitro induced by *Nepeta rtanjensis* and *N. cataria* essential oils. 2nd International Conference on Plant Biology (21st Symposium of the Serbian Plant Physiology Society) & COST Action FA1106 Qualityfruit Workshop, Serbian Plant Physiology Society, June 17-20, Petnica, Serbia, pp. 159.
5. Škorić, M., Todorović, S., Ristić, M., Soković, M., Glamočlija, J., Živković, S., **Stojić, A.**, Puač, N., Kanellis, A. K., 2013. In vitro culture of Cistus creticus subsp. creticus-a source of biological active compounds. 11th International Meeting on Biosynthesis, Function and Biotechnology of Isoprenoids in Terrestrial and Marine Organisms (TERPNET 2013), June 1-5, Thessaloniki, Greece, pp. 132.

6. Škorić, M., Nestorović-Živković, J., Ristić, M., **Stojić, A.**, Puač, N., Kanellis, A.K., Todorović, S., 2013. PTR MS and GC/MS analysis of volatile compounds in shoot cultures of Cistus creticus subsp. creticus. 1st International Conference on Plant Biology (20th Symposium of the Serbian Plant Physiology Society), Serbian Plant Physiology Society, July 4-7, Subotica, Serbia, pp 107.

7. **Stojić, A.**, Perišić, M., Todorović, M., Nikitović, Ž., Jotić, A., Lalić, N., Petrović, Z.Lj., 2013. Application of PTR-MS measurements of volatile organic compounds (VOC) in medical science. 15th annual conference of YUCOMAT, September 2-6, Herceg Novi, Montenegro, pp. 68.

8. **Stojić, A.**, Mijić, Z., Perišić, M., Rajšić, S., Tasić, M., 2011. Ambient VOCs measurement in Belgrade semi-urban area: winter case study. 16th European conference on analytical chemistry Challenges in modern analytical chemistry, EUROanalysis, September 11-15, Belgrade, Serbia, pp. 102.

9. Mijić, Z., Kuzmanoski, M., **Stojić, A.**, Žekić, A. Rajšić, S., Tasić, M., 2011. Investigation of regional transport and health risk effects of metals in $PM_{2.5}$ air particulate matter in Belgrade. 3rd International WeBIOPATR Workshop & Conference, November 15-17, Belgrade, Serbia.

10. Mijić, Z., Tasić, M., Rajšić, S., **Stojić, A.**, 2011. Receptor modeling studies for the characterization of PM_{10} pollution sources in Belgrade, Proceedings of the 3rd International WeBIOPATR Workshop & Conference, November 15-17, Belgrade, Serbia.

11. Perišić, M., **Stojić, A.**, Mijić, Z., Rajšić, S., 2010. Source apportionment of volatile organic compounds in Belgrade semi-urban area. 11th European Meeting on Environmental Chemistry EMEC 11, December 8-11, Portorož, Slovenia, pp. 232.

12. **Stojić, A.**, Perišić, M., Mijić, Z., Rajšić, S., Ristić, D., 2010. Ambient VOCs measurement In Belgrade semi urban area using Proton Transfer Reaction Mass Spectrometer, 1st Center of Excellence for Food Safety and Emerging Risk (CEFSER) Workshop "Regional perspectives in food safety", 12th Danube-Kris-Mures-Tisa (DKMT) Euroregion Conference on Food, Environment and Health, Faculty of Technology, University of Novi Sad, September 14-15, Novi Sad, Serbia, CD Book of Abstracts.

13. **Stojić, A.**, Perišić, M., Mijić, Z., Rajšić, S., 2010. Proton Transfer Reaction Mass Spectrometry: ambient air VOCs measurement in Belgrade semi-urban area, 20th ESCAMPIG, July, Novi Sad, Serbia.

14. Nestorović, J., Mišić, D., Šiler, B., Živkovic, S., Malović, G., Perišić, M., **Stojić, A.**, Grubišić, D., 2010. Application of PTR-MS in detection of volatile compounds: in vitro culture of three nepeta species, 20th ESCAMPIG, July, Novi Sad, Serbia.

15. Nestorović, J., Mišić, D., Šiler, B., Živkovic, S., **Stojić, A.**, Perišić, M., Grubišić, D., 2009. PTR-MS detection of nepetalactone in shoot cultures of three Nepeta species grown under different carbohydrate source. New research in biotechnology, 2nd International Symposium, November 19-20, Bucharest, Romania, pp. 138.

16. **Stojić, A.**, Nešić, M., Mijić, Z., Novaković, V., Rajšić, S., Tasić, M., 2008. Heavy metal concentrations in streat dust and soils adjacent to roads in Belgrade, Serbia. 9th Highway and Urban Environment Symposium, June 9-11, Madrid, Spain, pp. 87.

17. Nešić, M., **Stojić, A.**, Mijić, Z., Novaković, V., Rajšić, S., 2007. First results of outdoor and indoor VOCs measurements using PTR-MS in Belgrade, Serbia, 8th European Meeting on Environmental Chemistry (EMEC8), Book of abstracts, December 5-8, Inverness, Scotland, pp. 37.

6.11 Монографија националног значаја (M42)

Радови објављени након претходног избора у звање

1. Perišić, M., Jovanović, G., **Stojić A.**, 2024. Kompleksnost životne sredine integralni pristup razumevanju fenomena. ISBN: 978-86-7912-826-3, Univerzitet Singidunum. https://singipedia.singidunum.ac.rs/izdanje/44592-kompleksnost-zivotne-sredine-integralni-pristup-razumevanju-fenomena

6.12 Рад у истакнутом националном часопису (M53)

Радови објављени пре претходног избора у звање

1. Mijić, Z., Tasić, M., Rajšić, S., **Stojić, A.**, 2012. Primena hibridnih receptorskih modela za ispitivanje transporta PM_{10} čestica na područje Beograda, Glasnik Hemičara, Tehnologa i ekologa Republike Srpske, 4(7), 41-48.

6.13 Саопштење са скупа националног значаја штампано у целини (M63)

Радови објављени пре претходног избора у звање

1. Perišić, M., **Stojić, A.**, Todorović, M., Mijić, Z., Rajšić, S., 2013. Analiza dinamike i transporta CO, NO_x i SO_2 u urbanoj sredini Beograda. XII Kongres fizičara Srbije, April 28-May 2, Vrnjačka Banja, Srbija, str. 444-447.
2. **Stojić, A.**, Perišić, M., Mijić, Z., Todorovic, M., Rajšić, S., 2013. Određivanje izvora emisije isparljivih organskih jedinjenja u Beogradu. XII Kongres fizičara Srbije, April 28-May 2, Vrnjačka Banja, Srbija, str. 453-456.
3. **Stojić, A.**, Perišić, M., Petrović, N., 2008. Merenje isparljivih organskih jedinjenja u realnom vremenu masenim spektrometrom (PTR-MS) Naučnostrucni skup sa međunarodnim učešćem, Zbornik radova, Institut zaštite, ekologije i informatike, Novembar 14-15, Banja Luka, Bosna i Hercegovina, str. 257-262.

6.14 Саопштења с међународних скупова штампана у изводу (M64)

Радови објављени пре претходног избора у звање

1. Stojić, S.S., **Stojić, A.**, Šoštarić, A., Mijić, Z., Todorović, M., 2015. Contribution of transported pollution to traffic-related VOC concentrations in Belgrade urban area. 7th Symposium Chemistry and Environmental Protection, June 9-12, Palić, Serbia, pg. 167-168.

2. Todorović, M., Perišić, M., **Stojić, A.**, Rajšić, S., 2013. Concentrations trend of NO, NO_2 and O_3 during the 2011 in Belgrade urban area. 6th Symposium Chemistry and Environmental Protection, May 21-24, Vršac, Serbia, pg. 320-321.
3. Perišić, M., Todorović, M., **Stojić, A.**, Kuzmanoski, M., Rajšić, S., 2013. Health risk assessment of VOCs in Belgrade semi-urban area, 6th Symposium Chemistry and Environmental Protection, May 21-24, Vršac, Serbia, pg. 378-379.
4. Nestorović, J., Mišić, D., Dević, M., **Stojić, A.**, Malović, G., Grubišić, D., 2009. PTR-MS and HPLC analysis of nepetalactone in shoots cultures of *Nepeta rtanjensis* Diklić & Milojević. XVIII Symposium of Biological Society, May 27-29, Vršac, Serbia.
5. Nešić, M., **Stojić, A.**, Mijić, Z. Rajšić, S., Tasić, M., 2008. First results of ambient VOCs measurements using PTR-MS in Belgrade. 5th Symposium Chemistry and Environmental Protection, Ed. The Serbian Chemical Society, Book of abstracts, June, 27-30, Tara, Serbia, pp. 41.
6. Nestorović, J., Mišić, D., Šiler, B., Grubišić, D., Nešić, M., **Stojić, A.**, Tasić, M., 2008. Uticaj isparljivih jedinjenja rtanjske metvice (*Nepeta rtanjensis*) na klijanje semena *Lepidum sativum*: alelopatski potencijal. IX dani lekovitog bilja, Septembar 17-20, Kosmaj, Srbija, pp. 138.

6.15 Одбрањена докторска теза (M71)

Радови објављени пре претходног избора у звање

1. **Stojić**, **A.**, 2015, Spatio-temporal Distribution of Volatile Organic Compounds and Aerosols in Troposphere: Lidar and Mass Spectrometry, Faculty of Physics, University of Belgrade.

6.16 Newsletters

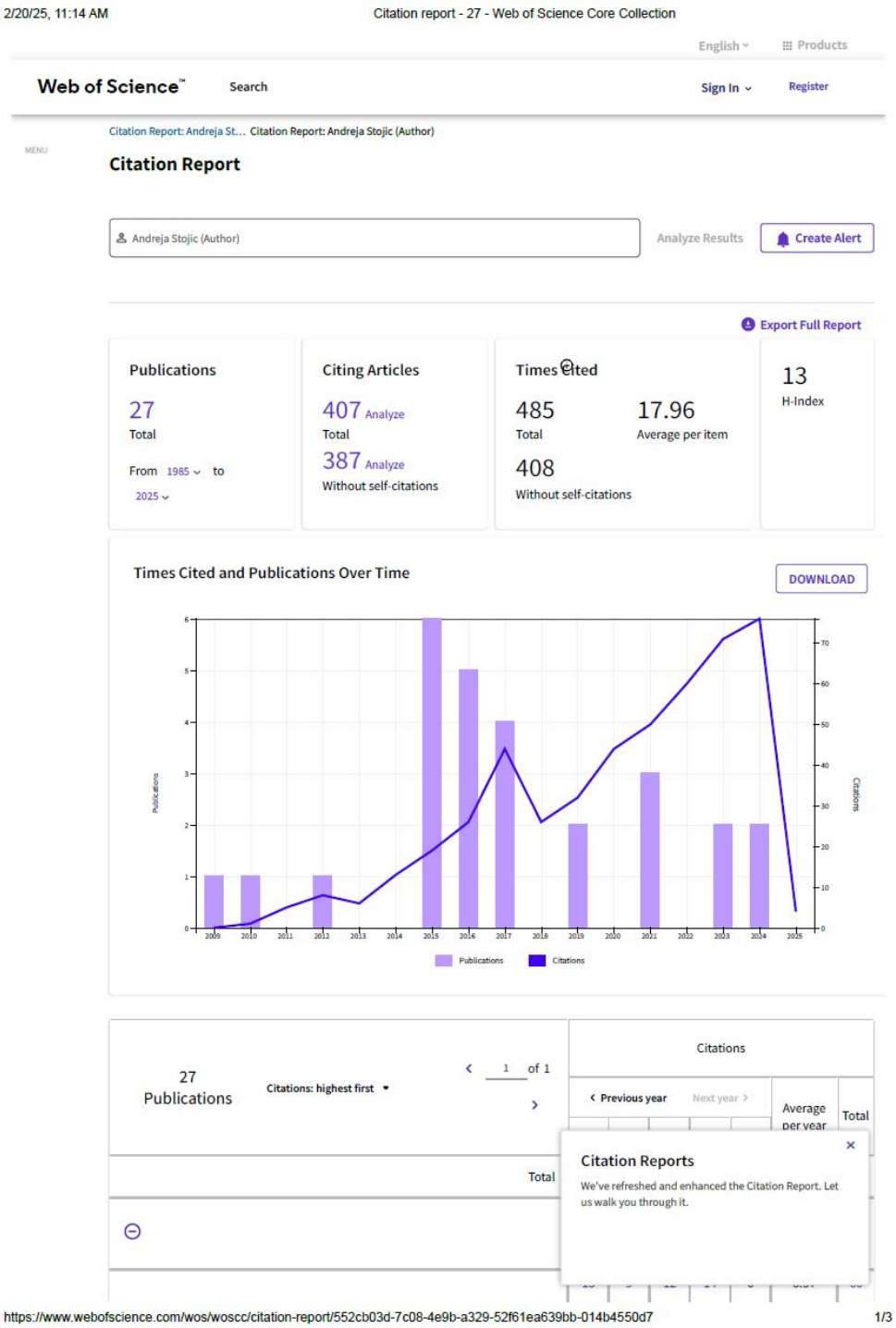
1. Šoštarić, A., Mladenović, S., Stanišić Stojić, S., **Stojić, A.**, Stanišić, N., Slepčević V., 2015. Health burden of air pollutant exposure in Belgrade: a European region with high circulatory and malignant mortality rates, Newsletter, WHO Collaborating Centre for Air Quality Management and Air Pollution Control, No. 56, December 2015, pp. 3-9.

6.17 Симпозијуми

1. **Stojić**, **A.**, 2018. Principles underlying Agile approach and highly sophisticated machine learning algorithms. 1st Agile Humans Belgrade Day, October 24, Belgrade, Serbia. https://agilehumans.city/agile-humans-belgrade-day/
2. Pavlović, N., Ristić, J., Šoštarić, A., Slepčević, V., **Stojić, A.**, Stanišić Stojić, S., Stanišić, N., 2016. Analiza podataka redovnog merenja PM_{10}, O_3 i NO_2 u vazduhu i smrtnost od kardiovaskularnih i respiratornih bolesti i *Diabetes mellitusa* u Beogradu, Stručna konferencija – simpozijum „Dani Zavoda 2016. godine", sa temom: „Kvalitet vazduha – monitoring, modelovanje, unapređenje", Novembar 25, Beograd, Srbija, pp. 99-114.

Прилози

Цитираност



Publishing Summary

For manuscripts published from date range February 2007 - February 2025

(4) Atmospheric Environment	(2) Chemosphere
(2) Air Quality, Atmosphere and Hea...	(2) Science of the Total Environment
(2) Environmental Science and Pollu...	(1) Stanovnistvo
(1) Environmental Health	(1) Journal of Physics: Conference S...
(1) Journal of Toxicology and Environ...	(1) Environmental Research
(1) International Journal of Environ...	(1) Journal of Environmental Scienc...
(1) Ecotoxicology and Environmenta...	(1) Marine Pollution Bulletin
(1) Atmosphere	(1) Chemical Industry & Chemical E...
(1) Plant Biosystems	(1) Acta Physiologiae Plantarum

Publications

For manuscripts published from date range February 2007 - February 2025 (25)	Times Cited (All time)
The Explainable Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing VOCs' Environmental Fate Published: Jan 2023 in Atmosphere DOI: 10.3390/ATMOS14010109	26
The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo[a] pyrene levels Published: Feb 2022 in Chemosphere DOI: 10.1016/J.CHEMOSPHERE.2021.133154	6
Fatty acids, persistent organic pollutants, and trace elements in small pelagic fish from the eastern Mediterranean Sea Published: Sep 2021 in Marine Pollution Bulletin DOI: 10.1016/J.MARPOLBUL.2021.112654	13
The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources Published: Feb 2021 in Environmental Research DOI: 10.1016/J.ENVRES.2020.110520	14
Introducing of modeling techniques in the research of POPs in breast milk - A pilot study Published: May 2019 in Ecotoxicology and Environmental Safety DOI: 10.1016/J.ECOENV.2019.01.087	10

Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition Published: Feb 2019 in Science of the Total Environment DOI: 10.1016/J.SCITOTENV.2018.10.368	61
Rainwater capacities for BTEX scavenging from ambient air Published: Nov 2017 in Atmospheric Environment DOI: 10.1016/J.ATMOSENV.2017.08.045	9
Levels of PM10-bound species in Belgrade, Serbia: spatio-temporal distributions and related human health risk estimation Published: Jan 2017 in Air Quality, Atmosphere and Health DOI: 10.1007/S11869-016-0411-6	14
Forecasting hourly particulate matter concentrations based on the advanced multivariate methods Published: 2017 in International Journal of Environmental Science and Technology DOI: 10.1007/S13762-016-1208-8	10
The innovative concept of three-dimensional hybrid receptor modeling Published: 2017 in Atmospheric Environment DOI: 10.1016/J.ATMOSENV.2017.06.009	18
Temperature-related mortality estimates after accounting for the cumulative effects of air pollution in an urban area Published: Jul 2016 in Environmental Health DOI: 10.1186/S12940-016-0164-6	12
Comprehensive analysis of PM10 in Belgrade urban area on the basis of long-term measurements Published: Jun 2016 in Environmental Science and Pollution Research DOI: 10.1007/S11356-016-6266-4	13
Quantification and mechanisms of BTEX distribution between aqueous and gaseous phase in a dynamic system Published: Feb 2016 in Chemosphere DOI: 10.1016/J.CHEMOSPHERE.2015.09.042	17
Heavy metal accumulation in wheat and barley: The effects of soil presence and liquid manure amendment Published: Jan 2016 in Plant Biosystems DOI: 10.1080/11263504.2014.976288	8
Single and combined effects of air pollutants on circulatory and respiratory system-related mortality in Belgrade, Serbia Published: Jan 2016 in Journal of Toxicology and Environmental Health - Part A: Current Issues DOI: 10.1080/15287394.2015.1101407	27

Seasonal mortality variations of cardiovascular, respiratory and malignant diseases in the City of Belgrade Published: 2016 in Stanovnistvo DOI: 10.2298/STNV151019001S	Not indexed in the Web of Science
Assessment of PM10 pollution level and required source emission reduction in Belgrade area Published: Nov 2015 in Journal of Environmental Science and Health, Part A DOI: 10.1080/10934529.2015.1059110	4
Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling Published: Sep 2015 in Environmental Science and Pollution Research DOI: 10.1007/S11356-015-4540-5	26
Estimation of required PM10 emission source reduction on the basis of a 10-year period data Published: Aug 2015 in Air Quality, Atmosphere and Health DOI: 10.1007/S11869-014-0292-5	9
Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods Published: Jul 2015 in Science of the Total Environment DOI: 10.1016/J.SCITOTENV.2015.03.098	20
Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling Published: Apr 2015 in Atmospheric Environment DOI: 10.1016/J.ATMOSENV.2015.01.071	20
Essential oils of two Nepeta species inhibit growth and induce oxidative stress in ragweed (Ambrosia artemisiifolia L.) shoots in vitro Published: Mar 2015 in Acta Physiologiae Plantarum DOI: 10.1007/S11738-015-1810-2	17
RECEPTOR MODELING STUDIES FOR THE CHARACTERIZATION OF PM10 POLLUTION SOURCES IN BELGRADE Published: Oct 2012 in Chemical Industry & Chemical Engineering Quarterly DOI: 10.2298/CICEQ120104108M	15
Seasonal variability and source apportionment of metals in the atmospheric deposition in Belgrade Published: Sep 2010 in Atmospheric Environment DOI: 10.1016/J.ATMOSENV.2010.06.045	95
Source Apportionment of Atmospheric Bulk Deposition in the Belgrade Urban Area Using Positive Matrix Factorization Published: Apr 2009 in Journal of Physics: Conference Series DOI: 10.1088/1742-6596/162/1/012018	8

Scopus

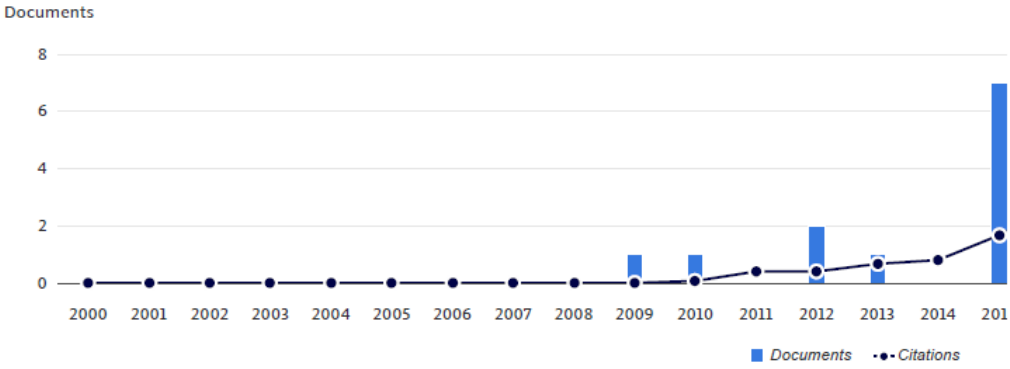
← Back to author results

Citation overview

Stojić, Andreja M.

33	623	15
Documents	Citations	h-index

Date range: 2000 to 2025 | Exclude self citations of selected author | Exclude self citations of all authors | Exclude book citations | Hide documents with 0 citations | Export



Sort by Date (newest)

	Documents	Year	<2000	2000	2001	2002	2003	2004	Total
	Total		**0**	**0**	**0**	**0**	**0**	**0**	623
1	The explainable potential of coupling hybri...	2024	0	0	0	0	0	0	10
2	Potential of Coupling Metaheuristics-Opti...	2023	0	0	0	0	0	0	11
3	The Explainable Potential of Coupling Met...	2023	0	0	0	0	0	0	45
4	The PM_{2.5}-bound polycyclic a...	2022	0	0	0	0	0	0	8

Scopus

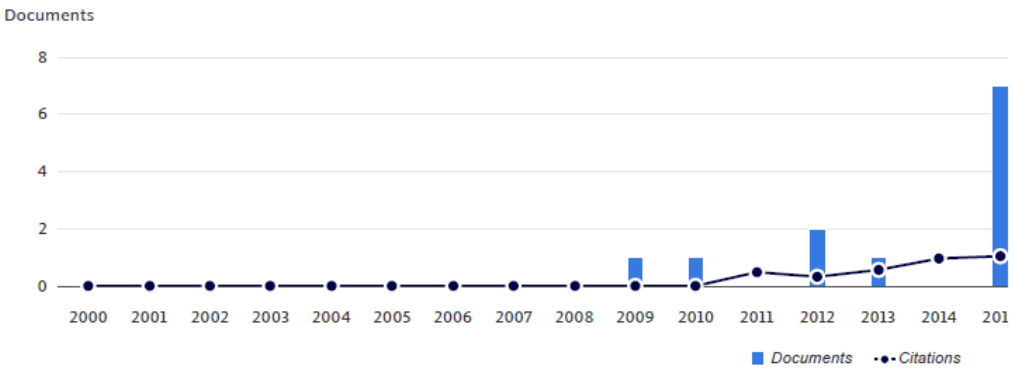
Back to author results

Citation overview

Stojić, Andreja M.

33	506	14
Documents	Citations	h-index

Date range: 2000 to 2025 Exclude self citations of selected author Exclude self citations of all authors Exclude book citations Hide documents with 0 citations Export



Sort by Date (newest)

	Documents	Year	<2000	2000	2001	2002	2003	2004	Total
	Total		**0**	**0**	**0**	**0**	**0**	**0**	506
1	The explainable potential of coupling hybri...	2024	0	0	0	0	0	0	9
2	Potential of Coupling Metaheuristics-Opti...	2023	0	0	0	0	0	0	9
3	The Explainable Potential of Coupling Met...	2023	0	0	0	0	0	0	42
4	The PM_{2.5}-bound polycyclic a...	2022	0	0	0	0	0	0	4

Одлука о избору звање

Република Србија
МИНИСТАРСТВО ПРОСВЕТЕ,
НАУКЕ И ТЕХНОЛОШКОГ РАЗВОЈА
Комисија за стицање научних звања

Број: 660-01-00001/1422
15.09.2020. године
Б е о г р а д

На основу члана 24. став 2. и члана 76. став 6. Закона о науци и истраживањима ("Службени гласник Републике Србије", број 49/19), члана 3. ст. 1. и 3. и члана 40. Правилника о поступку, начину вредновања и квантитативном исказивању научноистраживачких резултата истраживача ("Службени гласник Републике Србије", број 24/16, 21/17 и 38/17) и захтева који је поднео

Институт за физику у Београду

Комисија за стицање научних звања на седници одржаној 15.09.2020. године, донела је

ОДЛУКУ
О СТИЦАЊУ НАУЧНОГ ЗВАЊА

Др Андреја Стојић
стиче научно звање
Виши научни сарадник

у области природно-математичких наука - физика

О Б Р А З Л О Ж Е Њ Е

Институт за физику у Београду

утврдио је предлог број 206/1 од 11.02.2020. године на седници Научног већа Института и поднео захтев Комисији за стицање научних звања број 234/1 од 13.02.2020. године за доношење одлуке о испуњености услова за стицање научног звања ***Виши научни сарадник***.

Комисија за стицање научних звања је по претходно прибављеном позитивном мишљењу Матичног научног одбора за физику на седници одржаној 15.09.2020. године разматрала захтев и утврдила да именовани испуњава услове из члана 76. став 6. Закона о науци и истраживањима ("Службени гласник Републике Србије", број 49/19), члана 3. ст. 1. и 3. и члана 40. Правилника о поступку, начину вредновања и квантитативном исказивању научноистраживачких резултата истраживача ("Службени гласник Републике Србије", број 24/16, 21/17 и 38/17) за стицање научног звања ***Виши научни сарадник***, па је одлучила као у изреци ове одлуке.

Доношењем ове одлуке именовани стиче сва права која му на основу ње по закону припадају.

Одлуку доставити подносиоцу захтева, именованом и архиви Министарства просвете, науке и технолошког развоја у Београду.

ПРЕДСЕДНИК КОМИСИЈЕ
Др Ђурђица Јововић,
научни саветник

МИНИСТАР
Младен Шарчевић

Позивно предавање

Sinteza

Singidunum University International Scientific Conference

SINTEZA 2020- Information Technology and Data Related Research, provides an ideal platform for the exchange of information and dissemination of best practices, advancements in the state-of-the-art and technical improvements in the domain of ICT and e-business related research in today's ubiquitous and virtual environment.

Rapid advances in information and communication technology (ICT) in recent decades have had a huge impact on numerous facets of everyday life and have created tremendous opportunities for economic, technological and social gains at a global scale. New technologies and scientific breakthroughs have altered the working and living environments making them safer, more convenient and more connected. As a key infrastructure of knowledge-based economies, ICT is a driving force for rapidly growing new sectors, including advanced computing and software development, business process out-sourcing and various Internet services.

Certificate of Appreciation

is awarded to

ANDREJA STOJIĆ, *Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade*

for participating as a keynote speaker and delivering a lecture on the following topic

UNDERSTANDING POLLUTION – CONTEXTUAL ENVIRONMENTAL DATA ANALYSIS

at the online International Scientific Conference SINTEZA 2020
"Information Technology and Data Related Research"

October 17, 2020
Singidunum University
Belgrade, Serbia

Professor Mladen Veinović, PhD
Dean of Faculty of Informatics and Computing
Member of Organizing and Scientific Committee

УНИВЕРЗИТЕТ СИНГИДУНУМ
IV
БЕОГРАД

Professor Nemanja Stanišić, PhD
Rector of Singidunum University
Member of Scientific Committee

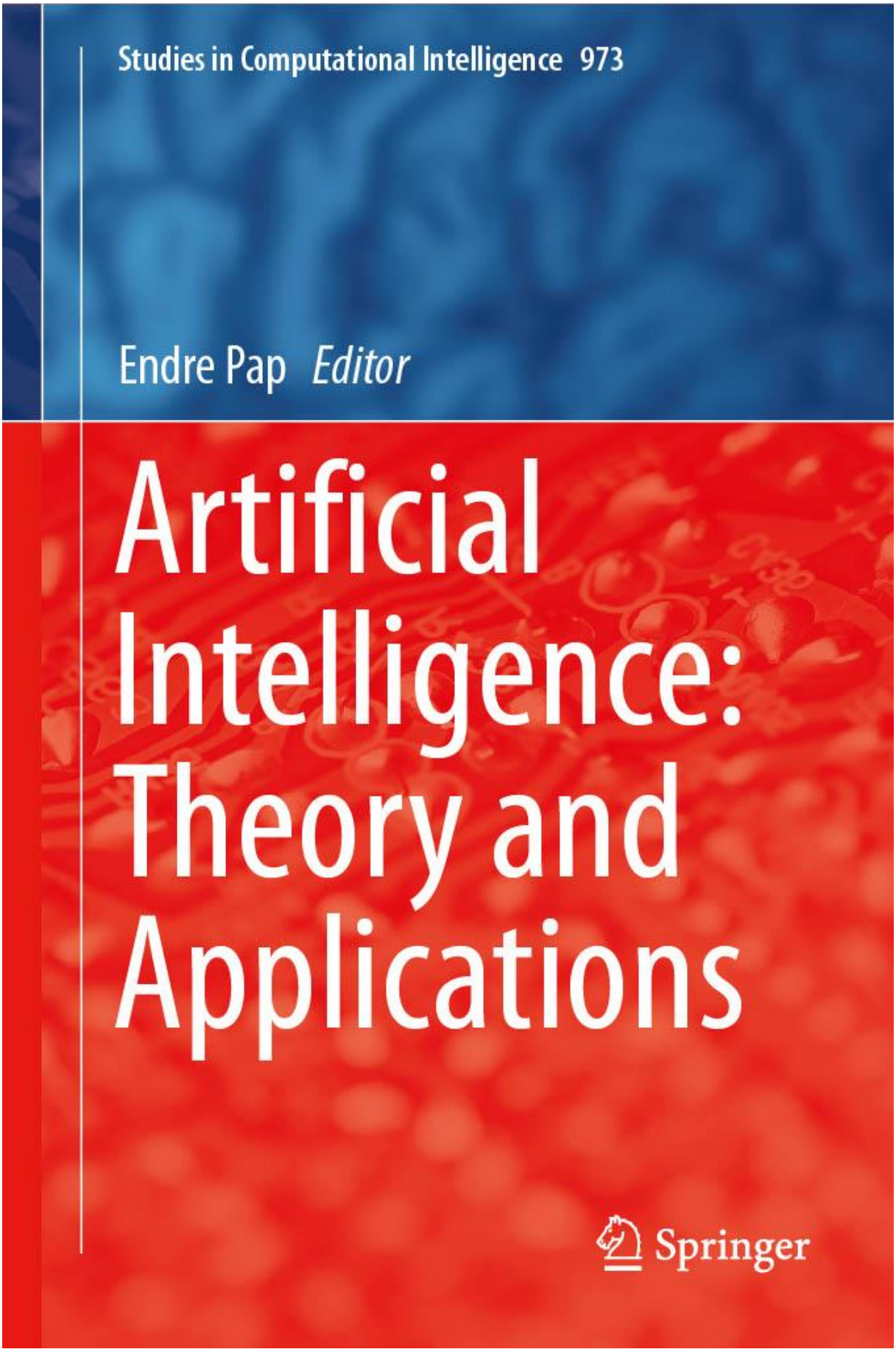
Singidunum University was founded on January 17th, 2005, by decision No 612-00-663/2004-04 issued by the National Council for University Education Development, giving approval to the feasibility study for the establishment of Singidunum University. From the very beginnings, we have been governed by the ambition to offer modern knowledge to future business elite of Serbia and neighbouring countries and to educate competent experts by applying scientific methods and techniques used at the renowned European universities.

Singidunum University

Публикације

Публикације категорије М10

Публикације након избора у претходно звање


Studies in Computational Intelligence 973
Endre Pap Editor
Artificial Intelligence: Theory and Applications
Springer

The series "Studies in Computational Intelligence" (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/7092

Studies in Computational Intelligence

Volume 973

Series Editor
Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

Endre Pap
Editor

Artificial Intelligence: Theory and Applications

Springer

Editor
Endre Pap
Singidunum University
Belgrade, Serbia

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-3-030-72710-9 ISBN 978-3-030-72711-6 (eBook)
https://doi.org/10.1007/978-3-030-72711-6

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Applications

Patterns of PCB-138 Bioaccumulation in Small Pelagic Fish from the Eastern Mediterranean Sea Using Explainable Machine Learning Prediction

Andreja Stojić, Bosiljka Mustać, Gordana Jovanović, Jasna Đinović Stojanović, Mirjana Perišić, Svetlana Stanišić, and Snježana Herceg Romanić

Abstract Fish consumption, especially consumption of oily marine species, is increasing globally due to its recommendation by dieticians. This is due to high polyunsaturated ω-3 and ω-6 (PUFAs) fatty acid content in the tissue of the fish. The health benefits of PUFA ingestion coincide with the risk of intaking hazardous lipophilic persistent pollutants including organochlorine pesticides (OCPs) and related polychlorinated biphenyls (PCBs). We examined the impact of 17 fatty acids (FAs) and 36 toxic organic and inorganic contaminants on the behavior patterns of the indicator congener PCB-138 in marine fish using eXtreme Gradient Boosting (XGBoost), SHapley Additive exPlanations (SHAP), and SHAP value fuzzy clustering. XGBoost indicated non-linear relationships between PCB-138 and other investigated variables that were explained by SHAP values. The ten obtained fuzzy

A. Stojić (✉) · G. Jovanović · M. Perišić
Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia
e-mail: andreja@ipb.ac.rs

G. Jovanović
e-mail: gordana.vukovic@ipb.ac.rs

M. Perišić
e-mail: mirjana.perisic@ipb.ac.rs

B. Mustać
Department of ecology, agronomy and aquaculture, University of Zadar, Zadar, Croatia
e-mail: bmustac@unizd.hr

J. Đinović Stojanović
Institute of Meat Hygiene and Technology, Kaćanskog 13, 11 000 Belgrade, Serbia
e-mail: jasna.djinovic@inmes.rs

S. Stanišić
Environment and Sustainable Development, Singidunum University, Belgrade, Serbia
e-mail: sstanisic@singidunum.ac.rs

S. Herceg Romanić
Institute for Medical Research and Occupational Health, Ksaverska cesta 2, PO Box 291, 10001 Zagreb, Croatia
e-mail: sherceg@imi.hr

E. Pap (ed.), *Artificial Intelligence: Theory and Applications*,
Studies in Computational Intelligence 973,
https://doi.org/10.1007/978-3-030-72711-6_10

clusters of SHAP values revealed that a higher intake of saturated myristic-C14:0 and margaric-C17:0 acids followed by the intake of nutritionally beneficial eicosadienoic acid (C20:2n-6) mostly do not increase the bioaccumulation of PCB-138. Important effects on PCB-138 behavior patterns were also recorded for the chemically allied indicator congeners (−153, −180, −118 and−101) and organochlorines' metabolite p,p'-DDE. Associations between the target congener and the toxicologically relevant PCBs (−123 and −170) were less prominent.

Keywords Persistent organic pollutants (pops) · (omega-3-6) fatty acids · Heavy metals · Shapley additive explanations (shap)

1 Introduction

Whether intentionally or unintentionally, anthropogenic activities have led to the release of numerous man-made synthetic chemicals which continuously harm human health. The contaminants have also caused several global, regional, and local environmental problems related to air, water and soil pollution, a gradual decrease of the stratospheric ozone layer, a decrease in biodiversity, etc. Organic and inorganic contaminant emissions and dispersion caused by human activities refer to different classes of polychlorinated biphenyls (PCBs), polycyclic aromatic pollutants, trace metals, and natural radioactivity. Organochlorine pesticides (OCPs) and PCBs are well-known persistent organic pollutants (POPs). They are long-lived contaminants, which resist photolytic, biological, and chemical degradation to differing extents. Because of their resistance to metabolic breakdown and high lipid solubility, POPs are bioaccumulated in the fatty tissues of living organisms and cause numerous adverse effects [20]. Although their production and use are limited or prohibited in most countries, investigating the behavior of POPs is still challenging as long-term monitoring of their impacts was not mandatory during their early use. Recognized as hydrophobic, POPs bind to particles in soil and sediment tightly, which can act as a secondary source of contamination for environmental media (water, air and living organisms).

At a global level, the final sink/destination of POPs are marine environments, which are polluted via coastal outfalls, rivers, and deposition from the atmosphere. Oceans act as a secondary source of contamination because POPs are slowly degraded and adsorbed on suspended particles or bioaccumulated in benthic marine organisms, which are at the bottom of the food chain and represent a source of toxins in human nutrition. The assessment of environmental exposure to marine toxins is based on data on the concentrations of POPs in samples of water, plants, and food, surrounding conditions (e.g., salinity, temperature), the pollutant physico-chemical properties and biotic factors (e.g., organisms' feeding modality, trophic position). Heavy metals can occur naturally in marine environments and some of them are essential to living organisms. However, levels are increased by anthropogenic activities and all of them can be toxic above threshold levels.

Small pelagic oily fish are a highly recommended source of nutrients worldwide due to their content of protein, minerals and healthy fats, including long-chain omega-3 (ω-3) and 6 (ω-6) polyunsaturated fatty acids (PUFAs), such as eicosapentaenoic (EPA) and docosahexaenoic acid (DHA) [8]. According to FAO [2018a] small pelagic fish are considered low-priced fish, which means they are an important food source in many developing countries, whereas in other countries they are largely processed into fishmeal and fish oil. The Mediterranean Sea is an important commercial fishing ground whose catches are primarily small pelagic fish, accounting for around 50% of all catches [9]. However, it is a semi-enclosed sea, which is particularly vulnerable to chemical contamination from the surrounding heavily industrialized and agricultural countries. The benefits of fatty acid consumption are associated with normal growth and development, the prevention of cardiovascular and inflammatory diseases, as well as the prevention of cognitive decline and dementia. However, these FAs represent a suitable matrix for the bioaccumulation of highly lipophilic xenobiotics such as POPs. More than 90% of human exposure to POPs is through food; mainly meat, dairy products, fish, and shellfish [24].

Although the profiles and organochlorine content of FAs have been evaluated in numerous marine fish species worldwide, there are few data on their interrelations. In our previous studies, a preliminarily investigation was conducted on the presence of OCPs and PCBs in different edible fish species [12], followed by elaboration on the joint use of advanced algorithms (Kohonen self-organizing maps and Decision Tree Learning) to study the spatio-temporal distribution of POPs in fish from the Croatian Adriatic [23]. In this study, the presence of OCPs and PCBs was examined in small pelagic edible fish species: sardine *Sardina pilchardus* (Walbaum, 1792), anchovy *Engraulis encrasicolus* (Linnaeus, 1758), round sardinella *Sardinella aurita* (Valenciennes, 1847), chub mackerel *Scomber japonicus* (Houttuyn, 1782) and horse mackerel *Trachurus trachurus* (Linnaeus, 1758). We applied eXtreme Gradient Boosting (XGBoost), SHapley Additive exPlanations (SHAP), and SHAP value fuzzy clustering aiming to obtain a detailed insight into the distribution of the indicator congener PCB-138 in the fish species. The impacts of the following factors including the level of OCPs, PCBs, saturated fatty acids (SFAs), monounsaturated fatty acids (MUFAs), PUFAs and heavy metals were evaluated using SHAP since the method offers uniquely consistent and locally accurate solutions that have been confirmed in previous studies of environmental phenomena [22]. An earlier version of this paper was presented at the International Scientific Conference on Information Technology and Data Related Research (Sinteza 2020), Singidunum University [21].

2 Materials and Methods

2.1 Sampling

Fish samples were collected along the Croatian part of the Eastern Adriatic Sea in 2014 and 2016. Details about sampling were described previously [23]. A total of 107 fish samples were collected in various fisheries in coastal (A, E, and F) and off-coast (B and C) zones. Approximately 50 specimens were sampled randomly using purse seine catches (mesh size: 8 mm/bar length) totaling 107 pooled samples from the fillet of the specimens.

2.2 Chemical Analyses of POPs

The chemical analysis of POPs was described previously elsewhere [12, 23]. In brief, from each pool approximately 5 g of two aliquots was ground with 2 g of sodium sulphate in a glass mortar and extracted with 40 mL of n-hexane. After passing through filter paper (Whatman No. 1) into a pre-weighted test tube, the extract was reduced under a nitrogen steam. The lipid extracts were dissolved in 5 mL of n-hexane and purified three times with 4 mL of 96% sulphuric acid; the solvent was evaporated under a gentle nitrogen stream. Finally, the tissue lipid content was determined gravimetrically.

Seven OCPs (HCBα-, β-, and γ-HCH, *p,p'*-DDT, *p,p'*-DDE, and *p,p'*-DDD), six indicator PCB congeners (PCB-28, PCB-52, PCB-101, PCB-138, PCB-153, and PCB-180), eight mono ortho congeners (PCB-105, PCB-114, PCB-118, PCB-123, PCB-156, PCB-157, PCB-167, PCB-189) and PCB-60, PCB-74 and PCB-170 were analyzed. The extracted residues were dissolved in 1.0 mL of n-hexane and 5 μ L of the sample was injected. The pollutant content was determined by high-resolution gas chromatography with an electron capture detector (s) performed on a CLARUS 500 chromatograph [23]. Two capillary columns (Restek, Bellefonte, PA, USA) were used simultaneously: (1) 60 m × 0.25 mm, Rtx-5 film thickness of 0.25 μ m, and (2) 30 m × 0.25 mm, Rtx-1701 film thickness of 0.25 μ m. The column temperature was attained after the following steps: (1) initial heating from 100 °C to 110 °C at 4 °C min^{-1}, (2) isothermal heating for 5 min at 110 °C, (3) heating from 240 °C at 15 °C min^{-1}, and (4) 50 min of isothermal heating at 240 °C. Nitrogen was used as the carrier gas. The injector and detector temperatures were 250 and 270 °C, respectively. Only compounds identified on both columns were evaluated.

The certified reference material CRM 430 (Community Bureau of Reference, Commission of the European Communities, Brussels, Belgium) was used to determine the relative accuracy of the method. The results agreed with the reference range. The concentrations of analytes in blanks were below the detection limit of the instrument.

To test the recovery and reproducibility of the method, standard addition was used. A known amount of all of the analyzed compounds (between 0.61 and 0.68 ng g^{-1} of fresh weight for OCPs and between 0.42 and 0.7 ng g^{-1} of fresh weight for PCBs) were added to the five aliquots of homogenized samples before extraction. The recoveries for the PCBs ranged between 75% and 89% while a relative standard deviation (RSD) of 1% to 11% was obtained. The recoveries for OCPs ranged from 76% to 86%, with RSD from 1% to 11% for both PCBs and OCPs, while the determination limits were 0.01 ng g^{-1} of fresh weight. The presented data were recalculated based on the recovery values.

2.3 Chemical Analyses of Elements

Sixteen macro- and micro-elements including toxic metals (Na, Mg, K, Ca, As, Cd, Co, Cr, Cu, Mn, Fe, Hg, Ni, Zn, Pb and Se) were analyzed using the previously described procedure [5]. Prior to analysis, the frozen fish samples were thawed at 4 °C. The homogenized fish samples (0.5 g) were basted with 5 mL of nitric acid (67% TraceMetal grade, Fisher Scientific, Bishop, UK) and 1.5 mL of hydrogen peroxide (30% analytical grade, Sigma-Aldrich, St. Louis, MA, USA). Subsequently, microwave digestion (Start D Microwave Digestion System, Milestone, Sorisole, Italy) was performed under the following conditions: maximum power (1000W); heating ramp to 180 °C in 5 min; hold at 180 °C for 15 min; cooling in an oven for 20 min and then at room temperature for 15 min. After cooling at room temperature, the digests were diluted to 100 mL with deionized water into polypropylene volumetric flasks. Analysis of the elements was performed by inductively coupled plasma mass spectrometry (ICP-MS), (iCap Q mass spectrometer, Thermo Scientific, Bremen, Germany). The operating conditions were: RF power (1550 W); cooling gas flow (14 L min^{-1}); nebulizer flow (1 L min^{-1}); collision gas flow (1 mL min^{-1}); operating mode (Kinetic Energy Discrimination-KED); and dwell time (10 ms). The most abundant isotopes were used for quantification.

For five-point calibration, solutions of Fe, Zn, Cu, Mn, Se, Cr, Co, Ni, Na, K, Mg, Ca, Cd, Pb, Hg and As were prepared from standard stock solutions (VGH labs, Manchester, UK) containing 1000 mg L^{-1} of each element. The concentration of the calibration-solution for Fe, Zn, Cu, Mn, Se, Cr, Co, Ni, Na, K, Mg and Ca was in a concentration range of 0.2–2.0 mg L^{-1}, for Cd, Hg and As, it ranged from 0.2 to 2.0 µg L^{-1} and for Pb it ranged between 2.0 and 20.0 µg L^{-1}.

To check the accuracy of the analysis, the certified reference material NIST SRM 1577c (bovine liver, Gaithersburg, MD, USA) was analyzed in the same manner as the fish samples. For all elements, the obtained results were within a satisfactory range of the certified values.

2.4 Chemical Analyses of Fatty Acids

The concentrations of 6 SFAs (myristic, pentadecylic, palmitic, palmitoleic, margaric, and stearic acid), 3 MUFAs (oleic, paullinic and arachidonic acid) and 8 PUFAs, ω-6 and ω-3 families (linoleic, α-linolenic, eicosadienoic, dihomo-γ-linolenic, eicosatrienoic, eicosapentaenoic, docosapentaenoic and docosahexaenoic) were determined according to the previously described procedure [18]. Prior to analysis, the samples were partially thawed at +4 °C. To determine the total lipids for fatty acid content, accelerated solvent extraction (ASE 200, Dionex, Sunnyvale, CA), using a mixture of n-hexane and iso-propanol (60:40 v/v) was applied to an extraction cell heated at 100 °C and under nitrogen pressure of 10.3 MPa. The samples were analyzed as fatty acid methyl esters (FAME) obtained by transesterification using trimethyl sulfonium hydroxide (ISO 2000 standard). Afterwards, gas-liquid chromatography (Shimdzu, Japan) was performed. A fused silica cianopropyl HP-88 column (length 100 m, i.d. 0.25 mm, film thickness 0.20 μ m, J&W Scientific, USA) and a flame ionization detector (GC/FID) were used. Hydrogen was used as a carrier gas at a 1.3 mL flow rate and an injector split ratio of 1:50. The following temperature program was used: (1) heating from 125 °C to 175°C at a rate of 10°C min^{-1}, (2) isothermal heating at 175 °C for 10 min, (3) a temperature increase from 175 °C to 210°C (rate 5 °C min^{-1}), (4) isothermal heating at 210 °C for 5 min, and (5) final temperature incline to 230 °C at a rate of 2 °C min^{-1}. The injector and detector temperatures were 250 °C and 280 °C, respectively. The chromatographic peaks were identified and quantified using Supelco 37 Component mix standard and internal standard (heneicosanoic acid methyl ester), respectively.

The chromatographic peaks were identified and quantified using Supelco 37 Component mix standard and internal standard (heneicosanoic acid methyl ester), respectively.

2.5 Data Analysis

The relationships between PCB-138 and all the other measured parameters were modeled through XGBoost regression. Details on the method were given elsewhere [23]. Briefly, it is a decision-tree-based ensemble machine learning algorithm that applies a gradient boosting framework. The method boosts weak learners by sequentially correcting the errors made by existing models. XGBoost is based on a gradient descent algorithm, used to minimize loss when adding new models. The algorithm includes system optimization and algorithmic enhancements through parallelized sequential tree building, tree pruning, regularization, weighted quantile sketch algorithm implementation, cross-validation, etc. In this study, we used Python XGBoost implementation. The dataset was split into training (80%) and validation (20%) sets. Hyperparameter tuning was implemented using a brute-force grid search and 10-fold

stratified cross-validation. The best performing hyperparameter values were used for the final model.

The explainability of the produced XGBoost model, that operates with high-dimensional input data in a non-linear fashion, was obtained using the explainable artificial intelligence method SHapley Additive exPlanations (SHAP) [2, 13]. Based on game theory, Shapley explanations represent the only possible locally accurate and globally consistent feature attribution values. In this study, we applied fuzzy clustering of absolute SHAP attributions to identify and characterize the relations among the measured parameters responsible for PCB-138 behavior.

3 Results and Discussion

3.1 Fatty Acid Content

Descriptive statistics of the content of the FAs in the examined fish species is presented in Fig. 1. The molecules of FAs are classified according to the presence and sum of double bonds, which determine their physico-chemical properties, functional characteristics, and metabolic fate in living organisms. The most prevalent and equally abundant FAs in sardine, anchovy and mackerel species were palmitic acid (C16:0), which is an SFA with no double bond, and oleic acid (C18:1n-9), a MUFA, which

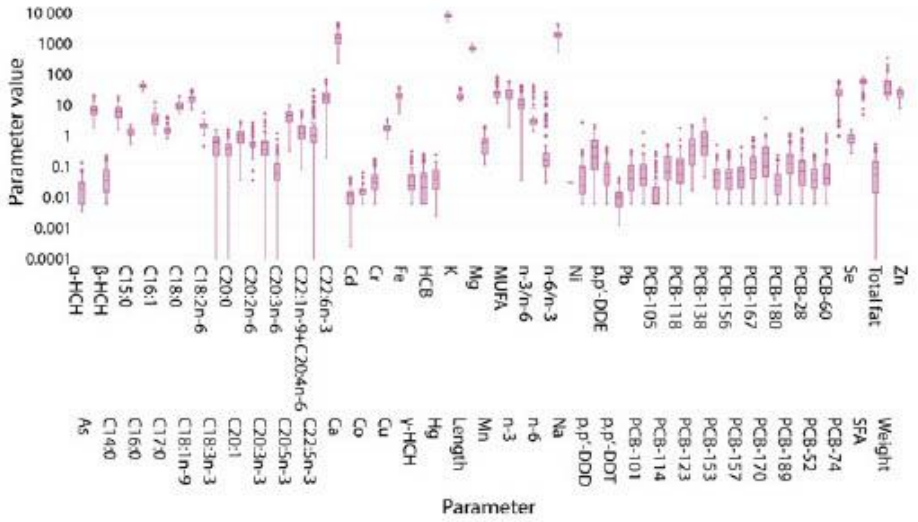


Fig. 1 Box plot

has a single double bond. In the studied fish, the highest percentage of these acids are usually found in the form of free FAs in which they are transported to the large intestine and extracted. In addition, the absorption of palmitic acid by human enterocytes is limited since the best absorption rate of SFAs is achieved when they are presented as sn-2 monoacylglycerols [16, 17].

Although some SFAs and MUFAs can be produced in the human body from carbon groups, which are constituents of carbohydrate and protein macromolecules, humans cannot synthesize some ω-3 and ω-6 PUFAs since they lack the delta 12 and delta 15 desaturase enzymes, which arrange a double bond at the n-3 and n-6 positions of FAs. These PUFAs are known as essential FAs and their uptake must be provided via food intake including fish and seafood. From the parent acids, ω-3 α-linoleic (C18:3n-3) and ω-6 linoleic (C18:2n-6) acid, 20 carbon long-chain (or more) FAs could be synthesized through successive reactions of desaturation and elongation [11]. The possible products are dihomo-γ-linolenic acid (20:3n-6), arachidonic acid (20:4n-6), EPA (20:5n-3) and DHA (22:6n-3). However, the metabolite-conversion efficiency is low and the consumption of EPA- and DHA-enriched food is recommended to ensure an optimal level of FAs in human blood, which reflects both dietary intake and metabolic production [4].

Eicosapentaenoic acid and DHA are the most recognized n-3 PUFAs with two or more double bonds because of their nutritionally and physiologically beneficial effects for healthy adults, pregnant women, and children. As presented in Fig. 1, the content of DHA was significant in the studied marine fish and varied from 10.4% to 21.3% while EPA was less abundant, from 3.3% to 4.6%. The accumulation pattern of both FAs was as follows: sardines>anchovy>chub mackerel>horse mackerel while the DHA:EPA ratio was approximately 2.5:1 implying that a variety of different fish species should be consumed to gain a sufficient supply of nutritionally relevant FAs. A considerably higher amount of EPA and DHA has been reported in the class of phospholipids compared to other classes including monoacylglycerols, diacylglycerols, free fatty acids, triacyl-glycerols, sterol esters, neutral lipids, and glycolipids. In this way, EPA and DHA are more bioavailable and less susceptible to oxidative degradation whereas EPA and DHA free forms of FA are less efficiently absorbed than other lipid classes and are easily eliminated from human intestines where they are dispersed into mixed micelles and bound to soluble lipid-binding proteins [3, 17]. Among the other essential FAs, the content of linoleic acid (1.5%–2.2%) was three times higher than the levels (0.4%–0.7%) of α-linoleic acid while the metabolite of linoleic acid, dihomo-γ-linoleic acid, where levels of up to 1% were found (Fig. 1).

Many previous studies have indicated that the ω-6 to ω-3 PUFA ratio in the human diet evolved from 1:1 to the current 20:1. While increased amounts of ω-6 PUFAs can cause the pathogenesis of many diseases, high levels of ω-3 PUFAs show suppressive effects [15]. In this study, the ω-6 to ω-3 ratio ranged from between 0.025 to 22.63 (average value = 1.80) and showed pronounced maximums in sardines and chub mackerel. As the results demonstrated, small pelagic fish species are a good source of FAs and their nutritional benefits are comparable to those of Atlantic salmon [14] or Atlantic Bluefin tuna [19].

3.2 Pollutant Toxicological Profile

Small edible pelagic fish live briefly at the bottom of the marine food chain (plankton<sardine species, anchovy<mackerel species), and are expected to bioaccumulate low levels of environmental contaminants. The transfer of accumulated toxins to the human body via fish intake is generally considered to be low. As the results of the study indicated, inorganic compounds, macro-elements, and toxic heavy metals, were notably more abundant in fish tissue than organic xenobiotics, OCPs and PCBs (Fig. 1).

The biological effects of inorganic elements depend on processes which include absorption, accumulation, elimination, and biotransformation into less or more toxic metabolites. Several minutes after absorption in the gut, the elements reach the internal organs such as the heart, liver, kidney, and the brain, while their penetration into the muscles and adipose tissue may take up to several hours [1]. However, when they are present in low concentrations in aquatic environments, their accumulation is also less common. In the studied fish species, the concentrations of macro- (K > Na > Ca > Mg) and micro-elements (Fe > Zn > Cu) were within the normal physiological range, below levels that could potentially cause pathological changes in tissues and organs, because of the absence of leading sources of pollution. In addition, levels of the following elements: Hg, Pb and Cd, were lower than the maximum thresholds prescribed by existing EU regulations [6].

Among POPs, *p,p'*-DDE, PCB-153, PCB-138 and PCB-180 were the most prevalent. Although it is expected that larger, longer-lived species that are at higher trophic levels, such as chub mackerel and horse mackerel, will uptake more organic contaminants than sardine species and/or anchovy. Toxicological parameter values, such as the total quantity of indicator PCBs and WHO-dioxin-like PCBs toxic equivalents, were below the threshold concentrations of 75 ng g^{-1} w.w. and 6.5 pg g^{-1} [7]. Therefore, it can be concluded that the examined fish species appear to be safe for human consumption.

3.3 PCB-138 Patterns

As demonstrated by the Pearson's correlation analysis (Fig. 2), significant linear correlation coefficients (r>0.90) were found between the following pairs of investigated variables: *p,p'*-DDE-PCB-118; *p,p'*-DDE-PCB-138; *p,p'*-DDD-PCB-105; *p,p'*-DDD-PCB-180; PCB-138-PCB-118; PCB-138-PCB-153; PCB-153-PCB-118; PCB-153-PCB-170 and PCB-156-PCB-180.

The correlations between pairs of POPs: γ-HCH-PCB-170; γ-HCH-PCB-156; γ-HCH-PCB-118; γ-HCH-PCB-114; γ-HCH-PCB-180; *p,p'*-DDE-PCB-153; *p,p'*-DDT-PCB-170; *p,p'*-DDT-PCB-156; *p,p'*-DDT-PCB-118; *p,p'*-DDT-PCB-153; PCB-138-PCB-170; PCB-180-PCB-170; PCB-180-PCB-156; PCB-105-PCB-180; PCB-105-PCB-156 and PCB-105-PCB-170 ranged from 0.80 to 0.90. The listed

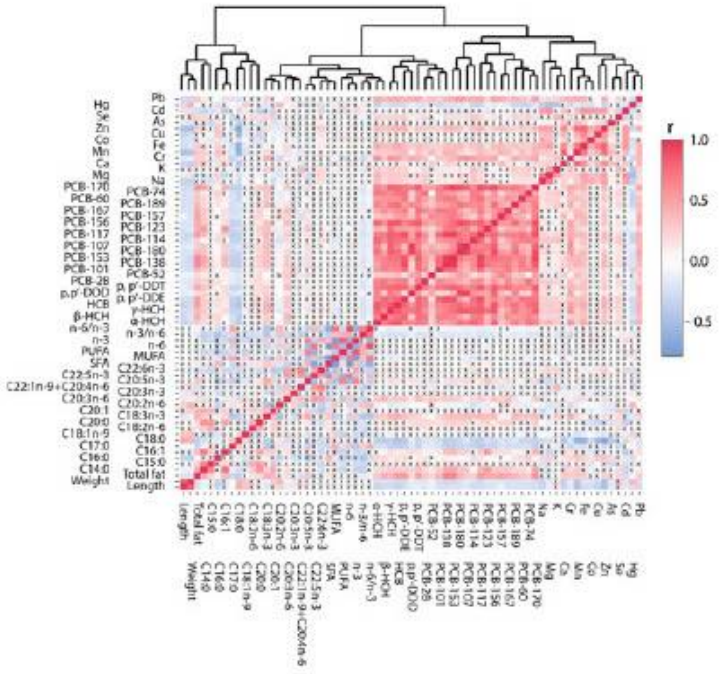


Fig. 2 Parameter value correlations matrix

species with similar chemical structures and common origins display similar behavior patterns, as discussed below. Since OCPs and PCBs are highly liphophilic chemicals, their toxicodynamics are considered to be dependent on the amount of adipose tissue and various fatty compounds in living organisms. No significant Pearson's correlation coefficients were observed between the 18 FAs and the studied POPs(Fig. 2). Therefore, we assume that methods other than the typical correlation matrices could be utilized to describe the associations between POPs and FAs or inorganic contaminants in more detail.

In this study, XGBoost was utilized for the investigation of non-linear relationships between PCB-138 levels and key variables that shape its behavior patterns in marine fish tissue. The predicted/observed calculated relative error was 11.2%, while the r2 was 0.99 (Fig. 3).

The fuzzy clustering of SHAP values extracted ten groups (relative error < 10%) of similar variables that shape the dynamics of PCB-138 (Fig. 4). The constituents of six clusters, C2 and C6-C10, had a highly positive impact on PCB-138 concentrations from the mean to maximum values (0.6–1.9 ng g^{-1}). Prominent impacts, described by relative SHAP values (%), are attributed to the following compounds: *p,p'*-DDE (C2: 23, C6: −10, C7: 33, C8: 25, C9: 31, and C10: 23), PCB-153 (C2: 39, C6: 36, C7: 13, C8: 34, C9: 25, and C10: 28), PCB-118 (C2: 8, C6: 13, C7: 10, C8: 7.8, C9: 10, and C10: 12), PCB-180 (C2: 6.3, C6: 5.3, C9: 3.9, and C10: 5.0), and PCB-170 (C2: 6.0, C6: 4.3, C8: 6.6, C9: 5.0, and C10: 6.2). Less significant effects

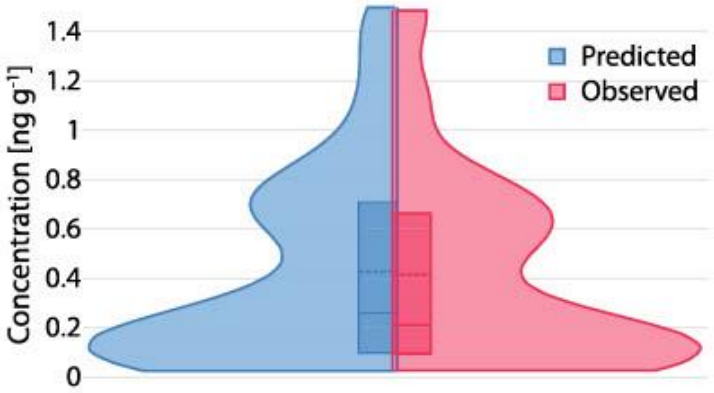


Fig. 3 Model evaluation

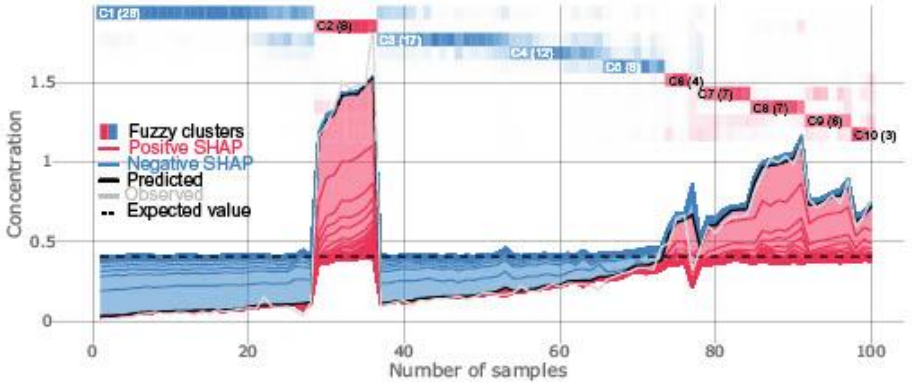


Fig. 4 SHAP value force plot

were recorded for PCB-101 (C2: 1.2 and C9: 2.1) and PCB-123 (C2: 1.8, C8: 1.8, and C9: 2.0). The average and high PCB-138 concentrations were accompanied by levels (ng g^{-1}) of listed pollutants, which ranged from the median and 75th percentile values to the upper fence values: *p,p'*-DDE (C2: 1.3, C6: 0.3, C7: 0.7, C8: 0.9, C9: 0.8, and C10: 0.5), PCB-153 (C2: 2.5, C6: 1.4, C7: 1.0, C8: 1.7, C9: 1.4, and C10: 1.2), PCB-118 (C2: 0.4, C6: 0.2, C7: 0.2, C8: 0.2, C9: 0.2, and C10: 0.1), PCB-180 (C2: 0.8, C6: 1.0, C9: 0.4, and C10: 0.3), PCB-170 (C2: 0.4, C6: 0.3, C8: 0.1, C9: 0.2, and C10: 0.2), PCB-101 (C2: 0.2 and C9: 0.1), and PCB-123 (C2: 0.4, C8: 0.2, and C9: 0.2). Although indicator congeners (−28, −52, −101, −138, −153, and −180), which are often classified as non-dioxin-like PCBs, overshadowed dioxin-like PCBs, both classes were significant in the evaluation of PCB-138 patterns. The relationships between PCB congeners and their presence at higher concentration levels are due to their rigid molecular structure and number of attached halogen atoms, and consequently, their persistence and prolonged half-lives in marine environments.

The more chlorinated congeners are also more resistant to endogenous metabolic degradation in fish in comparison with less chlorinated congeners.

In addition, similar pollution sources contribute to the relationships between "heavy" long-lived penta- to hepta-chloro congeners. The studied fish species live in the epipelagic zone of the Adriatic Sea, which extends from the surface to 200 m depth and shows similar trophodynamics. Although certain differences in nutrition cycles were reported between the species investigated in this study, the main prey of all small pelagic fish include: calanoid copepods, fish larvae, decapod larvae, copepod eggs, amphipods and isopods [10]. Both the overlapping diet, which contains low-niche marine organisms, and inputs from the surrounding environment contributed to the POP burden in the fish samples. In addition, the presence of *p,p'*-DDE reflects the metabolic degradation of toxicology relevant persistent xenobiotics.

Of the 17 examined FAs, saturated myristic (C14:0) acid impacts the PCB-138 bioaccumulation patterns represented in C2 and C6-C10 the most, followed by the nutritionally beneficial eicosadienoic (C20:2n-6) and dihomo-γ-linolenic (C20:2n-3) acids. The interrelations between PCB-138, and the SFA and PUFAs were indicated by relative SHAP values (%): C14:0 (C7: −0.6, C9: −0.4, and C10: −0.5), C20:2n-6 (C6: −0.8, C8: −1.5, and C10: −1.2), and C20:3n-6 (C10: 1.3). We have noted that these FAs were not the most abundant in the studied fish species, and the prevalence of other FAs was recorded in the following order: palmitic > oleic > docosahexaenoic > stearic > myristic > eicosapentaenoic > palmitoleic > linoleic acid. Saturated acids such as C14:0 predominantly occurred as esterified polar phospholipids and therefore, SFAs are more bioavailable than FAs in non-esterified free form. However, negative relationships indicated by relative SHAP values imply that low content (%) of FAs (C14:0-C7: 6.5, C9: 6.1, and C10: 8.0; C20:2n-6-C6: 0.4, C8: 0.7, and C10: 0.8; and C20:3n-6-C10: 0.07) do not increase PCB-138 uptake.

Eicosapentaenoic acid and DHA are widely known as the most nutritionally relevant ω-3 PUFAs in oily blue fish such as mackerel, sardines, and anchovies, but they appear to have no impact on PCB-138 accumulation. Chub mackerel has been known to contain high percentages of PUFAs in the bioavailable phospholipid form, which is of nutritional and physiological relevance [17]. Therefore, consumption of the studied fish species, which ensure a continuous supply of dietarily beneficial levels of EPA and DHA, does not coincide with POPs burden in humans.

Fuzzy clustering identified four clusters (C1, C3–C5) representing groups of the variable interrelations associated with low to minimal levels of PCB-138, from 0.013 to 0.1965 ng g^{-1} (Fig. 4). The same parameters as discussed above shaped the clusters, but their negative influences on PCB-138 distribution were recorded: *p,p'*-DDE (C1: −22, C3: -25, C4: −25, and C5: −27), PCB-153 (C1: −42, C3: −36, C4: −30, and C5: −8.5), PCB-118 (C1: −7.8, C3: −8.4, C4: −8.7, and C5: −5.4), PCB-180 (C1: −6.2 and C3: −5.2), PCB-170 (C1: −7.7, C3: −6.7, and C4: −7.8), PCB-101 (C3: −1.0 and C5: −1.6), and PCB-123 (C1: −1.8, C3: −2.0, and C4: −2.3). Low concentrations of PCB-138 were related with POPs levels, which ranged from the mean to minimum values (ng g^{-1}): *p,p'*-DDE (C1: 0.056, C3: 0.117, C4: 0.023, and C5: 0.212), PCB-153 (C1: 0.125, C3: 0.279, C4: 0.408, and C5: 0.660), PCB-118 (C1: 0.023, C3: 0.041, C4: 0.057, and C5: 0.079), PCB-180 (C1: 0.031 and C3:

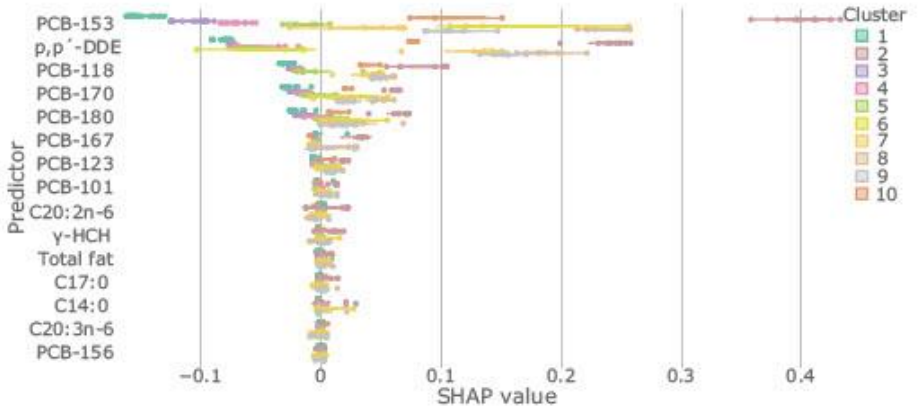


Fig. 5 SHAP value top ten clusters

0.055), PCB-170 (C1: 0.028, C3: 0.047, and C4: 0.071), PCB-101 (C3: 0.026 and C5: 0.030), and PCB-123 (C1: 0.023, C3: 0.034, and C4: 0.048). The results indicate that the uptake of POPs from different sources is found to have variable impacts depending on the sampling time, season, and other fishery zone-related factors.

As discussed above, the negative impact of myristic (C14:0) and margaric (C17:0) acids was also evident in clusters C1, C3–C5. The negative relative SHAP values, %, (C14:0-C1: −0.54 and C5: −0.85; C17:0-C4: −0.76 and C5: −1.7) were accompanied by the following concentrations (%): C14:0 (C1: 3.8 and C5: 5.5) and C17:0 (C4: 1.3 and C5: 1.3).

The plot in Fig. 5 shows the impact each cluster has on the output of the model. High levels of the constituents of clusters (C2, C6–C10) primarily PCB-153, *p,p'*-DDE, PCB-118, PCB-170, PCB-180 and myristic acid, had a high and positive impact on PCB-138 patterns, as shown by the right-oriented long distribution tail. Lower concentrations of these variables, grouped in C1 and C3–C5 clusters, are negatively correlated with the target variable, as indicated by the negative relative SHAP value. To a lesser extent, the influences are also of importance for bioaccumulation fate of *p,p'*-DDE, PCB-118, PCB-170, and PCB-180, while their relevance is weaker for describing environments which shape the occurrence of volatile POPs such as γ-HCH and FAs.

4 Conclusion

Small edible fish species can be considered both a nutritionally valuable food source and a source of hazardous organochlorine pollutants which negatively affect human health. Significant indications of the relationships between the uptake of POPs and fatty acid in fish tissue have been reported worldwide. In this study, we presented a promising methodology, explainable artificial intelligence methods (XGBoost and

SHAP), with the aim of gaining a better understanding of the specific interrelations between the fatty acid content and contaminants in consumable marine fish species. Out of 17 fatty acids, two saturated (myristic and margaric) and two ω-3 and 6 (eicosadienoic and dihomo-γ-linolenic) acids were identified as crucial for the bioaccumulation of PCB-138 in sardine, anchovy, and mackerel species. However, nutritionally beneficial EPA and DHA are assumed to have no impact on the uptake of contaminants. The content of macro-elements and heavy metals is not related to PCB-138 chemodynamics in fish tissue, while the influences of *p,p'*-DDE and both indicator and toxicological congeners (−101, −118, −123, −153, and −180) were evident. Finally, the methods have been successfully verified as a reliable means for examining the relationships between POPs and FAs that improve upon commonly employed statistical approaches.

Acknowledgements The authors acknowledge the funding provided by the Institute of Physics, Belgrade, through research supported by the Ministry of Education, Science and Technological Development and the Science Fund of the Republic of Serbia, #GRANT No. 6524105, AI - ATLAS. In addition, this study was supported by the institutional financing of scientific activity 2018-2020 Project "Persistent Organic Pollutants-Environmental Impact Assessment and Stability of Human Genetic Material" (IMROH, 2018–2021, Zagreb, Croatia).

References

1. Brucka-Jastrzêbska, E., Kawczuga, D., Rajkowska, M., Protasowick, M.: Levels of microelements (Cu, Zn, Fe) and macroelements (Mg, Ca) in freshwater fish. J. Elementol. **14**, 437–447 (2009)
2. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 85–794 (2016)
3. Cook, C.M., Hallaråker, H., Sæbø, P.C., et al.: Bioavailability of long chain omega-3 polyunsaturated fatty acids from phospholipid-rich herring roe oil in men and women with mildly elevated triacylglycerols. Prostaglandins Leukot Essent Fatty Acids **111**, 17–24 (2016)
4. Davidson, M.H.: Omega-3 fatty acids: new insights into the pharmacology and biology of docosahexaenoic acid, docosapentaenoic acid, and eicosapentaenoic acid. Curr. Opin. Lipidol. **24**, 467–474 (2013)
5. Đinović-Stojanović, J., Nikolić, D., Vranić, D., Babić, J., Milijašević, M., Pezo, L., Janković, S.: Zinc and magnesium in different types of meat and meat products from the Serbian market. J. Food Compost. Anal. **59**, 50–54 (2017)
6. European Commission: Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs 20.12.2006 Off J Eur Union L 364, 5-24 (2006)
7. Commission, European: Commission Regulation (EU) No 1259/2011 of 2 December 2011 amending Regulation (EC) No 1881/2006 as regards maximum levels for dioxins, dioxin-like PCBs and non-dioxin-like PCBs in foodstuffs. Off. J. Eur. Union L **320**, 18–23 (2011)
8. Food and Agriculture Organization of the United Nations, FAO: The state of world fisheries and aquaculture worldwide 2018-Meeting the sustainable development goals. 1-227 (2018)
9. Food and Agriculture Organization of the United Nations, FAO: The State of Mediterranean and Black Sea Fisheries. General Fisheries Commission for the Mediterranean 172, Licence: CC BY-NC-SA 3.0 IGO (2018)

10. Hure, M., Mustać, B.: Feeding ecology of Sardina pilchardus considering co-occurring small pelagic fish in the eastern Adriatic Sea. Mar. Biodivers. **50**, 40 (2020)
11. Jump, D.B., Depner, C.M., Tripathy, S.: Omega-3 fatty acid supplementation and cardiovascular disease. J. Lipid Res. **53**, 2525–2545 (2012)
12. Kljaković-Gašpić, Z., Herceg Romanić, S., Klinić, D., Tiina, V.: Chlorinated compounds in the muscle tissue of fish from the Croatian Adriatic: preliminary data on contamination. Arh. Hig. Rada Toksikol. **66**, 299–308 (2015)
13. Lundberg, S., Lee, S.: A unified approach to interpreting model predictions. Adv. Neural Inf .Process Syst. **4765–4774** (2017)
14. Lundebye, A.K., Lock, E.J., Rasinger, J.D., et al.: Lower levels of Persistent Organic Pollutants, metals and the marine omega 3-fatty acid DHA in farmed compared to wild Atlantic salmon (Salmo salar). Environ. Res. **155**, 49–59 (2017)
15. Nasir, M., Bloch, M.H.: Trim the fat: the role of omega-3 fatty acids in psychopharmacology. Ther. Adv. Psychopharmacol. **9**, 2045125319869791 (2019)
16. Ramírez, M., Amate, L., Gil, A.: Absorption and distribution of dietary fatty acids from different sources. Early Hum. Dev. **65**, S95–S101 (2001)
17. Rincón-Cervera, M.Á., González-Barriga, V., Valenzuela, R., López-Arana, S., Romero, J., Valenzuela, A.: Profile and distribution of fatty acids in edible parts of commonly consumed marine fishes in Chile. Food Chem. **274**, 123–129 (2019)
18. Špirić, A., Trbović, D., Vranić, D., Đinović, J., Petronijević, R., Matekalo-Sverak, V.: Statistical evaluation of fatty acid profile and cholesterol content in fish (common carp) lipids obtained by different sample preparation procedures. Anal. Chim. Acta **672**, 66–71 (2010)
19. Sprague, M., Dick, J. R., Medina, A., Tocher, D. R., Bell, J. G., Mourente, G.: Lipid and fatty acid composition, and persistent organic pollutant levels in tissues of migrating Atlantic bluefin tuna (Thunnus thynnus, L.) broodstock. Environ. Pollut. **171**, 61–71 (2012)
20. Stockholm Convention: Stockholm Convention for Persistent Organic Pollutants (2001). http://chm.pops.int/. Cited November 2020
21. Stojić, A., Mustać, B., Jovanović, G.: Explainable machine learning prediction of PCB-138 behavior patterns in edible fish from Croatian Adriatic. In: Book of Proceedings of International Scientific Conference on Information Technology and Data Related Research (Sinteza 2020), Singidunum University, pp. 23–28 (2020)
22. Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L.: Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Total Environ. **653**, 140–147 (2019)
23. Vuković, G., Herceg Romanić, S., Babić, Ž., Mustać, B., Štrbac, M., Deljanin, I., Antanasijević, D.: Persistent organic pollutants (POPs) in edible fish species from different fishing zones of Croatian Adriatic. Marine Pollut. Bull. **137**, 71–80 (2018)
24. World Health Organisation, WHO: Dioxins and their effects on human health. In: World Health Organization Fact Sheet, No. 225 (2014)

What Information on Volatile Organic Compounds Can Be Obtained from the Data of a Single Measurement Site Through the Use of Artificial Intelligence?

Svetlana Stanišić, Mirjana Perišić, Gordana Jovanović, Dimitrije Maletić, Dušan Vudragović, Ana Vranić, and Andreja Stojić

Abstract Increasing air pollutant concentrations over the last few decades have been a focus of contemporary scientific research due to adverse effects on public health, the environment and climate change. In this chapter, we used an innovative integrated methodology for spatio-temporal characterization of sources and concentration forecasts of toxic, mutagenic and carcinogenic representatives of volatile organic species—benzene, toluene, ethylbenzene and xylene, commonly referred to as BTEX. The methodology is based on receptor-oriented air circulation modeling and artificial intelligence implemented through machine learning and explainable artificial intelligence methods. The study covered two years of data obtained from a single monitoring station located at 54a Despota Stefana Boulevard (44°49'68" N, 20°28'04" E). This station was selected from the local and state network for air quality monitoring in the territory of Belgrade. The receptor-oriented modeling was effective for classifying sources of BTEX and the assessment of BTEX concentra-

S. Stanišić (✉)
Environment and Sustainable Development, Singidunum University, Belgrade, Serbia
e-mail: sstanisic@singidunum.ac.rs

M. Perišić · G. Jovanović · D. Maletić · D. Vudragović · A. Vranić · A. Stojić
Institute of Physics Belgrade, National Institute of the Republic of Serbia, Environment and Sustainable Development, Singidunum University, Belgrade, Serbia
e-mail: mirjana.perisic@ipb.ac.rs

G. Jovanović
e-mail: gordana.vukovic@ipb.ac.rs

D. Maletić
e-mail: dimitrije.maletic@ipb.ac.rs

D. Vudragović
e-mail: dusan.vudragovic@ipb.ac.rs

A. Vranić
e-mail: ana.vranic@ipb.ac.rs

A. Stojić
e-mail: andreja.stojic@ipb.ac.rs

E. Pap (ed.), *Artificial Intelligence: Theory and Applications*,
Studies in Computational Intelligence 973,
https://doi.org/10.1007/978-3-030-72711-6_12

tions in the Belgrade urban area surrounding the receptor site that was not regularly monitored. The correlations and ratios between BTEX compounds were used for estimating their interrelationships and presence in the air, which contributed to the identification of their origin. Also, this study evaluated the possibilities of BTEX spatio-temporal forecasts based on the integrated methodology. For this purpose, XGBoost was efficient at forecasting BTEX levels, with estimated errors (6–15%) significantly below the uncertainty obtained by conventional models for the evaluation of average annual pollutant concentrations. The results suggest that temperature, wind speed and wind direction represented the main parameters which explain the spatio-temporal distribution of BTEX, while the impact of other factors showed significant variations depending on the locations of the receptor and the compound.

Keywords BTEX · Volatile organic compounds · Machine learning · Explainable artificial intelligence (xai) · XGBoost

1 Introduction

Growing urban populations, economic development, and transport have a significant impact on environmental pollution. Air pollution constitutes a major underestimated cause of non-communicable diseases, being responsible for 19% of all cardiovascular deaths and 23% of lung cancer deaths globally [19]. Around 91% of the world's population lives in areas where air pollution levels exceed the World Health Organization's recommended values [43]. People often perceive air pollution as an issue that affects people living in middle- and low- income countries or people living in megacities. However, it has been estimated that out of 8.8 million global deaths associated with air pollution in 2015, 8% were citizens of high-income countries, mostly those living in small urban areas where the topography and climate contribute to high air pollutant concentrations [3, 7]. Holgate [17] emphasizes that annually, 40,000 excess deaths in the UK can be attributed to low air quality. Society might be much more aware of this issue if the mortality was a result of drinking polluted water. Air pollution in households has decreased since the 1990's as new fuels such as petroleum gas and renewable sources of energy have replaced biomass. In developing countries, residential combustion of solid fuels for cooking and heating remains a significant source of air pollution and a major concern due to its detrimental health effects, particularly in rural areas [9]. In contrast, in developed countries, indoor air pollution is less important because adverse health effects are associated with exposure to outdoor pollutant concentrations that have decreased over previous decades. However, it has also been estimated that the rapid expansion of metropolises, industrial production, increasing pesticide use, toxic chemicals and motor vehicles will offset the effects of air pollution mitigation measures [13].

Among the air pollutants that are of interest for current and future research due to their detrimental effects on both human health and the environment are volatile organic compounds (VOCs). VOCs are a heterogeneous group of organic species

with boiling points <250 °C. Their representatives are benzene and its alkylated derivatives toluene, ethylbenzene and xylene, commonly referred to as BTEX. Over the last few decades in developed countries, reducing BTEX levels has been a challenge [26]. This is due to their abundance, their numerous emission sources, their complex atmospheric chemistry, insufficient funds for the establishment and maintenance of monitoring networks, and the fact that abatement programs might have negative impacts on economic output because they are among the most abundantly produced compounds worldwide. They are used as feedstock for several materials and products upon which modern society has become dependent. These compounds are naturally found in crude oil, while in urban and suburban areas they originate from traffic emissions, commercial and industrial uses of petroleum, gasoline, adhesives, coatings, degreasers, solvents, detergents, explosives, pesticides, resins, ink, paints, and varnishes [30, 33, 35]. With the exception of m-/p-xylene, concentrations of BTEX were mostly reported to be higher in indoor air than outdoor air [4].

The health effects of BTEX are wide-ranging. Studies have shown that BTEX can be found in cord blood and the blood of children and adults, particularly adults that are occupationally exposed [18]. Research has shown that long-term exposure to high benzene concentrations increases the risk of developing malignant blood disorders, while long-term exposure to high toluene concentrations causes renal tubular acidosis [42]. Some studies have shown that after benzene concentrations in industrial and urban areas are reduced, lifetime cancer risk decreases by one order of magnitude [20]. In addition, exposure to ambient levels of BTEX, which in many cases were orders of magnitude below the reference concentrations, can be dangerous. This is particularly true during the susceptibility period when exposure can lead to the disruption of endocrine signaling (which is essential for the growth and development), immune responses, reproduction, cardiovascular function and aging. Therefore, populations in highly industrialized areas, the socioeconomically deprived, as well as children, pregnant women and elderly people, appear to be more susceptible to pollution-related morbidity and mortality [5, 32]. While traffic emissions are known to be the main source of outdoor BTEX, converting to alternative renewable energy sources will not reduce demand for consumer products, and the replacement of benzene with safer alternatives, such as toluene and xylene, might ultimately be a poor solution once long-term toluene and xylene exposure scenarios are considered [4]. According to a United States Environmental Protection Agency [40] report, ethylbenzene is sixth on a list of the top 20 chemicals used in children's products, primarily food packaging, toys, sport equipment, arts, crafts and hobby materials. Toluene, on the other hand, is the seventh most used chemical in consumer products, primarily fuels, paints, and coatings. Although toluene and xylenes are less harmful than benzene, it should be kept in mind that the products of photochemical reactions in which BTEX are involved often have more harmful effects on human health than their precursors [14]. It should be also mentioned that limited ventilation in closed premises can often be the reason why indoor BTEX levels are higher than outdoor levels [29].

Apart from their impacts on human health, BTEX and other VOCs are associated with increases in the oxidation capacity of the atmosphere, as well as with the

generation of secondary pollutants, such as tropospheric ozone, polycyclic aromatic hydrocarbons and secondary aerosol through photochemical reactions [8, 36]. As regards the impact on global warming, not only do volatile species directly and indirectly contribute to climate change, but their emission and fate are expected to be influenced and increased by global warming.

Despite vast changes in the development and integration of different approaches in the field of environmental science, spatio-temporal air pollution modeling remains a challenge. Two main approaches are typically utilized to forecast air quality and to identify the factors that govern the concentrations of certain pollutants. The first approach relies on atmospheric diffusion models, while the second refers to statistical models that capture the essential relationships between the variables [24]. Multidimensionality and the size of data sets, as well as the complexity of air pollutant processes and interactions, set requirements that exceed the capabilities of conventional statistical methods. For this reason, machine learning (ML) methods and explainable artificial intelligence (XAI), subfields of artificial intelligence (AI) that enable automatized big data analysis and the development of learning algorithms, have been introduced into environmental science research. In this paper, we used an innovative and integrated methodology based on artificial intelligence and implemented through ML ana XAI methods for the modeling of spatio-temporal air pollution and the characterization of BTEX sources in a wider region surrounding the receptor site that was not covered by regular monitoring. The obtained BTEX correlations and ratios were used for estimating the interrelationships between the species in the air, while XGBoost was utilized for efficient spatio-temporal BTEX forecasting. The present paper is an extended version of the study presented at the International Scientific Conference on Information Technology and Data Related Research (Sinteza 2020), Singidunum University [38].

2 Materials and Methods

Nowadays, there are a large number of libraries implemented in different programming languages (R, Python, JS) that deal with interactive display (plotly, leaflet) and spatial data analysis. In addition to various spatial autocorrelation possibilities, the analysis of spatial data patterns, the interpolation of data by statistical methods such as Kriging, Spline or Inverse Distance Weighting, and machine learning methods such as Random Forests are increasingly used. Furthermore, spatial distributions of air pollutant sources can be estimated using general and local hybrid receptors, as well as by analyzing clustered data [9, 33, 34].

Significant improvements to the general hybrid receptor modeling approach have been made in recent decades. Receptor-oriented methods, based on conditional probability and analyzing the residence time of pollutants in an area, have become widely accepted both for studying the dynamic processes and circulation patterns of pollutants in the atmosphere and for investigating the spatial distribution of potential

emission sources and their impact on the receptor site without relying on an emission inventory.

The identification of potential emission sources at a local level, as well as their contribution to the measured concentrations of pollutants at the receptor site, can be determined by models analogous to the general hybrid receptor model such as the Potential Source Contribution Function (PSCF), Concentration-Weighted Trajectories (CWT), sQTBA (Simplified Quantitative Transport Bias Analysis) and RTWC (Residence Time Weighted Concentration). The general models, which are based on the analysis of trajectory end points, can be improved by using local wind parameters. This way it is possible to determine pollution circulation patterns very accurately in the area around a given measuring point. A local three-dimensional hybrid receptor model similar to the 3D CWT model was developed for the purposes of this paper [34]. The study involved two-years of regular measurements of BTEX, suspended particles, inorganic gaseous oxides, and meteorological parameters within the Belgrade City Institute for Public Health's automatic monitoring network (44°49'68" N, 20°28'04" E).

Machine learning is an area of artificial intelligence that involves the development of algorithms that can learn based on input data and can thus be trained to predict value variations. Machine learning algorithms are based on the extraction of patterns and the selection of specific attributes from a large number of data, while eliminating irrelevant information. Through identifying most important prediction attributes, machine learning methods acquire knowledge and define the substantial relationships that exist between input and output parameters by focusing on the aspect of the data that is the most useful for efficient forecasting.

Forecasting the concentration dynamics of BTEX in the air was done using the Xtreme Gradient Boosting (XGBoost) method [11], with meteorological data used as predictors. XGBoost is a general-purpose ensemble method of supervised machine learning which combines the results of many decision trees and achieves high accuracy in a wide range of practical applications. It usually outperforms support vector machines, random forests, and deep learning neural networks [31]. The main advantage of the XGBoost method is its obtainment of more precise predictions than those provided by single constitutive decision algorithms. XGBoost is based on a boosting technique that sequentially defines a series of decision trees for classifying input data into two or more attribute-defined classes. Each consecutive decision tree is trained through iterations, taking into account the previously registered classification errors. The datasets from each of the grid cells used for spatio-temporal BTEX forecasting was split into training (80%) and validation (20%) sets. Hyperparameter tuning was implemented using a brute-force grid search and a 10-fold stratified cross-validation. The best performing hyperparameter values were used in the final model.

Methods based on decision trees, such as Gradient Boosting and Random Forests, have been shown to provide inconsistent attribute contributions. This has led to the development of SHAP (SHapley Additive exPlanation), a method that estimates the contribution of each instance of an attribute, which enables interpretation of the model's outputs [15, 21, 31]. To explain the contribution of each feature to the individual XGBoost prediction, the SHAP method was utilized [21]. It is based on

coalitional game theory and provides a distribution of each prediction among the features represented as additive attributions. In this study, we used Python SHAP implementation (SHAP Python package). The captured importance of a feature is visually presented as a SHAP summary plot.

3 Results and Discussion

Previous studies have demonstrated strong gradients and pronounced intra-urban spatial variability of pollutant levels. This depends on traffic density, street configuration and prevailing wind direction [41]. In addition to this, Ning et al. [27] emphasized the importance of terrain complexity for air quality because topographic features can, to a certain extent, limit pollutant dispersion under different weather conditions. The results indicate that BTEX concentrations were significantly higher in areas with dense traffic, in the vicinity of busy streets and intersections, where traffic emission sources have a high impact, where local topography limits natural ventilation and vertical dispersion of pollution, and in areas with narrow streets lined by tall buildings and trees.

3.1 BTEX Levels Surrounding the Receptor Site

Benzene concentrations were estimated to range from 1.2 to 2.6 $\mu g\ m^{-3}$, with an average level of 2.6 $\mu g\ m^{-3}$. In a few locations, estimates of pollutant concentrations were extremely high, exceeding the recommended limit of 5 $\mu g\ m^{-3}$. Extreme benzene levels in areas distant from the receptor site should be taken with caution. This is due to a relatively small number of events and therefore limited data produces the calculated values. It has been estimated that the area surrounding the receptor site was mostly influenced by traffic emissions. Other studies have reported that similar average benzene concentrations are sometimes discovered in urban areas where traffic is the predominant source of pollutant emissions [2]. Concentrations of toluene were estimated to range from 0.9 to 8.9 $\mu g\ m^{-3}$ with an average value of 8.7 $\mu g\ m^{-3}$, while the average m/p-xylene concentrations were assessed to be several times higher (8.4 $\mu g\ m^{-3}$) than the concentrations of o-xylene (1.8 $\mu g\ m^{-3}$). Concentrations of ethylbenzene were estimated to range from 0.2 to 2.1 $\mu g\ m^{-3}$ with an average value of 1.8 $\mu g\ m^{-3}$. As predicted, TEX levels were highest in the city center and the north of the city. The assessed TEX levels were also consistent with the reported values for outdoor TEX concentrations in various studies published over the last two decades [1, 10, 16, 23, 25], with somewhat lower levels of toluene than typically expected in an urban area.

3.2 Seasonal and Daily BTEX Variations

While the intensity of BTEX emission sources tend to be higher in winter, evaporations in warmer parts of the year make an important contribution to total pollutant concentrations. On the other hand, the stability of atmospheric conditions and decrease in chemical reactivity of BTEX in winter results in longer retention in ambient air [29]. Furthermore, BTEX are known to exhibit variations on a daily basis. For instance, a previous study that contained a risk assessment of an accidental benzene release in an urban area using an atmospheric dispersion model, showed that benzene spreads over a much larger area during the nighttime due to a stable boundary layer. In contrast, enhanced vertical mixing results in limited dispersion of the pollutant over the study area during the daytime [39]. Similar results were obtained in our study.

3.3 BTEX Forecasting Based on Meteorological Variables as Predictors

As can be seen in Figs. 1 and 2, high correlation coefficients between predicted and observed values (>0,80) were obtained for most of the analyzed data, it can therefore be concluded that XGBoost was a successful and efficient method for forecasting air pollution in an urban area. It should be emphasized that the estimated method errors (6–15%) were significantly lower than the uncertainty (50%) which is required for the evaluation of average annual pollutant concentrations using conventional modeling.

It should be also emphasized that due to the relatively short atmospheric lifetimes of BTEX, the area that can be affected by BTEX emissions is approx. 15–20 km [29], but the consistency of meteorological conditions significantly affects the extent of volatile pollutant dispersion.

The results of this study suggest that low temperatures and weak to moderate wind represent the main parameters which govern the spatio-temporal distribution of benzene in a majority of locations. The impacts of other factors display significant variation depending on the characteristics of the receptor's location (Fig. 1). Namely, one can note that the horizontal axis marked with wind speed is the longest, which means that benzene concentrations were mostly affected by this parameter. Each axis is composed of a series of points that represent the measured values of the predictor. As can be seen in the example for wind speed, extremely high wind speeds, represented by points located on the far right of the axis, have relatively little impact on benzene concentrations in the air. This suggests that benzene concentrations were mostly influenced by weak to moderate wind, which further suggests that emission sources located in the vicinity of the selected receptor site have the highest impact.

Furthermore, the importance of a certain wind direction for pollutant level prediction is related to the position of the emission source which affects the receptor location the most. Namely, in cases where a single pollutant emission source has a

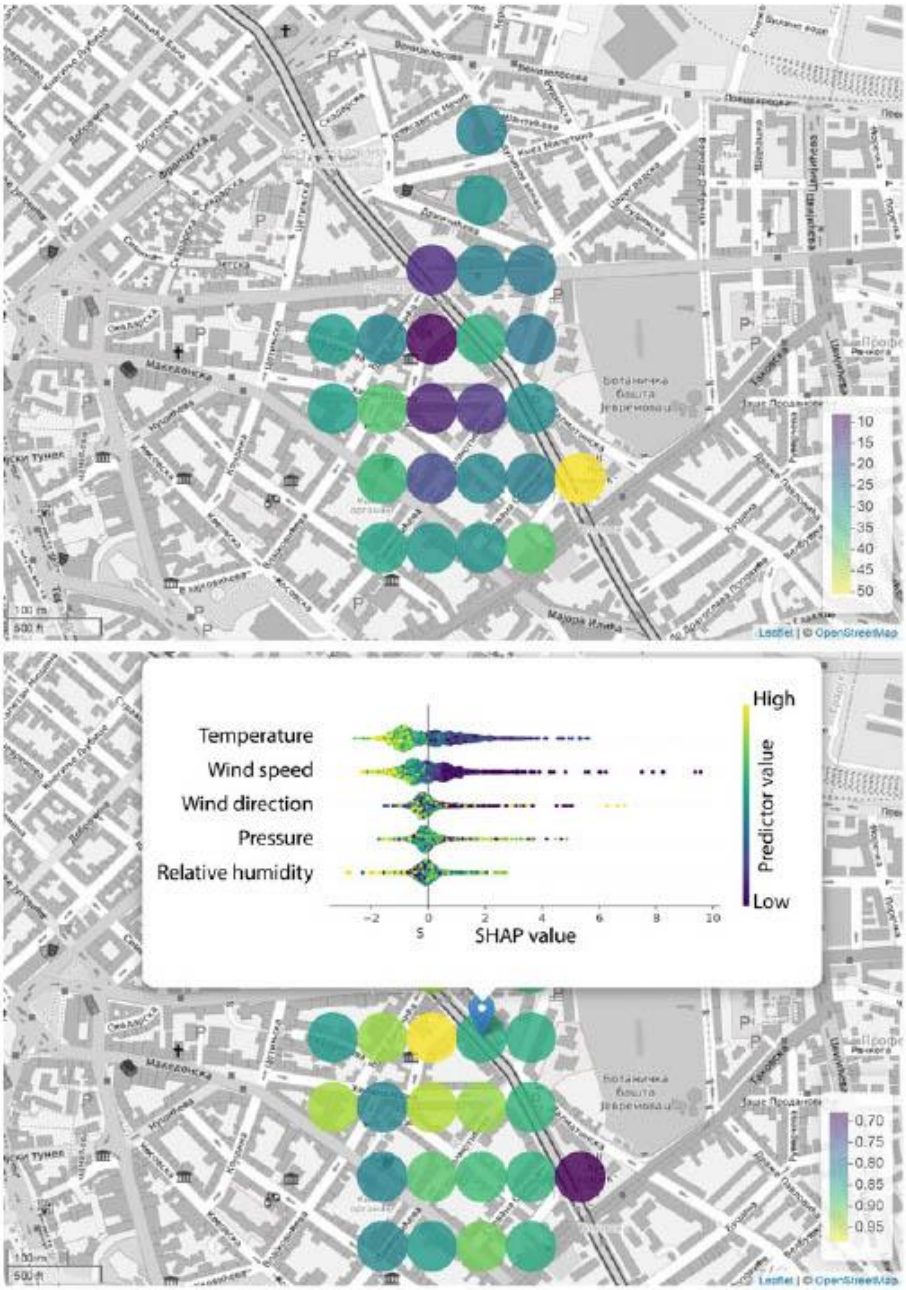


Fig. 1 Benzene forecast based on meteorological parameters—a relative error [%] (top) and SHAP values and predicted/observed correlation coefficients (bottom)

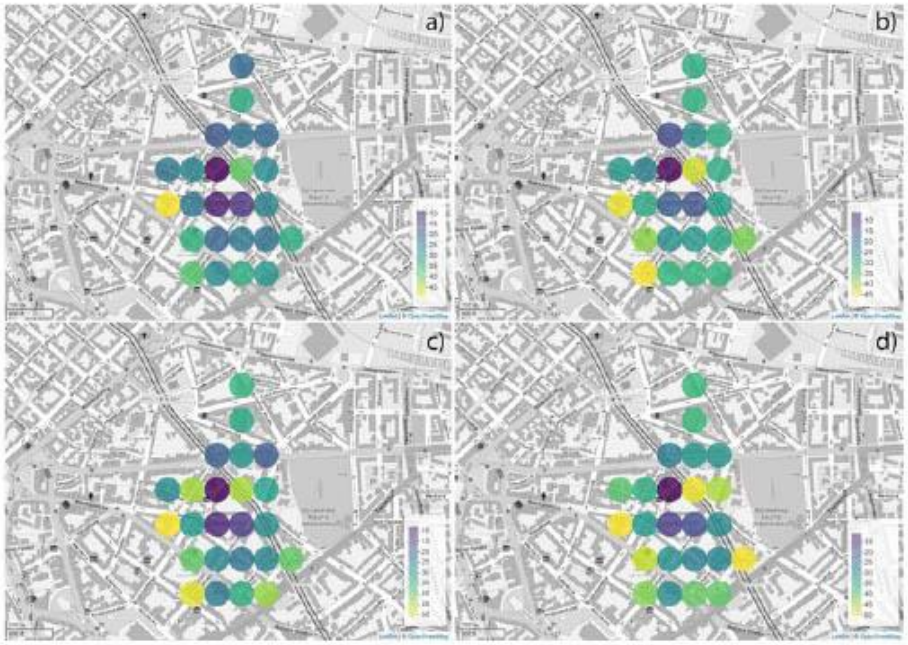


Fig. 2 Toluene (**a**), m,p-xylene (**b**), o-xylene (**c**), and ethylbenzene (**d**) relative error [%] forecasts based on meteorological parameters

major effect on the receptor area, the importance of wind direction as a predicting factor is particularly accentuated. Additionally, the importance of wind direction is evident in situations where tall buildings and trees along the roadside form a corridor that either hinders or assists polluted air masses from the surrounding emission sources, which was the case in this study as well. The importance of relative humidity as predictor for benzene concentrations was predicted to be high in the event of extreme benzene levels. This is because benzene and TEX are partially water-soluble compounds, which suggests that both high and low levels of relative humidity can impact their ambient air levels [28].

The presented figures displaying TEX distributions can be interpreted in the following manner. Atmospheric pressure and wind parameters appear to be the major predictors of TEX behavior in general. The impacts of air temperature and pressure were significantly lower for forecasting toluene levels (Fig. 2) than for forecasting benzene levels, which can be explained by the fact that benzene evaporates at significantly lower temperatures than toluene.

The correlations between the observed and forecasted concentrations of benzene ranged from 0.70 to 0.95, but most often ranged from 0.80 to 0.85. At the location where the correlation was the highest (0.95), benzene concentrations were forecasted with a relative error of 8%, while the most influential SHAP-revealed factor was wind direction. The correlations between the observed and forecasted concentrations of toluene ranged from 0.65 to 0.95, but mostly around 0.80.

At the location where the correlation was the highest (0.95), toluene concentrations were forecasted with a relative error of 9%, while the most influential SHAP-revealed factors were wind speed and direction. The correlations between the observed and forecasted concentrations of m/p-xylenes ranged from 0.65 to 0.95, but averaged approximately 0.75. At the locations where the correlations were highest (0.97 and 0.98), the concentrations of m/p-xylenes were forecasted with relative errors of 6% and 15% respectively, while the most influential SHAP-revealed factors were wind speed, direction, and air pressure. The correlation between the observed and forecasted concentrations of o-xylene ranged from 0.60 to 0.90, with an average of approximately 0.85. At the locations where the correlations were the highest (0.94 and 0.95), o-xylene concentrations were forecasted with relative errors of 12% and 8% respectively, while the most influential SHAP-revealed factors were wind direction and air pressure. The correlations between the observed and forecasted concentrations of ethylbenzene ranged from 0.60 to 0.95, but averaged approximately 0.80. At the location where the correlation was the highest (0.95), ethylbenzene concentrations were forecasted with a relative error of 11%, while the most influential SHAP-revealed factors were wind speed, wind direction, and air pressure.

Although the results demonstrate the capacity of an innovative methodology to identify the importance of certain meteorological factors as predictors of air pollutant concentrations, the fact that variations of meteorological parameters cause changes in other related parameters makes it difficult to distinguish their actual impact on air pollution phenomena. Thus, the impact of meteorological factors should not be observed as the isolated effect of a single parameter and its variations, but as the effect of a certain type of weather. For instance, Liao et al. [22] identified ten typical air circulation types within one of the most polluted areas of China and explored how their synergetic relationship with topography affected the local air quality.

3.4 The Importance of Other Pollutants as Predictors for BTEX Levels

Apart from meteorological parameters, other analyzed factors can be considered important for predicting BTEX concentrations. Namely, for forecasting benzene levels, high CO concentrations appear to be the most important (Fig. 3), with the importance of the other predictors stated here in decreasing order: toluene> ethylbenzene> m/p-xyneles> o-xylene> PM_{10} > NO_x > NO_2 > NO> SO_2.

As regards forecasting toluene levels, high m/p-xylenes, o-xylene and CO concentrations appear be the most important, with the importance of the other predictors stated here in decreasing order: ethylbenzene> benzene> PM_{10} > NO> NO_x > SO_2 > NO_2. As regards forecasting m/p-xylenes levels, high toluene, o-xylene, ethylbenzene and CO concentrations appear to be the most important, with the importance of the other predictors stated here in decreasing order: benzene> NO_2 > NO_x > PM_{10} > SO_2 > NO. As regards forecasting o-xylene levels, high m/p-

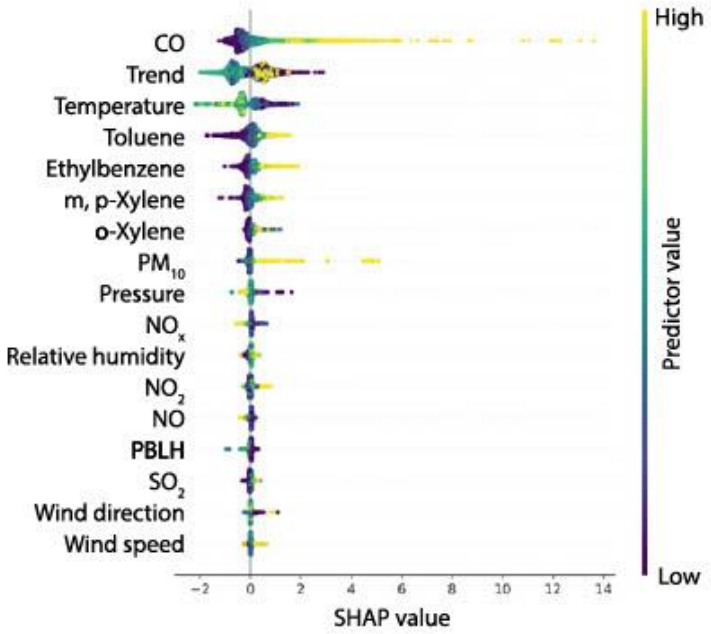


Fig. 3 SHAP summary plot for benzene

xylenes, toluene, ethylbenzene and NO concentrations appear to be the most important, with the importance of the other predictors stated here in decreasing order: NO_x > benzene> CO> PM_{10} > NO_2 > SO_2. As regards forecasting ethylbenzene levels, high m/p-xylenes, o-xylene and benzene concentrations appear to be the most important, with the importance of the other predictors stated here in decreasing order: toluene> NO_2 > SO_2 > CO> PM_{10} > NO> NO_x.

3.5 The Interdependence of BTEX Level Predictors

As part of forecasting pollutant concentrations, an examination of the interdependence between individual predicting factors and their combined effect on BTEX concentrations in the air was performed. As shown in the Fig. 4, during the cold part of the year, when temperatures were below 14 °C, concentrations of benzene are better predicted by toluene levels. Conversely, on days when the temperature exceeded 14 °C, benzene and toluene didn't share the same emission sources. Furthermore, during the cold part of the year, benzene concentrations are either significantly higher or lower than average, depending on whether the location is affected by the burning of fossil fuels for heating or not, while during the warmer part of the year benzene concentrations at different locations tended to be more uniform. This is because higher temperatures cause benzene to evaporate at all locations.

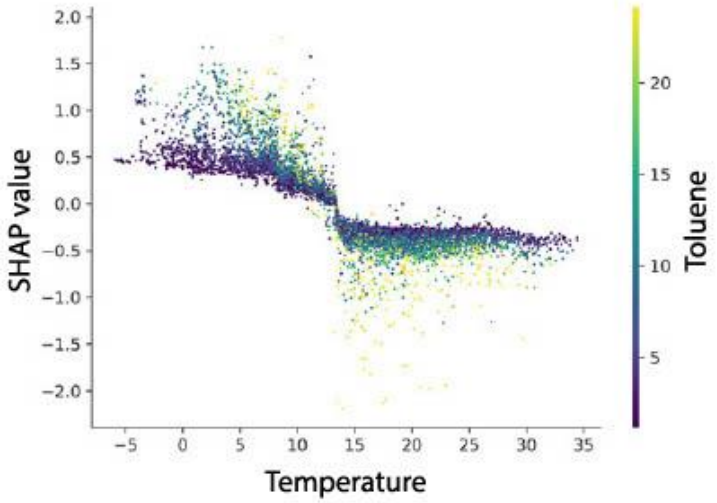


Fig. 4 Benzene SHAP dependency on temperature and toluene

In addition to temperature-toluene interactions, the results suggest that the concentrations of benzene in the ambient air depended on interactions between temperature and carbon monoxide as well. While the effects of interactions between ethylbenzene-m/p-xylenes, ethylbenzene-toluene, ethylbenzene-carbon monoxide and m/p-xylene-carbon monoxide were found to be of minor importance as regards shaping benzene concentrations (Fig. 5).

3.6 Pollutant Correlations and BTEX Origin

The results show moderate correlations between toluene and NO_x and CO concentrations throughout the city. The locations with toluene-NO_x correlations exceeding 0.7 are considered to be affected by burning-related emissions and not by toluene evaporations (Fig. 6).

Relatively high correlations between benzene and carbon monoxide were registered throughout the city, with the notable exception of the south (Fig. 7). The correlations between benzene and SO_2, NO_x and PM_{10} were lower in the eastern, and higher in the western, part of the urban region. These results suggest that benzene in the eastern area of the city is associated with evaporations and emissions from the petrochemical industry such as the Pančevo Oil refinery and the Petrohemija chemical plant. Furthermore, high correlations ($r > 0{,}70$) between benzene and inorganic oxides in the western region of the city may suggest that traffic emissions, as well as remote air pollution sources such as TPP Nikola Tesla A and B in Obrenovac, have a detrimental impact (Fig. 7). The significantly lower correlations between benzene and inorganic oxides, that can be considered indicators of combustion processes in

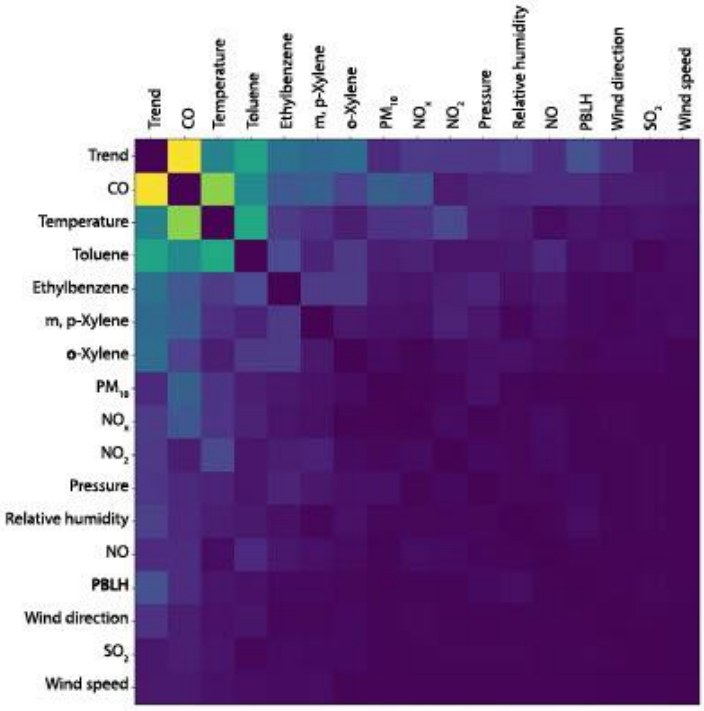


Fig. 5 SHAP interaction plot

the southern part of the city center, probably reflect the impact of gasoline evaporation. This conclusion appears more likely when we take into account that benzene concentrations were assessed to be the lowest in the southern area.

The m/p-xylenes exhibited the highest correlations with NO_x, somewhat lower correlations with CO, while the correlations with SO_2 and PM_{10} tended to be significantly lower throughout the city. A similar calculation was obtained for o-xylene and ethylbenzene.

3.7 BTEX Ratios

The benzene-to-toluene (B/T) ratio is often used as an index for identifying emission sources, while xylenes-to-benzene (X/B) and ethylbenzene-to-benzene (E/B) ratios are generally applied as indices of photochemical reactivity [12]. According to the literature, a B/T ratio below 0.5 suggests that vehicle emissions are the predominant source of BTEX, while X/B and E/B ratios below 1 suggest that sampled air masses are photochemically aged. As the results show, the B/T ratios were approx. 0.3, while the X/B and E/B ratios ranged from 0.7 to 4 with lower ratio values calculated for

Fig. 6 Toluene and NO_x correlations

the central area of the city where narrow streets and tall buildings restrict pollutant dispersion and support the photochemical aging of the emissions. Buczynska et al. [6] showed that B/T ratios ranging from 0.22 to 0.26 for locations that were affected by traffic emissions, while the authors mentioned that ratios below 0.1 reported in previous research suggest additional emission sources of toluene, like industrial emissions, that contribute to high values. It could be concluded from the results that in most locations, BTEX concentrations in ambient air are a result of recent emissions, while traffic emissions might be the major contributor to BTEX concentrations in the analyzed area. Furthermore, BTEX have a similar chemical structure, but there is a difference in the reactivity of these typically unreactive compounds. Namely, benzene is more stable and has a longer atmospheric lifetime [14] which can significantly modify the starting BTEX concentration ratios, and thus another possibility indicated by the B/T ratio is related to aging air masses. Namely, after being emitted from a common source, toluene is about 5 times more reactive than benzene, which is why a high B/T ratio is an indicator of aged air masses. The results suggest that the examined area can be divided diagonally into two segments:

- The north-western and north-eastern area which covers the old city center where the B/T ratio is relatively high. The north-western part is subject to emissions from the industrial area, which includes the petrochemical industry comprised of the Pančevo Oil refinery and the Petrohemija chemical plant. The north-western part is subject to the retention of aged air masses in narrow streets which form an urban

Fig. 7 Benzene correlation coefficients with CO (**a**) and NO_x (**b**)

canyon. The south-western part is subject to the influence of aged air masses from Obrenovac.
- The southern area, covers a large area of the city which displays a relatively low B/T ratio. This indicates the predominant effect of traffic emissions.

Although diagnostic concentration ratios are a commonly used tool for identifying and distinguishing emission sources, they should be used with caution since the calculated values often exhibit seasonal variations and can be susceptible to distortions from a number of environmental factors [37].

4 Conclusions

Identifying pollutant emission sources, investigating the temporal dynamics and spatial distributions of pollutant concentrations, and estimating the contribution of certain emission sources to local air quality are of great importance. They can help to develop an essential understanding of the impact key factors have on processes within natural and anthropogenic ecosystems and their resilience, but are also crucial for defining strategies aimed at improving air quality, resolving environmental issues, and improving human health. While soil and water pollution is mainly of local significance, air pollution can affect very remote areas given meteorological conditions that enable the transportation of pollutants. In addition to this, a number of other factors can contribute to the final effect of meteorological conditions on air pollutant concentrations, including the distribution of pollutant emission sources, local topography, street geometry and the distribution of all elements and surfaces that can be of significance for the air flow regime, pollutant dispersion conditions, their transport pathways and thus, the spatio-temporal variability of their levels. The multidimensionality of modern data, the scope of time series, as well as the complexity of processes and interactions in which air pollutants participate, are too demanding for conventional statistical methods. For this reason, recent decades have seen research conducted to find alternative methods for data analysis. One of the approaches that has proven effective in many professional and scientific fields, including environmental protection, is machine learning which provides the tools for automated analysis of a large amount of data.

As can be concluded, we have demonstrated the use of an efficient methodology for spatio-temporal BTEX concentration modeling in the Belgrade area, based on receptor-oriented air circulation modeling and artificial intelligence implemented through machine learning and explainable artificial intelligence methods. The estimated method errors ranged from 6 to 15%, which is significantly lower than the required uncertainty for conventional models. The presented methodology has the potential to provide a basis for the establishment of a unique and sustainable system for identifying sources of air pollution and enhanced air pollution data coverage that does not require additional investments in monitoring equipment. In the long term, the results of such an approach would provide a solid basis for establishing a sustain-

able system for improving the management and control of air pollution. As the results show, temperature, pressure, wind speed and wind direction were the main parameters which governed the spatio-temporal distribution of BTEX, while the impact of other factors showed significant variations depending on the characteristics of receptor's location and the compound. In addition, spatial correlations and ratios between different air pollutant concentrations were considered for determining their origin in all the locations covered by our analysis. All examples illustrate how the application of SHAP and other supplemental methods can provide a systematic insight into the impact of emission sources and environmental factors on the presence of BTEX in the air.

Acknowledgements Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through a grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI - ATLAS, and Green Fund of the Ministry of Environmental Protection of the Republic of Serbia (No. 401-00-1219/2018-05).

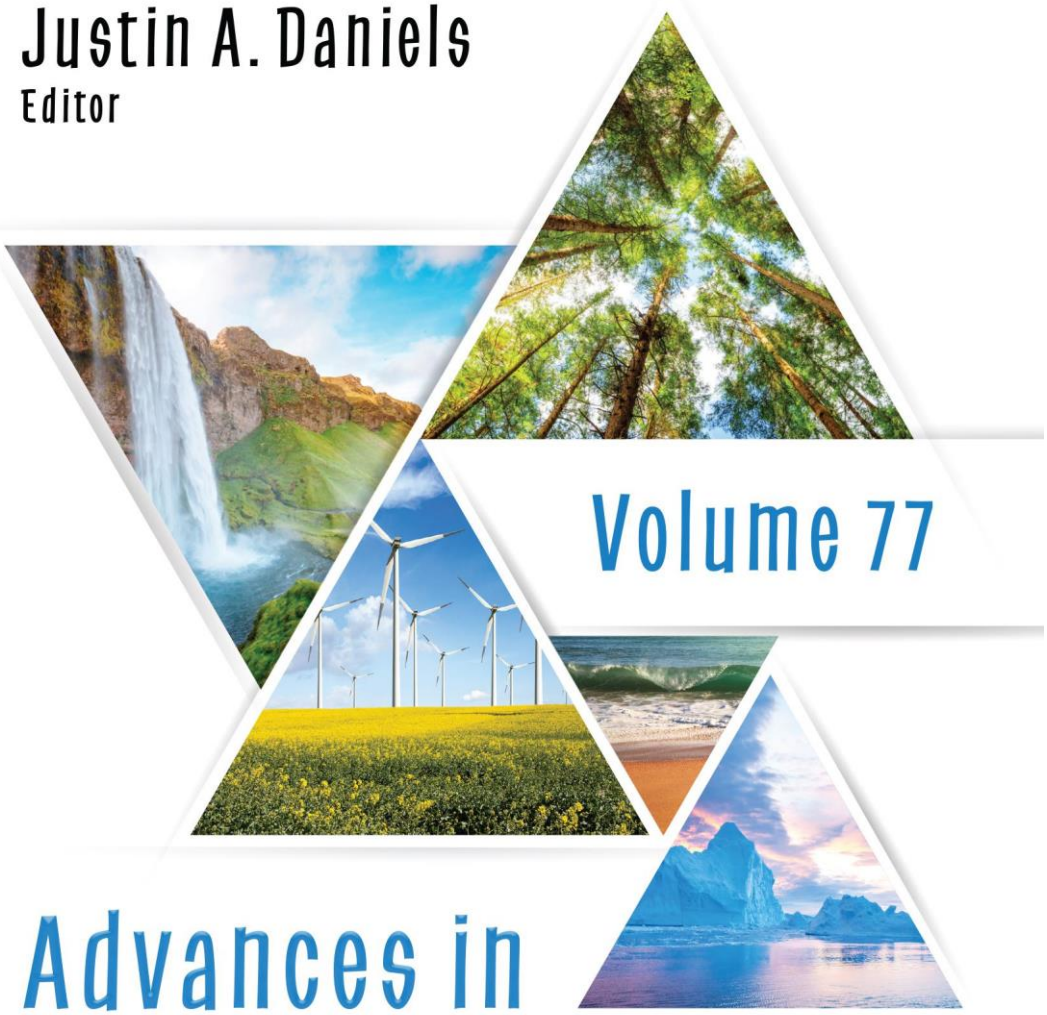
References

1. Abd Hamid, H., Latif, M.T., Nadzir, M.S.M., Uning, R., Khan, M.F., Kannan, N.: Ambient BTEX levels over urban, suburban and rural areas in Malaysia. Air Qual. Atmos. Health **12**, 341–351 (2019)
2. Baltrėnas, P., Baltrėnaitė, E., Šerevičienė, V., Pereira, P.: Atmospheric BTEX concentrations in the vicinity of the crude oil refinery of the Baltic region. Environ. Monit. Assess. **182**, 115–127 (2011)
3. Behrens, D.A., Koland, O., Leopold-Wildburger, U.: Why local air pollution is more than daily peaks: modelling policies in a city in order to avoid premature deaths. Cent. Eur. J. Oper. Res. **26**, 265–286 (2018)
4. Bolden, A.L., Kwiatkowski, C.F., Colborn, T.: New look at BTEX: are ambient levels a problem? Environ. Sci. Technol. **49**, 5261–5276 (2015)
5. Bose, S., Diette, G.B.: Health disparities related to environmental air quality. In: Health Disparities in Respiratory Medicine, pp. 41–58. Humana Press, Cham (2016)
6. Buczynska, A.J., Krata, A., Stranger, M., Godoi, A.F.L., Kontozova-Deutsch, V., Bencs, L., Naveau, I., Roekens, E., Van Grieken, R.: Atmospheric BTEX-concentrations in an area with intensive street traffic. Atmos. Environ. **43**, 311–318 (2009)
7. Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C.A., Apte, J.S., Brauer, M., Cohen, A., Weichenthal, S., Coggins, J.: Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. PNAS USA **115**, 9592–9597 (2018)
8. Campbell, P., Zhang, Y., Yan, F., Lu, Z., Streets, D.: Impacts of transportation sector emissions on future US air quality in a changing climate. Part II: Air quality projections and the interplay between emissions and climate change. Environ. Pollut. **238**, 918–930 (2018)
9. Carter, E., Archer-Nicholls, S., Ni, K., Lai, A.M., Niu, H., Secrest, M.H., Sauer, S.M., Schauer, J.J., Ezzati, M., Wiedinmyer, C., Yang, X.: Seasonal and diurnal air pollution from residential cooking and space heating in the Eastern Tibetan Plateau. Environ. Sci. Technol. **50**, 8353–8361 (2016)
10. Cerón Bretón, J.G., Cerón Bretón, R.M., Martínez Morales, S., Kahl, J.D., Guarnaccia, C., Lara Severino, R.D.C., Rangel Marrón, M., Ramírez Lara, E., Espinosa Fuentes, M.D.L.L., Uc Chi, M.P., Sánchez, G.L.: Health risk assessment of the levels of BTEX in ambient air of one

urban site located in Leon, Guanajuato, Mexico during two climatic seasons. Atmosphere **11**, 165 (2020)
11. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
12. Dehghani, M., Fazlzadeh, M., Sorooshian, A., Tabatabaee, H.R., Miri, M., Baghani, A.N., Delikhoon, M., Mahvi, A.H., Rashidi, M.: Characteristics and health effects of BTEX in a hot spot for urban pollution. Ecotoxicol. Environ. Saf. **155**, 133–143 (2018)
13. Fowler, D., Brimblecombe, P., Burrows, J., Heal, M.R., Grennfelt, P., Stevenson, D.S., Jowett, A., Nemitz, E., Coyle, M., Lui, X., Chang, Y.: A chronology of global air quality. Philos. Trans. R. Soc. A **378**, 20190314 (2020)
14. Gallego, E., Roca, F.X., Guardino, X., Rosell, M.G.: Indoor and outdoor BTX levels in Barcelona city metropolitan area and Catalan rural areas. J. Environ. Sci. **20**, 1063–1069 (2008)
15. García, M.V., Aznarte, J.L.: Shapley additive explanations for NO2 forecasting. Ecol. Inform. **56**, (2020)
16. Hajizadeh, Y., Mokhtari, M., Faraji, M., Mohammadi, A., Nemati, S., Ghanbari, R., Abdolahnejad, A., Fard, R.F., Nikoonahad, A., Jafari, N., Miri, M.: Trends of BTEX in the central urban area of Iran: a preliminary study of photochemical ozone pollution and health risk assessment. Atmos. Pollut. Res. **9**, 220–229 (2018)
17. Holgate, S.T.: Every breath we take: the lifelong impact of air pollution-a call for action. Clin. Med. **17**, 8 (2017)
18. Kirman, C.R., Aylward, L.L., Blount, B.C., Pyatt, D.W., Hays, S.M.: Evaluation of NHANES biomonitoring data for volatile organic chemicals in blood: application of chemical-specific screening criteria. J. Expo. Sci. Environ. Epidemiol. **22**, 24–34 (2012)
19. Landrigan, P.J.: Air pollution and health. Lancet Public Health **2**, e4–e5 (2017)
20. Lerner, J.E.C., Kohajda, T., Aguilar, M.E., Massolo, L.A., Sánchez, E.Y., Porta, A.A., Opitz, P., Wichmann, G., Herbarth, O., Mueller, A.: Improvement of health risk factors after reduction of VOC concentrations in industrial and urban area. Environ. Sci. Pollut. Res. **21**, 9676–9688 (2014)
21. Lundberg S., Lee, S.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, pp. 4765–4774 (2017)
22. Liao, Z., Gao, M., Sun, J., Fan, S.: The impact of synoptic circulation on air quality and pollution-related human health in the Yangtze River Delta region. Sci. Total Environ. **607**, 838–846 (2017)
23. Liu, A., Hong, N., Zhu, P., Guan, Y.: Understanding benzene series (BTEX) pollutant load characteristics in the urban environment. Sci. Total Environ. **619**, 938–945 (2018)
24. Ly, H.B., Le, L.M., Phi, L.V., Phan, V.H., Tran, V.Q., Pham, B.T., Le, T.T., Derrible, S.: Development of an AI model to measure traffic air pollution from multisensor and weather data. Sensors **19**, 4941 (2019)
25. Marć, M., Namieśnik, J., Zabiegała, B.: BTEX concentration levels in urban air in the area of the Tri-City agglomeration (Gdansk, Gdynia, Sopot), Poland. Air Qual. Atmos. Health **7**, 489–504 (2014)
26. Milazzo, M.J., Gohlke, J.M., Gallagher, D.L., Scott, A.A., Zaitchik, B.F., Marr, L.C.: Potential for city parks to reduce exposure to BTEX in air. Environ. Sci. Process. Impacts **21**, 40–50 (2019)
27. Ning, G., Yim, S.H.L., Wang, S., Duan, B., Nie, C., Yang, X., Wang, J., Shang, K.: Synergistic effects of synoptic weather patterns and topography on air quality: a case of the Sichuan Basin of China. Clim. Dyn. **53**, 6729–6744 (2019)
28. Rattanajongjitrakorn, P., Prueksasit, T.: Temporal variation of BTEX at the area of petrol station in Bangkok, Thailand. APCBEE Procedia **10**, 37–41 (2014)
29. Słomińska, M., Konieczka, P., Namieśnik, J.: The fate of BTEX compounds in ambient air. Crit. Rev. Environ. Sci. Technol. **44**, 455–472 (2014)
30. Stingone, J.A., McVeigh, K.H., Claudio, L.: Early-life exposure to air pollution and greater use of academic support services in childhood: a population-based cohort study of urban children. Environ. Health **16**, 2 (2017)

31. Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L.: Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Total Environ. **653**, 140–147 (2019)
32. Stojić, A., Stojić, S.S., Mijić, Z., Šoštarić, A., Rajšić, S.: Spatio-temporal distribution of VOC emissions in urban area based on receptor modelling. Atmos. Environ. **106**, 71–79 (2015b)
33. Stojić, A., Stojić, S.S., Šoštarić, A., Ilić, L., Mijić, Z., Rajšić, S.: Characterization of VOC sources in urban area based on PTR-MS measurements and receptor modelling. Environ. Sci. Pollut. Res. **22**, 13137–13152 (2015a)
34. Stojić, A., Stojić, S.S.: The innovative concept of three-dimensional hybrid receptor modeling. Atmos. Environ. **164**, 216–223 (2017)
35. Šoštarić, A., Stojić, S.S., Vuković, G., Mijić, Z., Stojić, A., Gržetić, I.: Rainwater capacities for BTEX scavenging from ambient air. Atmos. Environ. **168**, 46–54 (2017)
36. Šoštarić, A., Stojić, A., Stojić, S.S., Gržetić, I.: Quantification and mechanisms of BTEX distribution between aqueous and gaseous phase in a dynamic system. Chemosphere **144**, 721–727 (2016)
37. Stanišić, S., Perišić, M., Jovanović, G., Milićević, T., Romanić, S.H., Jovanović, A., Šoštarić, A., Udovičić, V., Stojić, A.: The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, Part I: emission sources. Environ. Res. 110520 (2020b)
38. Stanišić, S., Perišić, M., Stojić, A.: The use of innovative methodology for the characterization of benzene, toluene, ethylbenzene and xylene sources in the Belgrade area. In: Sinteza International Scientific Conference on Information Technology and Data Related Research, Belgrade, Serbia (2020a)
39. Truong, S.C., Lee, M.I., Kim, G., Kim, D., Park, J.H., Choi, S.D., Cho, G.H.: Accidental benzene release risk assessment in an urban area using an atmospheric dispersion model. Atmos. Environ. **144**, 146–159 (2016)
40. U.S. EPA.: Chemical Data Reporting. United States Environmental Protection Agency, Washington, DC, USA (2016). https://www.epa.gov/sites/production/files/2014-11/documents/2nd_cdr_snapshot_5_19_14.pdf
41. Vardoulakis, S., Solazzo, E., Lumbreras, J.: Intra-urban and street scale variability of BTEX, NO2 and O3 in Birmingham, UK: implications for exposure assessment. Atmos. Environ. **45**, 5069–5078 (2011)
42. Werder, E.J., Engel, L.S., Blair, A., Kwok, R.K., McGrath, J.A., Sandler, D.P.: Blood BTEX levels and neurologic symptoms in Gulf states residents. Environ. Res. **175**, 100–107 (2019)
43. Wright, C.Y., Millar, D.A.: A global statement for air pollution and health. Clean Air J. **29**, 1–2 (2019)

Justin A. Daniels
Editor



Volume 77

Advances in Environmental Research

ADVANCES IN ENVIRONMENTAL RESEARCH

VOLUME 77

ADVANCES IN ENVIRONMENTAL RESEARCH

Additional books and e-books in this series can be found on Nova's website under the Series tab.

ADVANCES IN ENVIRONMENTAL RESEARCH

VOLUME 77

JUSTIN A. DANIELS
EDITOR

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-53618-759-3(eBook)
ISSN: 2158-5717

Published by Nova Science Publishers, Inc. † New York

CONTENTS

In: Advances in Environmental Research ISBN: 978-1-53618-708-3
Editor: Justin A. Daniels

Chapter 4

ADVANCED MODELING OF PERSISTENT ORGANIC POLLUTANT (POPS) PATTERNS IN BIOMATRICES

Gordana Jovanović[1,2,*], Snježana Herceg Romanić[3] and Andreja Stojić[1,2]

[1]Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia
[2]Singidunum University, Belgrade, Serbia
[3]Institute for Medical Research and Occupational Health, Zagreb, Croatia

ABSTRACT

Persistent organic pollutants (POPs) are man-made xenobiotics highly resistant to environmental and metabolic degradation. Because of their high lipophilicity, POPs tend to accumulate in the fatty tissues of living organisms inducing negative health effects to different extents. Current research on POP biomonitoring has mainly relied on traditional

* Corresponding Author's E-mail: gordana.vukovic@ipb.ac.rs.

statistics such as factor analysis or principle component analysis to provide information on the relationship between pollutants and accessory variables. The obtained results required a better understanding of complex functional non-linear dependencies of the contaminants in various environmental samples. For this reason, over recent years, innovative modeling methods have been applied in the field of environmental science to deepen the study of environmental pollutants.

This chapter aims to demonstrate a promising methodology, including machine learning classification and regression (self-organizing maps, decision trees, neural networks, extreme gradient boosting, etc.), and explainable artificial intelligence methods (Shapley additive explanations), to gain a better understanding of specific interrelations of organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs) in biomatrices. We reviewed our investigations aimed at capturing: i) pollutant non-linear interrelations in breast milk and the importance of their association with mother's age and parity; ii) the relationship of fatty acid composition and contaminants in consumable marine fish species; iii) spatio-temporal distribution of POPs in marine fish and iv) species- and seasonal-specific POP dependencies in lake fish. All of the methods have been successfully verified as a reliable means that overcomes traditionally applied statistical approaches. The presented findings should be primarily considered as results of pilot studies that should serve as a valuable base for further research involving a large number of respondents.

Keywords: organochlorine pesticides (OCPs), polychlorinated biphenyls (PCBs), breast milk, fish, machine learning, explainable artificial intelligence

INTRODUCTION

Regardless of the fact that persistent organochlorine compounds have been explored widely over the past several decades, they are still present with equal intensity in scientific research around the world. Persistent organochlorine compounds, polychlorinated biphenyls (PCBs) and organochlorine pesticides (OCPs), belong to a group of compounds known as Persistent Organic Pollutants (POPs) regulated under the Stockholm Convention.

PCBs are a group of 209 aromatic compounds (congeners) varying in position and number of chlorine atoms. They were commercially produced as mixtures with different chlorine content. PCB mixtures have been used in "closed systems" such as dielectrics in transformers and capacitors, and in "open systems" such as plasticizers in paints and plastics, carriers in carbonless copy paper, adhesives, flame retardants, adhesives, etc. PCB production started in 1930 and at the end of 1984, the total world production was in excess of 1 million tonnes (Danse et al., 1997). Intensive usage of OCPs in agriculture and public health started following World War II. The most prominent were DDT (1,1,1-trichloro-2,2-di(4-chlorophenyl)ethane, HCB (hexachlorobenzene) and α-, β-, ɣ-HCH (alpha-, beta-, gamma-hexachlorocyclohexane) (Voldner and Li, 1995). PCBs and OCPs have excellent physico-chemical properties for commercial application but also properties that are considered to be relevant for environmental behavior. Under environmental conditions they persist for a long time because they are stable towards physical, chemical and biological breakdown mechanisms or transform to persistent and toxic products like DDT into DDE (1,1-dichloro-2,2-di(4-chlorophenyl)ethene) and DDD (1,1-dichloro-2,2-di(4-chlorophenyl)ethane). Studies on laboratory animals show broad adverse effects upon animal health. PCBs and OCPs act as endocrine disruptors and according to the IARC, some of these compounds are considered carcinogens (Fisher and Schmidt, 1999). The most prominent effects are chloracne, liver damage following endocrine and reproductive disfunction, immunotoxic and neurobehavioural effect. Humans upon accidental and occupational exposure are susceptible to some of these effects. At environmental background levels, human health effects are often unclear because data are not available to demonstrate the specific relationship between exposure and an endocrine adverse effect. Some epidemiological studies have suggested possible damage on neural function, reproductive and immune system and temporal increases in the incidence of cancers in hormonally sensitive tissues (Longnecker et al., 1997; Ross, 2004; Tsygankov and Lukyanova, 2019). Due to world-wide public concern about adverse outcomes in wildlife and the presence and persistence of organochlorines

in environment, the use of PCBs and OCPs was banned or restricted during the 1970s and 1980s in many industrial countries. International instruments are also established with the purpose of protection of the environment. In 2001, representatives of 92 countries signed the Stockholm Convention on Persistent Organic Pollutants and committed themselves to taking measures to stop or reduce the release of POPs into the environment.

HUMAN BIOLOGICAL MONITORING – HUMAN MILK

Historically, the determination of organochlorine compounds in breast milk began as early as 1950, when the organochlorine pesticide DDT was measured at a higher level in breast milk than in cows. Also, one of the important starting points in the study of persistent organochlorine compounds was the publication "The PCB story" (Jensen, 1972), which initiated research about the accumulation of PCBs in the food chain. A major poisoning incident of about 1000 people by PCBs in Japan in 1968 (Yusho disease) triggered scientific research of PCBs in humans. One of the first results about PCBs in breast milk was published in 1967 within research conducted in Sweden. Around 1970, the results of a human milk study carried out in Croatia, Serbia and Slovenia were published (Krauthacker, Ph.D. Thesis, 1984). A comprehensive review published in 1983 showed data of persistent compounds in breast milk worldwide. The results are shown for countries in Europe, some African, South American and Asian, and the US, Canada, and some data for Australia (Jensen, 1983). A higher levels of PCBs were observed in urban centers than in rural areas.

The common properties of these compounds are persistence, bioaccumulation, toxicity, and the ability to transfer over long distances by air masses. Due to their persistence and lipophilic properties, they accumulate in the tissues of humans and animals containing fat and their elimination is very slow. Therefore, human exposure monitoring would require invasive techniques such as adipose tissue sampling during surgery or by autopsy. The alternative is breast milk because it is a non-invasive

way of obtaining a sample. Because of relatively high fat content in breasts, these compounds are excreted via milk. When human milk is in use in biological monitoring, persistent organic compounds are distributed equally within the fatty tissues of a body, and reflect their levels in adipose tissue as well (as expressed by the pollutant concentratuions per fat content). This allows human exposure determination regardless of the type of sample. The disadvantage of the use of human milk for exposure monitoring is the limited age group of one sex. On the other hand, human milk analyses allow the assessment of the exposure of infants. The factors that can influence PCB and OCP levels in human milk are parity, length of lactation, milk volume (some mothers have more milk than the baby needs), age, smoking and dietary habits and exposure through skin or inhalation. Due to this fact, numerous studies of human exposure have been conducted from 1950 to the present. From 1987 (and followed by 1992-1993, 2005-2007 and 2008-2012), WHO has been continuously performing global surveys to identify quantitative geographical differences in the contamination of mother's milk by POPs and in relation to safety standards (UNEP, 2013). In line with these initiatives, in more than 250 scientific articles (Fång et al., 2015, and references therein), the common way of interpreting milk constituents was with respect to their spatial trend or toxic equivalents. Despite global tendencies, a part of the literature has reported on a dependence of POP occurrence in milk samples and external parameters such as mother's age and weight, childbirth, dietary habits and occupation (e.g., Antignac et al., 2016; Dimitriadou et al., 2016; Hassine et al., 2012; Mannetje et al., 2013; Polder et al., 2009), as well as POPs breaking out during lactation (Vigh et al., 2013). Maternal age and parity have been among the most frequently appraised parameters as their increase has the opposite effects on the POPs in the mother's body. The concentrations of these chemicals have been expected to increase in the body fat of elder mothers, whereas parity is considered to reduce their levels since lactation is recognized as the main path for their excretion from the body.

Even so, these studies are still intense due to several reasons, mainly beacouse of the many adverse effects on humans and the environment,

considering the fact that their concentrations have decreased but reached "steady state level". In bodies as single entities, PCB and OCP (and many other compounds) are present in complex mixtures, which means that humans (and the environment) are exposed to the synergism of compounds. From the epidemiological point of view, the synergistic effect of the compounds has not been sufficiently investigated.

Recent papers provide some insight into new data. Regardless of whether their use is prohibited or restricted, scientific publications expose on elevated levels (e.g., in breast milk). Thus, the investigation of the factors that affect the levels of compounds in breast milk, conducted in Pakistan, revealed possible recent exposure to PCBs. Levels of estimated daily intake to infants through lactation were a hundred-fold higher than the recommended to tolerable daily intake limits established by the World Health Organization (WHO), USEPA, and other reputable regional organizations (Naqvi et al., 2019). Aerts and co-workers (2019) published results from the 6th WHO Coordinated Survey on POPs in human milk in Belgium and recommended that exposure routes other than diet deserve more attention in future research. Potential health impacts caused by e-waste recycling operations should be investigated through epidemiological studies, because high PCBs levels were reported in human milk samples (Man et al. 2017). Food contamination was investigated in the Campania region (Italy) where PCB concentrations are above the maximum residue limit fixed by the EU in goat milk due to the proximity of illegal waste dumps of urban and industrial origins (Ferrante et al, 2017). An investigation of PCBs and OCPs in human milk from Eastern Siberia (Russia) indicated a relatively homogeneous contamination across almost the entire region with the exception of some parts. Some other interesting facts are: high exposure to PCBs and DDTs via the consumption of fat and meat of the Lake Baikal seal, a negative correlation between the duration of lactation and PCBs and location of food production is one of the important factors in exposure to POPs in industrialized areas (Mamontova et al., 2017). WHO coordinates a comprehensive global monitoring program on PCBs, polychlorinated dibenzo-p-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs) and a decreasing temporal trend

has been indicated, with large global and regional differences found; for example, PCB levels were highest in East and West Europe. They indicate significant in utero exposure in comparison to lactational exposure, and pointed out the importance of breastfeeding advantages (van den Berg et al., 2017).

However, the complexities that arise among POP interrelations in the milk samples or their dependence on the mother's age and number of born children have not been evaluated by data modelling methods. In our studies, we monitored the pollutant levels in the breast milk sampled in Croatia (Klinčić et al., 2014; 2016) and investigated non-linear dynamics of POP pattern (Jovanović et al., 2019).

ENVIRONMENTAL BIOLOGICAL MONITORING

PCBs and OCPs released into the environment are subject to a variety of transfer and transformation processes. Transmission processes involve transitions from one part of the environment to another (air - surface of water; air - soil, plants; gaseous phase in air - particles in air; water - soil, sediment). They are not retained in one, but are constantly transferred from one part of the environment to the other or circulating in the environment. The circulation between individual parts of the environment depends on the physicochemical properties of the compound, the ambient temperature and the content of organic matter in each part of the environment. Transformation processes refer to changes in the structure of a compound by photochemical, chemical or biotransformation reactions. To what extent a process of transfer and/or transformation will take place depends on the chemical structure of the compound and the part and properties of the environmental compartment (Wania et al., 1998; Wania and Mackay,1996).

In the marine part of the environment, PCBs and OCPs are bound to a suspended substance and are deposited on the bottom, whereas the levels of these compounds are higher in sediment than in water. In this part of the environment, organochlorine compounds are very persistent because they

are not subject to oxidation and hydrolysis, but are slowly degraded microbiologically, while in sea depths there is no microbial degradation. Consequently, deep sea environments act as a sink and as a secondary source of pollution. From the sea bottom, PCBs and OCPs accumulate and re-enter the food chain by dwelling organisms. Marine predators with long life spans may accumulate high POP concentrations and there is a risk to people who consume fish from the top (tuna e.g.) and other seafood, as well (Chiesa et al., 2016).

ENVIRONMENTAL DATA SCIENCE: MODELLING APPROACH

Since humans and the environment are exposed to a mixture of compounds, multi-chemical analysis is often used for exploring the exposure. Commonly, in PCDD/F and PCB congener analysis, sample toxicity is expressed in relation to the toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (2,3,7,8-TCDD), one of the most toxic compounds, using equivalent toxicity factors determined on the basis of in vivo and in vitro studies. Factor values are periodically re-evaluated based on new data on the activity and toxicity of certain compounds. Many epidemiological studies used toxic equivalent factor to evaluate health effects (van den Berg et al., 2006; Baba et al., 2018).

Different statistical approaches are used to clarify the impact of various factors on PCB and OCP levels. Most publications are dominated by the traditional statistical approach, while some models have only recently been developed. The need for model development arose because of the complexity of investigation around the world which includes a lot of new information, as for example, information on elevated levels, new inputs of compounds into humans and the environment, new possible sources of pollution or inputs from secondary sources of pollution, and a number of large compounds-specific datasets. All of these influence the need for a

better understanding of the health effects caused by exposure to mixtures of persistent compounds.

Development of flexible methods for data analysis is a new strategy for exploring the underlying structure of a persistent compound complex dataset in environmental systems. Algorithmic and computational breakthroughs make data science an inevitable approach to extract meaningful conclusions from complex and heterogeneous environmental data. Artificial intelligence (AI) algorithms implemented via machine learning (ML) has led to global transformations of science, technology and daily lives, and it will continue to have a huge influence on the world in the forthcoming years (Steinberg, 2020). AI paved the way for innovative approaches in many scientific disciplines and shifted research from a single discipline's analytical concept towards a transdisciplinary approach.

Factor analysis, principle component analysis and latent class analysis used to investigate the correlation among the variables of interest could not address the interaction of different POPs and the corresponding combined effects, particularly for the potential non-linear exposure-response relationship (Tonga et al., 2018). To tackle challenges of the complexity, heterogeneity and non-linearity, ML algorithms appear as promising methodology. By "letting the data speak for itself" (McCabe et al., 2017), ML methods enable sophisticated data processing and in combination with explainable artificial intelligence (XAI), the possibility of interpretation that could reveal associations between a target variable and a potentially unlimited number of explanatory predictors without explicit knowledge of underlying processes. In other words, interpretation methods overcome human intuition and often limit its domain knowledge (Montavon et al., 2018). Lundberg and Lee (2017a) have recently developed SHAP (Shapley Additive exPlanations), an additive feature attribution XAI method, which offers a unique model solution and surpasses inconsistent results and often contradictory explanations of ML algorithms.

Our scope is to present a promising methodology for studying pollutant interactions (Jovanović et al., 2019; Herceg Romanić et al., 2018a; Herceg Romanić et al., 2018b; Vuković et al., 2018), deliberately broad in considering various data from natural environment, which includes, but it

is not limited to the atmosphere, geosphere, hydrosphere, biosphere and anthroposphere. The introduction of the AI approach in environmental science is required to provide a broader insight into the research phenomena, increase understanding of complex chemodynamics and make a solid foundation for reasonable decision-making.

BREAST MILK: MODELLING APPROACH

In this section of the chapter, we will outline the statistical and ML methods that have been applied to explore PCBs and OCPs in breast milk (Jovanović et al., 2019).

To investigate a potential dynamic of POP concentrations, linear and a locally weighted Loess regression analyses obtained using a software environment for statistical computing R (R Team, 2012), were used. The Loess regression fits simple models to localized data subsets and provide a function that describes the deterministic part of the variation in the data (Li et al., 2014).

To select parameters (POPs and mother's age/parity) that are most relevant to one another, we applied the guided regularized random forest (GRRF) (Deng and Runger, 2013). The GRRF represents an ensemble learning method, which uses the importance scores from a preliminary random forest (RF) to guide the feature selection of regularized random forest (RRF). It is more robust and computationally efficient than RRF and it moderates the course of dimensionality. It avoids the effort to analyze irrelevant or redundant features. The variable importance was calculated as the average value of 1000 model runs while the method performances were tested by 100 times replicated 10-fold cross validation.

To investigate the most relevant features for a pollutant prediction (OCPs, PCBs, mother's age and childbirth), the autoWeka meta learner was applied on five best predicting features (Kotthoff et al., 2016). To maximize performance, the learner uses a Bayesian optimization method and searches through the Weka's learning algorithms and their respective hyperparameter settings. The following ML algorithms: Additive

Regression, Attribute Selected Classifier, IBk, K*, Bagging, Linear NN Search, Locally Weighted Learning (LWL), M5 Rules, Multilayer Perceptron (MLP), RF, Random Tree (RT), REPTree and Sequential Minimal Optimization Regression (SMOreg), were implemented in Weka (Weka 3.8 core and Weka Sourceforge, 2016) (Frank et al., 2005). Brief descriptions of the methods are given in Jovanović et al. (2019) and Šoštarić et al. (2017). We will describe herein only the relevant assumptions.

Additive Regression is a meta classifier that in each iteration fits a model to the residuals left on the previous iteration. Prediction is accomplished by adding the predictions of each classifier. Reducing the learning rate parameter helps prevent overfitting and has a smoothing effect. However, this increases the learning time. The method improves the presentation of a regression base classifier (Friedman, 2002).

Attribute Selected Classifier is a meta classifier that tests data using attribute selection before being passed on to a classifier and reduces the dimensionality of training.

Gaussian Processes uses lazy learning and a measure of the similarity between data to predict the value for a point from training data (Rasmussen, 2006) starting from any finite number of variables, which has (consistent) joint Gaussian distributions.

IBk (k nearest neighbors – k-NN) only locally approximates the function and gives the output value as the average of its k nearest neighbors (Aha et al., 1991).

Unlike other instance-based learners, K* uses an entropy-based distance function and classifies a test instance based on the class of training instances like it, created by some similarity function (Cleary and Trigg, 1995).

Bagging is a meta class for boosted aggregating (bagging) a classifier that reduces variance and the possibility of overfitting, which improve the overall stability and accuracy.

Linear NN Search employs the brute force search algorithm which involves all possible nearest neighbor candidates for the explanation.

Locally Weighted Learning (LWL) apportions instance weights based on a weighting function according to the fits of linear regression model (Frank et al., 2003).

M5 Rules uses separate-and-conquer to generate a model tree at an iteration using M5. Afterwards, the method transforms the "best" leaf into a rule in each cell of the input space (Wang and Witten, 1996).

Multilayer Perceptron (MLP) is a feed forward neural network, which comprises one or more hidden layer between input and output data (Haykin, 1994). It is trained by the back-propagation algorithm, which corrected the weights to each connection between the unseen and output layer.

Random Forest represents a collection of unpruned classification or regression trees. In the tree induction process, RF selection derived the trees from bootstrap samples of the training data while the prediction is made by aggregation (Breiman, 2001; Zhao and Zhang, 2008). The RF yields a generalization error rate making it more robust to noise.

Random Tree (RT) designs a tree without pruning and estimates class probabilities based on a holdout set (backfitting). The formed tree considers K randomly chosen attributes at each node (Zhao and Zhang, 2008) giving an equal chance to each tree in the set of trees to be sampled.

REPTree is a fast decision tree learner, which use variance and prunes to construct a regression tree (with backfitting).

The Sequential Minimal Optimization Regression (SMOreg) algorithm (Shevade et al., 1999) uses non-linear function and maps the input data into a high dimensional feature space, where linear regression is performed.

Apportionment of dominant pollutant sources was obtained using Unmix (USEPA, 2007). The maximum number of pollutants was chosen as input variables based on trials and errors, which yield the most physically meaningful results. In addition, GRRF was applied to discuss the importance of Unmix derived sources.

To investigate the interaction effects between indicator congener PCB-138 and other relevant contaminants in breast milk and the mother's features, we applied an additive feature attribution method – SHAP. Unlike the conventional attribution methods, SHAP has a unique solution aimed at

the post-hoc interpretation of ML methods and overcomes the main drawback of inconsistency that is presented in gradient boosting or RF methods. Obtained SHAP values attribute to each feature the change in the expected model prediction (Lundberg and Lee, 2017a). The SHAP runs from exponential to O (TLD2) for unbalanced trees and O (TL log2 L) for balanced trees, where T is the number of trees, L is the maximum number of leaves in any tree, and D is the maximum depth of any tree (Lundberg and Lee, 2017b). SHAP offers summary, partial dependence and interaction plots to better capture the interaction effects. Partial dependence characterizes the dependency of a model on a subset of features with all other features fixed and captures a feature's attributed importance and corresponding changes. Unlike standard partial dependence plots represented by lines, the SHAP dependence plots depict interaction effects as vertical dispersion.

In our investigations (Jovanović et al., 2019), we performed GRRF to supplement interpretations on the pollutant's non-linear interactions in addition to Pearson's correlation coefficients, commonly used as a measure of a linear explanatory linkage between the milk constituents and mother's conditions. The GRRF-derived variable importance implied a minor influence of α- and β-HCH on the other investigated POPs, which is a likely consequence of the pollutant's different metabolic pathway in the maternal body. The highest relative GRRF importance was observed for p,p′-DDE contrary to p,p′-DDT. The p,p′-DDE was the most important for the prediction of the following congeners: PCB-118 > PCB-170 > PCB−153 > PCB−180 > PCB-105 > PCB-167 > p,p′-DDT > PCB-138. Out of the 17 examined congeners, GRRF recognized a prominent influence of non-dioxin-like/indicator PCBs (−153 and −180) on the others including PCB−138 and dioxin-like PCBs (−118, −156 and −105). The advanced GRRF approach confirmed results of generic (physiologically based pharmacokinetic) PBPK modeling which indicated highly correlated predictions and toxicokinetic profiles of the listed compounds (Verner et al., 2009). What is common to both GRRF-derived classes and PCB-170 is the chlorine atom attached at the ortho position to the biphenyl ring. The result implied mono chlorine as an essential feature that drives PCB

profiles in breast milk regardless of the ring coplanar structure and toxicological properties. Our results support previous suggestions that the position of the halogen substitutes rather than lipophilicity, molecule diameter and weight, and the number of attached halogens, leads to the rigid molecular structure and consequently, influence the PCBs' partition between blood and breast milk (Mannetje et al., 2012; Vasios et al., 2016). We note that further studies with extensive datasets are required to confirm these indications not only for the noted compounds but also for similarly structured congeners such as −157, −114 and −189 whose concentrations in this study were lower than the limit of detection. In addition, the GRRF importances of β-HCH, p,p′-DDE, PCB-138, PCB-180, PCB-118, PCB-156 and PCB170 suggested a slight influence of the mother's age on the PCB and OCP content in the milk regardless of the number of born children.

The different origin of investigated contaminants was apportioned by Unmix, which extracted four factors, i.e., source profiles with high determination and correlation coefficients (>0.90). The first and the second factor (Source 1 and 2) were differentiated by the highest attributions of β-HCH (78%) and HCB (67%), and γ-HCH (73%) therefore representing byproducts of lindane (United Nations Environment, 2017). The overall highest shares (approximately 30%) were allocated to the "heavy" hexa- to hepta-chloro occupational congeners: PCB-180 (83%), PCB-170 (81%), PCB-156 (59%), PCB-153 (54%) and PCB-138 (53%). The PCBs possess a higher half-life than the "light" congeners, moderately contributed to the third factor (Source 3). The highest shares of p,p′-DDD (93%) and p,p′-DDE (53%) in the fourth factor (Source 4) showed a common metabolic pathway of p,p′-DDT xenobiotics and partially extracted the "light" tetra- to penta chloro congeners PCBs (around 46%). These associations are like those recognized by GRRF. In addition, GRRF did not identify a visible relationship between Unmix profiles and mother's age.

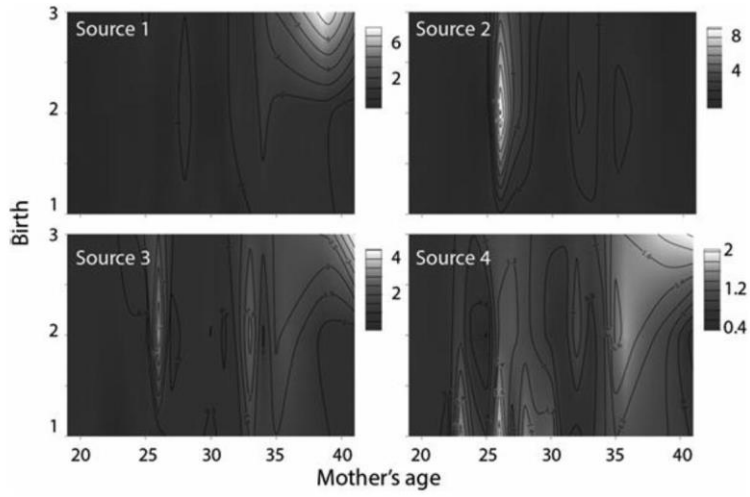


Figure 1a. Unmix-derived source contributions in relation to the maternal age and parity.

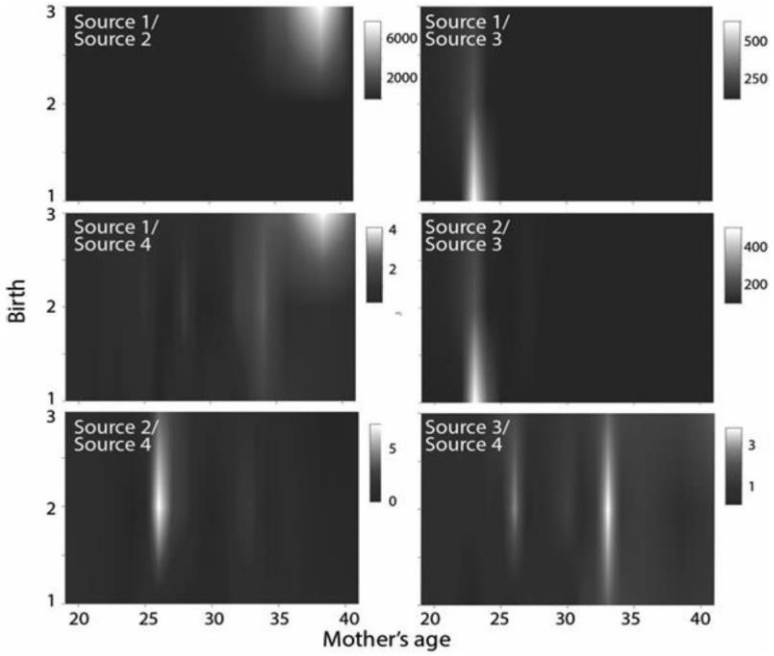


Figure 1b. Unmix-derived source contribution ratios in relation to the maternal age and parity.

Apart from mutual sources, Unmix grouped pollutants according to similar chemical structure and behavior (Figures 1a and 1b). Continuously increased levels of highly chlorinated congers were evident in the milk of older mothers due to slower elimination mechanisms compared to the more volatile OCPs. The ratio of HCH to DDT group as well as heavy PCBs did not mainly change with age and parity as noted in the representative range of maternal ages. We note that indications based on the milk samples of mothers in their mid-twenties should be approached cautiously due to the limited dataset. Ratios among the functionally dependent PCBs varied and exhibited maximums between first-time mothers and multiparae. The findings are in accordance with the conclusions of Hooper et al. (2007), who emphasized that the breast-feeding does not noticeably decrease PCBs in the milk samples of primiparae and multiparea. Conversely, some authors have favored pregnancy, lactation and, consequently, parity as factors affecting reduced levels of POPs in mature milk (e.g., Waliszewski et al., 2002).

Among ML methods, the MLR provided predictions of PCB-138, PCB-153, PCB-180, PCB-105, PCB-118, PCB-167 and PCB-170, based on the other congeners with relative errors ranging from 20% to 40%, and predicted/pbserved correlation coefficients higher than 0.90. Better performances were shown by RF, Gaussian Processes and LWL using IBk with relative errors lower than 30% and high correlation coefficients ($r > 0.90$). The methods confirmed prominent non-linear relationships between β-HCH, γ-HCH and HCB, as well as p,p´-DDT and its metabolites, underlining the possible complexity of their interactions. Moreover, given the MLR accuracy metrics, r and the relative absolute error, LWL, Attribute Selected Classifier and Additive Regression (SMOreg) appeared to be superior to MLR. To relate maternal age and pollutant contents in milk, Additive Regression (K*, $r = 0.98$, relative error=1.6%) based on GRRF attribute selection appeared to be the best performing classifiers. However, Bagging on Unmix derived source contribution (IBk, $r = 0.90$, relative error = 14.8%) was more accurate for the estimation of maternal parity influences.

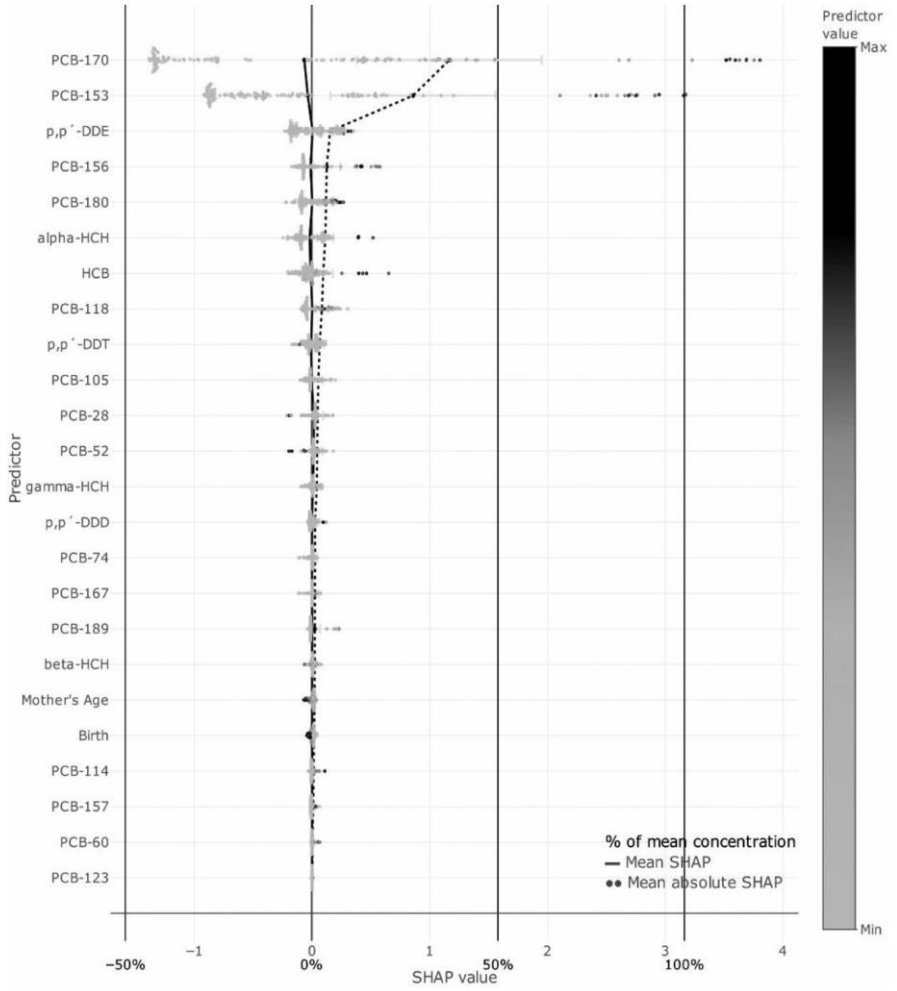


Figure 2. SHAP summary plots for PCB-138 concentrations in breast milk.

To get a detailed overview of which features are the most important for PCB 138 patterns in breast milk, SHAP allows to plot the SHAP values of every feature for every sample. The plot below (Figure 2) sorts investigated pollutants, mother's age and parity by the sum of SHAP value magnitudes of all samples and shows the distribution of the impacts each variable has on the model output. A high level of PCB-170 and PCB-152 has a high and positive impact on the PCB-138 patterns as indicated by dark-labeled tailed distribution to the right. Lower concentrations of these

congeners (in light colour) are negatively correlated with the target variable as indicated by negative SHAP value. Less prominent attributions of p,p'-DDE, PCB-156, PCB-180, HCB, HCH, PCB-118, PCB-28, PCB-52 and PCB-189 governed the PCB-138 distribution. The influence of both non-dioxin-like/indicator (−28, −52 and −180) and toxicologically relevant dioxin-like PCBs (−118 and −189) was evident.

LAKE AND SEA FISH: A MODELLING APPROACH

The interpretation of absolute OCP and PCB concentrations in fish has been approached rather generally, perhaps because numerous factors such as diet, elimination/metabolisms processes, and growth rate, age and lipid affect the POP bioaccumulation in fish (e.g., Polder et al., 2014; Ren et al., 2017; Yang et al., 2010). We discussed how advanced statistical models could be used as a promising approach to tackle challenging OCP and PCB issues through analyses of fish samples and POPs based on their similarities. To investigate i) spatio-temporal distribution of POPs in marine fish (Vuković et al., 2018) and ii) species- and seasonal-specific POP dependencies in lake fish (Herceg Romanić et al., 2018a), we applied advanced classification and clustering methods – the Kohonen self-organizing maps (SOM) and Decision Tree Learning (DT). SOM was performed using MATLAB Neural network toolbox 10.0 (Ballabio and Vasighi, 2012) while a STATISTICA (StatSoft Inc., 2012) Classification and Regression Trees (C&RT) routine (Breiman et al., 1984) was applied for DT analyses.

Briefly, the SOM analysis represents an unsupervised neural network clustering technique (Kohonen, 2013). It uses high-dimensional data and alters the non-linear statistical relationship between them into a one- or two-dimensional grid of a geometric relationship. Each neuron in an input layer is weighted by n-dimensional vectors, $m = [m_1, \ldots, m_n]$. The Euclidean distances between each data vector and all the weight vectors are calculated based on the best-matching model. The output layer is represented by the map unit containing as many components as the number

of input variables and data with similar properties are close to each other. In our investigations, the SOM inputs were the concentrations of the analyzed pollutants that were normalized to the range of 0–1. The number of neurons in the output layer (map) was chosen following the empirical rule $M = 5\sqrt{n}$ (Wang et al., 2015), where M is the number of map units and n is the number of samples.

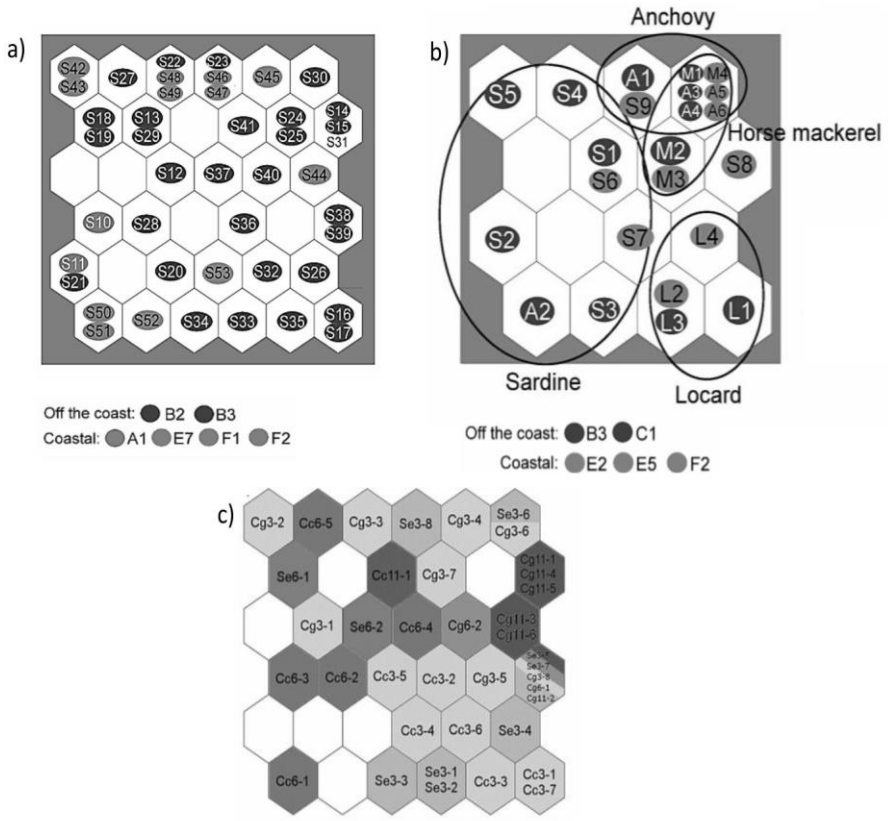


Figure 3. The Kohonen self-organizing maps (SOM) clusters by: a) marine fish species sampled in 2014 and fisheries zone A – anchovy, L – chub mackerel, M – horse mackerel, and S – sardine and round sardinella; b) sardines sampled in 2016 and fisheries zone, and c): lake fish species: Cc – *C. carpio* (carp) and Cg – *C. gibelio* (Prussian carp); numbers 3, 6 and 11 refer to the sampling month, March, June and November while the rest represent the label of samples.

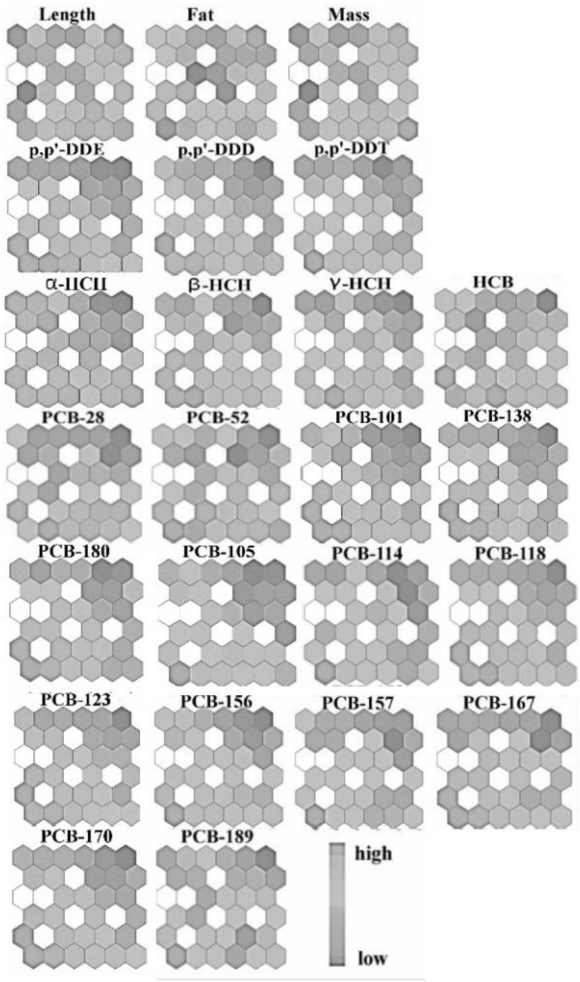


Figure 4. An example of the component planes obtained for organochlorine pollutants and physical characteristics using the Kohonen self-organizing maps (SOM) for marine fish (sardines) sampled in 2016.

A DT is a widely used recursive partition method which classifies independent variables based on one root node, several internal nodes – "leaves", and a set of terminal nodes (Kim and Koehler, 1995; Otukei and Blaschke, 2010). To find values of single variables that best divide the data into groups, the decision rule is applied at each node – "branches". Firstly, DT models were created by splitting until no further split can be performed

and the maximal tree is obtained. Secondly, a set of subtrees is gained by pruning the maximal tree based on a cost-complexity parameter. Finally, the optimal tree was selected from the set of subtrees using cross validation. We used a 5-fold cross-validation to obtain a model with the smallest overall misclassification rate.

The SOM analysis concerning marine fishes clearly distinguished species sampled in different periods and from various zones but it could not provide a straightforward explanation on the obtained observations because many external parameters such as data size and sample characteristics affect contaminant levels (Vuković et al., 2018). As shown in Figure 3a, pollutant content was different in the sardines caught in the off coast and coastal zone. In addition, the HCH compounds along with lighter congeners (PCB-52 and PCB-60) and PCB-105, PCB-118, PCB-156, PCB-167, PCB170 and PCB-180 contributed to the clear distinction of off-coast samples (Figure 4). The physicochemical features such as age, length and weight along with sampling period due to spawning probably cause variations in the OCP and PCB concentrations in sardines (Antunes et al., 2007). The estimation of the influence of sampling location, the evident pollutant sources and emission levels are susceptible to numerous uncertainties since sardines actively move/swim between different zones and the pollutant source apportionment is.

DT analysis supplemented the SOM results and indicated that species mass dominantly varied between coastline and off-coast sardines with pronounced heavier off-shore species.

As indicated by SOM (Figure 5b), differences in the length and mass of sardine and chub mackerel appeared to be distinctive properties among the following fish species: horse mackerel, chub mackerel, anchovies and sardines. DT analysis underlined the anchovy with the smallest *p,p'*-DDD levels and the lightest mass whereas the mackerel was classified as the heaviest species with the highest levels of PCB-189 (Figure 5a). The chub mackerel was distinguished from the others by certain shares of α-HCH and PCB-74 separated. Since horse mackerel and chub mackerel are important predators, higher POP concentrations are expected in these species, whose diet is composed mainly of smaller prey fishes like

anchovies and sardines. Mackerels mostly occupied the open sea, whereas sardines and anchovies move between coastal and open sea regions which contributed to different levels of PCB-74, PCB-101, PCB-118 and PCB-157 in the samples (Figure 5b).

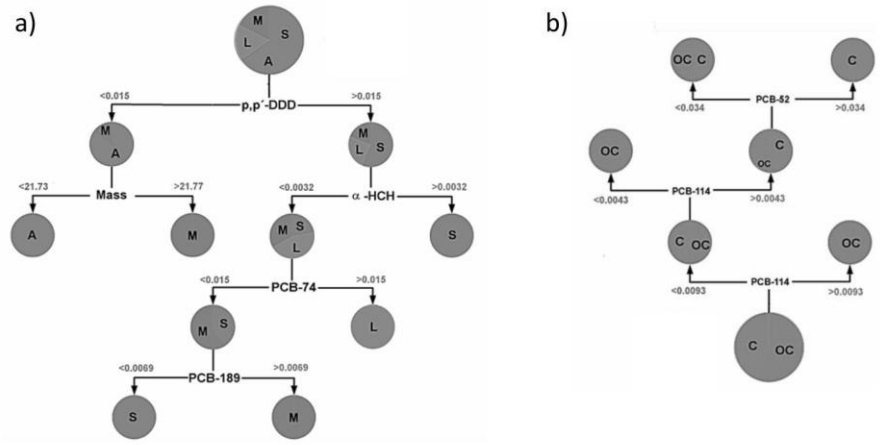


Figure 5. An example of DT classification of the POPs according to: a) fish species sampled in 2014 and b) fisheries zones; A – anchovy, L – chub mackerel, M – horse mackerel, and S – sardine and round sardinella; OC – off coast zones, and C – coastal zones.

The SOM analysis applied in our investigation regarding *Cyprinidae* lake fish (Herceg Romanić et al., 2018a) separated clusters regarding fish species and sampling periods (Figure 3c). The season-related changes and correspondent meteorological parameters (temperature, atmospheric pressure and relative air humidity) affected similarities of the POP contents in the tissue of carp sampled in March. Conversely, samples of the same species from June showed the strongest dissimilarities due to the volatilization of light PCBs which is hard to estimate. Uptake of OCPs and heavier PCB congeners in fish have been usually reported as temperature-dependent (e.g., Gallego et al., 2007); higher elevation and low temperature favors the entrapment of less volatile congeners. These compounds have also been proven to poorly re-evaporate from water into the surrounding air during warm seasons, which is why they are likely to bioaccumulate in fish tissue during spring or autumn months (Yang et al.,

2010). Moreover, the lower temperature slows the metabolic eliminations of chlorine atoms of PCBs.

The most similar SOM clusters of the Prussian carp sampled in March, June and November, suggested this species as appropriate for studying the general patterns of organochlorine and biphenyl pollutants in lakes whereas the rudd samples appeared to be the least suitable *Cyprinidae* (Figure 3c).

The results of DT analysis supported the SOM indications underlying the following pollutants: PCB-126, PCB-156, PCB-105, PCB-28, PCB-114, PCB-105, PCB-180, PCB-52, PCB-101 and PCB-170 as crucial for differentiation between fish species and seasons (Figure 6).

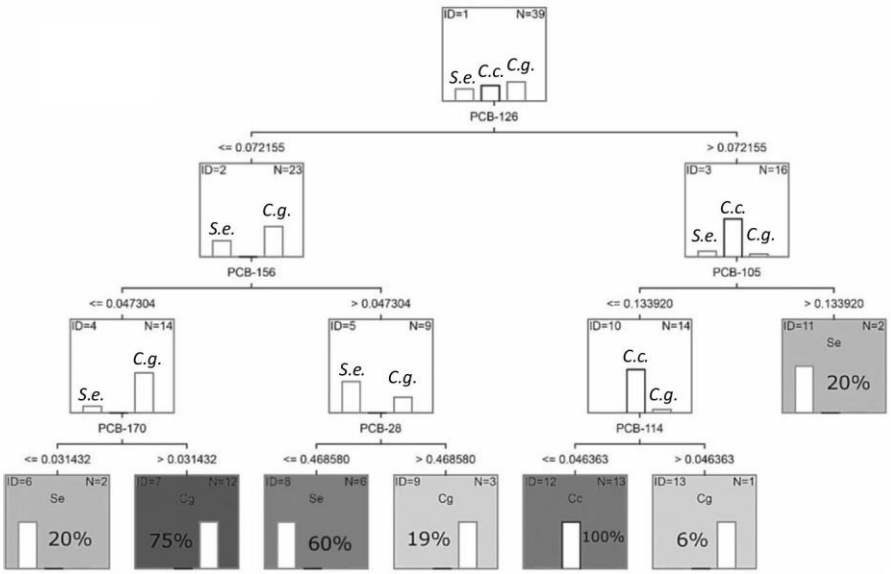


Figure 6. DT classification of the POPs according to the lake-fish species: *S.e.* – *S. erythrophthalmus* (rudd), *C.c.* – *C. carpio* (carp) and *C.g.* – *C. gibelio* (Prussian carp).

The presented results indicate that SOM analysis could complement traditional statistics approaches and offer a detailed interpretation of the pollutant behavior. Moreover, the joint use of models significantly improved the investigation of POP as indicated herein by the SOM and DT (Herceg Romanić et al., 2018a; Vuković et al., 2018).

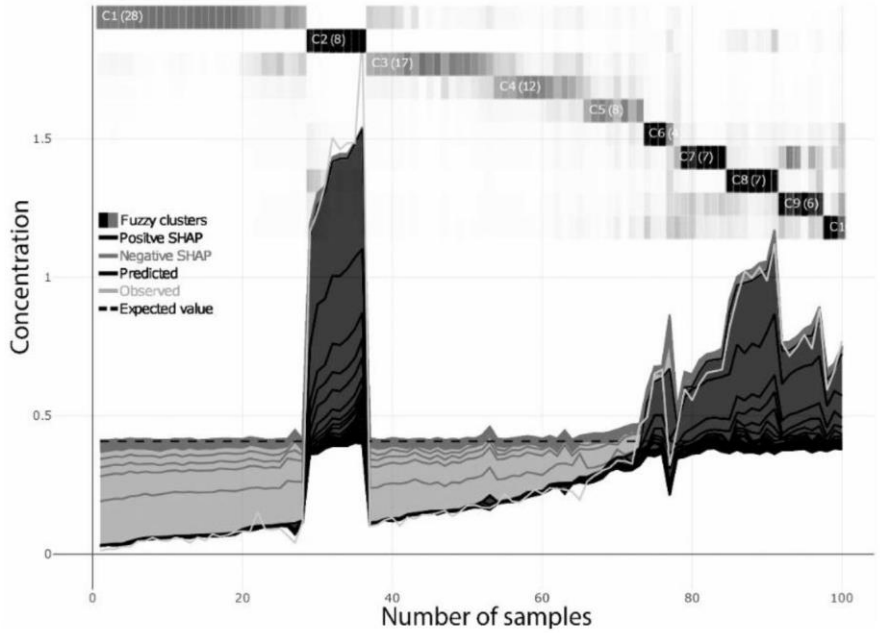


Figure 7. An example of SHAP fuzzy clusters of variables that shape PCB-138 concentrations in fish.

The relationship of fatty acid composition and POPs in consumable marine fish species were studied since oily fish represent a major source of omega-3 and 6 long-chain polyunsaturated fatty acids (n-3/ω-3 and n-6/ω-6 PUFA), which are known for various health benefits. We applied fuzzy clustering of SHAP values to find groups of similar variables that shape the dynamics of PCB-138. Fuzzy clustering works by clustering on Shapley values of each instance, i.e., by explanation similarity overcoming difficulties to compute distances between instances on different, non-comparable scales. All SHAP values have the same unit – the unit of the prediction space. Ten clusters were extracted in total among which six (C2 and C6–C10) strongly positively influence PCB-138 concentrations higher than 0.5 ng g-1. As labelled by lighter colour in Figure 7, four clusters (C1 and C3–C5) dominantly (indicated by shares given in brackets) showed negative correlations with PCB-138 presented in levels lower than 0.5 ng g-1. Out of 18 investigated fatty acids (saturated, monounsaturated–MUFA and PUFA), MUFA (myristic acid-C14:0 and margaric-C17:0) followed by

nutritionally beneficial eicosadienoic acid (C20:2n-6) mostly contributed to the PCB-138 bioaccumulation. Indicator congeners (–153 and –180) and metabolite p,p'-DDE showed significant impact on PCB-138 patterns, whereas less important relationships were indicated between target and the most toxic dioxin-like congeners.

CONCLUSION

At the beginning of the 21st century, a rapid development of personal computers occurred and new software and applications were created to solve various problems and finally, AI was implemented. AI does not need to specify every subroutine for every possible scenario that may or may not happen. Therefore, it has been experiencing a rapid rise in recent years in the deployment and development of intelligent systems in areas such as communications, internet searches, science, technology and daily lives, and it will continue to have a huge influence on the world in the forthcoming years. AI sub-disciplines such as ML can substantially contribute to detailed investigation and interpretation of incomplete, imprecise, and vague scientific data generated by heterogeneous sources.

From the elaborated data in this chapter, it is evident that many doubts have been present for a long time in the field of environmental chemistry where traditional statistics has been applied. Some of the issues that could be approached by the advanced AI include modeling of natural complex dynamics and spatio-temporality; quality, uncertainty, and the representativeness of the environmental data; and information interpretation and its long-term storage. This is why our research into the application of advanced statistical methods and AI to environmental chemistry has been initiated. The research presented herein can be understood as one of the first attempts to link AI and biomonitoring. The chapter demonstrated a promising AI/ML/XAI methodology to enhance understanding of specific non-linear interrelations of OCPs and PCBs in biomatrices–human milk and marine and lake edible fish, and their dependences on variables like mother's age and parity, spatio-temporality

and fatty acid composition, respectively. All of the methods have been successfully verified as a reliable means that overcomes traditionally applied statistics. The indications underlined the need for the urgent shift from traditional approaches in environmental science, because the environmental field is the one of extreme complexity and has great importance for the mankind and its future.

REFERENCES

Aerts, R., Van Overmeire, I., Colles, A., Andjelković, M., Malarvannan, G., Poma, G., Den Hond, E., Van de Mieroop, E. Dewolf, M-C., Charlet, F., Van Nieuwenhuyse, A., Van Loco, J., Covaci, A. (2019). Determinants of persistent organic pollutant (POP) concentrations in human breast milk of a cross-sectional sample of primiparous mothers in Belgium. *Environmental International,* 131:104979.

Aha, D.W., Kibler, D., Albert, M.K. (1991). Instance-based learning algorithms. *Machine Learning*, 6:37–66.

Antignac, J.P., Main, K.M., Virtanen, H.E., Boquien, C.Y., Marchand, P., Venissieux, A., Guiffard, I., Bichon, E., Wohlfahrt-Veje, C., Legrand, A., Boscher, C., Skakkebæk, N.E., Toppari, J., Le Bizec, B. (2016). Country-specific chemical signatures of persistent organic pollutants (POPs) in breast milk of French, Danish and Finnish women. *Environmental Pollution,* 218:728–738.

Antunes, P., Amado, J., Vale, C., Gil, O. (2007). Influence of the chemical structure on mobility of PCB congeners in female and male sardine (*Sardina pilchardus*) from Portuguese coast. *Chemosphere,* 69:395–402.

Baba, T., Ito, S., Yuasa, M., Yoshioka, E., Miyashita, C., Araki, A., Sasaki, S., Kobayashi, S., Kajiwara, J., Hori, T., Kato, S., Kishi, R. (2018). Association of prenatal exposure to PCDD/Fs and PCBs with maternal and infant thyroid hormones: the Hokkaido Study on Environment and Children's Health. *Science of the Total Environment*, 615:1239–1246.

Ballabio, D., Vasighi, M. (2012). A MATLAB Toolbox for Self Organizing Maps and supervised neural network learning strategies. *Chemometrics and Intelligent Laboratory Systems,* 118:24−32.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Chiesa, L.M., Labella, G.F., Panseri, S., Pavlović, R., Bonacci, S., Arioli, F. (2016). Distribution of persistent organic pollutants (POPs) in wild Bluefin tuna (*Thunnus thynnus*) from different FAO capture zones. *Chemosphere,* 153:162–169.

Cleary, J.G., Trigg, L.E. (1995). K*: An instance-based learner using an entropic distance measure. In: *Proceedings of the 12th International Conference on Machine learning,* 108–114.

Danse, I.R., Jaeger, R.J., Kava, R., Kroger, M., London, W.M., Lu, F.C., Maickel, R.P., McKetta, J.J., Newell, G.W., Shindell, S., Stare, F.J., Whelan, E.M. (1997). Position paper of the American council on science and health: Public health concerns about environmental polychlorinated biphenyls (PCBs). *Ecotoxicolology and Environmental Safety*, 38:71–84.

Deng H., Runger G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition,* 46:3483–3489.

Dimitriadou, L., Malarvannan, G., Covaci, A., Iossifidou, E., Tzafettas, J., Zournatzi-Koiou, V., Kalantzi, O.I. (2016). Levels and profiles of brominated and chlorinated contaminants in human breast milk from Thessaloniki, Greece. *Science of the Total Environment*, 539:350–358.

Ferrante, M.C., Fusco, G., Monnolo, A., Saggiomo, F., Guccione, J., Mercogliano, R., Clausi. M.T. (2017). Food contamination by PCBs and waste disposal crisis: evidence from goat milk in Campania (Italy). *Chemosphere*, 186:396–404.

Fisher, B.E., Schmidt, C.W. (1999). Most unwanted persistent organic pollutants. *Environmental Health Perspect*ives, 107: A18–A25.

Fång, J., Nyberg, E., Winnberg, U., Bignert, A., Bergman, Å. (2015). Spatial and temporal trends of the Stockholm Convention POPs in mothers' milk — a global review. *Environmental Science and Pollution Research,* 22:8989–9041.

Frank, E., Hall, M., Pfahringer, B. (2003). Locally Weighted Naive Bayes. In: *19th Conference in Uncertainty in Artificial Intelligence*, 249–256.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H., Trigg, L. (2005). Weka." In: *Data Mining and Knowledge Discovery Handbook* 1305–1314. Springer US.

Frank, E., Mayo, M., Kramer, S. (2015). Alternating model trees. In: *Proceedings of the 30th Annual ACM Symposium on Applied Computing* 871–878. ACM.

Friedman, J.H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38:367–378.

Gallego, E., Grimalt, J.O., Bartrons, M., Lopez, J.F., Camarero, L., Catalan, J., Stuchlik, E., Battarbee, R. (2007). Altitudinal gradients of PBDEs and PCBs in fish from European high mountain lakes. *Environmental Science and Technology*, 21:2196–2202.

Graczyk, M., Lasota, T., Trawiński, B. (2009). Comparative analysis of premises valuation models using KEEL, RapidMiner, and WEKA. In: *International Conference on computational Collective Intelligence* 800–812. Springer Berlin Heidelberg.

Hassine, S.B., Ameur, W.B., Gandoura, N., Driss, M.R. (2012). Determination of chlorinated pesticides, polychlorinated biphenyls, and polybrominated diphenyl ethers in human milk from Bizerte (Tunisia) in 2010. *Chemosphere,* 89:369–377.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. 1st Ed., Upper Saddle River, NJ, USA: Prentice Hall PTR.

Herceg Romanić, S., Vuković. G., Klinčić, D., Matek Sarić, M., Župan, I., Antanasijević, D., Popović, A. (2018a). Organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs) in *Cyprinidae* fish: Towards hints of their arrangements using advanced classification methods. *Environmental Research,* 165:349–357.

Herceg Romanić, S., Vuković, G., Klinčić, D., Antanasijević, D. (2018b). Self-organizing maps for indications of airborne polychlorinated biphenyl (PCBs) and organochlorine pesticide (OCPs) dependence on spatial and meteorological parameters. *Science of the Total Environment,* 628–629, 198–205.

Hooper, K., She, J., Sharp, M., Chow, J., Jewell, N., Gephart, R., Holden, A. (2007). Depuration of polybrominated diphenylethers (PBDEs) and polychlorinated biphenyls (PCBs) in breast milk from California first-time mothers (primiparae). *Environmental Health Perspectives*, 115:1271–1275.

Jensen, A.A. (1983). Chemical contaminants in human milk. *Residue Reviews,* 89:1–128.

Jensen, S. (1972). The PCB Story. *Ambio,* 1:123–131.

Jovanović, G., Herceg Romanić, S., Stojić, A., Klinčić, D., Matek Sarić, M., Grzunov Letinić, J., Popović, A. (2019). Introducing of modeling techniques in the research of POPs in breast milk – A pilot study. *Ecotoxicology and Environmental Safety,* 172:341–347.

Kim, H., Koehler, G.J. (1995). Theory and Practice of Decision Tree Induction. *Omega-International Journal of Management Science*, 23:637–652.

Klinčić, D., Herceg Romanić, S., Matek Sarić, M., Grzunov, J., Dukić, B. (2014). Polychlorinated biphenyls and organochlorine pesticides in human milk samples from two regions in Croatia. *Environmental Toxicology and Pharmacology*, 37:543–552.

Klinčić, D., Herceg Romanić, S., Brčić Karačonji, I., Matek Sarić, M., Grzunov Letinić, J., Brajenović N. (2016). Organochlorine pesticides and PCBs (including dl-PCBs) in human milk samples collected from multiparae from Croatia and comparison with primiparae. *Environmental Toxicology and Pharmacology,* 45:74–79.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37:52–65.

Kotthoff, L., Thornton, C., Hoos, H.H., Hutter, F., Leyton-Brown, K. (2016). Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. *Journal of Machine Learning Research,* 17:1–5.

Krauthacker, B. (1984). *Gas chromatographic determination of organochlorine compounds in human blood and milk, and their distribution in populations*. Faculty of Science University of Zagreb, PhD Thesis.

Li, L., Qian, J., Ou, C.Q., Zhou, Y.X., Guo, C. and Guo, Y. (2014). Spatial and temporal analysis of Air Pollution Index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001–2011. *Environmental Pollution,* 190:75–81.

Li, Y.F. (1999). Global technical hexachlorocyclohexane usage and its contamination consequences in the environment: from 1948 to 1997. *Science of the Total Environment,* 232:121–158.

Longnecker, M.P., Rogan, W.J.L.G. (1997). The human health effects of DDT dichlorodiphenyl-trichloroetane) and PCBs (polychlorinated biphenyls) and an overview of organochlorines in public health. *Annual Review of Public Health,* 18:211–44.

Lundberg, S.M., Lee, S.I. (2017a). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765–4774.

Lundberg, S.M., Lee, S.I. (2017b). *Consistent feature attribution for tree ensembles*. arXiv: 1706.06060 (arXiv preprint).

Lundberg, S.M., Erion, G.G., Lee, S.I. (2018). *Consistent individualized feature attribution for tree ensembles.* arXiv:1802.03888 (arXiv preprint).

Mamontova, E.A., Tarasova, E.N., Mamontov, A.A. (2017). PCBs and OCPs in human milk in Eastern Siberia, Russia: Levels, temporal trends and infant exposure assessment. *Chemosphere,* 178:239–248.

Man, Y.B., Chow, K.L., Xing, G.H., Chan, J.K.Y., Wu, S.C., Wong, M.H. (2017). A pilot study on health risk assessment based on body loadings of PCBs of lactating mothers at Taizhou, China, the world's major site for recycling transformers. *Environmental Pollution,* 227:364–371.

Mannetje, A.'t, Coakley, J., Jochen, M.F., Harden, F., Toms, L.M., Douwes, J. (2012). Partitioning of persistent organic pollutants (POPs) between human serum and breast milk: A literature review. *Chemosphere,* 89:911–918.

Mannetje, A.'t, Coakley, J., Bridgen, P., Brooks, C., Harrad, S., Smith, A.H., Pearce, N., Douwes, J. (2013). Current concentrations, temporal trends and determinants of persistent organic pollutants in breast milk

of New Zealand women. *Science of the Total Environment,* 458–460:399–407.

McCabe, M.F., Rodell, M., Alsdorf, D.E., Gonzalez Miralles, D., Uijlenhoet, R., Wagner, W., ... and Shi, J. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences*, 21:3879–3914.

Montavon, G., Samek, W., Müller, K.R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15.

Naqvi, A., Qadir, A., Mahmood, A., Baqar, M., Aslam, I., Jamil, N., Mumtaz, M., Saeed, S, Zhang, G. (2020). Screening of human health risk to infants associated with the polychlorinated biphenyl (PCB) levels in human milk from Punjab Province, Pakistan. *Environmental Science and Pollution Research*, 27:6837–6850.

Otukei, J.R., Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation,* 12S:S27–S31.

Polder, A., Müller, M.B., Lyche, J.L., Mdegela, R.H., Nonga, H.E., Mabiki, F.P., Mbise, T.J., Skaare, J.U., Sandvik, M., Skjerve, E., Lie, E. (2014). Levels and patterns of persistent organic pollutants (POPs) in tilapia (*Oreochromis* sp.) from four different lakes in Tanzania: geographical differences and implications for human health. *Science of the Total Environment*, 488–489:252–260.

Polder, A., Skaare, J.U., Skjerve, E., Løken, K.B., Eggesbø, M. (2009). Levels of chlorinated pesticides and polychlorinated biphenyls in Norwegian breast milk (2002–2006), and factors that may predict the level of contamination. *Science of the Total Environment,* 407:4584–4590.

Rasmussen, C.E. (2006). *Gaussian processes for machine learning*. The MIT Press, Massachusetts Institute of Technology.

Ren, J., Wang, X., Wang, C., Gong, P., Wang, X., Yao, T. (2017). Biomagnification of persistent organic pollutants along a high altitude aquatic food chain in the Tibetan Plateau: processes and mechanism. *Environmental Pollution*, 220:636–643.

Ross, G. (2004). The public health implications of polychlorinated biphenyls (PCBs) in the environment. *Ecotoxicology and Environmental Safety,* 59:275–291.

Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., Murthy, K.R.K. (2000). Improvements to the SMO algorithm for SVM regression." *IEEE transactions on neural networks*, 11:1188–1193.

Steinberg, J. (2020). *The 2020s Will be the Decade of Artificial Intelligence – And Huge Related Societal Challenges*. J. Retrieved January 5, 2020, from https://josephsteinberg.com/the-2020s-will-be-the-decade-of-artificial-intelligence-and-major-accompanying-societal-challenges/

Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L. (2019). Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Science of the Total Environment*, 653:140–147.

Stojić, A., Maletić, D., Stojić, S.S., Mijić, Z., Šoštarić, A. (2015). Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. *Science of the Total Environment,* 521:19–26.

Šoštarić, A., Stojić, S.S., Vuković, G., Mijić, Z., Stojić, A., Gržetić, I. (2017). Rainwater capacities for BTEX scavenging from ambient air. *Atmospheric Environment,* 168:46–54.

Team, (2012). *R: A Language and Environment for Statistical Computing.* http://cran.case.edu/web/packages/dplR/vignettes/timeseries-dplR.pdf.

Tonga, Y., Luo, K., Runkui, L., Lu, P., Ang, L., Mingan, Y., Qun, X., (2018). Association between multi-pollutant mixtures pollution and daily cardiovascular mortality: an exploration of exposure-response relationship. *Atmospheric Environment*, 186:136–143.

Tsygankova, V.Y., Lukyanova, O.N. (2019). Current Levels of Organochlorine Pesticides in Marine Ecosystems of the Russian Far Eastern Seas. *Contemporary Problems of Ecology,* 12:562–574.

United Nations Environment, (2017). *The 16 New POPs, An introduction to the chemicals added to the Stockholm Convention on Persistent Organic Pollutants by the Conference of the Parties*. Available at: ⟨http://chm.pops.int/⟩. Last (accessed November 2019).

United Nations Environment Programme, (2013). *Results of the global survey on concentrations in human milk of persistent organic pollutants by the United Nations Environment Programme and the World Health Organization, Conference of the Parties to the Stockholm Convention on Persistent Organic Pollutants*. Sixth meeting, Geneva, 28 April–10 May.

United States Environmental Protection Agency (USEPA), (2007). *EPA Unmix 6.0 Fundamentals and User guide*. USEPA Office of Research and Development.

Van den Berg, M., Birnbaum, L.S., Denison, M., De Vito, M., Farland, W., Feeley, M., Fiedler, H., Hakansson, H., Hanberg, A., Haws, L., Rose, M., Safe, S., Schrenk, D., Tohyama, C., Tritscher, A., Tuomisto, J., Tysklind, M., Walker, N., Peterson, R.E. (2006). The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicological Sciences*, 93:223–241.

Van den Berg, M., Kypke, K., Kotz, A., Tritscher, A., Lee, S.Y., Magulova, K., Fiedler, H., Malisch, R. (2017). WHO/UNEP global surveys of PCDDs, PCDFs, PCBs and DDTs in human milk and benefit–risk evaluation of breastfeeding. *Archives of Toxicology*, 91:83–96.

Vasios, G., Kosmidi, A., Kalantzi, O.-I., Tsantili-Kakolidou, A., Kavantzas, N., Theocharis, S., Giaginis, C. (2016). Simple physicochemical properties related with lipophilicity, polarity, molecular size and ionization status exert significant impact on the transfer of drugs and chemicals into human breast milk. *Expert Opinion on Drug Metabolism and Toxicology,* 11:1273–1278.

Vigh, É., Colombo, A., Benfenati, E., Håkansson, H., Berglund, M., Bódis, J., Garai, J. (2013). Individual breast milk consumption and exposure to PCBs and PCDD/Fs in Hungarian infants: a time-course analysis of the first three months of lactation. *Science of the Total Environment,* 449:336–344.

Voldner, E.C., Li, Y.L. (1995). Global usage of selected persistent organochlorines. *Science of the Total Environment*, 160:201–210.

Vuković, G., Herceg Romanić, S., Babić, Ž., Mustać, B., Štrbac, M., Deljanin, I., Antanasijević, D. (2018). Persistent organic pollutants (POPs) in edible fish species from different fishing zones of Croatian Adriatic. *Marine Pollution Bulletin* 137:71–80.

Waliszewski, S.M., Aguirre, A.A., Infanzon, R.M., Siliceo, J. (2002). Persistent organochlorine pesticide levels in maternal blood serum, colostrum, and mature milk. *Bulletin of Environmental Contamination and Toxicology,* 68:324–331.

Wang, Y., Witten, I.H. (1996). Induction of model trees for predicting continuous classes. In: *Proceedings of Poster Papers, 9th European Conference on Machine Learning* 1997.

Wania, F. and Mackay, D. (1996). Tracking the distribution of persistentorganic pollutants. *Environmenatal Science and Technology*, 30:390–396.

Wania, F., Axelman, J., Broman, D. (1998). A review of processes involved in the exchange of persistent organic pollutants across the air-sea interface. *Environmental Pollution*, 102:3–23.

Weka Sourceforge (2016). http://weka.sourceforge.net/packageMetaData, Accessed date: June 2016.

Yang, R., Wang, Y., Li, A., Zhang, Q., Jing, C., Wang, T., Wang, P., Li, Y., Jiang, G. (2010). Organochlorine pesticides and PCBs in fish from lakes of the Tibetan Plateau and the implications. *Environmental Pollution*, 158:2310–2316.

Zhao, Y., Zhang, Y. (2008). Comparison of decision tree methods for finding active objects. *Advances in Space Research,* 41:1955–1959.

BIOGRAPHICAL SKETCH

Gordana Jovanović

Affiliation: Institute of Physics Belgrade, a National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia; Environment and Sustainable Development Studies, Singidunum University, Belgrade, Serbia

Education:

2017−2018: Postdoctoral research; Biochemistry and Organic Analytical Chemistry Unit, Institute for Medical Research and Occupational Health, Zagreb, Croatia.

2012−2015: Doctoral academic studies in Chemistry; Department of Applied Chemistry, Faculty of Chemistry, University of Belgrade.

2011−2012: Master academic studies in Environmental Chemistry; Department of Applied Chemistry, Faculty of Chemistry, University of Belgrade.

2007−2011: Basic academic studies in Environmental Chemistry; Department of Applied Chemistry, Faculty of Chemistry, University of Belgrade.

Research and Professional Experience: Atmospheric chemistry and air pollution; Biomonitoring using moss, leaves, fish and human milk; Environmental multiphase systems; Heavy metals, Polycyclic Aromatic Hydrocarbons (PAHs), Persistent Organic Pollutants (POPs), Inorganic gases, Volatile Organic Compounds (VOCs); Analytical Chemistry (ICP-OES, ICP-MS, GC-MS, PTR-MS); Environmental data analysis (Advanced Statistics, Machine Learning, explainable Artificial Intelligence)

Research Appointments:

2017−present: Professor, Singidunum University, Belgrade, Serbia (Courses: Physical Chemistry, Analytical methods for environmental quality).

2016−present: Assistant Research Professor, Institute of Physics Belgrade, a National Institute of the Republic of Serbia, University of Belgrade.

2014−2016: Research Assistant, Institute of Physics Belgrade, a National Institute of the Republic of Serbia, University of Belgrade.

2013−2014: Research Trainee, Institute of Physics Belgrade, a National Institute of the Republic of Serbia, University of Belgrade.

Honors:

2017: Fellowship of Ministry of Science and Technological Development of the Republic of Serbia for postdoctoral research.

2015: The First award of the Foundation "Dr Milena Dalmacija", for the doctoral thesis which has made the greatest scientific contribution in the field of environmental protection at the universities in the Republic of Serbia during the period from October 2012 to October 2015.

2012: Special recognition of Serbian Chemical Society for remarkable achievement during the Basic academic studies at the Faculty of Chemistry.

Publications from the Last 3 Years:

Book Chapters

Stojić, A., Vuković, G., Perišić, M., Stanišić, S., Šoštarić, A., 2018. Urban air pollution: an insight into its complex aspects. In: *A Closer Look at Urban Areas,* Editor: Sahar Romero, Nova Science Publishers, NY, USA, ISBN: 978-1-63485-375-0. pp. 69-123.

Vuković G., Aničić Urošević, M., 2017. Is moss bag biomonitoring suitable method for assessment of intricate urban air pollution?, In: Aničić Urošević, M., Vuković G., Tomašević, M. (Eds.), *Biomonitoring of Air Pollution Using Mosses and Lichens, A Passive and Active Approach, State of the Art Research and Perspectives.* Nova Science Publishers, New York, NY, ISBN: 978-1-63485-375-0, pp. 27-74.

Book (Editor)

Aničić Urošević, M., Vuković G., Tomašević, M., 2017. *Biomonitoring of Air Pollution Using Mosses and Lichens, A Passive and Active Approach, State of the Art Research and Perspectives.* Nova Science Publishers, New York, NY, ISBN: 978-1-63485-375-0.

International Journals

Aničić Urošević, M., Jovanović, G., Stević, N., Deljanin, I., Nikolić, M., Tomašević, M., Samson, R., 2019. Leaves of common urban tree species (*Aesculus hippocastanum, Acer platanoides, Betula pendula and Tilia cordata*) as a measure of particle and particle-bound pollution: a 4-year study. *Air Qual. Atmos. Health* 12, 1081-1090.

Aničić Urošević, M., Krmar, M., Radnović, D., Jovanović, G., Jakšić, T., Vasić, P., Popović, A., 2020. The use of moss as an indicator of rare earth element deposition over large area. *Ecol. Indic.* 109, 105828.

Aničić Urošević, M., Vuković, G., Jovanović, P., Vujičić, M., Sabovljević, A., Sabovljević, M., Tomašević, M., 2017. Urban background of air pollution: Evaluation through moss bag biomonitoring of trace elements in Botanical garden. *Urban For. Urban Green.* 25, 1-10.

Aničić Urošević, M., Vuković, G., Vasić, P., Jakšić, T., Nikolić, D., Škrivanj, S., Popović, A., 2018. Environmental implication indices by elemental characterisation of the collocated topsoil and moss samples. *Ecol. Indic.* 90, 529-539.

Herceg Romanić, S., Vuković, G., Klinčić, D., Antanasijević, D., 2017. Self-organizing maps for indications of airborne polychlorinated biphenyl (PCBs) and organochlorine pesticide (OCPs) dependence on spatial and meteorological parameters. *Sci. Total Environ.* 628–629:198–205.

Herceg Romanić, S., Vuković. G., Klinčić, D., Matek Sarić, M., Župan, I., Antanasijević, D., Popović, A., 2018. Organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs) in *Cyprinidae* fish: Towards hints of their arrangements using advanced classification methods. *Environ. Res.* 165, 349–357.

Jovanović, G., Herceg Romanić, S., Stojić, A., Klinčić, D., Matek Sarić, M., Grzunov Letinić, J., Popović, A., 2019. Introducing of modeling techniques in the research of POPs in breast milk – A pilot study. *Ecotox. Environ. Saf.* 172, 341-347.

Milićević, T., Aničić Urošević, M., Relić, D., Vuković, G., Škrivanj, S., Popović, A., 2018. Integrated approach to environmental pollution investigation – spatial and temporal patterns of potentially toxic elements and magnetic particles in vineyard through entire grapevine season. *Ecotox. Environ. Saf.* 163, 245-254.

Milićević, T., Aničić Urošević, M., Relić, D., Vuković, G., Škrivanj, S., Popović, A., 2017. Bioavailability of potentially toxic elements in soil−grapevine (leaf, skin, pulp, and seed) system and environmental and health risk assessment. *Sci. Total Environ.* 626, 528-545.

Milićević, T., Aničić Urošević, M., Vuković G., Škrivanj, S., Relić, D., Frontasyeva, M, Popović, A., 2017. Assessment of species-specific and temporal variations of major, trace and rare earth elements in vineyard ambient using moss bags. *Ecotox. Environ. Saf.* 144, 208-215.

Šimić, I., Jovanović, G., Herceg Romanić, S., Klinčić, D., Matek Sarić. M., Popović, A., 2020. Optimization of Gas Chromatography-electron ionization-tandem Mass Spectrometry for Determining Toxic Non-ortho Polychlorinated Biphenyls in Breast Milk. *Biomed. Environ. Sci.* 33, 58-61.

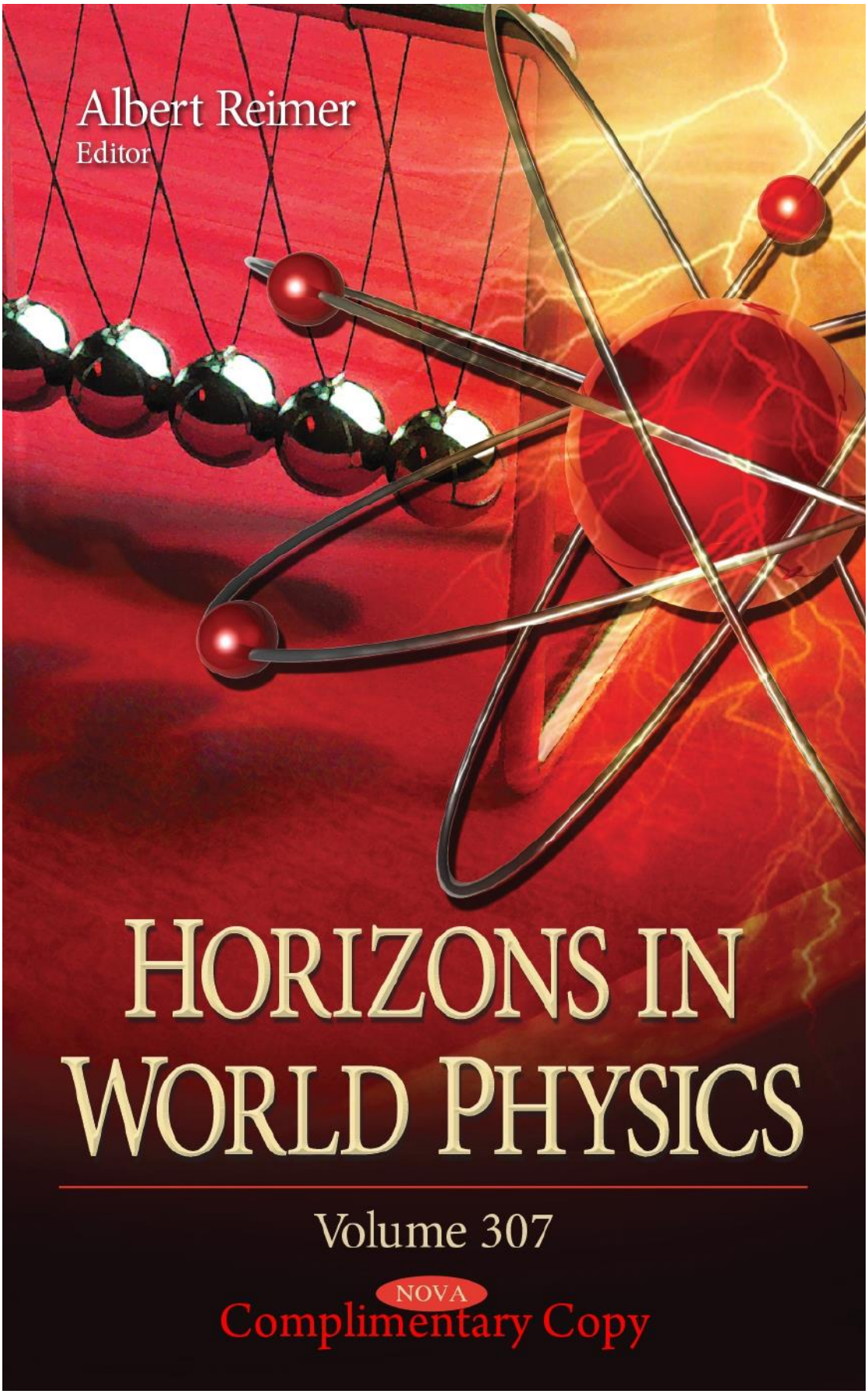
Šoštarić, A., Stojić, S.S., Vuković, G., Mijić, Z., Stojić, A., Gržetić, I., 2017. Rainwater capacities for BTEX scavenging from ambient air. *Atmos. Environ.* 168, 46-54.

Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Sci. Total Environ.* 653, 140-147.

Turgut, E., Gaga, E., Jovanović, G., Odabasi, M., Artun, G., Ari, A., Aničić Urošević, M., 2019. Elemental characterization of general aviation aircraft emissions using moss bags. *Environ. Sci. Pollut. Res.* 1-14.

Vuković, G., Aničić Urošević, M., Škrivanj, S., Vergel, K., Tomašević, M., Popović, A., 2017. The first survey of airborne trace elements at airport using moss bag technique. *Environ. Sci. Pollut. Res*.

Vuković, G., Herceg Romanić, S., Babić, Ž., Mustać, B., Štrbac, M., Deljanin, I., Antanasijević, D., 2018. Persistent organic pollutants (POPs) in edible fish species from different fishing zones of Croatian Adriatic. *Mar. Pollut. Bull.* 137, 71-80.


Albert Reimer
Editor
HORIZONS IN
WORLD PHYSICS
Volume 307
NOVA
Complimentary Copy

HORIZONS IN WORLD PHYSICS

VOLUME 307

HORIZONS IN WORLD PHYSICS

Additional books and e-books in this series can be found on Nova's website under the Series tab.

HORIZONS IN WORLD PHYSICS

VOLUME 307

ALBERT REIMER
EDITOR

nova
science publishers
New York

https://doi.org/10.52305/DKIM5863

NOTICE TO THE READER

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-68507-574-3 (eBook)
ISSN: 2159-2004

Published by Nova Science Publishers, Inc. † New York

Contents

In: Horizons in World Physics. Volume 307 ISBN: 978-1-68507-549-1
Editor: Albert Reimer

Chapter 3

EXPLAINING XYLENE WET DEPOSITION USING ARTIFICIAL INTELLIGENCE

Gordana Jovanović[1,2], Mirjana Perišić[1,2], Svetlana Stanišić[2] and Andreja Stojić[1,2,*]

[1]Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
University of Belgrade, Belgrade, Serbia
[2]Singidunum University, Belgrade, Serbia

ABSTRACT

The modern world is facing many environmental challenges. All environmental compartments are being affected by pollution and even with the immediate actions and measures it will take them a long time to recover. Currently, air pollution represents the big global burden due to the fact that it accounts for an estimated 4.2 million deaths and 103.1 million disability-adjusted life years annually.

The commonly monitored pollutants include toxic and mutagenic species such as volatile organic compounds and their representatives –

* Corresponding Author's Email: andreja.stojic@ipb.ac.rs.

benzene, toluene, ethylbenzene, and xylene. Apart from photochemical reactions or dispersion, BTEX removal from ambient air can be affected by atmospheric precipitation. However, the current understanding of the wet deposition role in biogeochemical cycles of volatile organic compounds is insufficient. The removal process depends on the ambient concentrations and water solubility of the pollutant, surface interactions, Henry's law, physicochemical properties of the water, meteorological factors, as well as on the frequency and intensity of precipitation events.

Modeling of natural environment complexity, including the complexity that results from the interrelations between the environmental compartments, spatio-temporality, and high order interactions, is hardly achievable without advanced data processing. Therefore, artificial intelligence, enhancing data-driven research, becomes one of the most powerful frameworks capable of capturing defining factors and processes that govern the environmental evolution of polluting species.

This chapter resumes the findings of several multi-perspective analyses of the role of the ultrapure water, urban environment rainwater, and the phase-specific factors in shaping the xylene environmental fate, and focuses on the processes that govern xylene partition between the gaseous and aqueous phase. In this study, we have employed machine learning for revealing the transfer of the xylene from ambient air to rainwater and an explainable artificial intelligence framework for investigating the relevance of the monitored parameters and identification of the key factors that govern wet deposition of the xylene.

Keywords: multiphase systems, rain, volatile organic compounds (VOC), xylenes, extreme gradient boosting machine (XGBoost), Shapley Additive exPlanations (SHAP)

INTRODUCTION

Xylene Occurrence in the Environment

Xylene, also identified as xylol and dimethyl benzene ($C_6H_4(CH_3)_2$), belongs to volatile organic compounds (VOCs) and it is a well-recognized synthetic pollutant from the group of aromatic hydrocarbons known as BTEX (benzene, toluene, ethylbenzene, and xylenes). Xylene occurs in three different forms (meta-xylene, ortho-xylene, and para-xylene (m-, o-, and p-xylene)) referred to as isomers depending on the position of the

methyl group on the benzene ring. Commercially available xylene represents a mixture that usually contains about 45–65% m-xylene, 10–25% o-xylene, and 6–15% of each p-xylene, and a small amount of ethylbenzene (US EPA, 2003). In the available literature regarding environmental investigations, the terms xylene, xylenes, and total xylenes are usually used with the same meaning. In this chapter, the term xylene will refer to the sum of all three isomers.

In 2019, approximately 56 million metric tons of xylene were supplied globally while in 2021, the growth to about 61 million metric tons was expected. The global xylene market consisting of several regions (North and South America, Asia-Pacific, Western and Eastern Europe) will experience considerable growth over the next five years, potentially increasing from 73.38 mtpa in 2020 to 113.43 mtpa in 2025, with the world's leading xylene-producing countries being United States, Brazil, Japan, South Korea, China, Russia, United Kingdom, France, and Germany. In terms of volume, xylene is among the top 30 chemicals manufactured in the United States (US EPA, 2000). It is widely used as a solvent to produce agricultural sprays, inks, lacquers, varnishes, adhesives, rubbers, and gums. It is also employed as a thinner for paints and as an ingredient in the coating of fabrics and papers coatings. Xylene represents a feedstock in manufacturing various polymers, polyesters, ethylbenzene, phthalic anhydride, isophthalic acid, terephthalic acid, and dimethyl terephthalate (Clough, 2014). A small portion of the xylene is added to gasoline to improve its octane ratings. Aviation fuel and liquid gasoline typically contain between 5.8 and 15.8%w/w xylene while a xylene content in the vapour of gasoline varies from 0 to 2.1%w/w (Chilcott, 2007). In addition, petroleum and coal naturally contain xylene and to a small extent, it can be emitted in the air during wildfires.

Xylene is a colorless, flammable liquid that has a sweet odor. It is released in the environment from industrial sources (factories and refineries), traffic exhaust, waste disposal sites, or spills, and during its use as a solvent in the production, packaging, and shipping processes (ATSDR, 2007). However, the most probable source of xylene emissions is related to leakage from underground gasoline storage tanks and leachate from

landfills (Su et al., 2010). As a liquid, xylene could contaminate soil, surface waters including streams, rivers, lakes, wetlands, and creeks as well as groundwaters and drinking water wells. Like the other chemicals from the BTEX group, xylene is sparingly soluble in water, but it mixes with alcohols and other solvents and possesses a high octanol/water partition coefficient (K_{OW}). Because of its physicochemical properties including volatility, it evaporates easily and enters the atmosphere, partially dissolves in atmospheric water, or migrates into the soil pores filled with air. The volatilization dynamics of xylene is complex and differs depending on the matrix; but when compared to benzene or toluene, the rate of xylene and ethylbenzene volatilization from water, sand and soil mixture is shown to be lower (Varona-Torres et al., 2017). In groundwater and soils at large depths, xylene may resist up to several months until it is finally decomposed by microorganisms while insignificant amounts could be uptaken by plants or fish species. If not trapped by underground constituents or within water bodies, xylene reaches the air rather than being retained in topsoil or surface water. Once in the atmosphere, xylene is expected to be decayed within several days into less harmful compounds via photochemical reactions governed by UV radiation. In addition, compared to benzene, toluene, and ethylbenzene, xylene is highly reactive, and it may be involved in the formation of tropospheric ozone.

In urban areas worldwide, mean xylene air concentrations typically range from 3 to 390 $\mu g\ m^{-3}$, while the outdoor levels up to 15.45 $\mu g\ m^{-3}$, 32.4 $\mu g\ m^{-3}$, 61 $\mu g\ m^{-3}$, and 100 $\mu g\ m^{-3}$ were measured in India (Singh et al., 2012), Algeria (Kerchich & Kerbachi, 2012), the United States (ASTDR, 2007; US EPA, 2000), and Greece (Alexopoulos et al., 2006), respectively. In the outdoor environment, the xylene peak concentrations were recorded near crossroads, roadsides, and gasoline stations (Alexopoulos et al., 2006; Kuranchie et al., 2019, Soltanpour et al., 2021). In the indoor environment, xylene concentrations can reach 48 $\mu g\ m^{-3}$ and even 200 $\mu g\ m^{-3}$ in cigarette smoke (WHO, 2003; Alexopoulos et al., 2006; Niaz et al., 2015).

The most common sources that cause xylene contamination of groundwaters are underground gasoline storage tanks, leakage from

landfills, and release from factories and refineries. Although xylene is relatively insoluble in water, levels up to 1000 mg L^{-1} have been recorded in groundwaters, which are far beyond the allowed maximum xylene levels of 10 mg L^{-1} (Su et al., 2010). However, information about the amount of xylene in surface water and soil is scarce. Levels of xylene in drinking water were far below those found in other types of water. As reported by World Health Organisation (WHO, 2003), xylene concentrations in drinking water in North America did not exceed US EPA maximum contaminant level of 10 ppm and were reported to be 5.2 µg L^{-1}, < 1 µg L^{-1}, < 0.5 µg L^{-1} and 8 µg L^{-1} in the USA, Canada, Ontario, and New Orleans (WHO, 2003).

Human Exposure to Xylene

According to the International Agency for Research on Cancer xylene belongs to Group 3, which means that the evidence of xylene carcinogenicity is inadequate in humans and experimental animals and further research is necessary (IARC, 2021). With respect to this, the regulatory standards were established regarding occupational exposure to xylene. The National Institute of Occupational Safety and Health, and Occupational Safety and Health Administration set threshold limit values of 150 ppm and 100 ppm in the workplace environment during an 8-hour workday and 40-hour week.

Human exposure to xylene mostly occurs in painting industry facilities, biomedical laboratories, distilleries of xylene, wood processing plants, gas stations, etc. Most people could smell xylene at concentrations ranging from 0.08 to 3.7 ppm in air and from 0.53 to 1.1 ppm in water (ATSDR, 2007). Exposure to xylene also refers to the inhalation of vapours or skin contact during the usage of gasoline, paints, solvents, varnishes and cigarettes. Rarely, people could be exposed to xylene through drinking water or contaminated food, due to the low xylene concentrations of 2 ppb and 100 ppb in these matrices, respectively.

Once inhaled, xylene primarily reaches the mucous membranes and lungs from where approximately 60% is immediately absorbed and passes into the bloodstream (Clough, 2014). In addition, xylene could be absorbed in the gastrointestinal tract including the gut, and further distributed to many lipid-rich organs and tissues in the body. Skin contact with xylene could be also followed by absorption, although this pathway accounts for only about 12% of the xylene absorption by the respiratory system. The biotransformation of xylene occurs in the liver and lungs where more water-soluble compounds (o-, m-, and p-toluic acids, methyl benzoic acids, and/or methyl hippuric acid) are produced through the enzyme-catalyzed oxidation of a side chain methyl group. Another possible transformation of xylene includes conjugation with glycine to yield methyl hippuric acid (Rajan & Malathi, 2014). A small amount of xylene is exhaled unaltered within a few seconds after absorption, while more than 95% is metabolized and the metabolites are rapidly excreted from the body via urine after 18 to 24 hours (Clough, 2014).

Up to 10% of the absorbed xylene may be deposited in the body fats, which may prolong the time taken for the xylene excretion (ASTDR, 2007). Xylene could cross the placenta and reach the fetus, or enter breast milk having adverse effects on embryogenesis, fetal development and newborns. In addition to this, studies have shown that the metabolic breakdown of xylene in fetal bodies is very limited (Sirotkin, 2019).

The research has revealed that all xylene isomers have similar effects on humans and animals, but specific isomers appeared to be not equally potent in terms of adverse health effects. There is no evidence on the health impacts when people are exposed daily to xylene background levels. Nevertheless, short-term (acute) exposure to high xylene levels can cause breathing difficulties and lung malfunctions, skin, eyes, nose, and throat irritations, gastrointestinal discomfort, neurological effects and possible alterations in the liver and kidney functioning, while long-term (chronic) exposure primarily affects the central nervous system causing headache, dizziness, tremors, muscle incoordination and spasms, impaired short-term memory, inability to concentrate, central auditory effects and poorer sound

detection, balance problems, and respiratory, cardiovascular, and kidney damages (Fuente et al., 2013; Niaz et al., 2015; Rajan & Malathi, 2014).

Information on xylene adverse health effects was mostly obtained from the studies in which subjects were occupationally exposed to very high levels of xylene in the air, the levels significantly exceeding those that are commonly registered in the outdoor and indoor environment. The type and strength of the effects, in particular depended on different parameters including exposure route, duration, and individual susceptibility (Niaz et al., 2015). Specific non-carcinogenic effects (hematological, cardiovascular, respiratory, neurological, immunological, reproductive, renal, and musculoskeletal) of the exposure to xylene have been reported in many studies, but it should be emphasized that the findings were not obtained from the exposure to xylene only and the effects of other volatile organic compounds could not be excluded (Bahadar et al., 2014).

Long-term exposure to pollutant mixtures, which contain benzene, xylene, and toluene at low concentrations, may lead to DNA damage in the peripheral blood cells and oxidative stress (Huang et al., 2010; Tang and Xu, 2005). With respect to hematological parameters, prolonged exposure to xylene causes a decrease of hemoglobin, total leukocytes, and lymphocytes as well as an increase in the blood enzyme levels like glutamatic oxaloacetic transaminase and alkaline phosphatase. The studies observing workers, who were exposed to xylene concentrations of up to 14 ppm, implied indirect effects on muscles like a decrease in muscle strength and mass (Niaz et al., 2015; Rajan and Malathi, 2014; and references therein). Furthermore, many studies have revealed that xylene exposure can cause disruption of nervous system functions. Well-recognized xylene-related neurological alterations include coma, anesthesia, reduced memory capacities, incoordination, spasms, cramps, and a slow response to external stimuli (e.g., Jovanović et al., 2004).

Out of impacts on the gastrointestinal tract, gut discomfort, vomiting, and nausea were observed after two-hour and two-week exposure to organic mixtures containing xylene (Ernstgård et al., 2002). Although adverse effects of other organic pollutants should be accounted for, it has been reported that acute exposure to xylene induced a number of

cardiovascular and respiratory dysfunctions. As reported in the review of Niaz et al. (2015), in one out of nine participants, tachycardia was detected in human subjects being exposed to 200 ppm of xylene mixture for 3 to 5 minutes. Irritation of respiratory mucosa and reduction of the cytochrome levels (CYP 2B1, 2E1, and 4B1) in lungs have been observed as well (Vaidyanathan et al., 2003).

In the liver, the common xylene metabolic pathway includes the enzyme CYP2E1, which converts xylene through a series of enzymatic reactions to methyl hippuric acid. Additionally, other enzymes govern the hydroxylation of xylene to 2,4 dimethyl phenol. Histopathological investigations have confirmed that exposure to high xylene levels could be associated with increased production and activity of hepatic enzymes such as CYP2E1. Other studies implied that inhalation, oral or dermal occupational exposure to xylene could lead to liver apoptosis, necrosis, elevated concentrations of serum transaminases, the increase in the endoplasmic reticulum and autophagous bodies, and change in the propagation of hepatocellular nuclei. Besides liver, kidney functioning can be also impaired by xylene exposure. Various studies have reported an increased acidity of kidney tubules, increased creatinine levels in the urine, and hematuria due to the accumulation of m-xylene in perirenal fats (Niaz et al., 2015; Rajan and Malathi, 2014; and references therein).

In terms of reproductive system function and toxicity of xylene, a few epidemiological studies have reported adverse impacts on reproductive hormones like reduction of pregnanediol 3-glucuronide, luteinizing hormone, and estrone 3-glucuronide, and higher follicle phase pregnanediol 3-glucuronide. Although participants were exposed to solvents containing different organic aliphatic and aromatic hydrocarbons, the contribution of xylene to the observed effects is considered to be over 50% (Reutman et al., 2002).

Xylene in Multiphase Systems

The environmental fate of VOCs including BTEX group is governed by complex processes of their physico-chemical transformations and transport as well as by meteorological conditions. The occurrence, behaviour, and distribution of xylene in the environment have attracted scientific attention over several decades because of its abundance and adverse effects on human health as described above. These processes are strongly dependent on the xylene partitioning in the multiphase systems, usually two-component systems consisting of different phases, like air-water, air-rainwater/snow, air-soil, air-suspended particles/aerosol and soil-water (Starokozhev et al., 2009; 2011). The role of atmospheric water appeared to be of great importance for the investigation of VOC distribution within multiphase systems but still the point of scientific contention (McNeill et al., 2012).

Theoretically, the extent of xylene partition between the gaseous and aqueous phases in the atmosphere depends on the thermodynamic properties of a system and partition process dynamics which could be characterized by interphase partitioning constants, primarily Henry's law constant (K_H), but also on temperature and salts, organic substances, and acids dissolved in water (Kampf et al., 2013; Kurtén et al., 2015; Sander, 2015). However, previous findings indicated that the concentrations of VOCs recorded in rainwater and fog samples are higher than those predicted by Henry's law (e.g., Okochi et al., 2005; 2004) and several explanations have been suggested with an attempt to explain the observed deviations. Okochi et al. (2004) assumed that surface-active substances in rain droplets tend to bond the pollutants while Sato et al. (2006) proposed that dissolved organic compounds from ambient air cause the supersaturation of VOCs in rainwater during wet deposition. In addition, some investigations implied that interactions between pollutant and water surface, such as hydrogen bonding, surface adsorption and van der Waals forces, as well as physico-chemical properties of the rainwater may play an important role in the pollutant partitioning (Allou et al., 2011; Goss, 2004; Roth et al., 2004). Conversely, some authors concluded that VOC

susceptibility to UV-induced photochemical reactions with free hydroxyl and nitrogen radicals and their involvement in aerosol formation appeared to be of importance for understanding their removal from the atmosphere and distribution in the air-water system (Słomińska et al., 2014; Mullaugh et al., 2015). In addition, Starokozhev et al. (2011) claimed that particles (mineral dust, diesel soot, urban dust) dispersed in the water phase can enhance adsorption on solid/liquid interfaces and thus, determine the VOC partition between aqueous and gaseous phases. The experimental data revealed that the interaction of VOCs with solid interfaces depended on the pollutant functional groups, the length of the hydrocarbon chain, the surface area, and the affinity between the surface and the compound.

Over recent years, the application of data science including artificial intelligence (AI) implemented in machine learning (ML) methods has drawn an increased attention in environmental sciences as well as the corresponding control and management systems (UN, 2018; Blair et al., 2019). Data-science driven approaches possess an enormous potential to improve the current understanding of environmental phenomena since conventional data analysis methods are usually insufficient to cope with the size and diversity of environmental data. The advanced methods offer modelling of the environmental complexity that arises from the interrelations between the compartments, pollutants and different variables, spatio-temporality, and high order interactions (Gibert et al., 2018).

Tremendous advances in interpretation modelling methods have been developed to provide valuable insights into scientific issues unattainable to human intuition and domain knowledge (Montavon et al., 2017). Until recently, interpretation frameworks lead to inconsistent results and often contradictory explanations for machine learning algorithms (Lundberg et al., 2018). Therefore, Shapley Additive exPlanations (SHAP), an additive feature attribution method, was introduced. It has a unique solution in the class of explanation models aimed at post-hoc interpreting machine learning methods and is more aligned with human intuition (Lundberg and Lee, 2017). Unlike approaches that provide a specific global predictor,

SHAP provides an explanation of the model overall behavior in the form of particular feature contributions.

In this chapter, we summarize the findings from several laboratory and field measurements regarding xylene partition between aqueous and gaseous phases (Šoštarić et al., 2016; Šoštarić et al., 2017; Stojić et al., 2018; Stojić et al., 2019). We employed the SHAP algorithm to investigate the functional dependency between xylene enrichment factor (EF) in rainwater and its concentrations in air, physico-chemical properties of rainwater (pH, turbidity, UV extinction, electrical conductivity, total organic carbon, anions: F^-, Cl^-, SO_4^{2-}, NO_2^-, and NO_3^-, and cations: Na^+, NH_4^+, K^+, Ca^{2+}, and Mg^{2+}) and a number of meteorological parameters (temperature, relative humidity, pressure, wind speed, wind direction, and rainfall intensity/amount).

METHODOLOGY

Measurements

BTEX concentrations in both gas and water phases were measured using a proton transfer reaction mass spectrometer (Standard PTR-quad-MS, Ionicon Analytik, GmbH, Austria), whose detailed description is given elsewhere (Lindinger and Jordan, 1998). Since PTR-quad-MS is not capable of distinguishing isobaric ions, the signal detected at m/z 107 referred to C_8 aromatic hydrocarbons, ethylbenzene, o-, m-, and p-xylene, while the signals detected at m/z 79 and m/z 93 referred to benzene and toluene, respectively (Warneke et al., 2003). The obtained PTR-MS signal was subjected to baseline fitting and, depending on the stage, the time interval required for equilibrium level to be achieved is taken for insufflation (t_i), or exsufflation time (t_e). For each insufflation cycle, the quantity of target compounds retained from the gaseous phase (Q_i^V) is calculated as the integral of the area between the obtained PTR-MS signal and the equilibrium level achieved for ti. The time required for the target compounds to be completely purged out from the aqueous phase (t_{eq}) is

also determined so that the quantity of target compounds during exsufflation stage could be calculated. The details on baseline extraction, best fitting baseline tests, and other statistical methods in data treatment were described in Šoštarić et al. (2016).

Field Campaign

Sample pairs of air and rainwater samples were collected simultaneously during several rain events that occurred over two distinct time periods in the summer and autumn seasons of 2015. The sampling was performed at the Institute of Physics Belgrade (Serbia; 44°49' N, 20°28' E), located in the vicinity of the Danube River, in the suburban residential area, with a number of local fireboxes active during the heating season. Determination of BTEX concentrations in rainwater was performed immediately after each sampling campaign. A detailed description of the measurement campaign is provided in Šoštarić et al. (2017).

To determine the extent to which Henry's law constant (K_H) describes BTEX distribution between the gaseous and aqueous phase, distribution coefficients (D_{OBS}) were calculated for each air/water sample pair and each species, as the ratio of the corresponding experimentally derived rainwater concentrations in nM (C_R) and ambient gas-phase mixing ratios in ppbV (p_g):

$$D_{OBS} = {}^{C_R}/_{p_g} \tag{1}$$

Considering the K_H temperature dependence, enrichment factors (EF) were calculated using D_{OBS} and temperature corrected K_HT for each rain sample by means of the following equation (Sander, 2015):

$$K_HT = K_H(298.15)exp\left\{\frac{-\Delta H}{R}\left(\frac{1}{298.15} - \frac{1}{T}\right)\right\} \tag{2}$$

where K_H is Henry's law constant at 298.15 K for pure water, ΔH is the

enthalpy change of air-water transfer, T is the rainwater temperature, and R is the universal gas constant (8.314 J K^{-1} mol^{-1}).

To assess the representativeness of ground-level conditions for the atmospheric conditions during rainfall, K_HT and EF altitude profiles were calculated using the temperature profiles obtained from GDAS1 (Global Data Assimilation System, 2015), by replacing the rainwater temperature value by the temperature at the corresponding altitude.

Data Analysis

Statistical Analysis

The relationships between enrichment factors (EFs), physicochemical characteristics, and wind characteristics were examined using the bivariate polar plot analyses (Carslaw and Beevers, 2013) implemented in the "openair" package (Carslawand and Ropkins, 2012) within the statistical software environment R (Team, 2012).

Potential remote source regions that might affect the observed BTEX mixing ratios were identified using HYSPLIT-derived 72-h back trajectories (Draxler and Rolph, 2014). The trajectories were computed and processed according to Stojić and Stojić (2017).

Rainwater source apportionment was performed using Unmix (US EPA, 2007). The maximum number of species selected as input variables was chosen using trial and error with the overall aim of yielding the most physically meaningful results.

Machine Learning

For estimating the relationships between TEX rainwater concentrations and EFs on one hand, and TEX ambient air concentrations, and physicochemical and meteorological parameters, on the other, regression analysis using eXtreme Gradient Boosting (XGBoost) in Python and many multivariate methods (MVA) implemented in Weka 3.8 (Alternating Model Tree, Conjunctive Rule, Decision Stump, Decision Table, Elastic Net, Gaussian Processes, IBk, IBkLG, Isotonic Regression,

K*, Least Median Squared, Linear Regression, Locally Weighted Learning, M5P, M5 Rules, Multilayer Perceptron, Pace Regression, Random Forest, Random Tree, Radial Base Function (RBF) Network, RBF Regressor, REP Tree, Simple Linear Regression, and SMOreg Support Vector Machine; Hall et al., 2009) was implemented. The details on the applied methods were provided in Šoštarić et al. (2017) and Stojić et al. (2019). Hyperparameter tuning for all regression methods was implemented by using a brute-force grid search and 10-fold stratified cross-validation which was replicated 10 times. The best performing hyperparameter values were used for the final model, while the best performing models were used for further analyses.

Explainable Artificial Intelligence

The ability to accurately interpret a model's prediction supports a deeper understanding of the process being modeled. For this purpose, we employed guided regularized random forest (GRRF) and the advanced explainable artificial intelligence method SHAP, which is capable to provide the straightforward and meaningful interpretation of the ML model-derived decisions, now being shifted towards user-readable logic rules. SHapley Additive exPlanations is a method based on Shapley values, calculated as a measure of feature importance using a game-theory approach (Lundberg and Lee, 2017a and 2017b). The Shapley value method provides fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). It perfectly apportions the difference between the prediction and the average prediction among the features (Molnar, 2018). The Shapley explanations represent the only possible locally accurate and globally consistent feature contribution values (Chen et al., 2019; Stojić et al., 2019). The method overcomes the main drawback of inconsistency and minimizes the possibility of underestimating the importance of a feature with a certain attribution value. We used Python SHAP implementation.

RESULTS AND DISCUSSION

Partition of Xylene between Gaseous and Aqueous Phase in a Laboratory-Designed Dynamic System

The mechanisms which govern the air-water distribution of xylene were investigated in the study with a dynamic dilution system (DDS), designed to resemble interactions between the two phases during rainfall, as described in the Material and Methods section. The overall aim of the research was to examine whether BTEX dissolution in rainwater is in accordance with Henry's law based on the empirically determined distribution coefficient (D_{OBS}), and calculated enrichment factors (EFs). Factors other than dissolution, which possibly influence the uptake of xylene in rainwater, were addressed as well. Herein, we present the main findings of the study while more details were given elsewhere (Šoštarić et al., 2016).

To describe xylene beheviour in the two-component system, we considered the following physico-chemical properties of the compound: molecular weight, water solubility, K_{OW}, ionization potential (IP), total van der Waals surface area (AvdW), the calculated quantity of uptaken species (Q_i^V), the pollutant concentrations in water (C_i^V), as well as K_H, D_{OBS}, and EFs as explained elsewhere (Šoštarić et al., 2016; Stojić et al., 2018). Xylene concentrations in ultrapure water were shown to be linearly dependent on its gas mixing ratio and these observations obey Henry's law. However, the ratio of xylene uptake in different rainwater samples was in the range of 1.7 and 1.9, which is lower than 2, the value predicted by Henry's law. In addition, the average value of EFs calculated for xylene was 42.6, which is significantly higher than 1, implying that the compound concentrations in the aqueous phase are higher than predicted by Henry's law. Although less soluble than benzene or toluene, xylene exhibited the highest EF value. This is in accordance with the findings of Starokozhev et al. (2009) who concluded that the removal efficiency of species like m/p-xylene and o-xylene from snow is much higher than expected having in mind their low water solubility and high volatility.

According to the available literature, scavenging of gaseous VOCs by atmospheric water is dependent on their K_H values, which is highly influenced by: temperature, pH, mixture constituents, and consequently, the presence of surfactants, suspended particles, dissolved salts such as chloride or ammonia, organic material including oil or humic-like substances, but also on the compound ability to form hydrogen bonds or undergo adsorption processes (Allou et al., 2011; Okochi et al, 2004; Mullaugh et al., 2015; Xue et al., 2015). However, since we performed the experiment at a constant temperature of 25ºC and used 18 MΩ ultrapure water, the influence of temperature and water composition are irrelevant for the explanation of the observed significant xylene enrichment. From the study of Valsaraj et al. (1993), several investigations reported that surface area between two phases (air and water) and surface-related processes could not be neglected when exploring VOC partition in multiphase systems. For instance, Goss (2004) noted that the chemisorption and formation of hydrogen bonds between VOC and atmospheric water are the main mechanisms promoting air-water partition, while Roth et al. (2004) drew out hydrogen and van der Waals bonding as the governing interactions between VOCs and snow crystals. We assumed that presented observations could offer a possible explanation for the xylene concentrations in water, which exceeded those predicted by Henry's law.

Chemical species like BTEX have a lone electron pair and participate as the acceptors during the establishment of hydrogen bonds with water. The acceptor capability of molecules depends on their IP, which reflects the tendency of a compound to release electrons, and among BTEX, xylene has the lowest IP value (Furutaka and Ikinawa, 2002). In our study, the IP of BTEX was negatively associated with EF values which implies that hydrogen bonding is not of relevance with respect to the partition of xylene between gaseous and aqueous phases. Contrary, a high correlation existed between EF and parameters that characterize interfacial adsorption like AvdW as well as hydrophobicity indicator–K_{OW}. The results are in compliance with the findings of Okochi et al. (2004), who reported that the EFs of chlorinated hydrocarbons and monocyclic aromatic hydrocarbons

in rainwater tend to increase with the rise of the Kow. As discussed above, Roth et al. (2004) suggested the relationship between interfacial adsorption and van der Waals interactions regarding BTEX uptake in snow, while Valsaraj et al. (1993) reported that raindrops possess smaller specific area compared to fog droplets and therefore, they are one to two orders of magnitude less enriched in hydrophobic compounds. In our study, the spreading of gas bubbles was more intense in a 250 mL gas-washing bottle compared to those of 500 mL due to the smaller diameter. This led to the increased frequency of collisions between water and gas molecules and consequently, to enhanced interactions on the surface of phases and a greater enrichment. Finally, we assumed that van der Waals interactions are probably the major phenomenon leading to the interfacial adsorption and enhanced xylene uptake in water.

Explainable Prediction of Rainwater Capacities for Xylene Scavenging from the Atmosphere

After exploring theoretically based assumptions related to xylene partitioning observed between aqueous and gaseous phases through laboratory simulations, our subsequent investigations aimed to better understand the contribution of rainwater to wet removal of BTEX from the atmosphere. For this purpose, field measurements were performed, and the xylene scavenging was discussed with respect to its EF in wet deposition and the relationship with the most relevant variables including physico-chemical properties of rainwater and meteorological parameters (Šoštarić et al., 2017) while the SHAP algorithm was employed to obtain a more detailed insight into the factors that govern the air-rainwater distribution of xylene (Stojić et al., 2019). As shown by Lundberg and Lee (2017), the SHAP method offers a unique, consistent, and accurate solution for the explanation and interpretation of ML algorithms. The SHAP feature attributions, SHAP summary plots and partial SHAP dependency and interaction plots enabled to reveal interaction patterns between xylene and

the environmental parameters, and moreover, to identify in which respect the studied predictors affected the investigated variable.

Field measurements during summer and autumn revealed higher xylene concentrations in both the aqueous and gaseous phases during the cold days. A decrease in the toluene and xylene air mixing ratios and rain concentrations was observed immediately after the rain began, during the first 2 h of the rainfall, but with the tendency to rise afterwards probably due to the decreased rainfall intensity. The cold period was characterized by the following meteorological conditions: light and sporadical rain events, air masses coming from N and NE direction, and scattered clouds below 300 m, while broken clouds were registered at levels from 100 to 900 m. According to the physico-chemical properties, the rainwater was mainly almost neutral (average pH = 6.01; pH range = 3.70 to 8.20) containing dominantly alkaline and neutralizing ions: Ca^{2+} followed by NH_4^+, Mg^{2+}, K^+, and SO_4^{2-} as well as moderate amounts of suspended particles (NTU < 10). A high abundance of most ions was associated with the high-speed wind (20–30 m s^{-1}) from SW direction, with exception of NH_4^+ which concentrations were increased during the episodes of moderate wind (ws ≈ 10 m s^{-1}).

As expected, individual combustion sources spread locally around the sampling site contributed to the xylene abundance in autumn whereas intense photochemical degradation during summer was associated with lower levels of xylene. Unmix source apportionment identified four emission source factors out of which xylene exhibited the highest share in two of them. The first factor was dominated by benzene (99%), toluene (44.5%), xylene (52.2%), and UV extinction (33.3%) indicating common sources of volatile organic species. The result was confirmed by a high correlation between aqueous and gaseous BTEX levels (>0.80), as well as by ratios between xylene and toluene (0.67), and xylene and benzene (0.67). In addition, significant shares of xylene (>30%) along with ions, gaseous oxides (SO_2 and NO_2), and other VOCs were apportioned to an aerosol fraction suggesting BTEX susceptibility to photooxidation in the presence of ozone and hydroxyl free radicals, but also its possible involvement into formation of secondary organic aerosol through reactions

with SO_2 and NO_X. Wind/air masses of moderate speed from the N direction were mostly related to the contributions of BTEX- and aerosol-related Unmix sources indicating their local origin.

Although BTEX concentrations in water are expected to be low because of low solubility and Henry's constant, xylene concentrations in the rainwater significantly exceeded theoretically predicted values as implied by the EF values from 25 to 295 while the ranges for benzene and toluene were from 61 to 128 and from 8 to 209, respectively. This is in accordance with the results of our laboratory simulations described in the previous section, which showed that that levels of BTEX in ultra-pure water exceeded theoretically calculated values of BTEX transfer between air-water phases. In addition, these findings accompany previous observations regarding the dissolution of organic volatiles in various aqueous environmental samples (groundwater, snow, fog, dew, rainwater). Almost two decades ago, Delzer et al. (1996) were among the first who demonstrated that total xylenes along with toluene are the most prevalent volatiles in urban stormwater. These observations were amended by Baehr et al. (1999), who reported that BTEX concentrations in the shallow groundwater were higher than expected. Later, some studies, which investigated the transfer of organic aromatics from air to rainwater or dew, reported BTEX enrichment from 2.4 to 34 (Okochi et al., 2004; Okochi et al., 2005; Sato et al., 2006), whereas fog samples were found to exhibit several hundred to a thousand-fold enrichment of hydrophobic organic species (Valsaraj et al., 1993).

Besides discussed observations, our studies aimed to strengthen the existing knowledge by exploring how meteorological conditions contribute to xylene partition in an air-water multiphase system. In addition, we investigated the relationships between the pollutant enrichment and the main inorganic constituents of rainwater (anions and cations), the rainwater's physico-chemical properties (pH, NTU, TOC, EC, and UV extinction TOC), and adsorption at the air/water interface as discussed in the previous section. As the raindrop falls to the ground, i.e., at lower altitudes, an increase in xylene enrichment and a decrease of K_H were observed. As suggested by the study of Okochi et al. (2005), which

compared dissolution of VOCs in urban dew and rainwater, the pollutant enrichment is expected to be higher in dewdrops generated near the ground with high content of humic-like species. The presence of these substances reduces the surface tension of water and promotes the transfer of VOCs into an aqueous phase. In contrast to this, Mullaugh et al. (2015) suggested that high volatility, hydrophobic nature, and short lifetimes of the organic aromatics weaken the capacity of wet deposition for the removal of the compounds from the lower atmospheric layers.

In order to explore nonlinear relationships between the xylene concentrations in rainwater and the other investigated factors, an XGBoost tree-based method was successfully applied with the predicted/observed calculated relative errors of 20% and the correlation coefficients of 0.93. As shown in Figure 1, the SHAP importance score (>15) suggested that the ambient air toluene, ethylbenzene and xylene concentrations followed by rain temperature, TOC, pressure, concentrations of Cl^-, ethylbenzene, and Ca^{2+} in rainwater appeared to be by far the most important predictors of for xylene enrichment in rainwater.

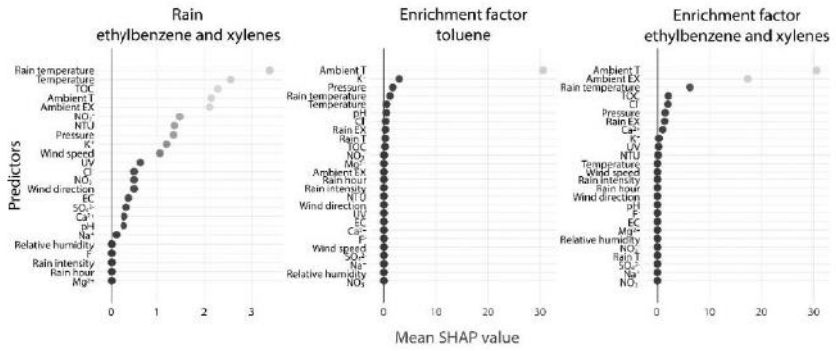


Figure 1. SHAP importance scores for rainwater ethylbenzene/xylene concentrations and toluene and ethylbenzene/xylene enrichment factors.

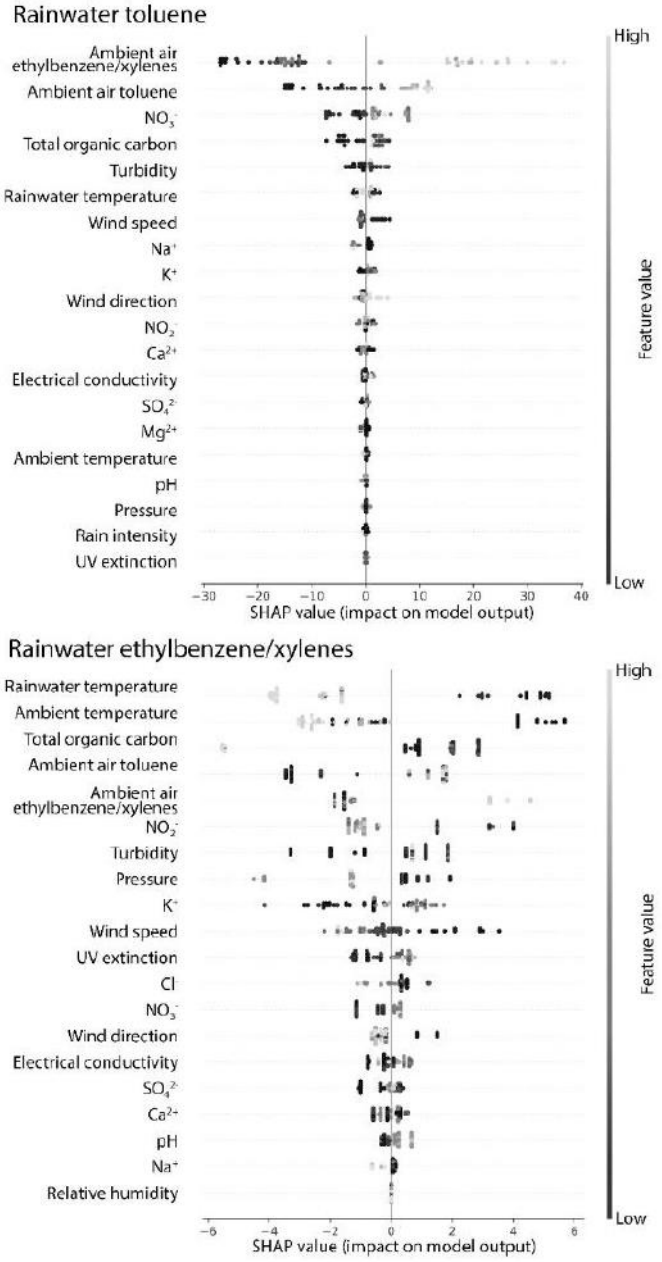


Figure 2. SHAP summary plots of toluene and ethylbenzene/xylene concentrations in rainwater.

In addition, SHAP importance score (>1) showed that xylene concentrations in rainwater are influenced by the factors in the following order: ambient and air temperature > TOC > ambient concentrations of toluene, ethylbenzene, and xylene > NO_3^- > NTU > pressure > K^+ > wind speed. The noted impacts were confirmed by SHAP summary plots and SHAP supervised clustering. As indicated by the long-tailed distribution to the right (Figure 2), high xylene ambient air concentrations (from 3.4 to 5.0 ng g^{-1}) were mainly associated with increased toluene concentrations in rainwater (up to 140 nM). Increased air and rainwater temperatures were negatively associated with xylene transfer to rainwater which is expected having in mind that the solubility of gases decreases as the temperature increases. Besides, a decline in rainwater temperature is complemented with an increased tendency of water molecules to form hydrogen bonds, which leads to lower concentrations of H^+ ions and increased pH. This process is also accompanied by lower acid-base dissociation of chemical species that originate from the gaseous oxides SO_2 and NO_X and produce SO_4^{2-}, NO_3^-, and NO_2^- ions in rainwater (Wang et al., 2015). Unsurprisingly, xylene transfer to rainwater was favored during several rainfall episodes characterized by neutral to alkaline rainwater. The rainfall events occurred under stable atmospheric conditions (ws < 6 m s^{-1}) when limited wind-induced dispersion of atmospheric constituents enabled prolonged contact between gaseous and aqueous phases. However, we noted that the opposite was observed for high wind speed (up to 30 m s^{-1}) and high air pressure which resulted in enhanced dispersion of volatile species and short-lasting contact between the gaseous phase and rainwater.

The relationships between xylene concentrations in rainwater and other parameters including the concentrations of Ca^{2+}, K^+, Cl^-, and NO_3^-, as well as NTU and TOC were arranged in non-gradually decreasing order implying that these factors can have an occasional impact on xylene phase partition (Figure 2). The influence of Na^+, K^+, and Cl^- on xylene fate in the aqueous solution can be theoretically explained by the saltingout effect. As stated by Sander (2015), the solubility of gases in a solution is decreased in the presence of electrolytes compared to solubility in pure water (Sander, 2015). For example, Allou et al. (2011) showed that Henry's

constants for formaldehyde and benzaldehyde were 27–66% and 12–21% lower, respectively, in 35 g L^{-1} NaCl salt solution, than in deionized water. Our results are mostly in accordance with this theoretical background, but in a few cases, an increase of Na^+, K^+, and Cl^- concentrations in rainwater was associated with higher xylene transfer to rainwater.

SHAP values were an order of magnitude higher for toluene than for xylenes as presented in detail in Stojić et al. (2019). Toluene and ethylbenzene/xylenes exhibit differences in solubility, volatility, and other physicochemical properties including length of alkyl chain and methyl substitutions, which can cause diversity in their distribution between environmental compartments and/or multiphase systems.

Although the significance of UV extinction in the analysis could serve as a good indication of the presence of conjugated systems including monoaromatic species such as xylene, an almost insignificant relationship between the xylene concentrations and UV extinction was recorded in our study. The results suggest that UV extinction could be associated with the presence of volatiles in rainwater in general, rather than particularly with xylenes. Similarly, an almost negligible impact of TOC, turbidity, and electrical conductivity was observed. These parameters are of importance since they indicate the presence of solid particles in rainwater that could create a three-phase system with air and water and modify the xylene partition. For example, urban aerosols, diesel soot or mineral dust, can reduce the gaseous and aqueous concentrations of ethylbenzene/xylenes by 20%, respectively (Starokozhev et al., 2011). The xylene was the most abundant in rainwater that contained the lowest amounts of particulate matter (NTU = 1–3), which theoretically should promote xylene transfer into the aqueous phase. However, the absence of direct relationships between UV extinction, NTU, EC, and xylene deposition in rainwater could be linked with the presence of ions and/or suspended particles in rainwater and the salting out effect described above. The highest values of all parameters (TOC, UV, EC, and NTU) were observed during the summer episodes of high wind speed (20–30 m s^{-1}), wind direction below 250°, and intensive rainfall whereas the lower parameter values were recorded during W, NW, and N winds.

SHAP interaction plots and zero SHAP values showed that the xylene deposition in rainwater does not depend on the mutual interactions between the physico-chemical parameters of rainwater. On the other hand, as indicated by SHAP values given in brackets, we assumed that the interrelations between the following pairs: ambient air toluene–NTU (−1.5 to 2), K^+–air temperature (−2 to 2), rainwater temperature–air pressure (−3 to 2), K^+–NO_2^- (−0.4 to 1), K^+–ambient air ethylbenzene/xylenes (−0.4 to 1), NO_2^-–NO_3^- (−0.4 to 1), rainwater temperature–TOC (−0.75 to 1), K^+–NTU (−1 to 1), K^+–ambient air toluene (−1 to 1), K^+–EC (−1 to 1), wind speed–Cl^- (−1.5 to 1), ambient air ethylbenzene/xylenes–NO_2^- (−1.5 to 0.5), wind speed–UV extinction (−0.6 to 0.8), wind speed–Ca^{2+} (−0.6 to 0.8), ambient temperature–NO_2^- (−1.5 to 0.5) and ambient temperature–NO_3^- (−0.4 to 0.4) could be of relevance for further investigations of the xylene partition between air and rainwater.

According to the SHAP analysis, TOC, UV extinction, and NTU could be considered the most important factors for describing xylene enrichment factor (EF_X) in rainwater. The findings are in accordance with those observed for EF_T and illustrate the ethylbenzene and xylene partition in the three-phase system as discussed above. Among the meteorological conditions, the dependency of xylene wet deposition on rainwater temperature and air pressure was the strongest recorded whereas air temperature and wind characteristics were found to be insignificant. As elaborated in the previous sections, interrelationships between the following pairs of variables: Cl^-–ambient air toluene, K^+–rainwater ethylbenzene/xylenes, K^+–ambient air/rainwater toluene, TOC–ambient air toluene, TOC–gaseous ethylbenzene/xylenes, and NTU–rainwater temperature could be of interest for further investigations of xylene enrichment in rainwater.

Finally, the results of SHAP analysis go far beyond those obtained by GRRF and MVA regression methods (Šoštarić et al., 2017), but they are not in collision. For instance, GRRF results and MVA regression methods when applied, with relative errors of approx. 20%, provided correlation coefficients between 0.87 and 0.95 implying that xylene enrichment is

functionally dependent on gaseous toluene concentrations and some rainwater properties rather than meteorological parameters.

CONCLUSION

High concentrations of BTEX are registered as a result of anthropogenic activities in urban areas and their scientific and practical importance is associated with the well-recognized effects on both the environment and human health. The transfer of these volatile organic species from the atmosphere to rainwater is an important process that affects the long-range pollutant transport of BTEX and their transfer to terrestrial and aquatic systems.

Since artificial intelligence and machine learning have been widely applied for the prediction of environmental processes and concentrations of pollutant species in environmental samples, it is important for researchers to gain an insight into the way the complex methodology arrives at their predictions. Several methods for interpretable approximations of complex models have been developed for this purpose. In this study, the extreme gradient boosting tree-based method was successfully applied (with relative errors lower than 20%) for predicting TEX concentrations and corresponding enrichment factors in rainwater, based on the pollutant levels in air and rainwater samples, physico-chemical properties of rainwater, and meteorological conditions. Furthermore, a novel feature attribution framework was applied to examine the relevance of the monitored parameters for the modeled predictions, and enable the insight into the main factors that govern deposition of TEX in rainwater.

As shown, BTEX concentrations observed in the aqueous phase exceeded the theoretically predicted values and ambient air TEX concentrations, with the rainwater and air temperatures being the main features that determined the spatio-temporal BTEX distribution in the environmental multiphase system. Occasional, but far less important impacts were assigned to wind speed, air pressure, turbidity, and total

organic carbon, NO_3^-, Cl^- and K^+ rainwater concentrations, whereas the impacts of other measured parameters were proved to be negligible. Given that the interfacial adsorption is assumed to be the major mechanism underlying the enhanced rain scavenging of BTEX, the removal process was observed to be more efficient for lower gas mixing ratios which might result in less accurate theoretical predictions in the area of lower gaseous concentrations. Finally, the current knowledge on TEX deposition in rainwater would benefit from a further in-depth exploration of the interactions between the examined features.

ACKNOWLEDGMENTS

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS.

REFERENCES

Alexopoulos, C. E., Chatzis, C. and Linos, A. (2006). An analysis of factors that influence personal exposure to toluene and xylene in residents of Athens, Greece. *BMC Public Health*, 6, 1-9.

Allou, L., El Maimouni, L., and Le Calvé, S. (2011). Henry's law constant measurements for formaldehyde and benzaldehyde as a function of temperature and water composition. *Atmospheric Environment*, 45, 2991-2998.

ATDSR, Public Health Service Agency for Toxic Substances and Disease Registry, Department of health and human services (2007). Xylene, Public Health Statement.

Baehr, A. L., Stackelberg, P. E., and Baker, R. J. (1999). Evaluation of the atmosphere as a source of volatile organic compounds in shallow groundwater. *Water Resources Research* 35, 127-136.

Bahadar, H., Mostafalou, S. and Abdollahi, M. (2014). Current understandings and perspectives on non-cancer health effects of benzene: a global concern. *Toxicology and Applied Pharmacology*, 276, 83-94.

Blair, G. S., Henrys, P., Leeson, A., Watkins, J., Eastoe, E., Jarvis, S. and Young, P. J. (2019). Data Science of the Natural Environment: A Research Roadmap. *Frontiers in Environmental Science*, 7, 121.

Carslaw, D. C. and Beevers, S. D. (2013). Characterising and understanding emission sources using bivariate polar plots and k-means clustering. *Environmental modelling & software*, *40*, pp.325-329.

Carslaw, D. C. and Ropkins, K. (2012). Openair—an R package for air quality data analysis. *Environmental Modelling & Software*, *27*, pp.52-61.

Chen, H., Lundberg, S. and Lee, S. I. (2019). Explaining Models by Propagating Shapley Values of Local Components. *arXiv preprint arXiv:1911.11888*.

Chilcott, R. P. (2007). Petrol toxicological overview. London, UK.

Clough, S. R. (2014). Xylene. In: *Encyclopedia of Toxicology*, *Reference Module in Biomedical Sciences,* Editor: Wexler, P., Academic Press, 989-992.

Delzer, G. C., Zogorski, J. S., Lopes, T. J., and Bosshart, R. L. (1996). Occurrence of the gasoline oxygenate MTBE and BTEX compounds in urban stormwater in the United States, 1991–95. *USGS Water-Resources Investigations Report*, 96-4145.

Draxler, R. R., Rolph, G. D. (2014). HYSPLIT (HYbrid Single-particle Lagrangian Integrated).

Ernstgård, L., Gullstrand, E., Löf, A. and Johanson, G. (2002). Are women more sensitive than men to 2-propanol and m-xylene vapours? *Occupational and Environmental Medicine*, 59, 759- 67.

Fuente, A., McPherson, B. and Cardemil, F. (2013). Xylene-Induced Auditory Dysfunction in Humans. *Ear and Hearing*, 34, 651-660.

Furutaka, S., and Ikawa, S. I. (2002). π-hydrogen bonding between water and aromatic hydrocarbons at high temperatures and pressures. *The Journal of chemical physics*, 117, 751-755.

Gibert, K., Horsburgh, J. S., Athanasiadis, I. N. and Holmes, G. (2018). Environmental data science. *Environmental Modelling & Software*, 106, 4-12.

Global Data Assimilation System (2015). https://www.ready.noaa.gov/gdas1.php. (Accessed 27 December 2015).

Goss, K. U. (2004). The air/surface adsorption equilibrium of organic compounds under ambient conditions. Critical Reviews in *Environmental Science and Technology*, 34, 339-389.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. (2009). The WEKA data mining software: an update. SIGKDD Explor 11, 10e18.

Huang, L. J., Fang, S. B., Chen, C. D. (2010). Effect of benzene, toluene, xylene occupational exposure on DNA damage of peripheral blood cells in female jewel processing workers. *Preventive Medicine*, 13, 006.

IARC, International Agency for Research on Cancer (2021). IARC monographs on the identification of carcinogenic hazards to humans. Available at: https://monographs.iarc.who.int/list-of-classifications. Last Accessed: October 2021.

Jovanović, J. M., Jovanović, M. M., Spasić, M. J and Lukić, SR. (2004). Peripheral nerve conduction study in workers exposed to a mixture of organic solvents in paint and lacquer industry. *Public Health*, 45, 769-74.

Kampf, C. J., Waxman, E. M., Slowik, J. G., Dommen, J., Pfaffenberger, L., Praplan, A. P., and Volkamer, R. (2013). Effective Henry's law partitioning and the salting constant of glyoxal in aerosols containing sulfate. *Environmental Science and Technology*, 47, 4236-4244.

Kerchich, Y. and Kerbachi, R. (2012). Measurement of BTEX (benzene, toluene, ethybenzene, and xylene) levels at urban and semirural areas

of Algiers City using passive air samplers. *Journal of the Air & Waste Management Association*, 62:12, 1370-1379.

Kuranchie, A. F., Angnunavuri, N. P., Attiogbe, F., and Nerquaye-Tetteh, E. N. (2019). Occupational exposure of benzene, toluene, ethylbenzene, and xylene (BTEX) to pump attendants in Ghana: Implications for policy guidance. *Cogent Environmental Science*, 5: 1603418.

Kurtén, T., Elm, J., Prisle, N. L., Mikkelsen, K. V., Kampf, C. J., Waxman, E. M., and Volkamer, R. (2014). Computational study of the effect of glyoxal–sulfate clustering on the Henry's law coefficient of glyoxal. *The Journal of Physical Chemistry A*, 119, 4509-4514.

Lindinger, W. and Jordan, A. (1998). Proton-transfer-reaction mass spectrometry (PTR–MS): on-line monitoring of volatile organic compounds at pptv levels. *Chemical Society Reviews*, *27*(5), pp.347-375.

Lundberg, S. M. and Lee, S. I. (2017a), December. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768-4777.

Lundberg, S. M. and Lee, S. I. (2017b). A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pp. 4765-4774.

McNeill, V. F., Grannas, A. M., Abbatt, J. P. D., Ammann, M., Ariya, P., Bartels-Rausch, T., Domine, F., Donaldson, D. J., Guzman, M. I., Heger, D., Kahan, T. F., Klán, P., Masclin, S., Toubin, C., and Voisin, D. (2012). Organics in environmental ices: sources, chemistry, and impacts. *Atmospheric Chemistry and Physics*, 12, 9653-9678.

Molnar, C. (2019). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. *URL https://christophm. github. io/interpretable-ml-book*.

Montavon, G., Samek, W. and Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, pp.1-15.

Mullaugh, K. M., Hamilton, J. M., Avery, G. B., Felix, J. D., Mead, R. N., Willey, J. D., and Kieber, R. J. (2015). Temporal and spatial variability of trace volatile organic compounds in rainwater. *Chemosphere*, 134, 203-209.

Niaz, K., Bahadar, H., Maqbool, F., and Abdollahi, M. (2015). A review of environmental and occupational exposure to xylene and its health concerns. *Experimental and Clinical Sciences Journal*, 14, 1167-1186.

Okochi, H., Kataniwa, M., Sugimoto, D., and Igawa, M. (2005). Enhanced dissolution of volatile organic compounds into urban dew water collected in Yokohama, Japan. *Atmospheric Environment,* 39, 6027-6036.

Okochi, H., Sugimoto, D., and Igawa, M. (2004). The enhanced dissolution of some chlorinated hydrocarbons and monocyclic aromatic hydrocarbons in rainwater collected in Yokohama, Japan. *Atmospheric Environment*, 38, 4403-4414.

Rajan, T. S. and Malathi, N. (2014). Health Hazards of Xylene: A Literature Review. *Journal of Clinical and Diagnostic Research*, 8, 271-274.

Reutman, S. R., LeMasters, G. K., Knecht, E. A., Shukla, R., Lockey, J.E., Burroughs, G. E. et al. (2002). Evidence of reproductive endocrine effects in women with occupational fuel and solvent exposures. *Environmental Health Perspectives*, 110, 805-11.

Roth, C. M., Goss, K. U., and Schwarzenbach, R. P. (2004). Sorption of diverse organic vapors to snow. *Environmental science and technology*, 38, 4078-4084.

Sander, R. (2015). Compilation of Henry's law constants (version 4.0) for water as solvent. *Atmospheric Chemistry and Physics*, *15*(8), pp.4399-4981.

Sato, E., Matsumoto, K., Okochi, H., and Igawa, M. (2006). Scavenging effect of precipitation on volatile organic compounds in ambient atmosphere. *Bulletin of the Chemical Society of Japan*, 79, 1231-1233.

Singh, A. K., Tomer, N. and Jain, C. L. (2012). Monitoring, Assessment and Status of Benzene, Toluene and Xylene Pollution in the Urban

Atmosphere of Delhi, India. *Research Journal of Chemical Sciences*, 2, 45-49.

Sirotkin, V. A. (2019). Reproductive Effects of Oil-Related Environmental Pollutants. *Encyclopedia of Environmental Health. Reference Module in Earth Systems and Environmental Sciences*, 493-498.

Słomińska, M., Konieczka, P., and Namieśnik, J. (2014). The fate of BTEX compounds in ambient air. Critical Reviews in *Environmental Science and Technology*, 44, 455-472.

Soltanpour, Z., Mohammadian, Y., Fakhri, Y. (2021). The concentration of benzene, toluene, ethylbenzene, and xylene in ambient air of the gas stations in Iran: A systematic review and probabilistic health risk assessment. *Toxicology and Industrial Health*, 37, 134-141.

Šoštarić, A., Stojić, A., Stojić, S. S., and Gržetić, I. (2016). Quantification and mechanisms of BTEX distribution between aqueous and gaseous phase in a dynamic system. *Chemosphere*, 144, 721-727.

Šoštarić, A., Stojić, S. S., Vuković, G., Mijić, Z., Stojić, A., and Gržetić, I. (2017). Rainwater capacities for BTEX scavenging from ambient air. *Atmospheric Environment*, 168, 46-54.

Starokozhev, E., Fries, E., Cycura, A. and Püttmann, W. (2009). Distribution of VOCs between air and snow at the Jungfraujoch high alpine research station, Switzerland, during CLACE 5 (winter 2006). *Atmospheric Chemistry and Physics*, 9, 3197-3207.

Starokozhev, E., Sieg, K., Fries, E., and Püttmann, W. (2011). Investigation of partitioning mechanism for volatile organic compounds in a multiphase system. *Chemosphere*, 82, 1482-1488.

Stojić, A. and Stojić, S. S. (2017). The innovative concept of three-dimensional hybrid receptor modeling. *Atmospheric Environment*, *164*, pp.216-223.

Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A. and Lazić, L. (2019). Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene, and xylene wet deposition. *Science of the Total Environment,* 653, 140-147.

Stojić, A., Vuković, G., Perišić, M., Stanišić, S., and Šoštarić, A. (2018). Urban air pollution: an insight into its complex aspects. In: *A Closer*

Look at Urban Areas, Editor: Sahar Romero, Nova Science Publishers, NY, USA, 69-123.

Su, F., Lu, C., Johnston, K. R., Hu, S. (2010). Kinetics, Thermodynamics, and Regeneration of BTEX Adsorption in Aqueous Solutions via NaOCl-Oxidized Carbon Nanotubes. In: *Environanotechnology,* Editors: Fan, M., Huang, C-P., Bland, A. E., Wang, Z., Slimane, R. and Wright, I., Elsevier, 71-97.

Tang, D. C. and Xu, L. (2005). Study on lipid peroxidation in workers exposed to benzene, toluene and xylene. *Chinese Occupational Medicine,* 2, 016.

Team (2012). R: a language and environment for statistical computing. http://cran.case.edu/web/packages/dplR/vignettes/timeseries-dplR. pdf (Accessed 4 December 2015).

UN, United Nations (2018). *The Sustainable Development Goals Report*. Available at: https://unstats.un.org/sdgs/files/report/2018/The SustainableDevelopmentGoalsReport2018-EN.pdf. Last Accessed: October 2021.

US EPA, United States Environmental Protection Agency (2003). *Toxicological review of xylenes*. Available at: https://iris.epa. gov/static/pdfs/0270tr.pdf. Last Accessed: October 2021.

US EPA, United States Environmental Protection Agency (2007). *EPA Unmix 6.0 Fundamentals and User Guide.* USEPA Office of Research and Development.

Vaidyanathan, A., Foy, J. W. D. and Schatz, R. (2003). Inhibition of rat respiratory-tract cytochrome P-450 isozymes following inhalation of m-xylene: possible role of metabolites. *Journal of Toxicology and Environmental Health*, 66, 1133-43.

Valsaraj, K. T., Thoma, G. J., Reible, D. D., and Thibodeaux, L. J. (1993). On the enrichment of hydrophobic organic compounds in fog droplets. *Atmospheric Environment. Part A. General Topics*, 27, 203-210.

Varona-Torres, E., Carlton Jr., D. D., Payne, B., Hildenbran, L. Z., and Schug K. A. (2017). The Characterization of BTEX in Variable Soil Compositions Near Unconventional Oil and Gas Development. In:

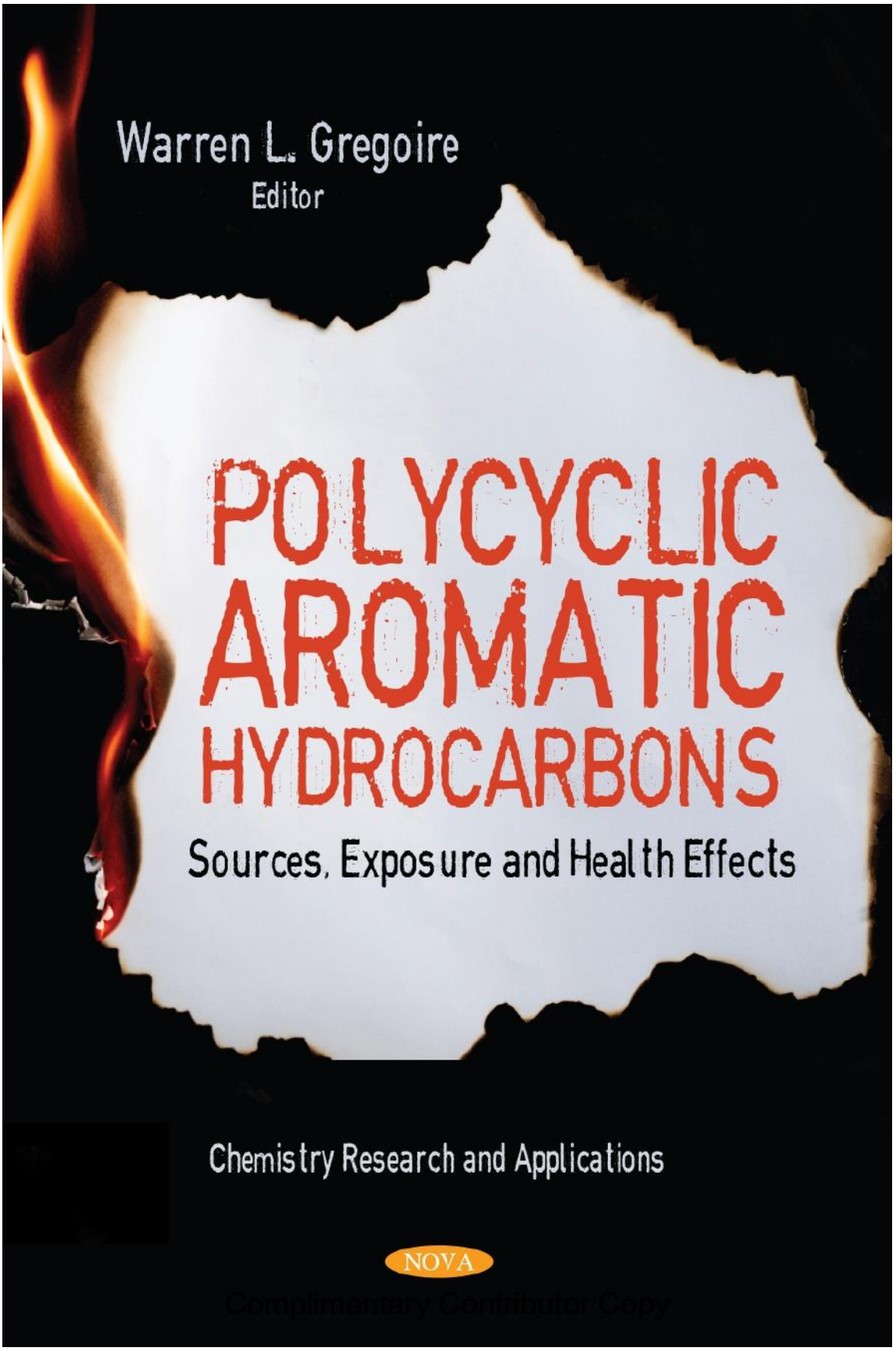
Advances in Chemical Pollution, Environmental Management and Protection, Editors: Hildenbran, L. Z., and Schug K. A., 1, 321-351.

Wang, L., Wen, L., Xu, C., Chen, J., Wang, X., Yang, L., and Zhang, Q. (2015). HONO and its potential source particulate nitrite at an urban site in North China during the cold season. *Science of the Total Environment*, 538, 93-101.

Warneke, C., De Gouw, J. A., Kuster, W. C., Goldan, P. D. and Fall, R. (2003). Validation of atmospheric VOC measurements by proton-transfer-reaction mass spectrometry using a gas-chromatographic preseparation method. *Environmental science & technology*, *37*(11), pp.2494-2501.

WHO, World Health Organisation (2003). *Xylenes in Drinking-water Background document for development of WHO Guidelines for Drinking-water Quality*. Available at: https://www.who.int/water_sanitation_health/dwq/chemicals/xylenes.pdf. Last Accessed: October 2021.

Xue, B. X., Wei, L., Li, C. Y., and Li, T. Y. (2015). A study of the distribution and composition of pollutants in snow collected from streets and a treatment system for recycling snow in winter cities. *Desalination and Water Treatment*, 54.


Warren L. Gregoire
Editor
POLYCYCLIC AROMATIC HYDROCARBONS
Sources, Exposure and Health Effects
Chemistry Research and Applications
NOVA
Complimentary Contributor Copy

Chemistry Research and Applications

Chemistry Research and Applications

A Closer Look at Carvacrol
Zak A. Cunningham (Editor)
2022. ISBN: 978-1-68507-627-6 (Softcover)
2022. ISBN: 978-1-68507-634-4 (eBook)

Fundamentals of Photocatalysis
Orva Auger (Editor)
2021. ISBN: 978-1-68507-374-9 (Softcover)
2021. ISBN: 978-1-68507-417-3 (eBook)

Polypropylene: Advances in Research and Applications
Théodore Marleau (Editor)
2021. ISBN: 978-1-68507-378-7 (Hardcover)
2021. ISBN: 978-1-68507-401-2 (eBook)

Boron: Advances in Research and Applications
Lynn Mcconnell (Editor)
2021. ISBN: 978-1-68507-231-5 (Hardcover)
2021. ISBN: 978-1-68507-259-9 (eBook)

Applications of Layered Double Hydroxides
Rajib Lochan Goswamee, PhD (Editor)
Pinky Saikia, PhD (Editor)
2021. ISBN: 978-1-68507-355-8 (Hardcover)
2021. ISBN: 978-1-68507-381-7 (eBook)

More information about this series can be found at
https://novapublishers.com/product-category/series/chemistry-research-and-applications/

Warren L. Gregoire
Editor

Polycyclic Aromatic Hydrocarbons

Sources, Exposure and Health Effects

nova
science publishers
New York

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-68507-685-6 (eBook)

Published by Nova Science Publishers, Inc. † New York

Contents

Chapter 1

Explaining the Environmental Fate of PAHs in Indoor and Outdoor Environments by the Use of Artificial Intelligence

Svetlana Stanišić[1], Gordana Jovanović[1,2], Mirjana Perišić[1,2], Snježana Herceg Romanić[3], Tijana Milićević[2] and Andreja Stojić[1,2,*]

[1]Singidunum University, Belgrade, Serbia
[2]Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia
[3]Institute for Medical Research and Occupational Health, Zagreb, Croatia

Abstract

The environmental fate of polycyclic aromatic hydrocarbons (PAH) is complex, as they are emitted from natural and anthropogenic sources into the atmosphere and further distributed to soil and, to a lesser extent, ground and surface water. People nowadays spend more than 80% of their time indoors in developed countries and 85-90% in Europe, which results in elevated human exposure to poor air quality in homes, working environments, public buildings, or the means of transportation.

We are witnessing the emerging need for the application of machine learning (ML) and explainable artificial intelligence (XAI) for the capturing of the unique and defining factors and processes responsible for the complexity, non-linearity, interactivity, or cross-compartment interconnectivity of environmental phenomena. The investigations of

* Corresponding Author's Email: andreja.stojic@ipb.ac.rs.

In: Polycyclic Aromatic Hydrocarbons
Editor: Warren L. Gregoire
ISBN: 978-1-68507-626-9

mutual interdependencies and relations among organic pollutants and environmental factors that govern their distribution in various matrices are supported by the large availability of high-dimensional data and successfully justified in environmental science.

In this chapter, we have used ML and XAI to explain the behavior and environmental fate of PAH compounds, as constituents of $PM_{2.5}$ measured in indoor and outdoor environments of a university building. We have investigated complex non-linear interactions between outdoor and indoor PAH levels, and inorganic gaseous pollutants, trace elements, ions, radon, 31 meteorological parameters, the number of people in the amphitheater, and the time they spent indoors. To illustrate the potential of the presented methodology, we have chosen to characterize environmental conditions that shape the occurrence of benzo[b]fluoranthene in both environments and determine the importance and degree of impact that certain environmental factors exhibit on its levels.

Keywords: PAHs, benzo[b]fluoranthene, air quality, PM-bound pollutants, artificial intelligence

The Environmental Fate of PAH Compounds

The Origin and Sources of PAHs

PAHs refer to a group of several hundred aromatic hydrocarbons with two to seven fused benzene rings bonded in either a linear, angular or clustered way. In general, PAHs are generated in the thermal decomposition of organic substances including high-temperature processes of incomplete combustion and/or long-term exposure of organic material to low temperatures up to 300°C. The lower the combustion temperatures are, the incomplete combustion process will release more PAHs. Additionally, under low temperatures, the generation of alkyl homologues is favored while the higher combustion temperatures are more associated with simple and condensed chemical structures. The group of low molecular weight compounds consists of two or three aromatic rings, while the PAH species with four or more benzene rings are assigned as high molecular weight PAHs (Zhao et al., 2020). According to the structure, PAHs can be divided into alternant compounds which are more planar and symmetrical. The compounds are derived by fusion of additional six-membered benzene rings and non-alternant species in which benzene rings can be connected by five-numbered structures, which are

typically emitted from low-temperature combustion sources (Wick et al., 2011).

At room temperature, PAHs are colored, crystalline solids that have low solubility in water, high boiling points, and low vapour pressure. These properties are pronounced with the increase in molecular weight. Some features that can be associated with PAHs are persistence, fluorescence, light sensitivity, heat, and corrosion resistance and conductivity. Additionally, each PAH ring structure possesses unique UV absorbance spectra and this is used for the PAH identification.

These air pollutants are emitted from both natural and anthropogenic sources. Natural sources encompass forest and vegetation fires, oil seeps, volcanic eruptions, diagenetic processes, and exudates from trees. Some PAH species are produced from pigments of fungi, insects, and marine species in anaerobic conditions that can be found in soil or subaquatic sediments (Paulik et al., 2018). While the natural PAH emissions represent a significant source of these contaminants to the environment, anthropogenic sources remain dominant in urban and industrialized areas. Nowadays the anthropogenic fossil fuel combustion is the major source of these compounds, including traffic emissions, coal-gasification sites, and smokehouses, while other sources of PAHs include aluminum production plants, burning of wood, garbage, tobacco, plant material and refuse, use of lubricating oil and oil filters, municipal solid waste incineration, petroleum spills and discharge, coke production, as well as activities that include the use of coal tar, asphalt, creosote (wood preservation) and roofing tar. In urban locations where the traffic speed is low and gear changing is frequent, higher PAH emissions are registered, followed by higher PAH levels in the soil next to the road (Teixeira et al., 2015). Thereby, diesel vehicle emissions are associated with lighter molecular weight PAHs, while gasoline emissions tend to contain more heavy molecular weight PAH. The study of Slezakova et al., (2013) has shown that diesel vehicular emissions can represent the major source of PAHs in urban areas. Once emitted, they are distributed to the atmosphere and deposited on terrestrial and water surfaces.

Taking into consideration the potential for toxicity and carcinogenicity, prevalence in hazardous waste, and persistence in environmental conditions, the U.S. EPA has listed 16 PAH compounds on the Priority Pollutant List created under the Clean Water Act. These PAH species were prioritized not only due to their toxicity but also due to their frequency of occurrence and potential for human exposure. For the purpose of certain location risk assessment, concentrations of these species, often referred to as the “priority

PAHs," are generally monitored: naphthalene (NAP), acenaphthylene (ACY), acenaphthene (ACE), fluorene (FLU), phenanthrene (PHEN), anthracene (ANTH), fluoranthene (FLTH), pyrene (PYR), benzo[a]anthracene (B[a]A), chrysene (CHRY), benzo[b]fluoranthene (B[b]F), benzo[k]fluoranthene (B[k]F), benzo[a]pyrene (B[a]P), benzo[g,h,i]perylene (B[ghi]P), indeno[1,2,3-c,d]pyrene (IND), and dibenz[a,h]anthracene (D[ah]A) (Samburova et al., 2017). Generally, phenanthrene, fluoranthene, and pyrene are dominating the urban air.

Human Exposure to PAH Compounds and Health Effects

As widely distributed environmental contaminants, PAHs were among the first atmospheric pollutants that were confirmed to be toxic, mutagenic, and carcinogenic species with detrimental effects on human health and living organisms. Due to their ubiquitous occurrence in air, water, terrestrial, and biological systems and physicochemical properties which support their distribution to all environmental compartments, resistance to biodegradation, a tendency for bioaccumulation, and carcinogenic potential, PAHs have gathered significant environmental concern. Due to their lipophilic nature, these compounds readily penetrate the skin, as well as cellular membranes. Among PAH contaminants, benzo(a)pyrene is considered the most carcinogenic and toxic, followed by benzo(a)anthracene and dibenz(ah) anthracene.

Other health-relevant possible/probable carcinogenic PAHs (IARC, 2012) include chrysene (Chy), benzo[b]fluoranthene (B[b]F), benzo[k]fluoranthene (B[k]F), and indeno[1,2,3-cd]pyrene (I[cd]P). Generally, with the increase in molecular weight, PAH carcinogenicity is more pronounced, while acute toxicity is reduced. Therefore, low molecular weight PAHs are considered directly toxic, while heavy molecular weight PAH compounds are considered genotoxic or capable of causing damage to DNA (Kim et al., 2013). Although the lighter-weight compounds are in some way less hazardous, their reactions with ozone and gaseous oxides result in the formation of highly toxic substances such as diones, nitro- and dinitro-PAHs, and sulfuric acids. On the other hand, the carcinogenic potential of certain heavy-weight PAHs depends on the compound and its metabolic route. Namely, as a result of PAH metabolic transformations in the human body, a complex mixture of quinones, quinines, cis- and trans-dihydrodiols, phenols, epoxides, and other oxidized metabolites are generated and these metabolites are able to covalently bind to

nucleic acids and induce strand breaks and DNA damage, which results in genetic mutation (Bansal and Kim, 2015).

Human exposure to PAHs is related to cigarette smoking, inhalation of polluted air, dermal exposure from occupational or non-occupational settings, eating agricultural products grown in the PAH contaminated soil, and even more by intake of food which has been grilled, smoked, or roasted.

Nevertheless, assessing human health risk can be very complex due to the carcinogenic or detrimental potential of PAHs in combination. Therefore, this procedure is simplified by the introduction of the toxicity equivalency factor (TEF). TEFs are assigned to different species based on their relative toxicity compared to known human carcinogen B[a]P, chosen to be a reference compound, and then the concentrations of carcinogenic PAHs are converted to an equivalent concentration of B[a]P mixtures (Hussar et al., 2012).

The extent of exposure, the contaminant type, and its concentration are the main predictors of adverse health effects. The factors that should be also considered are related to the route of exposure and pre-existing health conditions. The majority of studies have been investigated human exposure to a mixture of PAH compounds, and moreover to the mixture of air pollutants including non-PAH potentially carcinogenic contaminants, while the studies on animals have considered the effects of the exposure to higher levels of individual contaminant. The latter have shown that the exposure to benzo(a)anthracene, benzo(a)pyrene, and naphthalene has been linked to embryotic effects in experimental animals, while the ingestion of benzo(a)pyrene during pregnancy can lead to low birth weight and birth defects (Vignet et al., 2014).

As regards acute health effects in the human population, it has been recognized that PAHs can provoke an exacerbation of asthma and thrombosis in people with atherosclerotic formations. Furthermore, occupational PAH exposure during coke production, bituminous product use during roofing or oil refining can lead to eye irritation and gastrointestinal symptoms of intoxication. Additionally, anthracene and benzo(a)pyrene are direct skin irritants and sensitizers, responsible for skin irritations and allergies in people who are exposed to their high concentrations (Friesen et al., 2010). As regards chronic health effects, it has been already mentioned that long-term exposure to some PAH species is associated with the binding of electrophilic PAH metabolites to DNA, gene mutation, DNA damage, and consequentially higher incidence of skin, bladder, respiratory, and gastrointestinal cancer. Besides, certain PAHs have the potential to interfere with hormones, which increases the chances of hormonal disruption and immune and reproductive system

failure. In compliance with the findings of animal studies, it has been evidenced that exposure to PAH pollution during pregnancy might be related to adverse birth outcomes, low birth weight, premature delivery, low IQ, behavioral problems, childhood asthma, and delayed child development (Kim et al., 2013). Finally, it might be worth considering some indirect adverse health effects of PAH pollution. Namely, the study of Parajuli et al., (2017) has concluded that pollution-induced shifts in natural soil bacterial population which is registered in PAH-polluted areas can contribute to the prevalence of chronic diseases in the local population. These findings have been explained by the fact that he human microbiome, which has been shown to play a significant role in the immune system, is affected by the environmental microbiome.

In a number of studies, the PAH exposure assessment was conducted by the determination of certain pyrene metabolites in the urine. The level of this exposure biomarker in the urine can be two times higher in smoking compared to the non-smoking population.

Dietary Exposure to PAH Compounds

The current scientific evidence and knowledge suggest that dietary intake of PAHs is the major route of human exposure for the non-smoking population in particular. Besides, over the last three decades, human exposure to PAHs has been increased not only in developing countries where pollutant emissions are on the rise but also in developed countries due to their wide spread in food chain, modern lifestyle, and heavy reliance on fast food. In compliance with the aforementioned, dietary intake of PAHs is suggested to be one of the major factors contributing to skin and lung cancer, but also metabolic disorders and non-genotoxic diseases including diabetes mellitus and cardiovascular disorders.

The common nutrition sources of PAH refer to a wide variety of foods, including raw fruits and vegetables, smoked and grilled meat, refined fats and oils, fatty fish and seafood, etc. (Zelinkova and Wenzl, 2015).

The food that contains the highest PAH concentrations is the one processed by grilling, roasting, smoking, and frying. The amount of PAH in food depends on the heat source, the distance of heating, the design of the food device, and the type of fuel being used. Previous research has shown that PAH formation in the open-flame grilled meat was reduced when pre-heating with steam and microwave or wrapping with aluminum and banana leaf were

applied, with the latter being more effective. The removal of smoke during meat grilling can lead to a 74% reduction in PAH content (Lee et al., 2016). In addition to this, PAH formation is prevented by the consumption of lean meat and fish, by the avoidance of open flame and direct food contact of meat and flame during barbequing, by cooking at lower temperatures, by using electric or gas meat broilers over charcoal, and by using the acid-base marinade as lemon juice before grilling of meat.

The process of food smoking results in PAH emissions during the incomplete combustion of wood. The use of poplar and hickory can result in a decrease of up to 55% in the PAH food contents compared to the commonly used beech wood (Hitzel et al., 2013). In developed countries, the majority of food products are treated with liquid smokes, and this alternative to the smoking process doesn't cause air pollution and allows for better control over PAH concentrations in the final product (Varlet et al., 2010). The possible approach to PAH reduction in smoked food products includes the immediate use of low-density polyethylene (LDPE) layer over smoked meat. This procedure has been proved to reduce the initial PAH level in food by more than 50% because the surface of the smoked/grilled foods contains most of the PAHs which can diffuse to packaging film of similar polarity immediately after wrapping. Moreover, polytetrafluoroethylene (PET) has been also shown to be useful for PAH concentration reduction in refined seed oil (Bansal and Kim, 2015).

Fruits and vegetables which are grown in the vicinity of industrial sources or areas with dense traffic contain higher amounts of PAH, but rarely more than 5 $\mu g\ kg^{-1}$. Thereby, it has been evidenced that leafy vegetables are more contaminated than stem vegetables such as cucumber, eggplant, and tomato, due to their larger, waxy, and cuticle surface area, which is vulnerable to the PAH deposition. Similarly, the root vegetables are more prone to PAH uptake from contaminated soil. Trace levels of phenanthrene, fluoranthene, and pyrene have been found in every raw fruit and vegetable (Paris et al., 2018). The concentrations of PAHs are lower in raw food grown in the areas not affected by volcanoes, forest fires, traffic, and industrial emissions. Additionally, washing fruits and vegetables with oxidizing agents can be useful for lowering the PAH content.

Seafood also contains certain PAH amounts, depending on the lifetime accumulation and water contamination, disposal of industrial effluents near coastal waters, proximity of oils spills, and use of creosote-treated wood for mussel cultivation (Gohlke et al., 2011; Rotkin-Ellman et al., 2012). Zhao et al., (2014) have shown that PAH concentrations can reach up to 513 $\mu g\ kg^{-1}$

in different tissues of bighead carp and silver carp. Fats and oils are also significant dietary sources of PAHs (Hao, Li and Yao, 2016). Furthermore, chocolate sweets contain PAH due to the drying, roasting, winnowing, blending, and fermenting of the cocoa seeds (Lowor et al., 2012).

Soil as the Ultimate Sink of PAHs

Although they have been investigated as air pollutants, for PAHs soil represents the ultimate sink. Levels of PAH compounds in a particular soil can typically range from 1 to 10 μg kg^{-1} depending on the proximity of the emission sources. Nevertheless, the total concentrations of 16 US EPA priority PAHs (ΣPAHs) can reach much higher levels. As shown in the study of Zhu et al., (2019), the average ΣPAHs concentrations were 274 ng m^{-3} in the air, 255 μg kg^{-1} in the soil, and 15 μg kg^{-1} in vegetation at a measurement site located in the economic and industrial center The Yangtze River Delta.

Most PAHs are deposited from the atmosphere in the soil and leached from overlying horizons to greater depth and lower aggregate surfaces. Due to their low solubility in water, they are found in low concentrations in soil water and more often are bound to soil particles that are dispersed in soil solution (Wilcke, 2000).

In the environment, PAHs undergo adsorption on soil particles, absorption by plant roots or animal ingestion, volatilization from soil, plant or water surfaces, leaching or translocation laterally or downward through the soil, photo-oxidation, chemical and microbial degradation, alteration by chemical processes such as oxidation-reduction reactions, adsorption through interaction with soil and sediments and diffusion into soil micropores where they become unavailable for microbial degradation (Johnsen, Wick and Harms, 2005). The processes of transfer, degradation, and sequestration occur through a variety of mechanisms and are highly dependent on PAH molecular weight, structure, water solubility, and vapour pressure. For instance, their persistence increases with molecular weight (Yukhimets et al., 2019).

The levels of PAHs are highest at the specifically contaminated sites, such as gasworks sites due to waste materials, coke ovens, petroleum refineries, and wood conservation plants, followed by urban soils, where the contribution of industrial and traffic emissions and emissions from fossil fuel burning for heating operations are significant. Lower PAH concentrations were registered in permanent grassland, forest, and arable soils. The PAH mobility that occurs in the soil doesn't correspond with molecular weight as it would be expected

due to the fact that high molecular weight PAH species are less water-soluble. Sorption and desorption are considered to be the main processes for PAH transport in the soil, which suggests that high molecular weight PAHs are mostly transported as particle-adsorbed. The significant share of organic pollutants that remain in prolonged contact with soil is adsorbed on soil particles in the process of contaminant sequestration which makes them also less immobilized and less available for biodegradation (Haritash and Kaushik, 2009). Thus, the transport of PAHs in the soil is described by a model which includes the solid, the dissolved and the particle-bound phase.

In the organic horizons of forests and urban soils, individual PAH concentrations can reach several 100 $\mu g\ kg^{-1}$. The presence of different compounds depends on the climate, but in temperate soils, the most dominant species are benzofluoranthenes, chrysene, and fluoranthene (Wilcke, 2000). Furthermore, meteorological features exhibit a strong impact on PAH dynamic patterns in soil. For instance, moisture and temperature affect PAH decomposition and volatilization.

PAH levels are assumed to be 10 times higher than the concentrations that were present prior to industrialization, which suggests the share of anthropogenic contribution over the last century. In most remote areas like the Arctic, benzo(a)pyrene levels are in the range of those from the preindustrial era.

Nowadays, various techniques have been developed to treat PAH-contaminated soil, including soil washing, chemical oxidation, electrokinetic, and phytoremediation. The study of Gitipour et al., (2018) reported that the surfactant-aided washing process had a 90% efficiency, while compost-amended phytoremediation removed from 58 to 99% of pyrene from the soil over the 90-day period. Additionally, the most efficient treatment procedure was shown to be chemical oxidation, while the electrokinetic treatment has been proved successful in removing specific PAH contaminants.

The major PAH degradation natural process is related to microbial degradation generally catalyzed by enzymatic systems of microorganisms. Microbial degradation refers to PAH transformation into less complex and less hazardous/non-hazardous compounds and depends on weather conditions which can be aerobic and anaerobic. The final products of the degradation process are inorganic minerals, water, carbon dioxide, and methane. In addition to the fact that microbial degradation is a natural process, contaminated locations can be remediated by microbial manipulations and for this purpose, it is very important to understand the environmental fate of specific PAH compounds (Zhang et al., 2006). The rate of biodegradation

depends on: 1) the environmental conditions, including pH, temperature, nutrients, metals, moisture, and oxygen presence, 2) microbial species such as algae, bacteria, and fungi and related factors including their population, degree of acclimation, accessibility of nutrients, cellular transport properties, and chemical partitioning in the growth medium, and 3) nature and chemical structure of the PAH compound being degraded. For instance, previous research has reported anaerobic degradation of two- and three-ring PAHs, but it has not been shown for PAHs containing more than three rings. Furthermore, higher molecular weight PAHs have been shown to be more resistant to biodegradation compared to low molecular weight PAHs. The studies have been performed *in situ* (contaminated sites), *ex situ* (bioreactors), or in laboratory settings with soil samples being PAH spiked. As shown, some microorganisms are capable of using PAHs as a source of carbon and energy, transforming the contaminants into nontoxic products (Ghosal et al., 2016).

Some "dead-end products" of many biological and chemical degradation pathways such as oxygenated PAHs are also known for their toxicity and carcinogenicity, which can be particularly important in cases where microflora is used for bioremediation purposes. Namely, it has been reported that badly designed bioremediation soil treatments resulted in the microbial formation of new, more toxic, more persistent, and more water-soluble and thus, mobile co-contaminants than those initially detected (Lundstedt et al., 2007). Finally, the bioremediation process might be limited by the supply of nutrients for bacterial population, non-optimal conditions related to temperature, pH, oxygen presence, and salt content, as well as lack of bacterial species that can degrade PAH compounds or low PAH availability due to their physicochemical features (Wick et al., 2011). Furthermore, the rate of PAH degradation in the soil can be also decreased due to contaminant adsorption which is dependent on soil features including cation exchange capacity, micropore volume, soil texture, and surface area. The biodegradation PAH rate and remediation of soil can be enhanced by increasing PAH bioavailability by plant establishment, increasing metabolic potential of the bacterial population through the addition of specific bacterial strains, or supplementation of contaminated sites with light oil, straw, fertilizer, manure, and compost material which serve to improve soil texture, oxygen transfer, and provide nutrients for the bacterial population. In general, without supplementation, only three-ring aromatics can be degraded.

The uptake of PAHs by plants is not significant because the plants are unable to transport hydrophobic substances and most of these species detected in plant tissue originate from atmospheric deposition, although the uptake by

above-ground parts of the plant can also contribute. Vegetation in rural regions can contain from 50 to 80 μg kg^{-1} of PAHs, while the urban vegetation can have up to 10 times more PAH levels. Nevertheless, living beings in the soil accumulate considerable amounts of PAHs in a short period either by soil ingestion directly or by plant ingestion indirectly.

PAHs in the Atmosphere

The levels of PAHs in the atmosphere are dependent on season, meteorological conditions, time of the day, measurement site, but also on certain factors that in addition to weather conditions have an impact on atmospheric chemistry, dry or wet deposition, and finally, PAH half-lives and their interactions with other pollutants including ozone, SO_2, SO_3, NO_x, or OH radicals (Lee, 2003). In the cold season, PAH concentrations are higher due to increased fossil fuel burning for heating, reduced thermal and photo-decomposition, and less atmospheric mixing associated with the lower planetary boundary layer.

In the atmosphere, PAHs are found in the vapor phase or adsorbed in particulate matter. The partitioning of PAHs between particles and gas is mainly dependent on the atmospheric conditions and nature of the contaminants and aerosol including PM, soot, dust, fly ash, pyrogenic metal oxides, pollens, etc. At ambient temperatures, most atmospheric PAHs are particle-bound. The atmospheric transformations and degradations of PAHs in the atmosphere are reduced or inhibited when they are adsorbed to particles. Thereby, the lower molecular weight contaminants are more volatile and found in the gas phase, while particle-bound species are mostly those of higher molecular weight. In the cold season, particulate phase-bound PAHs are dominant, while in the warm season, gas-phase PAHs are often registered (Vuong et al., 2020). The sorption to particulate matter which increases with a decline in PAH volatility enables long-range transport of these species. As well as for the particulate matter, atmospheric precipitations appear to have a particularly scavenging efficiency for atmospheric PAHs (Gaga and Ari, 2019).

According to the report of NAEI (Brown et al., 2012), out of 621 tons of 16 priority PAHs being emitted from anthropogenic sources, 3.23 tons referred to benzo(a)pyrene, while the contribution of natural sources to total benzo(a)pyrene was negligibly lower (2.88 tons). Thereby, the absolutely highest contribution of all anthropogenic sources was attributed to fossil fuel

combustion for heating purposes. While the PAH levels have declined over the previous decades due to the prevention measures and environmental protection policies, the air quality in developing countries where energy production relies on biomass and coal combustion remains low (Zhu et al., 2019). In areas where industrial emissions represent the major PAH sources, registered contaminant levels exhibit no seasonal variations. However, in the residential areas which are dominated by coal and wood burning for heating, higher PAH concentrations are registered in winter and the seasonal concentration dynamic is pronounced (Gune et al., 2019). Among the factors that have an impact on outdoor PAH concentrations, the study of Slezakova et al., (2013) has emphasized the significance of other gaseous air pollutants and solar radiation, while ozone, temperature, relative humidity, and wind speed were recognized to be negatively correlated with PAH concentrations. The study of Elorduy et al., (2016) emphasized that out of meteorological parameters, wind speed, temperature, and atmospheric pressure were recognized as significant for governing distribution of PAHs in the air whereas the influence of relative humidity appeared to be negligible.

The $PM_{2.5}$-bound PAH Behavior in Indoor and Outdoor Environments: the Explainable Prediction of Benzo(b)fluoranthene Levels

People nowadays spend more than 80% of their time indoors in developed countries and 85-90% in Europe, which results in elevated human exposure to poor air quality in homes, working environment, public buildings, or the means of transportation (Ferguson et al., 2020; González-Martín et al., 2020). However, literature sources on the occurrence and behavior of air pollutants in indoor environments are still scarce, although the levels of indoor PAH compounds can exceed their concentrations in the outdoor environment. For instance, the study of Cao et al., (2019) aimed at investigating spatial variations and corresponding health risks of PAHs included composite settled dust samples collected from four types of microenvironments including offices, hotels, dormitories, and kindergartens in Beijing. As shown, the total concentrations of explored PAH species ranged from 388 μg kg^{-1} (kindergarten dust) to 8140 μg kg^{-1} (hotel dust). Indoor air quality clearly depends on outdoor contaminants intrusion and on the activity of endogenous sources such as emissions from building materials, cleaning and disinfection

products, cooking and heating, and an individual's activities (González-Martín et al., 2020).

The previous literature findings regarding benzo(b)fluoranthene (B[b]F) indoor levels (González-Martín et al., 2020; Oliveira et al., 2019) reported highly variable data which can be attributed to the fact that different sampling periods and measurement sites were chosen, as well as to the fact that different methodologies were applied. Additionally, the features of a certain measurement location including surrounding emission sources, topography, and meteorological conditions, make a regional and global comparison of data less feasible. According to the study of Pereira et al. (2019), B[b]F was was among the most abundant PAH in outdoor (0.04 ng m^{-3}) and indoor (0.06 ng m^{-3}) school environments and based on the B[b]F/B[k]F diagnostic ratio, its' origin was recognized to be associated with light vehicular diesel emissions. Unlike in Europe or North America, PAH imposes a significant burden in the school environment in Asia (Oliveira et al., 2019), where average indoor and outdoor $PM_{2.5}$-bound B[b]F levels were 25.3 and 27.1 ng m^{-3}, and 0.95 and 1.08 ng m^{-3}, during the cold and non-heating season, respectively (Zhang et al., 2020). Extremely high levels of PM-bound B[b]F reaching 174.7 ng m^{-3} were found in indoor water pipe cafés of Ardabil city, Iran (Rostami et al., 2021).

The emerging need for the application of (explainable) artificial intelligence (AI) and machine learning (ML) methods is supported by the large availability of high-dimensional data (Blair et al., 2019; Gilbert et al., 2018) and successfully justified in environmental science for the investigations of mutual interdependencies and relations among organic pollutants and environmental factors that govern their distribution in various matrices (Jovanović et al., 2019; Stojić et al., 2019; Stanišić et al., 2021). In this study, we used eXtreme Gradient Boosting (XGBoost) to study the behavior of B[b]F in the indoor and outdoor environments of a university building. We also employed SHapley Additive exPlanations (SHAP), an additive feature attribution method that provides a *post-hoc* explanation of the ML models' overall behavior in the form of feature contributions, which is more aligned with human intuition (Chen et al., 2021; Stojić et al., 2019). The methods were applied to: 1) explain complex non-linear interactions between outdoor and indoor B[b]F levels, and inorganic gaseous pollutants, trace elements, ions, radon, 31 meteorological parameters, the number of people in the amphitheater, and the time they spent indoor, 2) characterize environmental conditions that shape B[b]F occurrence in both environments, and 3)

determine the importance and degree of impact that certain environmental factor exhibit on B[b]F levels.

Methodology

Measurement Campaign

During the three-month study campaign (March 1st – May 31st), the sampling of organic and inorganic criteria air pollutants was performed in indoor and outdoor measurement sites of the building of Singidunum University (44°45'33.8"N, 20°29'47.6"E), situated in the urban area of Belgrade, the capital of Serbia. The University building is placed within large residential areas, which contain households with individual fireboxes using coal and wood; heating plants and fuel oil heating plant using natural gas and crude oil around 800 m to the W and SW; several small-scale chemical plants and food factories located up to 600 m in the NW and S direction; a boulevard with public transport and moderate vehicle flow approximately 250 m in the SW direction and a road with intense traffic about 500 m away in the W–NW direction.

The outdoor sampling of $PM_{2.5}$, air and ambient air temperature, outdoor relative humidity, outdoor air pressure, wind and rain characteristics was carried out at the rooftop of the University building, around 10 m above the ground while additional data on 24 meteorological parameters were obtained from Global Data Assimilation System (GDAS1). The indoor air and $PM_{2.5}$ sampling were performed at a height of 6 m and 2 m from the floor, respectively, in an amphitheater where the number of students ranged from 50 to 80 during the study. Indoor ambient air temperature, indoor relative humidity, and indoor air pressure were registered as well.

Air sampling system consisted of diaphragm vacuum pump Pfeiffer MVP, manifolds with openings for the continuous monitoring of inorganic gaseous pollutants (O_3, CO, SO_2, and NO_x) using Horiba devices (APOA, APMA, APSA, and APNA, 370 series) with 2-min resolution, and electronically controlled valves, which operated in alternating indoor/outdoor air sampling mode in the 10-min cycles. Over 24 h period, $PM_{2.5}$ was collected daily on quartz filters (Whatman QMA, 47 mm) using Svan Leckel LVS6-RV devices operating at a flow rate of 2.3 $m^3\ h^{-1}$. The concentrations of $PM_{2.5}$ and their constituents, including trace elements (As, Cd, Cr, Mn, Ni, and Pb), ions (Cl^-, Na^+, Mg^{2+}, Ca^{2+}, K^+, NO_3^-, SO_4^{2-}, and NH_4^+) and 16 US EPA PAHs were

determined at the reference laboratory of the Institute of Public Health of Belgrade. More details on the study area, sampling campaign, chemical analyses, quality assurance/quality control, and the relevant references are described previously (Stanišić et al., 2021).

Air Pollutant Sampling and Chemical Analyses

Inorganic gases (O_3, CO, SO_2, and NO_x) were continuously measured at indoor and outdoor sampling sites during a three-month campaign using Horiba devices (APOA, APMA, APSA, and APNA, 370 series) with 2-min resolution. The measurements based on ultraviolet absorption, infrared spectroscopy, ultraviolet fluorescence, and chemiluminescence methods for O_3, CO, SO_2, and NO_x, respectively, were performed according to the following Standards EN 14211:2012, EN 14212:2012, EN 14625:2012, and EN 14626:2012. The limit of detection (LOD) was 0.1 mg m^{-3} for CO and 1 μg m^{-3} for O_3, SO_2, and NO_x.

The radon concentrations (Bq m^{-3}) were measured by SN1029 radon monitor (Sun Nuclear Corporation, NRSB approval code 31822) with a time resolution of 30 min. The monitor contains two diffused junction photodiodes, which simultaneously detect radon, temperature, barometric pressure, and relative humidity. The LOD for radon was 0.1 Bq m^{-3}.

$PM_{2.5}$ was collected on quartz filters (Whatman QMA, 47 mm) for 24 h every day during a three-month campaign. For indoor and outdoor sampling, Svan Leckel LVS6-RV devices with a flow rate of 2.3 m^3 h^{-1} were used. Gravimetric measurements of $PM_{2.5}$ were performed according to the Standard EN 12341:2014. After preconditioning for 48 hours, the filters were weighed twice using a micro-balance (Precisa XR 125 SB) in Class 100 cleanroom and average values of $PM_{2.5}$ mass concentrations were further used. Prior to chemical analysis, filters from the sampling sites were stored in a cool room at 4°C. After gravimetric measurements, each filter (approximately 13.85 cm^2) was divided. Each half has a surface of approximately 1.76 cm^2 and one was used for the analysis of anions and cations (Cl^-, Na^+, Mg^{2+}, Ca^{2+}, K^+, NO_3^-, SO_4^{2-}, and NH_4^+). The remaining half 12.09 cm^2 were cut into two pieces which were used for the determination of trace elements (As, Cd, Cr, Mn, Ni, and Pb) and 16 US EPA PAHs.

For the determination of anion and cation concentrations, the filter extraction by ultra-pure water was carried out for 24 h and further analyzed by ion chromatography (Dionex DX500 IC system, MDL 064 Standard operating

procedure). The LOD was as follows: 2 μg m^{-3} for Cl^- and NO_3^-, 1 μg m^{-3} for SO_4^{2-}, 0.2 μg m^{-3} for NH_4^+, and 8 μg m^{-3} for Ca^{2+}.

The trace elements were determined following the Standard EN 14902:2005. A mixture of HNO_3 (30%):H_2O_2:H_2O (3:2:5) (analytical grade reagents, Merck) and distilled/deionized water (MiliQ, 18.2 MΩ) were used for the extraction (CEN/TC 264 N779) in a microwave accelerated digestor (Anton Paar 3000). The concentrations of As, Cd, Cr, Mn, Ni, and Pb were determined by inductively coupled plasma–mass spectrometry, ICP-MS (Agilent 7500ce with Octopole Reaction System). Standard reference material 2783 NIST (National Institute of Standard and Technology, MD, USA) was used for quality control and the recovery values were within a satisfactory range of ±20%. The LOD was as follows: 0.4 ng m^{-3} for As, 0.05 ng m^{-3} for Cd, and 2 ng m^{-3} for Cr, Mn, Ni, and Pb.

The concentrations of the following 16 priority PAHs (US EPA, 2005): naphthalene (Nap), acenaphthylene (Ace), acenaphthene (Ane), fluorene (Flu), phenanthrene (Phe), anthracene (Ant), fluoranthene (Fla), pyrene (Pyr), benzo[a]anthracene (B[a]A), chrysene (Chy), benzo[b]fluoranthene (B[b]F), benzo[k]fluoranthene (B[k]F), benzo[a]pyrene (B[a]P), dibenz[a,h]anthracene (DB[ah]A), benzo[g,h,i]perylene (B[ghi]P), and indeno[1,2,3-cd]pyrene (I[cd]P) were determined (ISO 12884:2010). Further details have previously been described (Stanišić et al., 2021; Cvetković et al., 2015). A mixture of n-hexane and acetone, 12.5 mL:12.5 mL (US EPA, 1999) was used for the microwave extraction of filters. Extracts were rotary evaporated to 1 mL under reduced pressure (55.6 kPa and with 0.2 ml isooctane) and to 0.25 mL under a nitrogen stream. Gas chromatography–mass selective detector (Agilent GC 6890/5973 MSD) with a DB-5 MS capillary column (30 m × 0.25 mm × 25 μm) was used for the analysis of PAHs (EPA Compendium Method TO-13A). The oven temperature was programmed as follows: (1) isothermal heating for 4 minutes at 70°C, (2) gradient heating from 70°C to 310°C at 8°C min^{-1}, and (3) isothermal heating for 5 minutes at 310°C. The solvent delay was 5 min and the time of run was 46 min. Helium was used as the carrier gas while the temperature of the injector was set to 300°C.

Prior to the analysis, calibration curves ($R^2 > 0.995$) with the concentration range between 5 and 200 ng mL^{-1} were obtained using Ultra Scientific PAH Mixture PM-831, which contains 16 US EPA PAHs. Ultra Scientific PAH Mixture PM-831, which consists of 16 compounds, each of 500.8 ± 2.5 μg mL^{-1} concentration was used as an external standard for the calibration curve. We used Ultra Scientific Semi-Volatiles Internal Standard Mixture ISM-560 (Ane-d10, Chy-d10, 1,4-dichlorobenzene, Nap-d8,

Perylene-d12, and Phe-d10) as an internal standard for the estimation of method recovery. Recovery values ranged from 85% to 110% for all the PAHs in the internal standard. The LOD was calculated as three times signal/noise and it was 0.01 ng m^{-3} while the limit of quantification was determined as 3.3 times of LOD. All data were corrected with reference to the field and laboratory blanks, which were prepared and analyzed as well.

Data Analysis

Machine Learning

To estimate the relationships between B[b]F levels and other PAHs, inorganic gaseous pollutants, trace elements, ions, and meteorological parameters in indoor and outdoor environments, the regression analysis was implemented using eXtreme Gradient Boosting. XGBoost is a technique of building a complex prediction model by iteratively combining ensembles of weak prediction models into a single strong learner possessing advantages such as computational efficiency and competitive accuracy, even when sparse and unstructured data are used (Hartmann, 2019, Lundberg et al., 2020). It builds a sequential series of smaller decision trees, where each tree efforts to complement all others and correct for the residuals in the predictions made by all previous trees (Sheridan et al., 2016). In this study, we used Python (Python Software Foundation) XGBoost implementation (XGBoost Python Package). The dataset was split into training (80%) and validation (20%) sets. Hyperparameter tuning was implemented using a brute-force grid search and stratified cross-validation that was replicated ten times. The best performing hyperparameter values were used for the final model.

Explainable Artificial Intelligence

The explainability of the ML model behavior that operates with high-dimensional input data in a non-linear and nested fashion is crucial for understanding the process being modeled. For this purpose, we employed the advanced explainable artificial intelligence method, which is capable to provide a straightforward and meaningful interpretation of ML model-derived decisions, now being shifted towards user-readable logic rules (Stanišić et al., 2021).

SHapley Additive exPlanations

SHapley Additive exPlanations is a method based on Shapley values, calculated as a measure of feature importance using a game-theory approach (Lundberg and Lee, 2017). In this study we used Python SHAP implementation (SHAP Python package). The captured attributed importance of a feature, the change of feature importance over its value range, and its interaction effects with other features are visually presented as SHAP summary plots, SHAP dependency plots, and SHAP interaction plots, respectively.

Fuzzy Clustering

The fuzzy clustering of absolute SHAP attributions was performed to identify and characterize indoor and outdoor ambient conditions responsible for B[b]F behavior. It was chosen because each event will not necessarily belong to a single class of environmental conditions that shape it. Fuzzy clustering was performed by using an R 'cluster' package (Maechler et al., 2019). The obtained results were presented as force plots. A detailed analysis of each cluster was performed based on the statistical character of its absolute and relative SHAP values, as well as the measured parameter values.

Results and Discussion

Descriptive Statistics

The results have shown that the mean B[b]F level was negligibly higher in the outdoor environment (0.9 ng m^{-3}) compared to the indoor (0.7 ng m^{-3}). Thereby, minimum B[b]F levels were similar both indoor and outdoor (0.08 vs. 0.10 ng m^{-3}), while the maximum levels were registered indoor (7.1 vs. 3.8 ng m^{-3}). The study of Živković et al., (2015) made a comparison of total gas and particle-bound B[b]F concentrations registered at different school locations throughout the year. The average winter levels were in the range from 1.72 to 9.11 ng m^{-3}, and from 3.36 to 13.68 ng m^{-3}, for indoor and outdoor environments, respectively, while the concentrations were significantly lower during the warm season, ranging from 0.15 to 1.49 ng m^{-3}.

According to the Pearson's correlation analysis, indoor B[b]F levels significantly correlated (r>0.90) with the levels of B[a]A, Chy, I[cd]P, B[ghi]P, B[a]P, B[k]F and Pyr, while in the outdoor environment B[b]F levels exhibited significant correlations ($r > 0.90$) with B[a]A, Chy, I[cd]P, B[a]P, B[k]F, and B[ghi]P levels. While the listed pollutants were observed to share common sources and display similar behavior patterns within the environment they were analyzed in, the correlation between indoor and outdoor B[b]F levels was not too high (0.62; $p < 0.05$). The outdoor PAH air burden can be associated with emissions from fossil fuel combustion for numerous purposes, including transport, and have an impact on indoor air quality. Nevertheless, in addition to this, indoor PAH levels are influenced by the reemission, resuspension, and sorption of PAHs from indoor dust and human activities (Morawska et al., 2013; Stanišić et al., 2021). Furthermore, no significant correlation was observed between B[b]F levels and meteorological variables, concentrations of inorganic gases, trace elements, and ions. In addition to this, the number of people, the time they spent indoors, temporal trends in concentrations, weekday, and weekend were also found to be irrelevant for predicting B[b]F levels.

In order to further investigate the relationships between B[b]F and other variables, we have amended these findings with the results of the advanced machine learning methods. For this purpose, XGBoost was successfully employed for the exploration of non-linear relationships between B[b]F levels and key variables that shape its distribution in indoor and outdoor air. As indicated by the predicted and observed relative errors (12.5% and 10.5%, respectively) and high correlation coefficients ($r^2 = 0.97$ and 0.98 for indoor and outdoor, respectively), the model was successfully applied (Figure 1).

As shown by SHAP analysis, the most important variables that shaped B[b]F levels in the indoor environment were attributed the highest positive (up to 0.8) and negative (up to -0.2) SHAP values (Figures 2 and 3), according to the following order: B[k]F, Chy, B[a]A, I[cd]P, B[a]P, Pyr, Rh (relative humidity), CO, and Fla. Besides, the analysis revealed less significant impacts of D[ah]A, B[ghi]P, and inorganic elemental and ionic $PM_{2.5}$ constituent levels (Cr, Rn, As, SO_4^{2-}, and NH_4^+) on B[b]F behavior patterns in indoor environment. The similar applies to the outdoor environment as the highest impacts on the B[b]F distribution, described by SHAP values (from -0.15 to 0.5) were attributed to the following compounds: B[a]P, B[a]A, B[k]F, Chy, I[cd]P, and B[ghi]P. Less significant effects on B[b]F outdoor levels (SHAP = -0.1 – 0.1) were attributed to both PAHs and inorganic pollutants (Pyr, CO, NH_4^+, Phen, As, NO_3^-, Fla, Mn, D[ah]A, Cr, and Ant), as well as to

meteorological parameters including cinh (convective inhibition), WD (wind direction), temperature (temp), and lib4 (best 4-layer lifted index).

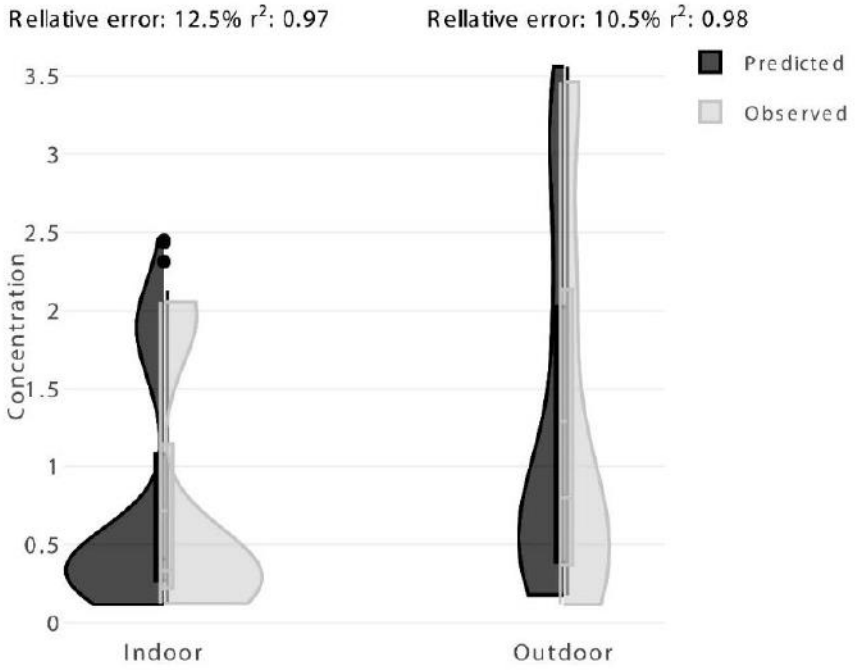


Figure 1. XGBoost model evaluation.

Coal combustion and related pyrogenic processes were identified as dominant sources that contributed to the registered levels of B[b]F, B[k]F, Chy, B[a]A, B[a]P, I[cd]P, and B[ghi]P at the measurement site. As indicated by SHAP dependence plots, the highest impacts of CO, As, Cr, and Mn, corresponding to the highest concentrations of these compounds, suggest common sources and processes which shape B[b]F levels.

The increases in the concentrations of ionic species (outdoor NH_4^+: up to 11.9 μg m^{-3}; indoor NH_4^+: up to 7.2 μg m^{-3}; outdoor NO_3^-: up to 17 μg m^{-3}; and indoor SO_4^{2-}: up to 6.9 μg m^{-3};) had a positive impact on B[b]F levels both indoor and outdoor. As indicated by SHAP dependence plots (Figure 4 and 5), higher concentrations of nitrate ions (> 8 μg m^{-3}) were related to lower B[b]F levels (< 2 ng m^{-3}), whereas the increase in SO_4^{2-} and NH_4^+ levels corresponded to the increase in B[b]F concentrations.

As regards meteorological parameters, the impact of northern winds, convective inhibition and lifted index on B[b]F outdoor levels was evidenced (cinh: up to 0 J kg^{-1}; lib4: up to 10.5 K), as well as the minor impact of lower temperatures (down to 4.8°C). The impact of wind direction on $PM_{2.5}$ concentrations and the associated compounds depends on local geographical and topographic conditions and may lead to the reduction or enhanced accumulation of pollutants (Chen et al., 2020). The relationship between wind direction and outdoor B[b]F levels confirmed the findings of previous studies that have shown that, in addition to many individual local emissions, long-range transport, and emissions from distant sources including thermal power plants, petrochemical industry, and oil refineries affect air quality over the study area (Perišić et al., 2017; Stojić et al., 2016). The relationship between B[b]F concentrations and convective inhibition, which represents a measure of the energy required by the atmosphere to inhibit the rising of air, can be explained by the fact that vertical air mass movements affect the retention of particle-bound PAHs in the ground layers. Besides, under low-temperature conditions, atmospheric convection weakens and enhances the accumulation of $PM_{2.5}$. In contrast to this, high temperatures would induce turbulent movements that would further accelerate the dispersion of $PM_{2.5}$ (Li et al., 2015; Yang et al., 2016). Finally, the importance of the positive lifted index suggests that the outdoor B[b]F concentration dynamics of is shaped by stable tropospheric conditions with respect to boundary layer-based convection.

In the indoor environment, a variable impact of relative humidity on B[b]F levels has been shown. Namely, a negative impact on B[b]F levels corresponded with higher values of relative humidity (up to 57%), probably because the gas to particle partitioning of pollutants is favored under these conditions, as well as the increase in $PM_{2.5}$ concentrations and the uptake of B[b]F and other high-ring PAHs onto particles with high organic content (Wang and Ogawa, 2015; Liao et al., 2017).

In contrast to this, lower humidity (down to 20%) had a positive impact on B[b]F levels because it led to an enhanced evaporation loss of $PM_{2.5}$ and associated contaminants (Liu et al., 2015). Madruga et al., (2019) also emphasized relative humidity as a prominent factor, which has an impact on the PAH behavior.

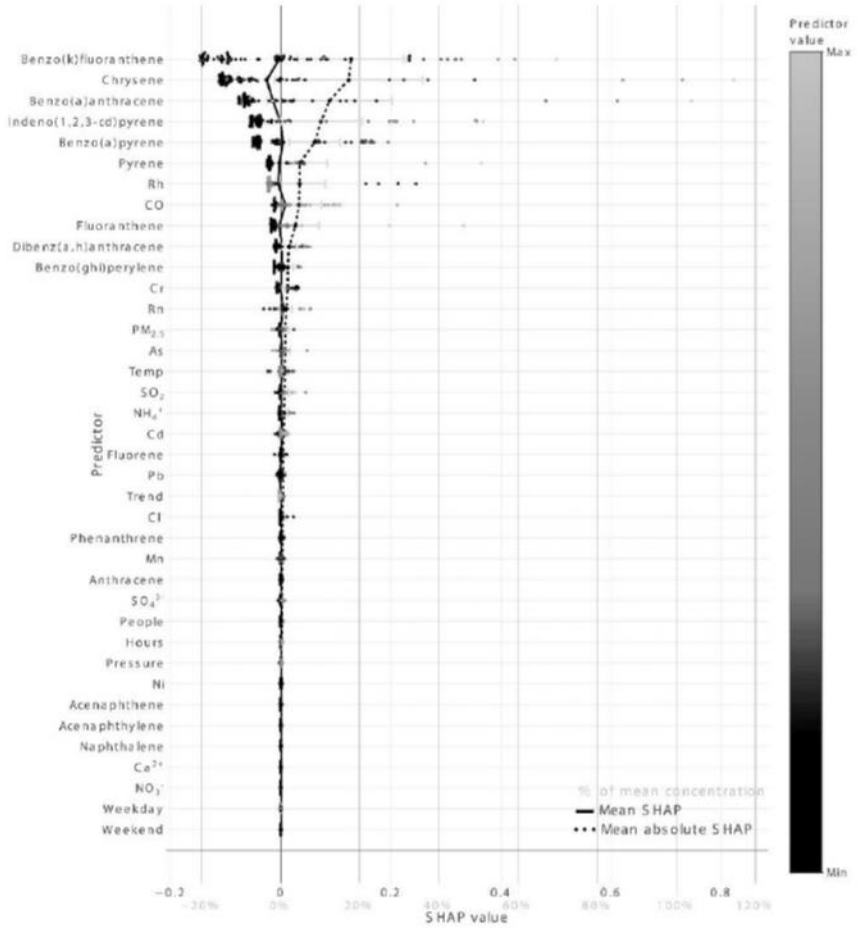


Figure 2. Indoor B[b]F SHAP summary plot.

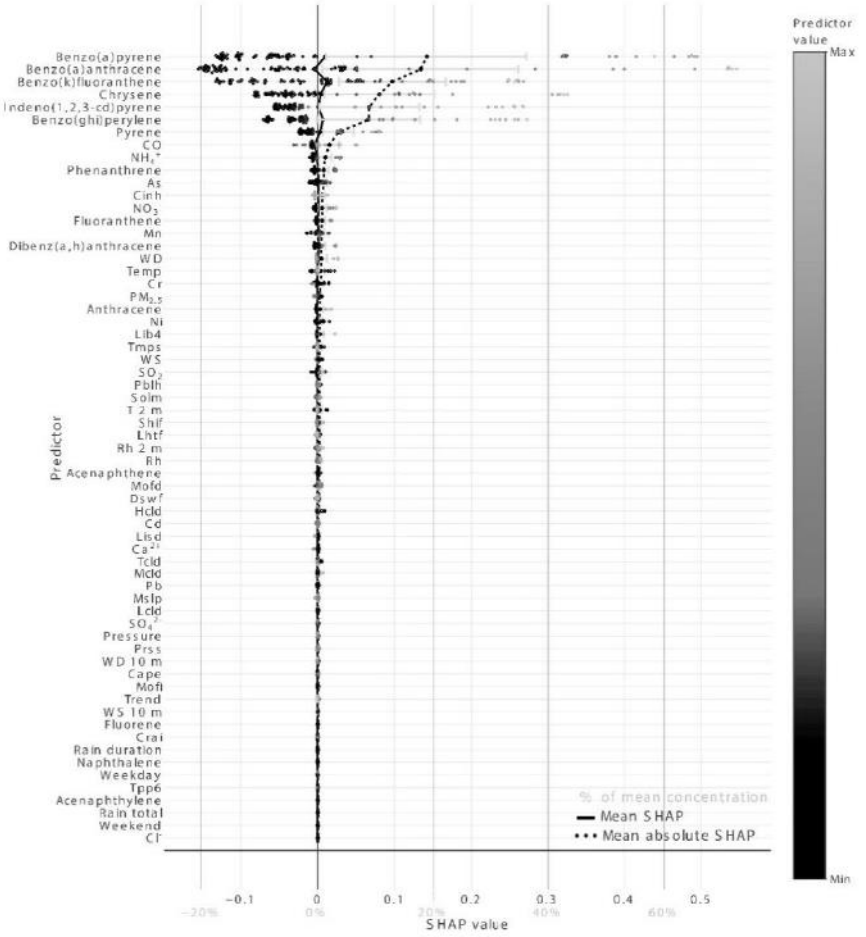


Figure 3. Outdoor B[b]F SHAP summary plot.

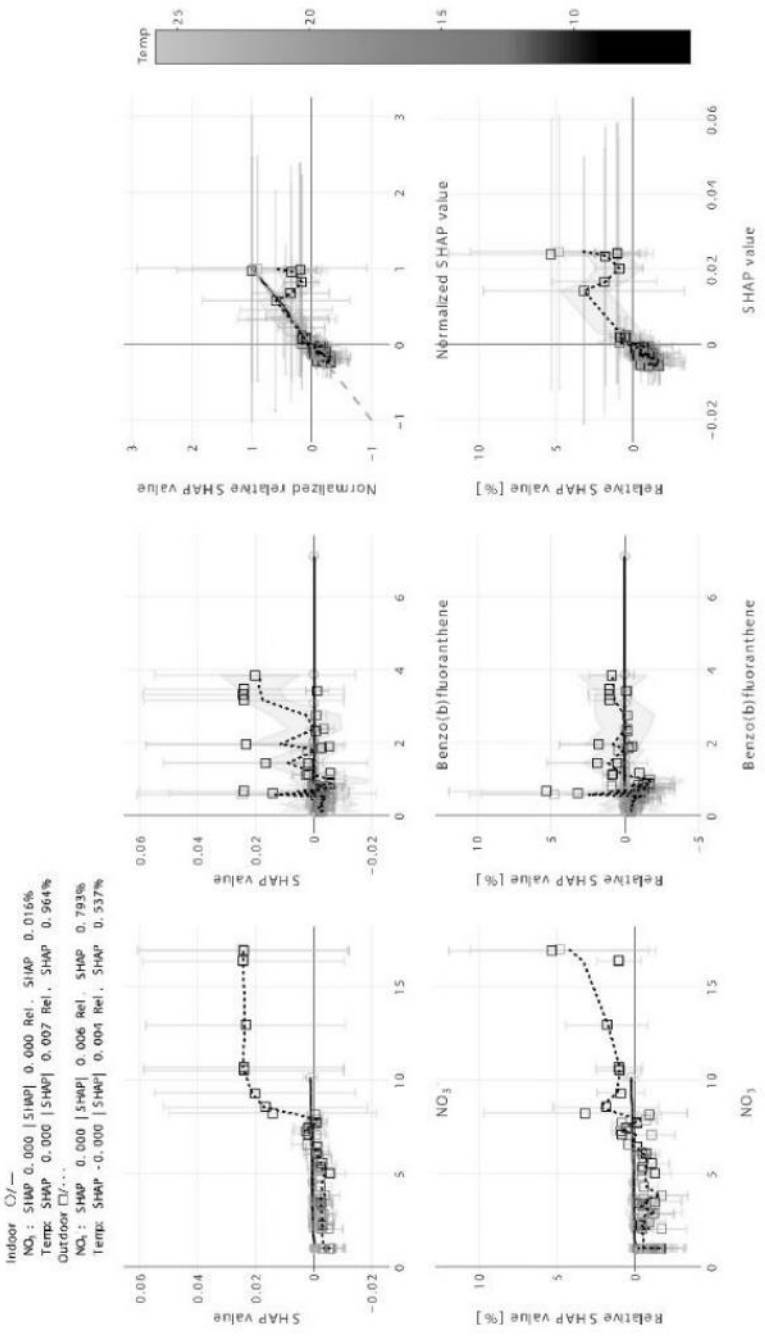


Figure 4. B[b]F SHAP dependence on NO_3^- and temperature.

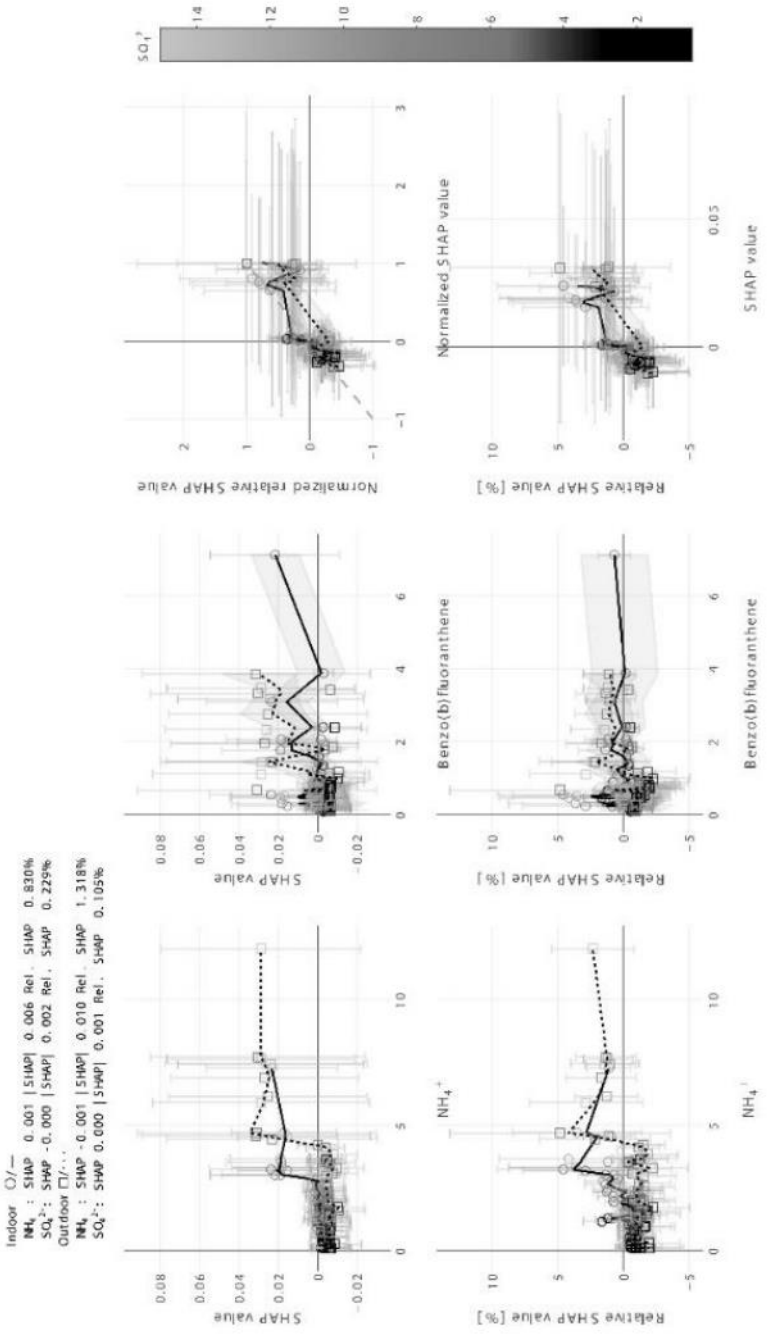


Figure 5. B[b]F SHAP dependence on NH_4^+ and SO_4^{2-}.

Interactions between Environmental Factors and Indoor and Outdoor B[b]F

Although many investigations evidenced that indoor air quality depends on outdoor pollutant concentrations, it does not represent a simple reflection of the outdoor ambiance. Figure 6 represents the fuzzy clustering of SHAP values, according to which eight groups of variables that shape B[b]F dynamics in indoor and outdoor (relative error < 25% and 30%, respectively) environments were identified.

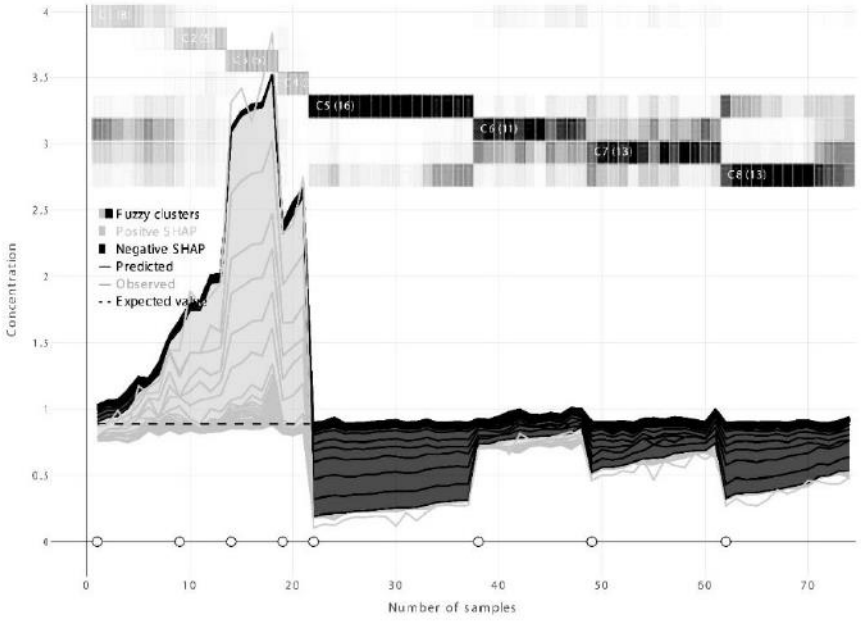


Figure 6. Outdoor B[b]F SHAP force plot.

The constituents of four clusters (°C1: March 4th – April 30th; °C2: March 7th – April 11th; °C3: March 10th – March 25th; and °C4: March 20th – May 17th) had a positive impact on outdoor B[b]F levels from the mean to 95th quantile values (0.87–3.31 ng m^{-3}). Thereby, the highest impacts indicated by relative SHAP values (%) were attributed to the following PAHs: B[a]A, B[k]F, Chy, B[a]P, I[cd]P, B[ghi]P, and Pyr. Additionally, minor impacts were revealed for convective inhibition, NH_4^+, NO_3^-, and As. The highest B[b]F concentrations during the period assigned to °C3 were accompanied by

the increased levels of the pollutants (B[ghi]P: 3.44 ng m^{-3}; Chy: 2.73 ng m^{-3}; B[k]F: 2.33 ng m^{-3}; B[a]A: 2.04 ng m^{-3}; B[a]P: 2.02 ng m^{-3}; I[cd]P: 2.00 ng m^{-3}; Pyr: 0.95 ng m^{-3}; As: 1.84 ng m^{-3}; NH_4^+: 6.18 µg m^{-3}; and NO_3^-: 10.66 µg m^{-3}) originating from the joint pollution sources. Low-temperature periods were associated with intensified pollutant emissions from heating sources. Furthermore, the absence of PAH photolytic degradation also contributed to the elevated outdoor pollutant levels (Stanišić et al., 2021). The periods assigned to °C1, °C2, and °C4 were characterized by changeable weather conditions with occasional precipitations, strong winds, and penetration of warmer and dry air masses. During these periods, reduced activities of heating plants and individual fireboxes using coal and wood, as well as dispersion and photooxidative degradation caused by higher temperatures and stronger winds led to a decrease in PAH outdoor levels (from 0.4 ng m^{-3} to 2.5 ng m^{-3}, belonging from 25th to 75th quantile values).

Fuzzy clustering also identified clusters (°C5: March 26th – May 31st; °C6: March 30th – May 20th; °C7: April 1st – May 24th; and °C8: April 7th – May 25th) representing the groups of variables related to 25th to 75th quantile outdoor B[b]F values (from 0.27 ng m^{-3} to 0.28 ng m^{-3}) (Figure 6). The same predictors as discussed above shaped the clusters, but relative SHAP values (%) revealed negative interrelations between B[b]F and certain pollutant (B[a]A, B[k]F, Chy, B[a]P, I[cd]P, B[ghi]P, and CO) behavior patterns. The presence of CO can be attributed to burning wood or coal in individual heating units in densely populated residential areas surrounding the measurement site. The contribution of traffic emissions to PAH burden was lower compared to the contributions of fossil fuel burning for heating operations, and it appeared to remain stable over the measurement campaign.

Four clusters (^{i}C1: March 4th – May 9th; ^{i}C4: March 7th – March 24th; ^{i}C5: March 10th – March 28th; and ^{i}C6: March 18th – March 20th), grouping variables with a positive impact on B[b]F indoor distribution, were identified (Figure 7). Two of them (^{i}C5 and ^{i}C6) were more differentiated with the highest B[b]F indoor levels (from 1.41 to 7.11 ng m^{-1}). The periods assigned to these clusters corresponded to °C3, which implies the association between outdoor and indoor air quality. During the mentioned periods, the episodes of cold, dry weather (temperature below 10°C) were linked to the intensified emissions from heating sources. As shown by relative SHAP values (%), the higher indoor B[b]F concentrations were shaped by the increased levels of 4- (Chy and B[a]A; Fla and Pyr), 5- (B[a]P and B[k]F), and 6-ring (B[ghi]P and I[cd]P) structural PAH isomers. The impact of fossil fuel combustion for heating purposes was additionally confirmed by non-negligible attributions of

CO to certain clusters (^{i}C4: 4%; ^{i}C5: 6%; and ^{i}C6: 6%) related to the elevated concentration of this gas (^{i}C4: 0.82 mg m^{-3}; ^{i}C5: 1.13 mg m^{-3}; and ^{i}C6: 5.53 mg m^{-3}). The impact of relative humidity increased with the increase of its value. In general, relative humidity was twice lower indoor (36.8%) than outdoor (61.5%) during the measurement campaign, due to the heating of indoor spaces. Under reduced humidity conditions, evaporation of pollutants from particle surface may occur (Liu et al., 2015) in contrast to their enhanced adsorption/absorption in circumstances of higher relative humidity. This was evident in a significantly lower abundance of semi-volatile Chy in ^{i}C4 compared to ^{i}C6, later being characterized by higher humidity.

Low B[b]F indoor concentrations (< 0.53 ng m^{-3}) indoor were associated with four clusters (^{i}C2: March 5th – May 26th; ^{i}C3: March 6th – May 24th; ^{i}C7: April 10th – May 27th; and ^{i}C8: April 15th – May 23rd) which comprised the negative impacts of B[a]A, Chy, B[a]P, Fla, I[cd]P, Pyr, B[k]F, D[ah]A, relative humidity and Rn on B[b]F indoor behavior. The strength of their effects is indicated by the relative SHAP values (%). It is worth noting that the four analyzed clusters were related to the lower occurrence of B[a]A, Chy, B[a]P, Fla, I[cd]P, Pyr, B[k]F, D[ah]A (< 0.5 ng m^{-3}) compared to ^{i}C1, ^{i}C4, ^{i}C5, and ^{i}C6.

Radon is usually presented indoors as a buildup impurity originating from underlying soil, building material and its granular size, pore air spaces, and moisture content (Burghele et al., 2021). In a microstructure level, Rn occupies pore spaces inside mineral surface layers and could migrate by diffusion and advection until it releases from the opened pathways into the atmosphere (Syuryavin et al., 2020). The low positive impact of Rn, occurring in high concentrations (> 70 Bq m^{-3}), implies that dust resuspension, containing both particle-bound and pore-entrapped pollutants, is a non-negligible pollution source in the indoor environment apart from emissions of fossil fuel combustion for the industrial, heating and traffic purposes, and pollutant diffusion from outdoors.

Generarly, the clusters distinguished by high to excessive concentrations of PAHs could be considered the most influential for the observed high outdoor and indoor B[b]F levels, as indicated by the absolute and relative SHAP values.

Absolute and relative SHAP interaction values implied the most significant interactions among the studied variables that shape indoor and outdoor B[b]F behaviour. The most important pairs of predictors indoors were: B[a]P–I[cd]P, B[k]F–SO_4^{2-}, Chy–Cr, I[cd]P–CO, D[ah]P–NH_4^+, I[cd]P–Pb, Cr–Pb, B[a]A–CO, B[k]F–CO, Chy–CO, Chy–B[a]P, Chy–B[k]F,

Chy–$PM_{2.5}$, Chy–Rn, and Rh–CO, while in the outdoor environment the following pairs were recognized as important: B[a]P–B[a]A, B[a]P–Chy, B[a]P–B[k]F, B[k]F–B[a]A, I[cd]P–B[a]A, I[cd]P–B[a]P, I[cd]P–B[ghi]P, I[cd]P–Chy, B[ghi]P–B[a]A, and B[ghi]P–CO.

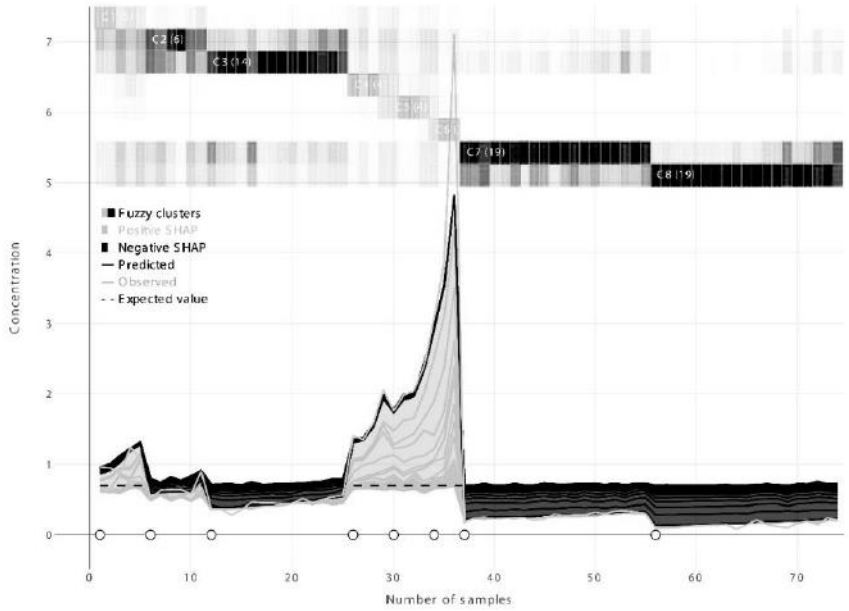


Figure 7. Indoor B[b]F SHAP force plot.

Acknowledgments

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS.

Conclusion

Although B[a]P is commonly used as a marker of exposure to mixtures of carcinogenic PAHs in indoor and outdoor environments, emerging research points the need to shift towards the investigation of atmospheric behavior of

other health-relevant possible/probable carcinogenic PAHs. This chapter resumes the finding based on the application of machine learning and explainable artificial intelligence (as indicated by the model evaluation statistics) for investigating the relationships between B[b]F levels and dozens of variables including inorganic and organic $PM_{2.5}$ constituents, meteorological parameters, and the number of people in the amphitheater, and the time they spent indoor. According to the results, high molecular weight particle-bound PAHs (B[a]P, B[k]F, B[ghi]P, I[cd]P, and D[ah]A), followed by 4-ring isomers (Chy and B[a]A; Fla and Pyr) were dominantly related with the presence of B[b]F in both indoor and outdoor environments. Under stable tropospheric conditions, temperature, convective inhibition, wind direction, and lifted index exhibited less significant impacts on B[b]F fate outdoors while relative humidity appeared to be the most distinctive meteorological parameter indoors. Less important functional dependencies were observed between B[b]F and inorganic pollutants (CO, Rn, As, Cr, SO_4^{2-}, NH_4^+, and NO_3^-) suggesting their similar origin and participation in oxidative degradation and aerosol formation. The herein presented explainable artificial intelligence methodology represents a promising tool for exploring and understanding the complexity of environmental conditions which shape air pollution processes.

References

Bansal, V. and Kim, K. H. (2015). Review of PAH contamination in food products and their health hazards. *Environment International*, 84, 26-38.

Blair, G. S., Henrys, P., Leeson, A., Watkins, J., Eastoe, E., Jarvis, S. and Young, P. J. (2019). Data science of the natural environment: a research roadmap. *Frontiers in Environmental Science*, 7, 121-129.

Brown, A. S., Sarantaridis, D., Butterfield, D. M., Brown, R. J., Whiteside, K. J., Hughey, P., Goddard, S. L., Hussain, D. and Williams, M. (2012). *Annual Report for 2011 on the UK PAH Monitoring and Analysis Network.*

Burghele, B. D., Botoș, M., Beldean-Galea, S., Cucoș, A., Catalina, T., Dicu, T., Dobrei, G., Florică, Ș., Istrate, A., Lupulescu, A. and Moldovan, M. (2021). Comprehensive survey on radon mitigation and indoor air quality in energy efficient buildings from Romania. *Science of The Total Environment*, *751*, 141858.

Cao, Z., Wang, M., Chen, Q., Zhu, C., Jie, J., Li, X., Dong, X., Miao, Z., Shen, M. and Bu, Q. (2019). Spatial, seasonal and particle size dependent variations of PAH contamination in indoor dust and the corresponding human health risk. *Science of the Total Environment*, 653, 423-430.

Chen, H., Lundberg, S. and Lee, S. I. (2021). Explaining models by propagating Shapley values of local components. In *Explainable AI in Healthcare and Medicine*, 261-270. Springer, Cham.

Chen, Z., Chen, D., Zhao, C., Kwan, M. P., Cai, J., Zhuang, Y., Zhao, B., Wang, X., Chen, B., Yang, J. and Li, R. (2020). Influence of meteorological conditions on $PM_{2.5}$ concentrations across China: A review of methodology and mechanism. *Environment International*, 139, 105558.

Cvetković, A., Jovašević-Stojanović, M., Marković, D. and Ristovski, Z. (2015). Concentration and source identification of polycyclic aromatic hydrocarbons in the metropolitan area of Belgrade, Serbia. *Atmospheric Environment*, 112, 335-343.

Elorduy, I., Elcoroaristizabal, S., Durana, N., García, J. A. and Alonso, L. (2016). Diurnal variation of particle-bound PAHs in an urban area of Spain using TD-GC/MS: influence of meteorological parameters and emission sources. *Atmospheric Environment*, 138, 87-98.

European Standards (EN) 12341:2014 (2014). *Ambient air. Standard gravimetric measurement method for the determination of the PM_{10} or $PM_{2.5}$ mass concentration of suspended particulate matter*. https://cds.cern.ch/record/ 2624772 (Last accessed February 4, 2021).

European Standards (EN) 14211:2012 (2012). *Ambient air. Standard method for the measurement of the concentration of nitrogen dioxide and nitrogen monoxide by chemiluminescence*. https://www.en-standard.eu/bs-en-14211-2012-ambient-air.-standard-method-for-the-measurement-of-the-concentration-of-nitrogen-dioxide-and-nitrogen-monoxide-by-chemiluminescence/ (Last accessed February 4, 2021).

European Standards (EN) 14212:2012 (2012). *Ambient air. Standard method for the measurement of the concentration of sulphur dioxide by ultraviolet fluorescence*. https://www.sis.se/en/produkter/ environment-health-protection-safety/air-quality/ ambient-atmospheres/ssen142122012/ (Last accessed February 4, 2021).

European Standards (EN) 14625:2012 (2012). *Ambient air. Standard method for the measurement of the concentration of ozone by ultraviolet photometry*. https://shop. bsigroup.com/ProductDetail? pid=000000000030210754 (Last accessed February 4, 2021).

European Standards (EN) 14626:2012 (2012). *Ambient air. Standard method for the measurement of the concentration of carbon monoxide by non-dispersive infrared spectroscopy*. https://www.en-standard.eu/bs-en-14626-2012-ambient-air.-standard-method-for-the-measurement-of-the-concentration-of-carbon-monoxide-by-non-dispersive-infrared-spectroscopy/ (Last accessed February 4, 2021).

European Standards (EN) 14902:2005 (2005). *Ambient air quality. Standard method for the measurement of Pb, Cd, As, and Ni in the PM_{10} fraction of suspended particulate matter*. https://www. antpedia.com/standard/pdf/ 13.040.20/1703/EN%2014902-2005_ 6234.pdf (Last accessed February 4, 2021).

Ferguson, L., Taylor, J., Davies, M., Shrubsole, C., Symonds, P. and Dimitroulopoulou, S. (2020). Exposure to indoor air pollution across socio-economic groups in high-income countries: A scoping review of the literature and a modelling methodology. *Environment International*, 143, 105748.

Friesen, M. C., Demers, P. A., Spinelli, J. J., Eisen, E. A., Lorenzi, M. F. and Le, N. D. (2010). Chronic and acute effects of coal tar pitch exposure and cardiopulmonary mortality among aluminum smelter workers. *American Journal of Epidemiology*, 172, 790-799.

Gaga, E. O. and Ari, A. (2019). Gas-particle partitioning and health risk estimation of polycyclic aromatic hydrocarbons (PAHs) at urban, suburban and tunnel atmospheres: Use of measured EC and OC in model calculations. *Atmospheric Pollution Research*, 10, 1-11.

Ghosal, D., Ghosh, S., Dutta, T. K. and Ahn, Y. (2016). Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): a review. *Frontiers in Microbiology*, 7, 1369-1382.

Gitipour, S., Sorial, G. A., Ghasemi, S. and Bazyari, M. (2018). Treatment technologies for PAH-contaminated sites: a critical review. *Environmental Monitoring and Assessment*, 190, 1-17.

Gohlke, J. M., Doke, D., Tipre, M., Leader, M. and Fitzgerald, T. (2011). A review of seafood safety after the Deepwater Horizon blowout. *Environmental Health Perspectives*, 119, 1062-1069.

González-Martín, J., Kraakman, N., Pérez, C., Lebrero, R. and Muñoz, R. (2020). A state-of-the-art review on indoor air pollution and strategies for indoor air pollution control. *Chemosphere*, 262, 128376.

Gune, M. M., Ma, W. L., Sampath, S., Li, W., Li, Y. F., Udayashankar, H. N., Balakrishna, K. and Zhang, Z. (2019). Occurrence of polycyclic aromatic hydrocarbons (PAHs) in air and soil surrounding a coal-fired thermal power plant in the south-west coast of India. *Environmental Science and Pollution Research*, 26, 22772-22782.

Hao, X., Li, J. and Yao, Z. (2016). Changes in PAHs levels in edible oils during deep-frying process. *Food Control*, 66, 233-240.

Haritash, A. K. and Kaushik, C. P. (2009). Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review. *Journal of Hazardous Materials*, 169, 1-15.

Hitzel, A., Pöhlmann, M., Schwägele, F., Speer, K. and Jira, W. (2013). Polycyclic aromatic hydrocarbons (PAH) and phenolic substances in meat products smoked with different types of wood and smoking spices. *Food Chemistry*, 139, 955-962.

Hussar, E., Richards, S., Lin, Z. Q., Dixon, R. P. and Johnson, K. A. (2012). Human health risk assessment of 16 priority polycyclic aromatic hydrocarbons in soils of Chattanooga, Tennessee, USA. *Water, Air, & Soil Pollution*, 223, 5535-5548.

International Agency for Research on Cancer (2012). A review of human carcinogens. Part F: Chemical agents and related occupations. *IARC monographs on the evaluation of carcinogenic risks to humans.* https://www. thelancet.com/journals/lanonc/article/ PIIS1470-2045(09)70358-4/fulltext (Last accessed: February 2, 2021).

International Standardization Organization (ISO) 12884:2000 (2020). Ambient air — Determination of total (gas and particle-phase) polycyclic aromatic hydrocarbons — *Collection on sorbent-backed filters with gas chromatographic/mass spectrometric analyses*. https://www.iso.org/ standard/ 1343.html (Last accessed February 4, 2021).

Johnsen, A. R., Wick, L. Y. and Harms, H. (2005). Principles of microbial PAH-degradation in soil. *Environmental Pollution*, 133, 71-84.

Jovanović, G., Romanić, S. H., Stojić, A., Klinčić, D., Sarić, M. M., Letinić, J. G. and Popović, A. (2019). Introducing of modeling techniques in the research of POPs in breast milk–A pilot study. *Ecotoxicology and environmental safety*, *172*, 341-347.

Kim, K. H., Jahan, S. A., Kabir, E. and Brown, R. J. (2013). A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. *Environment International*, 60, 71-80.

Lee, J. G., Kim, S. Y., Moon, J. S., Kim, S. H., Kang, D. H. and Yoon, H. J. (2016). Effects of grilling procedures on levels of polycyclic aromatic hydrocarbons in grilled meats. *Food Chemistry*, 199, 632-638.

Lee, R. F. (2003). Photo-oxidation and photo-toxicity of crude and refined oils. *Spill Science & Technology Bulletin*, 8, 157-162.

Li, J., Chen, H., Li, Z., Wang, P., Cribb, M. and Fan, X. (2015). Low-level temperature inversions and their effect on aerosol condensation nuclei concentrations under different large-scale synoptic circulations. *Advances in Atmospheric Sciences*, 32, 898-908.

Liao, T., Wang, S., Ai, J., Gui, K., Duan, B., Zhao, Q., Zhang, X., Jiang, W. and Sun, Y. (2017). Heavy pollution episodes, transport pathways and potential sources of $PM_{2.5}$ during the winter of 2013 in Chengdu (China). *Science of the Total Environment*, 584, 1056-1065.

Liu, C. N., Lin, S. F., Tsai, C. J., Wu, Y. C. and Chen, C. F. (2015). Theoretical model for the evaporation loss of $PM_{2.5}$ during filter sampling. *Atmospheric Environment*, 109, 79-86.

Lowor, S. T., Jacquet, M., Vrielink, T., Aculey, P., Cros, E. and Takrama, J. (2012). Post-harvest sources of polycyclic aromatic hydrocarbon contamination of cocoa beans: a simulation. *International Journal of AgriScience*, 2, 1043-1052.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 56-67.

Lundberg, S. M., Lee, S. I. (2017). A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 30* (NIPS 2017), 4765-4774.

Lundstedt, S., White, P. A., Lemieux, C. L., Lynes, K. D., Lambert, I. B., Öberg, L., Haglund, P. and Tysklind, M. (2007). Sources, fate, and toxic hazards of oxygenated polycyclic aromatic hydrocarbons (PAHs) at PAH-contaminated sites. *AMBIO: A Journal of the Human Environment*, 36, 475-485.

Madruga, D. G., Ubeda, R. M., Terroba, J. M., Dos Santos, S. G. and García-Cambero, J. P. (2019). Particle-associated polycyclic aromatic hydrocarbons in a representative urban location (indoor-outdoor) from South Europe: assessment of potential sources and cancer risk to humans. *Indoor Air*, 29, 817-827.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2019). Cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0. https://cran.r-project.org/web/packages/cluster/cluster.pdf (Last accessed: January 27, 2021).

Morawska, L., Afshari, A., Bae, G. N., Buonanno, G., Chao, C. Y. H., Hänninen, O., Hofmann, W., Isaxon, C., Jayaratne, E. R., Pasanen, P. and Salthammer, T. (2013). Indoor aerosols: from personal exposure to risk assessment. *Indoor Air*, 23, 462-487.

Oliveira, M., Slezakova, K., Delerue-Matos, C., Pereira, M. C. and Morais, S. (2019). Children environmental exposure to particulate matter and polycyclic aromatic hydrocarbons and biomonitoring in school environments: a review on indoor and outdoor exposure levels, major sources and health impacts. *Environment International*, 124, 180-204.

Parajuli, A., Grönroos, M., Kauppi, S., Płociniczak, T., Roslund, M. I., Galitskaya, P., Laitinen, O. H., Hyöty, H., Jumpponen, A., Strömmer, R. and Romantschuk, M. (2017). The abundance of health-associated bacteria is altered in PAH polluted soils—Implications for health in urban areas? *PLoS One*, 12, e0187852.

Paris, A., Ledauphin, J., Poinot, P. and Gaillard, J. L. (2018). Polycyclic aromatic hydrocarbons in fruits and vegetables: Origin, analysis, and occurrence. *Environmental Pollution*, 234, 96-106.

Paulik, L. B., Hobbie, K. A., Rohlman, D., Smith, B. W., Scott, R. P., Kincl, L., Haynes, E. N. and Anderson, K. A. (2018). Environmental and individual PAH exposures near rural natural gas extraction. *Environmental Pollution*, 241, 397-405.

Pereira, D. C. A., Custódio, D., de Andrade, M. D. F., Alves, C. and de Castro Vasconcellos, P. (2019). Air quality of an urban school in São Paulo city. *Environmental Monitoring and Assessment*, 191, 1-13.

Perišić, M., Rajšić, S., Šoštarić, A., Mijić, Z. and Stojić, A. (2017). Levels of PM_{10}-bound species in Belgrade, Serbia: spatio-temporal distributions and related human health risk estimation. *Air Quality, Atmosphere & Health*, 10, 93-103.

Rostami, R., Kalan, M. E., Ghaffari, H. R., Saranjam, B., Ward, K. D., Ghobadi, H., Poureshgh, Y. and Fazlzadeh, M. (2021). Characteristics and health risk assessment of heavy metals in indoor air of waterpipe cafés. *Building and Environment*, 190, 107557.

Rotkin-Ellman, M., Wong, K. K. and Solomon, G. M. (2012). Seafood contamination after the BP Gulf oil spill and risks to vulnerable populations: a critique of the FDA risk assessment. *Environmental Health Perspectives*, 120, 157-161.

Samburova, V., Zielinska, B. and Khlystov, A. (2017). Do 16 polycyclic aromatic hydrocarbons represent PAH air toxicity? *Toxics*, 5, p. 17.

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. CRC Press.

Slezakova, K., Pires, J. C. M., Castro, D., Alvim-Ferraz, M., Delerue-Matos, C., Morais, S. and Pereira, M. D. C. (2013). PAH air pollution at a Portuguese urban area: carcinogenic risks and sources identification. *Environmental Science and Pollution Research*, 20, 3932-3945.

Stanišić, S., Perišić, M., Jovanović, G., Milićević, T., Romanić, S. H., Jovanović, A., Šoštarić, A., Udovičić, V. and Stojić, A. (2021). The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources. *Environmental Research*, 193, 110520.

Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A. and Lazić, L. (2019). Explainable extreme gradient boosting tree-based prediction of toluene,

ethylbenzene and xylene wet deposition. *Science of The Total Environment*, 653, 140-147.

Stojić, A., Stojić, S. S., Reljin, I., Čabarkapa, M., Šoštarić, A., Perišić, M. and Mijić, Z. (2016). Comprehensive analysis of PM_{10} in Belgrade urban area on the basis of long-term measurements. *Environmental Science and Pollution Research*, 23, 10722-10732.

Syuryavin, A. C., Park, S., Nirwono, M. M. and Lee, S. H. (2020). Indoor radon and thoron from building materials: Analysis of humidity, air exchange rate, and dose assessment. *Nuclear Engineering and Technology*, *52*(10), 2370-2378.

Teixeira, E. C., Agudelo-Castañeda, D. M. and Mattiuzi, C. D. P. (2015). Contribution of polycyclic aromatic hydrocarbon (PAH) sources to the urban environment: a comparison of receptor models. *Science of the Total Environment*, 538, 212-219.

United States Environmental Protection Agency (US EPA) (1999). *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air. Determination of Polycyclic Aromatic Hydrocarbons (PAHs) in Ambient Air Using Gas Chromatography/Mass Spectrometry (GC/MS)*. Center for Environmental Research Information, Office of Research and Development U.S. Environmental Protection Agency, Cincinnati, USA. https://www.epa. gov/sites/production/files/ 2019-11/ documents/to-13arr.pdf. (Last accessed: February 4, 2021).

United States Environmental Protection Agency (US EPA) (2005*). Guidelines for carcinogen risk assessment*, EPA/630/P-03/001F, Washington, D.C., USA.

Varlet, V., Serot, T. and Prost, C. (2010). Smoke flavoring technology in seafood. *Handbook of seafood and seafood products analysis*, 233-254.

Vignet, C., Le Menach, K., Mazurais, D., Lucas, J., Perrichon, P., Le Bihanic, F., Devier, M. H., Lyphout, L., Frère, L., Bégout, M. L. and Zambonino-Infante, J. L. (2014). Chronic dietary exposure to pyrolytic and petrogenic mixtures of PAHs causes physiological disruption in zebrafish-part I: Survival and growth. *Environmental Science and Pollution Research*, 21, 13804-13817.

Vuong, Q. T., Thang, P. Q., Nguyen, T. N. T., Ohura, T. and Choi, S. D. (2020). Seasonal variation and gas/particle partitioning of atmospheric halogenated polycyclic aromatic hydrocarbons and the effects of meteorological conditions in Ulsan, South Korea. *Environmental Pollution*, 263, 114592.

Wang, J. and Ogawa, S. (2015). Effects of meteorological conditions on $PM_{2.5}$ concentrations in Nagasaki, Japan. *International Journal of Environmental Research and Public Health*, 12, 9089-9101.

Wick, A. F., Haus, N. W., Sukkariyah, B. F., Haering, K. C. and Daniels, W. L. (2011). *Remediation of PAH-contaminated soils and sediments: a literature review*. CSES Department, internal research document, 102.

Wilcke, W. (2000). Synopsis polycyclic aromatic hydrocarbons (PAHs) in soil—a review. *Journal of Plant Nutrition and Soil Science*, 163, 229-248.

Yang, X., Zhao, C., Guo, J. and Wang, Y. (2016). Intensification of aerosol pollution associated with its feedback with surface solar radiation and winds in Beijing. *Journal of Geophysical Research: Atmospheres*, 121, 4093-4099.

Yukhimets, A., Kuzu, S. L., Akyüz, E. and Saral, A. (2019). Investigation of geospatial distribution of PAH compounds in soil phase and determination of soil–air exchange direction in a megacity. *Environmental Geochemistry and Health*, 1-14.

Zelinkova, Z. and Wenzl, T. (2015). The occurrence of 16 EPA PAHs in food–a review. *Polycyclic Aromatic Compounds*, 35, 248-284.

Zhang, L., Morisaki, H., Wei, Y., Li, Z., Yang, L., Zhou, Q., Zhang, X., Xing, W., Hu, M., Shima, M. and Toriba, A. (2020). $PM_{2.5}$-bound polycyclic aromatic hydrocarbons and nitro-polycyclic aromatic hydrocarbons inside and outside a primary school classroom in Beijing: Concentration, composition, and inhalation cancer risk. *Science of the Total Environment*, 705, 135840.

Zhang, X. X., Cheng, S. P., Cheng-Jun, Z. H. U. and Shi-Lei, S. U. N. (2006). Microbial PAH-degradation in soil: degradation pathways and contributing factors. *Pedosphere*, 16, 555-565.

Zhao, L., Zhao, Y., Nan, H., Yang, F., Qiu, H., Xu, X. and Cao, X. (2020). Suppressed formation of polycyclic aromatic hydrocarbons (PAHs) during pyrolytic production of Fe-enriched composite biochar. *Journal of Hazardous Materials*, 382, 121033.

Zhao, Z., Zhang, L., Cai, Y. and Chen, Y. (2014). Distribution of polycyclic aromatic hydrocarbon (PAH) residues in several tissues of edible fishes from the largest freshwater lake in China, Poyang Lake, and associated human health risk assessment. *Ecotoxicology and Environmental Safety*, 104, 323-331.

Zhu, Y., Tao, S., Sun, J., Wang, X., Li, X., Tsang, D. C., Zhu, L., Shen, G., Huang, H., Cai, C. and Liu, W. (2019). Multimedia modeling of the PAH concentration and distribution in the Yangtze River Delta and human health risk assessment. *Science of the Total Environment*, 647, 962-972.

Živković, M. M., Jovašević-Stojanović, M., Cvetković, A., Lazović, I., Tasić, V., Stevanović, Ž. and Gržetić, I. A. (2015). PAHs levels in gas and particle-bound phase in schools at different locations in Serbia. *Chemical Industry and Chemical Engineering Quarterly/CICEQ*, 21, 159-167.

Публикације категорије М20

Публикације након избора у претходно звање

Science of the Total Environment 929 (2024) 172195

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere

Nebojsa Bacanin [a,b], Mirjana Perisic [a,c], Gordana Jovanovic [a,c], Robertas Damaševičius [d,*], Svetlana Stanisic [a], Vladimir Simic [e,f,g], Miodrag Zivkovic [a], Andreja Stojic [a,b]

[a] Informatics and Computing, Singidunum University, Danijelova 32, Belgrade 11010, Serbia
[b] Sinergija University, Raje Banjicica, Bjeljina 76300, Bosnia and Herzegovina
[c] Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade 11010, Serbia
[d] Centre of Real Time Computer Systems, Kaunas University of Technology, Barsausko 59, Kaunas 51423, Lithuania
[e] Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, Belgrade 44249, Serbia
[f] Yuan Ze University, College of Engineering, Department of Industrial Engineering and Management, Taoyuan City 320315, Taiwan
[g] Department of Computer Science and Engineering, College of Informatics, Korea University, Seoul 02841, Republic of Korea

HIGHLIGHTS

- The study presents a novel approach to air quality research, utilizing advanced algorithms and AI techniques.
- It contributes to our understanding of toluene's environmental impact and interactions with other pollutants.
- The research demonstrates the potential of using enhanced metaheuristic algorithms in environmental studies.

GRAPHICAL ABSTRACT

ARTICLE INFO

Editor: Jianmin Chen

Keywords:
Explainable AI
Swarm intelligence
Metaheuristics

ABSTRACT

Toluene is a neurotoxic aromatic hydrocarbon and one of the major representatives of volatile organic compounds, known for its abundance, adverse health effects, and role in the formation of other atmospheric pollutants like ozone. This research introduces the enhanced version of the reptile search metaheuristics algorithm which has been utilized to tune the extreme gradient boosting hyperparameters, to investigate toluene atmospheric behavior patterns and interactions with other polluting species within defined environmental conditions. The study is based on a two-year database encompassing concentrations of inorganic gaseous contaminants every hour (NO, NO_2, NOx, and O_3), particulate matter fractions (PM_1, $PM_{2.5}$, and PM_{10}), m,p-xylene, toluene, benzene, total non-methane hydrocarbons, and meteorological data. The experimental outcomes were validated against the results of extreme gradient boosting models optimized by seven other recent powerful metaheuristics algorithms. The best-performing model has been interpreted by employing Shapley additive explanations

* Corresponding author.
E-mail addresses: nbacanin@singidunum.ac.rs (N. Bacanin), mirjana@ipb.ac.rs (M. Perisic), gordana.vukovic@ipb.ac.rs (G. Jovanovic), robertas.damasevicius@ktu.lt (R. Damaševičius), sstanisic@singidunum.ac.rs (S. Stanisic), vsima@sf.bg.ac.rs (V. Simic), mzivkovic@singidunum.ac.rs (M. Zivkovic), andreja.stojic@ipb.ac.rs (A. Stojic).

https://doi.org/10.1016/j.scitotenv.2024.172195
Received 2 December 2023; Received in revised form 1 April 2024; Accepted 1 April 2024
Available online 15 April 2024

method. In the study, we have focused on the relationship between toluene and benzene, as its most important predictor, and provided a detailed description of environmental conditions which directed their interactions.

1. Introduction

Our knowledge and understanding of air pollution-related processes, including dispersion, accumulation, deposition, or removal of polluting species are somewhat limited, because pollutant interactions, emission sources, measurement site characteristics, meteorological and other factors, all contribute to the complexity of air pollutant environmental fate (Shah et al., 2023). Therefore, data-driven research and conclusions are substantial for mitigating air pollution's deteriorating effects on human health and for guaranteeing future global environmental sustainability.

Toluene is an abundant, mono-substituted aromatic hydrocarbon and one of four representatives of organic substances that are volatile in the air. Together with benzene, toluene has been detected in the air of the distant arctic regions, yet its main sources are urban, i.e., traffic exhaust, cigarette smoke, and anthropogenic activities associated with fuel, paint, adhesive, cleaner, polish, rubber, and lacquer production and use. Previous studies have shown that the levels below 5 $\mu g\ m^{-3}$ were mostly observed in rural areas, while a wide range of toluene concentrations from 5 to 150 $\mu g\ m^{-3}$ have been measured in urban locations, with extreme values registered in the very vicinity of volatile pollutant sources (Murindababisha et al., 2021). As regards the adverse health effects, toluene is not prone to bioaccumulation due to the lack of persistent functional groups. When inhaled, it is rapidly absorbed, spread in the body, and accumulated mostly in vascularized organs, particularly within the brain, due to its affinity for lipid-rich tissues (Davidson et al., 2021). The trace levels in the blood are rapidly metabolized by the liver and eliminated in the form of hippuric acid, which can be detected and determined in urine (Clough, 2014). Furthermore, toluene is not labeled as carcinogenic, yet its concentrations are regularly monitored due to the lipoaffinity, toxic effects on the nervous system, and permanent brain damage observed in adhesive abusers. Namely, acute exposure to high toluene concentrations results in confusion, coordination, and cognitive decline, dizziness, headaches, and eventually coma and death, while repeated, often occupational exposure to low toluene concentrations in a long term, results in neurological manifestations, including ataxia, tremor, anosmia, hearing loss, dementia and epileptic seizures (Cheremisinoff and Rosenfeld, 2010), which are revealed in neuroimaging studies as diffuse atrophy and decreased gray–white differentiation (Eicher, 2009). The severe consequences including impaired mental growth and retarded development of newborns, as well as pregnancy loss are observed after pregnant women are exposed to elevated toluene concentrations.

The atmospheric half-life of toluene rarely exceeds two days, due to the rapid reaction with hydroxyl radicals. In urban areas, volatile aromatics significantly contribute to secondary organic aerosol (SOA) and ozone creation (Whitten et al., 2010). For instance, the amount of ozone being generated during and after the use of paint for road marking is estimated to be more than twice the amount of emitted VOCs (Burghardt et al., 2016). Furthermore, Skorokhod et al. (Skorokhod et al., 2017) have estimated that benzene and toluene contribute equally to photochemical ozone generation was only 16 % in rural areas, while in urban locations, the anthropogenic benzene and toluene emission sources play a substantial role in ozone photochemistry, with total contribution ranging from 60 to 70 %. Similarly, the SOA-forming contribution is estimated to be 36 and 30 % for benzene and toluene, respectively, mostly in low-NOx regimes (Wohl et al., 2023).

Our previous studies have addressed air pollution analysis in complex urban environmental conditions, some of which were related to adequate contextualization of data (Stojić et al., 2018; Šoštarić et al., 2017; Stanišić et al., 2022), and those related to the use of statistical methods and algorithms of artificial intelligence (AI) in this field (Andreja Stojić et al., 2019; Stanišić et al., 2021; Perišić et al., 2017; Andreja Stojić et al., 2015). As regards data modeling, NP-hard problems, especially concerning machine learning hyperparameter optimization, are commonly addressed by metaheuristics algorithms due to their stochastic nature.

In this study, we adopted the reptile search algorithm (RSA) (Abualigah et al., 2022), one of the most recent metaheuristics algorithm, determined empirically for modeling toluene concentrations. We propose an enhanced variant of RSA, being hybridized with the firefly algorithm (FA) for resolving the well-known shortcomings of the elementary RSA. The herein presented enhanced version of RSA metaheuristics was applied as an integral component of the AI framework, with the aim to optimize the set of the XGBoost hyper-parameters for toluene atmospheric fate research. Tuned XGBoost model by proposed improved RSA was compared to other models, optimized by other well-known heuristics and the best model was explained using Shapley Additive exPlanations (SHAP).

The main contributions of proposed study can be summarized as:

- devised improved RSA approach that overcomes cons of the original algorithm;
- tuned XGBoost model can obtain satisfactory performance for regression problems;
- environmental lifecycle of toluene was analyzed for the first time using XGBoost model, and non-linear relationships between toluene and other pollutants and meteorological factors were determined using explainable AI.

The novelty of this paper is the adoption of the improved RSA algorithm to investigate toluene atmospheric behavior patterns and interactions with other polluting species within defined environmental conditions.

Further parts of the article are structured in the following. Section 2 shows details related to algorithms and methods used in the proposed research, introduced improved RSA metaheuristics along with its basic implementation are presented in Section 3, while Sections 4 and 5 exhibit obtained experimental results, comparative analysis with other approaches and best model interpretation by using SHAP explainable AI technique, that provides insights into the influence of predictors (pollutants) to toluene behavior in the atmosphere. Section 6 finalizes this article.

2. Background and related works

This section introduces basic methods used in this research along with relevant literature review. The XGBoost model was described first, followed by basics of metaheuristics optimizers and SHAP method.

2.1. Metaheuristics optimization

Stochastic algorithms like metaheuristics are necessary for tackling NP-hard challenges, which are commonly encountered in computer science. Deterministic methods are often not feasible in such scenarios. Metaheuristics algorithms are frequently classified following natural phenomena they mimic to control the search procedure, like the social interactions of insects or evolution (Stegherr et al., 2020; Emmerich

et al., 2018; Fausto et al., 2020). The primary families of metaheuristics methods include nature-inspired methods (that can be further classified as swarm intelligence and evolutionary algorithms), physical phenomena-based methods (e.g. inspired by storms, gravitational field, or electromagnetic properties), human behavior-based methods (e.g. brainstorming, social networks activity), and mathematically guided approaches (e.g. inspirited by the main trigonometric functions' properties).

Swarm intelligence is focused on the coordinated behavior exhibited by big groups (swarms) of otherwise modest subjects such as insects or birds when they engage in activities like hunting, feeding, mating, or migrating (Beni, 2020; Abraham et al., 2006). These methods have proven to be highly effective in resolving real-world NP-hard challenges. Some famous examples of swarm intelligence metaheuristics methods include particle swarm optimization (PSO) (Kennedy and Eberhart, 1995), ant colony optimization (ACO) (Dorigo et al., 2006), bat algorithm (BA) (Yang, 2010; Yang and Gandomi, 2012) and firefly algorithm (FA) (Yang, 2009). In recent years, a new class of highly efficient algorithms has emerged that utilize algebraic functions and their attributes to guide the search process, such as arithmetic optimization algorithm (AOA) (Abualigah et al., 2021).

Such a diversity of population-based techniques is explained by no-free-lunch theorem (NFL) (Wolpert and Macready, 1997), which claims that no single method can be universally optimal for every single optimization problem that exists. As a result, an algorithm that performs well on one particular problem may not be effective for another, necessitating a range of metaheuristics methods and the requirement to carefully select the appropriate approach for every individual optimization problem.

Population methods have become a popular option for tackling multiple real-world problems in recent times. Successful applications include green energy production forecasting (Pavlov-Kagadejev et al., 2024; Damaševičius et al., 2024), wireless sensors and IoT applications (Alharbi et al., 2021; Dobrojevic et al., 2023), feature selection (Bacanin et al., 2023b; Helmi et al., 2021), medical images classification (Zivkovic et al., 2022a), software defect prediction (Khurma et al., 2021), air pollution monitoring (Bacanin et al., 2022a; Jovanovic et al., 2023b), machine scheduling (Jouhari et al., 2020; Makhadmeh et al., 2021), intrusion detection in computer networks (Bacanin et al., 2022d; Alzaqebah et al., 2022; Salb et al., 2023), and lastly, the optimization of the wide range of ML structures (Bacanin et al., 2022c; Bacanin et al., 2022b; Jovanovic et al., 2022; Bukumira et al., 2022; Bacanin et al., 2023a; Minic et al., 2023; Todorovic et al., 2023; Jovanovic et al., 2023a; Cuk et al., 2024).

2.2. Shapley additive explanations

To gain an understanding of the modeling process and the best-performing model, the explainable artificial intelligence technique SHAP was utilized. This method avoids the typical trade-off betwixt accuracy and interpretability by providing a clear and meaningful explanation of the model's decisions. The SHAP method calculates feature importance through Shapley values, which are derived from a gaming theory that measures the impact of individual features on the output (Lundberg and Lee, 2017). These values apportion the difference between predictions and the average predictions between the features and denote a fair distribution of payouts across cooperating features based on their individual contributions to the joint payout. It also captures interactions among features by generalizing Shapley values and allows interpretation of the global behavior while maintaining local fidelity (Lundberg et al., 2020; Andreja Stojić et al., 2019).

Aiming to provide insight into the relative relationships between features' attributions for every prediction, the relative SHAP values have been utilized in this research, which were proposed by (Andreja Stojić et al., 2022). These values represent the portion of absolute SHAP in the total attributed significance of all features for a single forecast, indicating the relative influence every feature has on each prediction. To achieve this, Python SHAP implementation (SHAP Python package) and TreeExplainer (Lundberg et al., 2020) were used to derive SHAP values, that were later utilized to generate the SHAP dependency plots. These plots represent the shift of the feature's importance across its range of values (Stoean et al., 2023).

3. Methods

This sections first exhibits the inner functional principles of the basic RSA, followed by its observed drawbacks and details of improved approach devised for the purpose of this research study.

3.1. Reptile search algorithm

Reptile search algorithm (RSA) (Abualigah et al., 2022) was inspired by the collaborative and highly coordinated hunting style of crocodiles. The algorithm was chosen for this research through empirical trials along with several other metaheuristics, as it exhibited very promising outcomes on the observed dataset during short simulations that were executed prior to the main experiments. It is one of the latest algorithms, attractive due to its excellent exploration capabilities in case of high-dimensional search spaces. Moreover, it was recently used to solve different complex challenges, including software defect prediction Zivkovic et al., 2023, data clustering (Almotairi and Abualigah, 2022a), intrusion detection (Dahou et al., 2022), feature selection (Abualigah and Diabat, 2023), image classification (Abualigah et al., 2023), and general optimization problems (Almotairi and Abualigah, 2022b), to name the few.

This technique involves crocodiles surrounding the prey before attacking. The RSA's initialization phase involves creating a matrix X of random individuals $x_{i,j}$ based on Eq. (1). In this equation, i refers to the individual's index, j denotes their current location, N is the total solution number, and n is the specific problem's dimensionality.

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & x_{1,n-1} & x_{1,n} \\ x_{2,1} & \cdots & x_{2,j} & \cdots & x_{2,n} \\ \cdots & \cdots & x_{i,j} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-1,1} & \cdots & x_{N-1,j} & \cdots & x_{N-1,n} \\ x_{N,1} & \cdots & x_{N,j} & x_{N,n-1} & x_{N,n} \end{bmatrix} \tag{1}$$

Eq. (2) has the role to generate random individuals, where *rand* represents a random number inside $[0, 1]$, and *LB* and *UB* mark lower and upper boundaries of the respective searching realm for the specific task (Abualigah et al., 2022).

$$x_{ij} = rand \times (UB - LB) + LB, j = 1, 2, \ldots, n \tag{2}$$

To promote both exploration and exploitation, the search stage was divided into two key functions, namely targeting and hunting, which are complemented by four distinct behaviors. Crocodiles exhibit two types of walking to facilitate exploration: the elevated walk and the stomach walk. The primary objective during this phase is to expand the search area and prepare for the 2nd hunting stage. Elevated walk is utilized when $t \leq \frac{T}{4}$, whereas stomach walk is employed when $t > \frac{T}{4}$ and $t \leq 2\frac{T}{4}$. Update of each crocodile's location is managed by Eq. (3) (Abualigah et al., 2022):

$$x_{(i,j)}(t+1) = \begin{cases} \text{Best}_j(t) \times -\eta_{(i,j)}(t) \times \beta - R_{(i,j)}(t) \times \text{r and}, t \leq \dfrac{T}{4} \\ \text{Best}_j(t) \times x_{(r_1,j)} \times ES(t) \times \text{r and}, t > \dfrac{T}{4} \text{ and } t \leq 2\dfrac{T}{4} \end{cases} \tag{3}$$

$$\eta_{(i,j)} = \mathrm{Best}_j(t) \times P_{(i,j)} \tag{4}$$

The latest best individual at location j is denoted by $Best_j$, while t represents the current iteration, and T represents the largest number of iterations. Eq. (4) defines hunting operator $\eta_{(i,j)}$, with variable β set at 0.1 to control exploring precision (Abualigah et al., 2022).

To reduce searching space, the reduction function described in Eq. (5) is applied. Here, r_1 is a random value within the range $[1, N]$, $x_{r1,j}$ represents the solution's random location for the i_{th} individual, and ε represents a negligible value.

$$R_{(i,j)} = \frac{\mathrm{Best}_j(t) - x_{(r_1,j)}}{\mathrm{Best}_j(t) + \varepsilon} \tag{5}$$

Eq. (6) is used to obtain the probability rate, labeled as "Evolutionary Sense", and utilized to randomly alternate from −2 to 2 over the iterations, as specified by (Abualigah et al., 2022):

$$ES(t) = 2 \times r_2 \times \left(1 - \frac{1}{T}\right) \tag{6}$$

where r_2 is a random number drawn from the $[-1, 1]$ interval.

Eq. (7) has the role to establish the difference (in percent's) among the locations of the current and latest best solution (Abualigah et al., 2022):

$$P_{(i,j)} = \alpha + \frac{x_{(i,j)} - M(x_i)}{\mathrm{Best}_j(t) \times \left(UB_{(j)} - LB_{(j)}\right) + \varepsilon} \tag{7}$$

where α represents the sensitive parameter, that has been preset to 0.1, steering the fluctuations between potential solutions that are appropriate to execute the cooperated hunt. The corresponding upper and lower bounds of the position j_{th} are given by $UB_{(j)}$ and $LB_{(j)}$.

Average location $M(x_i)$ belonging to the ith solution can be determined by applying Eq. (8).

$$M(x_i) = \frac{1}{n}\sum_{j=1}^{n} x_{(i,j)} \tag{8}$$

RSA's exploitation mechanism consists of two methods: hunting coordination (for $t \le 3\frac{T}{4}$ and $t > \frac{T}{2}$) and cooperation (for $t \le T$ and $t > 3\frac{T}{4}$). The objective is to enhance local exploration of the search space and converge towards the best individual. Eq. (9) summarizes the hunting behavior demonstrated by crocodiles.

$$x_{(i,j)}(t+1) = \begin{cases} \mathrm{Best}_j(t) \times P_{(i,j)}(t) \times \mathrm{r} \text{ and}, t \le 3\dfrac{T}{4} \text{ and } t > \dfrac{T}{2} \\ \mathrm{Best}_j(t) - \eta_{(i,j)}(t) \times \varepsilon - R_{(i,j)}(t) \times rand, \ t \le T \text{ and } t > 3\dfrac{T}{4} \end{cases} \tag{9}$$

3.2. Enhanced RSA algorithm

RSA is a new metaheuristic and has demonstrated its effectiveness as a powerful optimizer in its basic version (Abualigah et al., 2022). However, there have been some noted drawbacks. Intensive experiments using benchmark functions (including bound-constrained and constrained problems from CEC dataset) have revealed that the RSA lacks sufficient exploitation power, particularly in the later iterations of execution, despite owning excellent exploration behavior.

Extensive simulations with the original RSA have further revealed that while the algorithm's search process is able to identify a suitable domain in most runs, it lacks the necessary raw exploitation power to conduct a precise search inside that region. Consequently, the original RSA may require more iterations to converge. This problem can be regarded in terms of the diversification-intensification trade-off, that is a common issue with metaheuristics algorithms (Bacanin et al., 2020; Bacanin et al., 2022b). In the case of basic RSA, it is observed that the trade-off has a bias directed to exploration.

However, the firefly algorithm (FA) is well known for its remarkable exploitation capabilities, as discussed in (Yang, 2009). Consequently, a plausible strategy to enhance the RSA's performance is to integrate it with the FA in the form of the low-level hybrid, since both methods have complementary strengths. However, when constructing hybrid metaheuristics, one must be cautious in achieving the appropriate trade-off of exploration vs. exploitation. When the run starts, the main objective is to identify the optimal search region, which requires more emphasis on exploration. As iterations pass by, a more focused search in the proximity of the latest best individual is required, necessitating stronger exploitation. If the FA search strategy is initiated at the beginning of the execution, some runs may fail to find the optimal region, thus the introduced low-level hybrid approach addresses this by incorporating hard-coded control parameters.

The hybrid technique suggested in this article fuses the RSA and FA metaheuristics to address the RSA's weaknesses. At the start of the algorithm, solutions are being updated using the RSA search Eq. (3). However, as the search progresses and a promising area is identified, the FA search procedure (10) is employed to improve exploitation. The FA is known for its potent exploitation due to its randomization and Gaussian distribution parameters (α and κ, respectively), which are used to update individuals based on their distance from one another ($r_{i,j}$) (Yang, 2009).

$$X_i^{t+1} = X_i^t + \beta_0 \cdot e^{-\gamma r_{ij}^2}\left(X_j^t - X_i^t\right) + \alpha^t(\kappa - 0.5) \tag{10}$$

Algorithm's transition between RSA and FA search mechanisms in later phase is regulated by two extra parameters. First parameter, called the varying search (vs), activates the hybrid search procedure when $t > vs$. The parameter has a set value of $maxIter/5$, and in the conducted simulations, it was set to 4 since there are twenty rounds. The appropriate value of vs was determined experimentally.

Second parameter introduced in the proposed hybrid metaheuristic is called "search mode" (sm). It is used to decide whether each individual in the populace should use the RSA or FA mechanism for the search. Each solution generates a stochastic number rnd between 0 and 1, and if $rnd < sm$, the RSA search is performed, otherwise, the FA search is executed. The value of sm is reduced as the algorithm progresses through the rounds, giving more emphasis to the more powerful FA search after the process has converged to promising areas of the search domain. The initial value of sm is preset to 0.8, and it is decreasing with respect to Eq. (11), that was determined empirically.

$$sm_t = sm_{t-1} - (sm_{t-1}/10) \tag{11}$$

A supplementary mechanism utilized in this study was inspirited by the *trial* parameter used by ABC algorithm (Karaboga and Basturk, 2008). Initially, this parameter is preset to 0, and if no improvement in the solution is observed in this iteration, *trial* is incremented. Once *trial* hits the threshold $limit = 1$ (determined empirically as $2^*(solutions/maxIter)$ for this problem), this individual is taken out from the populace and replaced with another solution x_{qrl} generated using the quasi-reflection-based learning (QRL) approach (Rahnamayan et al., 2007). Previous studies has shown that QRL is efficient in generating solutions in the opposite region of the search area (Bacanin et al., 2021).

Algorithm 1. Pseudo-code of the proposed HRSA algorithm.

Algorithm 1 Pseudo-code of the proposed HRSA algorithm

▷ **Initialization procedure**
Configure RSA parameters α, β, etc.
Configure $vs = maxIter/5$
Set $trial = 0$ for every individual
Set the random positions for the solutions. $X : i = 1, ..., N$.
while $(t < T)$ **do**
  Ascertain the fitness function for the candidate solutions (X)
  Ascertain the best solution until now
  Update the ES by applying Eq. (6)
▷ **The start of the RSA**
  for $(i$=1 to $N)$ **do**
    for $(j$=1 to $n)$ **do**
      if $t > vs$ **then**
        Generate arbitrary number rnd
      end if
      if $(t > vs)$ & $!(rnd < sm)$ **then**
        Proceed with FA search by applying Eq. (10)
      else
        Continue with RSA search
        Update the η, R, P and values by applying Eqs. (4), (5) and (7), respectively
        if $(t \leq \frac{T}{4})$ **then**
          Execute high walking by applying top case of Eq. (3)
        else if $(t \leq 2\frac{T}{4}$ *and* $t > \frac{T}{4})$ **then**
          Execute belly walking by applying bottom case of Eq. (3)
        else if $(t \leq 3\frac{T}{4}$ *and* $t > 2\frac{T}{4})$ **then**
          Execute hunting coordination by applying top case of Eq. (9)
        else
          Execute hunting cooperation by applying bottom case of Eq. (9)
        end if
      end if
    end for
    if individual was not enhanced **then**
      $trial = trial + 1$
      if $trial = limit$ **then**
        Replace the observed individual with quasi-reflective opposite one x_{qrl}
        $trial = 0$
      end if
    end if
  end for
  $t = t + 1$
end while
Return the most fit individual $(Best(X))$

The suggested hybrid was plainly named hybrid RSA - HRSA, while its pseudo-code is given by the Algorithm 1.

3.3. The XGBoost model

The XGBoost model is based on an adaptive training approach to tune its fitness function, and every step in the process depends on the outcome of the previous step (Chen et al., 2015; Chen and Guestrin, 2016). XGBoost model's fitness function can be mathematically represented by Eq. (12).

$$F_o^{\,i} = \sum_{k=1}^{n} l\left(y_k, \widehat{y}_k^{\,i-1} + f_i(x_k)\right) + R(f_i) + C, \tag{12}$$

where y denotes the input, $\widehat{y}$ represent the residual errors, the loss of the t-th iteration is given as l, C denotes the constant, and the XGBoost regularization parameter R is calculated as:

$$R(f_i) = \gamma T_i + \frac{\lambda}{2}\sum_{j=1}^{T} w_j^2 \tag{13}$$

The tree structure's complexity typically decreases as the customization parameters γ and λ increase. This means that higher values of these parameters lead to a simpler tree. The expressions for the model's 1st and 2nd derivatives are denoted by g and h respectively, are given with Eq. (14 and Eq. (15):

$$g_j = \partial_{\widehat{y}_k^{\,i-1}} l\left(y_j, \widehat{y}_k^{\,i-1}\right) \tag{14}$$

$$h_j = \partial^2_{\widehat{y}_k^{\,i-1}} l\left(y_j, \widehat{y}_k^{\,i-1}\right) \tag{15}$$

The solution can be determined by applying the following two expressions:

$$w_j^* = -\frac{\sum g_t}{\sum h_t + \lambda} \tag{16}$$

$$F_o^{\,*} = -\frac{1}{2}\sum_{j=1}^{T}\frac{(\sum g)^2}{\sum h + \lambda} + \gamma T, \tag{17}$$

where the merit of the loss function is represented by $F_o^{\,*}$, while the w_j^* denotes the weights of the solution.

4. Simulation results and comparative analysis

In this section, we prove whether or not performance improvements of proposed approach over other methods is statistically significant, this section concludes with the conducted statistical tests.

4.1. Employed dataset and evaluation metrics

This study obtained data on various air pollutants and atmospheric parameters from the air quality observation station Vatrogasni Dom in Pančevo, Serbia. The air pollutants measured included inorganic gaseous pollutants such as NO, NO_2, NOx, and O_3, as well as particulate matter like PM_1, $PM_{2.5}$, and PM_{10}, and certain hydrocarbons. The researchers collected data for two years (2019–2020) and gathered a total of 11,368 hourly pollutant concentration measurements. Additionally, atmospheric parameters were retrieved from the Global Data Assimilation System (GDAS1). The same dataset was used in our previous research, where only basic metaheuristics were evaluated for VOC's environmental fate, therefore for more details regarding the data, please refer to (Jovanovic et al., 2023b).

Aiming to evaluate the results attained by the proposed XGBoost model, a collection of standard ML benchmark criteria has been applied. The observed criteria included mean squared error (MSE) calculated by Eq. (18), root mean squared error (RMSE) according to Eq. (19), mean absolute error (MAE) defined as Eq. (21), and the coefficient of determination (R^2) that is obtainable by applying Eq. (21).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{z}_i - z_i)^2 \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{z}_i - z_i)^2} \tag{19}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{z}_i - z_i| \tag{20}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(z_i - \hat{z}_i)^2}{\sum_{i=1}^{n}(z_i - \bar{z})^2}, \tag{21}$$

where z_i denotes the array of observed actual values being predicted, while $\hat{z}_i$ represents the array of predictions, both arrays being comprised of N elements. MSE indicator has been set as the objective function in this research, with a goal to minimize it.

4.2. Settings

The HRSA algorithm was assigned to tune the XGBoost structure's level of performance on this peculiar dataset. The XGBoost is well-known for its efficiency and speed in different classification tasks, however, the performance directly depends on the appropriate settings of its hyperparameters. According to the previous research, and thorough empirical tests conducted by the authors of this study, the most influential hyperparameters have been identified (Probst et al., 2019; Budholiya et al., 2022; Zivkovic et al., 2022b). The XGBoost model's array of hyperparameters that were tuned in this research, along with their associated search confines and variable types have been arranged in the following way as presented in (Jovanovic et al., 2023c):

- *gamma*, range: $[0, 0.8]$,
- max_depth, range: $[3, 10]$,
- collsample_bytree, range: $[0.01, 1]$,
- subsample, range: $[0.01, 1]$,
- *min_child_weight*, range: $[1, 10]$,
- learning rate (η), range: $[0.1, 0.9]$,

Finally, other XGBoost parameters were calibrated to the default values of the XGBoost throughout the conducted simulations.

The proposed framework was developed in Python and common set of machine learning libraries, namely numpy, scipy, and pandas, whereas the XGBoost model itself was derived from the scikit-learn library.

The metaheuristics solutions were encoded as the array of size l, representing the count of tuned hyperparameters. Consequently, as a total of six hyperparameters were optimized in the conducted experiments, l was set to 6.

The simulation outcomes of the XGBoost optimized by the proposed HRSA method were validated against the scores achieved by seven modern algorithms, namely elementary RSA, PSO (Kennedy and Eberhart, 1995), harris hawks' optimization (HHO) (Heidari et al., 2019), ABC (Karaboga and Basturk, 2008), FA (Yang, 2009), whale optimization algorithm (WOA) (Mirjalili and Lewis, 2016), and chimp optimization algorithm (ChOA) (Khishe and Mosavi, 2020). All metaheuristics that were included in the comparisons have been implemented separately by the authors for the sake of this research, by using a recommended set of control parameters' values as described by their corresponding authors. Every contender was assigned the identical task, to optimize the very same XGBoost hyperparameters as the proposed HRSA.

All metaheuristics methods have been evaluated with 10 individuals in the population and 20 iterations in each execution, over the course of 8 independent executions. The fitness function for the observed problem was MSE, and all algorithms were aiming to minimize it.

4.3. Results and comparative analysis

Simulation outcomes of the proposed HRSA method and other metaheuristics are presented in this section. Tables 1 and 2 bring forward

Table 2
Results for the best run of each algorithm.

	R^2	R	MSE	MAE	RMSE
XG-HRSA	**0.918963**	**0.958626**	**0.636515**	**0.448479**	**0.797819**
XG-RSA	0.916737	0.957464	0.654000	0.455532	0.808703
XG-PSO	0.914975	0.956543	0.667839	0.467261	0.817214
XG-ABC	0.917826	0.958032	0.645446	0.455709	0.803397
XG-FA	0.918430	0.958348	0.640698	0.450520	0.800436
XG-HHO	0.917319	0.957768	0.649428	0.459348	0.805871
XG-WOA	0.910043	0.953962	0.706577	0.478858	0.840581
XG-ChOA	0.917592	0.957910	0.647280	0.460521	0.804537

Table 1
MSE results of the metaheuristics algorithms.

Method	XG-HRSA	XG-RSA	XG-PSO	XG-ABC	XG-FA	XG-HHO	XG-WOA	XG-ChOA
Best	**0.636515**	0.654000	0.667839	0.645446	0.640698	0.649428	0.706577	0.647280
Worst	**0.651034**	0.848581	0.735295	0.727634	0.651614	0.794969	0.778161	0.702748
Mean	**0.642792**	0.712121	0.697220	0.699582	0.646773	0.693997	0.739410	0.667459
Median	**0.640927**	0.699782	0.684052	0.707975	0.646327	0.664924	0.735603	0.662052
Std	0.005242	0.059714	0.028533	0.025374	**0.004080**	0.052107	0.027783	0.016544
Var	0.000027	0.003566	0.000814	0.000644	**0.000017**	0.002715	0.000772	0.000274

Table 3
Best solutions' found with XGBoost hyperparameters.

Method	l.r. (μ)	max_child_weight	subsample	collsample_bytree	max_depth	gamma
XG-HRSA	0.336192	1.000000	1.000000	0.994616	10	0.670826
XG-RSA	0.346958	1.913642	1.000000	1.000000	10	0.000000
XG-PSO	0.395272	4.466178	0.996013	1.000000	10	0.269886
XG-ABC	0.349842	7.069863	1.000000	0.992345	9	0.323110
XG-FA	0.347431	2.239742	1.000000	0.972526	10	0.077855
XG-HHO	0.359068	5.817867	1.000000	1.000000	10	0.363443
XG-WOA	0.298181	4.665303	0.851359	0.867922	10	0.160447
XG-ChOA	0.347070	5.243854	0.940824	0.949336	10	0.381203

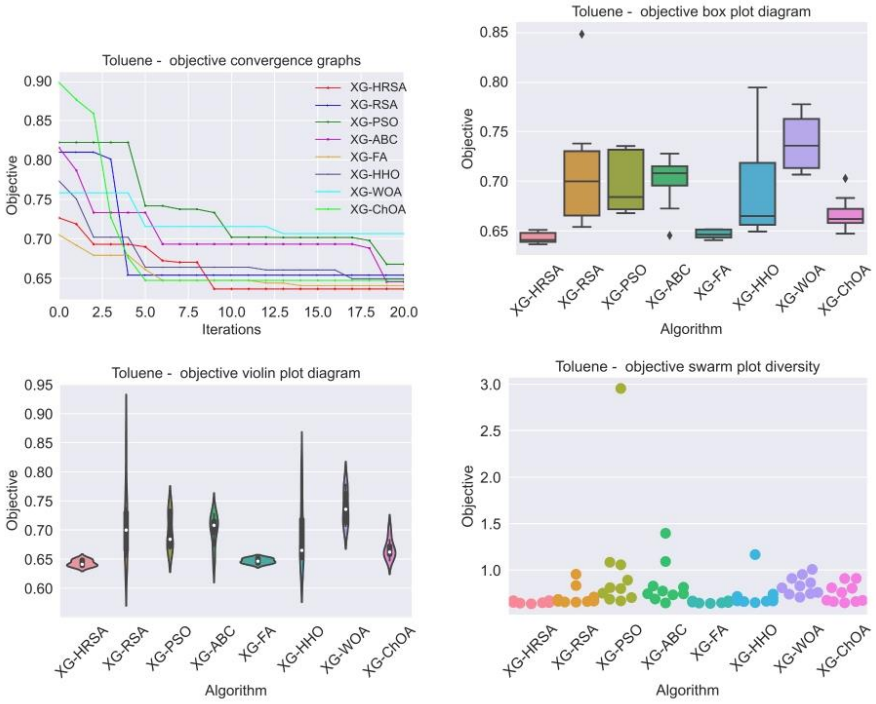


Fig. 1. Visualizations of the XGBoost experiments for all eight observed metaheuristics for the fitness function (MSE) on the Toluene dataset.

the experimental results, with addition of the detailed metrics attained by the each algorithm's best individual run, where the best scores are highlighted by bold text.

Table 1 shows comparative analysis related to the fitness function (MSE) obtained by the XGBoost model tuned by each of the 8 algorithms (the suggested HRSA and seven contending algorithms). The experimental outcomes strongly indicate that the HRSA algorithm exhibited a superior performance, as it achieved the best results for the most important performance metrics (mean and median, worst and best values). FA achieved the best scores with respect to the standard deviation and variance, as it delivered the steadiest results. FA was the second-best algorithm considering the best, worst, mean and median run values as well, followed closely by ChOA on the third place. The best achieved results by the proposed HRSA XGBoost model were the MSE of 0.636515, and R^2 of 0.918963.

Table 2 brings forward the detailed metrics scored by the best single run of each metaheuristics. Again, it must be noted that the suggested HRSA distinctively outperformed other algorithms with respect to all monitored indicators - R, R^2, MAE MSE, and RMSE. Taking into the consideration that the MSE was the objective function, HRSA achieved the score of 0.636515, followed by FA on the second place with the score of 0.640698 and ABC that finished third with the MSE of 0.645446.

Finally, the collections of the XGBoost model's hyperparameters discovered in the best run of each regarded metaheuristics are presented in the Table 3. The model with the best level of performance, namely the suggested HRSA, generated the XGBoost structure that has max_child_-weight of 1, max_depth of 10, a learning rate of 0.336192, a subsample of 1, collsample_bytree of 0.994616, and gamma value of 0.670826. The second best XGBoost model was generated by the FA algorithm, had the established learning rate of 0.347431, max_child_weight of 2.239742, a subsample score of 1, collsample_bytree of 0.972526, max_depth value of 10, and lastly, gamma of 0.077855.

The experimental outcomes are visually presented in Figs. 1 and 2 for all eight observed metaheuristics, considering both objective functions - MSE (Fig. 1) and R^2 (Fig. 2).

Closer inspection of the Figs. 1 and 2 shows that suggested HRSA

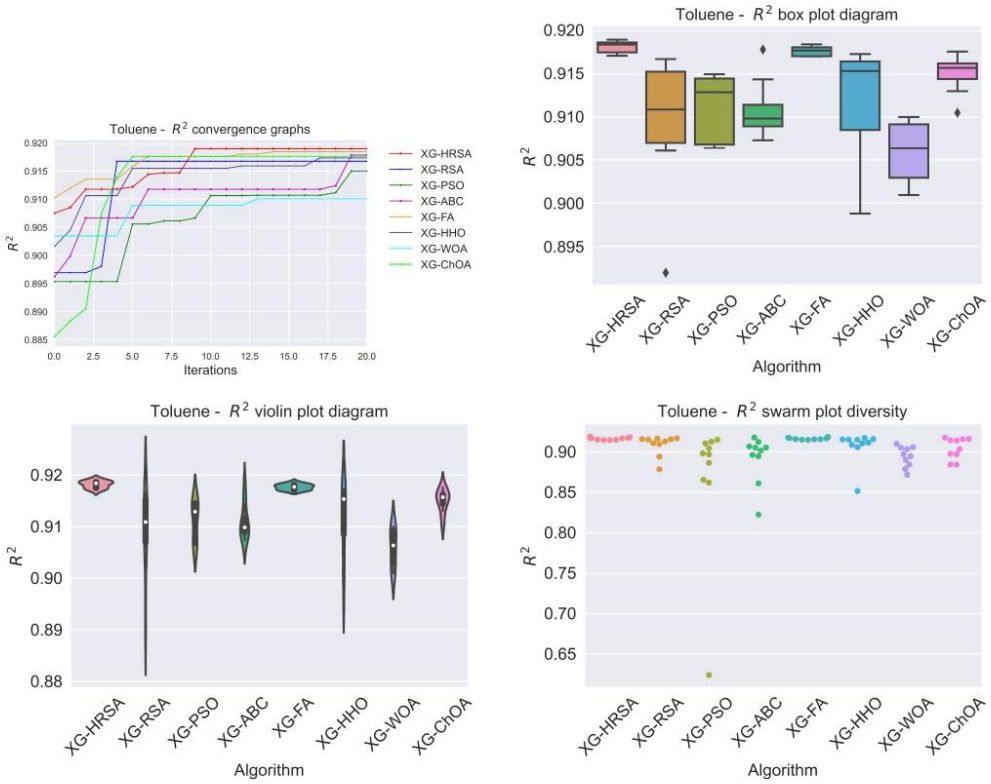


Fig. 2. Visualizations of the XGBoost experiments for all eight observed metaheuristics for the R^2 indicator on the Toluene dataset.

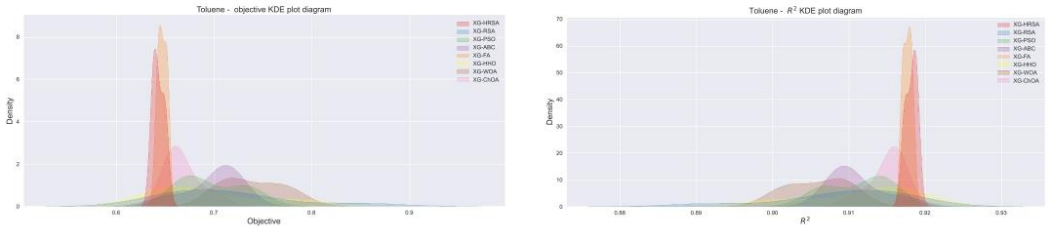


Fig. 3. Kernel density estimation (KDE) diagrams of objective function - MSE (left) and R^2 indicator (right) on the Toluene dataset.

algorithm demonstrates fast convergence, accompanied by the FA algorithm, which is faster early on, but nevertheless finishes behind HRSA. The box plots visually show that the FA provides the steadiest results, followed by HRSA and ChOA algorithms. The swarm plots provide the information about the diversity of the solutions during the terminal iteration of the best runs of all eight algorithms. It visually shows that every solution of the HRSA populace was positioned closely to the optimum value.

Fig. 3 brings forward the kernel density estimations (KDE) that provide the estimate of the probability density functions for the MSE and R^2. By observing these KDE diagrams, it can be concluded that the experimental results are not belonging to the normal distribution. As a supplement, the join plots of the objective function - MSE and R^2 of two best performing algorithms (HRSA and FA) are provided in Fig. 4.

4.4. Statistical tests

To confirm the experimental outcomes and verify their statistical significance, the top results from each of the eight independent runs of the algorithms were gathered and analyzed as data series. However, before selecting the appropriate statistical test, it is necessary to determine whether a parametric or non-parametric test is suitable. It is important to assess normality, homoscedasticity and independence of the data variances before deciding to use non-parametric tests (LaTorre

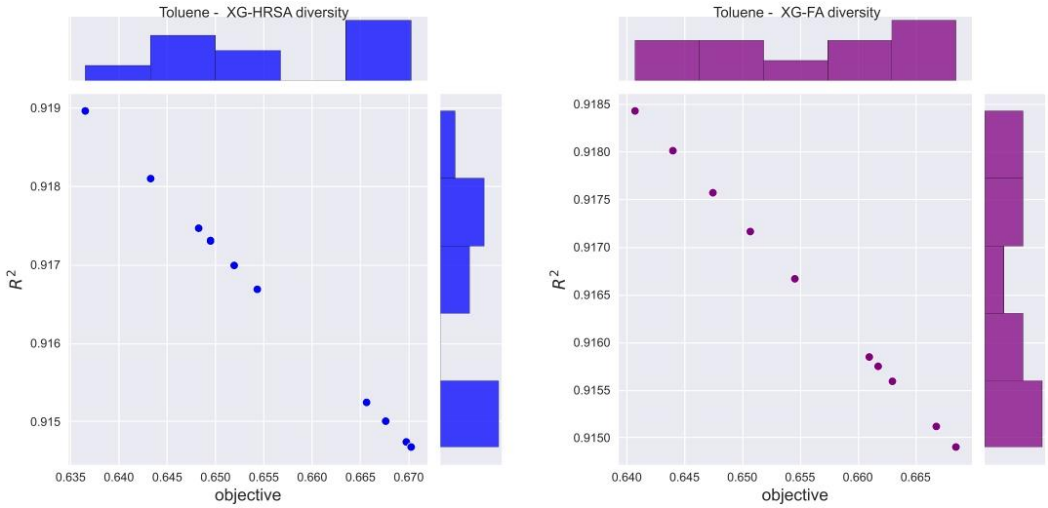


Fig. 4. Comparison of two best performing algorithms: HRSA (left) and FA (right) for the Toluene dataset.

Table 4
Shapiro-Wilk scores on Toluene problem used to check the normality requirement.

Methods	HRSA	RSA	PSO	ABC	FA	HHO	WOA	ChOA
Toluene	0.037	0.028	0.019	0.032	0.039	0.026	0.027	0.034

Table 5
$p-values$ calculated by the Wilcoxon signed-rank test on Toluene problem.

Methods	HRSA	RSA	PSO	ABC	FA	HHO	WOA	ChOA
Toluene	N/A	0.027	0.021	0.044	0.059	0.036	0.012	0.039

Table 6
SHAP values.

SHAP	Benzene	m, p-Xylene	TNMHC	PM_1	NOx	SOLM	MOFD	MOFI	NO_2	NO	PM_{10}	T02M
Absolute	1.28	0.85	0.25	0.11	0.09	0.07	0.05	0.04	0.03	0.03	0.03	0.03
Relative [%]	36.1	27.3	8.3	3.1	3.1	2.4	2.0	1.5	1.1	1.2	0.9	0.9

et al., 2021). The data meets the first requirement for independence as each metaheuristics execution begins by generating a collection of pseudo-random variables. The second condition, however, is not fulfilled according to the KDE diagrams presented in Fig. 3.

To confirm that the normality condition is not satisfied, the Shapiro-Wilk test was performed separately for each algorithm (Shapiro and Francia, 1972). The resulting p-values were compared against significance levels of 0.05 and 0.1, and it was found that H0 hypothesis can be rejected, as all p-values were below $alpha = 0.05$. Consequently, it was concluded that the observed results are not normally distributed. As already mentioned, this conclusion was also supported by examining the KDE diagrams in Fig. 3. The Shapiro-Wilk test results are summarized in Table 4.

As the Shapiro-Wilk test indicated that the normality assumption was not satisfied, it was deemed inappropriate to use parametric tests. Consequently, the Wilcoxon signed-rank test (Wilcoxon, 1992), was employed using the same data series composed of the top results for each run. The HRSA algorithm was chosen as the baseline algorithm. The results obtained from this analysis are presented in Table 5.

For the Toluene experiment, the obtained $p-values$ are in all cases below the threshold value 0.1 ($p-values$ were 0.027 vs RSA, 0.021 vs PSO, 0.044 vs ABC, 0.059 vs FA, 0.036 vs HHO, 0.012 vs WOA, and 0.039 vs ChOA). In conclusion, the proposed HRSA demonstrated statistical superiority over all other algorithms for the threshold limit of $alpha = 0.1$. In the case of threshold $alpha = 0.05$, suggested HRSA was superior to all other methods except FA (0.059 vs FA).

5. Best model interpretation and discussion

Since the XGBoost-HRSA obtained the best performance metrics, this model was taken for further interpretation by SHAP technique. This section puts special focus on benzene, as the most important predictor of toluene concentration.

5.1. Variable importance

As shown by mean absolute and relative SHAP values, the concentrations of benzene followed by m,p-xylene levels appeared to be the major factor that shapes the toluene dynamic in the air (Table 6). In addition, the concentrations of THNMC, PM_1, and NOx along with

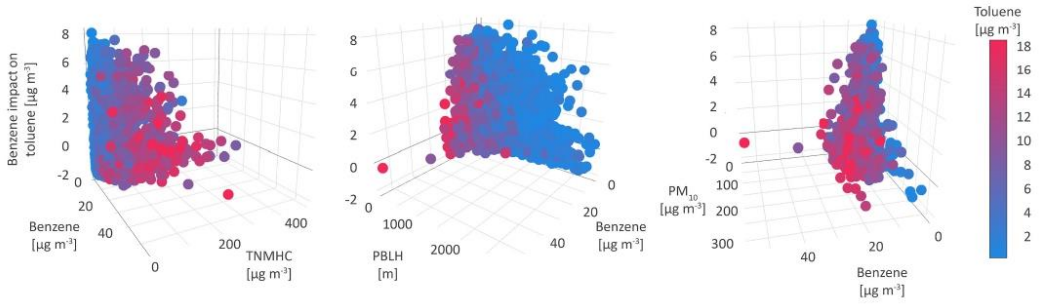


Fig. 5. Benzene impact on toluene.

meteorological parameters including volumetric soil moisture content (SOLM), direction, and intensity of momentum flux (MOFD and MOFI) affect toluene environmental fate. For this study, to demonstrate the potential of the applied methodology, we will focus on benzene as the most important predictor.

5.2. Benzene impact on toluene

The results suggest that benzene concentrations govern the toluene dynamics in the air more than any other pollutant or meteorological parameter, up to 1.28 $\mu g\ m^{-3}$ on average as shown by absolute SHAP values (Table 6). These findings are in compliance with our previous study identifying the strong interdependence between benzene and toluene (Jovanovic et al., 2023b). Benzene and toluene are emitted into the atmosphere from numerous anthropogenic mobile and stationary sources while traffic-related sources are believed to contribute to 60–85 % of atmospheric benzene (Hazrati et al., 2016). These organics have industrial applications as solvents or petroleum derivatives, and both of them are emitted during fossil fuel combustion, evaporative leaks, tank refill, and similar petroleum-related operations in the South industrial area of Pančevo in the vicinity of the monitoring station. After being released into the atmosphere, both benzene and toluene undergo complex chemical homogeneous and heterogeneous reactions in the air, mainly phototransformations characterized by several processes including oxidative addition of polar functional groups to the carbon skeleton–functionalization, the oxidative break of single bonds between carbon atoms–fragmentation, or the association of organic molecules–oligomerization (Kroll et al., 2011).

Three types of the environment were distinguished depending on the conditions and the way in which toluene and benzene were interrelated, as well as on their coexistence and interactions with other pollutants (Fig. 5). In all types of environments, the highest levels of polluted species are accompanied by low air and soil temperatures, high relative humidity, low planetary boundary layer (PBLH), and stable atmospheric conditions as indicated by the other meteorological parameters. The increase in benzene levels linearly correlated with the rise of PM_1 concentrations whereas a similar trend could not be attributed to the relations between benzene and PM_{10}, NOx, and THMNC. High levels of NOx and TNMHC are recorded when lower levels of benzene were obtained, which points to the influence of different emission sources in the surroundings of the measuring site as well as the enhanced capacity of finer particles for the adsorption of aromatic compounds. In addition, several events are characterized by an extremely positive influence of benzene on toluene patterns under low concentrations of both compounds. With the increase of pollutant concentrations, the type of relation changes, and the influence of benzene decreases (Fig. 5).

The first type of environment, characterized exclusively by the positive impact of benzene (average absolute SHAP 3.5 $\mu g\ m^{-3}$), increases toluene concentrations from 1.1 to 8.1 $\mu g\ m^{-3}$, while average SHAP relative to the other parameters was 54.2 %. When compared with the other two types of environmental conditions, this environment is also marked by low mean levels of m,p-Xylene, NO, NO_2, NOx, and O_3 (5.6, 18.9, 16.5, 45.6, and 49.1 $\mu g\ m^{-3}$, respectively), and medium mean concentrations of PM_1, $PM_{2.5}$, PM_{10}, and TNMHC (21.0, 23.1, 45.5, and 27.9 $\mu g\ m^{-3}$, respectively). Apart from benzene, the following pollutants: m,p-xylene (absolute SHAP ranges from −2.7 to 5.4 $\mu g\ m^{-3}$), PM_1 (absolute SHAP ranges from: −1.2 to 1.4 $\mu g\ m^{-3}$.0) and TNMHC (absolute SHAP ranges from −0.6 to 1.6 $\mu g\ m^{-3}$) also affect the dynamics of toluene in this environment. The apportioned meteorological conditions included slightly lower temperatures T02M and TMPS (average values 14.2 and 14.3 °C, respectively), lower value of PBLH (454.6 m on average), and slightly higher amount of precipitation, which is indicated based on the values of TPP6 and CPP6. The results suggest the influence of local petroleum-related anthropogenic emissions and the occasional result of fossil fuel burning emissions associated with cold weather. In addition, during the colder months of the year, radical-initiated photochemical removal of vapor-phase toluene and benzene is limited, which is reflected in the prolonged lifetime of pollutants in the air, up to several weeks.

Out of other meteorological parameters, low cloud cover (maximum SHAP value 0.98 $\mu g\ m^{-3}$) and relative humidity (maximum SHAP value 0.88 $\mu g\ m^{-3}$) contribute to the increase in toluene concentration, whereas SOLM (minimum SHAP value -1.14 $\mu g\ m^{-3}$) and T02M (minimum SHAP value -0.93 $\mu g\ m^{-3}$) lead to the highest reduction of toluene levels. After being deposited in the soil, toluene's tendency to volatilize is limited. Some studies have shown that dissolved organic compounds, like humic acids, could enhance VOC transfer to the aqueous phase while our previous findings demonstrated that BTEX levels in rainwater have overtaken calculated values of BTEX transfer between stages (Šoštarić et al., 2017; Andreja Stojić et al., 2019).

In the second type of environment, a moderately positive influence of benzene (0.6 $\mu g\ m^{-3}$ on average) is recorded, leading to the maximum increase in toluene concentration up to 2.1 $\mu g\ m^{-3}$. Considering the first type of environment, the mean concentrations of benzene, PM_{10}, and TNMHC (2.7, 43.8, and 27.8 $\mu g\ m^{-3}$, respectively) are slightly lower and slightly higher for all other pollutants: PM_1, $PM_{2.5}$, m,p-xylene, NO, NO_2, NOx, and O_3 (20.2, 22.8, 5.7, 19.5, 16.5, 46.3, and 50.6 $\mu g\ m^{-3}$, respectively). The influence of PM_{10}, NO, and NOx is twice higher as in the first type of environment. In this environment, the influence of m,p-xylene and TNMHC becomes significant with absolute SHAP values ranging from −1.9 to 3.7 $\mu g\ m^{-3}$ and from −0.6 to 1.7 $\mu g\ m^{-3}$, respectively. Together with benzene, the presence of the most abundant pollutants indicates the influence of traffic emission sources, which is dominant during the warmer part of the year. A non-negligible impact could be assigned to the constant emissions from

"HIP Azotara Pančevo", which is the largest regional plant of mineral fertilizers, and soil/dust resuspension during agricultural practices in the surrounding farming. Besides, NOx and VOCs like toluene are involved in the photochemical cycle of O_3 formation.

Apart from the increase in the mean value of PBLH (483.3 m), other meteorological parameters do not variate significantly. However, the impact of TPP6 (in the range from −0.3 to 1.7 $\mu g\ m^{-3}$), LHTF (from −0.3 to 1.1 $\mu g\ m^{-3}$), and PRSS (from −0.1 to 0.9 $\mu g\ m^{-3}$) changes from negative to positive, as shown by SHAP values given in brackets. Wet scavenging is one of the most important mechanisms for the removal of aerosol particles. The finest particles, so-called accumulation mode particles, are mainly removed during in-cloud scavenging, while ultrafine and coarse particles are removed via below-cloud scavenging, in which aerosol particles collide with falling raindrops or accumulated water (Isokääntä et al., 2022).

As shown by Li et al. (Li et al., 2022), toluene is susceptible to photoconversion in haze air matrix because of the abundant reactive oxygen species (O_{2^-}, O-, O_2^-• and •OH) and the high adsorption ability of PM. Compared to the coarse particles, $PM_{2.5}$ consists of small and heterogeneous aggregated spheres with a branching string structure, which provided larger pore volume and more active sites. This morphology enhances the adsorption of substrate molecules like toluene and oxygen (Wu et al., 2022). During haze days, $PM_{2.5}$ is found to contain more O_2 vacancies, active species such as O_2^-• and •OH and various Lewis's acid sites, which result in a series of toluene transformation into small molecular compounds (e.g., pentane, benzene, ethylbenzene, 4-methyl-1-pentanone and numerous macromolecules.

The third type of environmental conditions refers to a negative influence of benzene (0.9 $\mu g\ m^{-3}$ on average), whereby the reduction of toluene levels can be up to 2.0 $\mu g\ m^{-3}$. The mean concentration of benzene and toluene does not deviate significantly concerning the first and second environmental conditions, but the maximum benzene level is observed throughout the analyzed period (54.7 $\mu g\ m^{-3}$). The belonging values of TNMHC, m,p-xylene, PM_1, $PM_{2.5}$, and O_3 range from slightly higher mean to maximum values. The prominent change is recorded for the influence of TNMHC while the other polluting species showed the least effects. The impact of TNMHC increases in the second environment and causes changes in the toluene concentration from −0.6 to 2.2 $\mu g\ m^{-3}$. Like toluene, TNMHC originates from numerous sources that are active throughout the whole year including vehicular emissions, gasoline filling stations, petrochemical plants, refineries, and chemical plants, and both toluene and TNMHC are highly reactive VOCs that produce tropospheric O_3 through reactions with nitrogen oxides NOx and radicals (Majumdar and Gavane, 2020).

The observed pollutant levels have occurred under maximum levels of the indicators of convective movements in the atmosphere (CAPE and CINH), indicating that the conditions are more unstable compared to the other two environments. The greatest effects of the other meteorological parameters (CRAI, LIB4, PBLH, SHGT, and WS) also imply the atmospheric conditions that enable vertical mixing, wet scavenging, pollutant dispersion, and transport. As can be concluded, the conditions characterizing the third type of environment are not seasonally specific and relate to the consistent impacts of anthropogenic pollutant sources.

Based on the aforementioned, machine learning has emerged as a crucial tool in dissecting and understanding the patterns of environmental pollution, as evidenced by research spanning from satellite data analysis to source apportionment (Liu et al., 2022). The complexity of global environmental challenges highlights the need for an interdisciplinary collaborative approach, combining environmental science with computer science to devise effective solutions. In this study, we demonstrate that machine learning combined with explainable artificial intelligence does more than predict; it provides insights into the importance of variables and the nature of relationships within the data. Our modeling approach addresses limitations previously identified in research, particularly the challenge of hyperparameter optimization. By employing a hybrid reptile search algorithm and an XGBoost model, our study introduces an innovative methodology that may deliver more accurate interrelations of the non-linear relationships between pollutants and environmental factors than traditional models (Gu et al., 2022). Additionally, the use of SHAP values underscores the study's focus on model transparency and understandability, highlighting our collective commitment to making complex machine learning models more accessible and actionable for stakeholders.

5.3. Policy making

Given the global increase in urban populations and their exposure to high concentrations of air pollutants, the data-driven approach we present offers valuable insights for urban planning and long-term environmental management (Li et al., 2024; Iorember et al., 2023). Furthermore, the advanced models we introduce can predict and assess the effectiveness of pollution control measures, facilitating ongoing enhancements in air quality management.

5.4. Limitations of this study

The results of this study are contingent upon specific environmental conditions and may not be directly applicable to different settings or pollutants without appropriate adjustments and validations. Furthermore, the effectiveness of machine learning models is significantly influenced by the quality and volume of input data, and thus, biases or deficiencies in the dataset can markedly undermine the accuracy and dependability of the models. Additionally, the study may not comprehensively capture all the variables that affect urban air quality, including unpredictable weather conditions and diverse sources of emissions. These omissions could affect the predictive models' robustness and reliability.

6. Conclusion

The fate of air-pollutants depends upon some factors, such as emission source activity, topographic features, and meteorological conditions, which interact and provide the atmospheric environment for complex pollution-related processes and phenomena. This paper contributed to the field of air quality, atmosphere, and health by providing insights into the environmental fate of toluene in complex urban environments and proposing the methodology for investigating the behavior pattern of toluene across distinguished environmental settings and its interactions with other polluting species or meteorological factors, which were identified as influential within each setting. The proposed methodology can be used as a useful tool for addressing complex environmental issues and acquiring knowledge on stochastic nature processes. The proposed hybrid reptile search algorithm (HRSA) for air pollution modeling outperformed other well-known metaheuristics algorithms in optimizing the XGBoost model for toluene atmospheric fate research. The XGBoost model, optimized by HRSA, provided satisfactory performance for regression problems and can be used to analyze the environmental fate of toluene in complex urban environments.

The existing research highlights the transformative potential of machine learning in air quality research. Complex algorithms are praised for their accuracy in predicting pollutant levels, yet they often face criticism for their lack of interpretability and their potential to overlook the precision of peak values, which are essential for managing severe pollution episodes. Interpretability is crucial for developing practical recommendations for environmental and health policies. In response, we have utilized a reptile search-optimized XGBoost model for toluene analysis, emphasizing interpretability through the SHAP method. This approach provides detailed insights into the specific relationships and impacts of various features. We have identified nonlinear relationships between toluene and other pollutants and

meteorological factors. We demonstrated that benzene is the most important predictor of toluene dynamics, and the interaction between toluene and benzene varies depending on the environmental conditions. We also distinguished three types of environments based on the conditions and the way in which toluene and benzene are interrelated, as well as on their coexistence and interactions with other polluting species.

The principal findings of this article are: (1) The proposed HRSA algorithm can be applied to other air quality problems to improve the search performance and obtain better results. (2) The XGBoost model, optimized by HRSA, can be used to analyze the environmental fate of other pollutants in complex urban environments. (3) The SHAP method can be applied to interpret the results of other machine learning models and gain insights into the relationships between predictors and outcomes.

We are aware that these findings are specific to certain conditions and may not apply universally without adjustments, and that data quality and quantity significantly affect machine learning model performance, with biases or gaps impacting accuracy. Additionally, not all factors affecting urban air quality, like variable weather and diverse emissions, are fully accounted for, potentially limiting model robustness and reliability. Therefore, future research could benefit from extending the application of this methodology to diverse environmental settings, thereby capturing a broader spectrum of pollution dynamics. Additionally, conducting both a comprehensive seasonal analysis and a comparative analysis of pollution levels and sources between urban and rural areas would provide critical insights and a nuanced understanding of pollution fluctuations throughout the year. This approach would enhance our ability to address environmental challenges more effectively.

Ethical responsibilities of Authors

All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, as well as by the Science Fund of the Republic of Serbia, Grant No. #7373, Characterizing crises-caused air pollution alternations using an artificial intelligence-based framework - crAIRsis and Grant No. #7502, Intelligent Multi-Agent Control and Optimization Applied to Green Buildings and Environmental Monitoring Drone Swarms - ECOSwarm.

CRediT authorship contribution statement

Nebojsa Bacanin: Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mirjana Perisic:** Writing – original draft, Validation, Investigation, Formal analysis. **Gordana Jovanovic:** Writing – original draft, Validation, Investigation, Formal analysis. **Robertas Damaševičius:** Writing – review & editing, Validation, Investigation, Formal analysis. **Svetlana Stanisic:** Validation, Software, Investigation, Data curation. **Vladimir Simic:** Validation, Software, Resources, Investigation, Formal analysis. **Miodrag Zivkovic:** Writing – original draft, Visualization, Validation, Investigation, Formal analysis. **Andreja Stojic:** Writing – original draft, Visualization, Validation, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abraham, Ajith, He, Guo, Liu, Hongbo, 2006. Swarm intelligence: foundations, perspectives and applications. In: Swarm Intelligent Systems, pp. 3–25.
Abualigah, Laith, Diabat, Ali, 2023. Chaotic binary reptile search algorithm and its feature selection applications. J. Ambient. Intell. Humaniz. Comput. 14 (10), 13931–13947.
Abualigah, Laith, Diabat, Ali, et al., 2021. The arithmetic optimization algorithm. In: Computer Methods in Applied Mechanics and Engineering, 376, p. 113609.
Abualigah, Laith, Abd Elaziz, Mohamed, et al., 2022. Reptile Search Algorithm (RSA): a nature-inspired meta-heuristic optimizer. Expert Syst. Appl. 191, 116158.
Abualigah, Laith, Habash, Mahmoud, et al., 2023. Improved reptile search algorithm by salp swarm algorithm for medical image segmentation. J. Bionic Eng. 20 (4), 1766–1790.
Alharbi, A., et al., 2021. Botnet attack detection using local global best bat algorithm for industrial internet of things. Electronics 10 (11).
Almotairi, Khaled H., Abualigah, Laith, 2022a. Hybrid reptile search algorithm and remora optimization algorithm for optimization tasks and data clustering. Symmetry 14 (3), 458.
Almotairi, Khaled H., Abualigah, Laith, 2022b. Improved reptile search algorithm with novel mean transition mechanism for constrained industrial engineering problems. Neural Comput. & Applic. 34 (20), 17257–17277.
Alzaqebah, A., et al., 2022. A modified grey wolf optimization algorithm for an intrusion detection system. Mathematics 10 (6).
Bacanin, Nebojsa, Bezdan, Timea, Tuba, Eva, et al., 2020. Optimizing convolutional neural network hyperparameters by enhanced swarm intelligence metaheuristics. Algorithms 13 (3), 67.
Bacanin, Nebojsa, Bezdan, Timea, Venkatachalam, K., et al., 2021. Artificial neural networks hidden unit and weight connection optimization by quasi-refection-based learning artificial bee colony algorithm. IEEE Access 9, 169135–169155.
Bacanin, Nebojsa, Sarac, Marko, et al., 2022a. Smart wireless health care system using graph LSTM pollution prediction and dragonfly node localization. Sustain. Comput.: Inform. Syst. 35, 100711.
Bacanin, Nebojsa, Stoean, Catalin, Zivkovic, Miodrag, Jovanovic, Dijana, et al., 2022b. Multi-swarm algorithm for extreme learning machine optimization. Sensors 22 (11), 4204.
Bacanin, Nebojsa, Zivkovic, Miodrag, Al-Turjman, Fadi, et al., 2022c. Hybridized sine cosine algorithm with convolutional neural networks dropout regularization application. Sci. Rep. 12 (1), 1–20.
Bacanin, Nebojsa, Zivkovic, Miodrag, Stoean, Catalin, et al., 2022d. Application of natural language processing and machine learning boosted with swarm intelligence for spam email filtering. Mathematics 10 (22), 4173.
Bacanin, Nebojsa, Stoean, Catalin, Zivkovic, Miodrag, Rakic, Miomir, et al., 2023a. On the benefits of using metaheuristics in the hyperparameter tuning of deep learning models for energy load forecasting. Energies 16 (3), 1434.
Bacanin, Nebojsa, Venkatachalam, K., et al., 2023b. A novel firefly algorithm approach for efficient feature selection with COVID-19 dataset. Microprocess. Microsyst. 98, 104778.
Beni, Gerardo, 2020. Swarm intelligence. In: Complex Social and Behavioral Systems: Game Theory and Agent-based Models, pp. 791–818.
Budholiya, Kartik, Shrivastava, Shailendra Kumar, Sharma, Vivek, 2022. An optimized XGBoost based diagnostic system for effective prediction of heart disease. J. King Saud Univ. - Comput. Inf. Sci. 34 (7), 4514–4523.
Bukumira, Milos, et al., 2022. Carrot grading system using computer vision feature parameters and a cascaded graph convolutional neural network. J. Electron. Imaging 31 (6), 061815.
Burghardt, Tomasz E., Pashkevich, Anton, Żakowska, Lidia, 2016. Influence of volatile organic compounds emissions from road marking paints on ground-level ozone formation: case study of Kraków, Poland. Transp. Res. Procedia 14, 714–723.
Chen, Tianqi, Guestrin, Carlos, 2016. Xgboost: a scalable tree boosting system. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
Chen, Tianqi, He, Tong, et al., 2015. Xgboost: Extreme Gradient Boosting. *R package version 0.4-2* 1.4, pp. 1–4.
Cheremisinoff, N.P., Rosenfeld, P.E., 2010. Sources of air emissions from pulp and paper mills. In: Handbook of Pollution Prevention and Cleaner Production, 2, pp. 179–259.
Clough, S.R., 2014. Toluene A2-Wexler. Philip BT-Encyclopedia of Toxicology.

Cuk, Aleksa, et al., 2024. Tuning attention based long-short term memory neural networks for Parkinson's disease detection using modified metaheuristics. Sci. Rep. 14 (1), 4309.
Dahou, Abdelghani, et al., 2022. Intrusion detection system for IoT based on deep learning and modified reptile search algorithm. Comput. Intell. Neurosci. 2022.
Damaševičius, Robertas, et al., 2024. Decomposition aided attention-based recurrent neural networks for multistep ahead time-series forecasting of renewable power generation. PeerJ Comput. Sci 10.
Davidson, Cameron J., Hannigan, John H., Bowen, Scott E., 2021. Effects of inhaled combined benzene, toluene, ethylbenzene, and xylenes (BTEX): toward an environmental exposure model. Environ. Toxicol. Pharmacol. 81, 103518.
Dobrojevic, Milos, et al., 2023. Addressing internet of things security by enhanced sine cosine metaheuristics tuned hybrid machine learning model and results interpretation based on SHAP approach. PeerJ Comput. Sci 9, e1405.
Dorigo, Marco, Birattari, Mauro, Stutzle, Thomas, 2006. Ant colony optimization. IEEE Comput. Intell. Mag. 1 (4), 28–39.
Eicher, Tracy J., 2009. Toxic encephalopathies I: cortical and mixed encephalopathies. In: Clinical Neurotoxicology, Syndromes, Substances, Environments, 2009, pp. 69–87.
Emmerich, Michael, Shir, Ofer M., Wang, Hao, 2018. Evolution strategies. In: Handbook of Heuristics. Springer, pp. 89–119.
Fausto, Fernando, et al., 2020. From ants to whales: metaheuristics for all tastes. Artif. Intell. Rev. 53 (1), 753–810.
Gu, Yuanlin, Li, Baihua, Meng, Qinggang, 2022. Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing 468, 123–136.
Hazrati, Sadegh, et al., 2016. Benzene, toluene, ethylbenzene and xylene concentrations in atmospheric ambient air of gasoline and CNG refueling stations. Air Qual. Atmos. Health 9, 403–409.
Heidari, Ali Asghar, et al., March 2019. Harris hawks optimization: algorithm and applications. Futur. Gener. Comput. Syst. 97, 849–872.
Helmi, A.M., et al., 2021. A novel hybrid gradient-based optimizer and grey wolf optimizer feature selection method for human activity recognition using smartphone sensors. Entropy 23 (8).
Iorember, Paul Terhemba, et al., 2023. New insight into decoupling carbon emissions from economic growth: do financialization, human capital, and energy security risk matter? Rev. Dev. Econ.
Isokääntä, Sini, et al., 2022. The effect of clouds and precipitation on the aerosol concentrations and composition in a boreal forest environment. Atmos. Chem. Phys. 22 (17), 11823–11843.
Jouhari, H., et al., 2020. Modified Harris Hawks optimizer for solving machine scheduling problems. Symmetry 12 (9).
Jovanovic, Gordana, et al., April 2023. Potential of coupling metaheuristics- optimized-XGBoost and SHAP in revealing PAHs environmental fate. Toxics 11.4, 394. ISSN: 2305-6304. https://doi.org/10.3390/toxics11040394. URL: doi:10.3390/toxics11040394.
Jovanovic, Luka, Jovanovic, Dejan, et al., 2022. Multi-step crude oil price prediction based on LSTM approach tuned by salp swarm algorithm with disputation operator. Sustainability 14 (21), 14616.
Jovanovic, Luka, Jovanovic, Dijana, et al., 2023a. Improving phishing website detection using a hybrid two-level framework for feature selection and xgboost tuning. J. Web Eng. 22 (3), 543–574.
Jovanovic, Luka, Jovanovic, Gordana, et al., 2023b. The explainable potential of coupling metaheuristics-optimized-XGBoost and SHAP in revealing VOCs' environmental fate. Atmosphere 14 (1), 109.
Karaboga, Dervis, Basturk, Bahriye, 2008. On the performance of artificial bee colony (ABC) algorithm. Appl. Soft Comput. 8 (1), 687–697.
Kennedy, James, Eberhart, Russell, 1995. Particle swarm optimization. In: Proceedings of ICNN'95-International Conference on Neural Networks, vol. 4. IEEE, pp. 1942–1948.
Khishe, Mohammad, Mosavi, Mohammad Reza, 2020. Chimp optimization algorithm. Expert Syst. Appl. 149, 113338.
Khurma, R.A., et al., 2021. An enhanced evolutionary software defect prediction method using island moth flame optimization. Mathematics 9, 15.
Kroll, Jesse H., et al., 2011. Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. Nat. Chem. 3 (2), 133–139.
LaTorre, Antonio, et al., 2021. A prescription of methodological guidelines for comparing bio-inspired optimization algorithms. Swarm Evol. Comput. 67, 100973.
Li, Pengcheng, et al., 2024. Impact of sectoral mix on environmental sustainability: how is heterogeneity addressed? Gondwana Res. 128, 86–105.
Li, Qianqian, et al., 2022. PM2. 5-mediated photochemical reaction of typical toluene in real air matrix with identification of products by isotopic tracing and FT-ICR MS. Environ. Pollut. 313, 120181.
Liu, Xian, et al., 2022. Data-driven machine learning in environmental pollution: gains and problems. Environ. Sci. Technol. 56 (4), 2124–2133.
Lundberg, Scott M., Lee, Su-In, 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Proces. Syst. 30.
Lundberg, Scott M., Erion, Gabriel, et al., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.
Majumdar, Deepanjan, Gavane, Ashok Gangadhar, 2020. Diurnal-, seasonal- and site-dependent variability in ground-level total non-methane hydrocarbon in Nagpur City of Central India. Asian J. Atmos. Environ. 14 (1), 1–13.
Makhadmeh, S.N., et al., 2021. Smart home battery for the multi-objective power scheduling problem in a smart home using grey wolf optimizer. Electronics 10 (4), 1–35.
Minic, Ana, et al., 2023. Applying recurrent neural networks for anomaly detection in electrocardiogram sensor data. Sensors 23 (24), 9878.
Mirjalili, Seyedali, Lewis, Andrew, 2016. The whale optimization algorithm. Adv. Eng. Softw. 95, 51–67.
Murindababisha, David, et al., 2021. Current progress on catalytic oxidation of toluene: a review. Environ. Sci. Pollut. Res. 1–31.
Pavlov-Kagadejev, Marijana, et al., 2024. Optimizing long-short-term memory models via metaheuristics for decomposition aided wind energy generation forecasting. Artif. Intell. Rev. 57 (3), 45.
Perišić, Mirjana, et al., 2017. Forecasting hourly particulate matter concentrations based on the advanced multivariate methods. Int. J. Environ. Sci. Technol. 14 (5), 1047–1054.
Probst, Philipp, Boulesteix, Anne-Laure, Bischl, Bernd, 2019. Tunability: importance of hyperparameters of machine learning algorithms. J. Mach. Learn. Res. 20 (53), 1–32.
Rahnamayan, Shahryar, Tizhoosh, Hamid R., Salama, Magdy M.A., 2007. Quasi-oppositional differential evolution. In: 2007 IEEE Congress on Evolutionary Computation. IEEE, pp. 2229–2236.
Salb, Mohamed, et al., 2023. Enhancing internet of things network security using hybrid CNN and XGBoost model tuned via modified reptile search algorithm. Appl. Sci. 13 (23), 12687.
Shah, Syed Ale Raza, et al., 2023. Waste management, quality of life and natural resources utilization matter for renewable electricity generation: the main and moderate role of environmental policy. Util. Policy 82, 101584.
Shapiro, Samuel S., Francia, R.S., 1972. An approximate analysis of variance test for normality. J. Am. Stat. Assoc. 67 (337), 215–216.
Skorokhod, Andrey I., et al., 2017. Benzene and toluene in the surface air of northern Eurasia from TROICA-12 campaign along the Trans-Siberian Railway. Atmos. Chem. Phys. 17 (8), 5501–5514.
Šoštarić, A., et al., 2017. Rainwater capacities for BTEX scavenging from ambient air. Atmos. Environ. 168, 46–54.
Stanišić, Svetlana, Perišić, Mirjana, et al., 2021. What information on volatile organic compounds can be obtained from the data of a single measurement site through the use of artificial intelligence?. In: Artificial Intelligence: Theory and Applications. Springer, pp. 207–225.
Stanišić, Svetlana, Jovanović, Gordana, et al., Mar. 2022. Explaining the environmental fate of PAHs in indoor and outdoor environments by the use of artificial intelligence. In: Gregoire, Warren L. (Ed.), Polycyclic Aromatic Hydrocarbons. Nova Science, Hauppauge, NY, pp. 1–36. Chap. 1.
Stegherr, Helena, Heider, Michael, Hähner, Jörg, 2020. Classifying metaheuristics: towards a unified multi-level classification system. Nat. Comput. 1–17.
Stoean, Catalin, et al., March 2023. Metaheuristic-based hyperparameter tuning for recurrent deep learning: application to the prediction of solar energy generation. Axioms 12 (3), 266 issn: 2075-1680. https://doi.org/10.3390/axioms12030266. url: doi:10.3390/axioms12030266.
Stojić, A., et al., 2018. Urban air pollution: an insight into its complex aspects. In: A Closer Look at Urban Areas. Nova Science Publishers, NY, USA.
Stojić, Andreja, Maletić, Dimitrije, et al., 2015. Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. Sci. Total Environ. 521, 19–26.
Stojić, Andreja, Stanić, Nenad, et al., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Total Environ. 653, 140–147.
Stojić, Andreja, Jovanović, Gordana, et al., 2022. The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: explainable prediction of benzo [a] pyrene levels. Chemosphere 289, 133154.
Todorovic, Mihailo, et al., 2023. Improving audit opinion prediction accuracy using metaheuristics-tuned XGBoost algorithm with interpretable results through SHAP value analysis. Appl. Soft Comput. 149, 110955.
Whitten, Gary Z., et al., 2010. A new condensed toluene mechanism for carbon bond: CB05-TU. Atmos. Environ. 44 (40), 5346–5355.
Wilcoxon, Frank, 1992. Individual comparisons by ranking methods. In: Breakthroughs in Statistics, pp. 196–202.
Wohl, Charel, et al., 2023. Marine biogenic emissions of benzene and toluene and their contribution to secondary organic aerosols over the polar oceans. Sci. Adv. 9.4, eadd9031.
Wolpert, David H., Macready, William G., 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1 (1), 67–82.
Wu, Shipeng, et al., 2022. O-vacancy-rich porous MnO2 nanosheets as highly efficient catalysts for propane catalytic oxidation. Appl. Catal. B Environ. 312, 121387.
Yang, Xin-She, 2009. Firefly algorithms for multimodal optimization. In: International Symposium on Stochastic Algorithms. Springer, pp. 169–178.
Yang, Xin-She, 2010. A new metaheuristic bat-inspired algorithm. In: Nature Inspired Cooperative Strategies for Optimization (NICSO 2010). Springer, pp. 65–74.
Yang, Xin-She, Gandomi, Amir Hossein, 2012. Bat algorithm: a novel approach for global engineering optimization. Eng. Comput. 29 (5), 464–483.
Zivkovic, Miodrag, Bacanin, Nebojsa, et al., 2022a. Hybrid CNN and XGBoost model tuned by modified arithmetic optimization algorithm for COVID-19 early diagnostics from X-ray images. Electronics 11 (22), 3798.
Zivkovic, Miodrag, Jovanovic, Luka, et al., 2022b. Xgboost hyperparameters tuning by fitness-dependent optimizer for network intrusion detection. In: Communication and Intelligent Systems: Proceedings of ICCIS 2021. Springer, pp. 947–962.
Zivkovic, Tamara, et al., 2023. Software defects prediction by metaheuristics tuned extreme gradient boosting and analysis based on shapley additive explanations. Appl. Soft Comput. 146, 110659.

Marine Pollution Bulletin 170 (2021) 112654

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Baseline

Fatty acids, persistent organic pollutants, and trace elements in small pelagic fish from the eastern Mediterranean Sea

Snježana Herceg Romanić [a], Gordana Jovanović [b,c,*], Bosiljka Mustać [d], Jasna Stojanović-Đinović [e], Andreja Stojić [b,c], Tena Čadež [a], Aleksandar Popović [f]

[a] Institute for Medical Research and Occupational Health, Ksaverska cesta 2, PO Box 291, 10001 Zagreb, Croatia
[b] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
[c] Singidunum University, Danijelova 32, 11 000 Belgrade, Serbia
[d] Department of ecology, agronomy and aquaculture, University of Zadar, Trg Kneza Višeslava 9, 23000 Zadar, Croatia
[e] Institute of Meat Hygiene and Technology, Kaćanskog 13, 11 000 Belgrade, Serbia
[f] Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia

ARTICLE INFO

Keywords:
Saturated fatty acids
Omega fatty acids
Organochlorine pesticides (OCPs)
Polychlorinated biphenyls (PCBs)
Unmix
Self-organizing maps (SOM)

ABSTRACT

Fatty acids (FAs) composition, 24 persistent organic pollutants (POPs), and 16 trace elements were examined in small pelagic fish (sardine, anchovy, round sardinella, chub and horse mackerels) caught by a fishing fleet for more than three years in the eastern Mediterranean Sea. Five Unmix source profiles associated with both sources, such as overlapping diet, including low-niche marine organisms and inputs from the surrounding environmental compartments were resolved. Inorganic compounds were notably more abundant in fish tissue than organochlorine xenobiotics. Comparison with the values of toxicological parameters revealed that the examined fish species are safe for human consumption, while the content of FAs emphasized the studied species as a valuable source of nutrients. A significant linear correlation was not observed between the 18 FAs and lipophilic organochlorines. Based on the obtained database, future assessments of the quality of edible fish species and the aquatic environment of the eastern Mediterranean Sea, which is known as an important fishing ground, could be significantly improved.

Globally, marine environments are well known as a final sink of environmental contaminants, which are continuously burdened by coastal outfalls, rivers, and deposition from the atmosphere (UNEP/MAP-Barcelona Convention;, 2012). They also act as a secondary source of xenobiotics owing to the absence of or very slow degradation of the pollutants. The Mediterranean Sea is a semi-enclosed sea and an important commercial fishing ground. Catches are dominated by small pelagic fish and represent approximately 50% of all catches in Mediterranean countries (FAO, 2018a, 2018b). Small pelagic fish, e.g., sardine, anchovy, round sardinella, chub and horse mackerels, are considered to have a key role in marine ecosystems, since they transport organic matter from lower (e.g., planktonic production) to higher trophic levels, including predator species and humans (Hure and Mustać, 2020). The exploitation of small pelagic fish can have important effects on the energy channelling and the food web (Palomera et al., 2007). Small pelagic oily fish are a highly recommended nutrient source worldwide because of their content of proteins, minerals, and healthy fats including long-chain omega-3 (ω-3) and 6 ω-6) polyunsaturated fatty acids (PUFAs) (FAO, 2018a, 2018b). Nutritional benefits of fish consumption concur with the co-ingestion of lipophilic pollutants such as POPs and other contaminants such as trace elements. Many of the POPs and heavy metals are harmful to the human body and its proper functioning, thus representing a risk to human health if exposure is chronic. Among meat and dairy products, intake of fish and shellfish contributes to more than 90% of human exposure to POPs (WHO, 2014).

Many studies have been conducted to assess the levels of POPs and trace elements in the Mediterranean marine environment (Bilandžić et al., 2011; Kljaković-Gašpić et al., 2010; Kljaković Gašpić et al., 2015; Vuković et al., 2018; Ancora et al., 2020; Storelli et al., 2012; and references therein). Despite bans in many Mediterranean countries for more than 40 years and Stockholm Convention measures, the level of POPs has reached a "steady state level", which still possess adverse effects on humans. A recent study of POP levels in anchovy, sardine, and bogue from the Mediterranean Sea pointed to the importance of health

* Corresponding author at: Pregrevica 118, 11080 Belgrade, Serbia.
E-mail address: gordana.vukovic@ipb.ac.rs (G. Jovanović).

https://doi.org/10.1016/j.marpolbul.2021.112654
Received 21 October 2020; Received in revised form 4 June 2021; Accepted 17 June 2021
Available online 29 June 2021

Fig. 1. Mediterranean Sea map with indicated Adriatic Sea.

risk assessment of marine toxins based on their concentrations in fish species (Bartalini et al., 2020). The purpose of this study was to examine the content of fatty acids (FAs) and contaminants (POPs and trace elements) in five species of small pelagic fish (sardine *Sardina pilchardus* (Walbaum, 1792), anchovy *Engraulis encrasicolus* (Linnaeus, 1758), round sardinella *Sardinella aurita* (Valenciennes, 1847), chub mackerel *Scomber japonicus* (Houttuyn, 1782), and horse mackerel *Trachurus trachurus* (Linnaeus, 1758) caught in the Mediterranean Sea (Croatian Adriatic).

A total of 105 fish samples were collected from various fishery zones by a fishing fleet for three years (2014, 2015, and 2016) along the eastern Adriatic Sea (Fig. 1). Biometric data of the collected samples are presented in Supplementary data (Table S1). A random sample of cca 50 specimens was taken by purse seine catches (mesh size: 8 mm/bar length) and a total of 105 pooled samples from a fillet of specimens.

Seven organochlorine pesticides, OCPs, [hexachlorobenzene (HCB), hexachlorocyclohexane isomers (α-, β-, and γ-HCH), 1,1,1-trichloro-2,2-di(4-chlorophenyl)ethane (*p,p'*-DDT), 1,1-dichloro-2,2-di(4-chlorophenyl) ethylene(*p,p'*-DDE), 1,1-dichloro-2,2-di(4-chlorophenyl)ethane (*p,p'*-DDD)], and 17 PCB congeners were analysed. Out of 209 polychlorinated biphenyls(PCBs),six indicator PCBs (PCB-28, PCB-52, PCB-101, PCB-138, PCB-153, and PCB-180) were analysed and presented as ΣPCBi. Toxicologically relevant PCB congeners (PCB-60, PCB-74, PCB-105, PCB-114, PCB-118, PCB-123, PCB-156, PCB-157, PCB-167, PCB-170, PCB-189) were presented as ΣPCBtox. Chemical analysis, quality assurance, and quality control were described in detail elsewhere (Kljaković Gašpić et al., 2015; Vuković et al., 2018). Briefly, two aliquots were analysed from each pool, and approximately 5 g were homogenized with 2 g of sodium sulphate, cold extracted in 40 mL of n-hexane, repeatedly cleaned with 96% sulphuric acid, and analysed by high-resolution gas chromatography with an electron capture detector (s) (CLARUS 500). Two capillary columns (Restek, Bellefonte, PA, USA) were used simultaneously: (1) 60 m × 0.25 mm, Rtx-5, and (2) 30 m × 0.25 mm, Rtx-1701. Only the compounds identified on both columns were evaluated. The performance of the analytical procedure was validated through the analysis of the reference material IAEA-406 (fish homogenate) supplied by the International Atomic Energy Agency-Marine Environment Laboratory (IAEA-MEL), Monaco. The produced data on OCPs and PCBs were within the acceptable range. The recoveries ranged from 79% to 89%, with a relative standard deviation from 4% to 15%. The detection limits for PCBs and OCPs were 0.01 ng g^{-1} of wet weight.

The contents of 6 saturated fatty acids, SFAs, (myristic, pentadecylic, palmitic, palmitoleic, margaric, and stearic acid), 3 monounsaturated fatty acids, MUFAs, (oleic, paullinic, and arachidonic acid) and 9 PUFAs from ω -6 and ω-3 families (linoleic, α-linolenic, eicosadienoic, dihomo-γ-linolenic, eicosatrienoic, arachidonic, eicosapentaenoic, docosapentaenoic, and docosahexaenoic) were determined. Details on the analysis of FAs have been previously reported (Spirić et al., 2010). Briefly, total lipids were extracted from fish muscle by accelerated solvent extraction (ASE). Fatty acid methyl esters (FAMEs) were prepared by transesterification using trimethylsulfonium. The samples were analysed as FAMEs by gas–liquid chromatography with a flame ionisation detector (GC/FID).

Analysis of 16 elements (Fe, Zn, Cu, Mn, Se, Cr, Co, Ni, Na, K, Mg, Ca, Cd, Pb, Hg, and As) and its quality assurance and quality control have been described previously (Nikolić et al., 2017). In brief, homogenized fish samples were mineralized and then digested in a microwave digestion system. After cooling at room temperature, the digests were diluted with deionized water and analysed by inductively coupled plasma-mass spectrometry (ICP-MS).

To investigate any similarities among fish samples based on their tissue burden by POPs, inorganic elements, and FAs composition, the Kohonen self-organizing maps (SOM)were applied. Following the first description by Kohonen (2013), SOM analysis has been used in a broad variety of fields. It has been proven as an efficient clustering tool in environmental investigations (Herceg Romanić et al., 2018; Mari et al., 2010; Vuković et al., 2018) and is better than traditional approaches such as principal component and cluster analysis (Budayan et al., 2009; Wehrens and Buydens, 2007). Detailed elaborations on the SOM principles are beyond the scope of this study and are mentioned elsewhere (Wehrens and Buydens, 2007; Vuković et al., 2018). The SOM inputs were the values of each calculated parameter (69 in total) in the fish samples ($N = 105$). The values of all pollutants and FAs were normalized to the range of 0–1. The number of neurons in the output layer (map) was 16 (4 × 4) aiming for at least 5–10 samples per node when choosing the map size. A hexagonal grid was selected, while the iteration process was optimized until the distance from the weights of each node to the samples represented by that node reached a minimum plateau (Wehrens and Kruisselbrink, 2017). SOM and correlation analysis were performed

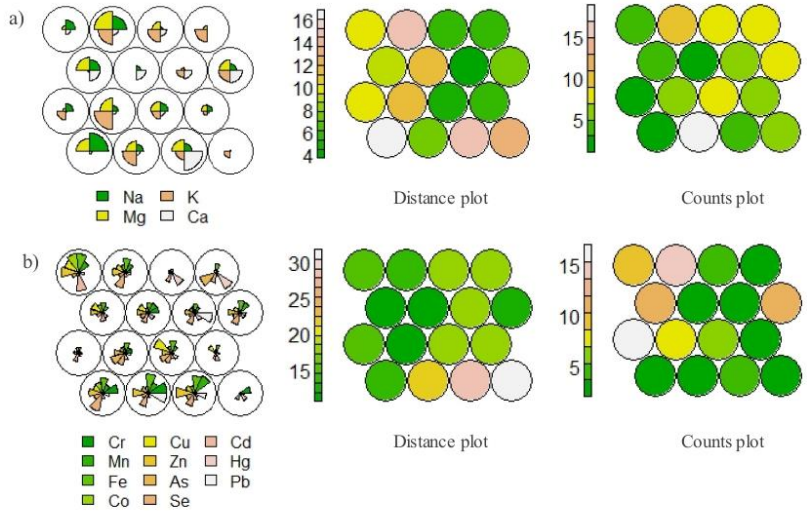


Fig. 2. Self-organizing maps, distance and count plots of marine fish species (anchovy, sardines, chub mackerel, round sardinella, and horse mackerel) classified based on the content of: a) macroelements and b) trace elements; warmer colour at distance plot means stronger differences between clusters, while counts plot represent the number of sample in each cluster.

using R, a free software environment for statistical computing (R Team, 2012). One-way ANOVA was performed using STATISTICA (data analysis software system), version 12 (StatSoft Inc., 2012) to test whether there was a significant difference ($p < 0.05$) in the content of the pollutant and FAs in the investigated species. To quantify the sources contributing to the fish sample burden by pollutants, Unmix (USEPA, 2007) was performed. The maximum number of pollutants was selected as input variables following the trials and errors, which enabled to gain the most meaningful and reliable results.

Descriptive statistics of the FAs content in the examined fish species is presented in Supplementary material (Table S2). Palmitic acid (C16:0) was the most prevalent and equally abundant FAs in sardine, anchovy, and mackerel species with a median value that ranged between 29.5% and 41%. The second most abundant FAs was oleic acid (C18:1n-9) exhibiting a median range from 11.3% to 21.2%. A similar range was found in tuna muscle (Sprague et al., 2012). In the studied species, palmitic and oleic acids are usually found in the form of free FAs in which they are transported to the large intestine and extracted (Rincón-Cervera et al., 2019). Stearic acid (C18:0) and docosahexaenoic acid (C22:6n-3) DHA were also dominant among investigated FAs; their median content ranged from 6.5% to 12.11% and from 10.4% to 21.3%, respectively. The content of eicosapentaenoic acid (C20:5n-3) EPA varied between 3.3% and 4.6%, while the DHA:EPA ratio was approximately 2.5:1. Among essential FAs, the levels of linoleic acid, C18:2n-6, (1.5%–2.2%) was three times higher than the content (0.4%–0.7%) of α-linoleic acid. A metabolite of linoleic acid, dihomo-γ-linoleic acid (C20:3n-6), which serves as an important constituent of neuronal membrane phospholipids, were found up to 1%. Eicosatrienoic fatty acid (ETA) (20:3n-3) is a rare PUFA of the ω-3 series and it was found within the range between 0.2% and 0.4%. In humans, it represents less than 0.25% of serum phospholipid FAs and it is one of the most active essential FAs. Eicosadienoic acid (EDA) (C20:2n-6) is a rare, naturally occurring n-6 FA that has attracted great interest because it acts as an ω-6 and ω-3 fatty acid analogue, and it was found within the range between 0.4% and 0.5%. Paullinic acid is an ω-7 fatty acid found in a variety of fish and its presence in the study samples falls between 0.5% and 1.4%. These results pointed to small pelagic fish as a good source of FAs, and their nutritional benefits are comparable to those of Atlantic salmon (Lundebye et al., 2017) and Atlantic Bluefin tuna (Sprague et al., 2012). In this study, the accumulation pattern of both DHA and EPA was as follows: sardines>anchovy = chub mackerel>horse mackerel. Many past studies have indicated that ω-6 to ω-3 PUFA ratio in the human diet shifted from 1:1 in the past to the current 20:1. High amounts of ω-6 PUFA can cause the pathogenesis of many diseases, whereas high levels of ω-3 PUFA exhibit suppressive effects (Nasir and Bloch, 2019). In this study, the ω-6 to ω-3 ratio varied from 0.025 to 22.63 (average value = 1.80) and demonstrated pronounced maximums in sardines and chub mackerel. As revealed by the results, a variety of different fish species should be consumed to gain a sufficient supply of nutritionally relevant EPA and DHA.

Table S3 (Supplementary data) summarizes the levels of organochlorine compounds and trace elements. The highest medians were determined for macroelements, which are involved in numerous metabolism reactions and are important for human health. The abundance pattern was as follows: K > Na > Ca > Mg in horse mackerel and round sardinella, and Mg > K > Ca > Na in sardine, anchovy, and chub mackerel. The concentration of microelements (Fe > Zn > Cu) was within the normal physiological range in the studied fish species, below the levels that could potentially cause pathological changes in tissues and organs. The levels of the most toxic elements including Hg, Pb, and Cd in all studied species were lower than the maximum thresholds prescribed by the existing EU regulations (European Commission, 2006, 2008). Several studies concerning the level of trace elements in the Adriatic marine environment and large pelagic fish, such as tuna or swordfish, were published previously (e.g., Bilandžić et al., 2011; Kljaković-Gašpić et al., 2010; Ancora et al., 2020; and references therein). Comparison with these investigations confirmed that contaminant levels found in the fish species sampled in this study reflect low or moderate pollution of the aquatic surrounding. In addition, according to literature data, metal concentrations in fish vary widely, depending on the location of capture, sex, size, water concentrations of metals, and exposure period (Zhang and Wong, 2007).

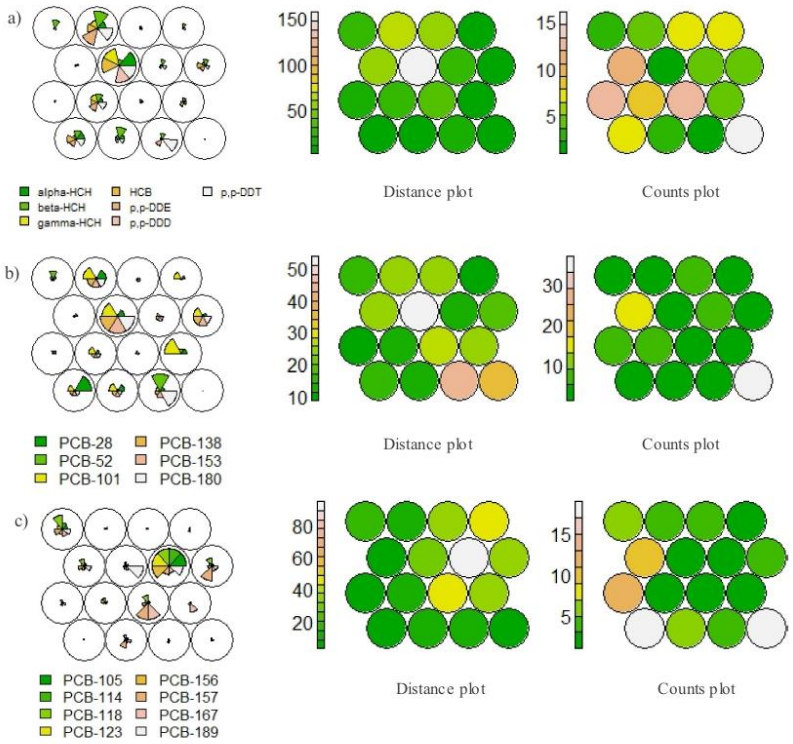


Fig. 3. Self-organizing maps, distance and count plots of marine fish species (anchovy, sardines, chub mackerel, round sardinella, and horse mackerel) classified based on the content of: a) organochlorine pesticides, b) indicator PCBs, and c) toxicologically relevant (dioxin-like) PCBs; warmer colour at distance plot means stronger differences between clusters, while counts plot represent the number of sample in each cluster.

The order of abundance of the major POP groups was as follows: PCB > DDT > HCH > HCB. Among PCBs, the most dominant congener was PCB-153, and among OCPs the most prominent compound was *p,p'*-DDE. Out of the investigated PCBs, the highest concentrations were found for PCB-153, PCB-138, and PCB-180, which is in line with previous findings concerning the POP levels in fish species from the Adriatic Sea (Baptista et al., 2013; Storelli et al., 2012). The results reflect the high stability and persistence of the pollutant. A dominant presence of β-HCH in the HCH group is in accordance with its stability in the aquatic environment compared to other isomers. Some researchers suggest using the ratio *p,p'*-DDT/*p,p'*-DDE as an indicator of fresh inputs of *p,p'*-DDT in the organism. In the studied fish species, the *p,p'*-DDE/ *p,p'*-DDT ratio ranged from 2.4 to 5.6 implying that the occurrence of these compounds may be associated with their previous use rather than the recent exposure. The values of *p,p'*-DDE/ *p,p'*-DDT ratio in the Adriatic Sea were discussed in detail in our previous research (Kljaković-Gašpić et al., 2010; Kljaković Gašpić et al., 2015). The median concentrations of PCB-118 and PCB-170 were higher than those of the other toxicologically significant congeners. Distinctions in POP occurrences among anchovy, horse mackerel, round sardinella, chub mackerel, and sardine were discussed in our previous study, where the following pattern of congener accumulation was suggested: anchovy < chub mackerel<horse mackerel<sardines (Vuković et al., 2018). The values of corresponding toxicological parameters, such as the sum of indicator PCBs (ICES-6) and WHO-dioxin-like PCBs toxic equivalents, were notably below the threshold levels of 75 ng g^{-1} w.w. and 6.5 pg g^{-1} (European Commission, 2011). The pollutant distribution found in this study represents a likely profile of pollutant occurrence reported as in previous research (Storelli et al., 2012).

Globally, marine environments are the final sink of pollutants, which are polluted by coastal outfalls, rivers, and direct or long-ranged deposition from the atmosphere. Oceans and deep seas act as a secondary source of POPs because these compounds are slowly degraded and adsorbed on suspended particles or bioaccumulated in benthic marine organisms/pelagic fish species, which are at the bottom of the food chain. The assessment of environmental exposure to marine toxins and the uptake of POPs from different sources depend on the sampling time, season, other fishery zone-related factors, surrounding conditions (salinity, temperature), and the physicochemical properties of the pollutant and biotic factors (e.g., organisms' trophic position). In this study, for source apportionment, the Unmix receptor model was used. Unmix identified five source profiles (Table S4, Supplementary data), which contributed to the POP burden in the studied fish species that could be associated with both diffuse sources such as the overlapping diet, which contains low-niche marine organisms and pollutant inputs from the surrounding environment. The studied fish species live in the epipelagic zone of the Adriatic Sea, which extends from the surface to 200 m depth. The species show similar trophodynamics with certain differences in nutrition cycles. However, the main prey of all small pelagic fish includes the following: calanoid copepods, fish larvae, decapod larvae, copepod eggs, amphipods, and isopods (Hure and Mustać, 2020). The highest Unmix shares were attributed to the second

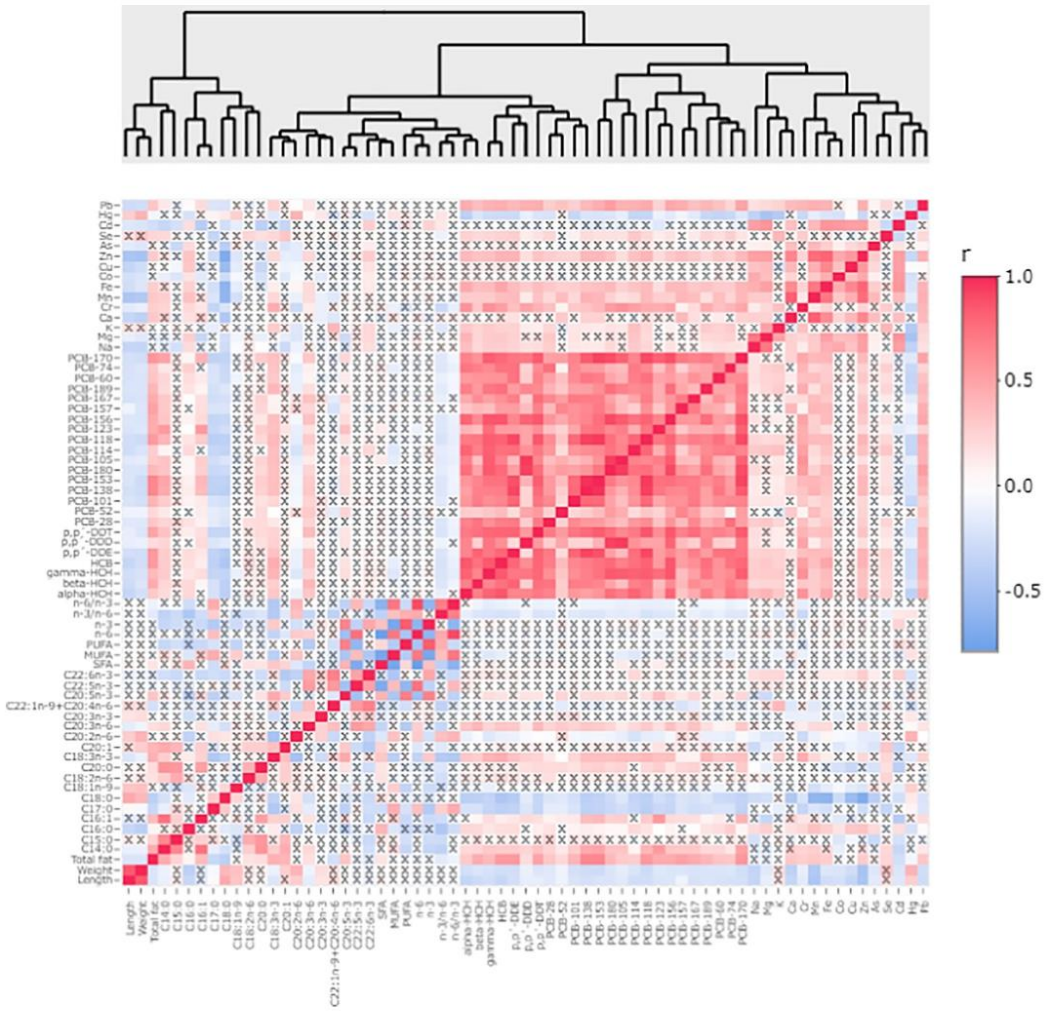


Fig. 4. Correlation matrix (p-value: <0.05) for the fatty acids, organochlorine pollutants and elements in the marine fish species (anchovy, sardines, chub, round sardinella and horse mackerel).

(31%) source which was dominated by the presence of toxicology relevant PCBs, particularly, PCB-114 (71%). Metabolic derivatives of *p,p′*-DDT, *p,p′*-DDD, *p,p′*-DDE, and γ-HCH contributed to this source, also reflecting the metabolic degradation of toxicology relevant persistent xenobiotics. The third source (23%) was characterized by the dominance of the long-lived penta- to hepta-chloro congeners, including indicator PCBs. The fourth (19%) and the first (12%) Unmix profile corresponded to the presence of more volatile short-living organochlorines as indicated by the dominant shares of γ-HCH, β-HCH, α-HCH, HCB, PCB-74, PCB-28, and PCB-60. The inorganic components were identified as the fifth source (15%) and included Cu, Na, Mn, and Ca, which are naturally found in the marine environment in the form of dissolved salts or as a part of ocean sediment.

Microelements (Cu and Mn) and macroelements (Na and Ca) were recognized by the SOM analyses as pronounced contaminants, which are distinguished among fish species when accumulated(Fig. 2). A highly variable level of macroelements was found in all species throughout all the sampling years as projected by the upper and low left-oriented SOM clusters (Fig. 2a). ANOVA results (Table S5, Supplementary data) confirmed that the concentrations of K, Na, Ca, and Mg were significantly different ($p < 0.05$) within certain fish species. Although it is expected that larger, long-lived species that are at higher trophic levels, such as chub mackerel and horse mackerel, will uptake more contaminants than sardines and/or anchovy, the expectation was not approved in this study. The result implies that the element uptake is not only species-dependent, but also vary because of other abiotic factors as discussed in previous paragraphs. The trace elements map and SOM clusters, which are extracted from the bottom of the map and have a

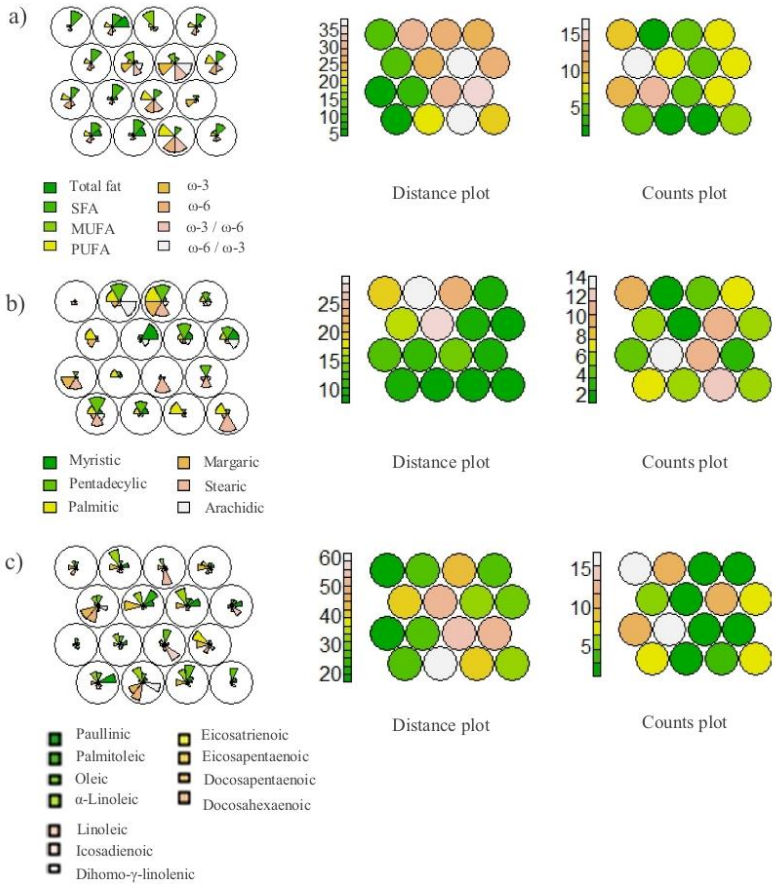


Fig. 5. Self-organizing maps, distance and count plots of marine fish species (anchovy, sardines, chub mackerel, round sardinella, and horse mackerel) classified based on a) fatty acid parameters, b) content of saturated fatty acids, and c) content of unsaturated fatty acids; warmer colour at distance plot means stronger differences between clusters, while counts plot represent the number of sample in each cluster.

distance value equal to or higher than 25, proved that the uptake of Cr, Mn, Zn, and Se is the most distinctive in sardines(Fig. 2b). However, it was noted that the greatest number of samples were grouped in the most similar clusters as illustrated by the counts plot (>10) and distance values (<15). As indicated by ANOVA (Table S5), variabilities in bioaccumulation among the studied species are attributable to the content of As, which dominated in mackerel, along with Zn and Fe, which were the most abundant in sardines and anchovy. Iamiceli et al. (2015) pointed interspecific variability of As in Mediterranean aquatic species, and that environmental conditions may increase the element contents in organisms. In addition, Bilandić et al. (2011) reported differences of As, Cd, Cu, Hg, and Pb among anchovy, mackerel, red mullet, and picarel from the Croatian Adriatic (Mediterranean Sea).

As represented by SOM clusters and uniformly coloured correspondent distance plots (value <50), the accumulation of chlorinated pesticides in the fish tissue was similar among species even when we consider the samples that exhibited maximum loads (the left upper nodes, Fig. 3a). Although total concentrations of indicator PCB congeners overshadowed those of toxicologically relevant PCBs, indicator compounds appeared to be insignificant in the evaluation of the differences in total POPs burden in studied marine fish species (Fig. 3b). The observation was accompanied by ANOVA results and confirm our previous findings based on a smaller dataset (Vuković et al., 2018). Regardless of the sampling year, sardines were isolated (the maps' upper right clusters, Fig. 3c) as the most important collector of toxicologically relevant PCBs. The indication could be of importance as sardines have a notable role in the human diet because it is protein-rich and has low fat composition. However, this observation was based on a limited dataset (the count plot value < 5) and should be further verified. The ANOVA results (Table S5) did not record toxicologically relevant congeners as important predictors to make a distinction between individual species. As demonstrated by Pearson's correlation analysis (Fig. 4), the correlation coefficients ($r > 0.80$) implied a relatively strong or very strong linear linkage between the studied organochlorines. A relatively high correlation ($r = 0.80–0.90$) was recognized between PCB-170 and *p,p'*-DDT, PCB-138, PCB-180, PCB-105, PCB-123, PCB-156; PCB-156 and γ-HCH, HCB, *p,p'*-DDD, *p,p'*-DDT, PCB-105, PCB-118; α-HCH and PCB-114; and *p,p'*-DDE and PCB-153. Very high associations ($r > 0.90$)

were found between the following pairs of congeners: PCB-170 and -153, -118, -105; PCB-118 and *p,p'*-DDE, *p,p'*-DDT, -138, -153; PCB-180 and *p,p'*-DDE, -156, -105; PCB-105 and *p,p'*-DDE, *p,p'*-DDT, -156; and PCB-138 and *p,p'*-DDE, -153. The pollutants belonged to both indicators (-138, -153, -180) and toxicologically relevant PCBs (-105, -118, -156). Compounds from both groups and PCB-170 possess chlorine atom at the ortho position implying that a similar chemical structure governs the PCB profiles in marine fish regardless of their toxicological properties. Moreover, similar pollution sources or metabolic pathways contribute to the relations between "heavy" hexa- to hepta-chloro congeners and "light" tetra- to penta-chloro PCBs, and HCH and DDT compounds. No significant Pearson's correlation coefficients were observed between the 18 FAs and lipophilic organochlorines, which indicate that methods other than the typical correlation matrices could be used to describe the associations between POPs and FAs.

According to ANOVA output (Table S5), myristic, palmitic, oleic, stearic acids, EPA, and DHA differed among the studied fish species. In addition to the listed FAs, the content of docosapentaenoic acid was different among sardine samples and palmitoleic acid in mackerels and sardines. The content of saturated and ω-FAs appeared to be the most distinctive among marine fish as indicated by SOM clusters grouped on the right side (Fig. 5a). Distance plots, which are characterized by the lightest colour and a value higher than 30, indicated strong dissimilarities among the extracted clusters, whereas the count plots showed that the result is based on a small number of samples (<10) (Fig. 5a). Regarding saturated FAs, the highest levels of margaric, stearic, palmitic and pentadecyl acids shaped the most dissimilar clusters (middle top of the SOM plot, distance plot value >25) and mainly highlighted anchovies as species rich in these FAs (Fig. 5b). The influences of myristic and arachidic acids were far lower. Out of the unsaturated FAs, oleic, eicosatrienoic acid, and DHA mostly contributed to the differences between species as indicated by clusters in the middle right bottom and distance plot value higher than 50 (Fig. 5c). In addition, distinctions were prevalently attributable to the high content of these acids in the tissue of some chub mackerel and anchovy samples (approximately 5). The results confirmed that the composition of FAs is species-specific and dependent on dietary, geographical, and environmental factors, including the reproductive cycle of marine fish and the fishery period (Metillo and Aspiras-Eya, 2014).

In conclusion, this study provides extensive data on the content of organic and inorganic contaminants and FAs in small pelagic fish (anchovy, sardines, chub mackerel, round sardinella, and horse mackerel) along with insights into their interrelations. Inorganic compounds, macroelements, and toxic heavy metals were notably more abundant in fish tissue than organic xenobiotics, OCPs, and PCBs. According to the toxicological parameter values, the examined fish species appear to be safe for human consumption and represent a valuable source of nutritionally beneficial FAs. The results could be of importance for future research of risk-benefit assessment, since studied edible fish species can be considered both a nutritionally valuable food source and a source of hazardous organochlorine pollutants, which negatively affect human health. Nevertheless, our results provide valuable data for the future evaluation of ecological/environmental profiles of the Mediterranean Sea.

CRediT authorship contribution statement

Snježana Herceg Romanić: Conceptualization, Investigation, Resources, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Gordana Jovanović:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision. **Bosiljka Mustać:** Investigation, Funding acquisition. **Jasna Stojanović-Đinović:** Investigation, Resources. **Andreja Stojić:** Methodology, Formal analysis, Visualization. **Tena Čadež:** Investigation. **Aleksandar Popović:** Resources, Writing – original draft, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the institutional financing of scientific activity 2018–2020 Project "Persistent Organic Pollutants - Environmental Impact Assessment and Stability of Human Genetic Material" and the funding was provided by the Institute of Physics, Belgrade, through research supported by the Ministry of Education, Science, and Technological Development.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2021.112654.

References

Ancora, S., Mariotti, G., Ponchia, R., Fossi, M.C., Leonzio, C., Bianchi, N., 2020. Trace elements levels in muscle and liver of a rarely investigated large pelagic fish: the Mediterranean spearfish Tetrapturus belone (Rafinesque, 1810). Mar. Pollut. Bull. 151, 110878 https://doi.org/10.1016/j.marpolbul.2019.110878.

Baptista, J., Pato, P., Pereira, E., Duarte, A.C., Pardal, M.A., 2013. PCBs in the fish assemblage of a southern European estuary. J. Sea Res. 76, 22–30. https://doi.org/10.1016/j.seares.2012.09.011.

Bartalini, A., Muñoz-Arnanz, J., Baini, M., Panti, C., Galli, M., Giani, D., Fossi, M.C., Jiménez, B., 2020. Relevance of current PCB concentrations in edible fish species from the Mediterranean Sea. Sci. Total Environ. 1 (737), 139520. https://doi.org/10.1016/j.scitotenv.2020.139520.

Bilandžić, N., Đokić, M., Sedak, M., 2011. Metal content determination in four fish species from the Adriatic Sea. Food Chem. 124, 1005–1010. https://doi.org/10.1016/j.foodchem.2010.07.060.

Budayan, C., Dikmen, I., Birgonul, M.T., 2009. Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. Expert Syst. Appl. 36, 11772–11781. https://doi.org/10.1016/j.eswa.2009.04.022.

European Commission, 2006. Commission Regulation (EC) No. 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs. Official Journal of the European Union L 364, 5–24.

European Commission, 2008. Commission Regulation (EC) No. 629/2008 of 2 July 2008 amending Regulation (EC) No. 1881/2006 setting maximum levels for certain contaminants in foodstuffs. Official Journal of the European Union L 173, 6–9.

European Commission, 2011. Commission Regulation (EU) No 1259/2011 of 2 December 2011 amending Regulation (EC) No 1881/2006 as regards maximum levels for dioxins, dioxin-like PCBs and non-dioxin-like PCBs in foodstuffs. European Union 320, 18–23.

Food and Agriculture Organization of the United Nations, FAO, 2018a. The State of Mediterranean and Black Sea Fisheries. General Fisheries Commission for the Mediterranean. Rome. 172 pp. Licence: CC BY-NC-SA 3.0 IGO.

Food and Agriculture Organization of the United Nations, FAO, 2018b. The State of World Fisheries and Aquaculture Worldwide 2018-Meeting the Sustainable Development Goals. 1–227.

Herceg Romanić, S., Vuković, G., Klinčić, D., Antanasijević, D., 2018. Self-organizing maps for indications of airborne polychlorinated biphenyl (PCBs) and organochlorine pesticide (OCPs) dependence on spatial and meteorological parameters. Sci. Total Environ. 628-629, 198–205. https://doi.org/10.1016/j.scitotenv.2018.02.012.

Hure, M., Mustać, B., 2020. Feeding ecology of Sardina pilchardus considering co-occurring small pelagic fish in the eastern Adriatic Sea. Mar. Biodivers. 50, 40 https://doi.org/10.1007/s12526-020-01067-7.

Iamiceli, A., Ubaldi, A., Lucchetti, D., Brambilla, G., Abate, V., De Felip, E., DeFilippis, S. P., Dellatte, E., De Luca, S., Ferri, F., Fochi, I., Fulgenzi, A., Iacovella, N., Moret, I., Piazza, R., Roncarati, A., Melotti, P., Fanelli, R., Fattore, E., diDomenico, A., Miniero, R., 2015. Metals in Mediterranean aquatic species. Mar.Pollut. Bull. 94, 278–283. https://doi.org/10.1016/j.marpolbul.2015.02.034.

Kljaković Gašpić, Z., Herceg Romanić, S., Klinčić, D., Tičina, V., 2015. Chlorinated compounds in the muscle tissue of fish from the Croatian Adriatic: preliminary data on contamination. Arhiv za Higijenu Rada i Toksikologiju 66, 299–308. https://doi.org/10.1515/aiht-2015-66-2667.

Kljaković-Gašpić, Z., Herceg Romanić, S., Kožul, D., Veža, J., 2010. Biomonitoring of organochlorine compounds and trace metals along the Eastern Adriatic coast (Croatia) using Mytilus galloprovincialis. Mar. Pollut. Bull. 60, 1879–1889. https://doi.org/10.1016/j.marpolbul.2010.07.019.

Kohonen, T., 2013. Essentials of the self-organizing map. Neural Netw. 37, 52–65.

Lundebye, A.K., Lock, E.J., Rasinger, J.D., Nøstbakken, O.J., Hannisdal, R., Karlsbakk, E., Wennevik, V., Madhun, A.S., Madsen, L., Graff, I.E., Ørnsrud, R., 2017. Lower levels of persistent organic pollutants, metals and the marine omega 3-fatty acid DHA in farmed compared to wild Atlantic salmon (Salmo salar). Environ. Res. 155, 49–59. https://doi.org/10.1016/j.envres.2017.01.026.

Mari, M., Nadal, M., Schuhmacher, M., Domingo, J.L., 2010. Applications of self-organizing maps for PCDD/F pattern recognition of environmental and biological samples to evaluate the impact of a hazardous waste incinerator. Environ. Sci. Technol. 44, 3162–3168. https://doi.org/10.1021/es1000165.

Metillo, E.B., Aspiras-Eya, A.A., 2014. Fatty acids in six small pelagic fish species and their crustacean prey from the Mindanao Sea, Southern Philippine. Tropical life sciences research 25 (1), 105–115.

Nasir, M., Bloch, M.H., 2019. Trim the Fat: The Role of Omega-3 Fatty Acids in Psychopharmacology. SAGE Publications, Therapeutic Advances in Psychopharmacology, p. 9. https://doi.org/10.1177/2045125319869791.

Nikolić, D., Đinović-Stojanović, J., Janković, S., Stanišić, N., Radović, C., Pezo, L., Laušević, M., 2017. Mineral composition and toxic element levels of muscle, liver and kidney of intensive (Swedish Landrace) and extensive (Mangulica) pigs from Serbia. Food Additives and Contaminants Part A 34 (6), 962–971. https://doi.org/10.1080/19440049.2017.1310397.

Palomera, I., Olivar, M.P., Salat, J., Sabatés, A., Coll, M., García, A., Morales-Nin, B., 2007. Small pelagic fish in the NW Mediterranean Sea: an ecological review. Prog. Oceanogr. 74, 377–396. https://doi.org/10.1016/j.pocean.2007.04.012.

Spirić, A., Trbović, D., Vranić, D., Đinović, J., Petronijević, R., Matekalo-Sverak, V., 2010. Statistical evaluation of fatty acid profile and cholesterol content in fish (common carp) lipids obtained by different sample preparation procedures. Anal. Chim. Acta 672, 66–71. https://doi.org/10.1016/j.aca.2010.04.052.

Sprague, M., Dick, J.R., Medina, A., Tocher, D.R., Bell, J.G., Mourente, G., 2012. Lipid and fatty acid composition, and persistent organic pollutant levels in tissues of migrating Atlantic bluefin tuna (Thunnus thynnus, L.) broodstock. Environ. Pollut. 171, 61–71. https://doi.org/10.1016/j.envpol.2012.07.021.

Storelli, M.M., Perrone, V.G., Busco, V.P., Spedicato, D., Barone, G., 2012. Persistent organic pollutants (PCBs and DDTs) in European conger eel, Conger conger L., from the Ionian Sea (Mediterranean Sea). Bull. Environ. Contam. Toxicol. https://doi.org/10.1007/s00128-012-0606-y.

Team, 2012. A Language and Environment for Statistical Computing. http://cran.case.edu/web/packages/dplR/vignettes/timeseries-dplR.pdf.

United Nations Environment Programme / Mediterranean Action Plan (UNEP/MAP), 2012. State of the Mediterranean Marine and Coastal Environment. UNEP/MAP-Barcelona Convention, Athens.

United States Environmental Protection Agency (USEPA), 2007. EPA Unmix 6.0 Fundamentals and User Guide. USEPA Office of Research and Development.

Vuković, G., Romanić, S.H., Babić, Ž., Mustać, B., Štrbac, M., Deljanin, I., Antanasijević, D., 2018. Persistent organic pollutants (POPs) in edible fish species from different fishing zones of Croatian Adriatic. Mar. Pollut. Bull. 137, 71–80. https://doi.org/10.1016/j.marpolbul.2018.10.014.

Wehrens, R., Buydens, L.M.C., 2007. Self- and super-organizing maps in R: the kohonen package. J. Stat. Softw. (5), 1–19. https://doi.org/10.18637/jss.v021.i05.

Wehrens, R., Kruisselbrink, J., 2017. Supervised and unsupervised self-organising maps. In: Package "Kohonen". https://cran.r-project.org/web/packages/kohonen/kohonen.pdf.

World Health Organisation, WHO, 2014. Dioxins and their effects on human health. In: World Health Organization Fact Sheet, No. 225 (Accessed date: April 2021).

Zhang, I., Wong, M.H., 2007. Environmental mercury contamination in China: sources and impacts. Environ. Int. 33, 108–121. https://doi.org/10.1016/j.envint.2006.06.022.

Environmental Research 193 (2021) 110520

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources

Svetlana Stanišić [a,*], Mirjana Perišić [a,b], Gordana Jovanović [a,b], Tijana Milićević [b], Snježana Herceg Romanić [c], Aleksandar Jovanović [b], Andrej Šoštarić [d], Vladimir Udovičić [b], Andreja Stojić [a,b]

[a] Singidunum University, 32 Danijelova Street, Belgrade, 11000, Serbia
[b] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 118 Pregrevica Street, 11000, Belgrade, Serbia
[c] Institute for Medical Research and Occupational Health, 2 Ksaverska Cesta Street, PO Box 291, 10001, Zagreb, Croatia
[d] Institute of Public Health Belgrade, 54 Despota Stefana Street, 11000, Belgrade, Serbia

ARTICLE INFO

Keywords:
Indoor/outdoor air quality
Polycyclic aromatic hydrocarbons
Source apportionment
XGBoost method
Explainable artificial intelligence

ABSTRACT

The previous research, aimed at exploring the relationships between the indoor and outdoor air quality, has evidenced that outdoor $PM_{2.5}$-bound polycyclic aromatic hydrocarbons (PAH) levels exhibit significant daily and seasonal variations which does not necessary corresponds with PAH indoor dynamics. For the purpose of this study, a three-month measurement campaign was performed simultaneously at indoor and outdoor sampling sites of a university building in an urban area of Belgrade (Serbia), during which the concentrations of O_3, CO, SO_2, NO_x, radon, $PM_{2.5}$ and particle constituents including trace metals (As, Cd, Cr, Mn, Ni and Pb), ions (Cl^-, Na^+, Mg^{2+}, Ca^{2+}, K^+, NO_3^-, SO_4^{2-} and NH_4^+) and 16 US EPA priority PAHs were determined. Additionally, the analysis included 31 meteorological parameters, out of which 24 were obtained from Global Data Assimilation System (GDAS1) database. The Unmix and PAH diagnostic ratios analysis resolved the source profiles for both indoor and outdoor environment, which are comparable in terms of their apportionments and pollutant shares, although it should be emphasized that ratio-implied solutions should be taken with caution since these values do not reflect emission sources only. The highest contributions to air quality were attributed to sources identified as coal combustion and related pyrogenic processes. Noticeable correlations were observed between 5- and 6-ring high molecular weight PAHs, but, except for CO, no significant linear dependencies with other investigated variables were identified. The PAH level predictions in the indoor and outdoor environment was performed by using machine learning XGBoost method.

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a complex group of pollutants generated during incomplete combustion of organic material. Only a small quantity of PAHs in the atmosphere originates from natural sources such as volcanic emissions, forest, and grassland fires. Their origin in ambient air of urban areas is associated with many emission sources, including fossil-fuel burning for power generation, transportation and heating, while in indoor environment, elevated concentrations of PAHs are related to tobacco smoke, the use of gas, coal or electric stoves, and candle burning (Gao et al., 2019). PAHs have received most attention as a major public health concern globally because a vast number of studies has confirmed that the increased exposure to high-molecular weight (5-ring and more) PAHs is associated with altered mitochondrial dynamics and cumulative oxidative cellular damage (Brehmer et al., 2020). Eight PAHs have been classified by USEPA (1997) as carcinogenic or potentially carcinogenic, including benz[a]anthracene, chrysene, benzo[a]pyrene, benzo[b]fluoranthene, benzo[k]fluoranthene, dibenz[a,h]anthracene, indeno[1,2,3-cd]pyrene, and benzo[g,h,i]perylene. Among them, benzo[a]pyrene belongs to Group 1 of hazardous species – carcinogenic to humans (International Agency for Research on Cancer, 2012), and its emissions are regulated

* Corresponding author.
E-mail addresses: sstanisic@singidunum.ac.rs (S. Stanišić), mirjana.perisic@ipb.ac.rs (M. Perišić), gordana.vukovic@ipb.ac.rs (G. Jovanović), tijana.milicevic@ipb.ac.rs (T. Milićević), sherceg@imi.hr (S.H. Romanić), aleksandar.jovanovic@ipb.ac.rs (A. Jovanović), andrej.sostaric@zdravlje.org.rs (A. Šoštarić), vladimir.udovicic@ipb.ac.rs (V. Udovičić), andreja.stojic@ipb.ac.rs (A. Stojić).

https://doi.org/10.1016/j.envres.2020.110520
Received 12 July 2020; Received in revised form 17 November 2020; Accepted 20 November 2020
Available online 28 November 2020

by the Directive 2004/107/Elie et al. (2015). Additionally, benzo[a] anthracene is assigned as probable carcinogenic to humans (Group 2A), and chrysene, benzo[b]fluoranthene, benzo[k]fluoranthene, dibenz[a, h]anthracene, and indeno[1,2,3-cd]pyrene as possibly carcinogenic to humans (Group 2B) (International Agency for Research on Cancer, 2012). The reactive PAH metabolites, emitted from combustion processes or formed in the heterogeneous reactions with atmospheric oxidants are also evidenced to have mutagenic and genotoxic potential (Elie et al., 2015).

Due to their sources and physico-chemical characteristics, PAHs are ubiquitous, found in all environmental compartments. Partitioning of PAHs between the gas and particulate phase has an important impact on their atmospheric fate, transport and chemical transformations of these compounds. It determines the extent of photo-degradation and photo-oxidation, and relative amounts of deposition that can occur by wet scavenging, dry deposition of particles, and by gas exchange between the air and water interface (Tasdemir and Esen, 2007). Generally, low-molecular weight PAHs are presented in a gas phase, while high-molecular weight PAHs are either sorbed to airborne coarse PM fraction and deposited close to the emission sources, or more often bound to fine suspended particles, $PM_{2.5}$ (Alves et al., 2014), which persist for a longer time in the atmosphere and penetrate deeper in the respiratory system. Beside the chemical complexity of PAHs, the gas/-particle partitioning of these species also depends on their vapor pressure and concentrations, ambient temperature and type of particles present in the atmosphere (Keyte et al., 2013).

The previous studies have shown that the concentrations of PAHs can be found within the relatively wide range of values and show large variations on a daily and seasonal basis, depending on the vicinity and strength of sources and sinks (Pehnec et al., 2020). Since the major emission source of PAHs is fossil fuel burning, the reports mostly confirm that PAH concentrations reach their maximum during cold season, both indoors and outdoors (Majd et al., 2019; Sarigiannis et al., 2015). Besides increased emissions associated with residential heating, PAH levels in winter season are expected to be high due to reduced vertical air mixing caused by inversion, less intensive atmospheric reactions and enhanced sorption to particles at lower temperature, as a result of reduced vapor pressure and/or shifting in the gas/particle distribution induced by ambient temperature variations (Ravindra et al., 2006).

A number of studies have also reported a strong correlation between indoor and outdoor concentrations of PAHs, as well as their indoor-to-outdoor (I/O) ratio <1, which indicates that indoor PAHs mostly originate from the outdoor environment (Krugly et al., 2014; Uchiyama et al., 2015). In these cases, the contributions of outdoor sources to indoor air concentrations were expected to follow seasonal variations, as shown in the study of Shi (2018), who estimated that indoor benzo(a) pyrene concentration in typical Beijing residence were outdoor source-affected by 72.3%, 60.3%, 65.2% and 82.9% in spring, summer, autumn and winter, respectively. However, there is also a number of studies which reported the I/O ratios of PAHs exceeding 1. For instance, the combustion of bituminous coal and unprocessed biomass in households in China, India, and other rural regions of Asia, remains the major source of indoor air pollution and PAHs, being present both in particle and gas phase (Wang et al., 2015b; Yao et al., 2019).

In this study, we have identified potential emission sources and investigated the relationships between meteorological parameters and indoor and outdoor O_3, CO, SO_2, NO_x, radon, $PM_{2.5}$ and particle constituents including trace metals (As, Cd, Cr, Mn, Ni and Pb), ions (Cl^-, Na^+, Mg^{2+}, Ca^{2+}, K^+, NO_3^-, SO_4^{2-} and NH_4^+) and 16 US EPA priority PAHs were determined, simultaneously collected at indoor and outdoor sites of a university building, located in the urban area of Belgrade (Serbia). At this location, lectures are visited by approximately 4000 students in total, and the indoor air sampling was conducted in an amphitheater having the capacity of 350 people. For this purpose, we used Unmix and eXtreme Gradient Boosting, the method that is highly adaptive to non-parametric data distributions, less sensitive to error term assumptions, and tolerable to noise, chaotic components and heavy tails (Šoštarić et al., 2017; Stojić et al., 2019). As shown, the PAH level predictions in the indoor and outdoor environment were successfully performed by using machine learning XGBoost method and the obtained results will be considered by using explainable artificial intelligence methods in the succeeding parts of this paper.

2. Materials and methods

2.1. Study variables

For the purpose of this study, a three-month (March 1st – May 31st) measurement campaign was performed simultaneously at indoor and outdoor sampling sites, during which the concentrations of inorganic gaseous pollutants, radon, $PM_{2.5}$ and particle constituents including trace metals (As, Cd, Cr, Mn, Ni, and Pb), ions (Cl^-, Na^+, Mg^{2+}, Ca^{2+}, K^+, NO_3^-, SO_4^{2-}, and NH_4^+) and 16 US EPA PAHs were regularly analyzed. For the analysis, several meteorological parameters were registered, including outdoor ambient air temperature, outdoor relative humidity, outdoor air pressure, wind and rain characteristics, indoor ambient air temperature, indoor relative humidity and indoor air pressure, while 24-parameter data were additionally obtained from Global Data Assimilation System (GDAS1). Additionally, the number of people in an amphitheater and the time they spent indoor was registered hourly.

2.2. Study area

Air sampling was performed at the rooftop and inside of the Singidunum University building (44°45′33.8″N, 20°29′47.6″E). At this location, lectures are visited by approximately 4000 students in total. The indoor air sampling was conducted in an amphitheater having the capacity of 350 people, and the number of students in the amphitheater during the study period most often ranged from 50 to 80. The University building is surrounded by large residential areas from W, SW and NE side, some of which encompass households with individual fireboxes, while small scale industry referring to the Road Institute of Belgrade, a building company and beverage factory stockroom are located in the nearest vicinity. Additionally, confectionery factory, footwear factory, and several small-scale chemical plants are located 600 m in the NW and S direction, respectively. Around 800 m to the W and SW from the measurement site a large district heating plant and fuel oil heating plant of urban forestry organization, used for the purposes of planting material production, are situated. A boulevard with public transport and moderate vehicle flow passes by approximately 250 m in the SW direction, while a road with intense traffic is about 500 m away in the W–NW direction. The old city center and river confluence are located at the distance exceeding 2 km in the NW direction, while the air quality at the sampling site was occasionally affected by the emissions from two large city municipalities situated just across the river.

2.3. Experimental settings

The outdoor $PM_{2.5}$ and air sampling, as well as meteorological measurements were performed at the rooftop of the building, around 10 m above the ground. The indoor air sampling inlet and $PM_{2.5}$ sampling device were placed at a height of 6 m and 2 m from the floor, respectively.

Air sampling system comprised diaphragm vacuum pump Pfeiffer MVP and manifolds with openings for measuring inorganic gaseous pollutants (O_3, CO, SO_2, and NO_x) by using Horiba APOA, APMA, APSA, and APNA, 370 series, and electronically controlled valves, which operated in alternating indoor/outdoor air sampling mode in the 10-min cycles.

The $PM_{2.5}$ sampling was performed by using Svan Leckel LVS6-RV devices operating at a nominal flow rate of 2.3 m^3 h^{-1}, over 24 h-

sampling period. The concentrations of $PM_{2.5}$ and their constituents, including trace metals, ions and PAHs were determined at the reference laboratory of the Institute of Public Health of Belgrade. The limit of detection was 1 μg m^{-3}.

The outdoor meteorological data were obtained by using Vaisala WXT530 monitoring station set at the building rooftop, while the indoor radon concentrations, ambient air temperature, relative humidity and air pressure were detected by SN1029 radon monitor (Sun Nuclear Corporation, NRSB approval code 31822) and corresponding integrated sensor devices, placed in the center of the amphitheater, at a height of 1 m from the floor.

2.4. Chemical analysis

$PM_{2.5}$ was collected on quartz filters (Whatman QMA, 47 mm) daily, each morning before the start of daily indoor activities and weighted, as described in the Standard SRPS EN 12341:2015 (Ambient air - Standard gravimetric measurement method for the determination of the PM_{10} or $PM_{2.5}$ mass concentration of suspended particulate matter, 2015). The filters were pre-fired to remove organic impurities, and the pre-conditioning of both non-exposed and loaded filters was performed prior to gravimetric measurements. After gravimetric measurements, the surface of each filter, amounting to 13.85 cm^2, was cut in two pieces – approximately 1.76 cm^2 each, which were used for the analysis of anions and cations, while the remaining 12.09 cm^2 were divided and used for the analysis of trace elements and 16 US EPA PAHs.

For ion concentration measurements, the sample pieces underwent an ultra-pure water extraction for 24 h and the aqueous extracts were further analyzed by standard ion chromatography using a Dionex DX-500 IC system according to the MDL 064 Standard operating procedure. The method detection limits are presented in Table 2.

The concentrations of As, Cd, Cr, Ni, and Pb as $PM_{2.5}$ constituents were determined as described in the SRPS EN 14902:2008/AC:2013 Standard (Ambient air quality - Standard method for the measurement of Pb, Cd, As and Ni in the PM fraction of suspended particulate matter, 2008). Firstly, CEN/TC 264 N779 procedure was applied for the extraction of the trace elements. In brief, the pieces of exposed quartz filters were treated with an acidic mixture of HNO_3(c)/30% H_2O_2/H_2O (3/2/5) using analytical grade reagents (Merck) and distilled/deionized water (MiliQ, 18.2 MΩ). The filters were digested in closed 100 ml Teflon vessels in the Anton Paar 3000 microwave accelerated reaction system and the concentrations of trace elements were determined by inductively coupled plasma–mass spectrometry (ICP-MS) (device Agilent 7500ce with Octopole Reaction System). Quality control and verification of the applied procedures for microwave digestion and multi-elemental trace analysis using ICP-MS was conducted by 2783 NIST (National Institute of Standard and Technology, MD, USA) standard reference material analysis containing a $PM_{2.5}$ fraction of urban dust from a mixed industrial urban area of Vienna, collected on a polycarbonate membrane filter. The recovery values were within satisfactory range of ±20% from the reference value. The method detection limits are presented in Table 2.

Sixteen US EPA priority PAHs including naphthalene (Nap), acenaphthylene (Ace), acenaphthene (Ane), fluorene (Flu), phenanthrene (Phe), anthracene (Ant), fluoranthene (Fla), pyrene (Pyr), benz[a] anthracene (B[a]A), chrysene (Chy), benzo[b]fluoranthene (B[b]f), benzo[k]fluoranthene (B[k]F), benzo[a]pyrene (B[a]P), dibenz[a,h] anthracene (Db[ah]A), benzo[g,h,i]perylene (B[ghi]P), and indeno [1,2,3-cd]pyrene (In[cd]P) were determined by the procedure described in the SRPS ISO 12884:2010 Standard (Ambient air — Determination of total (gas and particle-phase) polycyclic aromatic hydrocarbons — Collection on sorbent-backed filters with gas chromatographic/mass spectrometric analyses, 2010).

Parts of the exposed filters underwent microwave extraction procedure with a solvent mixture of n-hexane and acetone (12.5 ml n-hexane:12.5 ml acetone) according to EPA method 3546. After extraction, solution volume was reduced by rotary evaporation under reduced pressure (55.6 kPa and with 0.2 ml isooctane) to 1 ml. Afterwards, the n-hexane solution was reduced to 0.25 ml under a nitrogen stream. Known quantities of internal standards were added to estimate the method recovery. PAHs were analyzed using gas chromatography coupled with mass selective detector (Agilent GC 6890/5973 MSD) according to EPA compendium method TO-13A with a DB-5 MS capillary column (30 m × 0.25 mm × 25 μm). The oven temperature program started at 70 °C (duration 4 min), ramp 8 °C min^{-1} to the end temperature of 310 °C (duration 5 min). The solvent delay was 5 min and the time of run was 46 min. The calibration curves for all 16 PAHs were obtained by spiking seven different quantities of each PAH, all with an R^2 of the calibration curve above 0.995. Recovery values ranged from 85% to 110% for all the PAHs contained in the internal standard. The method detection limits are presented in Table 2.

Inorganic gaseous pollutant indoor and outdoor measurements were conducted by using Horiba 370 series devices which enabled continual pollutant concentration monitoring with a 2 min-resolution data and detection limit of 1 μg m^{-3} for all species except of CO with detection limit of 0.1 mg m^{-3}. More specifically, the CO concentrations were determined by non-dispersion cross modulation infrared spectroscopy method using APMA-370 device, as described in the SRPS EN 14626:2013 Standard. The concentrations of SO_2 were measured by UV fluorescence method using APSA-370 device, as described in the SRPS EN 14212:2013/AC:2015 Standard. The APNA-370 device was used for NO, NO_2, and NO_x concentration measurements by a combination of dual cross-flow modulation type chemiluminescence principle and the referential calculation method according to the SRPS EN 14211:2013

Table 1
Outdoor meteorological data used in analyses.

Outdoor meteorological data abbreviation	Origin	Meaning
WD	Vaisala	Wind direction
WS	Vaisala	Wind speed
Temp	Vaisala	Temperature
Rh	Vaisala	Relative humidity
Pressure	Vaisala	Pressure
Rain duration	Vaisala	Rain duration
Rain total	Vaisala	Rain intensity
Prss	GDAS1	Pressure at surface
Mslp	GDAS1	Pressure reduced to mean sea level
Tpp6	GDAS1	Accumulated precipitation (6 h accumulation)
Mofi	GDAS1	momentum flux intensity (3- or 6-h average)
Mofd	GDAS1	momentum flux direction (3- or 6-h average)
Shif	GDAS1	Sensible heat net flux at surface (3- or 6-h average)
Dswf	GDAS1	Downward short-wave radiation flux (3- or 6-h average)
Rh 2 m	GDAS1	Relative Humidity at 2m AGL
WD 10 m	GDAS1	wind direction at 10 m AGL
WS 10 m	GDAS1	wind speed at 10 m AGL
T0 2 m	GDAS1	Temperature at 2m AGL
Tcld	GDAS1	Total cloud cover (3- or 6-h average)
Cape	GDAS1	Convective available potential energy
Cinh	GDAS1	Convective inhibition
Lisd	GDAS1	Standard lifted index
Lib4	GDAS1	Best 4-layer lifted index
Pblh	GDAS1	Planetary boundary layer height
Tmps	GDAS1	Temperature at surface
Solm	GDAS1	Volumetric soil moisture content
Crai	GDAS1	Categorial rain (yes = 1, no = 0) (3- or 6-h average)
Lcld	GDAS1	Low cloud cover (3- or 6-h average)
Lhtf	GDAS1	Latent heat net flux at surface (3- or 6-h average)
Mcld	GDAS1	Middle cloud cover (3- or 6-h average)
Hcld	GDAS1	High cloud cover (3- or 6-h average)

Table 2
Descriptive statistics.

Variable	Mean	SD	Median	TM	MAD	Min	Max	Range	Skew	Kurtosis	SE	IQR	5th quantile	25th quantile	75th quantile	95th quantile	LOD
[i]Acenaphthylene [ng m^{-3}]	0.015	0.022	0.005	0.010	0	0.005	0.120	0.115	2.642	7.368	0.003	0.010	0.005	0.005	0.015	0.061	0.01
[o]Acenaphthylene [ng m^{-3}]	0.020	0.038	0.005	0.011	0	0.005	0.262	0.257	4.063	20.989	0.004	0.006	0.005	0.005	0.011	0.087	0.01
[i]Acenaphthene [ng m^{-3}]	0.025	0.039	0.005	0.017	0	0.005	0.232	0.227	2.727	9.474	0.005	0.034	0.005	0.005	0.039	0.094	0.01
[o]Acenaphthene [ng m^{-3}]	0.015	0.020	0.005	0.010	0	0.005	0.086	0.081	2.141	3.509	0.002	0.006	0.005	0.005	0.011	0.065	0.01
[i]Anthracene [ng m^{-3}]	0.029	0.043	0.010	0.020	0.008	0.005	0.286	0.281	3.500	16.369	0.005	0.034	0.005	0.005	0.039	0.089	0.01
[o]Anthracene [ng m^{-3}]	0.033	0.034	0.023	0.027	0.026	0.005	0.141	0.136	1.383	1.354	0.004	0.040	0.005	0.005	0.045	0.106	0.01
[i]As [ng m^{-3}]	0.705	0.396	0.659	0.666	0.336	0.200	1.865	1.665	0.787	0.222	0.046	0.463	0.200	0.442	0.905	1.492	0.4
[o]As [ng m^{-3}]	0.863	0.527	0.696	0.792	0.303	0.200	3.187	2.987	1.789	4.289	0.061	0.534	0.200	0.545	1.078	1.786	0.4
[i]Benzo(a)anthracene [ng m^{-3}]	0.326	0.901	0.093	0.157	0.061	0.020	7.249	7.229	6.349	44.574	0.105	0.128	0.028	0.060	0.189	0.786	0.01
[o]Benzo(a) anthracene [ng m^{-3}]	0.359	0.534	0.164	0.230	0.148	0.028	2.419	2.391	2.558	6.205	0.062	0.205	0.035	0.085	0.290	1.491	0.01
[i]Benzo(a)pyrene [ng m^{-3}]	0.504	0.918	0.220	0.309	0.195	0.033	6.844	6.811	4.771	28.370	0.107	0.349	0.050	0.113	0.461	1.915	0.01
[o]Benzo(a)pyrene [ng m^{-3}]	0.484	0.606	0.302	0.343	0.249	0.040	2.869	2.829	2.208	4.272	0.070	0.330	0.050	0.131	0.462	2.047	0.01
[i]Benzo(b) fluoranthene [ng m^{-3}]	0.699	1.040	0.331	0.475	0.271	0.084	7.106	7.022	3.821	18.276	0.121	0.435	0.106	0.208	0.643	2.173	0.01
[o]Benzo(b) fluoranthene [ng m^{-3}]	0.888	0.892	0.592	0.706	0.444	0.104	3.842	3.738	1.790	2.381	0.104	0.588	0.150	0.290	0.878	3.215	0.01
[i]Benzo(ghi)perylene [ng m^{-3}]	0.571	0.727	0.347	0.418	0.241	0.005	4.896	4.891	3.464	15.448	0.085	0.333	0.103	0.198	0.530	1.864	0.01
[o]Benzo(ghi)perylene [ng m^{-3}]	0.680	0.639	0.471	0.554	0.324	0.025	2.702	2.677	1.706	2.048	0.074	0.393	0.128	0.298	0.691	2.236	0.01
[i]Benzo(k) fluoranthene [ng m^{-3}]	0.595	0.927	0.293	0.395	0.252	0.061	6.472	6.411	4.067	20.567	0.108	0.402	0.080	0.141	0.543	1.866	0.01
[o]Benzo(k) fluoranthene [ng m^{-3}]	0.737	0.759	0.479	0.583	0.402	0.074	3.395	3.321	1.753	2.313	0.088	0.526	0.105	0.237	0.763	2.608	0.01
[i]Ca^{2+} [µg m^{-3}]	14.073	19.278	4.000	9.812	0	4.000	71.580	67.580	1.664	1.279	2.241	8.131	4.000	4.000	12.131	57.155	8
[o]Ca^{2+} [µg m^{-3}]	13.773	17.502	4.000	9.665	0	4.000	82.931	78.931	2.090	4.045	2.035	13.634	4.000	4.000	17.634	48.774	8
[o]Cape [J kg^{-1}]	51.783	82.647	12.956	32.783	19.209	0	376.429	376.429	2.187	4.580	9.607	76.688	0	0	76.688	228.632	/
[i]Cd [ng m^{-3}]	0.206	0.094	0.223	0.212	0.088	0.025	0.370	0.345	−0.524	−0.674	0.011	0.121	0.025	0.155	0.276	0.332	0.05
[o]Cd [ng m^{-3}]	0.236	0.118	0.236	0.232	0.088	0.025	0.610	0.585	0.518	1.129	0.014	0.118	0.042	0.173	0.291	0.425	0.05
[i]Chrysene [ng m^{-3}]	0.527	1.214	0.214	0.292	0.161	0.021	9.406	9.385	5.739	36.940	0.141	0.254	0.050	0.129	0.383	1.157	0.01
[o]Chrysene [ng m^{-3}]	0.645	0.806	0.357	0.464	0.281	0.025	3.781	3.756	2.271	4.777	0.094	0.432	0.087	0.178	0.609	2.602	0.01
[o]Cinh [J kg^{-1}]	−11.663	17.941	−4.383	−7.922	6.498	−103.941	0	103.941	−2.668	8.900	2.086	14.751	−43.827	−15.385	−0.634	0	/
[i]Cl^- [µg m^{-3}]	1.716	3.191	1.000	1.000	0	1.000	23.696	22.696	5.361	30.549	0.371	0	1.000	1.000	1.000	3.385	2
[o]Cl^- [µg m^{-3}]	1.312	1.048	1.000	1.017	0	1.000	7.484	6.484	3.862	16.270	0.122	0	1.000	1.000	1.000	3.796	2
[i]CO [mg m^{-3}]	0.300	0.089	0.277	0.292	0.066	0.163	0.552	0.389	0.854	0.087	0.010	0.086	0.187	0.244	0.329	0.490	0.1
[o]CO [mg m^{-3}]	0.287	0.093	0.259	0.278	0.067	0.154	0.549	0.395	0.879	−0.027	0.011	0.099	0.171	0.224	0.323	0.477	0.1
[o]Crai	0.318	0.342	0.250	0.283	0.371	0	1.000	1.000	0.594	−1.113	0.040	0.625	0	0	0.625	0.875	/
[i]Cr [ng m^{-3}]	11.878	5.985	11.564	11.264	4.470	3.298	42.497	39.199	1.958	7.436	0.696	5.745	4.350	8.221	13.966	21.841	2

(continued on next page)

Table 2 (*continued*)

Variable	Mean	SD	Median	TM	MAD	Min	Max	Range	Skew	Kurtosis	SE	IQR	5th quantile	25th quantile	75th quantile	95th quantile	LOD
[o]Cr [ng m^{-3}]	11.518	6.309	10.554	10.647	3.096	3.148	43.886	40.738	2.331	8.325	0.733	4.008	4.245	8.770	12.778	22.843	2
[i]Dibenz(a,h) anthracene [ng m^{-3}]	0.083	0.092	0.049	0.063	0.041	0.010	0.459	0.449	2.343	5.520	0.011	0.055	0.013	0.030	0.085	0.294	0.01
[o]Dibenz(a,h) anthracene [ng m^{-3}]	0.097	0.095	0.076	0.080	0.057	0.010	0.526	0.516	2.401	7.004	0.011	0.076	0.014	0.036	0.112	0.256	0.01
[o]Dwf [W m^{-2}]	213.026	83.677	233.059	218.129	76.098	31.203	340.581	309.379	−0.596	−0.624	9.727	107.664	59.545	162.581	270.246	331.919	/
[i]Fluoranthene [ng m^{-3}]	0.278	0.318	0.186	0.219	0.159	0.005	1.590	1.585	2.363	6.296	0.037	0.234	0.005	0.089	0.323	0.928	0.01
[o]Fluoranthene [ng m^{-3}]	0.303	0.289	0.200	0.256	0.187	0.005	1.339	1.334	1.618	2.381	0.034	0.249	0.005	0.129	0.378	0.934	0.01
[i]Fluorene [ng m^{-3}]	0.047	0.087	0.005	0.026	0	0.005	0.483	0.478	2.811	8.697	0.010	0.035	0.005	0.005	0.040	0.246	0.01
[o]Fluorene [ng m^{-3}]	0.044	0.090	0.005	0.022	0	0.005	0.565	0.560	3.445	14.293	0.010	0.030	0.005	0.005	0.035	0.214	0.01
[o]Hcld [%]	44.152	30.992	44.324	43.252	36.578	0	97.779	97.779	0.082	−1.261	3.603	52.120	0.910	14.834	66.953	96.015	/
[i]Hours [h]	4.050	3.328	5.075	3.958	4.571	0	9.917	9.917	−0.044	−1.571	0.387	6.779	0	0.050	6.829	8.571	0.02
[i]Indeno(1,2,3-cd) pyrene [ng m^{-3}]	0.468	0.579	0.266	0.345	0.201	0.030	3.600	3.570	2.892	10.465	0.067	0.297	0.079	0.156	0.452	1.559	0.01
[o]Indeno(1,2,3-cd) pyrene [ng m^{-3}]	0.544	0.533	0.378	0.436	0.287	0.011	2.287	2.276	1.779	2.434	0.062	0.352	0.087	0.220	0.572	1.899	0.01
[o]Lcld [%]	29.195	29.453	20.919	25.591	29.538	0	95.328	95.328	0.833	−0.546	3.424	45.488	0	3.450	48.938	88.440	/
[o]Lhtf [W m^{-2}]	78.533	37.242	76.203	76.206	41.577	13.514	165.274	151.760	0.452	−0.445	4.329	54.832	26.739	47.589	102.421	150.402	/
[o]Lhf [K]	3.491	3.261	2.958	3.290	3.425	−2.279	10.531	12.810	0.489	−0.731	0.379	4.803	−0.629	0.776	5.579	9.467	/
[o]Lid [K]	278.311	3.819	278.185	278.186	4.804	271.856	286.141	14.285	0.252	−1.029	0.444	6.296	272.866	274.967	281.263	284.601	/
[o]Mcld [%]	30.270	28.819	21.181	27.393	31.332	0	92.883	92.883	0.605	−0.943	3.350	48.475	0	3.841	52.316	82.871	/
[i]Mn [ng m^{-3}]	3.415	2.021	3.043	3.111	1.148	1.000	11.816	10.816	1.831	4.271	0.235	1.556	1.000	2.393	3.949	8.076	2
[o]Mn [ng m^{-3}]	3.525	1.479	3.102	3.318	0.845	1.000	9.188	8.188	1.458	2.332	0.172	1.478	2.106	2.580	4.058	6.322	2
[o]Mofd [°]	159.516	102.558	146.173	155.967	99.381	1.458	360.091	358.634	0.367	−0.821	11.922	131.346	5.670	90.766	222.112	336.635	/
[o]Mofi [N m^{-2}]	0.100	0.081	0.080	0.088	0.068	0.007	0.392	0.385	1.492	2.402	0.009	0.096	0.016	0.040	0.136	0.260	/
[o]Mslp [hPa]	1012.724	5.565	1013.741	1013.006	5.473	1000.900	1023.753	22.853	−0.418	−0.610	0.647	8.269	1002.450	1008.730	1016.999	1019.859	/
[i]Naphthalene [ng m^{-3}]	0.028	0.043	0.005	0.019	0	0.005	0.224	0.219	2.123	4.940	0.005	0.026	0.005	0.005	0.031	0.106	0.01
[o]Naphthalene [ng m^{-3}]	0.041	0.090	0.005	0.022	0	0.005	0.644	0.639	4.558	25.673	0.010	0.037	0.005	0.005	0.042	0.167	0.01
[i]NH_4 [µg m^{-3}]	1.513	1.319	1.213	1.350	1.278	0.100	7.277	7.177	1.476	3.455	0.153	1.686	0.100	0.553	2.239	3.574	0.2
[o]NH_4 [µg m^{-3}]	2.486	2.311	1.955	2.132	2.074	0.100	12.042	11.942	1.516	2.831	0.269	2.707	0.100	0.840	3.548	7.441	0.2
[i]Ni [ng m^{-3}]	7.926	8.198	5.118	5.999	2.363	1.000	45.963	44.963	2.833	8.139	0.953	3.851	2.555	4.019	7.869	26.248	2
[o]Ni [ng m^{-3}]	7.951	5.867	6.455	6.952	4.144	1.000	31.617	30.617	2.270	6.118	0.682	6.238	2.789	4.042	10.280	16.738	2
[i]NO_3 [µg m^{-3}]	1.279	1.198	1.000	1.000	0	1.000	10.142	9.142	5.889	38.534	0.139	0	1.000	1.000	1.000	2.441	2
[o]NO_3 [µg m^{-3}]	4.423	3.898	3.325	3.776	3.447	1.000	16.987	15.987	1.422	1.892	0.453	5.524	1.000	1.000	6.524	11.475	2
[i]Pb [ng m^{-3}]	4.079	2.924	3.621	3.743	1.309	1.000	23.951	22.951	4.305	26.298	0.340	1.706	1.000	2.886	4.592	7.425	2
[o]Phlh [m]	512.239	152.702	517.356	509.879	152.207	194.949	1070.878	875.929	0.439	0.935	17.751	202.364	279.869	407.687	610.051	733.540	/
[o]Pb [ng m^{-3}]	4.605	2.796	4.145	4.261	1.645	1.000	22.494	21.494	3.646	20.341	0.325	1.902	1.713	3.399	5.302	8.758	2
[i]People	185.973	208.428	167.500	154.183	244.629	0	1226.000	1226.000	2.066	6.876	24.229	275.500	0	0.500	276.000	504.950	1
[i]Phenanthrene [ng m^{-3}]	0.166	0.266	0.083	0.110	0.116	0.005	1.478	1.473	3.284	12.544	0.031	0.207	0.005	0.005	0.212	0.559	0.01
[o]Phenanthrene [ng m^{-3}]	0.170	0.226	0.116	0.124	0.165	0.005	1.215	1.210	2.302	6.289	0.026	0.222	0.005	0.005	0.227	0.595	0.01
[i]Pressure [mbar]	992.676	5.536	993.555	992.979	5.421	980.996	1003.425	22.429	−0.441	−0.604	0.643	8.281	982.196	988.749	997.030	999.778	1
[i]$PM_{2.5}$ [µg m^{-3}]	16.196	7.340	14.729	15.540	6.859	4.182	45.266	41.083	1.155	2.192	0.853	8.534	6.920	11.550	20.084	28.374	1
[o]$PM_{2.5}$ [µg m^{-3}]	17.469	8.003	15.862	16.527	5.411	5.785	50.718	44.933	1.509	3.193	0.930	8.335	7.728	12.493	20.828	31.164	1
[o]Pressure [mbar]	992.732	5.548	993.813	993.049	5.336	981.037	1003.086	22.049	−0.471	−0.611	0.645	7.895	982.043	989.185	997.080	999.670	1
[o]Prss [hPa]	988.226	5.417	989.418	988.564	5.342	976.491	998.371	21.880	−0.500	−0.585	0.630	8.186	977.957	984.653	992.839	994.732	/
[i]Pyrene [ng m^{-3}]	0.320	0.396	0.193	0.237	0.162	0.010	2.451	2.441	2.917	10.602	0.046	0.252	0.032	0.107	0.359	1.146	0.01

(continued on next page)

Table 2 (*continued*)

Variable	Mean	SD	Median	TM	MAD	Min	Max	Range	Skew	Kurtosis	SE	IQR	5th quantile	25th quantile	75th quantile	95th quantile	LOD
°Pyrene [ng m^{-3}]	0.338	0.313	0.216	0.282	0.190	0.011	1.426	1.415	1.580	1.835	0.036	0.269	0.053	0.129	0.397	1.046	0.01
°Rain duration [h]	1.536	2.707	0.025	0.908	0.037	0	12.233	12.233	2.031	3.698	0.315	2.094	0	0	2.094	8.071	0.02
°Rain total [h]	181.989	394.929	0.200	79.413	0.297	0	1897.600	1897.600	2.689	7.019	45.910	131.025	0	0	131.025	1082.645	0.02
°Rh 2 m [%]	67.950	13.061	70.443	68.355	17.070	42.081	89.196	47.115	−0.202	−1.155	1.518	21.989	45.916	56.685	78.674	86.306	/
iRh [%]	36.783	7.795	36.081	36.741	7.334	20.200	56.985	36.785	0.110	−0.233	0.906	9.109	23.257	32.017	41.126	50.248	0.1
°Rh [%]	61.534	15.717	59.228	61.545	19.299	31.208	89.668	58.460	0.050	−1.102	1.827	25.973	37.166	49.299	75.272	86.138	0.1
iRn Bq m^{-3}]	74.525	24.616	67.884	71.065	16.289	40.802	141.425	100.623	1.204	0.746	2.862	23.144	46.196	58.665	81.809	130.946	0.1
°Shif [W m^{-2}]	25.551	19.374	26.026	25.027	20.546	−23.071	80.690	103.761	0.289	0.197	2.252	27.280	−0.973	11.112	38.392	52.471	/
iSO_2 [µg m^{-3}]	1.849	1.453	1.204	1.588	0.683	0.573	6.913	6.340	1.572	1.605	0.169	1.324	0.632	0.924	2.248	4.867	1
°SO_2 [µg m^{-3}]	3.155	2.638	2.269	2.671	1.601	0.590	12.225	11.635	1.769	2.851	0.307	2.710	0.886	1.336	4.046	8.596	1
iSO_4^{2-} [µg m^{-3}]	5.319	3.961	4.466	4.819	3.106	0.500	19.170	18.670	1.223	1.752	0.460	4.221	0.500	2.906	7.127	12.942	1
°SO_4^{2-} [µg m^{-3}]	7.051	4.498	6.594	6.682	3.612	0.500	23.972	23.472	1.102	1.915	0.523	4.752	0.500	4.207	8.959	15.007	1
°Solm [frac.]	0.299	0.017	0.299	0.299	0.018	0.260	0.330	0.070	−0.066	−0.635	0.002	0.021	0.276	0.290	0.311	0.328	/
°T 2 m [°C]	12.766	4.421	13.353	12.725	4.534	4.220	20.853	16.633	0.000	−1.030	0.514	6.747	5.920	8.922	15.668	20.107	/
°Tcld [%]	58.645	31.046	67.126	60.005	42.229	1.370	99.905	98.535	−0.298	−1.258	3.609	55.332	8.101	28.569	83.901	99.435	/
iTemp [°C]	23.407	1.525	23.371	23.376	1.635	20.290	27.921	7.631	0.230	−0.095	0.177	2.224	21.076	22.235	24.459	25.821	0.1
°Tmps [°C]	13.020	4.231	13.439	13.006	4.591	5.015	20.795	15.780	−0.035	−1.020	0.492	6.691	6.356	9.146	15.837	19.966	/
°Temp [°C]	13.843	4.953	14.148	13.833	5.703	4.785	22.977	18.192	−0.012	−1.005	0.576	7.127	5.587	10.193	17.321	21.746	0.1
°Tpp6 [m]	0	0.001	0	0.000	0	0	0.003	0.003	1.922	2.835	0.000	0	0	0	0	0.001	/
°WD 10 m [°]	210.976	60.974	208.671	211.261	70.035	88.849	325.800	236.951	−0.021	−0.892	7.088	94.971	110.998	163.167	258.138	310.858	/
°WD [°]	220.327	74.935	225.833	222.170	87.768	23.289	359.233	335.944	−0.220	−0.638	8.711	116.300	112.880	166.137	282.437	326.362	1
°WS 10 m [m s^{-1}]	3.136	1.103	2.963	3.038	1.239	1.550	6.111	4.561	0.662	−0.178	0.128	1.693	1.736	2.143	3.835	5.318	/
°WS [m s^{-1}]	1.467	0.478	1.328	1.427	0.433	0.730	2.759	2.029	0.730	−0.284	0.056	0.589	0.905	1.115	1.704	2.299	0.1

*Abbreviations: standard deviation (SD), truncated mean (TM), median absolute deviation (MAD), standard error (SE), interquartile range (IQR), limit of detection (LOD).
**prefix: *i* – indoor, *o* – outdoor.

Standard. Continuous monitoring of O_3 concentrations was performed by the cross-flow modulated ultraviolet absorption method using APOA-370 device according to the SRPS EN 14625:2013 Standard.

The indoor concentrations of radon (Bq m^{-3}) were measured by using SN1029 radon monitor (Sun Nuclear Corporation, NRSB approval-code 31822). The device consists of two diffused junction photodiodes which serve as a radon detector and is equipped with sensors for temperature, barometric pressure and relative humidity. The device was set to simultaneously record radon concentration, temperature, atmospheric pressure, and relative humidity with a time resolution of 30 min.

2.5. Meteorological data

The outdoor meteorological data (air pressure, temperature, humidity, rainfall, and wind speed and direction) were obtained by using Vaisala weather station (Weather Transmitter WXT530 Series). Additionally, 24-parameter meteorological data for the sampling site location were obtained with a time resolution of 3 h from Global Data Assimilation System (GDAS1) database, by using MeteoInfo software for meteorological data visualization and analysis (Wang, 2014), Table 1.

2.6. Data analysis

After the exclusion of outliers and incomplete cases, a total of 74 samples were used for data analysis. Descriptive statistics (including box plots), probability density functions, correlation analysis (including hierarchical clustering) and time series analysis were obtained and presented by using R packages 'ggdendro' (de Vries and Ripley, 2016), 'Hmisc' (Harrell, 2019), 'ggplot2' (Wickham, 2016), and 'plotly' (Sievert, 2020).

For the purpose of source apportionment, the Unmix receptor model was applied (US EPA Unmix 6). The species were selected for the analysis by using an initial species function. Other pollutants were subsequently added to test stability of the minimal solution and explore whether any of them can lead to a better solution. Finally, a total of 14 and 11 pollutants were chosen as Unmix input variables resulting in a four-factor solution for both indoor and outdoor environments. The concepts underlying Unmix have been described in a geometrical and intuitive manner, and the mathematical details are presented elsewhere (Henry 2003).

Regression analysis by means of XGBoost was implemented for estimating the relationships between each PAH concentrations and all other PAHs, inorganic gaseous pollutants, radon, $PM_{2.5}$ and particle constituents (trace metals and ions), meteorological parameters (measured and GDAS1-modeled), the number of people in the amphitheater and the time they spent indoors, trend, weekday and weekend (39 and 64 parameters in total for indoor and outdoor environment, respectively).

XGBoost refers to a highly effective ML technique of building a complex prediction model by iterative combining ensembles of weak prediction models into a single strong learner. In the tree growing algorithm used by XGBoost each decision tree serves to complement all others and correct for the residuals in the predictions made by the previous ones (Sheridan et al., 2016). The XGBoost was successfully applied in a number of studies due to its core advantages being related to handling sparse data, excellent predictive performance, highly optimized multicore and the complexity penalization of the trees that was not commonly used for previous additive tree models (Mitchell and Frank, 2017; Nielsen, 2016). In this study, we used Python (Python Software Foundation) XGBoost implementation (XGBoost Python Package). The dataset was split into training (80%) and validation (20%) set. Hyperparameter tuning was implemented using a brute-force grid search and stratified 10-fold cross-validation that was replicated ten times. The best performing hyperparameter values were used for the final model. The obtained results will be considered in details by the application of explainable artificial intelligence methods in the succeeding parts of this paper.

Beside conventional images, we present all the relevant findings as interactive plots produced by using R package 'plotly' hosted at the web page designed to support this paper at www.envpl.ipb.ac.rs/papers/20/PAHs/.

3. Results

As can be seen in Table 2 and Fig. 1, mean indoor $PM_{2.5}$-related PAH concentrations (4.68 ng m^{-3}) were lower than the corresponding outdoor values (5.40 ng m^{-3}), although occasional extreme PAH concentration events were shown to reach almost two times higher values in the indoor compared to the outdoor environment (45.79 and 27.49 ng m^{-3}, respectively). Concentration distribution of all investigated PAHs, inorganic ions and trace elements in the indoor and outdoor environment, except CO, Cr and radon, appeared to be unimodal and positively skewed with a noticeable long right-sided tail (Fig. 2), which indicates that the majority of the measured pollutant concentrations are distributed within the first quartile of the registered range (Table 2). A sharp symmetrical bell-shaped curve of Rn can be inferred as a result of natural emissions, while the Cr concentration distribution suggests its levels are less affected by human activities (Pongpiachan and Iijima, 2016). Unsurprisingly, in both the indoor and outdoor environment, the concentrations of higher molecular weight (5-ring and more) PAHs in PM fine fraction exceeded the levels of volatile and semi-volatile low weight 2- and 3-ring aromatics, which are under normal ambient conditions almost entirely distributed in a gas phase (Table 2, Fig. 1).

Considering the meteorological factors, the same applies to the rain/precipitation parameters, convective potential energy, and low cloud coverage, whereas convective inhibition data exhibited the opposite, negatively skewed distribution with a long left-sided tail. The uniform to normal distribution of the relative humidity, temperature and PBL height data is evidenced, while wind direction, soil humidity, as well as relative humidity and temperature at 2 m data followed bimodal value distribution patterns implying two distinct ambiences which took turns over the study period. The beginning of the sampling campaign was marked by frequent cyclonic activity, mean daily temperatures below 12 °C, strong wind episodes and frequent precipitation events. A high-pressure system was established over the Balkans in mid-March (17th to 21st), bringing calm weather without precipitations and more sunshine hours, although mean daily temperature did not exceed 12 °C. The last days of March were marked by variable weather conditions, occasional precipitations and strong wind. An upper-level ridge in pressure prevailed in the first days of April causing the arrival of warmer and drier air masses. From April 8th recurrent changes in weather were caused by penetration of cold air fronts and frequent lowering of mean daily temperature. Changeable weather and temperatures below 12 °C, caused by the upper-level trough in the pressure field continued over the first few days of May. Calm and dry weather in the middle of May was followed by a sudden change, when a passage of a cold front led to intense rainfall, significant wind gusts and sharp decrease in temperature. The end of the study campaign was marked by the penetration of warmer and drier air masses from S direction. Details on meteorological conditions are presented in the Supplementary Fig. 1.

According to our results, total PAH concentrations exhibit sharp decrease from the start of the study campaign (March 1st) till April 1st, followed by a slower decline till the end of May. The total PAH concentrations exhibited weekly dynamics with the lowest values in the outdoor environment on Wednesday, which increased to their maximum levels on Friday and subsequently declined on Saturday and Sunday (Supplementary Fig. 2). It should also be noted that outdoor PAH weekly behavior pattern corresponded to indoor air quality variations, with an exception of weekend period, when the increase in outdoor concentrations preceded the rise in indoor PAH levels. The weekly dynamics of PAH levels suggest that the anthropogenic activities were intensified over the working days, particularly on Monday and Friday, while on

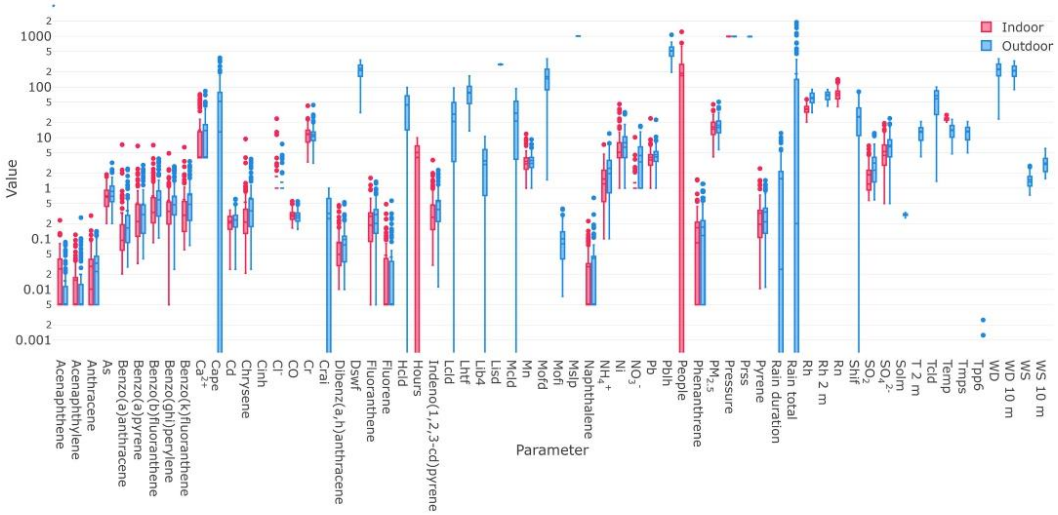


Fig. 1. The box plot of all indoor and outdoor measured parameters.

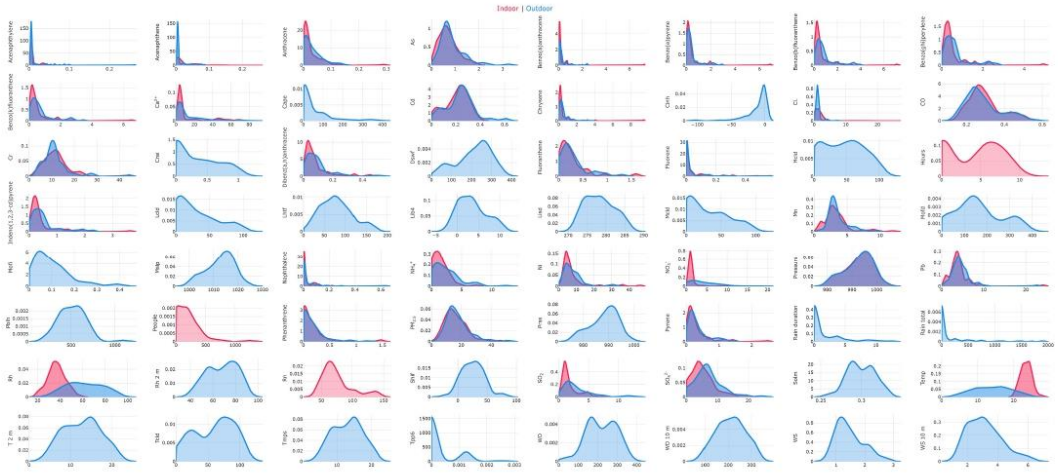


Fig. 2. The probabilty density function for indoor and outdoor measured parameters.

weekend, pollutant levels declined due to the decrease in industrial and traffic emissions, and the outdoor emission sources took on the role of major pollutant contributors.

As represented by correlation matrix (Fig. 3), significant linear correlation coefficients ($r > 0.90$) were found between the indoor concentrations of B[a]A, Chy, I[cd]P, B[ghi]P, B[k]F, B[b]F, B[a]P, and Fla, and the same applies to the outdoor environment. The correlations between the indoor and outdoor levels of the listed PAHs was in the range from 0.50 to 0.70. These results suggest similar behavior and common sources of the listed contaminants, being discussed in details below. Considering other investigated variables, the correlations were only registered between the indoor and outdoor levels of PAHs and CO ($r = 0.60–0.80$). We assumed that other functional dependencies apart from linear could be further investigated to describe associations between PAHs and inorganic pollutants or meteorological variables, which will be considered in details in the succeeding parts of this paper.

The Unmix resolved profiles are presented together with their contributions to the total observed pollutant concentrations in Table 3. A detailed description of profile identification-relevant factors is provided in the following text.

Also, several PAH diagnostic ratios are calculated and considered in the following text.

The XGBoost provided successful and reliable predictions of Pyr, I[cd]P, B[ghi]P, B[k]F, B[b]F, B[a]P, and Chy in indoor and outdoor ambient with relative errors (normalized mean gross error, NMGE) in the range from approx. 10%–20% and correlation coefficients higher

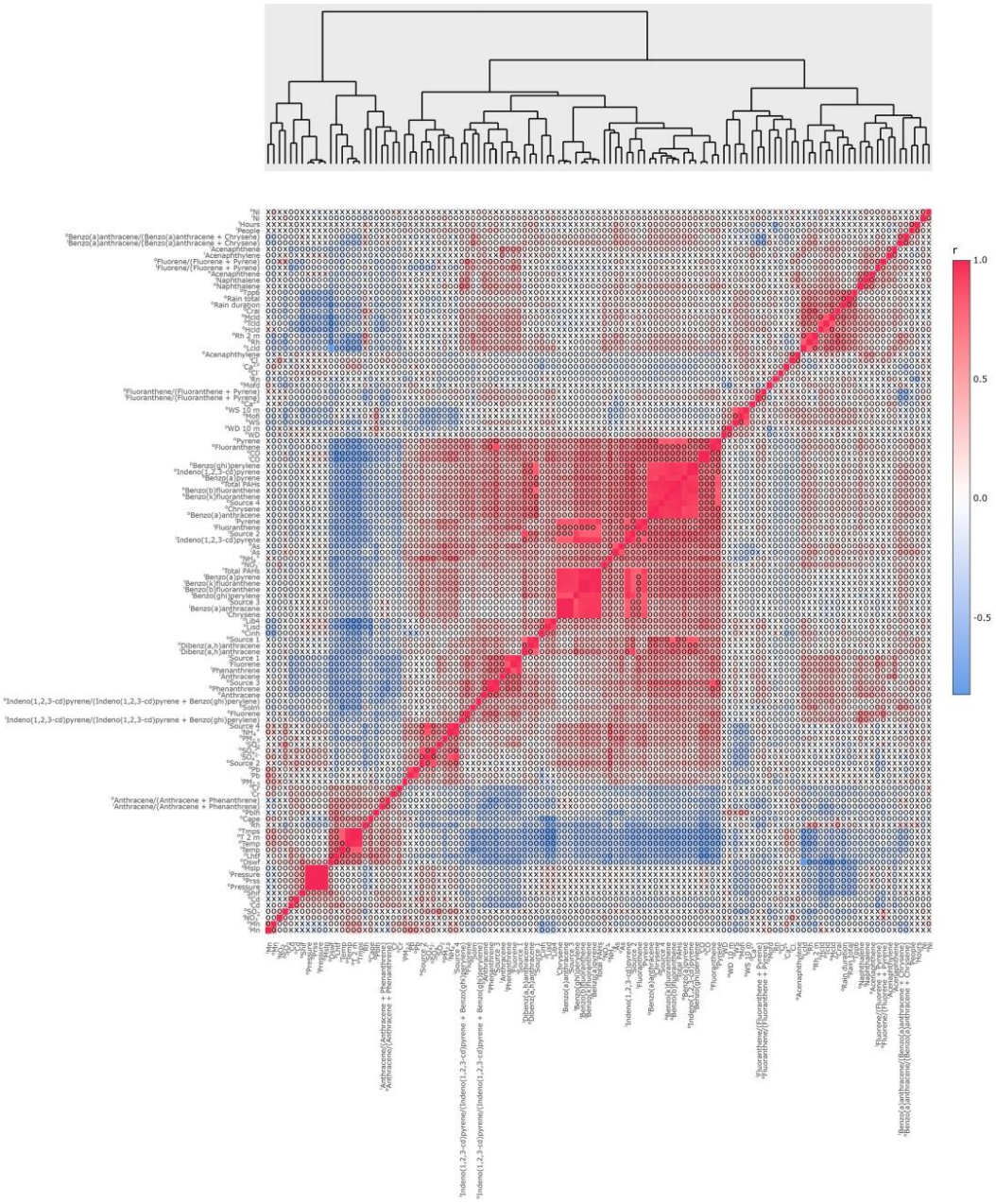


Fig. 3. The correlation matrix for indoor and outdoor measured parameter values.

Table 3
Unmix-resolved source profiles [%].

Species	Indoor				Outdoor			
	Source 1 ^{i}PE	Source 2 ^{i}CPC	Source 3 ^{i}PP	Source 4 ^{i}I/TE	Source 1 ^{o}CC	Source 2 ^{o}TE/PE	Source 3 ^{o}PE	Source 4 ^{o}PP
SO_4	0	0	0	100	7.6	74.1	18.3	0
NH_4	1.7	0	0	98.3	–	–	–	–
Flu	90.4	8.1	1.5	0	–	–	–	–
Phe	74.6	25.4	0	0	0	0	100	0
Fla	38.6	20.2	21.6	19.6	0	28.0	56.9	15.0
Pyr	28.5	15.5	31.8	24.2	–	–	–	–
B[a]A	3.1	8.2	88.7	0	27.7	0	1.2	71.1
Chy	3.8	5.3	80.1	10.9	27.5	0	11.8	60.8
B[b]F	3.9	44.6	34.9	16.6	29.1	16.9	5.6	48.3
B[k]F	5.5	39.8	39.0	15.8	19.2	23.0	7.3	50.5
B[a]P	6.9	49.0	44.1	0	8.7	18.7	0	72.6
I[cd]P	0	79.5	15.5	5.0	40.1	18.1	0	41.8
D[ah]A	0	100	0	0	84.6	13.7	1.7	0
B[ghi]P	0	79.6	18.8	1.6	38.0	21.8	0	40.2
Average contribution	**18.4**	**33.9**	**26.8**	**20.9**	**25.7**	**19.5**	**18.4**	**36.4**

*^{i} – indoor, ^{o} – outdoor.
**PE – petrogenic emissions, CPC – coal and petroleum combustion, PP – pyrogenic processes, I/TE – industrial and traffic emissions, CC – coal combustion, TE/PE – traffic and petrogenic emissions.

than 0.95 (Table 4). Satisfactory predictions were also evidenced for Fla, D[ah]A, and B[a]A, as described by model parameters (NMGE ≈ 30%, r ≥ 0.95). An effective model performance for the listed PAHs is indicated by the high values of coefficient of efficiency (COA = 0.83–0.88; a perfect model has a COA = 1) and index of agreement (IOA > 0.90; the values approaching 1 represent good model performances), as well as low values of mean bias and mean gross error. The prediction of 2- and 3-ring low molecular weight PAHs (Ane, Nap, Phe, Ace, Ant, and Flu) was less accurate (relative error = 35 %–70%; and r = 0.80–0.90), which is expected since those compounds are highly volatile and more gas-phase distributed.

4. Disscusion

As regards $PM_{2.5}$-related organic content, the study of Jedynska et al. (2014), aimed to investigate the levels of PAHs in fine particle fraction at street, urban and regional background, has shown that total PAH concentrations did not exceed 2.1 ng m^{-3} for all ten investigated sites in Europe. On the other hand, the studies performed in Asia have shown that the total $PM_{2.5}$-bound PAH concentrations were up to 50 ng m^{-3} (Wang et al., 2017), and even 70 ng m^{-3} (Xu et al., 2015). While the majority of studies have shown the outdoor $PM_{2.5}$-bound PAH levels to be higher than the corresponding values in the indoor environment in the same season, some studies have shown the opposite (Wang et al., 2020). In comparison with the previous reports, it becomes evident that PAH concentrations in Serbia are significantly higher than it would be expected, considering the air quality in neighbouring countries and air pollution in Serbia can be attributed to a rise in the number of vehicles over the last two decades, the use of outdated technologies in all production sectors, high number of local fireboxes and long-range transport of pollutants from surrounding industrial countries (Stojić et al., 2015, 2016; Perišić et al., 2017). The study of Cvetković et al. (2015) has investigated source contributions to the registered high concentrations of PAHs in Belgrade area, and shown that all sites were heavily influenced by diesel and gasoline emissions, as well as by stationary sources (combustion of oil, industry, residential heating).

The dominant shares of Flu (90%) and Phe (75%) considered to be associated with petrogenic emissions (^{i}PE) were apportioned to indoor Source 1, together with 2–3 times lower contributions of Fla and Pyr and negligible shares of other pollutants. Petrogenic PAHs mostly originate from the low temperature-combustion of crude oil and its products, including kerosene, gasoline, diesel fuel and lubricating oil, and their contributions was estimated to account for 18.4% of the total indoor pollutant concentrations. The highest contribution to the observed indoor pollutant concentrations, amounting to 33.9%, was associated with the indoor Source 2 dominated by D[ah]A (100%). The significant shares of other PAHs in the following order I[cd]P═B[ghi]P > B[a]P > B[b]F > B[k]F ≫ Phe > Fla > Pyr suggest that the source can be attributed to the coal and petroleum combustion (^{i}CPC). Apart from the dominant portions of Chy (88%) and B[a]A (80%), a smaller share of less alkylated and more stable PAHs including B[a]P, B[k]F, B[b]F, Pyr, Fla, B[ghi]P, and I[cd]P was apportioned to indoor Source 3. Considering the profile composition, this source with the estimated contribution of 26.8% was attributed to high temperature (350–1200 °C) pyrogenic processes (^{i}PP), which can be related to incomplete combustion of fossil fuels and biomass in power plants and local fireboxes, industrial coal and petroleum burning and traffic emissions. At high combustion temperatures organic compounds are cracked to reactive radicals that form stable gaseous high-weight PAHs during pyrosynthesis, which subsequently cool and condense on particles. The contribution of the indoor Source 4 attributed to a mix of industrial and traffic emissions (^{i}I/TE) accounted for 20.9% of total registered indoor pollutant concentrations. The species assigned to this source comprise dominant portions of inorganic ions SO_4^{2-} and NH_4^{+}, assigned together with several times lower shares of Fla, Pyr, B[b]F, and B[k]F, suggesting dual emission origin.

The dominant portion of D[ah]A (84.6%) and significant shares of other PAHs in the following order I[cd]P═B[ghi]P > B[b]F = B[a]A = Chy > B[k]F suggest that the outdoor Source 1 with the estimated contribution of 25.7% can be attributed to the coal combustion (^{o}CC). Apart from the dominant portion of SO_4^{2-} (74.1%), followed by Fla (28%), a smaller shares of B[k]F, B[ghi]P, B[a]P, I[cd]P, B[b]F, and D[ah]A was apportioned to the outdoor Source 2. Considering the profile composition, this source with the estimated contribution of 19.5% was attributed to a mix of traffic and petrogenic emissions (^{o}TE/PE). The contribution of the outdoor profile 3 attributed to petrogenic emissions (^{o}PE) was estimated to 18.4% of the registered pollutant concentrations, with the assigned species contributing in the following order Phe (100%) > Fla ≫ SO_4^{2-} and the absence of B[a]P. The outdoor Source 4 attributed to pyrogenic processes (^{o}PP) had the highest contribution of 36.4% to the total outdoor pollutant concentrations, and was distinguished by significant portions of B[a]P (72.6%) = B[a]A > Chy > B[k]F = B[b]F > I[cd]P = B[ghi]P, and the absence of D[ah]A. The origin of outdoor PAHs is mostly dependent on sampling location and surrounding emission sources. In industrial areas, such as the ones explored in the study of Kermani et al. (2019), industrial activities were identified as the main contributor to air quality deterioration. On the other hand,

Table 4
XGBoost model evaluation statistics.

PAH	Indoor								Outdoor							
	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE	IOA	FAC2	MB	MGE	NMB	NMGE	RMSE	r	COE	IOA
Ace	0.780	0.002	0.004	0.193	0.399	0.006	0.874	0.503	0.751	0.593	−0.004	0.010	−0.183	0.511	0.016	0.817	0.525	0.763
Ane	0.591	−0.006	0.016	−0.242	0.610	0.038	0.807	0.491	0.746	0.711	0.002	0.007	0.127	0.541	0.011	0.747	0.405	0.703
Ant	0.785	0.004	0.008	0.198	0.387	0.012	0.926	0.566	0.783	0.727	0.003	0.010	0.139	0.451	0.015	0.798	0.501	0.750
B[a]A	0.846	0.035	0.042	0.256	0.311	0.081	0.967	0.681	0.840	0.798	−0.019	0.058	−0.064	0.192	0.131	0.985	0.833	0.917
B[a]P	0.851	0.019	0.044	0.066	0.154	0.071	0.985	0.844	0.922	0.757	0.011	0.048	0.036	0.158	0.098	0.983	0.874	0.937
B[b]F	0.915	0.013	0.058	0.029	0.127	0.108	0.987	0.881	0.941	0.838	−0.014	0.089	−0.017	0.109	0.148	0.992	0.897	0.949
B[ghi]P	0.919	0.016	0.035	0.048	0.102	0.055	0.995	0.887	0.943	0.850	−0.049	0.062	−0.118	0.149	0.113	0.989	0.836	0.918
B[k]F	0.915	0.016	0.044	0.045	0.126	0.074	0.990	0.877	0.939	0.850	−0.064	0.097	−0.104	0.158	0.196	0.987	0.844	0.922
Chy	0.854	0.012	0.042	0.054	0.186	0.065	0.979	0.810	0.905	0.838	−0.060	0.120	−0.105	0.212	0.243	0.985	0.827	0.913
D[ah]A	0.818	−0.007	0.019	−0.101	0.286	0.035	0.969	0.713	0.857	0.753	−0.006	0.025	−0.075	0.293	0.039	0.945	0.647	0.824
Fla	0.798	−0.029	0.057	−0.159	0.307	0.153	0.970	0.716	0.858	0.794	0.002	0.021	0.014	0.125	0.029	0.987	0.851	0.926
Flu	0.667	0.003	0.010	0.104	0.398	0.019	0.942	0.686	0.843	0.487	−0.022	0.035	−0.442	0.695	0.086	0.939	0.495	0.748
I[cd]P	0.919	−0.002	0.032	−0.008	0.105	0.058	0.995	0.899	0.949	0.834	−0.016	0.046	−0.039	0.115	0.082	0.995	0.881	0.940
Nap	0.582	−0.003	0.014	−0.114	0.568	0.024	0.853	0.502	0.751	0.625	−0.005	0.015	−0.161	0.451	0.026	0.902	0.628	0.814
Phe	0.665	−0.035	0.056	−0.269	0.429	0.157	0.971	0.668	0.834	0.612	−0.008	0.036	−0.078	0.359	0.068	0.953	0.662	0.831
Pyr	0.870	0.019	0.029	0.086	0.131	0.050	0.991	0.871	0.935	0.850	−0.010	0.041	−0.035	0.152	0.080	0.971	0.833	0.916

the study of Liu et al. (2018) has shown that traffic emissions can be considered the most important PAH source, irrespectively of the sampling season, while some researchers (Zhang et al., 2020; Wang et al., 2015a) reported the dominance of coal combustion in colder part of the year.

As can be seen in Fig. 3, the linear correlations exceeding r = 0.70 were obtained for the following profile pairs: source profiles with smallest contributions to total indoor and outdoor PAH concentrations assigned to petrogenic emissions (^{i}PE and ^{o}PE); indoor and outdoor source profiles related to traffic emissions (^{i}I/TE and ^{o}TE/PE); and outdoor source profiles attributed to coal combustion and other pyrogenic processes (^{o}CC and ^{o}PP). Our results suggest that coal combustion and related pyrogenic processes are the dominant sources of PAHs in the study area, while the impact of traffic, industrial and gasoline emissions appear to be less significant. All Unmix resolved profiles with the exception of those associated with traffic exhaust, exhibited a sharp decrease in contributions in the range from 55 to 90%, from the start of the study campaign till the first days of April. In the period that followed, the source shares declined more slowly, while the contributions of profiles attributed to traffic emissions exhibited a continual decline over the entire study period reaching 40% and 70% of their initial shares for ^{i}I/TE and ^{o}TE/PE respectively, Supplementary Fig. 3.

As shown in the Supplementary Fig. 2, the contributions of almost all sources to the registered PAH concentrations decreased from the March to May. Thereby, the sharpest decline in emissions in the first weeks of the study period exhibited the indoor and outdoor sources related to coal combustion and other pyrogenic processes. The impact of traffic and industrial emissions remained relatively stable till the very end of measurement campaign.

The PAH isomeric pairs, *i.e.* the species with the same atomic structure such as Ant and Phe, are expected to behave similarly in the environment and thus, their concentration ratios are a commonly used tool for emission source identification and distinguishing of PAH pollution originating from pyrogenic and petrogenic processes, *i.e.* diesel and gasoline combustion emission, crude oil processing products and biomass or coal burning (Davis et al., 2019). For instance, the study of Yin and Xu (2018) has applied a diagnostic ratio to the particle-bound PAH source apportionment results, and identified diesel, gasoline, and coal combustion as the main emission sources affecting air quality, and the study of Khan et al. (2015) reported similar findings that were confirmed by the source apportionment analysis. Nevertheless, previous research, based on theoretical considerations and laboratory experiments, have suggested that, although different PAH ratios can be considered as valuable source apportionment indicators, the ratio-based conclusions should be drawn with caution since these values are often noticed to exhibit seasonal variations and can be affected by a number of environmental factors, such as the presence of free radicals, meteorological conditions which favor photoreactions, and particle size and characteristics (Tobiszewski and Namieśnik, 2012). Limited information in the literature regarding the specific conditions of partial or entire removal of PAHs from the environment can be found. However, it has been evidenced that their persistence increases with the molecular weight, which can be explained by the fact that higher molecular weight PAHs are mostly particle-bonded, predominantly (83–88%) found in fine fraction (Hassanvand et al., 2015) and more resistant to solar radiation and free radicals under natural conditions (Oliveira et al., 2019).

The Ant/(Ant + Phe) mean ratio higher than 0.1 with steady increase towards the end of the study period and occasional peaks reaching 0.5 indicated the dominance of pyrogenic sources over the three-month campaign, except the first days of April during which the warmer and drier air masses arrived and the corresponding ratio values were significantly below 0.1, suggesting the contribution of petrogenic emissions (Supplementary Fig. 2). However, the study of Kim et al. (2009) has shown that the Ant/(Ant + Phe) ratio could range from 0 to 1, depending on the extent of Ant photodegradation caused by the irradiation of different thickness layer-soot samples.

The B[a]A/(B[a]A + Chr) mean ratio just below the value of 0.35 for the major part of the study period indicated the impact of coal combustion with occasional contributions of vehicular emissions. In compliance with the aforementioned, in the warmer period around April 1st, the ratio values were lower which could be attributed to the impact of traffic emissions, but also to the fact that B[a]A decays faster when adsorbed on particles, which can add to B[a]A/(B[a]A + Chr) ratio decrease. From April 8th recurrent changes in weather were caused by penetration of cold air fronts and frequent lowering of mean daily temperature. It should be noted that indoor-outdoor B[a]A/(B[a]A + Chr) ratios exhibit stronger correlation (0.71) than other indoor-outdoor ratio values (Fig. 3), which suggests that, over the major part of study period, air quality in the indoor environment was strongly affected by the outdoor emissions related to coal combustion and traffic emissions.

The Fla/(Fla + Pyr) mean ratio in the range from 0.3 to 0.5 suggests the impact of petrogenic sources and combustion of fossil fuels other than coal. The significant decrease in the outdoor Fla/(Fla + Pyr) ratio values over the second part of the study period can be attributed to the changeable weather and temperatures below 12 °C, which caused less intense photoreactions that could otherwise lead to faster decay of particle-adsorbed Pyr (Kim et al., 2009).

The Fl/(Fl + Pyr) mean ratio bellow 0.5 can be attributed to gasoline emissions. In March, the Fl/(Fl + Pyr) mean ratio had the highest decrease in value of all ratios, which can be associated with calmer weather without precipitations and more sunshine hours, yet mean daily temperature did not exceed 12 °C, which resulted in shorter atmospheric life time of Fl.

The I[cd]P/(I[cd]P + B[ghi]P) mean ratio of 0.43, suggesting the impact of petroleum combustion and pyrogenic sources, was more or less steady throughout the entire study period, although some studies have shown that atmospheric lifetimes of these species and their ratio can be affected by UV radiation and ageing of particle fraction they are sorbed to.

As can be concluded, functional prediction of $PM_{2.5}$-bound PAHs in the indoor and outdoor environment can be achieved by using ML methods and further research of the pollutant dynamics and its dependency on meteorological factors or particle chemical composition would significantly benefit from the application of ML algorithms. The explained predictions of the obtained regression models by means of explainable artificial intelligence methods will be provided in the succeeding parts of this paper.

5. Conclusions

In this study, 16 US EPA priority PAHs were investigated in indoor and outdoor environment based on a three-month measurement campaign which included the concentrations of inorganic gaseous pollutants, radon, $PM_{2.5}$ and particle-bound trace metals, ions, and PAHs, along with 31 meteorological parameters. The correlation analysis showed noticeable relationships between 5- and 6-ring high molecular weight PAHs, but, except for CO, no significant linear dependencies with other investigated variables were identified. The Unmix source apportionment analysis resolved four source profiles for both indoor and outdoor environment, which are comparable in terms of their apportionments and pollutant shares. The highest contributions to air quality were attributed to sources identified as coal combustion and related pyrogenic processes. Except the impact of traffic and industrial emissions, which remained relatively stable over the study period, the contributions of other sources to the registered PAH concentrations decreased towards the end of the measurement campaign. The analysis of PAH diagnostic ratios revealed the emission sources similar to those identified by source apportionment, although it should be emphasized that ratio-implied solutions should be taken with caution since these values are not a reflection of pollutant sources only, but also point to the impact of environmental factors on air quality. As shown by the evaluation parameters of the XGBoost-obtained models, the prediction of PAH levels in the indoor and outdoor environment appears to be promising and their levels are partly determined by their molecular structure and physico-chemical properties including volatility and gas-particle phase partitioning. Although the presented methods are relevant for discriminating the origin of PAH emissions, supplementary approaches, such as machine learning and explainable artificial intelligence, are required to enhance the understanding of PAH dynamics and their functional relationships with influential factors in complex indoor and outdoor environments. The major contribution to air quality deterioration and high PAH concentrations in the study area was shown to be associated with coal combustion for heating purposes and other pyrogenic processes. It would be advisable to make a shift towards alternative heating sources which would be eco-friendlier.

CRediT authorship contribution statement

Svetlana Stanišić: Writing - original draft, Writing - review & editing. **Mirjana Perišić:** Data curation. **Gordana Jovanović:** Supervision. **Tijana Milićević:** Writing - original draft, Writing - review & editing. **Snježana Herceg Romanić:** Writing - original draft, Writing - review & editing. **Aleksandar Jovanović:** Conceptualization, Methodology, Software. **Andrej Šoštarić:** Visualization, Investigation. **Vladimir Udovičić:** Visualization, Investigation. **Andreja Stojić:** Conceptualization, Methodology, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS, as well as the Croatian Science Foundation – Project OPENTOX No. 8366.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2020.110520.

References

Alves, C.A., Urban, R.C., Pegas, P.N., Nunes, T., 2014. Indoor/outdoor relationships between PM_{10} and associated organic compounds in a primary school. Aerosol Air Qual. Res 14 (1), 86–98. https://doi.org/10.4209/aaqr.2013.04.0114.

Ambient Air — Determination of Total (Gas and Particle-phase) Polycyclic Aromatic Hydrocarbons — Collection on Sorbent-Backed Filters with Gas Chromatographic/mass Spectrometric Analyses, 2010 available at: https://iss.rs/en/project/show/iss:proj:24983.

Ambient Air - Standard Gravimetric Measurement Method for the Determination of the PM Mass Concentration of Suspended Particulate Matter, 2015 available at: https://iss.rs/en/project/show/iss:proj:49389.

Ambient Air Quality - Standard Method for the Measurement of Pb, Cd, as and Ni in the PM Fraction of Suspended Particulate Matter, 2008 available at: https://iss.rs/en/project/show/iss:proj:18667.

Brehmer, C., Norris, C., Barkjohn, K.K., Bergin, M.H., Zhang, J., Cui, X., Teng, Y., Zhang, Y., Black, M., Li, Z., Shafer, M.M., 2020. The impact of household air cleaners on the oxidative potential of PM2. 5 and the role of metals and sources associated with indoor and outdoor exposure. Environ. Res. 181, 108919. https://doi.org/10.1016/j.envres.2019.108919.

Cvetković, A., Jovašević-Stojanović, M., Marković, D., Ristovski, Z., 2015. Concentration and source identification of polycyclic aromatic hydrocarbons in the metropolitan area of Belgrade, Serbia. Atmos. Environ. 112, 335–345. https://doi.org/10.1016/j.atmosenv.2015.04.034.

Davis, E., Walker, T., Adams, M., Willis, R., Norris, G., Henry, R., 2019. Source apportionment of polycyclic aromatic hydrocarbons (PAHs) in small craft harbor

(SCH) surficial sediments in Nova Scotia, Canada. Sci. Total Environ. 691, 528–537. https://doi.org/10.1016/j.scitotenv.2019.07.114.
de Vries, A., Ripley, B., 2016. Ggdendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'. R Package Version 0.1-20. https://CRAN.R-project.org/package=ggdendro.
Directive 2004/107/EC of the Europen parliament and of the council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. Official Journal of the European Union 23 (1–16), 26/01/2005.
Elie, M.R., Choi, J., Nkrumah-Elie, Y.M., Gonnerman, G.D., Stevens, J.F., Tanguay, R.L., 2015. Metabolomic analysis to define and compare the effects of PAHs and oxygenated PAHs in developing zebrafish. Environ. Res. 140, 502–510. https://doi.org/10.1016/j.envres.2015.05.009.
Gao, X., Xu, Y., Cai, Y., Shi, J., Chen, F., Lin, Z., Chen, T., Xia, Y., Shi, W., Zhao, Z., 2019. Effects of filtered fresh air ventilation on classroom indoor air and biomarkers in saliva and nasal samples: a randomized crossover intervention study in preschool children. Environ. Res. 179, 108749. https://doi.org/10.1016/j.envres.2019.108749.
Harrell, F., with contributions from Charles Dupont and many others, 2019. Hmisc: Harrell miscellaneous. R package version 4.3-0. https://CRAN.R-project.org/package=Hmisc.
Hassanvand, M.S., Naddafi, K., Faridi, S., Nabizadeh, R., Sowlat, M.H., Momeniha, F., Gholampour, A., Arhami, M., Kashani, H., Zare, A., Niazi, S., 2015. Characterization of PAHs and metals in indoor/outdoor $PM_{10}/PM_{2.5}/PM_1$ in a retirement home and a school dormitory. Sci. Total Environ. 527, 100–110. https://doi.org/10.1016/j.scitotenv.2015.05.001.
Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. Chemometr. Intell. Lab. Syst. 65 (2), 179–189. https://doi.org/10.1016/S0169-7439(02)00108-9.
International Agency for Research on Cancer, 2012. A Review of Human Carcinogens. Part F: Chemical Agents and Related Occupations. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*.
Jedynska, A., Hoek, G., Eeftens, M., Cyrys, J., Keuken, M., Ampe, C., Beelen, R., Cesaroni, G., Forastiere, F., Cirach, M., de Hoogh, K., 2014. Spatial variations of PAH, hopanes/steranes and EC/OC concentrations within and between European study areas. Atmos. Environ. 87, 239–248. https://doi.org/10.1016/j.atmosenv.2014.01.026.
Kermani, M., Jonidi Jafari, A., Gholami, M., Shahsavani, A., Taghizadeh, F., Arfaeinia, H., 2019. Ambient air PM2. 5-bound PAHs in low traffic, high traffic, and industrial areas along Tehran, Iran, *Human and Ecological Risk Assessment*. Int. J. 1–18. https://doi.org/10.1080/10807039.2019.1695194.
Keyte, I., Harrison, R., Lammel, G., 2013. Chemical reactivity and long-range transport potential of polycyclic aromatic hydrocarbons – a review. Chem. Soc. Rev. 42, 9333–9391. https://doi.org/10.1039/c3cs60147a.
Khan, M.F., Latif, M.T., Lim, C.H., Amil, N., Jaafar, S.A., Dominick, D., Nadzir, M.S.M., Sahani, M., Tahir, N.M., 2015. Seasonal effect and source apportionment of polycyclic aromatic hydrocarbons in PM2. 5. Atmos. Environ. 106, 178–190. https://doi.org/10.1016/j.atmosenv.2015.01.077.
Kim, D., Kumfer, B.M., Anastasio, C., Kennedy, I.M., Young, T.M., 2009. Environmental ageing of polycyclic aromatic hydrocarbons on soot and its effect on source identification. Chemosphere 76, 1075–1081. https://doi.org/10.1016/j.chemosphere.2009.04.031.
Krugly, E., Martuzevicius, D., Sidaraviciute, R., Ciuzas, D., Prasauskas, T., Kauneliene, V., Stasiulaitiene, I., Kliucininkas, L., 2014. Characterization of particulate and vapor phase polycyclic aromatic hydrocarbons in indoor and outdoor air of primary schools. Atmos. Environ. 82, 298–306. https://doi.org/10.1016/j.atmosenv.2013.10.042.
Liu, Y., Yu, Y., Liu, M., Lu, M., Ge, R., Li, S., Liu, X., Dong, W., Qadeer, A., 2018. Characterization and source identification of PM2. 5-bound polycyclic aromatic hydrocarbons (PAHs) in different seasons from Shanghai, China. Sci. Total Environ. 644, 725–735. https://doi.org/10.1016/j.scitotenv.2018.07.049.
Majd, E., McCormack, M., Davis, M., Curriero, F., Berman, J., Connolly, F., Leaf, P., Rule, A., Green, T., Clemons-Erby, D., Gummerson, C., 2019. Indoor air quality in inner-city schools and its associations with building characteristics and environmental factors. Environ. Res. 170, 83–91. https://doi.org/10.1016/j.envres.2018.12.012.
Mitchell, R., Frank, E., 2017. Accelerating the XGBoost algorithm using GPU computing. PeerJ Computer Science 3, e127. https://doi.org/10.7717/peerj-cs.127.
Nielsen, D., 2016. Tree Boosting with XGBoost. NTNU Norwegian University of Science and Technology.
Oliveira, M., Slezakova, K., Delerue-Matos, C., Pereira, M.C., Morais, S., 2019. Children environmental exposure to particulate matter and polycyclic aromatic hydrocarbons and biomonitoring in school environments: a review on indoor and outdoor exposure levels, major sources and health impacts. Environ. Int. 124, 180–204. https://doi.org/10.1016/j.envint.2018.12.052.
Pehnec, G., Jakovljević, I., Godec, R., Štrukil, Z.S., Žero, S., Huremović, J., Džepina, K., 2020. Carcinogenic organic content of particulate matter at urban locations with different pollution sources. Science of The Total Environment, p. 139414. https://doi.org/10.1016/j.scitotenv.2020.139414.
Perišić, M., Rajšić, S., Šoštarić, A., Mijić, Z., Stojić, A., 2017. Levels of PM_{10}-bound species in Belgrade, Serbia: spatio-temporal distributions and related human health risk estimation. Air Quality, Atmosphere & Health 10 (1), 93–103. https://doi.org/10.1007/s11869-016-0411-6.
Pongpiachan, S., Iijima, A., 2016. Assessment of selected metals in the ambient air PM_{10} in urban sites of Bangkok (Thailand). Environ. Sci. Pollut. Control Ser. 23 (3), 2948–2961. https://doi.org/10.1007/s11356-015-5877-5.
Ravindra, K., Bencs, L., Wauters, E., De Hoog, J., Deutsch, F., Roekens, E., Bleux, N., Bergmans, P., Van Grieken, R., 2006. Seasonal and site specific variation in vapor and aerosol phase PAHs over Flanders (Belgium) and their relation with anthropogenic activities. Atmos. Environ. 40, 771–785. https://doi.org/10.1016/j.atmosenv.2005.10.011.
Sarigiannis, D.A., Karakitsios, S.P., Zikopoulos, D., Nikolaki, S., Kermenidou, M., 2015. Lung cancer risk from PAHs emitted from biomass combustion. Environ. Res. 137, 147–156. https://doi.org/10.1016/j.envres.2014.12.009.
Sheridan, R.P., Wang, W.M., Liaw, A., Ma, J., Gifford, E.M., 2016. Extreme gradient boosting as a method for quantitative structure–activity relationships. J. Chem. Inf. Model. 56 (12), 2353–2360. https://doi.org/10.1021/acs.jcim.6b00591.
Shi, S., 2018. Contributions of indoor and outdoor sources to airborne polycyclic aromatic hydrocarbons indoors. Build. Environ. 131, 154–162. https://doi.org/10.1016/j.buildenv.2018.01.001.
Sievert, C., 2020. Interactive Web-Based Data Visualization with R, Plotly, and Shiny. CRC Press.
Šoštarić, A., Stojić, S.S., Vuković, G., Mijić, Z., Stojić, A., Gržetić, I., 2017. Rainwater capacities for BTEX scavenging from ambient air. Atmos. Environ. 168, 46–54. https://doi.org/10.1016/j.atmosenv.2017.08.045.
Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Total Environ. 653, 140–147. https://doi.org/10.1016/j.scitotenv.2018.10.368.
Stojić, A., Stojić, S.S., Reljin, I., Čabarkapa, M., Šoštarić, A., Perišić, M., Mijić, Z., 2016. Comprehensive analysis of PM 10 in Belgrade urban area on the basis of long-term measurements. Environ. Sci. Pollut. Control Ser. 23 (11), 10722–10732. https://doi.org/10.1007/s11356-016-6266-4.
Stojić, A., Stojić, S.S., Šoštarić, A., Ilić, L., Mijić, Z., Rajšić, S., 2015. Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling. Environ. Sci. Pollut. Control Ser. 22 (17), 13137–13152. https://doi.org/10.1007/s11356-015-4540-5.
Tasdemir, Y., Esen, F., 2007. Urban air PAHs: concentrations, temporal changes and gas/particle partitioning at a traffic site in Turkey. Atmos. Res. 84 (1), 1–12. https://doi.org/10.1016/j.atmosres.2006.04.003.
Tobiszewski, M., Namieśnik, J., 2012. PAH diagnostic ratios for the identification of pollution emission sources. Environ. Pollut. 162, 110–119. https://doi.org/10.1016/j.envpol.2011.10.025.
Uchiyama, S., Tomizawa, T., Tokoro, A., Aoki, M., Hishiki, M., Yamada, T., Tanaka, R., Sakamoto, H., Yoshida, T., Bekki, K., Inaba, Y., 2015. Gaseous chemical compounds in indoor and outdoor air of 602 houses throughout Japan in winter and summer. Environ. Res. 137, 364–372. https://doi.org/10.1016/j.envres.2014.12.005.
United States Environmental protection Agency (USEPA), 1997. Office of Pesticide Programs: List of Chemicals Evaluated for Carcinogenic Potential (Washington, DC, USA).
Wang, F., Lin, T., Feng, J., Fu, H., Guo, Z., 2015a. Source apportionment of polycyclic aromatic hydrocarbons in PM2. 5 using positive matrix factorization modeling in Shanghai, China. Environ. Sci. J. Integr. Environ. Res.: Processes & Impacts 17 (1), 197–205. https://doi.org/10.1039/C4EM00570H.
Wang, G., Wang, Y., Yin, W., Xu, T., Hu, C., Cheng, J., Hou, J., He, Z., Yuan, J., 2020. Seasonal exposure to $PM_{2.5}$-bound polycyclic aromatic hydrocarbons and estimated lifetime risk of cancer: a pilot study, 702. Science of The Total Environment, p. 135056. https://doi.org/10.1016/j.scitotenv.2019.135056.
Wang, J., Guinot, B., Dong, Z., Li, X., Xu, H., Xiao, S., Ho, S.S.H., Liu, S., Cao, J., 2017. PM2. 5-bound polycyclic aromatic hydrocarbons (PAHs), oxygenated-PAHs and phthalate esters (PAEs) inside and outside middle school classrooms in Xi'an, China: concentration, characteristics and health risk assessment. Aerosol and Air Quality Research 17 (7), 1811–1824. https://doi.org/10.4209/aaqr.2017.03.0109.
Wang, T., Feng, W., Kuang, D., Deng, Q., Zhang, W., Wang, S., He, M., Zhang, X., Wu, T., Guo, H., 2015b. The effects of heavy metals and their interactions with polycyclic aromatic hydrocarbons on the oxidative stress among coke-oven workers. Environ. Res. 140, 405–413. https://doi.org/10.1016/j.envres.2015.04.013.
Wang, Y.Q., 2014. MeteoInfo: GIS software for meteorological data visualization and analysis. Meteorol. Appl. 21 (2), 360–368. https://doi.org/10.1002/met.1345.
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. https://doi.org/10.1007/978-3-319-24277-4, 2009.
Xu, H., Guinot, B., Niu, X., Cao, J., Ho, K.F., Zhao, Z., Ho, S.S.H., Liu, S., 2015. Concentrations, particle-size distributions, and indoor/outdoor differences of polycyclic aromatic hydrocarbons (PAHs) in a middle school classroom in Xi'an, China. Environ. Geochem. Health 37 (5), 861–873. https://doi.org/10.1007/s10653-014-9662-z.
Yao, Y., Wang, D., Ma, H., Li, C., Chang, X., Low, P., Hammond, S.K., Turyk, M.E., Wang, J., Liu, S., 2019. The impact on T-regulatory cell related immune responses in rural women exposed to polycyclic aromatic hydrocarbons (PAHs) in household air pollution in Gansu, China: a pilot investigation. Environ. Res. 173, 306–317. https://doi.org/10.1016/j.envres.2019.03.053.
Yin, H., Xu, L., 2018. Comparative study of PM10/PM2. 5-bound PAHs in downtown Beijing, China: concentrations, sources, and health risks. J. Clean. Prod. 177, 674–683. https://doi.org/10.1016/j.jclepro.2017.12.263.
Zhang, G., Ma, K., Sun, L., Liu, P., Yue, Y., 2020. Seasonal pollution characteristics, source apportionment and health risks of PM2. 5-bound polycyclic aromatic hydrocarbons in an industrial city in northwestern China. Hum. Ecol. Risk Assess. 1–18. https://doi.org/10.1080/10807039.2020.1799186.

toxics

MDPI

Article

Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing PAHs Environmental Fate

Gordana Jovanovic [1,2,*], Mirjana Perisic [1,2], Nebojsa Bacanin [2], Miodrag Zivkovic [2], Svetlana Stanisic [2], Ivana Strumberger [2], Filip Alimpic [1] and Andreja Stojic [1,2]

[1] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 11000 Belgrade, Serbia; mirjana.perisic@ipb.ac.rs (M.P.); filip.alimpic@ipb.ac.rs (F.A.); andreja.stojic@ipb.ac.rs (A.S.)

[2] Faculty of Informatics and Computing, Singidunum University, 11000 Belgrade, Serbia; nbacanin@singidunum.ac.rs (N.B.); mzivkovic@singidunum.ac.rs (M.Z.); istrumberger@singidunum.ac.rs (I.S.)

* Correspondence: gordana.vukovic@ipb.ac.rs

Abstract: Polycyclic aromatic hydrocarbons (PAHs) refer to a group of several hundred compounds, among which 16 are identified as priority pollutants, due to their adverse health effects, frequency of occurrence, and potential for human exposure. This study is focused on benzo(a)pyrene, being considered an indicator of exposure to a PAH carcinogenic mixture. For this purpose, we have applied the XGBoost model to a two-year database of pollutant concentrations and meteorological parameters, with the aim to identify the factors which were mostly associated with the observed benzo(a)pyrene concentrations and to describe types of environments that supported the interactions between benzo(a)pyrene and other polluting species. The pollutant data were collected at the energy industry center in Serbia, in the vicinity of coal mining areas and power stations, where the observed benzo(a)pyrene maximum concentration for a study period reached 43.7 ng m^{-3}. The metaheuristics algorithm has been used to optimize the XGBoost hyperparameters, and the results have been compared to the results of XGBoost models tuned by eight other cutting-edge metaheuristics algorithms. The best-produced model was later on interpreted by applying Shapley Additive exPlanations (SHAP). As indicated by mean absolute SHAP values, the temperature at the surface, arsenic, PM_{10}, and total nitrogen oxide (NOx) concentrations appear to be the major factors affecting benzo(a)pyrene concentrations and its environmental fate.

Keywords: machine learning; extreme gradient boosting; metaheuristics optimization; swarm intelligence; explainable artificial intelligence; sine cosine algorithm; benzo(a)pyrene

Citation: Jovanovic, G.; Perisic, M.; Bacanin, N.; Zivkovic, M.; Stanisic, S.; Strumberger, I.; Alimpic, F.; Stojic, A. Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing PAHs Environmental Fate. *Toxics* **2023**, *11*, 394. https://doi.org/10.3390/toxics11040394

Academic Editor: Daniel Alan Vallero

Received: 25 February 2023
Revised: 17 April 2023
Accepted: 19 April 2023
Published: 21 April 2023

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) refer to a group of several hundred species with two to seven fused benzene rings, generated via incomplete combustion of organic substances, in the high temperature or pressure process. The majority of these polluting species are persistent, bioaccumulative, light sensitive, heat and corrosion resistant, and emitted from both natural and anthropogenic sources, with the latter being dominant in urban areas.

The concentrations of PAHs in the atmosphere are dependent on the number and quality of air pollutant emission sources, regional meteorological conditions, season, measurement site characteristics, as well as other factors which contribute to their dispersion and have an impact on atmospheric chemistry, dry or wet deposition, and finally, pollutant half-lives and their mutual interactions [1].

After being released, mostly as part of vehicle exhaust and emissions from biomass and fossil fuel burning, PAHs are distributed to all environmental compartments, adsorbed

to airborne particle matter, and deposited on terrestrial and water surfaces. Their concentrations are particularly high in the cold season, as a result of increased fossil fuel burning, reduced thermal and photo-decomposition, and a low planetary boundary layer. Apart from their common sources in urban areas, Hoffer et al. [2] reported that municipal incineration of plastic waste in urban areas emits up to 750 more PAHs than the combustion of dry firewood under the same conditions, and estimated that these emissions were dominated by 4–6 ring PAHs, which are up to 4100 times more toxic than the ones emitted from wood combustion.

In the atmosphere, PAHs are found in the gaseous phase or, more often, adsorbed onto suspended particles. The U.S. EPA has listed 16 compounds as the "priority PAHs" due to their adverse health effects, frequency of occurrence, and potential for human exposure. As regards their impacts on human health, PAHs have obtained significant attention due to the toxicity of low molecular weight species, being the most abundant in the gas phase, and the carcinogenic potential of heavy molecular weight compounds, being mostly particle-bound [3]. The smaller the particle size, the higher the share of carcinogenic PAHs, and thus fine aerosol fraction poses excessive risks to human health. In addition to this, PAHs contribute to the high mutagenicity and carcinogenicity of suspended particles through reactions with atmospheric oxidants, which result in the formation of secondary species [4].

In Serbia, the use of low-quality lignite coal is the major cause of low air quality. While domestic fuel burning (wood, coal, and gas) contributes to global $PM_{2.5}$ and PM_{10} emissions with 20% and 15%, respectively. Karagulian et al. [5] estimated that these contributions amount to 32% and 45% in Central and Eastern Europe, respectively. Since it has been recognized as an indicator of total exposure to carcinogenic PAHs, the benzo(a)pyrene (B[a]P) presence is regularly monitored. To prevent and reduce harmful effects on human health and the environment, a European Directive has set a target value of 1 ng m^{-3} for the total content of B[a]P in the PM_{10} fraction, averaged over a calendar year. In this study, we have used the pollutant data from Lazarevac, an energy industry center in the vicinity of the coal mining areas and power stations, where the observed B[a]P concentrations have occasionally reached 30 ng m^{-3}. In comparison to this, Elzein et al. [6] reported the ∑17-PAHs concentrations have ranged from 2.6 and 31.2 ng m^{-3}, and 8.4 to 42.9 ng m^{-3}, in Beijing and Delhi, respectively. Previous studies have confirmed that the residents of the coal mining regions face an incremental lifetime cancer risk which is significantly higher than the target value [7].

In this study, based on our previous research [8–15], we have applied a novel approach based on the XGBoost model to identify the factors which are mostly associated with the observed B[a]P concentrations and the environmental conditions which support and facilitate B[a]P level dynamics and its interactions with other polluting species. The XGBoost itself is an efficient model; nevertheless, its hyperparameters require tuning for each particular prediction task in order to achieve good performance on the observed dataset. Manual tuning of the hyperparameters is an extremely slow, time-consuming, and error-prone task that is considered to be NP-hard by nature. To address this, a variant of the SCA metaheuristics algorithm [16] has been used to optimize the XGBoost hyperparameters. Metaheuristics algorithms, being stochastic by nature, have been established as a common choice to tackle NP-hard challenges. By performing simulations for the sake of this research, the most promising metaheuristics algorithm was determined to be the sine cosine algorithm (SCA); in other words, it was selected empirically. Moreover, this paper also proposes a modified version of SCA, by hybridizing it with another algorithm, to cancel the limitations of the elementary SCA. Modified SCA was later utilized as a part of the machine learning framework, and tasked to tune the collection of the XGBoost hyperparameters for this problem. The results attained by the proposed model have been compared to the results of XGBoost models tuned by eight other cutting-edge metaheuristics algorithms. The best-produced model was later on interpreted by applying Shapley Additive exPlanations (SHAP).

2. Background

2.1. XGBoost

The XGBoost algorithm utilizes an adaptive training method for optimizing its objective function, where each step in the optimization process relies on the outcome of the previous step. The mathematical representation of the XGBoost model's objective function has been defined as follows:

$$F_o{}^i = \sum_{k=1}^{n} l\left(y_k, \hat{y}_k^{i-1} + f_i(x_k)\right) + R(f_i) + C, \quad (1)$$

where the t-th round loss is denoted by l, y_k and $\hat{y}_k$ denote target observed values and predictions, respectively; f_i are additive functions from the space of the regression trees, constant term is marked as C, while the model's regularization parameter R can be defined as:

$$R(f_i) = \gamma T_i + \frac{\lambda}{2} \sum_{j=1}^{T} w_j^2 \quad (2)$$

where T corresponds to the number of leaves in the tree, while w values denote the scores in the corresponding leaves [17].

In general, the complexity of the tree structure is inversely proportional to the values of the customization parameters γ and λ. The larger the values of these parameters, the simpler the tree structure becomes. The model's first and second derivatives, represented by g and h, respectively, are expressed as follows:

$$g_j = \partial_{\hat{y}_k^{i-1}} l\left(y_j, \hat{y}_k^{i-1}\right) \quad (3)$$

$$h_j = \partial^2_{\hat{y}_k^{i-1}} l\left(y_j, \hat{y}_k^{i-1}\right) \quad (4)$$

The solution is obtained using the next two formulas:

$$w_j^* = -\frac{\sum g_t}{\sum h_t + \lambda} \quad (5)$$

$$F_o{}^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum g)^2}{\sum h + \lambda} + \gamma T, \quad (6)$$

where the loss function score is denoted by $F_o{}^*$, while the solution's weight values are marked by w_j^*.

2.2. Metaheuristics Optimization

NP-hard challenges, a common occurrence in computer science, necessitate the use of stochastic algorithms like metaheuristics because deterministic methods are impractical. Metaheuristics methods can be categorized into various families with respect to the natural phenomenon they utilize to steer the search process, such as evolution or insect behavior [18–20]. The most significant families of metaheuristic algorithms are nature-inspired methods (further divided into genetic algorithms and swarm intelligence), methods established upon certain physical phenomena (e.g., storm, gravity, electromagnetism), algorithms that imitate certain aspects of the human behavior (e.g., teaching and learning, or brainstorming, or actions taken on social media), and approaches based on mathematical laws to guide the search (e.g., trigonometric function oscillations).

Swarm intelligence is established upon the behavior manifested by massive groups comprised of relatively modest units; for example, insects or birds in swarms, that are able to manifest highly coordinated and sophisticated behavioral patterns while they hunt, feed, mate or migrate [21,22]. These algorithms have exhibited high efficiency in solving a variety of the real-world NP-hard challenges. Among many available algorithms, well-known samples are the particle swarm optimization (PSO) [23], the ant colony optimization

(ACO) [24], the firefly algorithm (FA) [25] and the bat algorithm (BA) [26,27]. Recently, a highly efficient group of algorithms were derived from the mathematical functions and their properties to steer the search procedure, including the sine-cosine algorithm (SCA) [16] and the arithmetic optimization algorithm (AOA) [28].

The reason why there is a range of population-based algorithms is due to the no-free-lunch theorem (NFL) [29]. The NFL discloses that no single method can be the best for all optimization problems. Consequently, one algorithm may excel in one task but entirely fail in another, leading to the need for diverse metaheuristics approaches and the requirement to choose a suitable method for each specific optimization challenge.

Recently, population-based algorithms have been a common choice to address numerous real-world problems. The application domains include predicting the number of COVID-19 cases [30,31], fog, cloud and cloud-edge computing systems organization [32–35], wireless sensors and IoT optimization [36–39], feature selection [40], image processing and classifying in medicine [41,42], global tuning challenges [43,44], credit card fraud identification [45,46], tracking and predicting air pollution [47,48], network and computer systems intrusion detection [49,50], and finally, tuning different ML structures [51–56].

2.3. Shapley Additive Explanations

To explain the obtained best-performing model, which is vital for understanding the process being modeled, we have applied the explainable artificial intelligence method SHAP. Avoiding the trade-off between accuracy and interpretability, SHAP provides a straightforward and meaningful interpretation of the model-derived decisions. It is based on Shapley values, calculated as a feature importance measure by a game-theory approach which provides an impact of features on individual predictions [57]. Apportioning the difference between the prediction and the average prediction among the features [58], Shapley values represent fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). SHAP assigns each feature importance as a measure of its contribution to a particular prediction and interprets the impact compared to a model's prediction if that feature took some baseline value (mean). This way, the method provides valuable insights into a model's behavior (1) overcoming the main drawback of inconsistency, (2) minimizing the possibility of underestimating the importance of a feature with a specific attribution value, and (3) capturing feature interaction effects based on a generalization of Shapley values and interpreting the model's global behavior while retaining local faithfulness [15,59]. The main challenges of the method include Shapley value computation and background data choice which can induce uncertain or unintuitive feature attributions.

In this study, we have used the relative SHAP values introduced by Stojic et al. [11] to gain an insight into relative relationships among feature attributions for each prediction. Relative SHAP values, defined as a share of absolute SHAP in total attributed importance of all features for the particular prediction, show the relative influence of a feature on the prediction.

We have used the Python SHAP implementation (SHAP Python package) and TreeExplainer [59] to obtain SHAP values that we have used to produce SHAP dependency plots, representing the change of feature importance over its value range.

3. Methods

3.1. Measurements Methods

For this study, the two-year daily concentrations (2018–2019; 645 observations) of particulate matter PM_{10}, its constituents (Pb, As, Cd, Ni, and B[a]P), and inorganic gaseous pollutants (NO, NO_2, NOx, and SO_2) were obtained from the regulatory air quality monitoring station (44°23′02″ N, 20°15′55″ E) in Lazarevac (Serbia). The meteorological data were obtained from the Global Data Assimilation System (GDAS1) with a 1.0° × 1.0° spatial resolution for the longitude and latitude of the monitoring station.

The Sven Leckel SEQ 47/50-RV sampler was used for collecting 24-h samples of particulate matter. The mass concentrations of PM_{10}, Pb, As, Cd, Ni, and B[a]P were determined according to the standards EN 12341, EN 14902, and EN 15549, while the concentrations of NO, NO_2, NOx, and SO_2 were obtained in accordance with the sampling procedures standardized in EN 14211 and EN 14212.

PM_{10} were collected on quartz filters (Whatman QMA, 47 mm) daily, as described in the Standard SRPS EN 12341:2015 (Ambient air—Standard gravimetric measurement method for the determination of the PM_{10} or $PM_{2.5}$ mass concentration of suspended particulate matter, 2015). The filters were pre-fired to remove organic impurities, and the pre-conditioning of both non-exposed and loaded filters was performed prior to gravimetric measurements.

The concentrations of As, Cd, Cr, Ni, and Pb as PM_{10} constituents were determined as described in the EN 14902:2008/AC:2013 Standard (Ambient air quality—Standard method for the measurement of Pb, Cd, As, and Ni in the PM fraction of suspended particulate matter, 2008). Firstly, the CEN/TC 264 N779 procedure was applied for the extraction of the trace elements. In brief, the pieces of exposed quartz filters were treated with an acidic mixture of HNO_3(c)/30% H_2O_2/H_2O (3/2/5) using analytical grade reagents (Merck) and distilled/deionized water (MiliQ, 18.2 MΩ). The filters were digested in closed 100 mL Teflon vessels in the Anton Paar 3000 microwave accelerated reaction system and the concentrations of trace elements were determined by inductively coupled plasma–mass spectrometry (ICP-MS) (device Agilent 7500ce with Octopole Reaction System). Quality control and verification of the applied procedures for microwave digestion and multi-elemental trace analysis using ICP-MS was conducted using the 2783 NIST (National Institute of Standard and Technology, MD, USA) standard reference material analysis, containing a PM_{10} fraction of urban dust from a mixed industrial urban area of Vienna, collected on a polycarbonate membrane filter. The recovery values were within the satisfactory range of ±20% from the reference value.

B[a]P was determined by the procedure described in the SRPS ISO 12884:2010 Standard (Ambient air—Determination of total (gas and particle-phase) polycyclic aromatic hydrocarbons—Collection on sorbent-backed filters with gas chromatographic/mass spectrometric analyses, 2010). Parts of the exposed filters underwent a microwave extraction procedure with a solvent mixture of n-hexane and acetone (12.5 mL n-hexane: 12.5 mL acetone) according to EPA method 3546. After extraction, the solution volume was reduced by rotary evaporation under reduced pressure (55.6 kPa and 0.2 mL iso-octane) to 1 mL. Afterward, the n-hexane solution was reduced to 0.25 mL under a nitrogen stream. Known quantities of internal standards were added to estimate the method recovery. B[a]P was analyzed using gas chromatography coupled with a mass selective detector (Agilent GC 6890/5973 MSD) according to the EPA compendium method TO-13A with a DB-5 MS capillary column (30 m × 0.25 mm × 25 µm). The oven temperature program started at 70 °C (duration of 4 min) and ramped 8 °C min-1 to the end temperature of 310 °C (duration of 5 min). The solvent delay was 5 min and the run time was 46 min. The calibration curve was obtained by spiking seven different quantities of B[a]P, all with an R2 of the calibration curve above 0.995. Recovery values ranged from 85% to 110% for all the PAHs contained in the internal standard.

The samples were collected at the suburban site located in the energy industry center of Lazarevac, a municipality of Belgrade (Serbia), and a home to 60,000 residents. The sampling location, surrounded by residential areas and sports facilities, is exposed to mining pollutant sources and emissions from household coal and wood fireboxes. Additionally, the nearest coal mine Vreoci and regional power station are located around 5 km east and northeast, while the 80 square kilometers large coal mining area Kolubara, which supplies 75% of Serbia's electricity generation and largest state power plants Nikola Tesla A and B, are located around 10 km north and 30 km northwest of the sampling site, respectively.

3.2. Original Sine Cosine Algorithm

The algorithm, proposed by Mirjalili in 2016 [16], is based on the properties of elementary trigonometrical functions. SCA belongs to the group of population-based metaheuristics that starts each run by producing a set of arbitrary initial solutions within the scope of the search realm. The individual positions update following the swinging behavior of the sine and cosine functions over time. SCA conducts the exploration and exploitation mechanisms steered by the set of four arbitrary control parameters. The fundamental SCA search is mathematically defined by Equation (7):

$$X_i^{t+1} = \begin{cases} X_i^{t+1} = X_i^t + r_1 \cdot sin(r_2) \cdot |r_3 \cdot P_i^{*t} - X_i^t|, & r_4 < 0.5 \\ X_i^{t+1} = X_i^t + r_1 \cdot cos(r_2) \cdot |r_3 \cdot P_i^{*t} - X_i^t|, & r_4 \geq 0.5, \end{cases} \quad (7)$$

where X_i^t and X_i^{t+1} define the individual's position in i-th dimension in a pair of consecutive iterations t and $i+1$, respectively, r_{1-4} are four generated above-mentioned control parameters, the P_i^* defines the target's position (the most recent estimation of the optimal solution) within i-th dimension. Additionally, the fresh values for r_{1-4} parameters are summoned for each component of each solution within the population.

3.3. Proposed Modified Sine Cosine Algorithm

The core implementation of SCA is considered to be an exceptional optimizer; however, like other metaheuristics methods, it also has some drawbacks. Testing using benchmark sets has shown that SCA is effective at exploring solutions, but lacks the ability to effectively exploit these solutions in the later stages of the process. This results in a limited exploration when the algorithm should be focusing on the most promising areas. In contrast, the firefly algorithm (FA) is known for its superior exploitation capability, as described by [25].

This manuscript suggests a hybrid solution by combining SCA and FA algorithms, aiming to profit from the advantages of both metaheuristics, aiming to cancel out each other's disadvantages. At the start of the execution, the solutions within the population will update according to the SCA search procedure, as described by Equation (7). However, in later stages, when it is necessary to narrow down and exploit the favorable regions of the search realm, the exploitation phase is backed up by employing the powerful FA search mechanism, defined by Equation (8):

$$X_i^{t+1} = X_i^t + \beta_0 \cdot e^{-\gamma r_{i,j}^2}(X_j^t - X_i^t) + \alpha^t(\kappa - 0.5) \quad (8)$$

where α represents the randomization variable, κ is an arbitrary value drawn from the Gaussian distribution. Finally, the space between solutions i and j is denoted as $r_{i,j}$.

A couple of new control parameters have been suggested to steer the alternation between the two search procedures in the later stage of the execution. The varying search vs control parameter is used to activate the combined search mode in the case where $t > vs$, when the suggested approach should alternate between SCA and FA search methods. Variable vs is initially set as $maxIter/5$, that has been determined empirically.

The second control parameter, named search mode sm, determines for every individual solution in the population whether to proceed with the SCA or FA search option. Each solution produces a random value rnd in range $[0, 1]$, and if $rnd < sm$ it will perform an SCA search, or otherwise continue with the FA search option. The value of this parameter is dynamically reduced over the iterations, giving an additional focus on a stronger FA search in the latter rounds. Initially, sm is set to 0.8, being reduced over time according to Equation (9).

$$sm_t = sm_{t-1} - (sm_{t-1}/10) \quad (9)$$

The hybrid algorithm is labeled hybrid self-adaptive SCA (HSA-SCA), and its pseudocode summarizing the most significant steps of the approach is provided by Algorithm 1.

Algorithm 1 Pseudocode of the HSA-SCA metaheuristics

Spawn a collection of starting solutions (X)
while $t < maxIter$ **do**
  validate each individual in terms of its fitness
  for each individual inside (X) **do**
    if $t < vs$ **then**
      Perform SCA search mechanism, provided by Equation (7)
    else
      if $rnd < sm$ **then**
        Perform SCA search mechanism, provided by Equation (7)
      else
        Perform FA search mechanism, provided by Equation (8)
      end if
    end if
  end for
end while
return The current fittest solution determined as the global optimum

4. Experimental Findings and Comparative Analysis

This section first provides insights into the dataset preprocessing, implementation technology, and evaluation metrics used to evaluate different tuned XGBoost models, followed by experimental setup, results, and comparative analysis. Finally, to validate improvements of devised hybrid metaheuristics over other baseline cutting-edge approaches, statistical tests were conducted, as suggested in the state-of-the-art AI literature [60].

4.1. Dataset Preprocessing, Implementation Technology and Evaluation Metrics

As already pointed out in Section 3, the employed dataset includes 645 observations. The challenge is formulated as a regression problem, where the feature with daily values for B[a]P was set as the target. Since the XGBoost is a tree-based method, scaling values, e.g., within the range $[0, 1]$, were not needed; therefore original measured values were used.

However, since the XGBoost requires training, the dataset was divided into train and test, where 70% of observations were used for training and 30% for testing. The same split was used for all metaheuristics considered for comparative analysis and the same pseudo-random number seed was employed, with the goal of establishing fair comparison conditions. It is noted that during the pre-experimentation, simulations with validation test were also conducted; however, improvements could not be achieved, and therefore it was decided to proceed with only training and testing data. Visual representation of the dataset split for the target variable is shown in Figure 1.

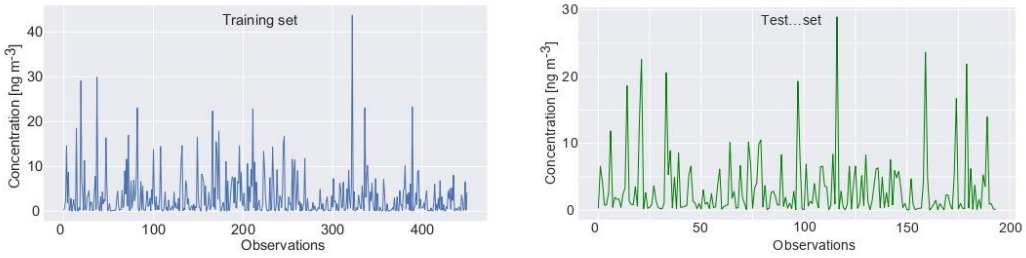


Figure 1. The Benzo(a)pyrene feature dataset split.

The analysis was conducted using daily concentrations and daily mean meteorological parameters, which have minimal to no autocorrelation in such a short period. Moreover, atmospheric processes relevant to air pollution dynamics usually occur within an hour, making autocorrelations even less prominent when using daily data.

The simulation environment, along with all methods, was implemented in Python using data science and ML libraries: *numpy*, *pandas*, *scikitlearn*, *xgboost*, *matplotlib*, *seaborn* and *shap*. Code snippets of the simulation framework along with the best generated XGBoost model by proposed HSA-SCA approach is available at the following URL: https://doi.org/10.5281/zenodo.7831739 (accessed on 25 February 2023).

The XGBoost model's experimental results have been evaluated by a set of traditional machine learning metrics, including mean squared error (MSE) defined by Equation (10), root mean squared error (RMSE) obtainable by Equation (11), mean absolute error (MAE) calculated by Equation (13), and the coefficient of determination (R2) described with Equation (13).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{a}_i - a_i)^2 \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{a}_i - a_i)^2} \tag{11}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{a}_i - a_i| \tag{12}$$

$$R2 = 1 - \frac{\sum_{i=1}^{n}(a_i - \hat{a}_i)^2}{\sum_{i=1}^{n}(a_i - \bar{a})^2}, \tag{13}$$

where a_i and $\hat{a}_i$ represent arrays comprised of the observed values that are predicted, and predicted values, both with length N. This paper utilizes MSE as the fitness function that is required to be minimized.

Additionally, according to [61,62], the index of agreement (IA) can be an insightful statistical measure used to evaluate the performance of a model or forecast in predicting a particular event or phenomenon, as well as the metric for the best-generated models. The IA can be calculated as the ratio of the MSE and the potential error that is varying in range $[0, 1]$, where the value of 1 suggests perfect agreement, while the value of 0 suggests no match at all. The Eq. 14 shows how the IA value is obtained:

$$IA = 1 - \frac{\sum_{i=1}^{n}(a_i - \hat{a}_i)^2}{\sum_{i=1}^{n}(|\hat{a}_i - \bar{a}| + |a_i - \bar{a}|)^2}, 0 \leq IA \leq 1, \tag{14}$$

where a_i and $\hat{a}_i$ again denote arrays comprised of the observed and predicted values, and the $\bar{a}$ are average observed values.

4.2. Experimental Setup

The proposed HSA-SCA algorithm was tasked to optimize the XGBoost model for this particular dataset. The set of optimized XGBoost hyperparameters, accompanied by their corresponding search limits and variable types are provided as follows:

- learning rate (η), search limits: $[0.1, 0.9]$, continuous variable,
- *min_child_weight*, search limits: $[1, 10]$, continuous variable,
- subsample, search limits: $[0.01, 1]$, continuous variable,
- collsample_bytree, search limits: $[0.01, 1]$, continuous variable,
- max_depth, search limits: $[3, 10]$, integer variable and
- *gamma*, search limits: $[0, 0.8]$, continuous variable.

The parameter counts of the *softprob objective function* ('num_class':self.no_classes)were passed as a parameter to the XGBoost, while the remainder of the XGBoost parameters were set to XGBoost defaults during the simulations.

The suggested method has been implemented in the Python programming language, accompanied by the standard collection of Python libraries related to machine learnin including scipy, numpy, and pandas, while the XGBoost model was acquired from the scikit-learn package.

The proposed setup utilizes the solutions' encoding scheme that observes each solution as an array with length l, where l denotes the number of optimized hyperparameters. Hence, the value l has been set to six, to match the tuned parameters.

Aiming to validate the performance of the XGBoost model tuned by the suggested HSA-SCA algorithm, the achieved results were compared to the results attained by eight other contending powerful algorithms. The comparisons were executed with elementary SCA, genetic algorithm (GA) [63,64], PSO [23], ABC [65], FA [25], whale optimization algorithm (WOA) [66], harris hawks' optimization (HHO) [67] and chimp optimization algorithm (ChOA) [68]. Every contending algorithm has been implemented independently by the authors of this manuscript, with the control parameters set to the recommended values from their respective publications. Each algorithm has been given the same task, to tune the same set of XGBoost hyperparameters.

All metaheuristics algorithms were tested with 40 solutions in the population and 20 iterations per run, over the course of 20 separate runs. As previously noted, MSE was set as the fitness function that needs to be minimized.

4.3. Experimental Findings and Comparative Analysis

This section yields the attained experimental outcomes, for the observed HSA-SCA algorithm and other contenders. Tables 1 and 2 show the simulation outcomes with respect to the fitness function, accompanied by the detailed metrics achieved in the best individual run of each algorithm, where the best results in each category are marked in bold.

Table 1 shows detailed comparisons with respect to the fitness function (MSE) attained by XGBoost models optimized by the nine regarded algorithms (the proposed HSA-SCA and eight contenders). The results suggest that the HSA-SCA method displayed a supreme performance level, by achieving the best scores for key performance indicators (best, worst, mean, and median). FA scored the best results for standard deviation and variance, by delivering the most stable results. The second best result with respect to the best, worst, mean and median run values was also FA, followed by PSO and ChOA. The best attained score by the HSA-SCA XGBoost model was the MSE of 2.468293, and R^2 of 0.892845.

Table 1. Comparative results of the objective function (MSE) of the observed metaheuristics.

Method	HSA-SCA	SCA	GA	PSO	ABC	FA	WOA	HHO	ChOA
Best	**2.468293**	3.184137	3.192827	3.077008	3.153481	2.590850	3.206221	3.180909	3.129932
Worst	**2.893362**	3.605363	3.635218	3.773639	3.774347	2.918338	3.685749	3.584440	3.561904
Mean	**2.731538**	3.443475	3.413906	3.390799	3.369339	2.771516	3.466122	3.363900	3.379943
Median	**2.725915**	3.479927	3.423980	3.413375	3.306318	2.779478	3.472304	3.358337	3.371960
Std	0.114964	0.112574	0.144976	0.214997	0.157270	**0.097592**	0.109602	0.130492	0.112995
Var	0.013217	0.012673	0.021018	0.046224	0.024734	**0.009524**	0.012013	0.017028	0.012768

Table 2 presents the detailed metrics achieved in the best single run of all regarded algorithms. Once more, it can clearly be seen that the proposed HSA-SCA dominantly outperformed contenders in terms of all indicators—R^2, R, MSE, RMSE and IA, except MAE, where FA achieved the best score. Looking into the MSE that has been employed as the fitness function with a goal to minimize it, HSA-SCA exhibited superior performance with the score of 2.468293, in front of the FA that scored 2.590850, PSO in third place that achieved 3.077008, and ChOA finishing fourth with the score of 3.129932. In terms of the

IA metric, the proposed HSA-SCA was also superior, attaining the value of 0.970348. FA finished second, with IA value of 0.967438, while WOA was third with IA value of 0.961817.

Table 2. Detailed metrics for the best individual run of the observed metaheuristics.

	R^2	R	MAE	MSE	RMSE	IA
HSA-SCA	**0.892845**	**0.944905**	0.987179	**2.468293**	**1.571080**	**0.970348**
SCA	0.861769	0.928315	1.081976	3.184137	1.784415	0.960579
GA	0.861392	0.928112	1.056114	3.192827	1.786848	0.958925
PSO	0.866420	0.930817	1.096070	3.077008	1.754140	0.959703
ABC	0.863100	0.929032	1.085311	3.153481	1.775804	0.959175
FA	0.887525	0.942085	**0.981363**	2.590850	1.609612	0.967438
WOA	0.860810	0.927799	1.036537	3.206221	1.790592	0.961817
HHO	0.861909	0.928391	1.143855	3.180909	1.783510	0.961245
ChOA	0.864122	0.929582	1.048337	3.129932	1.769161	0.960678

Lastly, the sets of the XGBoost hyperparameters that have been established by the best run of every algorithm are provided within the Table 3. The best performing method, that was the proposed HSA-SCA, produced the XGBoost model with a learning rate of 0.535844, max_child_weight of 4.768378, a subsample of 0.920331, collsample_bytree of 0.899994, max_depth of 5, and gamma value of 0.037125. The XGBoost structure produced by the FA algorithm, that finished in second place, consisted of the learning rate of 0.473028, max_child_weight of 5.459757, a subsample value of 0.937122, collsample_bytree of 1.000000, max_depth value of 7, and finally, gamma value of 0.318114.

Table 3. Best solutions' determined XGBoost hyper-parameters set.

Method	l.r. (μ)	Max_child_weight	Subsample	Collsample_bytree	Max_depth	Gamma
HSA-SCA	0.535844	4.768378	0.920331	0.899994	5	0.037125
SCA	0.424673	6.830426	0.903051	1.000000	10	0.800000
GA	0.515505	1.239850	0.921408	1.000000	4	0.000000
PSO	0.469675	5.890036	0.966332	0.732996	7	0.514569
ABC	0.424717	6.756257	0.910125	0.797826	7	0.349785
FA	0.473028	5.459757	0.937122	1.000000	7	0.318114
WOA	0.518772	6.961853	0.976281	0.978017	4	0.408959
HHO	0.533272	6.254540	1.000000	1.000000	10	0.800000
ChOA	0.388340	2.995555	0.766726	1.000000	8	0.000000

The performed simulations are visualized in Figures 2 and 3, showing the convergence graphs, box plots, violin plots and swarm plots of all nine algorithms, for both fitness function (Figure 2) and R2 (Figure 3).

While looking into the Figures 2 and 3, it is possible to see that the HSA-SCA method exhibits a very fast converging speed, together with FA metaheuristics, that is a little bit faster at the beginning, but finishing behind HSA-SCA at the end. One can note that FA also exhibits the most stable results, closely followed by the WOA and SCA, as can be seen from the box plot diagrams. Finally, the swarm plots show the diversity of the population within the last round of execution of the best run of each algorithm. It is possible to conclude that all solutions of the HSA-SCA population were proximal to the optimum value.

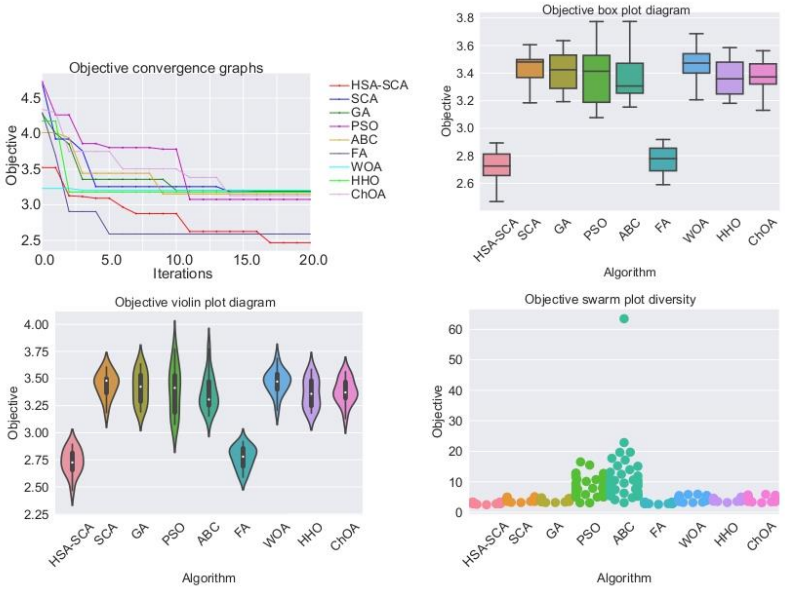


Figure 2. Visualized XGBoost results for all nine metaheuristics in terms of the convergence, box plot, violin diagrams, and swarm diversity plots for the fitness function (MSE).

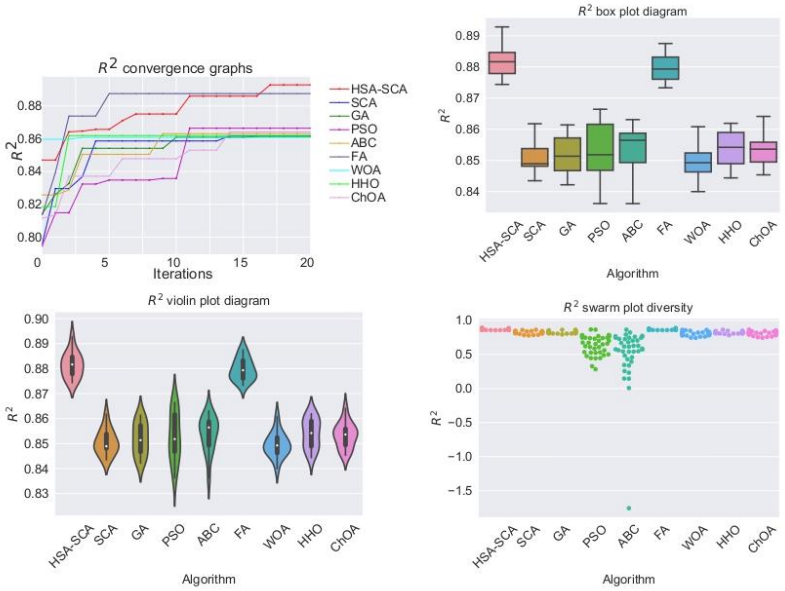


Figure 3. Visualized XGBoost results for all nine metaheuristics in terms of the convergence, box plot, violin diagrams, and swarm diversity plots for the R^2 indicator.

Figure 4 depicts the kernel density estimation (KDE), representing the estimation of the probability density function. It can be noted from these plots that the results originate from the normal distribution. Additionally, join plots of both fitness function (MSE) and R2 containing histograms for the two best algorithms (HSA-SCA and FA) are shown in Figure 5.

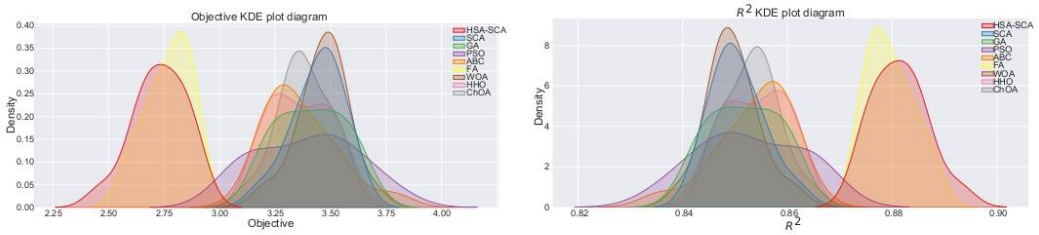


Figure 4. KDE diagrams for MSE (**left**) and R^2 indicator (**right**).

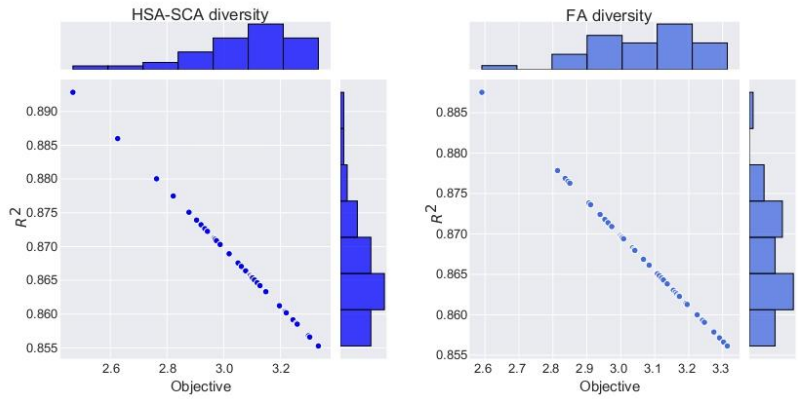


Figure 5. Join plots with histograms of two best methods: HSA-SCA (**left**) and FA (**right**).

Finally, the visualizations of the best-predicted outcomes attained by the best-produced model by four best algorithms is shown in Figure 6. Once more, it can be concluded that the model optimized by the HSA-SCA algorithm produced the best predictions for the observed problem.

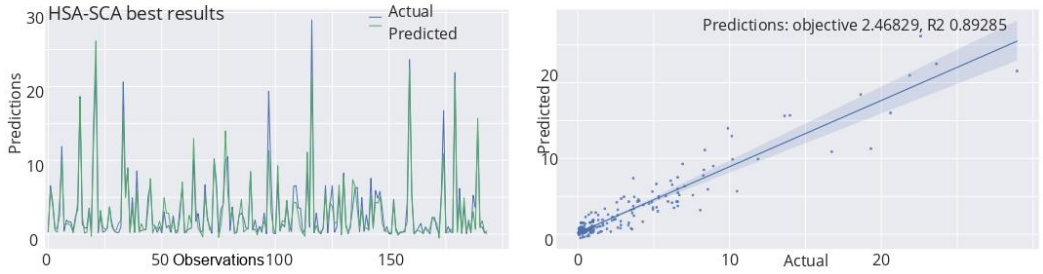


Figure 6. *Cont.*

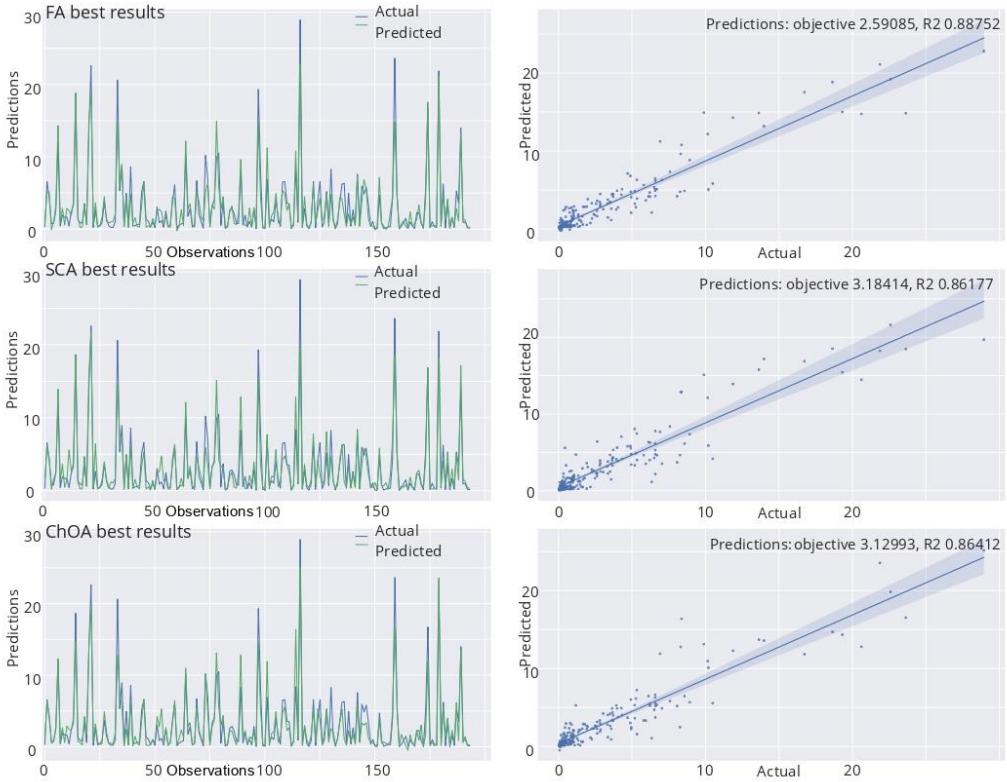


Figure 6. Best-predicted outcomes by the best produced models of HSA-SCA, FA, SCA and ChOA algorithms.

4.4. Results Improvements Validation—Statistical Tests

To further evaluate the obtained simulation results and determine whether or not they are statistically significant, the best scores of each of 20 runs from each observed metaheuristics were gathered and inspected as a data series. At the beginning, it was necessary to decide what sort of statistical tests was suitable—parametric or non parametric. First, the safe usage of parametric tests is checked, by evaluation of the independence, normality, and homoscedasticity of the data variances [69]. The independence condition is satisfied, because every run of the metaheuristics algorithms begins by producing a collection of random individuals. Considering the second condition, homoscedasticity, Levene's test [70] was executed, and since the p-value of 0.65 was obtained in every case, it was safe to assume that the homoscedasticity requirement was also fulfilled.

The normality condition has been investigated by employing the Shapiro-Wilk single problem analysis [71]. Shapiro-Wilk p-values were independently calculated in terms of each of the observed methods. The established p-values for every algorithm were greater than 0.05, therefore it was safe to conclude that it was not possible to reject the H0 hypothesis for both $alpha = 0.05$ and $alpha = 0.1$. As a consequence, it was also possible to conclude that the observed values originated from the normal distribution. One could establish a similar conclusion by simply taking a look at the KDE plots shown in Figure 4. The Shapiro-Wilk test scores are provided within Table 4.

Table 4. Shapiro-Wilk scores for validating the normality condition.

Methods	HSA-SCA	SCA	GA	PSO	ABC	FA	WOA	HHO	ChOA
	0.362	0.351	0.198	0.145	0.312	0.304	0.342	0.263	0.347

After verifying that the normality requirement was fulfilled, it was safe to conclude that one can proceed by applying the parametric tests. This paper utilizes the paired-*t* test [72], which is frequently selected to evaluate metaheuristics methods [73]. Paired-*t* test can be utilized if it is possible to observe the set of data points as paired measurements, and the differences among the pairs follows a normal distribution. In other words, the variances between samples for every pair of algorithms are required to be normally distributed. To inspect this condition, the Shapiro-Wilk test was employed one more time, over the absolute differences between distributions of the proposed algorithm and other contending methods. The obtained Shapiro-Wilk *p*-values were in every instance greater than the threshold value $alpha = 0.05$, meaning that H0 hypothesis cannot be rejected, and the set of observed values originates from the normal distribution. Since this is the prerequisite for using the paired-*t* test, it is safe to use it and compare the proposed algorithm against each of the opposing methods. The summarized results of both Shapiro-Wilk *p*-values calculated as the prerequisite for the paired-*t* test, and the paired-*t* test itself, are provided in Table 5.

The results of the paired-*t* test show that the *p*-values was smaller then 0.05 for all algorithms. Accordingly, it can be concluded that the introduced HSA-SCA approach is significantly superior over all contenders for both thresholds $alpha = 0.1$ and $alpha = 0.05$.

Table 5. Shapiro-Wilk scores over the mean differences between two samples as prerequisite for paired-*t* test, accompanied by the paired-*t* test results

Methods vs. HSA-SCA	SCA	GA	PSO	ABC	FA	WOA	HHO	ChOA
Shapiro-Wilk	0.164	0.202	0.241	0.195	0.213	0.224	0.189	0.207
paired-*t* test	0.021	0.024	0.025	0.031	0.041	0.025	0.031	0.033

5. Discussion

The average annual B[a]P concentrations of 3.73 ng m^{-3} and 2.78 ng m^{-3} in 2018 and 2019, respectively (Table 6), significantly exceeded the European Directive set level of 1 ng m^{-3}. The maximum pollutant level reached 43.71 ng m^{-3} in the first year of the study period. At the same time, no values exceeded the critical threshold for the concentration of PM_{10}, As, Cd, Ni, and Pb, and inorganic gaseous pollutants.

Table 6. Descriptive statistics.

Year	Statistics	B[a]P [ng m^{-3}]	PM_{10} [µg m^{-3}]	As [ng m^{-3}]	Cd [ng m^{-3}]	Ni [ng m^{-3}]	Pb [ng m^{-3}]	SO_2 [µg m^{-3}]	NO [µg m^{-3}]	NO_2 [µg m^{-3}]	NOx [µg m^{-3}]
2018	Average	3.73	36.18	1.56	0.41	2.78	4.35	10.82	10.63	17.88	29.96
	Minimum	0.07	10.2	0.5	0.05	1.5	2.5	2.09	0.5	0.5	0.5
	Maximum	43.71	179.7	11.2	27.1	19.6	30.7	25.02	101.49	98.5	189.71
	Median	1.15	29	0.5	0.2	1.5	2.5	12.45	4.61	14.33	19.3
2019	Average	2.78	33.01	1.38	0.2	2.26	4.26	17.21	4.98	11.91	19.43
	Minimum	0.03	4.9	0.5	0.05	1.5	2.5	5.7	1.1	1.75	6.1
	Maximum	23.63	180.7	13.6	1.9	28.5	40.6	66.3	91.2	48	146.81
	Median	1.07	23.3	0.5	0.1	1.5	2.5	15.45	2.75	10.45	14.6

As indicated by the mean absolute SHAP values, the temperature at surface (TMPS), As, PM_{10}, and total nitrogen oxide (NOx) concentrations appear to be major factors for governing B[a]P environmental fate (Table 7). In addition, the most important variables also include NO, SO_2, Pb, and Cd concentrations, as well as the temperature at 2 m (T02M) and

momentum flux intensity (MOFI)m have been shown to affect B[a]P dynamics. However, for this paper, we will focus on the aforementioned four.

Table 7. SHAP values.

Parameter	TMPS	As	PM_{10}	NOx	NO	SO_2	TO2M	Pb	MOFI	LIB4	SHIF	LHTF
Absolute	1.17	0.906	0.796	0.608	0.321	0.247	0.192	0.161	0.158	0.158	0.141	0.131
Relative [%]	22.7	14.36	13.75	9.99	4.99	4.25	3.98	3.38	2.45	2.99	2.18	2.33

5.1. Temperature at Surface

In this study, the temperature at the surface (TMPS) was estimated to be the most important parameter responsible for the B[a]P concentration increase of 1.17 ng m^{-3} on average, while mutual interrelations between TMPS and other studied parameters define three types of environmental conditions being responsible for shaping B[a]P levels. As a high-molecular-weight PAH, B[a]P is dominantly particle-bound in the atmosphere. The B[a]P partition between gas and particles is enhanced during colder months due to low temperature and high atmospheric pressure, which cause intense descending air movements and dry deposition of organic compounds [74]. Additionally, previous studies have shown that higher organic carbon content of particles in the cold season negatively affects the immobilization and biodegradation of PAHs [75], while high temperatures and light intensity in warm months enable both their photo- and biodegradation.

The first type of environment resulting in the increase of B[a]P concentrations up to 3.4 ng m^{-3} (Figure 7), was characterized by medium to low PM_{10}, B[a]P, As, Cd, and Ni levels (35.2 µg m^{-3}, and 3.4, 1.4, 0.3, and 2.5 ng m^{-3} on average, respectively), medium to high NO and NOx concentrations (6.9 and 23.4 µg m^{-3} on average, respectively), and meteorological parameters registered in a wide range of values. The observed constancy of the conditions suggests that this environment type might be related to anthropogenic sources, such as traffic and off-road vehicles.

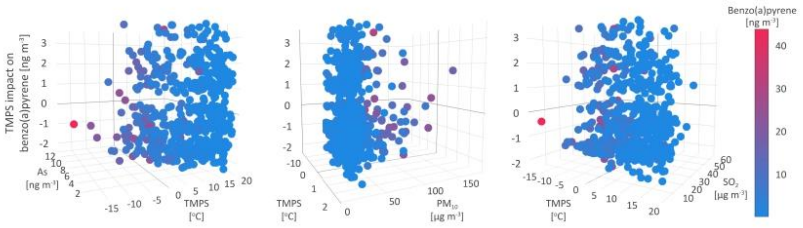


Figure 7. Temperature at surface impact on benzo(a)pyrene.

In the second type of environment, TMPS was ambivalently related to the B[a]P concentrations, leading both to their decrease by up to −1 ng m^{-3} and the increase by up to 0.7 ng m^{-3}. Compared to the first type, the second type of environment was characterized by lower B[a]P, As, Cd, Ni, Pb, PM_{10}, NO_2, and SO_2 (2.0, 1.1, 0.2, 1.8, 3.8 ng m^{-3}, and 32.8, 13.8, and 13.4 µg m^{-3}, respectively) and higher NO and NOx (about 9.5 and 28.8 µg m^{-3}, respectively) mean concentrations. The decrease in temperature range and wind speed (Figure 7) and the rise in relative humidity, alongside other meteorological parameters (MOFI, LIDS, and SHIF), indicate the stability of the atmosphere and cold weather-related conditions.

The third type of environment, leading to a decrease in B[a]P concentrations of 2.3 ng m^{-3}, was associated with medium mean PM_{10}, SO_2, and As levels (34.6 and 14.1 µg m^{-3}, and 1.6 ng m^{-3}, respectively), maximum observed B[a]P, As, and NO_2 concentrations (43.7 and 13.6 ng m^{-3}, and 98.5 µg m^{-3}, respectively), standard lifted index (304), and relative humidity (98%), minimum study period temperature (−15.3 °C), as well

as the highest number of precipitation events, i.e., non-zero TPP6, CPP6, and CRAI values (Figure 7).

The atmospheric stability and the intensity of anthropogenic emissions during the cold part of the year seem to result in high B[a]P concentrations. Since PAHs are mostly particle-bound, and precipitation scavenging plays a significant role in the PM removal from the atmosphere, it could be expected that wet deposition represents a way of PM-bound B[a]P elimination from the atmosphere. As shown by Liu et al. [76], wet removal and photodegradation are up to 10 and 5 times, respectively, more efficient in B[a]P elimination during summer than in winter. Additionally, wet scavenging dominates as a B[a]P removal path in summer, while the impact of photodegradation outweighs the wet removal in winter.

5.2. Arsenic

This study suggests that As concentrations affect B[a]P level dynamics up to 0.9 ng m^{-3} on average (Table 7), more than any other pollutant. A few types of environment were distinguished by analyzing the interrelations between As and B[a]P and their coexistence within certain conditions.

The obtained interrelation indicates similar emission sources of inorganic As, in a mixture of arsenite (AsIII) and arsenate (AsV), and organic B[a]P in the air, that could be identified as high-temperature combustion of fossil fuels and wood [77,78]. In addition, because of low volatility, both As and B[a]P mostly exist as particle-bound in the atmosphere, particularly associated with fine aerosol fractions. Up to approximately 10% of B[a]P occurs in the gaseous phase [79], although the multiphase B[a]P distribution was also highly dependent on ambient temperature [80].

In the first type of environment, B[a]P concentrations exhibited an increase in the range from 4 to 7 ng m^{-3} (Figure 8), with maximum concentrations reaching 30 ng m^{-3}. The relative impact of As, i.e., its association with B[a]P, compared to other studied parameters, reaches a maximum share of 43.6%. This environment was characterized by the lowest As and Cd concentrations, below 2 and 1.5 ng m^{-3}, respectively, and low to medium NOx, SO_2, and PM_{10} levels of below 55 µg m^{-3}, below 25 µg m^{-3}, and from 10 to 25 µg m^{-3}, respectively. Other PM-bound constituents, including Ni and Pb, were registered in higher concentrations of 8.2 and 24.6 ng m^{-3}, respectively, which suggests the impact of local anthropogenic source emissions and dust resuspension, as well as the impact of occasional fossil fuel burning emissions. The co-occurrence of As and B[a]P was observed in the wide range of temperatures at the surface and 2 m (from 1 to 20 °C), which indicates that the relationship between As and B[a]P concentrations was not seasonally dependent. Additionally, this type of environment was featured by PBLH below 150 m, humidity above 74%, wind speed below 2 m s^{-1}, and very low MOFI (Figure 8), all of which reflected extremely stable meteorological and atmospheric conditions, which were registered on a few occasions during the measurement campaign. Therefore, it can be assumed that in the first type of environment, the contributions of remote air pollution sources and atmospheric long-range transport to the observed B[a]P and As concentrations can be excluded.

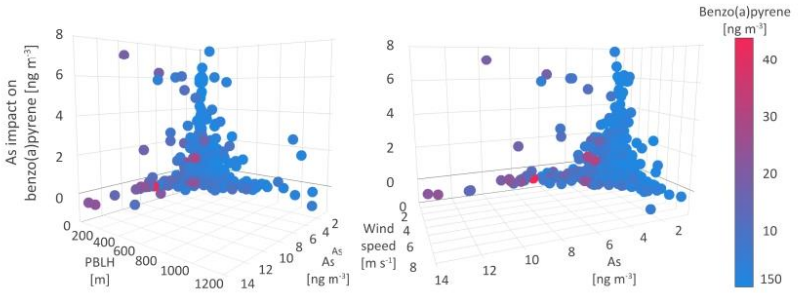


Figure 8. Arsenic impact on benzo(a)pyrene.

The second type of environment was characterized by an increase in B[a]P concentrations up to 4 ng m^{-3} on average and by the lower impact of As (5 to 20%), relative to other pollutants. In comparison to the previous one, this environment was also marked by up to three times higher PM_{10} levels (70 µg m^{-3}), up to two times higher As (5 ng m^{-3}) and NOx (up to 100 µg m^{-3}) levels, and somewhat higher SO_2 (30 µg m^{-3}) concentrations. The assigned meteorological conditions included low humidity, air and soil temperatures ranging from −5 to 20 °C, PBLH below 480 m (Figure 8), and wind speed below 3.7 m s^{-1}, as well as MOFI values typical for the cold season. As can be concluded, the second type of environment represented the cold season and its associated emissions of As and B[a]P as well as inorganic oxides from heating-related sources. In cold weather conditions, PM, NOx, SO_2, and As are slow-reacting and the atmospheric reactions associated with the generation of secondary air pollutants (other oxide forms, sulfates, nitrates, or ozone), reaction byproducts or fine particles require a prolonged time, which in this case contributed to high pollutant concentrations assigned to the second type of B[a]P environment.

The third type of environment referring to the majority of measured pollutant concentrations recognized more than one pattern of As-B[a]P interrelations. Depending on the wind speed and other meteorological factors, both high and low B[a]P and As concentrations were registered. Namely, wind speed below 2 m s^{-1} was associated with the highest pollutant concentrations, while the increase in wind speed above 5 m s^{-1} resulted in a significant decrease in both pollutant concentrations below 1 ng m^{-3}. These findings suggest a negligible contribution of regional pollutant sources to air quality at the sampling site, but also the presence of local pollution sources and processes, such as resuspension of ash from crude-oil and lignite-fired boilers, which strongly affect pollutant concentrations during the episodes of low wind speed.

SHAP values ranging from −0.6 to 0 ng m^{-3} referred to the situations in which As levels had a moderately negative or null impact on B[a]P dynamics. On these occasions, As, B[a]P, and PM_{10} levels were very high, 13.6 ng m^{-3}, 22 ng m^{-3} and 177 µg m^{-3}, respectively, while the SO_2 and NOx levels did not exceed 10 µg m^{-3}. Given these findings were associated with the T02M range from -3 to 5 °C, we can assume that As and B[a]P have separate sources during the cold season, which contribute to high concentrations of either one or another pollutant. More data and further analysis could provide detailed insight and confirm these assumptions.

5.3. Particulate Matter

The PM_{10} concentration is the third significant parameter that affects B[a]P concentrations, as shown by the mean absolute SHAP value of 0.8 ng m^{-3}. In the absence of meteorological conditions favoring the association of B[a]P and small particle fraction, the relationship between PM_{10} and B[a]P stands out.

The highest observed positive associations between PM_{10} levels and B[a]P concentrations, in compliance with a relative share of 57.52% and assigned an absolute SHAP value of 8.36 ng m^{-3}, was registered in the environmental conditions associated with the lowest

concentrations of all pollutants, including PM_{10} levels below 32 μg m^{-3}. As regards meteorological conditions, the strongest interrelation between PM_{10} and B[a]P concentrations was detected in the environment characterized by air and soil temperatures ranging from 0 to 20 °C and low wind speed (below 2 m s^{-1}). This type of environment is not seasonally specific and might indicate natural interactions in the atmosphere such as associations between PAHs and PM. Atmospheric PAHs such as high-ring B[a]P are easily adsorbed onto suspended particles with high organic content [76] while the degradation of particle-bound B[a]P fraction is minimized or inhibited. The gas-to-particle partitioning of pollutants and atmospheric removal by wet scavenging are favored depending on the atmospheric conditions, PM surface, its composition and size, and contaminant properties [81]. In the warm season, the increase in temperatures leads to increased B[a]P volatility, followed by its biodegradation. As the impact of PM_{10} on B[a]P levels weaken the environmental conditions change slightly towards higher pollutant concentrations and an increase in wind speed and PBLH (Figure 9).

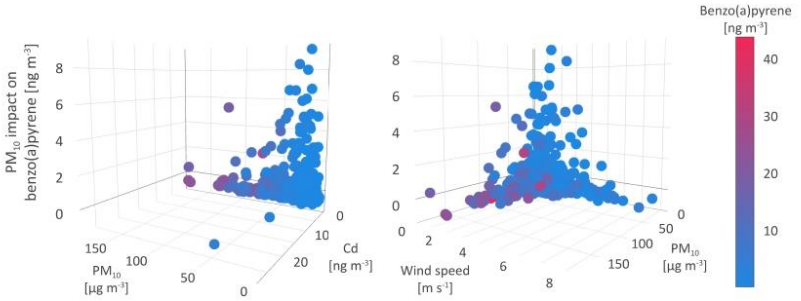


Figure 9. Particulate matter impact on benzo(a)pyrene.

Given the SHAP value of −0.87 PM_{10}, a high number of registered medium to high B[a]P concentrations was negatively associated with PM_{10}, particularly in the environment of high suspended particles As and low Cd, Ni, Pb, NOx, and SO_2 levels. As regards meteorological conditions, these interactions took place during the coldest days of the winter period, when low PBLH, high cloudiness, and wind speed up to 3 m s^{-1} were recorded (Figure 9).

As previously mentioned, the cold season was the period of intense emissions from power plants, domestic heating units, and commercial sources, resulting in elevated levels of PM, especially those of smaller diameter ($PM_{2.5}$ and PM_1) rather than PM_{10}. The finest particle fractions represented a highly suitable matrix for the adsorption of PAHs and these associations could be a possible explanation for the negative relation between B[a]P and PM. A number of studies have shown that small particle diameter plays an important role in the entrapment of PAHs, and thus more than 70% of high-molecular-weight PAHs with higher octanol-water partition coefficients, including B[a]P, is $PM_{2.5}$-bound [82,83]. Low air temperature, wind speed, solar radiation, and PBLH inhibited the vertical diffusion of pollutants and enhanced gas-to-particle pollutant partitioning [84,85]. In addition to this, the strong adsorption capacity of fine PM fraction prevailed over other environmental factors and suggests the particle partition of B[a]P to $PM_{2.5}$ and a smaller fraction rather than to PM_{10}. Lobscheid et al. [86] used multivariate linear regression models to predict relations of ambient B[a]P levels and $PM_{2.5}$ concentrations, spatial, temporal, and meteorological variates. The most significant variables included the average daily $PM_{2.5}$ concentration, wind speed, temperature, and relative humidity.

In contrast to this, during the warm and windy season, when the average temperatures, wind speed, and PBLH exceeded 15 °C, 4 m s^{-1}, and 450 m, respectively, the concentrations of PM_{10} and their constituents exhibited a significant decrease, although the same does not apply for NOx and SO_2. High solar radiation and temperature in warmer seasons lead to the

dispersion and photochemical degradation of the majority air polluting species [80,87], but the persistence of medium to high gaseous oxide levels during the warm season indicated the impact of intense and year-round continuous traffic emissions at the sampling site.

5.4. Nitrogen Oxides

Similar to PAHs, NOx (NO and smaller share of NO_2) emissions mainly resulted from the high-temperature combustion processes in power plants and motor vehicles. Both groups of compounds, PAHs and NOx, were subject to photochemical reactions in the atmosphere. Besides undergoing gas-particle phase distribution, PAHs are precursors for the generation of nitro-compounds. Namely, in the presence of free radicals, OH-PAH or NO_3-PAH are formed and subsequently, in the few-hour reaction with NO_2 upon release of nitric acid or water molecule, nitro-PAHs were generated [88,89].

The mean absolute SHAP value of 0.6 ng m^{-3} defines NOx as the third most significant parameter for shaping B[a]P levels in two distinguished types of environment, one of which strongly supports the increase in B[a]P concentrations. The polluted environment, with moderate to high B[a]P levels (average value of 3 ng m^{-3}) and attributed SHAP value of 6.78 ng m^{-3}, was characterized by a wide range of NOx, PM_{10}, and SO_2 concentrations, from 1.28 to 144 µg m^{-3}, up to 70 µg m^{-3} and up to 30 µg m^{-3}, respectively; however, the lowest levels of PM-bound As, Cd, Ni, and Pb (Figure 10).

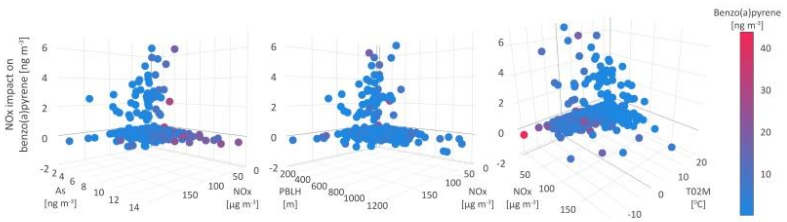


Figure 10. Nitrogen oxides impact on benzo(a)pyrene.

The meteorological conditions which enabled the positive impact of NOx on modelled B[a]P level dynamics and a wide range of B[a]P, NOx, PM_{10}, and SO_2 concentrations, refer to stable high-humidity cold weather without precipitations, with wind speed and PBLH below 3 m s^{-1} and 400 m, respectively; temperatures in the range from −7 to 20 °C, as well as with the corresponding CAPE, CPP6, CRAI, MOFI, and SHIF values (Figure 10). Under these conditions, common emission sources (fossil fuel burning for heating purposes) of the listed pollutants were intensified leading to their higher concentrations. In addition, the stagnant high-humidity conditions during heavy haze events enhanced the transformation of primary emitted particles containing PAHs to secondary organic aerosol (SOA), with the prominent presence of sulfate and nitrate water-soluble species dissolved in an aqueous outer particle layer [90].

The majority of studied pollutant events can be distinguished into two groups depending on the SHAP values and the strength of NOxs negative impact on modeled B[a]P concentration dynamics. The type of environment in which NOx and B[a]P interrelations are expressed by a lower SHAP value of −1.76 ng m^{-3}, refers to the warm season, with air temperatures from 15 to 20 °C, an occasional wind of high speed from 5 to 8 m s^{-1} and mean daily PBLH above 1000 m. As can be expected, these meteorological conditions have favored pollutant dispersion and resulted in low pollutant concentrations, as confirmed by measurements (Figure 10). During the warm season, PAHs undergo photolysis or processes which can yield their derivative compounds, such as oxygenated and nitrated PAHs. The UV-mediated ozone photolysis is a source of OH radicals in the troposphere, which react with PAHs to produce intermediate compounds OH-PAHs. After substitution with NO_2, OH-PAHs are further converted to nitro-PAHs, particularly at night, when the concentrations of NO are low [88]. Additionally, nitro-PAHs are also generated in the

chemical reactions between PAHs and NO_3-radicals, originating from reactions between O_3 and NO, and their formation can explain the negative NOx and B[a]P interrelations.

The SHAP value of -0.3 ng m^{-3} was attributed to the environment where no significant interactions between NOx and B[a]P were registered. Occasionally, these events were characterized either by high concentrations of B[a]P, PM_{10}, PM-bound constituents, and low NOx levels, or the opposite, the lowest concentrations of suspended particles, their constituents and high NOx levels exceeding 50 μg m^{-3}, which implies two different sources of origin.

6. Conclusions

In this study, we employed, coupled, and optimized advanced artificial intelligence-based modeling to accurately interrelate air pollution-related parameters to capture defining factors and processes that shape benzo(a)pyrene behavior. We have applied the XGBoost model optimized by metaheuristics and the Shapley Additive exPlanations explainable artificial intelligence method to a two-year database of pollutant concentrations and meteorological parameters to characterize types of environments that govern the interactions between benzo(a)pyrene, other polluting species, and meteorological conditions. The results suggest that the hybrid self-adaptive sine cosine algorithm method displayed a supreme performance level, by achieving the best scores for key performance indicators (mean square error of 2.5 and R^2 of 0.9), while the firefly algorithm scored the best results for standard deviation and variance, by delivering the most stable results. As shown, the temperature at the surface, arsenic, PM_{10}, and NOx were recognized to affect 22.7%, 14.4%, and 10.0% of benzo(a)pyrene concentrations, respectively. The observed interrelation between particulates and inorganic and organic pollutants could be associated with intensified fossil fuel burning such as low-quality lignite coal during the cold season. In the conditions of low temperature, PM, NOx, SO_2, and As are slow-reacting, and the atmospheric reactions in which the pollutants are involved require a prolonged time, which in this case enhanced the pollutant ambient levels. In addition, during cold seasons, photodegradation of B[a]P was weakened and its adsorption to the particles was favored. The results of this study have proved the potential of the applied methodology to improve the scientific knowledge and understanding of the complex factors that govern the environmental fate of air-polluting species.

Author Contributions: Conceptualization, A.S. and N.B.; methodology, A.S. and N.B.; software, N.B. and M.Z.; validation, G.J., M.P., N.B., I.S. and A.S.; investigation, N.B., M.Z., I.S., M.P., G.J., F.A., S.S. and A.S.; data curation, M.P.; writing—original draft preparation, N.B., M.Z., M.P., I.S., G.J., F.A., S.S. and A.S.; writing—review and editing, A.S. and N.B.; visualization, A.S. and N.B.; supervision, A.S. and N.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia GRANT No. #6524105, AI—ATLAS.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SHAP	SHapley Additive exPlanations	
XGBoost	eXtreme Gradient Boosting	
Label	**Meteorological paramater**	**Unit**
PRSS	Pressure at surface	hPa
MSLP	Pressure reduced to mean sea level	hPa
TPP6	Accumulated precipitation (6 h accumulation)	m
MOFI	Momentum flux intensity (3- or 6-h average)	N m^{-2}
MOFD	Momentum flux direction (3- or 6-h average)	°

SHTF	Sensible heat net flux at surface (3- or 6-h average)	W m^{-2}
DSWF	Downward short wave radiation flux (3- or 6-h average)	W m^{-2}
RH2M	Relative Humidity at 2 m AGL	%
WS	Wind speed at 10 m AGL	m s^{-1}
WD	Wind direction at 10 m AGL	°
TO2M	Temperature at 2 m AGL	°C
TCLD	Total cloud cover (3- or 6-h average)	%
SHGT	Geopotential height	gpm *
CAPE	Convective available potential energy	J Kg^{-1}
CINH	Convective inhibition	J Kg^{-1}
LISD	Standard lifted index	°C
LIB4	Best 4-layer lifted index	°C
PBLH	Planetary boundary layer height	m
TMPS	Temperature at surface	°C
CPP6 **	Accumulated convective precipitation (6 h accumulation)	m
SOLM	Volumetric soil moisture content	frac.
CSNO	Categorial snow (yes = 1, no = 0) (3- or 6-h average)	
CICE	Categorial ice (yes = 1, no = 0) (3- or 6-h average)	
CFZR	Categorial freezing rain (yes = 1, no = 0) (3- or 6-h average)	
CRAI	Categorial rain (yes = 1, no = 0) (3- or 6-h average)	
LHTF	Latent heat net flux at surface (3- or 6-h average)	W/m^2
LCLD	Low cloud cover (3- or 6-h average)	%
MCLD	Middle cloud cover (3- or 6-h average)	%
HCLD	High cloud cover (3- or 6-h average)	%

* geopotential meters

** Beginning with 00 UTC July 15, 2019, CPPA (total accumulation) instead of CPP6 (6-h accumulation)

References

1. Ravina, M.; Esfandabadi, Z.S.; Panepinto, D.; Zanetti, M. Traffic-induced atmospheric pollution during the COVID-19 lockdown: Dispersion modeling based on traffic flow monitoring in Turin, Italy. *J. Clean. Prod.* **2021**, *317*, 128425. [CrossRef]
2. Hoffer, A.; Jancsek-Turóczi, B.; Tóth, Á.; Kiss, G.; Naghiu, A.; Levei, E.A.; Marmureanu, L.; Machon, A.; Gelencsér, A. Emission factors for PM 10 and polycyclic aromatic hydrocarbons (PAHs) from illegal burning of different types of municipal waste in households. *Atmos. Chem. Phys.* **2020**, *20*, 16135–16144. [CrossRef]
3. Mahasakpan, N.; Chaisongkaew, P.; Inerb, M.; Nim, N.; Phairuang, W.; Tekasakul, S.; Furuuchi, M.; Hata, M.; Kaosol, T.; Tekasakul, P.; et al. Fine and ultrafine particle-and gas-polycyclic aromatic hydrocarbons affecting southern Thailand air quality during transboundary haze and potential health effects. *J. Environ. Sci.* **2023**, *124*, 253–267. [CrossRef]
4. Stanišić, S.; Jovanović, G.; Perišić, M.; Snježana, H.R.; Milićević, T.; Stojić, A. Explaining the Environmental Fate of PAHs in Indoor and Outdoor Environments by the Use of Artificial Intelligence. In *Polycyclic Aromatic Hydrocarbons*; Gregoire, W.L., Ed.; Nova Science: Hauppauge, NY, USA, 2022; Chapter 1; pp. 1–36.
5. Karagulian, F.; Belis, C.A.; Dora, C.F.C.; Prüss-Ustün, A.M.; Bonjour, S.; Adair-Rohani, H.; Amann, M. Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmos. Environ.* **2015**, *120*, 475–483. [CrossRef]
6. Elzein, A.; Stewart, G.J.; Swift, S.J.; Nelson, B.S.; Crilley, L.R.; Alam, M.S.; Reyes-Villegas, E.; Gadi, R.; Harrison, R.M.; Hamilton, J.F.; et al. A comparison of PM 2.5-bound polycyclic aromatic hydrocarbons in summer Beijing (China) and Delhi (India). *Atmos. Chem. Phys.* **2020**, *20*, 14303–14319. [CrossRef]
7. Marmett, B.; Carvalho, R.B.; Muccillo-Baisch, A.L.; Baisch, P.R.M.; Dos Santos, M.; Garcia, E.M.; Rhoden, C.R.; da Silva Júnior, F.M.R. Emissions monitoring and carcinogenic risk assessment of PM 10-bounded PAHs in the air from Candiota's coal activity area, Brazil. *Environ. Geochem. Health* **2023**, *45*, 899–911. [CrossRef] [PubMed]
8. Stojić, A.; Vuković, G.; Perišić, M.; Stanišić, S.; Šoštarić, A. Urban air pollution: An insight into its complex aspects. In *A Closer Look at Urban Areas*; Nova Science Publishers: Hauppauge, NY, USA, 2018.
9. Stojić, A.; Mustać, B.; Jovanović, G.; Đinović Stojanović, J.; Perišić, M.; Stanišić, S.; Herceg Romanić, S. Patterns of PCB-138 Bioaccumulation in Small Pelagic Fish from the Eastern Mediterranean Sea Using Explainable Machine Learning Prediction. In *Artificial Intelligence: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 175–189.

10. Stanišić, S.; Perišić, M.; Jovanović, G.; Maletić, D.; Vudragović, D.; Vranić, A.; Stojić, A. What Information on Volatile Organic Compounds Can Be Obtained from the Data of a Single Measurement Site Through the Use of Artificial Intelligence? In *Artificial Intelligence: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 207–225.
11. Stojić, A.; Jovanović, G.; Stanišić, S.; Romanić, S.H.; Šoštarić, A.; Udovičić, V.; Perišić, M.; Milićević, T. The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels. *Chemosphere* **2022**, *289*, 133154. [CrossRef] [PubMed]
12. Šoštarić, A.; Stojić, S.S.; Vuković, G.; Mijić, Z.; Stojić, A.; Gržetić, I. Rainwater capacities for BTEX scavenging from ambient air. *Atmos. Environ.* **2017**, *168*, 46–54. [CrossRef]
13. Perišić, M.; Maletić, D.; Stojić, S.S.; Rajšić, S.; Stojić, A. Forecasting hourly particulate matter concentrations based on the advanced multivariate methods. *Int. J. Environ. Sci. Technol.* **2017**, *14*, 1047–1054. [CrossRef]
14. Stojić, A.; Maletić, D.; Stojić, S.S.; Mijić, Z.; Šoštarić, A. Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. *Sci. Total. Environ.* **2015**, *521*, 19–26. [CrossRef]
15. Stojić, A.; Stanić, N.; Vuković, G.; Stanišić, S.; Perišić, M.; Šoštarić, A.; Lazić, L. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Sci. Total. Environ.* **2019**, *653*, 140–147. [CrossRef]
16. Mirjalili, S. SCA: A sine cosine algorithm for solving optimization problems. *Knowl.-Based Syst.* **2016**, *96*, 120–133. [CrossRef]
17. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
18. Stegherr, H.; Heider, M.; Hähner, J. Classifying Metaheuristics: Towards a unified multi-level classification system. *Nat. Comput.* **2020**, *21*, 155–171. [CrossRef]
19. Emmerich, M.; Shir, O.M.; Wang, H. Evolution strategies. In *Handbook of Heuristics*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 89–119.
20. Fausto, F.; Reyna-Orta, A.; Cuevas, E.; Andrade, Á.G.; Perez-Cisneros, M. From ants to whales: Metaheuristics for all tastes. *Artif. Intell. Rev.* **2020**, *53*, 753–810. [CrossRef]
21. Beni, G. Swarm intelligence. In *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*; Springer: New York, NY, USA, 2020; pp. 791–818.
22. Abraham, A.; Guo, H.; Liu, H. Swarm intelligence: Foundations, perspectives and applications. In *Swarm Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 3–25.
23. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December1995; Volume 4, pp. 1942–1948.
24. Dorigo, M.; Birattari, M.; Stutzle, T. Ant colony optimization. *IEEE Comput. Intell. Mag.* **2006**, *1*, 28–39. [CrossRef]
25. Yang, X.S. Firefly algorithms for multimodal optimization. In Proceedings of the International Symposium on Stochastic Algorithms, Sapporo, Japan, 26–28 October 2009; pp. 169–178.
26. Yang, X.S. A new metaheuristic bat-inspired algorithm. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 65–74.
27. Yang, X.S.; Gandomi, A.H. Bat algorithm: A novel approach for global engineering optimization. *Eng. Comput.* **2012**, *29*, 464–483. [CrossRef]
28. Abualigah, L.; Diabat, A.; Mirjalili, S.; Abd Elaziz, M.; Gandomi, A.H. The arithmetic optimization algorithm. *Comput. Methods Appl. Mech. Eng.* **2021**, *376*, 113609. [CrossRef]
29. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]
30. Zivkovic, M.; Bacanin, N.; Venkatachalam, K.; Nayyar, A.; Djordjevic, A.; Strumberger, I.; Al-Turjman, F. COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustain. Cities Soc.* **2021**, *66*, 102669. [CrossRef] [PubMed]
31. Zivkovic, M.; Venkatachalam, K.; Bacanin, N.; Djordjevic, A.; Antonijevic, M.; Strumberger, I.; Rashid, T.A. Hybrid Genetic Algorithm and Machine Learning Method for COVID-19 Cases Prediction. In *Proceedings of the International Conference on Sustainable Expert Systems: ICSES 2020*; Springer Nature: Singapore, 2021; Volume 176, p. 169.
32. Bacanin, N.; Bezdan, T.; Tuba, E.; Strumberger, I.; Tuba, M.; Zivkovic, M. Task scheduling in cloud computing environment by grey wolf optimizer. In Proceedings of the 2019 27th Telecommunications Forum (TELFOR), Belgrade, Serbia, 26–27 November 2019; pp. 1–4.
33. Bezdan, T.; Zivkovic, M.; Tuba, E.; Strumberger, I.; Bacanin, N.; Tuba, M. Multi-objective Task Scheduling in Cloud Computing Environment by Hybridized Bat Algorithm. In Proceedings of the International Conference on Intelligent and Fuzzy Systems, Istanbul, Turkey, 21–23 July 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 718–725.
34. Bezdan, T.; Zivkovic, M.; Antonijevic, M.; Zivkovic, T.; Bacanin, N. Enhanced Flower Pollination Algorithm for Task Scheduling in Cloud Computing Environment. In *Machine Learning for Predictive Analysis*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 163–171.
35. Zivkovic, M.; Bezdan, T.; Strumberger, I.; Bacanin, N.; Venkatachalam, K. Improved Harris Hawks Optimization Algorithm for Workflow Scheduling Challenge in Cloud–Edge Environment. In *Computer Networks, Big Data and IoT*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 87–102.

36. Zivkovic, M.; Bacanin, N.; Tuba, E.; Strumberger, I.; Bezdan, T.; Tuba, M. Wireless Sensor Networks Life Time Optimization Based on the Improved Firefly Algorithm. In Proceedings of the 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 15–19 June 2020; pp. 1176–1181.
37. Zivkovic, M.; Bacanin, N.; Zivkovic, T.; Strumberger, I.; Tuba, E.; Tuba, M. Enhanced Grey Wolf Algorithm for Energy Efficient Wireless Sensor Networks. In Proceedings of the 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), Novi Sad, Serbia, 26–27 May 2020; pp. 87–92.
38. Bacanin, N.; Tuba, E.; Zivkovic, M.; Strumberger, I.; Tuba, M. Whale Optimization Algorithm with Exploratory Move for Wireless Sensor Networks Localization. In Proceedings of the International Conference on Hybrid Intelligent Systems, Bhopal, India, 10–12 December 2019; pp. 328–338.
39. Zivkovic, M.; Zivkovic, T.; Venkatachalam, K.; Bacanin, N. Enhanced Dragonfly Algorithm Adapted for Wireless Sensor Network Lifetime Optimization. In *Data Intelligence and Cognitive Informatics*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 803–817.
40. Bezdan, T.; Cvetnic, D.; Gajic, L.; Zivkovic, M.; Strumberger, I.; Bacanin, N. Feature Selection by Firefly Algorithm with Improved Initialization Strategy. In Proceedings of the 7th Conference on the Engineering of Computer Based Systems, Novi Sad Serbia, 26–27 May 2021; pp. 1–8.
41. Bezdan, T.; Zivkovic, M.; Tuba, E.; Strumberger, I.; Bacanin, N.; Tuba, M. Glioma Brain Tumor Grade Classification from MRI Using Convolutional Neural Networks Designed by Modified FA. In Proceedings of the International Conference on Intelligent and Fuzzy Systems, Istanbul, Turkey, 21–23 July 2020; pp. 955–963.
42. Zivkovic, M.; Bacanin, N.; Antonijevic, M.; Nikolic, B.; Kvascev, G.; Marjanovic, M.; Savanovic, N. Hybrid CNN and XGBoost Model Tuned by Modified Arithmetic Optimization Algorithm for COVID-19 Early Diagnostics from X-ray Images. *Electronics* **2022**, *11*, 3798. [CrossRef]
43. Strumberger, I.; Tuba, E.; Zivkovic, M.; Bacanin, N.; Beko, M.; Tuba, M. Dynamic search tree growth algorithm for global optimization. In *Proceedings of the Doctoral Conference on Computing, Electrical and Industrial Systems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 143–153.
44. Preuss, M.; Stoean, C.; Stoean, R. Niching Foundations: Basin Identification on Fixed-Property Generated Landscapes. In Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, Dublin, Ireland, 12–16 July 2011; pp. 837–844. [CrossRef]
45. Jovanovic, D.; Antonijevic, M.; Stankovic, M.; Zivkovic, M.; Tanaskovic, M.; Bacanin, N. Tuning Machine Learning Models Using a Group Search Firefly Algorithm for Credit Card Fraud Detection. *Mathematics* **2022**, *10*, 2272. [CrossRef]
46. Petrovic, A.; Bacanin, N.; Zivkovic, M.; Marjanovic, M.; Antonijevic, M.; Strumberger, I. The AdaBoost Approach Tuned by Firefly Metaheuristics for Fraud Detection. In Proceedings of the 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 17–19 June 2022; pp. 834–839.
47. Bacanin, N.; Sarac, M.; Budimirovic, N.; Zivkovic, M.; AlZubi, A.A.; Bashir, A.K. Smart wireless health care system using graph LSTM pollution prediction and dragonfly node localization. *Sustain. Comput. Inform. Syst.* **2022**, *35*, 100711. [CrossRef]
48. Jovanovic, L.; Jovanovic, G.; Perisic, M.; Alimpic, F.; Stanisic, S.; Bacanin, N.; Zivkovic, M.; Stojic, A. The Explainable Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing VOCs' Environmental Fate. *Atmosphere* **2023**, *14*, 109. [CrossRef]
49. Bacanin, N.; Zivkovic, M.; Stoean, C.; Antonijevic, M.; Janicijevic, S.; Sarac, M.; Strumberger, I. Application of Natural Language Processing and Machine Learning Boosted with Swarm Intelligence for Spam Email Filtering. *Mathematics* **2022**, *10*, 4173. [CrossRef]
50. Stankovic, M.; Antonijevic, M.; Bacanin, N.; Zivkovic, M.; Tanaskovic, M.; Jovanovic, D. Feature Selection by Hybrid Artificial Bee Colony Algorithm for Intrusion Detection. In Proceedings of the 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 13–15 October 2022; pp. 500–505.
51. Milosevic, S.; Bezdan, T.; Zivkovic, M.; Bacanin, N.; Strumberger, I.; Tuba, M. Feed-Forward Neural Network Training by Hybrid Bat Algorithm. In Proceedings of the Modelling and Development of Intelligent Systems: 7th International Conference, MDIS 2020, Sibiu, Romania, 22–24 October 2020; Revised Selected Papers 7; Springer International Publishing: New York, NY, USA, 2021; pp. 52–66.
52. Gajic, L.; Cvetnic, D.; Zivkovic, M.; Bezdan, T.; Bacanin, N.; Milosevic, S. Multi-layer Perceptron Training Using Hybridized Bat Algorithm. In *Computational Vision and Bio-Inspired Computing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 689–705.
53. Bacanin, N.; Zivkovic, M.; Al-Turjman, F.; Venkatachalam, K.; Trojovskỳ, P.; Strumberger, I.; Bezdan, T. Hybridized sine cosine algorithm with convolutional neural networks dropout regularization application. *Sci. Rep.* **2022**, *12*, 6302. [CrossRef]
54. Bacanin, N.; Stoean, C.; Zivkovic, M.; Jovanovic, D.; Antonijevic, M.; Mladenovic, D. Multi-Swarm Algorithm for Extreme Learning Machine Optimization. *Sensors* **2022**, *22*, 4204. [CrossRef] [PubMed]
55. Jovanovic, L.; Jovanovic, D.; Bacanin, N.; Jovancai Stakic, A.; Antonijevic, M.; Magd, H.; Thirumalaisamy, R.; Zivkovic, M. Multi-Step Crude Oil Price Prediction Based on LSTM Approach Tuned by Salp Swarm Algorithm with Disputation Operator. *Sustainability* **2022**, *14*, 14616. [CrossRef]
56. Bukumira, M.; Antonijevic, M.; Jovanovic, D.; Zivkovic, M.; Mladenovic, D.; Kunjadic, G. Carrot grading system using computer vision feature parameters and a cascaded graph convolutional neural network. *J. Electron. Imaging* **2022**, *31*, 061815. [CrossRef]

57. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
58. Molnar, C. *Interpretable Machine Learning*; Lulu. com: Morrisville, NC, USA, 2020.
59. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]
60. Derrac, J.; García, S.; Molina, D.; Herrera, F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm Evol. Comput.* **2011**, *1*, 3–18. .: 10.1016/j.swevo.2011.02.002. [CrossRef]
61. de Mattos Neto, P.S.; Marinho, M.H.; Siqueira, H.; de Souza Tadano, Y.; Machado, V.; Antonini Alves, T.; de Oliveira, J.F.L.; Madeiro, F. A methodology to increase the accuracy of particulate matter predictors based on time decomposition. *Sustainability* **2020**, *12*, 7310. [CrossRef]
62. Neto, P.S.D.M.; Firmino, P.R.A.; Siqueira, H.; Tadano, Y.D.S.; Alves, T.A.; De Oliveira, J.F.L.; Marinho, M.H.D.N.; Madeiro, F. Neural-based ensembles for particulate matter forecasting. *IEEE Access* **2021**, *9*, 14470–14490. [CrossRef]
63. Goldberg, D.E.; Richardson, J. Genetic algorithms with sharing for multimodal function optimization. In *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*; Lawrence Erlbaum: Hillsdale, NJ, USA, 1987; Volume 4149.
64. Mirjalili, S. Genetic algorithm. In *Evolutionary Algorithms and Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 43–55.
65. Karaboga, D.; Basturk, B. On the performance of artificial bee colony (ABC) algorithm. *Appl. Soft Comput.* **2008**, *8*, 687–697. [CrossRef]
66. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]
67. Heidari, A.A.; Faris, H.; Aljarah, I.; Mirjalili, S.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]
68. Khishe, M.; Mosavi, M.R. Chimp optimization algorithm. *Expert Syst. Appl.* **2020**, *149*, 113338. [CrossRef]
69. LaTorre, A.; Molina, D.; Osaba, E.; Poyatos, J.; Del Ser, J.; Herrera, F. A prescription of methodological guidelines for comparing bio-inspired optimization algorithms. *Swarm Evol. Comput.* **2021**, *67*, 100973. [CrossRef]
70. Glass, G.V. Testing homogeneity of variances. *Am. Educ. Res. J.* **1966**, *3*, 187–190. [CrossRef]
71. Shapiro, S.S.; Francia, R. An approximate analysis of variance test for normality. *J. Am. Stat. Assoc.* **1972**, *67*, 215–216. [CrossRef]
72. Hsu, H.; Lachenbruch, P.A. Paired t test. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014.
73. Chen, H.; Yang, B.; Wang, S.j.; Wang, G.; Liu, D.y.; Li, H.; Liu, W.b. Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy. *Appl. Math. Comput.* **2014**, *239*, 180–197. [CrossRef]
74. Zhao, Y.; Wang, L.; Luo, J.; Huang, T.; Tao, S.; Liu, J.; Yu, Y.; Huang, Y.; Liu, X.; Ma, J. Deep learning prediction of polycyclic aromatic hydrocarbons in the high arctic. *Environ. Sci. Technol.* **2019**, *53*, 13238–13245. [CrossRef]
75. Cao, X.; Huo, S.; Zhang, H.; Ma, C.; Zheng, J.; Wu, F.; Song, S. Seasonal variability in multimedia transport and fate of benzo [a] pyrene (BaP) affected by climatic factors. *Environ. Pollut.* **2022**, *292*, 118404. [CrossRef]
76. Liu, C.; Guo, Y.; Shi, K.; Zhang, J.; Wu, B.; Du, J. Comparative analysis of contributions of wet deposition and photodegradation to the removal of atmospheric BaP by MFDCCA. *Sci. Rep.* **2021**, *11*, 5515. [CrossRef]
77. Moreno, N.; Viana, M.; Pandolfi, M.; Alastuey, A.; Querol, X.; Chinchon, S.; Pinto, J.F.; Torres, F.; Díez, J.M.; Saez, J. Determination of direct and fugitive PM emissions in a Mediterranean harbour by means of classic and novel tracer methods. *J. Environ. Manag.* **2009**, *91*, 133–141. [CrossRef] [PubMed]
78. Millán-Martínez, M.; Sánchez-Rodas, D.; de la Campa, A.S.; Alastuey, A.; Querol, X.; Jesús, D. Source contribution and origin of PM10 and arsenic in a complex industrial region (Huelva, SW Spain). *Environ. Pollut.* **2021**, *274*, 116268. [CrossRef]
79. Guerreiro, C.; Horálek, J.; de Leeuw, F.; Couvidat, F. Benzo (a) pyrene in Europe: Ambient air concentrations, population exposure and health effects. *Environ. Pollut.* **2016**, *214*, 657–667. [CrossRef] [PubMed]
80. Hunová, I.; Kurfürst, P.; Vlasáková, L.; Schreiberová, M.; Škáchová, H. Atmospheric Deposition of Benzo [a] pyrene: Developing a Spatial Pattern at a National Scale. *Atmosphere* **2022**, *13*, 712. [CrossRef]
81. Hu, T.; Mao, Y.; Ke, Y.; Liu, W.; Cheng, C.; Shi, M.; Zhang, Z.; Zhang, J.; Qi, S.; Xing, X. Spatial and seasonal variations of PAHs in soil, air, and atmospheric bulk deposition along the plain to mountain transect in Hubei province, central China: Air-soil exchange and long-range atmospheric transport. *Environ. Pollut.* **2021**, *291*, 118139. [CrossRef]
82. Jakovljević, I.; Pehnec, G.; Vađić, V.; Čačković, M.; Tomašić, V.; Jelinić, J.D. Polycyclic aromatic hydrocarbons in PM 10, PM 2.5 and PM 1 particle fractions in an urban area. *Air Qual. Atmos. Health* **2018**, *11*, 843–854. [CrossRef]
83. Jia, J.; Deng, L.; Bi, C.; Jin, X.; Zeng, Y.; Chen, Z. Seasonal variations, gas-PM2. 5 partitioning and long-distance input of PM2. 5-bound and gas-phase polycyclic aromatic hydrocarbons in Shanghai, China. *Atmos. Environ.* **2021**, *252*, 118335. [CrossRef]
84. Lv, Y.; Li, X.; Xu, T.T.; Cheng, T.T.; Yang, X.; Chen, J.M.; Iinuma, Y.; Herrmann, H. Size distributions of polycyclic aromatic hydrocarbons in urban atmosphere: Sorption mechanism and source contributions to respiratory deposition. *Atmos. Chem. Phys.* **2016**, *16*, 2971–2983. [CrossRef]

85. Nguyen, T.N.T.; Jung, K.S.; Son, J.M.; Kwon, H.O.; Choi, S.D. Seasonal variation, phase distribution, and source identification of atmospheric polycyclic aromatic hydrocarbons at a semi-rural site in Ulsan, South Korea. *Environ. Pollut.* **2018**, *236*, 529–539. [CrossRef] [PubMed]
86. Lobscheid, A.B.; McKone, T.E.; Vallero, D.A. Exploring relationships between outdoor air particulate-associated polycyclic aromatic hydrocarbon and PM2. 5: A case study of benzo (a) pyrene in California metropolitan regions. *Atmos. Environ.* **2007**, *41*, 5659–5672. [CrossRef]
87. Siudek, P. Polycyclic aromatic hydrocarbons in coarse particles (PM10) over the coastal urban region in Poland: Distribution, source analysis and human health risk implications. *Chemosphere* **2023**, *311*, 137130. [CrossRef]
88. Lee, Y.Y.; Hsieh, Y.H.; Huang, B.W.; Mutuku, J.K.; Chang-Chien, G.P.; Huang, S. An Overview: PAH and Nitro-PAH Emission from the Stationary Sources and their Transformations in the Atmosphere. *Aerosol Air Qual. Res.* **2022**, *22*, 220164. [CrossRef]
89. Lammel, G.; Mulder, M.D.; Shahpoury, P.; Kukučka, P.; Lišková, H.; Přibylová, P.; Prokeš, R.; Wotawa, G. Nitro-polycyclic aromatic hydrocarbons–gas–particle partitioning, mass size distribution, and formation along transport in marine and continental background air. *Atmos. Chem. Phys.* **2017**, *17*, 6257–6270. [CrossRef]
90. Wang, J.; Ye, J.; Zhang, Q.; Zhao, J.; Wu, Y.; Li, J.; Liu, D.; Li, W.; Zhang, Y.; Wu, C.; et al. Aqueous production of secondary organic aerosol from fossil-fuel emissions in winter Beijing haze. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2022179118. [CrossRef]

Chemosphere 289 (2022) 133154

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo[*a*]pyrene levels

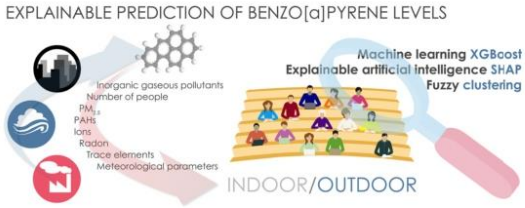
Andreja Stojić [a,b], Gordana Jovanović [a,b], Svetlana Stanišić [b,*], Snježana Herceg Romanić [c], Andrej Šoštarić [d], Vladimir Udovičić [a], Mirjana Perišić [a,b], Tijana Milićević [a]

[a] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 118 Pregrevica Street, 11000, Belgrade, Serbia
[b] Singidunum University, 32 Danijelova Street, 11000, Belgrade, Serbia
[c] Institute for Medical Research and Occupational Health, 2 Ksaverska Cesta Street, PO Box 291, 10001, Zagreb, Croatia
[d] Institute of Public Health Belgrade, 54 Despota Stefana Street, 11000, Belgrade, Serbia

HIGHLIGHTS

- Relative errors of the applied machine learning methods were below 15.1%.
- Explainable methodology characterized conditions which govern B[*a*]P fate.
- Key predictors of B[*a*]P dynamics were high-ring PAHs, Chy, CO, As, Cr, and $PM_{2.5}$.
- Out of 31 meteorological parameters, only one significantly affected outdoor B[*a*]P.
- 4 and 8 environmental condition types shape B[*a*]P behavior indoors and outdoors.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Volker Matthias

Keywords:
Indoor air pollution
Outdoor air pollution
benzo[*a*]pyrene
Machine learning
Explainable artificial intelligence

ABSTRACT

Among the polycyclic aromatic hydrocarbons (PAH), benzo[*a*]pyrene (B[*a*]P) has been considered more relevant than other species when estimating the potential exposure-related health effects and has been recognized as a marker of carcinogenic potency of air pollutant mixture. The current understanding of the factors which govern non-linear behavior of B[*a*]P and associated pollutants and environmental processes is insufficient and further research has to rely on the advanced analytical approach which averts the assumptions and avoids simplifications required by linear modeling methods. For the purpose of this study, we employed eXtreme Gradient Boosting (XGBoost), SHapley Additive exPlanations (SHAP) attribution method, and SHAP value fuzzy clustering to investigate the concentrations of inorganic gaseous pollutants, radon, $PM_{2.5}$ and particle constituents including trace metals, ions, 16 US EPA priority $PM_{2.5}$-bound PAHs and 31 meteorological variables, as key factors which shape indoor and outdoor $PM_{2.5}$-bound B[*a*]P distribution in a university building located in the urban area of Belgrade (Serbia). According to the results, the indoor and outdoor B[*a*]P levels were shown to be highly correlated and mostly influenced by the concentrations of Chry, B[*b*]F, CO, B[*a*]A, I[*cd*]P, B[k]F, Flt, D[ah]A, Pyr, B[ghi]P, Cr, As, and $PM_{2.5}$ in both indoor and outdoor environments. Besides, high B[*a*]P concentration events were recorded during the periods of low ambient temperature (<12 °C), unstable weather conditions with precipitation and increased soil humidity.

* Corresponding author.
E-mail address: sstanisic@singidunum.ac.rs (S. Stanišić).

https://doi.org/10.1016/j.chemosphere.2021.133154
Received 19 May 2021; Received in revised form 24 November 2021; Accepted 2 December 2021
Available online 3 December 2021

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a complex mixture of congeners originating from pyrogenic and petrogenic, as well as anthropogenic and natural sources (Velázquez-Gómez and Lacorte, 2020). Among PAHs, benzo[*a*]pyrene (B[*a*]P) has been recognized as a marker of carcinogenic potency of the air pollutant mixture (Liu et al., 2020). According to the IARC (2012), it has been assigned to a group 1 of hazardous species – mutagenic and carcinogenic to humans irrespective of the environment, and its emissions are regulated by the Directive, 2004/107/EC. While two-three ring low molecular weight PAHs mostly occur in the gas phase, the compounds with four aromatic rings and more, including B[*a*]P, are semi-volatile and 70–90% of their emitted content is adsorbed on particulate matter, overall on the fine inhalable particles with aerodynamic diameter less than 2.5 μm – $PM_{2.5}$ (Liu et al., 2014; Azari et al., 2020). Previous studies have been focused on levels, spatial and seasonal distribution, sources, local and regional source contributions, personal exposure, B[*a*]P equivalent toxicity and cancer risks of $PM_{2.5}$-bound PAHs (Liu et al., 2017; Han et al., 2019; Yan et al., 2019; Zhang et al., 2019a, 2019b, 2019c; Du et al., 2020; Lao et al., 2020; Ali-Taleshi et al., 2020; Gope et al., 2020). Some of the studies have also reported contrasting findings about carcinogenic potential of PAH mixture depending on the B[*a*]P content (Brehmer et al., 2020). A study aimed at characterizing indoor air quality in kindergartens located in urban and rural area of Poland has shown that no statistically significant differences exist in the concentrations of total PAHs in indoor versus outdoor air, although the mutagenic effect of outdoor $PM_{2.5}$ samples was twice as high as the effect of indoor samples (Błaszczyk et al., 2017). The review of Ma and Harrad (2015) has shown that even though the concentrations of both PAHs and B[*a*]P were higher in the indoor environment, indoor sources emitted proportionally less carcinogenic species than outdoor sources, which was implied by the comparison of I/O ratios for ΣPAH and B[*a*]P toxicity equivalents. Nevertheless, some studies proved the opposite (Oliveira et al., 2016; Sangiorgi et al., 2013). The previous studies also showed that, depending on a sampling location and environmental factors, B[*a*]P concentrations can be found within the relatively wide range of values (Rönkkö et al., 2020).

B[*a*]P atmospheric transformations and persistence are strongly affected by meteorological conditions, including temperature, precipitation, moisture, and solar radiation, as well as the presence of particles and oxidant species, e.g., ozone, nitrate, and hydroxyl radicals (Liu et al., 2017). In addition to this, the particles' chemical content and structure have significant impact on the B[*a*]P chemodynamics. Considering PAHs' heterogenous reactions with oxidizing agents such NO_2 and O_3, B[*a*]P has been reported to be among the most reactive PAH congener when bound to soot, silica, diesel, or graphite particles together with pyrene (Pyr), anthracene (Ant), benz[*a*]anthracene (B[*a*]A), and dibenz[a,l]pyrene (DB[al]P) (Keyte et al., 2013 and the references therein). Conversely, on the soot and ammonium sulfate particles a formation of monolayer coverage makes B[*a*]P less exposed to surface reactions and thus, more persistent.

The study of Lodovici et al. (2003) showed that the level of total particle-bound B[*a*]P was as low as 0.02 ng m^{-3} at the regional background site locations, while the study of Hassanvand et al. (2015) registered B[*a*]P concentrations of 5 or more ng m^{-3}, depending on the season, traffic impact, type of sampling location, and particle fraction. As regards $PM_{2.5}$-related organic content, the study of Jedynska et al. (2014), aimed at investigating the levels of PAHs in fine particle fraction at street, urban and regional background, showed that mean B[*a*]P levels were below 0.2 ng m^{-3} for all ten investigated sites in Europe. On the other hand, the studies performed in Asia showed that the total $PM_{2.5}$-bound B[*a*]P concentrations considerably exceeded the maximum permissible risk level of 1 ng m^{-3} (Yury et al., 2018).

In this study, we present promising advanced machine learning (ML) and explainable artificial intelligence (XAI) methodologies (eXtreme Gradient Boosting – XGBoost and SHapley Additive exPlanations – SHAP) for studying complex, heterogeneous, and non-linear interactions between indoor and outdoor B[*a*]P levels and $PM_{2.5}$, PAHs, inorganic gaseous pollutants, trace elements, ions, radon, 31 meteorological parameters, the number of people in the amphitheater, and the time they spent indoor that could not be addressed by traditional approaches. The methods have become increasingly recognized and successfully applied when predicting environmental phenomena (Blair et al., 2019; Gibert et al., 2018; Stojić et al., 2019; Stanišić et al., 2021; Ye et al., 2020). The study aims to provide an insight into the B[*a*]P behavior by attributing environmental factor importance (SHAP values), impacts (SHAP dependency), mutual relations (relative SHAP values), and interactions (SHAP interactions). Moreover, we aim to identify and characterize governing environmental conditions responsible for shaping the levels of B[*a*]P concentrations in both environments (SHAP force).

2. Materials and methods

2.1. Measurement campaign

The measurements of inorganic gaseous pollutants, radon, $PM_{2.5}$ and particle constituents including trace metals, ions, and PAHs were performed from March 1st – May 31st in a building of Singidunum University (44°45′33.8″N, 20°29′47.6″E), situated in the urban area of Belgrade, Serbia. In the residential area surrounding the measurement site, there is a large number of households with individual fireboxes using coal and wood, while approx. 1 km in the W/SW direction and W/NW direction, there are two heating plants operating with the total production capacity of 230 MW and 50 MW, respectively, mainly fueled with natural gas and crude oil. During the three-month study campaign, the outdoor pollutants were sampled at the rooftop of the University building, at the open space 10 m above ground. For the indoor sampling, sampling inlets and $PM_{2.5}$ sampling device were placed at a height of 6 m and 2 m off the floor in an amphitheater with a capacity of 350 seats where lectures for often 50 to 80 students were given. During the study campaign, the number of people and the time they spent in the amphitheater was registered hourly.

$PM_{2.5}$ was collected daily on quartz filters (Whatman QMA, 47 mm) by Svan Leckel LVS6-RV devices with a flow rate of 2.3 m^3 h^{-1}, over 24 h sampling period. Inorganic gases (O_3, CO, SO_2, and NO_x) were measured by Horiba devices APOA, APMA, APSA, and APNA, 370 series, for the continuous monitoring of pollutants with 2-min resolution using ultraviolet absorption, infrared spectroscopy, ultraviolet fluorescence, and chemiluminescence methods, respectively. The measurements were performed according to the following European Standards EN 14211:2012, EN 14212:2012, EN 14625:2012 and EN 14626:2012. The limit of detection (LOD) for O_3, SO_2, and NO_x was 1 μg m^{-3} while for CO it was 0.1 mg m^{-3}.

The outdoor meteorological data were obtained by using Vaisala WXT530 monitoring station, while the indoor radon concentrations, ambient air temperature, relative humidity and air pressure were detected by SN1029 radon monitor (Sun Nuclear Corporation, NRSB approval-code 31822). The LOD for radon was 0.1 Bq m^{-3}. More details on the study area, sampling campaign and chemical analyses are described in the Part 1 of this paper (Stanišić et al., 2021).

2.2. Chemical analyses of $PM_{2.5}$ constituents, quality assurance and quality control

In brief, gravimetric measurements of $PM_{2.5}$ were conducted according to the European Standard EN 12341:2014. Prior to gravimetric determination, the pre-fired and preconditioned non-exposed filters were measured representing control blanks. After preconditioning for 48 h in a Class 100 clean room with automatic temperature and pressure regulation, the filters were weighed twice using a micro-balance (Precisa XR 125 SB). Mass concentrations of $PM_{2.5}$ were calculated as average

values. Loaded filters were stored in a cool room at 4 °C prior to chemical analysis. After gravimetric measurements, the surface of each filter amounting to 13.85 cm^2 was cut in two pieces - approximately 1.76 cm^2 each, which were used for the analysis of anions and cations (Cl^-, Ca^{2+}, K^+, NO_3^-, SO_4^{2-}, and NH_4^+), while the remaining 12.09 cm^2 were divided and used for the analysis of trace elements (As, Cd, Cr, Mn, Ni, and Pb) and 16 US EPA PAHs.

The inorganic PM constituents were determined by the standard methods for elements (European Standards (EN) 14902:2005). The extraction of the trace elements was performed by a mixture of HNO_3 (30%):H_2O_2:H_2O (3:2:5) using analytical grade reagents (Merck) and distilled/deionized water (MiliQ, 18.2 MΩ) (CEN/TC 264 N779). After microwave accelerated digestion (Anton Paar 3000), the concentrations of trace elements were determined by inductively coupled plasma–mass spectrometry (ICP-MS) (Agilent 7500ce with Octopole Reaction System). Quality control was conducted by 2783 NIST standard reference material (National Institute of Standard and Technology, MD, USA). The recovery values were within satisfactory range of ±20% in relation to the reference value while method LOD was: 0.4 ng m^{-3} for As, 0.05 ng m^{-3} for Cd, and 2 ng m^{-3} for Cr, Mn, Ni, and Pb.

For the determination of ion concentrations, the filter pieces were extracted by ultra-pure water for 24 h. The aqueous extracts were further analyzed by standard ion chromatography (Dionex DX500 IC system, MDL 064 Standard operating procedure). The LOD was: 2 μg m^{-3} for Cl^- and NO_3^-, 1 μg m^{-3} for SO_4^{2-}, 0.2 μg m^{-3} for NH_4^+, 2 μg m^{-3} for K^+ and 8 μg m^{-3} for Ca^{2+}.

The concentrations of priority PAHs including naphthalene (Nap), acenaphthylene (Acy), acenaphthene (Ace), fluorene (Flu), phenanthrene (Phe), anthracene (Ant), fluoranthene (Flt), pyrene (Pyr), benz[*a*]anthracene (B[*a*]A), chrysene (Chry), benzo[*b*]fluoranthene (B[*b*]F), benzo[k]fluoranthene (B[k]F), benzo[*a*]pyrene (B[*a*]P), dibenz[a,h] anthracene (DB[ah]A), benzo[g,h,i]perylene (B[ghi]P), and indeno [1,2,3-*cd*]pyrene (I[*cd*]P) were determined following the Standard ISO 12884:2010. Further details have previously been illustrated in the studies preceding this one (Stanišić et al., 2021; Cvetković et al., 2015). The filters were microwave-extracted by a solvent mixture of n-hexane and acetone, 12.5 mL:12.5 mL (US EPA, 2007). Solution was rotary evaporated to 1 mL under reduced pressure (55.6 kPa and with 0.2 mL isooctane) and to 0.25 mL under a nitrogen stream. The PAHs were analyzed using gas chromatography coupled with mass selective detector (Agilent GC 6890/5973 MSD) with a DB-5 MS capillary column (30 m × 0.25 mm × 25 μm) according to EPA Compendium Method TO-13 A. The oven temperature was attained by applying the following steps: (1) isothermal heating for 4 min at 70 °C, (2) heating from 70 °C to 310 °C at 8 °C min^{-1}, and (3) 5 min of isothermal heating at 310 °C. Solvent delay was 5 min and the time of run was 46 min. Helium was used as the carrier gas. The injector was set to 300 °C. Prior to the analysis, calibration curves ($R^2 > 0.995$) were obtained using Ultra Scientific PAH Mixture PM-831, which contains 16 priority PAHs. The concentration of calibration solutions was between 5 and 200 ng mL^{-1}.

We used Ultra Scientific PAH Mixture PM-831, which consists of 16 compounds, each of 500.8 ± 2.5 μg/mL concentration as external standard for calibration curve. We determined concentrations of 16 priority USEPA PAHs: Nap, Acy, Ace, Flu, Phe, Ant, Flt, Pyr, B[*a*]A, Chry, B[*b*]F, B[k]F, B[*a*]P, I[*cd*]P, DB[ah]A, and B[ghi]P.

To estimate method recovery, Ultra Scientific Semi-Volatiles Internal Standard Mixture ISM-560 containing: Ace-d_{10}. Chry-d_{10}, 1,4-dichlorobenzene, Nap-d_8, Perylen-d_{12}, and Phe-d_{10} was used as internal standard. Recovery values ranged from 85% to 110% for all the PAHs in the internal standard. The LOD was calculated as three times signal/noise and it was 0.01 ng m^{-3} for all PAH species. The limit of quantification was determined as 3.3 times of LOD. Field and laboratory blank were also prepared and analyzed, and all data were corrected with reference to the blanks.

2.3. Data analysis

2.3.1. Machine learning

The relationships between indoor and outdoor levels of B[*a*]P (74 samples) and other investigated parameters (other PAHs, inorganic gaseous pollutants, radon, $PM_{2.5}$ and particle constituents including trace metals and ions, meteorological parameters including measured and GDAS1-modeled, the number of people and the time they spent indoor, trend, weekday and weekend – 39 and 64 parameters in total for indoor and outdoor environment, respectively) were explored by the regression analysis, implemented by eXtreme Gradient Boosting. Briefly, XGBoost is a highly effective ensemble method of supervised machine learning based on a sequential tree-growing algorithm. Iteratively reweighing the training data to improve regression performance, each decision tree aims to complement all the others and correct for residuals in the predictions made by the previous trees. XGBoost is based on a gradient descent algorithm, used to minimize loss when adding new models. The method includes system optimization and algorithmic enhancements through parallelized sequential tree building, tree pruning, regularization, weighted quantile sketch algorithm implementation, cross-validation, etc. Outperforming standard deep neural network models on tabular-style datasets, XGBoost was successfully applied across various domains especially due to its core advantages referring to computational efficiency and competitive accuracy, even when data is sparse and unstructured (Hartmann, 2019; Lundberg et al., 2020). In this study, we used Python (Python Software Foundation) XGBoost implementation (XGBoost Python Package). The data were split into training (80%) and validation (20%) sets. The criterion for splitting the data set into training and test set was that both data subsets should follow the same probability distribution. In this study, we have identified the PAH outliers according to the tradeoff between the split criteria and the necessity to maximize the total size of the data. The same indoor and outdoor events were used for training/testing. Hyperparameter tuning was implemented by using a brute-force grid search and 10-fold stratified cross-validation which was replicated 10 times. The best performing hyperparameter values were used for the final model.

2.3.2. Explainable artificial intelligence

The explainability of ML model behavior which operates with high-dimensional input data in a non-linear and nested fashion is crucial for understanding the process being modeled. Until recently, the inability to explain the predictions from accurate, but complex models, posed a serious limitation in understanding the governing factors that shape a prediction. For this purpose, we employed the advanced explainable artificial intelligence method, which is capable to avoid the trade-off between accuracy and interpretability and provide the straightforward and meaningful interpretation of the ML model-derived decisions, now being shifted towards user-readable logic rules to match human intuition.

2.3.2.1. Shapley additive exPlanations. SHapley Additive exPlanations (SHAP) is a method based on Shapley values, calculated as a measure of feature importance using a game-theory approach, that provide an impact of features on individual predictions (Lundberg and Lee, 2017). The Shapley value method provides fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). It perfectly apportions the difference between the prediction and the average prediction among the features (Molnar, 2019). Thus, SHAP assigns each feature the importance as a measure of its contribution to a particular prediction and interpret the impact of having a certain value for a given feature in comparison to the prediction of a model if that feature took some baseline value. The SHAP explanations represent the only possible locally accurate and globally consistent feature contribution values (Chen et al., 2019; Stojić et al., 2019). The method provides valuable insights into a model's behavior

by overcoming the main drawback of inconsistency and minimizes the possibility of underestimating the importance of a feature with a certain attribution value, capturing feature interaction effects based on generalization of Shapley values, and interpreting the model's global behavior while retaining local faithfulness (Lundberg et al., 2020).

In this study we used Python SHAP implementation (SHAP Python package) and the TreeExplainer which reduces the complexity of exact Shapley value computation from exponential to low-order polynomial time by leveraging the internal structure of tree-based models (Lundberg et al., 2020). The captured attributed importance of a feature, the change of a feature importance over its value range, as well as its interaction effects with other features are visually presented as SHAP summary plots, SHAP dependency plots, and SHAP interaction plots, respectively.

A change in the absolute SHAP value of a feature does not clearly indicate its relationships with other features. To gain an insight into relative relationships among feature attributions for each individual prediction, we introduced the relative SHAP values. They show the relative influence of a feature to the prediction and are defined as a share of absolute SHAP in total attributed importance of all features for the particular case.

The stabilities of the obtained absolute and relative SHAP values were evaluated by 50 times-replicated bootstrap method. The stabilities are presented in figures as error bars.

2.3.2.2. Fuzzy clustering. The fuzzy clustering of absolute SHAP attributions was performed to identify and characterize indoor and outdoor ambient conditions responsible for B[*a*]P behavior. It was chosen because each B[*a*]P concentration will not necessarily belong to a single class of environmental conditions which shapes it. Fuzzy clustering was performed by using R (R: A language and environment for statistical computing) 'cluster' package (Maechler et al., 2019). The obtained results were presented as force plots. A detailed analysis of each cluster was performed based on the statistical character of its absolute and relative SHAP values, as well as the measured parameter values.

Beside conventional images, we present all relevant findings as interactive plots by using R package 'plotly' (Sievert, 2020) hosted at the web page designed to support this paper at www.envpl.ipb.ac.rs/papers/20/PAHs/.

3. Results and discussion

The mean $PM_{2.5}$ and B[*a*]P levels in the indoor and outdoor environments were 16.2 vs. 17.5 μg m^{-3} and 0.50 vs. 0.48 ng m^{-3}, respectively, while both indoor and outdoor B[*a*]P mean concentrations were below the recommended level of 1 ng m^{-3} (Directive, 2004/107/EC).

As it can be seen in Fig. 1, extreme concentration events (ranging from 2.0 to 6.8 ng m^{-3}) were mostly registered over a few days in March and May, when mean daily temperature did not exceed 12 °C. As a higher molecular weight PAH, B[*a*]P is almost completely particle-bonded which makes it less reactive vs. solar radiation and free radicals, and its affinity towards particle phase increases with ambient air temperature lowering and decrease of vapor pressure. Beside the intense fossil fuel combustion for heating purposes, the described gas-particle phase distribution additionally contributes to higher B[*a*]P concentrations in cold season and significant seasonal differences in mean pollutant levels. For instance, the study of Jedrychowski et al. (2007) showed that winter B[*a*]P concentrations were 4.3 ng m^{-3} and 6.1 ng m^{-3}, while summer levels were 0.8 ng m^{-3} and 0.9 ng m^{-3} for indoor and outdoor environment, respectively.

Similarly to the study of Jedrychowski et al. (2007), our results showed that indoor/outdoor (I/O) B[*a*]P ratio was mainly below 1, with a few exceptions when the calculated values were in the range from 2 to 10 (Fig. 1). Conversely, some studies (Romagnoli et al., 2014) reported the outdoor B[*a*]P concentrations to be significantly lower throughout the year than the corresponding indoor levels. For instance, the study aimed at characterizing levels of PAHs at preschool environment in Portugal and assessing the exposure-related health risk, showed that carcinogenic risk due to indoor PAH-related exposure was 4–18 times higher than for outdoors (Oliveira et al., 2016). In compliance with this, the research focused on the impact of outdoor environment on indoor air quality in office buildings in Milano (Italy) throughout the year confirmed a strict correlation between indoor and outdoor PM concentrations. However, the reported B[*a*]P concentrations were higher indoor (I/O = 2.3) suggesting that indoor sources did not contribute to higher PM mass emissions but significantly affected human health through the apportionment of the particles enriched by carcinogenic species (Sangiorgi et al., 2013).

According to our results, the extreme B[*a*]P I/O ratio values (Fig. 1) were detected in the days when the indoor B[*a*]P remained concentrated while the frequent changes in the weather conditions, followed by significant wind gusts, precipitation events and sunshine hours considerably affected the outdoor air quality (Part 1 of this paper, Stanišić et al., 2021). The differences between reported B[*a*]P I/O ratio values can be explained by the fact that variable meteorological conditions govern more rapid outdoor concentration variations, while indoor air quality remains less affected. This further suggests that the long-term I/O ratio calculations excluding the extremely low or high pollutant values could better reflect an environment in which B[*a*]P pollution occurs.

In this study, XGBoost was successfully employed for exploring complex, heterogeneous, and non-linear relationships between B[*a*]P concentrations and key factors which shape their indoor and outdoor distribution including inorganic gaseous pollutants, radon, $PM_{2.5}$ and their constituents including trace metals, ions, all other US EPA priority $PM_{2.5}$-bound PAHs, 31 meteorological variables, the number of people and the time they spent indoor, trend, weekday, and weekend. The predicted/observed calculated relative errors were 15.1% and 14.5%, while the r^2 were 0.96 and 0.95 for indoor and outdoor, respectively (Fig. 2).

The data analysis revealed a correlation of 0.67 ($p<0.05$) between indoor and outdoor B[*a*]P levels (Figure S1). Further, both the indoor and outdoor B[*a*]P concentrations exhibited correlations above 0.9 with

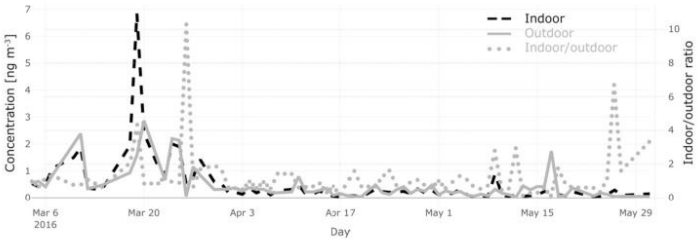


Fig. 1. Benzo(*a*)pyrene concentrations and indoor/outdoor ratio.

B[*a*]A, Chry, I[*cd*]P, B[ghi]P, B[*b*]F, and B[k]F levels from the corresponding environment, while also significant correlations of 0.89 and 0.82, respectively, were found with indoor and outdoor Pyr. The persistent influence of the listed pollutants on B[*a*]P behavior was also confirmed by the correlation of SHAP values higher than 0.8 (Figure S2a). However, relative SHAP value correlations, slightly above 0.8, reveal a more variable impact of B[*a*]A, B[*b*]F, B[k]F, and Chry (Figure S2b), and even more for indoor and outdoor Pyr (<0.7). It is known that Chy and Pyr can be distributed between gas and particle phase upon emission, which distinguishes them slightly from B[*a*]P, B[*a*]A, B[k]F, I[*cd*]F, B[ghi]P, and B[*b*]F, which exhibit high mutual correlations because they are found to be mostly PM-bonded (Oliveira et al., 2016).

Yet, according to absolute and relative SHAP values (Fig. 3), Chy appears to be the most important B[*a*]P predictor in the indoor environment (absolute SHAP: 0.11 ng m^{-3}; relative SHAP: 18.23%), followed by B[*b*]F, CO, B[*a*]A, I[*cd*]P, B[k]F, Flt, D[ah]A, Pyr, B[ghi]P, Cr, and $PM_{2.5}$, while in the outdoor environment B[*a*]P levels could be more accurately predicted by B[*b*]F concentrations, although the significance of B[*a*]A, B[k]F, B[ghi]P, I[*cd*]P, Pyr, CO, As, and Flt, as well as the significance of Chry (absolute SHAP: 0.01 ng m^{-3}; relative SHAP: 3.61%) were also evidenced. It can be assumed that in the indoor environment, semivolatile Chry is more particle-bonded and has similar fate to B[*a*]P, while outdoor Chry has been shown to be less important for B[*a*]P level prediction probably since it is, unlike B[*a*]P, more distributed in the gas phase, as well as more resistant to atmospheric reactions with oxidative species (Estève et al., 2004; Perraudin et al., 2007). On the other hand, Pyr has a comparable influence on B[*a*]P levels in both indoor and outdoor environment, which is supported by the fact that these species have very similar molecular structure and environmental behavior, and thus, they exhibit equal reactivity with hydroxyl and nitrate radical species. Also, the increased levels of Chry, Pyr, and B[*a*]P suggest that these species share the same source in both indoor and outdoor environment, which can probably be attributed to fuel-burning for heating purposes as identified by Unmix source apportionment and PAH diagnostic ratios (Part 1 of this paper, Stanišić et al., 2021), or more specifically to coal combustion, when considering the association between outdoor As and B[*a*]P levels. In the indoor environment, the evidenced relationship between B[*a*]P and Cr levels indicates the contribution of diesel and gasoline emissions, which are the major source of PAHs in the warm season. Namely, the study area was located 80 m from the main road, and thus, the impact of motor vehicle emission on indoor air quality could be registered. Unsurprisingly, SHAP values suggested no associations between indoor and outdoor B[*a*]P levels and highly volatile PAHs including Nap, Acy, and Ace, which are normally distributed in a gas phase.

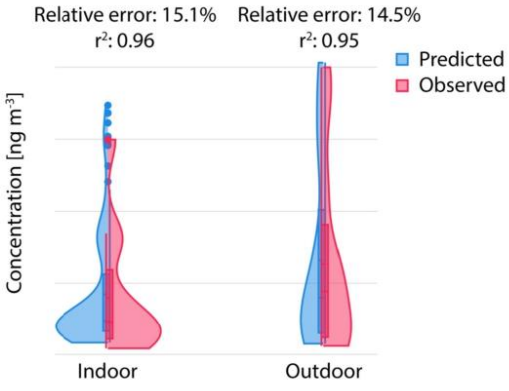


Fig. 2. XGBoost model evaluation.

The results of SHAP analysis revealed that the impact of $PM_{2.5}$ on B[*a*]P levels is less evident in the outdoor than indoor environment which can be explained by the inconsistent matrix-specific interactions of PAHs and particles at the molecular level. Namely, B[*a*]P is mostly entirely found within the fine particle fraction, but the chemical nature and the amount of bonding varies with particle composition and environmental factors (Lammel et al., 2010). To particles with higher organic content, B[*a*]P is most often bonded by solvation, and this process is enhanced in the presence of moisture and usually less temperature dependent compared to the weaker bond of adsorption type that occurs on particles with higher inorganic content. It appears that adsorption was the dominant mechanism involved in pollutant particle distribution in the outdoor environment, which made B[*a*]P more prone to transformations and oxidation with free OH radicals and less resistant to UV decomposition that also takes part in the outdoor environment.

The low SHAP values (Fig. 3) and dependence plots (Figures S3) showed that neither indoor nor outdoor B[*a*]P behavior exhibit significant weekend dynamic pattern or meteorologically-driven trend. While previous studies (Jung et al., 2014) mostly reported that B[*a*]P atmospheric persistence and levels were affected by the seasonal variations of temperature, relative humidity, and pressure, our results (the SHAP values from 0.2 to 0.4, Fig. 3, and SHAP dependencies, Figures S3) showed that only the increase in soil moisture (Solm>0.3) was positively associated with the increased outdoor B[*a*]P levels (1–3 ng m^{-3}). Beside temperatures below 12 °C, changeable weather conditions with precipitations, and consequently increased soil moisture, characterized the episodes of high B[*a*]P levels in the beginning and the end of March along with mid-April and May (Fig. 1). As already mentioned, B[*a*]P and other high-weight PAHs are hydrophobic and almost entirely found within the fine particle fraction, which implies that their atmospheric removal is regularly affected by dry deposition (Keyte et al., 2013). After being deposited in the soil, B[*a*]P decay on particles via heterogenous reactions is reduced with the increase of soil moisture, which leads to higher B[*a*]P levels in the soil and thus, larger pollutant pool for volatilization. Although the volatility of B[*a*]P is generally low (vapor pressure, $p = 7.9\ 10^{-6}$ Pa at 298 K) and its tendency to volatilize upon being bond to solid surfaces is limited, volatilization from the particles' surfaces still occurs and appears to contribute up to 9% to total B[*a*]P emissions as shown by the study conducted in the European region (Keyte et al., 2013), which can explain the positive association between soil moisture and outdoor B[*a*]P levels. In addition, soil moisture is recognized as one of the most important factors for controlling particulate/dust resuspension because it enhances the strength of inter-particle bonds by promoting the development of a humid film between soil grains that makes soil an important secondary emission source of particles depending on the texture and mechanical composition (Nieder et al., 2018). We note that the SHAP analysis indicated that the indoor levels of B[*a*]P were independent of the number of attendants and employees, as well as the time they spend indoors.

As shown by the dependence plot (Figures S3), the non-proportional relationships are evident: the elevated outdoor and indoor B[*a*]P concentrations are followed by increased levels of the main predicting species (B[*a*]A, B[*b*]F, I[*cd*]P, Chy, B[ghi]P, B[k]F, and Pyr). According to the relative SHAP values (Fig. 3), B[*a*]A (21%), B[*b*]F (21%), and B[k]F (15%) isomers are recognized as the main compounds which explain the outdoor B[*a*]P dynamics, i.e., contribute to the environment which is associated with B[*a*]P specific behavior and fate. The relationships between B[*a*]P and other PAH species are less evident, as shown by relative SHAP values below 10%. The non-proportional relationships are also confirmed between B[*a*]P and other considered parameters in the following order: CO, As, Cr, $PM_{2.5}$, and Rn (Fig. 3). Considering the relative SHAP values, their significance appears to be more evident in the indoor than outdoor environment. The main compounds which explain the indoor B[*a*]P dynamics were Chry (18%), B[k]F (17%), and CO (10%).

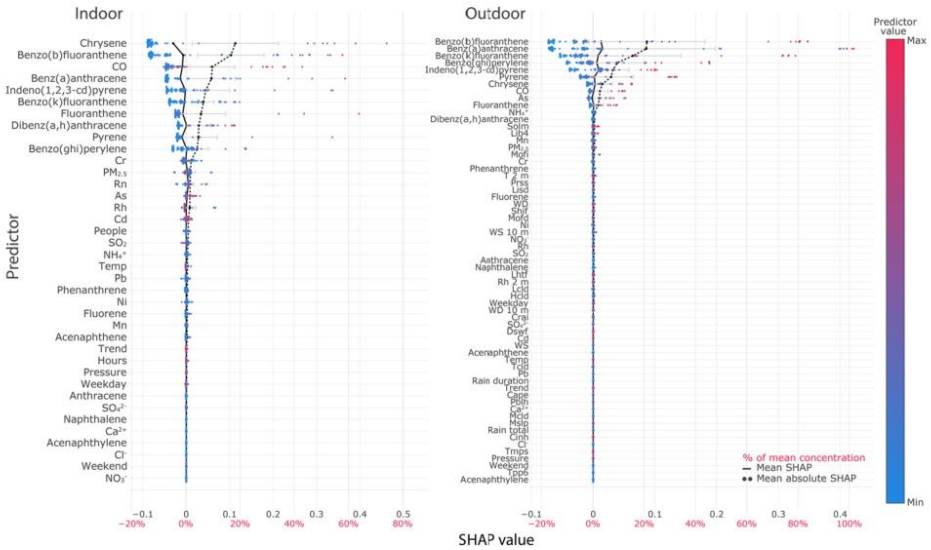


Fig. 3. Indoor (left) and outdoor (right) benzo(*a*)pyrene SHAP summary plots.

The impact of B[*b*]F on B[*a*]P levels is very similar in both environments and increases with B[*b*]F concentrations (Figure S3.3). Its share to the other governing factors decreases above B[*b*]F concentration of 2 ng m^{-3} which means that other factors overtake the leading role in shaping B[*a*]P levels above 0.8 ng m^{-3}. The influence of B[k]F on B[*a*]P environmental fate is more pronounced for low levels indoors and high levels outdoors, but we have not found the explanation for this observation.

The impacts of Chry and B[*a*]A on B[*a*]P levels are very similar in shape (Figures S3.6 and S3.2), with relative SHAP 18% and 21% in the indoor and outdoor environments, respectively. This implies that, although semi-volatile, Chry is more prone to be particle-bound in the

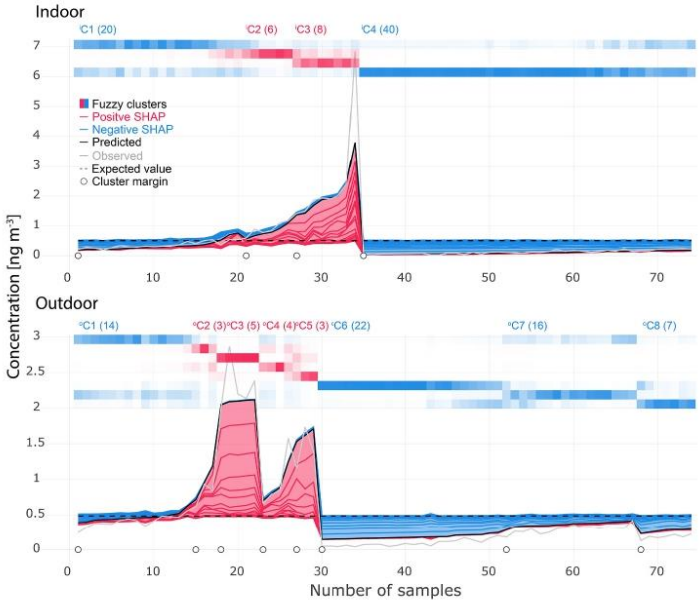


Fig. 4. Indoor (above) and outdoor (below) benzo(*a*)pyrene SHAP force plots.

conditions of limited photo oxidative reactions. This is confirmed by SHAP dependency analysis which reveals two types of environmental conditions interrelating these compounds characterized by Chry levels below/above 0.9 ng m^{-3} corresponding to the occurrence of this compound predominantly in a gaseous/particulate phase.

^{i}C1-^{i}C4 refers to four clusters being identified for indoor environment by the SHAP value fuzzy clustering, while ^{o}C1-^{o}C8 refers to eight clusters that were identified for outdoor environment. Most of the investigated days were assigned to ^{o}C6 (Figs. 4 and 5). The predictors used for cluster differentiation in this study (13 for indoor and 9 for outdoor environment) explain the approximately 90% of B[*a*]P level dynamics in total. The number of clusters indicate the complexity of the ambient, diversity of emission sources, and abundance of environmental influences. As can been seen in all clusters, mostly particle-bound 4-, 5- and 6-ring PAHs (B[*a*]A, B[*b*]F, B[k]F, D[ah]A, and I[*cd*]P) characterized the environment in which both low and high concentrations of B[*a*]P were registered, but non-negligible is the impact of 4-ring semi-volatile PAH congeners Chry and Pyr, which show ambivalent chemodynamics depending on the molecular structure, and CO (absolute value of relative SHAP ranging from 6.3 to 9.3%). The pollutants D[ah]A, Cr, As, Rn, and $PM_{2.5}$, for which low SHAP values (<0.1) were observed, had minor potential for explaining the indoor and/or outdoor environment that shaped B[*a*]P level dynamics.

During the 60 days in total, being attributed to ^{i}C4 (3rd April - 31st May) and ^{i}C1 (4th March - 26th May), the investigated parameters defined the indoor environment in which low concentrations of PAHs and particularly B[*a*]P were registered (mean B[*a*]P concentrations 0.07 and 0.36 ng m^{-3} for clusters ^{i}C4 and ^{i}C1, respectively). The range of the lowest concentrations (^{i}C4: 0.03–0.24 ng m^{-3}) is well-separated from the others (fuzzy cluster membership 88%), indicating that the specific environmental conditions governing their occurrence were associated with the dominant influence of Chy > B[*b*]F > CO > B[*a*]A > I[*cd*]P > B[k]F (relative SHAP ranging from −7.4 to −18%). Indoor low levels clustered in ^{i}C4 were observed during the warmer part of the measurement campaign and thus can be predominantly attributed to the reduction of the intensity of the outdoor emission sources related to heating and enhanced photodegradation of PAHs. The range of slightly higher concentrations (^{i}C1: 0.2–0.54 ng m^{-3}) is not well-differentiated (fuzzy cluster membership 69%), which can be probably attributed to the influences that do not originate from the features used in this study. The decrease of B[*a*]P concentrations was mostly affected by Chry > B[*a*]A > B[*b*]F (in average 18.9, 10.3, and 8.5%, respectively).

Conversely, during the period attributed to ^{i}C2 (7th March - 9th May) and ^{i}C3 (10th March - 28th March), the investigated parameters dominantly defined the indoor environment in which high concentrations of B[*a*]P, B[*a*]A, B[*b*]F, B[k]F, Chry, D[ah]A, and I[*cd*]P (0.5–3 ng m^{-3}), as well as CO were registered (>0.35 mg m^{-3}), whereas the concentrations of Cr were noticeably lower (<6.5 ng m^{-3}) than during the events assigned to ^{i}C4 and ^{i}C1. The ^{i}C2 and ^{i}C3 periods were associated with changeable weather, low temperature, occasional precipitations, and pronounced cold front breakthroughs that led to the intensified PAH emissions from heating sources and reduced B[*a*]P ambient decomposition by photolytic reactions. The observed B[*a*]P mean concentrations and ranges were ^{i}C2: 0.90; 0.6–1.2 ng m^{-3} and ^{i}C3: 2.48; 1.4–2.4 ng m^{-3}. Both clusters were characterized by the dominant impact of B[*b*]F (relative SHAP 18.9 and 25.3%) and differentiated by the impact of Chry

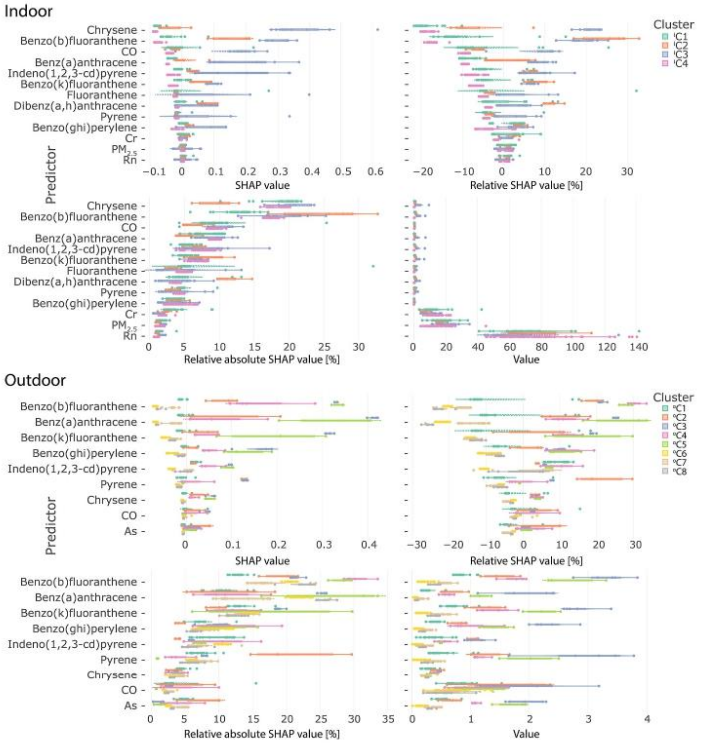


Fig. 5. Indoor (above) and outdoor (below) benzo(*a*)pyrene SHAP force plot cluster statistics.

and CO which is negative for ^{i}C2 negative, and positive for ^{i}C3. We did not observe the cause of differentiation within the features used in this study.

During the days attributed to clusters ^{o}C2, ^{o}C3, ^{o}C4, and ^{o}C5, the elevated concentrations of predictors including I[*cd*]P, Chry, B[ghi]P, and B[k]F (>1 ng m^{-3}), B[*a*]A, B[*b*]F, and particularly Pyr (>0.5 ng m^{-3}), As (>1.2 ng m^{-3}), and CO (>0.35 mg m^{-3}) shaped the outdoor environment (Figs. 4 and 5) in which high B[*a*]P concentrations were registered (0.8, 2.3, 1.0, and 1.4 ng m^{-3}, respectively). Over the 14 days assigned to ^{o}C1 (4th March - 20th May), the increase of B[*a*]P levels was mostly affected by I[*cd*]P concentrations, although its impact was not dominating. Over the days assigned to ^{o}C6 (22 days in total in the period 26th March - 31st May), ^{o}C7 (16 days in the period 30th March - 24th May) and ^{o}C8 (7 days in the period 8th April - 3rd May), the concentrations of investigated parameters including As (<0.7 ng m^{-3}), B[*a*]A, B[*b*]F, B[ghi]P, Pyr, B[k]F, Chry, and I[*cd*]P (<1 ng m^{-3}) defined the environment in which low B[*a*]P levels were registered (0.1, 0.3, and 0.2 ng m^{-3}, respectively). Additionally, the attributions of all predictors to the observed B[*a*]P concentrations were low, as implied by negative SHAP values and high errors calculated for ^{o}C6 (absolute error 0.09 and relative error 97.8%) and ^{o}C8 (absolute error 0.09 and relative error 46.6%). This further indicates that other environmental factors (e.g., UV PAH degradation or photochemical formation of PAH derivates initiated by the presence of peroxides, O_3, and nitrate and hydroxyl radicals) played an important role in B[*a*]P environmental fate over the corresponding days.

Based on the absolute SHAP interaction values (Figure S4), the interactions between the following pairs of pollutants: $PM_{2.5}$–Chry, CO–DB[ah]A, CO–Rn, CO–B[*b*]F, Chry–B[k]F and As–Chry appeared to be the most prominent features that shape the indoor environment while B[*a*]A–B[ghi]P and B[*a*]A–B[*b*]F were extracted as the most significant interactions in the outdoor ambience. Additionally, the SHAP relative values point to CO–D[ah]A indoor as the most influential interaction, and the potential explanation has been already discussed in the previous text.

4. Conclusions

The indoor air quality has attracted growing attention since the research has shown that it does not represent a simple reflection of the outdoor pollutant concentrations. Additionally, the findings that some pollutants, including carcinogenic B[*a*]P, can be more concentrated indoors emphasizes the significance of the internal air quality for human health and well-being. In this study, the machine learning and explainable artificial intelligence methods were successfully employed (relative errors ≤15.1%) for exploring the sources of indoor and outdoor B[*a*]P in a university building, and examine its relationships with other air pollutants and meteorological factors. According to the results, Chry and B[*b*]F concentrations were found to be the main factors which explain the environment, associated with B[*a*]P specific behavior and fate, followed by other long-lived particle-bound PAHs, including B[*a*]A, I[*cd*]P, B[k]F, Flt, D[ah]A, Pyr, and B[ghi]P. Less important associations were recorded between B[*a*]P concentrations and the levels of inorganic contaminants (CO, As, and Cr), $PM_{2.5}$, as well as soil moisture, whereas the impacts of other investigated parameters appeared to be negligible. As can be concluded, the ongoing developments and advances in machine learning and artificial intelligence in general, have resulted in complex modeling which have the potential to enhance our understanding of air pollution and related environmental processes.

Author contribution

Andreja Stojić; , Tijana Miličević: Conceptualization, Methodology, Software. Mirjana Perišić; : Data curation. Andrej Šoštarić; , Vladimir Udovičić; : Visualization, Investigation. Gordana Jovanović; : Supervision. Svetlana Stanišić; , Snježana Herceg Romanić; : Writing – original draft, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS, as well as "Analysis of organic pollutants in biological systems and the environment", institutional financing of scientific activity, Croatia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2021.133154.

References

Ali-Taleshi, M.S., Moeinaddini, M., Bakhtiari, A.R., Feiznia, S., Squizzato, S., Bourliva, A., 2020. A one-year monitoring of spatiotemporal variations of $PM_{2.5}$-bound PAHs in Tehran, Iran: source apportionment, local and regional sources origins and source-specific cancer risk assessment. Environ. Pollut. 274, 115883. https://doi.org/10.1016/j.envpol.2020.115883.

Azari, M.R., Mohammadian, Y., Pourahmad, J., Khodagholi, F., Mehrabi, Y., 2020. Additive toxicity of Co-exposure to pristine multi-walled carbon nanotubes and benzo α pyrene in lung cells. Environ. Res. 183, 109219. https://doi.org/10.1016/j.envres.2020.109219.

Blair, G.S., Henrys, P., Leeson, A., Watkins, J., Eastoe, E., Jarvis, S., Young, P.J., 2019. Young data science of the natural environment: a research roadmap. Front. Environ. Sci. 121, 1–14. https://doi.org/10.3389/fenvs.2019.00121.

Błaszczyk, E., Rogula-Kozlowska, W., Klejnowski, K., Kubiesa, P., Fulara, I., Mielżyńska-Švach, D., 2017. Indoor air quality in urban and rural kindergartens: short-term studies in Silesia, Poland. Air Qual. Atmos. Health 10, 1207–1220. https://doi.org/10.1007/s11869-017-0505-9.

Brehmer, C., Norris, C., Barkjohn, K.K., Bergin, M.H., Zhang, J., Cui, X., Teng, Y., Zhang, Y., Black, M., Li, Z., Shafer, M.M., 2020. The impact of household air cleaners on the oxidative potential of $PM_{2.5}$ and the role of metals and sources associated with indoor and outdoor exposure. Environ. Res. 181, 108919. https://doi.org/10.1016/j.envres.2019.108919.

Chen, H., Lundberg, S., Lee, S.I., 2019. Explaining Models by Propagating Shapley Values of Local Components arXiv preprint arXiv:1911.11888.

Cvetković, A., Jovašević-Stojanović, M., Marković, D., Ristovski, Z., 2015. Concentration and source identification of polycyclic aromatic hydrocarbons in the metropolitan area of Belgrade, Serbia. Atmos. Environ. 112, 335–345. https://doi.org/10.1016/j.atmosenv.2015.04.034.

Directive, 2004. 107/EC of the Europen parliament and of the council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. Off. J. European Union 23 (1–16), 26/01/2005.

Du, W., Yun, X., Chen, Y., Zhong, Q., Wang, W., Wang, L., Qi, M., Shen, G., Tao, S., 2020. PAHs emissions from residential biomass burning in real-world cooking stoves in rural China. Environ. Pollut. 267, 115592. https://doi.org/10.1016/j.envpol.2020.115592.

Estève, W., Budzinski, H., Villenave, E., 2004. Relative rate constants for the heterogeneous reactions of OH, NO_2 and NO radicals with polycyclic aromatic hydrocarbons adsorbed on carbonaceous particles. Part 1: PAHs adsorbed on 1–2 μm calibrated graphite particles. Atmos. Environ. 38, 6063–6072. https://doi.org/10.1016/j.atmosenv.2004.05.059.

European Standards (EN) 12341, 2014. 2014. Ambient Air. Standard Gravimetric Measurement Method for the Determination of the PM_{10} or $PM_{2.5}$ Mass Concentration of Suspended Particulate Matter. https://cds.cern.ch/record/2624772. (Accessed 27 January 2021).

European Standards (EN) 14211, 2012a. 2012. Ambient Air. Standard Method for the Measurement of the Concentration of Nitrogen Dioxide and Nitrogen Monoxide by Chemiluminescence. https://www.en-standard.eu/bs-en-14211-2012-ambient-air.-standard-method-for-the-measurement-of-the-concentration-of-nitrogen-dioxide-and-nitrogen-monoxide-by-chemiluminescence/. (Accessed 27 January 2021).

European Standards (EN) 14212, 2012b. 2012. Ambient Air. Standard Method for the Measurement of the Concentration of Sulphur Dioxide by Ultraviolet Fluorescence. https://www.sis.se/en/produkter/environment-health-protection-safety/air-quality/ambient-atmospheres/ssen14212012/. (Accessed 27 January 2021).

European Standards (EN) 14625, 2012c. 2012. Ambient Air. Standard Method for the Measurement of the Concentration of Ozone by Ultraviolet Photometry. https://shop.bsigroup.com/ProductDetail?pid=000000000030210754. (Accessed 27 January 2021).
European Standards (EN) 14626, 2012d. 2012. Ambient Air. Standard Method for the Measurement of the Concentration of Carbon Monoxide by Non-dispersive Infrared Spectroscopy. https://www.en-standard.eu/bs-en-14626-2012-ambient-air.-standard-method-for-the-measurement-of-the-concentration-of-carbon-monoxide-by-non-dispersive-infrared-spectroscopy/. (Accessed 27 January 2021).
European Standards (EN) 14902, 2005. 2005. Ambient Air Quality. Standard Method for the Measurement of Pb, Cd, AS, and Ni in the PM_{10} Fraction of Suspended Particulate Matter. https://www.antpedia.com/standard/pdf/13.040.20/1703/EN%2014902-2005_6234.pdf. (Accessed 27 January 2021).
Gibert, K., Horsburgh, S.J., Athanasiadis, N.I., Holmes, G., 2018. Environmental data science. Environ. Model. Software 106, 4–12. https://doi.org/10.1016/j.envsoft.2018.04.005.
Gope, M., Masto, R.E., Basu, A., Bhattacharyya, D., Saha, R., Hoque, R.R., Khillare, P.S., Balachandran, S., 2020. Elucidating the distribution and sources of street dust bound PAHs in Durgapur, India: a probabilistic health risk assessment study by Monte-Carlo simulation. Environ. Pollut. 267, 115669. https://doi.org/10.1016/j.envpol.2020.115669.
Han, J., Liang, Y., Zhao, B., Wang, Y., Xing, F., Qin, L., 2019. Polycyclic aromatic hydrocarbon (PAHs) geographical distribution in China and their source, risk assessment analysis. Environ. Pollut. 251, 312–327. https://doi.org/10.1016/j.envpol.2019.05.022.
Hartmann, J., 2019. Classification Using Decision Tree Ensembles. https://doi.org/10.2139/ssrn.3484009, 27 January 2021.
Hassanvand, M.S., Naddafi, K., Faridi, S., Nabizadeh, R., Sowlat, M.H., Momeniha, F., Gholampour, A., Arhami, M., Kashani, H., Zare, A., Niazi, S., 2015. Characterization of PAHs and metals in indoor/outdoor $PM_{10}/PM_{2.5}/PM_1$ in a retirement home and a school dormitory. Sci. Total Environ. 527, 100–110. https://doi.org/10.1016/j.scitotenv.2015.05.001.
International Agency for Research on Cancer, 2012. A Review of Human Carcinogens. Part F: Chemical Agents and Related Occupations. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(09)70358-4/fulltext. (Accessed 27 January 2021).
International Standardization Organization (ISO) 12884, 2000. 2020. Ambient Air — Determination of Total (Gas and Particle-phase) Polycyclic Aromatic Hydrocarbons — Collection on Sorbent-Backed Filters with Gas Chromatographic/mass Spectrometric Analyses. https://www.iso.org/standard/1343.html. (Accessed 27 January 2021).
Jedrychowski, W., Pac, A., Choi, H., Jacek, R., Sochacka-Tatara, E., Dumyahn, T.S., Spengler, J.D., Camann, D.E., Perera, F.P., 2007. Personal exposure to fine particles and benzo [a] pyrene. Relation with indoor and outdoor concentrations of these pollutants in Krakow. IJOMEH 20, 339–348. https://doi.org/10.2478/v10001-007-0035-z.
Jedynska, A., Hoek, G., Eeftens, M., Cyrys, J., Keuken, M., Ampe, C., Beelen, R., Cesaroni, G., Forastiere, F., Cirach, M., De Hoogh, K., 2014. Spatial variations of PAH, hopanes/steranes and EC/OC concentrations within and between European study areas. Atmos. Environ. 87, 239–248. https://doi.org/10.1016/j.atmosenv.2014.01.026.
Jung, K.H., Liu, B., Lovinsky-Desir, S., Yan, B., Camann, D., Sjodin, A., Li, Z., Perera, F., Kinney, P., Chillrud, S., Miller, R.L., 2014. Time trends of polycyclic aromatic hydrocarbon exposure in New York City from 2001 to 2012: assessed by repeat air and urine samples. Environ. Res. 131, 95–103. https://doi.org/10.1016/j.envres.2014.02.017.
Keyte, I.J., Harrison, R.M., Lammel, G., 2013. Chemical reactivity and long-range transport potential of polycyclic aromatic hydrocarbons–a review. Chem. Soc. Rev. 42, 9333–9391. https://doi.org/10.1039/C3CS60147A.
Lammel, G., Klánová, J., Ilić, P., Kohoutek, J., Gasić, B., Kovačić, I., Skrdlíková, L., 2010. Polycyclic aromatic hydrocarbons in air on small spatial and temporal scales e II. Mass size distributions and gas-particle partitioning. Atmos. Environ. 44, 5022–5027. https://doi.org/10.1016/j.atmosenv.2010.07.034.
Lao, J.Y., Wang, S.Q., Chen, Y.Q., Bao, L.J., Lam, P.K., Zeng, E.Y., 2020. Dermal exposure to particle-bound polycyclic aromatic hydrocarbons from barbecue fume as impacted by physicochemical conditions. Environ. Pollut. 260, 114080. https://doi.org/10.1016/j.envpol.2020.114080.
Liu, B., Xue, Z., Zhu, X., Jia, C., 2017. Long-term trends (1990–2014), health risks, and sources of atmospheric polycyclic aromatic hydrocarbons (PAHs) in the US. Environ. Pollut. 220, 1171–1179. https://doi.org/10.1016/j.envpol.2016.11.018.
Liu, D., Xu, Y., Chaemfa, C., Tian, C., Li, J., Luo, C., Zhang, G., 2014. Concentrations, seasonal variations, and outflow of atmospheric polycyclic aromatic hydrocarbons (PAHs) at Ningbo site, Eastern China. Atmos. Pollut. Res. 5, 203–209. https://doi.org/10.5094/APR.2014.025.
Liu, W., Wang, D., Wang, Y., Zeng, X., Ni, L., Tao, Y., Wu, J., Liu, J., Zou, Y., He, R., Zhang, J., 2020. Improved comprehensive ecological risk assessment method and sensitivity analysis of polycyclic aromatic hydrocarbons (PAHs). Environ. Res., 109500 https://doi.org/10.1016/j.envres.2020.109500.
Lodovici, M., Venturini, M., Marini, E., Grechi, D., Dolara, P., 2003. Polycyclic aromatic hydrocarbons air levels in Florence, Italy, and their correlation with other air pollutants. Chemosphere 50, 377–382. https://doi.org/10.1016/S0045-6535(02)00404-6.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 2522–5839. https://doi.org/10.1038/s42256-019-0138-9.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, vol. 30. NIPS 2017, pp. 4765–4774. In: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html. (Accessed 27 January 2021).
Ma, Y., Harrad, S., 2015. Spatiotemporal analysis and human exposure assessment on polycyclic aromatic hydrocarbons in indoor air, settled house dust, and diet: a review. Environ. Int. 84, 7–16. https://doi.org/10.1016/j.envint.2015.07.006.
Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2019. Cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0. https://cran.r-project.org/web/packages/cluster/cluster.pdf. (Accessed 27 January 2021).
Molnar, C., 2019. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. URL https://christophm. github. io/interpretable-ml-book. (Accessed 27 January 2021).
Nieder, R., Benbi, D.K., Reichl, F.X., 2018. Soil-borne particles and their impact on environment and human health. In: Nieder, R., Benbi, D.K., Reichl, F.-X. (Eds.), Soil Components and Human Health. Springer, Dordrecht, pp. 99–177. https://doi.org/10.1007/978-94-024-1222-2.
Oliveira, M., Slezakova, K., Delerue-Matos, C., Do Carmo Pereira, M., Morais, S., 2016. Assessment of polycyclic aromatic hydrocarbons in indoor and outdoor air of preschool environments (3–5 years old children). Environ. Pollut. 208, 382–394. https://doi.org/10.1016/j.envpol.2015.10.004.
Perraudin, E., Budzinski, H., Villenave, E., 2007. Kinetic study of the reactions of ozone with polycyclic aromatic hydrocarbons adsorbed on atmospheric model particles. J. Atmos. Chem. 56, 57–82. https://doi.org/10.1007/s10874-006-9042-x.
Romagnoli, P., Balducci, C., Perilli, M., Gherardi, M., Gordiani, A., Gariazzo, C., Gatto, M.P., Cecinato, A., 2014. Indoor PAHs at schools, homes and offices in Rome, Italy. Atmos. Environ. 92, 51–59. https://doi.org/10.1016/j.atmosenv.2014.03.063.
Rönkkö, T.J., Hirvonen, M.R., Happo, M.S., Leskinen, A., Koponen, H., Mikkonen, S., Bauer, S., Ihantola, T., Hakkarainen, H., Miettinen, M., Orasche, J., 2020. Air quality intervention during the Nanjing youth olympic games altered PM sources, chemical composition, and toxicological responses. Environ. Res., 109360 https://doi.org/10.1016/j.envres.2020.109360.
Sangiorgi, G., Ferrero, L., Ferrini, B.S., Porto, C.L., Perrone, M.G., Zangrando, R., Gambaro, A., Lazzati, Z., Bolzacchini, E., 2013. Indoor airborne particle sources and semi-volatile partitioning effect of outdoor fine PM in offices. Atmos. Environ. 65, 205–214. https://doi.org/10.1016/j.atmosenv.2012.10.050.
Sievert, C., 2020. Interactive Web-Based Data Visualization with R, Plotly, and Shiny. CRC Press.
Stanišić, S., Perišić, M., Jovanović, G., Milićević, T., Romanić, S.H., Jovanović, A., Šoštarić, A., Udovičić, V., Stojić, A., 2021. The $PM_{2.5}$-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: emission sources. Environ. Res. 193, 110520. https://doi.org/10.1016/j.envres.2020.110520.
Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Total Environ. 653, 140–147. https://doi.org/10.1016/j.scitotenv.2018.10.368.
United States Environmental Protection Agency (US EPA), 2007. Test Method 3546: Microwave Extraction, pp. 1–13. https://www.epa.gov/sites/production/files/2015-12/documents/3546.pdf. (Accessed 27 January 2021).
Velázquez-Gómez, M., Lacorte, S., 2020. Organic pollutants in indoor dust from Ecuadorian Amazonia areas affected by oil extractivism. Environ. Res., 109499 https://doi.org/10.1016/j.envres.2020.109499.
Yan, D., Wu, S., Zhou, S., Tong, G., Li, F., Wang, Y., Li, B., 2019. Characteristics, sources and health risk assessment of airborne particulate PAHs in Chinese cities: a review. Environ. Pollut. 248, 804–814. https://doi.org/10.1016/j.envpol.2019.02.068.
Ye, Z., Yang, J., Zhong, N., Tu, X., Jia, J., Wang, J., 2020. Tackling environmental challenges in pollution controls using artificial intelligence: a review. Sci. Total Environ. 699, 134279 https://doi.org/10.1016/j.scitotenv.2019.134279.
Yury, B., Zhang, Z., Ding, Y., Zheng, Z., Wu, B., Gao, P., Jia, J., Lin, N., Feng, Y., 2018. Distribution, inhalation and health risk of $PM_{2.5}$ related PAHs in indoor environments. Ecotoxicol. Environ. Saf. 164, 409–415. https://doi.org/10.1016/j.ecoenv.2018.08.044.
Zhang, J., Liu, W., Xu, Y., Cai, C., Liu, Y., Tao, S., Liu, W., 2019b. Distribution characteristics of and personal exposure with polycyclic aromatic hydrocarbons and particulate matter in indoor and outdoor air of rural households in Northern China. Environ. Pollut. 255, 113176 https://doi.org/10.1016/j.envpol.2019.113176.
Zhang, J., Yang, L., Ledoux, F., Courcot, D., Mellouki, A., Gao, Y., Jiang, P., Li, Y., Wang, W., 2019a. $PM_{2.5}$-bound polycyclic aromatic hydrocarbons (PAHs) and nitrated PAHs (NPAHs) in rural and suburban areas in Shandong and Henan Provinces during the 2016 Chinese New Year's holiday. Environ. Pollut. 250, 782–791. https://doi.org/10.1016/j.envpol.2019.04.040.
Zhang, Y., Zheng, H., Zhang, L., Zhang, Z., Xing, X., Qi, S., 2019c. Fine particle-bound polycyclic aromatic hydrocarbons (PAHs) at an urban site of Wuhan, central China: characteristics, potential sources and cancer risks apportionment. Environ. Pollut. 246, 319–327. https://doi.org/10.1016/j.envpol.2018.11.111.

atmosphere

MDPI

Article

The $PM_{2.5}$-Bound Polycyclic Aromatic Hydrocarbon Behavior in Indoor and Outdoor Environments, Part III: Role of Environmental Settings in Elevating Indoor Concentrations of Benzo(a)pyrene

Gordana Jovanović [1,2], Mirjana Perišić [1,2], Timea Bezdan [2], Svetlana Stanišić [2], Kristina Radusin [3], Aleksandar Popović [3] and Andreja Stojić [1,2,*]

[1] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia; gordana.jovanovic@ipb.ac.rs (G.J.); mirjana.perisic@ipb.ac.rs (M.P.)
[2] Software and Information Engineering, Singidunum University, Danijelova 32, 11000 Belgrade, Serbia; tbezdan@singidunum.ac.rs (T.B.); sstanisic@singidunum.ac.rs (S.S.)
[3] Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia; kristina.radusin@zdravlje.org.rs (K.R.); apopovic@chem.bg.ac.rs (A.P.)
* Correspondence: andreja.stojic@ipb.ac.rs

Abstract: This study aims to investigate the impact of indoor sources and outdoor air on indoor $PM_{2.5}$-bound benzo(a)pyrene, with a focus on identifying emission sources and understanding the influence of environmental variables. For this purpose, we collected indoor and outdoor data on $PM_{2.5}$-bound PAHs, inorganic gaseous pollutants, trace metals, ions, radon, and meteorological parameters, resulting in a comprehensive dataset of 100 variables from an urban site in Belgrade, Serbia. We applied seven regression tree ensemble algorithms to interrelate the variables alongside six metaheuristic optimization algorithms to refine model accuracy and robustness. Subsequently, we explained the best-performing model locally using Shapley additive explanations and clustered variables with similar impacts into distinct groups. These groups were systematically characterized, defining them as environmental settings that shape benzo(a)pyrene dynamics. The setting resulting in the highest indoor benzo(a)pyrene concentrations (197% to 297% relative to the expected value) was dominated by outdoor emissions associated with residential heating and traffic (up to 140%) and indoor source identified as cooking. This integrated approach uniquely enables a quantitative assessment of the contributions from both indoor and outdoor emission sources to pollutant concentrations in indoor spaces, underscoring the importance of both in shaping indoor air quality. Unlike traditional source apportionment methods that assume linear source mixing, our approach integrates nonlinear interactions and contextual variables, such as meteorological conditions and outdoor pollutants, to better capture indoor air quality dynamics. The results also highlight the need for further studies to explore broader contextual factors and refine source attribution in complex urban settings.

Keywords: benzo(a)pyrene; machine learning; metaheuristics; explainable artificial intelligence; AI-based source apportionment

Citation: Jovanović, G.; Perišić, M.; Bezdan, T.; Stanišić, S.; Radusin, K.; Popović, A.; Stojić, A. The $PM_{2.5}$-Bound Polycyclic Aromatic Hydrocarbon Behavior in Indoor and Outdoor Environments, Part III: Role of Environmental Settings in Elevating Indoor Concentrations of Benzo(a)pyrene. *Atmosphere* **2024**, *15*, 1520. https://doi.org/10.3390/atmos15121520

Academic Editor: Célia dos Anjos Alves

Received: 13 November 2024
Revised: 16 December 2024
Accepted: 18 December 2024
Published: 19 December 2024

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs), and their representative benzo(a)pyrene (B[a]P), are ubiquitous environmental pollutants recognized as a substantial health risk for their carcinogenic and mutagenic properties. With individuals spending most of their time indoors, exposure to PAHs in enclosed environments is a growing concern. Recent studies have underscored the dual origin of indoor pollutants in general, as well as PAHs, whose concentrations are shaped by a complex interplay of outdoor and indoor pollution sources, environmental conditions, housing characteristics, and occupant behaviors [1,2].

The primary indoor sources of PAHs include cooking, residential heating, and tobacco smoking, with their impact on indoor air quality being especially pronounced in settings with limited ventilation [3]. Seasonal variations further influence the dynamics between outdoor and indoor pollutant levels, with colder months showing elevated indoor PAH concentrations due to restricted ventilation and increased use of heating fuels, which facilitate the infiltration and accumulation of pollutants indoor [4]. While coal combustion for cooking in developing countries is a major contributor to global premature mortality due to indoor air pollution, practices in those countries also adversely impact indoor air quality. For example, although gas stoves are considered cleaner than biomass stoves in many countries, they still contribute to indoor PAH emissions and associated health risks, particularly respiratory issues such as childhood asthma. A study by Gruenwald et al. [5] in the United States estimated that 12.7% of current childhood asthma cases could be attributed to domestic gas stove use, with variations observed across different states.

Also, Vardoulakis et al. [6] observed that indoor PAH levels can exceed outdoor levels, especially in homes with smoking, kerosene heating, or specific cooking practices. Thereby, specific emission sources contribute differently to indoor PAH levels, based on their molecular weight. It has been shown that low-molecular-weight PAHs are often associated with smoking, cooking and moth repellents, while high-molecular-weight PAHs predominantly originate from outdoor sources.

Even during significant outdoor pollution events, indoor PAH concentrations can surpass outdoor levels, as indoor environments trap pollutants due to limited ventilation and the persistence of indoor sources. As shown by the study of Ghetu et al. [2], who investigated vapor-phase PAHs in indoor and outdoor air before, during, and after wildfire events, indoor PAH concentrations were higher in 77% of samples across all sampling events, with 58% of locations showing increased indoor PAH levels even during wildfires.

Research on pollutant dynamics has conventionally centered on analyzing pollutant concentrations through Principal component analysis (PCA), Unmix, and Positive matrix factorization (PMF), to identify primary emission sources. It is widely acknowledged, however, that numerous factors beyond emissions, such as meteorology, human activities, topography, and various environmental aspects, significantly influence air quality and related health effects.

Traditional source apportionment methods, while useful, face notable limitations. These include the assumption of linear mixing, which oversimplifies the complexity of pollutant interactions, and the inability to capture nonlinear relationships or interactions between variables. Moreover, these methods fail to incorporate critical contextual variables, such as meteorological parameters and outdoor air pollutants, that profoundly affect indoor air quality.

To address this complexity, we introduced the concept of "environmental settings" which encompasses a broad range of natural and anthropogenic factors, which independently, interactively, or collectively impact the distribution and behavior of pollutants, offering a more comprehensive approach to understanding air quality dynamics beyond conventional source apportionment methods. The concept combines advanced machine learning optimized by metaheuristics applied to contextualize variables [7,8] and explainable artificial intelligence and clustering of the resulting local impacts, to group variables by similar impact characteristics that determine distinct environmental settings [9–11]. By resolving issues inherent to linear mixing assumptions and incorporating meteorological and outdoor factors, our approach contextualizes pollutant dynamics to reflect real-world conditions. The approach centered on environmental settings enables a more nuanced understanding of pollutant behavior under varying ambient conditions, whereby the same factor may exert different effects on pollutant depending on the surrounding environmental context [12,13].

Building on our prior research into PAH emission sources across indoor and outdoor environments [14,15], this study narrows its focus to B[a]P dynamics, particularly regarding the complex and under-researched interactions between indoor and outdoor environments.

While outdoor air pollution has been extensively studied, the exchange and behavior of B[a]P across these spaces remain insufficiently understood.

2. Methodology

A three-month measurement campaign (1 March–31 May 2016) was conducted at indoor and outdoor sites at Singidunum University in Belgrade, Serbia (44°45′33.8″ N, 20°29′47.6″ E). The experimental setup and all relevant methodological details are thoroughly explained in our previously published papers [14,15].

2.1. Data

Briefly, concentrations of inorganic gaseous pollutants, radon, $PM_{2.5}$, and particle constituents (trace metals, ions, and PAHs) were measured. Meteorological data, including temperature, humidity, air pressure, wind, and rain characteristics, were recorded, with additional 24-parameter data from the Global data assimilation system (GDAS1). Air sampling was conducted on the rooftop and inside the university building, which hosts around 4000 students. Indoor sampling took place in an amphitheater with a 350-person capacity, where student numbers ranged from 50 to 80. The university is surrounded by residential areas with individual fireboxes (W, SW, NE) and nearby small-scale industries, including the Road Institute of Belgrade, a building company, and a beverage factory stockroom. Additional sources include a confectionery factory, footwear factory, and chemical plants (600 m NW and S), as well as a district heating plant and a fuel oil heating plant used by an urban forestry organization (800 m W and SW). A boulevard with moderate traffic lies 250 m SW, and a high-traffic road is 500 m W-NW. The old city center and river confluence are over 2 km NW, with occasional air quality impacts from emissions across the river.

Outdoor $PM_{2.5}$ and meteorological measurements were conducted at the rooftop (10 m above ground), while indoor air sampling was set 6 m from the amphitheater floor, with the $PM_{2.5}$ device positioned 2 m high. The air sampling system included a Pfeiffer MVP diaphragm vacuum pump (Pfeiffer Vacuum GmbH, Wetzlar, Germany) and manifolds with ports for inorganic gaseous pollutant measurement (O_3, CO, SO_2, NOx) using Horiba 370 series analyzers (Horiba, Ltd., Kyoto, Japan), and electronically controlled valves for alternating indoor/outdoor sampling in ten-minute cycles. $PM_{2.5}$ sampling was performed with Sven Leckel LVS6-RV devices (Sven Leckel, Ingenieurbüro GmbH, Berlin, Germany) at a flow rate of 2.3 m^3/h over 24 h, while the concentrations and constituents (trace metals, ions, PAHs) were analyzed at the Institute of Public Health Belgrade. Outdoor meteorological data were collected via a Vaisala WXT530 station (Vaisala, Vantaa, Finland) on the rooftop, while indoor radon levels, temperature, humidity, and air pressure were monitored using an SN1029 radon monitor and sensors (Sun Nuclear Corporation, Melbourne, FL, USA), placed centrally in the amphitheater at a 1 m height.

$PM_{2.5}$ samples were collected daily on pre-fired quartz filters, with concentrations determined by gravimetric analysis following SRPS EN 12341:2015 [16]. Ion concentrations (Cl^-, Na^+, Mg^{2+}, Ca^{2+}, K^+, NO_3^-, SO_4^{2-}, NH_4^+) were measured via ion chromatography, while trace elements (As, Cd, Cr, Mn, Ni, Pb) were quantified by ICP-MS following SRPS EN 14902:2008/AC:2013 [17]. Sixteen US EPA PAHs were analyzed using GC-MS in accordance with SRPS ISO 12884:2010 [18] after microwave extraction, while inorganic gaseous pollutants O_3, CO, SO_2, NOx were monitored indoors and outdoors according the standards SRPS EN 14625:2013, SRPS EN 14626:2013, SRPS EN 14212:2013/AC:2015, and SRPS EN 14211:2013 [19–22].

The only difference from the dataset used in our previous studies is that the instance recorded on 19 March 2016 was excluded from the analysis conducted in this study. This adjustment was made because the indoor concentrations of benzo(a)anthracene (B[a]A), chrysene (Chr), benzo(b)fluoranthene (B[b]F), benzo(k)fluoranthene (B[k]F), B[a]P, indeno(1,2,3-cd)pyrene (I[cd]P), dibenzo(a,h)anthracene (D[ah]A), and benzo(g,h,i)perylene (B[ghi]P) measured on that day were identified as outliers. In total, 100 indoor and outdoor variables

were included in the analysis to ensure comprehensive assessment of B[a]P dynamics across varying environmental conditions.

2.2. Data Analysis

Seven regression tree ensemble algorithms (AdaBoost, CatBoost, ExtraTrees, Gradient Boosting, Histogram Gradient Boosting, LightGBM, and XGBoost) were used to analyze data. These algorithms enhance prediction accuracy and robustness by combining multiple models to form a more precise ensemble, effectively reducing overfitting [23]. AdaBoost [24] refines model performance by focusing on previously misclassified instances. CatBoost [25] enhances gradient boosting by efficiently handling categorical features and minimizing overfitting with ordered boosting. ExtraTrees builds multiple decision trees by randomly selecting splits, thereby increasing model robustness and reducing variance [26]. LightGBM [27] uses gradient-based one-side sampling and exclusive feature bundling for computational efficiency. XGBoost [28] constructs trees in parallel with regularization to prevent overfitting, supporting various loss functions. Additionally, Gradient boosting and histogram gradient boosting (from Python's sklearn package [29]) improve modeling efficiency: gradient boosting builds models in a forward stage-wise manner, while histogram gradient boosting applies a binning method to reduce memory usage and accelerate training.

To ensure robust evaluation, we applied 5-fold cross-validation, maximizing dataset utilization for both training and validation to reduce overfitting risk. The top three performing models, identified based on r-squared values, underwent further refinement through metaheuristic optimization.

To optimize machine learning method hyperparameters, six metaheuristic algorithms were employed: Firefly Algorithm [30], Artificial Bee Colony [31], Harris Hawks Optimization [32], Sine Cosine Algorithm [33], Slime Mould Algorithm [34], and Quantum Search Algorithm [35]. Each algorithm efficiently navigates the search space to identify near-optimal hyperparameter values. The best-performing model was selected based on its r-squared value after hyperparameter optimization.

After identifying the best-performing model, we applied SHAP (SHapley Additive exPlanations [36]) to interpret the model's predictions. SHAP assigns each feature an importance value for a particular prediction, providing a unified measure of feature importance across a model, enabling comprehensive analysis of feature impact across the dataset. We further refined SHAP values into relative and normalized for clearer interpretation. Relative SHAP values represent the proportionate impact of each feature relative to all other features [14], while normalized SHAP values (adjusted to the expected value) simplify the understanding of impact magnitude.

To deepen the analysis of variable impacts and relationships, we applied dimensionality reduction with UMAP and clustering using HDBSCAN to the obtained SHAP values. UMAP [37] preserved both local and global data structures, making it suitable for complex datasets, while HDBSCAN [38] allowed for hierarchical clustering and the identification of clusters with varying densities. This combination enabled effective identification and categorization of localized impacts.

Based on data and observed impacts, we defined three ranges of normalized predictor levels—low (0–33%), medium (33–66%), and high (66–100%) of the variable's absolute value—and three ranges for normalized impacts: low (1–5%), medium (5–15%), and high (15–100%). Variables with high or medium impacts in a given cluster were identified as key determinants of that cluster, ensuring the meaningful grouping of variables. This ranking system enables the identification of variables that dominate specific environmental settings, offering a systematic approach to differentiate their roles in influencing pollutant dynamics.

3. Results and Discussion

According to the evaluation statistics, out of the top three best performing models, the extra trees algorithm optimized by the sine cosine algorithm provided the best MAPE-to-r-

squared ratio (Table 1). The model metrics include mean absolute error 0.05, mean squared error 0.01, root mean squared error 0.09, mean absolute percentage error 0.24, explained variance 0.98, max error 0.24, and r-squared 0.98.

Table 1. The top three best performing model evaluation statistics.

Metrics	Gradient Boosting/ Harris Hawks Optimization	Extra Trees/ Sine Cosine Algorithm	XGBoost/ Firefly Algorithm
Mean absolute error (MAE)	0.048	0.051	0.058
Mean squared error (MSE)	0.005	0.008	0.011
Root mean squared error (RMSE)	0.070	0.088	0.107
Mean absolute percentage error (MAPE)	0.315	0.240	0.281
Explained variance	0.986	0.979	0.969
Max error	0.213	0.238	0.332
r-squared	0.985	0.976	0.965

3.1. Environmental Settings

In previous studies, we separately examined the impact of indoor and outdoor sources on B[a]P concentrations [14,15]. However, given that indoor concentrations of B[a]P averaged 0.50 ng m^{-3}, closely matching outdoor levels of 0.48 ng m^{-3}, and that there is a limited number of emission sources indoors, the question arose as to the extent to which pollutants from outdoor sources contribute to observed indoor B[a]P levels. To address this, we applied our methodology to a combined dataset from both outdoor and indoor environments, with the aim of distinguishing the respective contributions of outdoor and indoor sources to indoor B[a]P concentrations, and identifying the seasonal and meteorological conditions influencing the compound's dynamics.

By clustering the obtained local impacts, four distinct indoor environmental settings (E1–E4) responsible for the environmental fate of B[a]P were identified (Table 2). Unclustered instances, which may result from the combined influence of multiple sources and sinks, outliers, or transitional regimes, require more detailed analysis and are thus fall outside the scope of this study.

Table 2. Local impact clustering statistics.

Environmental Setting	Mean Impact	Mean Normalized Impact [%]	Mean Absolute Impact	Population Percentage [%]
Unclustered	0.01	2.9	0.7	27.0
E1	1.18	262.9	2.4	16.8
E2	−0.10	−22.8	0.5	24.7
E3	−0.27	−59.7	0.6	11.0
E4	−0.34	−76.2	0.7	20.5

The mean absolute impacts, reflecting the overall influence of variables within each environmental setting on B[a]P dynamics, decreased in the following order: E1 (2.4 ng m^{-3}), E4 (0.7 ng m^{-3}), E3 (0.6 ng m^{-3}), and E2 (0.5 ng m^{-3}). Of the four identified settings, only E1 was associated with elevated B[a]P levels (1.18 ng m^{-3}, 262.9% normalized to the expected value 0.55 ng m^{-3}). Although E1 accounts for a relatively small portion of data instances (approximately 17%), its significance lies in revealing the specific conditions under which B[a]P concentrations notably increase. The elevated B[a]P concentrations observed in E1 were attributed to the combined influence of high indoor levels of B[b]F, B[k]F, I[cd]P, and B[a]A and infiltration of outdoor B[a]P from fossil fuel combustion. These conditions were likely exacerbated during colder periods due to increased heating activities in the vicinity, resulting in enhanced emissions of PAHs from residential heating and traffic sources. The elevated B[a]P levels in E1 are particularly concerning due to their implications for indoor air quality and potential health risks. Long-term exposure to B[a]P,

a known carcinogen, is associated with increased risks of respiratory illnesses and cancer. Therefore, E1 will be examined in detail in this study, as it offers valuable insights into the environmental dynamics and potential sources linked to elevated B[a]P levels. The remaining settings, comprising the other 55.2% of instances, contribute to the reduction of B[a]P levels normalized to the expected value 0.55 ng m^{-3}: E4 (−0.34 ng m^{-3}, −76.2%), E3 (−0.27 ng m^{-3}, −59.7%), and E2 (−0.1 ng m^{-3}, −22.8%).

It is important to note the distinction between an environmental setting and a single emission source, or more generally, between the methodological approach presented here and traditional source apportionment techniques. Compounds from the same source typically show temporal correlations, at least over short periods following their emission into the atmosphere. However, an environmental setting represents a broader context where a set of variables, beyond concentrations alone, jointly influences the dynamics of the target compound—in this case, B[a]P.

Moreover, individual environmental settings usually comprise a synergy of various sub-settings, with each one contributing differently to the observed pollutant levels. To elucidate B[a]P dynamics within a specific environmental sub-setting, it is essential to analyze the relationships and impacts of the most important variables in detail. The impacts of variables defining a sub-setting should correlate; that is, any change in the sub-setting's overall impact should be mirrored by changes in the impacts of its constituent variables. The term "constituent variables" refers to variables classified as having moderate or high impacts, according to the categorization in Section 2. In contrast, variable levels may not necessarily correlate, as they can exhibit independent relationships within the sub-setting, arising from shared emission sources, coexisting sources, or common physico-chemical interactions. If both impacts and levels correlate, however, this suggests that the sub-setting itself represents an emission source, as stable relationships between compound levels and impacts indicate a single, identifiable source. This setting-based approach allows for the identification of consistent patterns in target compound behavior, shaped by the combined effects of multiple variables linked to various sources within a shared environmental context.

3.2. Setting Associated with Elevated Indoor B[a]P Concentrations—E1

The distribution of the most important variable levels and impacts, which collectively constitute the environmental setting E1, is presented in Figure 1. Fluctuations in B[a]P levels relative to the expected value ranged from 0.9 ng m^{-3} (normalized impact 197.4%) to 1.3 ng m^{-3} (296.9%). These fluctuations were predominantly associated with both indoor and outdoor compounds, with notable correlations to indoor species, including B[b]F (relative impact 20.7%), B[k]F (18.9%), I[cd]P (16.6%), B[a]A (8.1%), B[ghi]P (5.7%), Chr (5.0%), pyrene (Pyr) (2.9%), fluoranthene (Fla) (2.8%), and Pb (1.0%). Outdoor-origin compounds, primarily B[a]P (7.0%), B[k]F (4.6%), B[b]F (2.4%), CO (1.4%), and Pyr (0.7%), also show a significant influence on these fluctuations. The influence of other variables was considered negligible.

The degree of impact on indoor B[a]P closely corresponded with the concentrations of five key variables, revealing a consistent alignment between high impacts and concentration levels for indoor B[b]F, B[k]F, I[cd]P, B[a]A, and outdoor-origin B[a]P. Conversely, moderate impacts were associated with high levels of indoor B[ghi]P, Chr, Fla, and Pyr, as well as outdoor-origin B[k]F and B[b]F. Low impacts were paired with high levels of outdoor Pyr and CO, and indoor Pb. The variety of these relationships suggests the presence of several sub-settings within E1, some of which may be attributed to specific emission sources.

Strong correlations in levels of indoor B[b]F, B[k]F, B[a]A, and Chr, with indoor B[a]P (Table 3), along with correlated impacts of these compounds on indoor B[a]P (Table 4), indicate that they originate from a shared emission source, likely linked to the incomplete combustion of organic materials [39]. This combination of high correlations in both levels and impacts points to the significant influence of this emission source on indoor air quality, with the potential to raise indoor B[a]P concentrations by up to 140%.

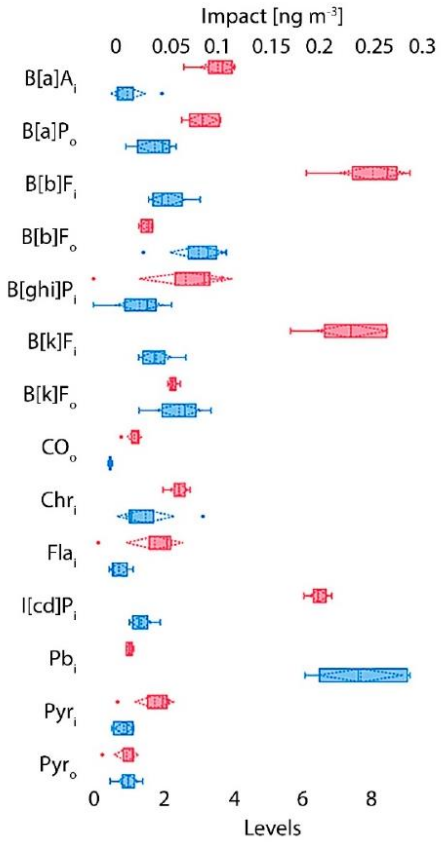


Figure 1. Statistical distribution (box-plot) of variable levels (blue) and impacts (red) characterizing the environmental setting E1 (medium and high impact variables only). The indexes next to the compound abbreviations indicate indoor (i) or outdoor (o) pollutant origin.

Table 3. Correlation of the most important parameter levels in E1. The indexes next to the compound abbreviations indicate indoor (i) or outdoor (o) pollutant origin.

	B[b]F$_i$	**B[k]F$_i$**	**I[cd]P$_i$**	**B[ghi]P$_i$**	**B[a]A$_i$**	**Chr$_i$**	**B[a]P$_o$**	**Fla$_i$**	**Pyr$_i$**	**B[k]F$_o$**	**Pyr$_o$**	**B[b]F$_o$**	**CO$_o$**	**Pb$_i$**
B[b]F$_i$														
B[k]F$_i$	0.98													
I[cd]P$_i$	0.05	−0.09												
B[ghi]P$_i$	0.26	0.11	0.53											
B[a]A$_i$	0.93	0.95	−0.23	0.22										
Chr$_i$	0.92	0.95	−0.28	0.17	1									
B[a]P$_o$	−0.54	−0.58	0.17	−0.18	−0.75	−0.73								
Fla$_i$	−0.33	−0.26	−0.56	−0.59	−0.35	−0.3	0.7							
Pyr$_i$	0.4	0.5	−0.81	−0.46	0.49	0.54	−0.05	0.63						
B[k]F$_o$	−0.71	−0.72	−0.01	−0.3	−0.83	−0.81	0.96	0.76	−0.02					
Pyr$_o$	0.37	0.45	−0.25	−0.67	0.2	0.24	0.24	0.56	0.66	0.16				
B[b]F$_o$	−0.74	−0.82	0.51	0.08	−0.92	−0.93	0.82	0.24	−0.59	0.81	−0.23			
CO$_o$	0.82	0.86	−0.42	−0.12	0.79	0.82	−0.24	0.25	0.84	−0.35	0.67	−0.72		
Pb$_i$	−0.71	−0.59	−0.34	−0.7	−0.54	−0.51	0.02	0.18	−0.13	0.24	−0.06	0.18	−0.54	

Table 4. Correlation of the most important parameter impacts in E1. The indexes next to the compound abbreviations indicate indoor (i) or outdoor (o) pollutant origin.

	$B[b]F_i$	$B[k]F_i$	$I[cd]P_i$	$B[ghi]P_i$	$B[a]A_i$	Chr_i	$B[a]P_o$	Fla_i	Pyr_i	$B[k]F_o$	Pyr_o	$B[b]F_o$	CO_o	Pb_i
$B[b]F_i$														
$B[k]F_i$	0.91													
$I[cd]P_i$	0.41	0.46												
$B[ghi]P_i$	−0.08	−0.13	0.77											
$B[a]A_i$	0.99	0.94	0.39	−0.12										
Chr_i	0.98	0.95	0.48	−0.09	0.98									
$B[a]P_o$	0.59	0.45	−0.09	−0.09	0.62	0.44								
Fla_i	0.93	0.82	0.07	−0.35	0.94	0.87	0.76							
Pyr_i	0.94	0.79	0.1	−0.36	0.93	0.9	0.63	0.97						
$B[k]F_o$	0.57	0.63	0.12	−0.01	0.65	0.48	0.87	0.66	0.48					
Pyr_o	0.92	0.86	0.05	−0.44	0.94	0.89	0.65	0.98	0.97	0.6				
$B[b]F_o$	0.22	0.12	−0.44	−0.25	0.27	0.06	0.9	0.49	0.33	0.76	0.4			
CO_o	0.97	0.85	0.22	−0.21	0.97	0.92	0.73	0.99	0.97	0.65	0.97	0.42		
Pb_i	−0.04	0.14	0.79	0.57	−0.06	0.11	−0.66	−0.39	−0.31	−0.35	−0.32	−0.85	−0.27	

Among indoor sources containing these compounds, the university kitchen adjacent to the amphitheater, although the only significant indoor source of B[a]P, has been assessed to have a relatively minor contribution, as high-temperature cooking techniques, such as grilling and frying, release PAH compounds [40].

Given the absence of other indoor combustion sources and the building's use of electric boilers and heat pumps, it can be inferred that the origin of these compounds lies in fossil fuel combustion in the surrounding outdoor environment, with pollutants subsequently infiltrating the indoor space. Outdoor emissions can permeate through ventilation systems, structural cracks, and other openings, thereby influencing indoor air quality. During colder seasons, such as the period associated with E1 (Figure 2), heightened use of heating fuels in the vicinity correlates with increased indoor levels of these PAHs.

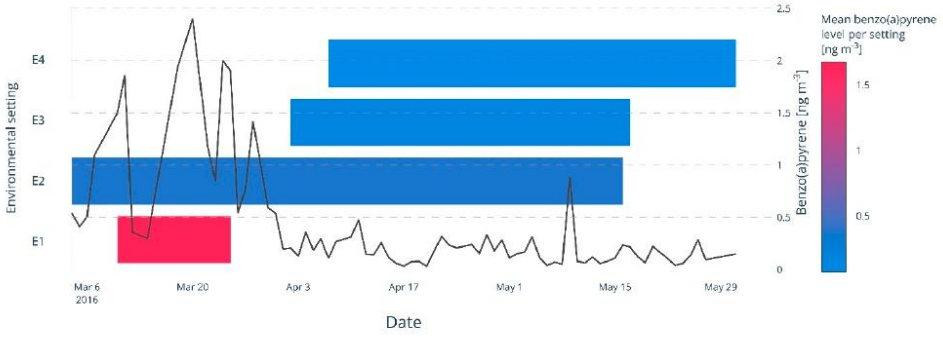


Figure 2. Environmental settings timeline.

Correlations between indoor and outdoor levels of B[k]F, B[b]F, and B[a]P in E1 further substantiate this conclusion, suggesting that outdoor sources contribute substantially to indoor B[a]P concentrations. This indoor-outdoor association might be more readily apparent if not for the behavior of these "infiltrated" compounds, which likely reflect residual concentrations from previous episodes of outdoor air infiltration, with a lag period equal or exceeding one day (measurement frequency). Namely, a shorter lag would likely result in stronger correlations between indoor and outdoor levels for each compound individually; however, the observed dynamics suggest a gradual accumulation of about one day as confirmed by the cross-correlation analysis rather than an immediate reflection of outdoor levels (correlation coefficients for indoor and outdoor B[k]F, B[b]F, and B[a]P

in E1 provided in Table 3 were approximately equal to their cross-correlation coefficients for lag 1). This gradual accumulation implies that outdoor B[a]P contributions to indoor concentrations are moderated by the infiltration process, which delays the direct impact of outdoor emissions and emphasizes the importance of indoor retention and re-suspension of particles.

The degradation rate of PAHs in outdoor environments varies significantly based on the specific PAH compound, the presence of oxidants such as ozone (O_3), hydroxyl radicals (•OH), and nitrogen dioxide (NO_2), as well as environmental factors like UV radiation, temperature, and humidity [41]. On average, the half-life of PAHs in outdoor environment ranges from a few hours to several days. Lower molecular weight PAHs, such as naphthalene, are more reactive and can degrade within hours to a day, while higher molecular weight PAHs, such as B[a]P, tend to be more stable, with half-lives extending from one to several days. The fastest degradation occurs during sunny conditions with high levels of UV radiation and oxidants, as these factors accelerate the transformation of PAHs into oxidized derivatives. However, in indoor environments, PAH compounds tend to remain stable and unaltered due to the absence of key reactive factors present in outdoor settings. Firstly, oxidants such as ozone (O_3), hydroxyl radicals (•OH), and nitrogen dioxide (NO_2), which facilitate PAH transformations in the atmosphere, are typically present at much lower concentrations indoors, significantly reducing the potential for oxidation and chemical transformation of PAHs. Secondly, photolytic reactions, which are common outdoors due to UV radiation from sunlight, rarely occur indoors due to limited exposure to UV light. In enclosed spaces, any UV light exposure is minimal unless there is direct sunlight through windows, further limiting photolytic transformations. As a result, while outdoor environments expose PAHs to reactive and photochemical conditions that can lead to transformations, indoor spaces lack these factors, allowing PAH compounds to persist largely unchanged, aside from gradual physical processes such as deposition on surfaces.

Additionally, in indoor environments, deposition processes play a crucial role in the accumulation and persistence of PAHs. When outdoor pollutants infiltrate indoor spaces, higher molecular weight PAH compounds, such as Pyr, tend to adhere more readily to various surfaces, including walls, furniture, and ventilation ducts, due to their hydrophobic nature and lower volatility. This creates a "reservoir effect", where deposited PAHs gradually accumulate on indoor surfaces and, over time, can re-enter the air influenced by factors such as temperature fluctuations, airflow, and mechanical disturbances. In contrast, CO, due to its gaseous state and chemical stability, remains predominantly in the air and does not contribute to surface deposition. Consequently, while CO levels decline as ventilation or infiltration varies, PAHs bound to surfaces continue to be re-emitted, sustaining elevated indoor concentrations over extended periods and contributing to prolonged exposure risks in indoor settings.

The correlation of indoor B[a]P with I[cd]P and B[ghi]P suggests an influence from high-temperature combustion processes, particularly fossil fuel burning. These specific compounds are commonly associated with the incomplete combustion of heavy fossil fuels, such as fuel oil, frequently used for heating in nearby buildings during colder seasons. The presence of I[cd]P and B[ghi]P, in conjunction with B[a]P, is characteristic of emissions from industrial boilers, heating systems, and diesel engines, indicating contributions from both local heating activities and potential traffic sources. This pattern of PAH compounds supports the conclusion that nearby fossil fuel combustion is a significant contributor to the observed indoor B[a]P levels. The observed correlations and cross-correlation analysis highlight how outdoor B[a]P contributions are moderated by infiltration dynamics, with retention and surface deposition further sustaining elevated indoor concentrations beyond immediate outdoor influences.

4. Limitations

This study has several limitations. Its findings are based on data collected under specific environmental conditions over a defined period in Belgrade, Serbia, and further

research is needed to validate these results across different contexts for a broader understanding of B[a]P dynamics. Furthermore, extending the research to other locations and incorporating longer time series data, covering various seasons and a wider range of emission sources, would provide a more comprehensive insight into seasonal variations and source contributions. While the study accounts for various factors represented by 100 variables, it does not incorporate other significant influences, such as long-range pollutant transport and regional topographical differences, which could impact air quality and B[a]P environmental fate. Expanding the dataset to include these variables could provide a more comprehensive analysis. The discrepancies observed among certain variables within E1 indicate the existence of several sub-settings, each potentially associated with specific emission sources. While further sub-clustering could provide more detailed insights, the limited amount of data in this study restricts this possibility. Future research with expanded datasets could enable a more granular analysis, allowing for the identification of distinct sub-settings and their corresponding sources. Additionally, although the use of ensemble machine learning algorithms, metaheuristics, dimensionality reduction, and clustering is an innovative approach, the study does not compare these methods with other modeling techniques—such as recurrent neural networks, hybrid metaheuristics, or alternative dimensionality reduction and clustering methods—which could enhance the generalizability of the findings. Lastly, further refining the identified emission sources and sinks with advanced statistical and artificial intelligence techniques would strengthen the robustness of the conclusions.

5. Conclusions

This study provides a detailed assessment of elevated indoor B[a]P levels within the context of $PM_{2.5}$, highlighting the complex interactions that shape its environmental fate. Among the machine learning algorithms applied, the ExtraTrees method optimized by the Sine cosine algorithm emerged as the best performer, achieving a high r-squared value alongside low mean absolute percentage error and root mean squared error values, demonstrating robust predictive capability for air quality data.

By clustering local impacts, four distinct indoor environmental settings (E1–E4) associated with B[a]P dynamics were identified, each characterized by varying mean absolute impacts on B[a]P levels. The only setting linked with elevated B[a]P concentrations, accounted for 17% of events, indicating a limited yet significant influence of these conditions on indoor air quality.

The environmental setting E1 is characterized by specific variable levels and impacts, resulting in notable fluctuations in B[a]P concentrations from 0.9 to 1.3 ng m^{-3}. These fluctuations are influenced by both indoor and outdoor compounds, particularly high levels of indoor B[b]F, B[k]F, I[cd]P, B[a]A, and outdoor B[a]P, suggesting shared sources potentially linked to incomplete combustion of organic materials. Although the university kitchen is the only significant indoor source of B[a]P, its contribution remains minor, with PAH levels primarily attributed to infiltration from outdoor fossil fuel combustion sources. This influence is more pronounced during colder periods, corresponding with increased local heating activities. Additionally, interrelations between indoor and outdoor levels of B[a]P, B[k]F, and B[b]F support the conclusion that outdoor sources substantially impact indoor air quality. This association may reflect the accumulation of residual PAH concentrations indoors due to infiltration with a lag period exceeding one day. Outdoors, PAHs degrade relatively quickly due to UV exposure and oxidants, but indoor PAH levels remain stable in the absence of these reactive agents, allowing them to persist through surface deposition. Unlike PAHs, CO levels decline more rapidly indoors, highlighting the sustained correlation of PAHs with outdoor sources. Furthermore, the correlation of indoor B[a]P with I[cd]P and B[ghi]P indicates significant contributions from nearby fossil fuel combustion used for heating, underlining the influence of local combustion activities on indoor PAH levels.

This integrated approach, which incorporates machine learning, metaheuristics, explainable artificial intelligence, and further treatment of the obtained impacts for interpreting complex environmental settings, underscores the critical role of comprehensive contextual analysis in understanding pollutant dynamics and offers a framework for future studies on indoor air quality. The flexibility of this methodology allows its application to diverse geographical and structural environments, provided the contextual variables appropriately describe the pollutant of interest, thus enabling meaningful clustering and variable-impact analyses. Further research with extended datasets and additional contextual variables is recommended to refine these insights and validate findings across broader settings.

Author Contributions: G.J.: conceptualization, formal analysis, funding acquisition, investigation, writing—original draft, review, and editing. M.P.: conceptualization, data curation, formal analysis, funding acquisition, investigation, writing—original draft, review, and editing, project administration. T.B.: methodology, resources, software. S.S.: investigation, writing—original draft, review, and editing. K.R.: investigation, writing—original draft, review, and editing. A.P.: conceptualization, writing—original draft, review. A.S.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, review, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, as well as by the Science Fund of the Republic of Serbia, Grant No. #7373, Characterizing crises-caused air pollution alternations using an artificial intelligence-based framework—crAIRsis.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, J.; Wang, Z.; Wei, Y.; Yang, S.; Yao, S.; Yang, B.; Yang, L. Characteristics, sources, and health risks of PAHs and their derivatives in indoor dust in Zhengzhou. *Atmos. Pollut. Res.* **2024**, *15*, 102246. [CrossRef]
2. Ghetu, C.C.; Rohlman, D.; Smith, B.W.; Scott, R.P.; Adams, K.A.; Hoffman, P.D.; Anderson, K.A. Wildfire impact on indoor and outdoor PAH air quality. *Environ. Sci. Technol.* **2022**, *56*, 10042–10052. [CrossRef] [PubMed]
3. WHO. *Human Health Effects of Polycyclic Aromatic Hydrocarbons as Ambient Air Pollutants: Report of the Working Group on Polycyclic Aromatic Hydrocarbons of the Joint Task Force on the Health Aspects of Air Pollution*; World Health Organization Regional Office for Europe: Copenhagen, Denmark, 2021.
4. Li, C.; Bai, L.; Wang, H.; Li, G.; Cui, Y. Characteristics of indoor and outdoor Polycyclic Aromatic Hydrocarbons (PAHs) pollution in TSP in rural Northeast China: A case study of heating and non-heating periods. *J. Environ. Health Sci. Eng.* **2022**, *20*, 899–913. [CrossRef] [PubMed]
5. Gruenwald, T.; Seals, B.A.; Knibbs, L.D.; Hosgood, H.D. Population attributable fraction of gas stoves and childhood asthma in the United States. *Int. J. Environ. Res. Public Health* **2023**, *20*, 75. [CrossRef]
6. Vardoulakis, S.; Giagloglou, E.; Steinle, S.; Davis, A.; Sleeuwenhoek, A.; Galea, K.S.; Dixon, K.; Crawford, J.O. Indoor exposure to selected air pollutants in the home environment: A systematic review. *Int. J. Environ. Res. Public Health* **2020**, *17*, 8972. [CrossRef]
7. Bacanin, N.; Perisic, M.; Jovanovic, G.; Damaševičius, R.; Stanisic, S.; Simic, V.; Zivkovic, M.; Stojic, A. The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere. *Sci. Total Environ.* **2024**, *929*, 172195. [CrossRef]
8. Bezdan, T.; Perišić, M.; Jovanović, G.; Bačanin, N.; Stojić, A. Artificial Intelligence-Based Framework for Analyzing Crises-Caused Air Pollution. In *Sinteza 2024—International Scientific Conference on Information Technology, Computer Science, and Data Science*; Singidunum University: Beograd, Serbia, 2024; pp. 281–287.
9. Stojić, A.; Vuković, G.; Perišić, M.; Stanišić, S.; Šoštarić, A. Urban air pollution: An insight into its complex aspects. In *A Closer Look at Urban Areas*; Nova Science Publishers: New York, NY, USA, 2018.
10. Stojić, A.; Stanić, N.; Vuković, G.; Stanišić, S.; Perišić, M.; Šoštarić, A.; Lazić, L. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Sci. Total Environ.* **2019**, *653*, 140–147. [CrossRef] [PubMed]

11. Stanišić, S.; Jovanović, G.; Perišić, M.; Herceg Romanić, S.; Milićević, T.; Stojić, A. *Explaining the Environmental Fate of PAHs in Indoor and Outdoor Environments by the Use of Artificial Intelligence*; Nova Science Publishers: New York, NY, USA, 2022.
12. Jovanovic, G.; Perisic, M.; Bacanin, N.; Zivkovic, M.; Stanisic, S.; Strumberger, I.; Alimpic, F.; Stojic, A. Potential of coupling metaheuristics-optimized-xgboost and shap in revealing pahs environmental fate. *Toxics* **2023**, *11*, 394. [CrossRef] [PubMed]
13. Jovanovic, L.; Jovanovic, G.; Perisic, M.; Alimpic, F.; Stanisic, S.; Bacanin, N.; Zivkovic, M.; Stojic, A. The explainable potential of coupling metaheuristics-optimized-xgboost and shap in revealing vocs' environmental fate. *Atmosphere* **2023**, *14*, 109. [CrossRef]
14. Stojić, A.; Jovanović, G.; Stanišić, S.; Romanić, S.H.; Šoštarić, A.; Udovičić, V.; Perišić, M.; Milićević, T. The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels. *Chemosphere* **2022**, *289*, 133154. [CrossRef] [PubMed]
15. Stanišić, S.; Perišić, M.; Jovanović, G.; Milićević, T.; Romanić, S.H.; Jovanović, A.; Šoštarić, A.; Udovičić, V.; Stojić, A. The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: Emission sources. *Environ. Res.* **2021**, *193*, 110520. [CrossRef] [PubMed]
16. *SRPS EN 12341:2015*; Ambient Air—Standard Gravimetric Measurement Method for the Determination of the PM10 or PM2,5 Mass Concentration of Suspended Particulate Matter. Available online: https://iss.rs/en/project/show/iss:proj:49389 (accessed on 29 May 2015).
17. *SRPS EN 14902:2008/AC:2013*; Ambient Air—Standard Method for the Measurement of Pb, Cd, As and Ni in the PM10 Fraction of Suspended Particulate Matter. Available online: https://iss.rs/en/project/show/iss:proj:46826 (accessed on 27 August 2013).
18. *SRPS ISO 12884:2010*; Ambient air—Determination of Total (Gas and Particle-Phase) Polycyclic Aromatic Hydrocarbons—Collection on Sorbent-Backed Filters with Gas Chromatographic/Mass Spectrometric Analyses. Available online: https://iss.rs/en/project/show/iss:proj:24983 (accessed on 29 January 2010).
19. *SRPS EN 14625:2013*; Ambient Air—Standard Method for the Measurement of the Concentration of Ozone by Ultraviolet Photometry. Available online: https://iss.rs/en/project/show/iss:proj:40991 (accessed on 31 May 2013).
20. *SRPS EN 14626:2013*; Ambient Air—Standard Method for the Measurement of the Concentration of Carbon Monoxide by Non-Dispersive Infrared Spectroscopy. Available online: https://iss.rs/en/project/show/iss:proj:40993 (accessed on 31 May 2013).
21. *SRPS EN 14212:2013/AC:2015*; Ambient Air—Standard Method for the Measurement of the Concentration of Sulphur Dioxide by Ultraviolet Fluorescence. Available online: https://iss.rs/en/project/show/iss:proj:54201 (accessed on 31 August 2015).
22. *SRPS EN 14211:2013*; Ambient Air—Ambient Air—Standard Method for the Measurement of the Concentration of Nitrogen Dioxide and Nitrogen Monoxide by Chemiluminescence. Available online: https://iss.rs/en/project/show/iss:proj:40985 (accessed on 31 May 2013).
23. Dietterich, T.G. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000.
24. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
25. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*. Available online: https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html (accessed on 12 November 2024).
26. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
27. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (accessed on 12 November 2024).
28. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 118–1232. [CrossRef]
29. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
30. Yang, X.S. Firefly algorithms for multimodal optimization. In *International Symposium on Stochastic Algorithms*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 169–178.
31. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization*; Technical Report-tr06; Erciyes University, Engineering Faculty, Computer Engineering Department: Kayseri, Türkiye, 2005; Volume 200, pp. 1–10.
32. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]
33. Mirjalili, S. SCA: A sine cosine algorithm for solving optimization problems. *Knowl.-Based Syst.* **2016**, *96*, 120–133. [CrossRef]
34. Li, S.; Chen, H.; Wang, M.; Heidari, A.A.; Mirjalili, S. Slime mould algorithm: A new method for stochastic optimization. *Future Gener. Comput. Syst.* **2020**, *111*, 300–323. [CrossRef]
35. Zhang, J.; Xiao, M.; Gao, L.; Pan, Q. Queuing search algorithm: A novel metaheuristic algorithm for solving engineering optimization problems. *Appl. Math. Model.* **2018**, *63*, 464–490. [CrossRef]
36. Lundberg, S. A unified approach to interpreting model predictions. *arXiv* **2017**, arXiv:1705.07874.
37. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
38. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density-based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [CrossRef]

39. Abdel-Shafy, H.I.; Mansour, M.S. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egypt. J. Pet.* **2016**, *25*, 107–123. [CrossRef]
40. Sampaio, G.R.; Guizellini, G.M.; da Silva, S.A.; de Almeida, A.P.; Pinaffi-Langley, A.C.C.; Rogero, M.M.; de Camargo, A.C.; Torres, E.A. Polycyclic aromatic hydrocarbons in foods: Biological effects, legislation, occurrence, analytical methods, and strategies to reduce their formation. *Int. J. Mol. Sci.* **2021**, *22*, 6010. [CrossRef] [PubMed]
41. Gbeddy, G.; Goonetilleke, A.; Ayoko, G.A.; Egodawatta, P. Transformation and degradation of polycyclic aromatic hydrocarbons (PAHs) in urban road surfaces: Influential factors, implications and recommendations. *Environ. Pollut.* **2020**, *257*, 113510. [CrossRef] [PubMed]

atmosphere

MDPI

Article

The Explainable Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing VOCs' Environmental Fate

Luka Jovanovic [1], Gordana Jovanovic [2,3], Mirjana Perisic [2,3], Filip Alimpic [2], Svetlana Stanisic [3], Nebojsa Bacanin [1], Miodrag Zivkovic [1] and Andreja Stojic [2,3,*]

[1] Faculty of Informatics and Computing, Singidunum University, Danijelova 32, 11010 Belgrade, Serbia
[2] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
[3] Environment and Sustainable Development, Singidunum University, Danijelova 32, 11010 Belgrade, Serbia
* Correspondence: andreja.stojic@ipb.ac.rs

Abstract: In this paper, we explore the computational capabilities of advanced modeling tools to reveal the factors that shape the observed benzene levels and behavior under different environmental conditions. The research was based on two-year hourly data concentrations of inorganic gaseous pollutants, particulate matter, benzene, toluene, m, p-xylenes, total nonmethane hydrocarbons, and meteorological parameters obtained from the Global Data Assimilation System. In order to determine the model that will be capable of achieving a superior level of performance, eight metaheuristics algorithms were tested for eXtreme Gradient Boosting optimization, while the relative SHapley Additive exPlanations values were used to estimate the relative importance of each pollutant level and meteorological parameter for the prediction of benzene concentrations. According to the results, benzene levels are mostly shaped by toluene and the finest aerosol fraction concentrations, in the environment governed by temperature, volumetric soil moisture content, and momentum flux direction, as well as by levels of total nonmethane hydrocarbons and total nitrogen oxide. The types of conditions which provided the environment for the impact of toluene, the finest aerosol, and temperature on benzene dynamics are distinguished and described.

Keywords: machine learning; extreme gradient boosting; metaheuristics; swarm intelligence; explainable artificial intelligence; volatile organic compounds; BTEX; benzene

Citation: Jovanovic, L.; Jovanovic, G.; Perisic, M.; Alimpic, F.; Stanisic, S.; Bacanin, N.; Zivkovic, M.; Stojic, A. The Explainable Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing VOCs' Environmental Fate. *Atmosphere* **2023**, *14*, 109. https://doi.org/10.3390/atmos14010109

Academic Editor: Jacek Koziel

Received: 14 December 2022
Revised: 26 December 2022
Accepted: 28 December 2022
Published: 4 January 2023

1. Introduction

The existing scientific evidence clearly shows that there are no air pollutant concentration limits below which the adverse health effects are excluded [1]. The current estimations warn that exposure to air pollution causes 7 million premature deaths each year globally and contributes to the early onset and exacerbations of noncommunicable diseases, including asthma and other respiratory disorders, coronary and neurodegenerative diseases, and diabetes [2]. The relevant authorities including United Nations (UN) and World Health Organization (WHO) consider air quality control and research imperative to public health and environmental sustainability. As a result, air quality is monitored in more than 6000 cities in 117 countries worldwide compared to 1100 cities in 91 countries a decade ago [3].

In addition to commonly monitored suspended particulate matter (PM) and inorganic gaseous pollutants, a group of volatile organic compounds (VOCs), represented by BTEX (benzene, toluene, ethylbenzene, and xylene) has been the focus of scientific and public interest, due to a large body of literature addressing the benzene carcinogenicity, as well as relationships of its homolog compounds with the respiratory, hematologic, reproductive, and nervous system disorders (e.g., [4,5] and references therein). Seasonal variations in BTEX concentrations have been reported worldwide, with the highest levels registered

during the cold season due to the lower atmospheric boundary layer and intensified fossil fuel combustion for heating purposes. On a daily basis, both natural and anthropogenic factors result in pronounced diurnal BTEX concentration dynamics, with peaks occurring in the morning and evening, and the lower values being registered in the meantime [6].

In addition to their detrimental effects on human health, benzene and its homologs are highly reactive compounds and the key precursors for the generation of secondary organic aerosol (SOA) and tropospheric ozone in the atmosphere. Photochemical reactions in which BTEX are involved depend on the sunlight, and the presence of oxidative species such as nitrogen oxides and many short-lived radicals, e.g., OH, alkyl peroxide, and hydrogen peroxide radicals [7,8]. In photochemical reactions, VOCs act as fuel, while NO_x has a role of a catalyst. At low VOC/NO_x ratios, which is often recognized as a VOC-limited regime and observed in urban polluted areas, maximum ozone concentrations that result from initial pollutant mixtures are defined by VOC levels, whereas, at high VOC/NO_x ratios, the O_3 production rate is limited by the supply of NO_x [9]. In the presence of suspended particles (PM), the photochemical oxidation and subsequent condensation of VOCs lead to the production of SOA. Among volatile organics, alkenes and aromatics including BTEX appeared to possess the highest O_3 formation potential of 59.6%, and 65.3%, respectively, while aromatics also largely contribute to the SOA formation (95%), as reported by Li et al. [10]. Similarly, Zhan et al. [11] concluded that out of 51 VOCs, benzene homologs, toluene and xylenes, were the main species responsible for SOA formation, as well as that alkenes and aromatics dominantly facilitated the production of ground-level O_3 (56.8% and 30.3%, respectively).

The nonlinear nature of BTEX behavior and the seasonal dynamics of their particle–gas partitioning require a multidisciplinary scientific approach and advanced computational capabilities of modeling tools and methods [12] which enable us to research relationships in the environment, enhance our current knowledge and provide the basis for future sustainability. Galán-Madruga and García-Cambero [13] performed multiple linear regression analysis to predict benzene concentration based on the independent atmospheric pollutants and meteorological factors, while Jephcote and Mah [14] used Bayesian multilevel models to provide analysis of benzene exposures from the petrochemical industry, connecting industrial emissions to pollution episodes and disparities in regional mortality rates. Generally, the researchers have combined a variety of methods such as eXtreme Gradient Boosting (XGBoost), Generalized AutoRegressive Conditional Heteroskedasticity (GARCH), neural networks, or light gradient boosting machine (LightGBM) algorithm to predict VOCs, PM, and PAHs concentrations, or haze hazards (e.g., [15–20]).

Machine learning methods, capable of interrelating pollutant evolution with the environmental conditions in which it occurs, require adaptations for each individual problem (dataset), which is considered to be a nondeterministic polynomial-hard (NP-hard) challenge by nature. It is extremely time-consuming if this task is done manually. Additionally, NP-hard challenges are not possible to be solved by applying traditional deterministic approaches, as it would require an impractical amount of time and resources. On the other hand, stochastic algorithms, where swarm intelligence metaheuristics algorithms belong, can be used to determine satisfactory solutions within a reasonable time. Thus, some papers have addressed the application of metaheuristic methods for the optimization and training of artificial intelligence techniques or statistical regression models, with the aim to improve their performance and reveal novel information on air quality and its impact on human health [21–23].

This paper proposes an XGBoost model tuned by the metaheuristics algorithms for this particular problem, as the survey of the available literature shows that it hasn't been done before. Eight well-known metaheuristics algorithms were employed to tune the XGBoost hyperparameters with the goal to determine the model that will be capable of achieving a superior level of performance on the observed dataset. Moreover, building on our previous experience in modeling environmental phenomena and machine learning method hyperparameter optimization [15–17,24–28], this study aimed to explore the potential of

coupling the advanced methodologies to capture and characterize defining factors and processes that shape benzene's fate in an urban environment. By conducting simulations for the purpose of this research, some of the most promising metaheuristics for tackling combined integer and continuous NP-hard challenges, such as XGBoost tuning, were employed in contextualizing the air pollutant concentrations and meteorological data, and the results of the best-generated model were interpreted using SHapley Additive exPlanations (SHAP). Taking into account that there is a research gap in AI applications for environmental sciences, the proposed research represents a significant contribution to this scientific domain, especially because two families of AI methods, machine learning and metaheuristics, were coupled together to generate a model that can be used to reveal causal relationships between air pollutants and factors governing their behavior, as well as to emphasize what novel conclusions on the environmental phenomena can be drawn by a multidisciplinary approach.

2. Background and Preliminaries

2.1. The XGBoost Algorithm

The XGBoost model utilizes the adaptive training technique to tune the objective function. Accordingly, every step of the tuning procedure is depending on the previous step with respect to the result. The objective function of the XGBoost approach can be mathematically articulated as follows:

$$F_o{}^i = \sum_{k=1}^{n} l\left(y_k, \hat{y}_k^{i-1} + f_i(x_k)\right) + R(f_i) + C, \tag{1}$$

where the loss of the tth round is denoted as l, constant term C, and the regularization parameter R of XGBoost, that can be obtained by:

$$R(f_i) = \gamma T_i + \tfrac{\lambda}{2} \sum_{j=1}^{T} w_j^2 \tag{2}$$

In general, the complexity of the XGBoost tree structure is correlated to the values of γ and λ tuning parameters. The increase of the parameters will result in a more simple tree structure of the model. The first and the second derivatives of the XGBoost, named g and h, are calculated in the following way:

$$g_j = \partial_{\hat{y}_k^{i-1}} l\left(y_j, \hat{y}_k^{i-1}\right) \tag{3}$$

$$h_j = \partial^2_{\hat{y}_k^{i-1}} l\left(y_j, \hat{y}_k^{i-1}\right) \tag{4}$$

Finally, the solution is determined by applying the following two equations:

$$w_j^* = -\tfrac{\sum g_t}{\sum h_t + \lambda} \tag{5}$$

$$F_o{}^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum g)^2}{\sum h + \lambda} + \gamma T, \tag{6}$$

where the loss function value is denoted by $F_o{}^*$, while the solution's weight values are represented by w_j^*.

2.2. Metaheuristic Optimization

Metaheuristics belong to the group of stochastic algorithms, that are commonly used to address NP-hard challenges. These problems are not possible to be solved with traditional deterministic methods, due to their complexity and an impractical amount of required resources. Metaheuristics algorithms are grouped into several distinctive families, still, many authors use different classifications. The mostly adopted taxonomy classifies the metaheuristics based on the natural phenomenon that inspired the search procedure of

the given algorithm [29–31]. According to this taxonomy, metaheuristics algorithms are classified into nature-inspired algorithms (with two distinctive subgroups, swarm intelligence and genetic algorithms), approaches inspirited by physical phenomena (for example, gravitational or electromagnetic force, waves, etc.), approaches based on human behavior (brainstorming, social networks, teaching, etc.) and mathematics-based approaches (derived from the properties of the sine, cosine, and arithmetical operators).

The swarm intelligence family is based on the animal behavior, typically exhibited by the large groups of individual beings (swarms) that tend to show very coordinated and complex actions while hunting, foraging, migrating and mating [32,33]. The methods belonging to this particular family have established themselves as very efficient optimizers that have been employed recently in a wide spectrum of practical NP-hard challenges. The most notable swarm intelligence algorithms are ant colony optimization (ACO) [34], (PSO) [35], artificial bee colony (ABC) [36], bat algorithm (BA) [37,38], and firefly algorithm (FA) [39], among many others.

The algorithms inspired by fundamental mathematical functions have recently gained popularity. The most important exemplars include the sine–cosine algorithm (SCA) [40] and the arithmetic optimization algorithm (AOA) [41]. The former mimics the mathematical fluctuation of the sine and cosine functions, and the latter relies on basic mathematical operations. Another recent metaheuristic that belongs to this family is the golden sine algorithm (Gold-SA) [42].

The biggest challenge in the application of population-based metaheuristics is summed in the no free lunch (NFL) theorem [43], stating that the universal algorithm that is capable to obtain the best results for all optimization tasks is not existing. In other words, one algorithm can be superior for one particular optimization problem, but completely fail when applied to other problems. Therefore, there is a large diversity in the metaheuristics and their implementations, and the appropriate algorithm must be tailored for every use case.

Some of the most successful contemporary applications of the population-based methods include COVID-19 case number prediction [44,45], cloud computing [46–48], cloud-edge computing [49], wireless sensor network tuning [50–53], feature selection task [54,55], classification of MRI scans and medical applications in general [56,57], global optimization problems [58], credit card frauds [59,60], pollution estimation [61], network security [62,63], as well as general tuning of the machine learning models [64–67].

The XGBoost model employed in this research has also been subjected to tuning by metaheuristics approaches. The paper [68] tests the classification capabilities of several metaheuristics methods alongside XGBoost, [69] utilizes PSO to address the network intrusion problem, and [70] employs the XGBoost and genetic algorithm (GA) for stock prices forecasting. Moreover, the XGBoost model optimized by metaheuristics has been extensively used as a part of intrusion detection and network security solutions [71–74].

2.3. Shapley Additive Explanations

The explainability of machine learning model behavior is crucial for understanding the process being modeled. The inability to explain the predictions derived from accurate but complex models posed a severe limitation in understanding the governing factors that shape predictions until recently.

To explain the obtained best-performing model, we applied the advanced explainable artificial intelligence method SHAP. SHAP avoids the trade-off between accuracy and interpretability and provides a straightforward and meaningful interpretation of the machine learning model-derived decisions. The method is based on Shapley values, calculated as a feature importance measure by a game theory approach that provides an impact of features on individual predictions [75]. In brief, Shapley values represent fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). They apportion the difference between the prediction and the average prediction among the features [76]. Thus, SHAP assigns each feature importance as

a measure of its contribution to a particular prediction and interprets the impact compared to a model's prediction if that feature took some baseline value. This way, the method provides valuable insights into a model's behavior and (1) overcomes the main drawback of inconsistency, (2) minimizes the possibility of underestimating the importance of a feature with a specific attribution value, and (3) captures feature interaction effects based on a generalization of Shapley values and interpreting the model's global behavior while retaining local faithfulness [24,77].

We use the relative SHAP values introduced by Stojic et al. [17] to gain insight into relative relationships among feature attributions for each prediction. Relative SHAP values show the relative influence of a feature on the prediction. They are defined as a share of absolute SHAP in total attributed importance of all features for the particular prediction.

This study used Python SHAP implementation (SHAP Python package) and TreeExplainer [77] to obtain SHAP values that we used to produce SHAP dependency plots representing the change of feature importance over its value range.

3. Materials and Methods

3.1. Data

For this study, the concentrations of inorganic gaseous pollutants (NO, NO_2, NO_x, O_3), particulate matter (PM_1, $PM_{2.5}$, and PM_{10}), and benzene, toluene, m, p-xylene, and total nonmethane hydrocarbons (TNMHC) were obtained from the station of regulatory air quality monitoring Vatrogasni Dom in Pančevo (Serbia). A two-year database (2019–2020) of air pollutants (11,368 hourly concentrations) was complemented by meteorological parameters obtained from the Global Data Assimilation System (GDAS1).

Hourly concentrations of organic (benzene, toluene, and m, p-xylene) and inorganic gaseous pollutants (NO, NO_2, NO_x, and O_3) were obtained using referent sampling devices, according to the European standards EN 14662-3, EN 14211, and EN 14625. GRIMM EDM 180 was used to determine hourly concentrations of particulate matter conferring the standards EN 12341 and EN 14907. Gas chromatograph Syntech Spectras GC955, which separates methane from other hydrocarbons and measures the concentration of methane and other total nonmethane hydrocarbons in the air, was used for measuring concentrations of TNMHC.

3.2. Study Area

Pančevo is a city of over 100,000 inhabitants located on the left bank of the Danube, 20 km east and northeast of Belgrade, the largest Serbian metropolitan area. The sampling site (44°51′31″ N, 20°38′56″ E), characterized as an urban background station, is situated about 500 m south of the city center, surrounded by the residential areas from E and NE sides, and small-scale industry referring scrap metal sorting and storage center, and factory for the flour production in the nearest vicinity. The E70 European corridor, with public transport and intensive vehicle flow, passes by approximately 200 m in the S-SW direction from the sampling site. The confluence of two navigable rivers, Tamiš and Danube, is located at a distance of approximately 500 m in the SW direction. Two kilometers SE stands one of the largest industrial complexes in this part of Serbia. It includes three main factories for producing artificial fertilizers (HTP Azotara), manufacturing chemical products (HTP Petrohemija), and the largest center for oil processing in Serbia (Pančevo Oil Refinery), and a few smaller chemical industrial plants.

3.3. Metaheuristics

This section describes the eight metaheuristics algorithms that were utilized to tune the XGBoost model for the purpose of this research. The chosen algorithms are well-known optimizers, that have been employed to solve various NP-hard challenges in the past, with a great deal of success. All algorithms were used in their original versions, with the control parameters' values recommended in their respective publications.

3.3.1. Genetic Algorithm

Genetic algorithm (GA) is an evolutionary algorithm inspired by natural selection. The processes of selection, inheritance, crossover, and mutation at the level of cells are simulated. The creator of genetic algorithms to solve various optimization problems was Goldberg [78,79]. The author of the GA is Mirjalili [80].

Initial population individuals have a set of properties which represent chromosomes that are alterable and mutative. Similar to the biological evolution process, GA is modifying the population of individuals over the iterations. In every round of execution, GA takes the best solutions from the population, to produce offspring. As the iterations pass, over several generations, the population as a whole is evolving in the direction of the optimum.

Based on the individual's fitness, the parents are selected for the creation of the individual in the next generation. Additionally, they can be crossed over by selecting two individuals and exploiting their advantages to create a better one. Finally, the process of mutation can be applied to a single individual to alter its previous properties for better fitness in the next generation. The interested reader can refer to the [80] for more details.

3.3.2. Particle Swarm Optimization

Kennedy and Eberhart developed a heuristic optimization method called Particle Swarm Optimization (PSO) in 1995 [35]. Birds and fish flocking are the main inspirations of the algorithm. The particles, which are considered individuals in the population act as search agents. Their goal is to provide satisfactory solutions for discrete and continuous optimization problems.

The collective experience is shared in search of the best solution, which consists of the individual best experience and those of neighboring solutions. After the evaluation of the gathered experiences, the next move is decided. Initially, random velocities are given to each particle in the generated population, which are represented as initial positions. The particles move over iterations and the best position of each one is stored.

The velocity with which the particle moves is a sum of three components' weights: the old velocity, the velocity that leads in direction of the best solution so far, and the velocity toward the best solution obtained by neighboring particles.

$$\begin{cases} \overrightarrow{v_i} \leftarrow \overrightarrow{v_i} + \overrightarrow{U}(0, \phi_1) \otimes (\overrightarrow{p_i} - \overrightarrow{x_i}) + \overrightarrow{U}(0, \phi_2) \otimes (\overrightarrow{p_g} - \overrightarrow{x_i}) \\ \overrightarrow{x_i} \leftarrow \overrightarrow{x_i} + \overrightarrow{v_i} \end{cases} \tag{7}$$

where the $\overrightarrow{U}(0, \phi_1)$ shows a vector consisting of uniformly distributed random numbers in the range of 0 to ϕ_i, randomly generated during each iteration for each particle. The $\otimes$ represents the componentwise multiplication. Each component of v_i is inside the range of $[-V_max, +V_max]$. More details about this algorithm are available in [35].

3.3.3. Artificial Bee Colony

The artificial bee colony (ABC) metaheuristics were developed by Karaboga with the goal to target continuous optimization problems, based on the honey-collecting behavior of the bees [36]. The original ABC implementation puts into use three control parameters, and models three types of bees: workers, onlookers, and scouts, where 50 % of the colony is allocated as worker bees.

This approach models the food sources as the possible solutions of the problem, and each individual food source is allocated with exactly one worker bee. Workers execute the search procedure by investigating the area in the proximity of the solution (food source). The onlookers pick the food source to be exploited with respect to the data collected by workers. Lastly, if the food source is not improved after a defined number of iterations, scout bees will replace it with a novel, arbitrary food source. This procedure is controlled by the *limit* parameter, as described in [81].

As stated above, the worker investigates the neighborhood, and if it comes upon a food source, it will evaluate its fitness value. The process of discovering the novel solution in the neighborhood is modeled by ABC search equation given by Equation (8):

$$v_{i,j} = \begin{cases} x_{i,j} + \phi * (x_{i,j} - x_{k,j}), R_j < MR \\ x_{i,j}, otherwise \end{cases} \quad (8)$$

where $x_{i,j}$ denotes the jth component of the previous solution i, $x_{k,j}$ denotes jth component of a discovered solution k, ϕ denotes an arbitrary value in range $[0, 1]$, while MR denotes a control parameter that controls the modification rate. In case that the fitness of the novel solution is better than the old solution, the worker continues the exploitation of the novel food source. Worker bees share gathered data with onlookers, who will choose a food source i with a probability that is proportional to the solution's fitness:

$$p_i = \frac{fitness_i}{\sum_{i=1}^{N} fitness_i} \quad (9)$$

For more details about the entire algorithm please refer to the original publication [36].

3.3.4. Firefly Algorithm

The firefly metaheuristics was proposed in 2009 by Yang [82]. The algorithm was based on the flashing behavior exhibited by fireflies. The algorithm utilizes the brightness and attractiveness of these insects, where the brightness is calculated by the value of the objective function, while the attractiveness property is depending on the brightness. When the distance between two fireflies is reduced, the attractiveness is increasing, and vice versa.

The FA search equation defined for an arbitrary solution i, that traverses to the new position x_i in iteration $t + 1$, toward the solution j that is more attractive (brighter), is provided by Equation (10):

$$x_i^{t+1} = x_i^t + \beta_0 \cdot e^{-\gamma r_{i,j}^2}(x_j^t - x_i^t) + \alpha^t(\kappa - 0.5) \quad (10)$$

where α denotes the randomization parameter, κ is an arbitrary value from the Gaussian or uniform distribution, and $r_{i,j}$ marks the distance between solutions i and j. Common values for β_0 and α parameters are determined as 1 and $[0, 1]$, respectively, and are suitable for most of the optimization problems. For more details about the entire algorithm please refer to the original publication [82].

3.3.5. Bat Algorithm

Bat algorithm (BA) was proposed in 2010 by Yang [37], and it is inspired by the hunting behavior and echolocation utilized by bats to catch prey or avoid trees and similar obstacles. Bats use sound wave reflection to estimate the distance to nearby objects and to form their images.

Bats execute the search phase by utilizing the Equation (11):

$$x_i^t = x_i^{t-1} + v_i^t, \quad (11)$$

that is used to define the solution's current position where the locations of the solution x_i in two consecutive rounds of execution are given by x_i^{t-1} and x_i^t, respectively. The speed of the solution x_i is determined by v_i^t, calculated with Equation (12):

$$v_i^t = v_i^{t-1} + (x_i^{t-1} - x_*)f_i, \quad (12)$$

where x_* represents the latest global best location, and f_i specifies the frequency utilized by ith bat in the population. During the exploitation phase, the algorithm utilizes a random

walk procedure to update the position of the current fittest bat. When the prey is located, the bats update their loudness by changing the pulse-emitting rate.

The exploitation relies on the random walk that updates the current best solution, described with the Equation (13):

$$x_{new} = x_{old} + \epsilon A^t, \tag{13}$$

where the mean loudness of all solutions is given by A^t, while the ϵ is a scaling factor given as an arbitrary number in the range $[0, 1]$.

Finally, when the prey is located, the loudness of bats is updated by utilizing the Equation (14):

$$A_i^t = \alpha A_i^{t-1},\ r_i^t = r_i^0[1 - exp(-\gamma t)] \tag{14}$$

$$A_i^t \to 0,\ r_i^t \to r_i^0,\ \text{while } t \to \infty \tag{15}$$

where A_i^t describes the loudness of ith bat in round t, and r represents the pulse emission rate. The parameters α and γ are constants. Additional information about this process is available in [37].

3.3.6. Whale Optimization Algorithm

The whale optimization algorithm was introduced by Mirjalili et al. in 2016 [83], and the algorithm mimics the unique hunting tactic of humpback whales, known as the bubble net method. While executing this maneuver, whales cooperatively dive under the flock of fish, and move upward in spirals while simultaneously blowing bubbles that trap the prey and force it to also swim toward the surface, where it can easily be caught by hunters.

The algorithm implements this bubble net strategy in the exploitation phase, while the exploration is implemented as the pseudorandom search for the fish. As the WOA is a population-based method, the latest best-candidate solution describes the prey, while the remaining solutions represent the whales.

The exploitation bubble net approach assumes that whales circle around the target by moving in the spirals and simultaneously decrease the circle radius, and it can be modeled by switching with equal probability p (determined in each iteration by a random value from 0 to 1), between two options described by the Equation (16):

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D}\ \text{, if } p < 0.5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t)\ \text{, if } p \geq 0.5 \end{cases}. \tag{16}$$

where $\vec{D}'$ denotes the distance between the ith solution and the global best, obtainable by $\vec{D}' = |\vec{X}^*(t) - \vec{X}(t)|$, and b is a fixed value used to control the dimensions of the logarithmic spiral. Finally, parameter l denotes the arbitrary value within $(-1, 1)$.

While executing the exploration, each individual whale updates its location with respect to the location of a random solution rather than to the global best. Vector $\vec{A}$ is employed in a way that if the produced arbitrary values are greater or equal to 1 ($|A| \geq 1$), the new location of the whale is directed to the random solution, thus the global search is performed. This is modeled by the Equation (17), as proposed in [83]:

$$\vec{X}(t+1) = \vec{X}_{rnd}(t) - \vec{A} \cdot \vec{D}, \tag{17}$$

where $\vec{D}$, representing the distance from the ith solution to the arbitrary solution rnd in iteration t, is obtained as $\vec{D} = |\vec{C} \cdot \vec{X}_{rnd}(t) - \vec{X}(t)|$. More details about the WOA are available in the original publication [83].

3.3.7. Harris' Hawks Optimization

The Harris' Hawks optimization metaheuristics was inspired by a variety of hunting techniques employed by these hawks to attack and capture the prey in nature. It is one of the most recent algorithms, being introduced by Heidari et al. in 2019 [84].

During the exploration stage, the algorithm tries to discover the solution nearest to the global optimum. The solutions are arbitrarily produced on several positions, and they move closer to the prey in each step, imitating the hawks perching behavior. The algorithm makes use of two methods having the same probability, decided by the parameter q as proposed in [84]:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2X(t)|, q \geq 0 \\ (X_{best}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)), q < 0.5, \end{cases} \quad (18)$$

where q, as well as r_1, r_2, r_3 and r_4, denote arbitrary values within the range $[0, 1]$, that are being updated in each round, $X(t+1)$ denotes the solutions' location in next round, while $X_{best}(t)$, $X(t)$ and $X_m(t)$ represent the best, current and average solutions' locations in the current round t. Lastly, LB and UB denote the lower and upper boundaries of the search domain. The average position of the solutions $X_m(t)$ is determined by:

$$X_m(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t), \quad (19)$$

where N is the total count of individuals and $X_i(t)$ denotes the location of solution X in round t.

HHO can transit from exploitation to exploration and vice versa multiple times, with respect to the solution's strength (representing the prey escaping energy). The strength of the solution is updated in every round in the following way:

$$E = 2E_0(1 - \frac{t}{T}), \quad (20)$$

where T denotes the maximal number of iterations and E_0 is the starting prey's energy, changing arbitrarily inside the $[-1, 1]$ range.

In the exploitation stage, the hawks start attacking the prey, who is trying to flee. The hawks therefore must employ various strategies to overtire the prey for an easy catch. They are moving closer to the target, and mix between soft and hard besiege, based on the target's remaining energy as follows. If $|E| \geq 0.5$, hawks employ the soft besiege approach, else if $|E| < 0.5$, hard besiege is used.

In the situation where $r \geq 0.5$ and $|E| \geq 0.5$, the prey is still not tired, and the hawks will surround it in a soft manner to exhaust it, as described in the following equations:

$$X(t+1) = \Delta X(t) - E|JX_{best}(t) - X(t)| \quad (21)$$

$$\Delta X(t) = X_{best}(t) - X(t), \quad (22)$$

where $\Delta X(t)$ represents a vector difference from the best solution (representing the hawks' prey) and solution position in round t. Variable J is modified randomly in every iteration to mimic the prey's escaping technique:

$$J = 2(1 - r_5), \quad (23)$$

where r_5 represents an arbitrary produced value in range $[0, 1]$). If $r \geq 0.5$ and $|E| < 0.5$, the prey is fatigued and hawks move to a hard attack strategy, where the current locations are being updated by:

$$X(t+1) = X_{best}(t) - E|\Delta X(t)| \quad (24)$$

If the pray yet has a certain amount of energy remaining, the hawks will use zigzag patterns before the attack, modeled by the next equation:

$$Y = X_{best}(t) - E|JX_{best}(t) - X(t)|, \quad (25)$$

followed by hawks' dives in leapfrog patterns defined in the following way:

$$Z = Y + S \times LF(D), \tag{26}$$

where D represent the dimensionality of the problem and S denotes a random vector of $1 \times D$ size, and LF represent the levy flight function, determined by:

$$LF(x) = 0.01 \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}}, \sigma = (\frac{\Gamma(1+\beta) \times sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{(\frac{\beta-1}{2})}})^{\frac{1}{\beta}} \tag{27}$$

Therefore, the overall strategy to update the locations of the solutions can be obtained as follows:

$$X(t+1) = \begin{cases} Y, \text{ if } F(Y) < F(X(t)) \\ Z, \text{ if } F(Z) < F(X(t)), \end{cases} \tag{28}$$

where Y and Z are obtained by employing the Equations (25) and (26).

Lastly, if $r \leq 0.5$ and $|E| < 0.5$ (denoted as hard besiege with progressive rapid dives strategy), the prey is entirely exhausted, and hawks start with a hard attack prior to the final catch of the prey. The hawks start decreasing their mean distance to the target, modeled by:

$$X(t+1) = \begin{cases} Y, \text{ if } F(Y) < F(X(t)) \\ Z, \text{ if } F(Z) < F(X(t)), \end{cases} \tag{29}$$

where, opposite to Equation (28), Y and Z are calculated as follows:

$$Y = X_{best}(t) - E|JX_{best}(t) - X(t)| \tag{30}$$

$$Z = Y + S \times LF(D) \tag{31}$$

3.3.8. Sine Cosine Algorithm

The sine cosine algorithm (SCA) was introduced by Mirjalili in 2016 [40], and it draws inspiration from the mathematical behavior of the fundamental trigonometric functions. The locations of the solutions are being updated according to the sine and cosine functions' fluctuations, oscillating in the proximity of the best individual. Similar to other population-based methods, the algorithm begins by producing a collection of arbitrary candidate solutions inside the limits of the search phase. The exploration and exploitation processes are steered during the run by four random modifiable variables. The SCA search equations are defined by Equation (32):

$$X_i^{t+1} = \begin{cases} X_i^{t+1} = X_i^t + r_1 \cdot sin(r_2) \cdot |r_3 \cdot P_i^{*t} - X_i^t|, & r_4 < 0.5 \\ X_i^{t+1} = X_i^t + r_1 \cdot cos(r_2) \cdot |r_3 \cdot P_i^{*t} - X_i^t|, & r_4 \geq 0.5, \end{cases} \tag{32}$$

where X_i^t and X_i^{t+1} represent the observed solution's location over the ith dimension at successive rounds t and $i+1$, respectively, r_{1-4} are values produced in a pseudorandom manner, the P_i^* determines the target point's location (the best currently available approximation of the optimal value) in the ith dimension. It should be stated that, for each part of every individual solution belonging to the population, the new pseudorandom numbers r_{1-4} are produced. For more details about the entire algorithm, please refer to the original publication [40].

4. Results

4.1. Metrics

The simulation outcomes of each XGBoost model were evaluated by mean squared error (MSE) calculated as Equation (33), root mean squared error (RMSE) that can be

obtained by Equation (34), mean absolute error (MAE) determined by Equation (36), and the coefficient of determination (R2) given by Equation (36).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{33}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{34}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{35}$$

$$R2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{36}$$

where y_i and $\hat{y}_i$ denote the vectors of the observed values that are predicted and predicted values with size N, respectively. In this work, MSE was observed as the objective function that needs to be minimized.

4.2. Experimental Setup

The eight described metaheuristics algorithms were utilized to tune the XGBoost model for the observed dataset. The XGBoost hyperparameters that were tuned, together with their respective bounds and types are listed below:

- learning rate (η), range: $[0.1, 0.9]$, continuous parameter,
- min_child_weight , range: $[0, 10]$, continuous parameter,
- subsample, range: $[0.01, 1]$, continuous parameter,
- collsample_bytree, range: $[0.01, 1]$, continuous parameter,
- max_depth, range: $[3, 10]$, integer parameter and
- gamma, range: $[0, 0.8]$, continuous parameter.

The count of parameters for softprob objective function ('num_class':self.no_classes) was also provided as the parameter to the XGBoost model. The rest of the XGBoost parameters were fixed in the experiments to the default XGBoost values.

The proposed approach was developed in Python programming language, and the common set of machine learning libraries was employed, including scikit-learn, scipy, numpy, and pandas. The XGBoost model was obtained from the scikit-learn package.

In the proposed implementation, the standard solutions' encoding scheme has been employed as follows. Each metaheuristics solution has been constituted as a vector of size l, where l represents the count of tuned hyperparameters. Consequently, the l for the XGBoost solutions' was set to six.

All metaheuristics algorithms were implemented independently by the authors, and tested with 40 individuals in the population and 20 iterations per run, over the course of 15 independent runs. As stated before, MSE was set as the objective function that is required to be minimized.

4.3. Experimental Outcomes

This section presents the results obtained through conducted simulations. Tables 1 and 2 contain the results for the objective function and detailed metrics of the best individual runs, and the best outcomes are noted in bold.

Table 1 presents detailed comparative metrics for the objective function (MSE) achieved by XGBoost models tuned by the eight observed algorithms. It can be noted that the FA algorithm obtained the best results for all performance indicators (best, worst, mean, median, standard deviation, and variance). The second-best performing algorithm in terms of the best run metric was HHO, while SCA obtained the second-best results for worst, mean, and

median indicators. Additionally, baseline XGBoost with default hyperparameters was also evaluated, and it obtained r2 of 0.8762324 with 1.3572 MSE, which is significantly worse than the models' performance generated by metaheuristics.

Table 1. Comparative results of the objective function (MSE) of the observed metaheuristics.

Method	GA	PSO	ABC	FA	BA	WOA	HHO	SCA
Best	0.977957	1.028479	0.964857	**0.933440**	1.012142	1.062729	0.951989	0.980756
Worst	1.160223	1.138054	1.191736	**0.981954**	1.114687	1.149946	1.116308	1.034898
Mean	1.050802	1.081808	1.094197	**0.953529**	1.056144	1.108541	1.033249	1.010514
Median	1.040954	1.083018	1.113793	**0.949829**	1.055508	1.111382	1.032899	1.014567
Std	0.042178	0.030035	0.065893	**0.014926**	0.025749	0.023924	0.039545	0.017681
Var	0.001779	0.000902	0.004342	**0.000223**	0.000663	0.000572	0.001564	0.000313

Table 2 shows the detailed metrics of the best individual runs of each observed algorithm. Again, it can be noted that the FA outperformed other metaheuristics for all indicators—R2, R, MSE, MAE, and RMSE. In terms of MSE, which was used as the objective function to be minimized, FA was superior with the result of 0.933440, followed by the HHO, which achieved 0.951989, ABC that scored 0.964857, and SCA that finished fourth with the result of 0.980756.

Table 2. Detailed metrics for the best individual run of the observed metaheuristics.

Method	R2	R	MSE	MAE	RMSE
GA	0.909619	0.953740	0.977957	0.526498	0.988917
PSO	0.904950	0.951289	1.028479	0.514747	1.014139
ABC	0.910830	0.954374	0.964857	0.515527	0.982272
FA	**0.913734**	**0.955894**	**0.933440**	**0.506238**	**0.966147**
BA	0.906460	0.952082	1.012142	0.509854	1.006053
WOA	0.901785	0.949624	1.062729	0.561811	1.030887
HHO	0.912019	0.954997	0.951989	0.514525	0.975699
SCA	0.909361	0.953604	0.980756	0.533190	0.990331

The collection of the XGBoost hyperparameters determined by the best runs of each metaheuristics are shown in Table 3. The best performing algorithm in this scenario was FA, that determined the XGBoost model with a learning rate of 0.338502, max_child_weight of 2.465529, a subsample of 0.895580, collsample_bytree of 1.000000, max_depth of 9, and gamma value of 0.562947. The second best performing method was HHO, that obtained the XGBoost model with a learning rate of 0.365808, max_child_weight of 7.374436, a subsample of 0.943227, collsample_bytree of 0.994926, max_depth of 10, and gamma value of 0.420847.

Table 3. Best solutions' determined XGBoost hyperparameters set.

Method	l.r. (μ)	Max_CHILD_WEIGHT	Subsample	Collsample_BYTREE	Max_DEPTH	Gamma
GA	0.393890	4.403968	0.863233	1.000000	10	0.584005
PSO	0.342858	3.883601	1.000000	1.000000	9	0.274638
ABC	0.344956	5.671974	0.787743	0.969200	10	0.195545
FA	0.338502	2.465529	0.895580	1.000000	9	0.562947
BA	0.313576	3.183511	1.000000	1.000000	10	0.000000
WOA	0.330494	4.867128	0.898843	0.956548	7	0.261317
HHO	0.365808	7.374436	0.943227	0.994926	10	0.521701
SCA	0.452871	2.288884	0.892422	1.000000	9	0.420847

The visualization of the executed experiments is given in Figure 1, presenting the objective convergence graphs, box plots, and violin plots of all eight observed methods

for both objective function and R2. Figure 2 first presents the swarm plots for both the objective function and R2, showing the diversity of the population in the last iteration of the best run of each algorithm. Additionally, join plots of both objective function and R2 with histograms for the two best algorithms are also shown in Figure 2.

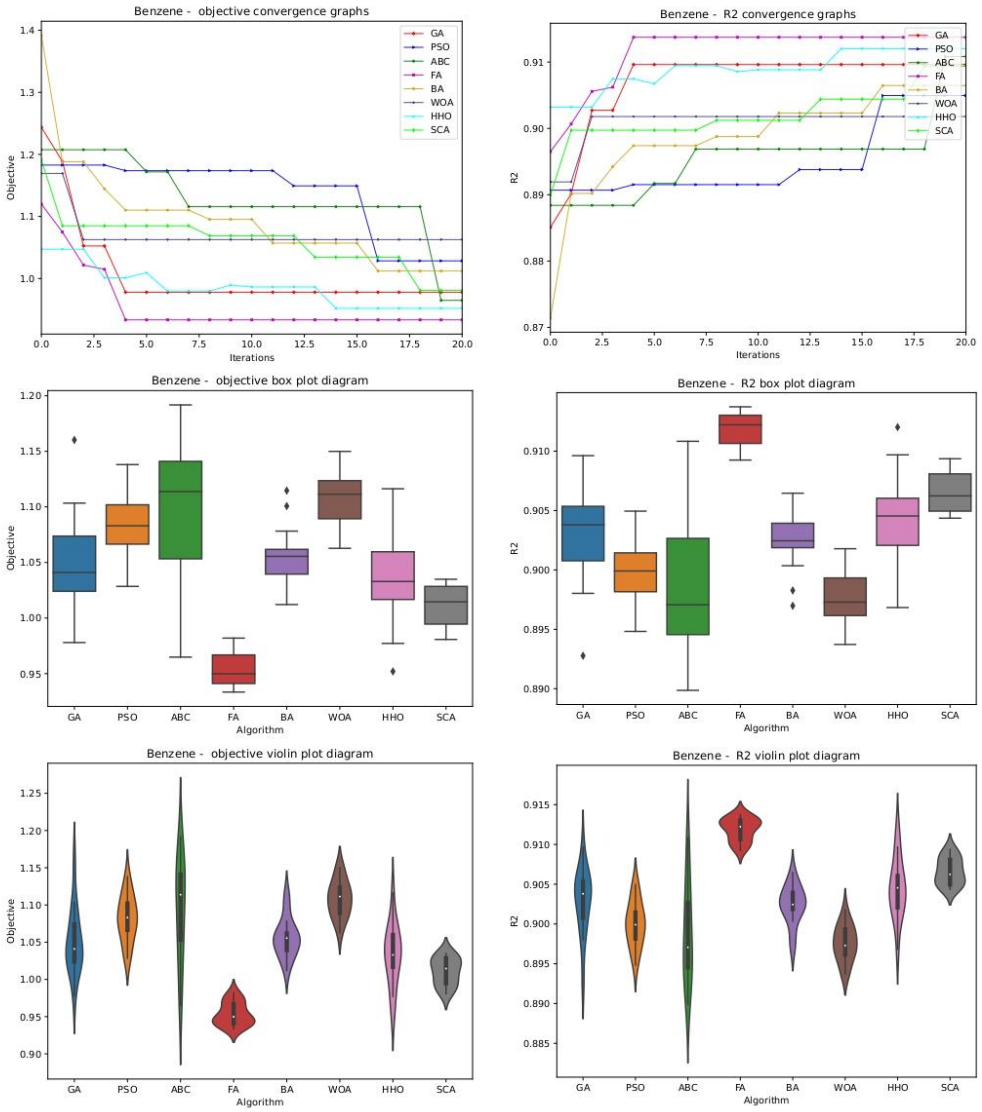


Figure 1. Visualisations of the conducted simulations for all algorithms in terms of convergence, box plot, and violin diagrams for both objective function and R2.

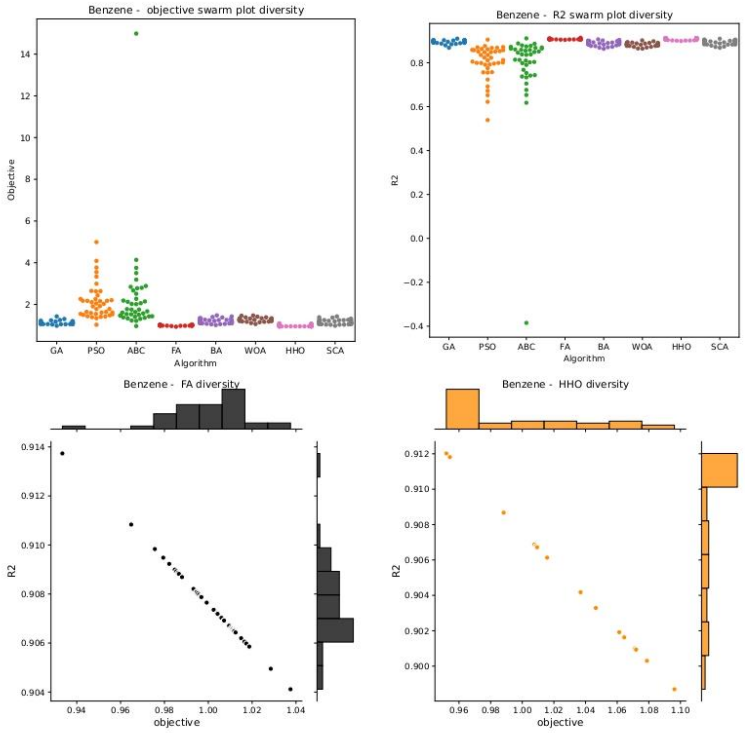


Figure 2. Visualisations of the conducted simulations for all algorithms in terms of swarm plot diagrams for both objective function and R2, and join plots with histograms of two best methods (FA and HHO).

While observing Figures 1 and 2, it can be concluded that FA algorithm exhibits the fastest converging speed, followed by the HHO algorithm. FA also achieves the most stable results, followed by the HHO and GA, as can be noted from the box plot and swarm plot diversity diagrams.

Finally, the visualization of the best-predicted results achieved by the best-produced model by each one of the eight observed algorithms is shown in Figure 3. Again, it is possible to note that the model tuned by the FA algorithm produced the best forecasts of the observed time series.

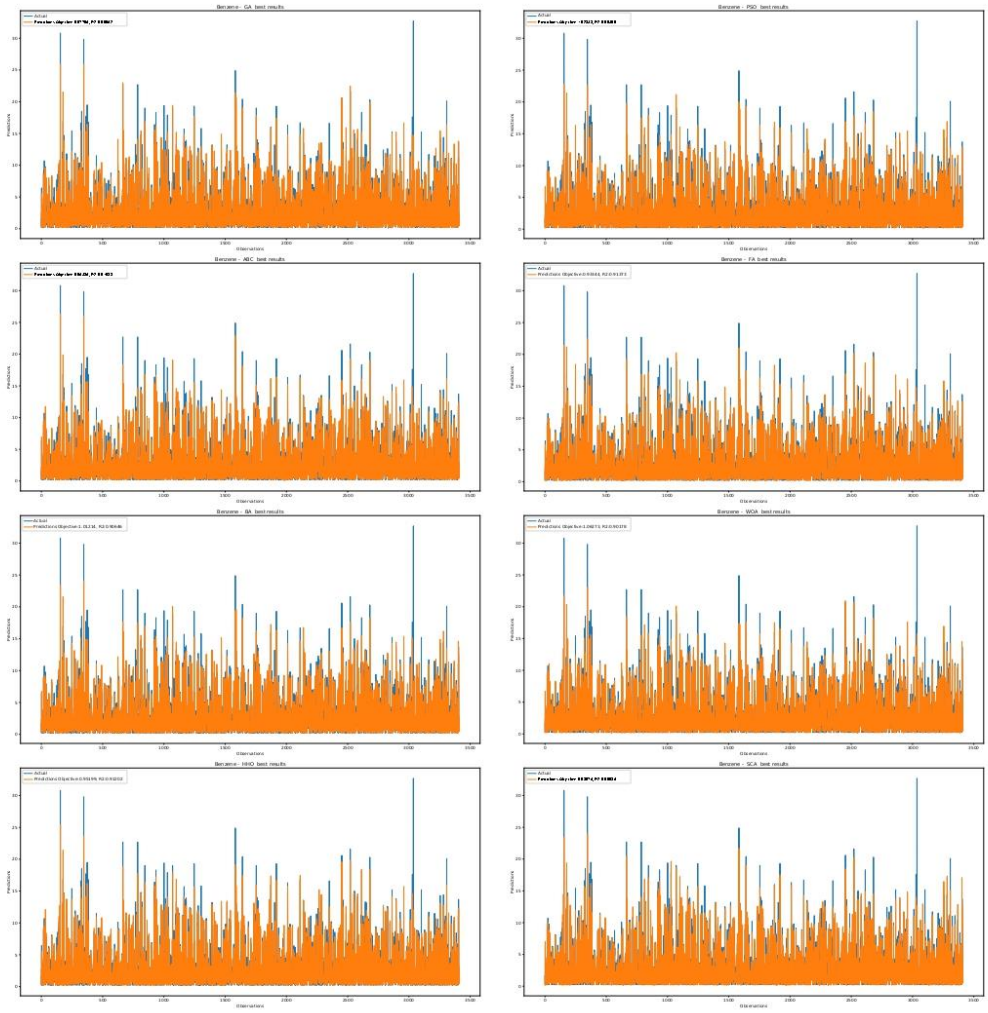


Figure 3. Best predictions by XGBoost model tuned by each observed algorithm.

5. Discussion

Descriptive statistics for the observed pollutant concentrations is provided in Table 4, while the obtained SHAP importance is provided in Table 5. In addition to their use for air quality forecasting [19,85], machine learning models, when supplemented by explainable and model interpretability analyses, provide insight into the significance and impact of considered prediction variables. By applying deep SHAP analysis to NO_2 LSTM forecasting, García and Aznarte [18] registered a significant influence of meteorological parameters on the modeled pollutant concentrations, which has been shown to be in compliance with the well-known natural phenomena in the investigated area. Further, Kang et al. [86] have used SHAP analysis to investigate the seasonal impacts of meteorological factors on the spatiotemporal prediction of NO_2 and O_3 levels. In this study, we have made a step

forward by analyzing the air pollutant behavior in a certain type of environment affected by different meteorological conditions and the presence of other polluting species.

To demonstrate the potential of the applied methodology, we provide details on the three most important predictors that describe the evolution of benzene concentration. The interrelations with toluene and finest aerosol fraction dominantly shape benzene concentrations, while other important variables include meteorological parameters temperature (T02M and TMPS), volumetric soil moisture content (SOLM), and momentum flux direction (MOFD), as well as concentrations of total nonmethane hydrocarbons (TNMHC) and total nitrogen oxides (NO_x).

Table 4. Descriptive statistics for the observed pollutant concentrations [μg m^{-3}].

Compound	Mean	Min	Max	Median	25th Percentile	75th Percentile
Benzene	2.82	0.25	54.70	1.74	0.67	3.64
Toluene	3.00	0.25	60.60	1.76	1.03	3.77
m, p-Xylene	5.74	0.25	82.80	3.71	1.68	8.94
TNMHC	45.79	6.26	806.00	33.20	24.20	51.50
NO	20.05	0.25	337.00	6.75	3.49	12.40
NO_2	16.81	0.55	129.00	12.90	8.58	21.40
NO_x	47.54	2.28	589.00	24.50	16.20	43.70
PM_1	21.66	1.00	283.00	12.40	6.79	24.50
$PM_{2.5}$	23.89	1.00	290.00	14.40	8.28	26.73
PM_{10}	28.89	1.00	318.00	19.20	11.60	32.33

Table 5. Variable SHAP importance.

Parameter	SHAP	Absolute SHAP	Relative SHAP [%]
Toluene	−0.087	1.453	35.03
PM_1	0.098	0.593	16.22
T02M	−0.027	0.538	15.82
SOLM	−0.021	0.175	4.95
TNMHC	−0.040	0.114	3.33
MOFD	0.002	0.078	2.57
TMPS	−0.011	0.072	1.98
NO_x	0.036	0.064	1.85
LIB4	0.007	0.052	1.66
m, p-Xylene	−0.001	0.051	1.56
NO	−0.006	0.051	1.75
WD	−0.001	0.041	1.14
PRSS	−0.006	0.029	0.85
MSLP	−0.006	0.029	0.93
RH2M	0.004	0.027	0.78
SHTF	−0.008	0.027	0.75
NO_2	0.008	0.025	0.69
$PM_{2.5}$	0.007	0.024	0.67
O_3	0.001	0.023	0.62
LHTF	−0.005	0.021	0.70
CPP6	−0.001	0.021	0.68
PM_{10}	−0.003	0.020	0.51
DSWF	−0.001	0.020	0.62
PBLH	−0.006	0.018	0.54
MOFI	−0.001	0.018	0.49
TCLD	−0.004	0.018	0.46
LISD	0.004	0.017	0.45

Table 5. *Cont.*

Parameter	SHAP	Absolute SHAP	Relative SHAP [%]
WS	0.005	0.016	0.45
CINH	0.002	0.013	0.34
LCLD	0.000	0.010	0.29
HCLD	−0.002	0.010	0.30
MCLD	0.000	0.009	0.30
TPP6	−0.002	0.008	0.25
SHGT	0.000	0.005	0.16
CAPE	0.001	0.005	0.18
CRAI	−0.001	0.004	0.12
CSNO	0.0001	0.0003	0.01
CFZR	0	0	0
CICE	0	0	0

5.1. Toluene

Although benzene and toluene most often appear as copollutants sharing the same sources (traffic, petrochemical industry, commercial product manufacturing, etc.), their reactivity and atmospheric half-life differ. Namely, toluene contains an electron-releasing methyl group attached to a benzene ring, which makes it more reactive and results in different environmental behavior of toluene and benzene, which can be indicative when distinguishing between emission sources.

The results have shown that benzene and toluene are strongly interrelated, with toluene levels affecting an average of 35% of the benzene concentrations and the decrease of 1.4 μg m^{-3} relative to the expected levels (Table 5). According to the findings, three types of environmental conditions that shape the benzene behavior depending on the toluene concentration range can be identified.

The first environment is characterized by toluene concentrations below 2 μg m^{-3} (Figure 4), as well as low concentrations of all pollutants including aerosol fractions, m, p-xylenes, and NO, except ozone, with the registered values exceeding 70 μg m^{-3}. Within these conditions, the considerable variability of all analyzed meteorological parameters and diverse weather conditions were observed, which indicates consistent emission sources. Additionally, the obtained organic/inorganic gas relations and a toluene-to-benzene ratio above 2 (Figure 4), suggests that the majority of low benzene concentrations assigned to this environment originate from evaporation processes related to petrochemical refinery sources in the southern zone of Pančevo, including equipment leaks from valves or steam power units and leakage during the transport.

The second environment refers to toluene concentrations ranging from 2 to 4.5 μg m^{-3} with no registered impact on benzene levels and SHAP values around zero. The belonging concentrations of NO were below 30 μg m^{-3} and above 100 μg m^{-3}, NO_x levels were below 100 μg m^{-3} and above 170 μg m^{-3}, while PM_1 and TNMHC were below 64 μg m^{-3} and 110 μg m^{-3}, respectively. The observed extreme concentrations of nitrogen oxides, as well as relatively high concentrations of fine particles and lacking relationship between toluene and benzene all point to the strong and intermittent emission source, which can be attributed to HIP Azotara Pančevo, one of the most important regional plant of mineral fertilizers and nitrogen compounds, and agricultural practices on the surrounding farming areas related to heavy use of fertilizers and livestock waste.

Toluene concentrations ranging from 4.5 to 15 μg m^{-3} describe the third environment, which appears to be favorable for elevating benzene concentrations (Figure 4). Within the defined conditions, two subenvironments related to separate emission sources can be distinguished based on the benzene-to-toluene ratio.

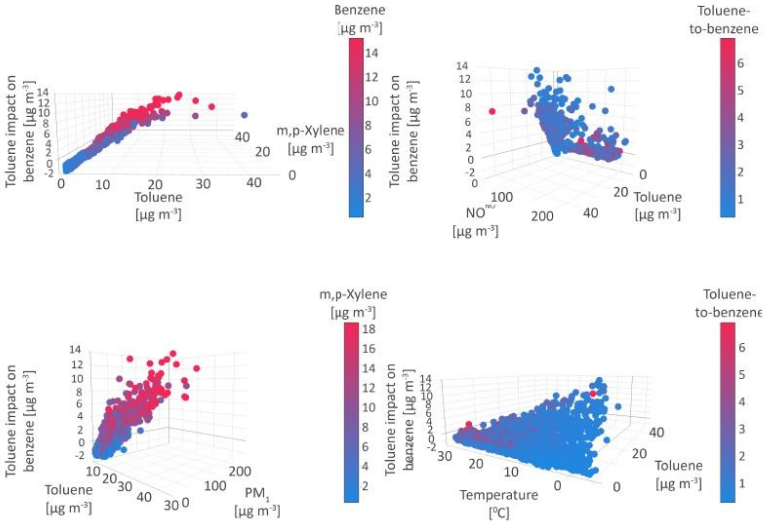


Figure 4. Toluene impact on benzene.

The first subenvironment is characterized as highly affecting conditions responsible for driving the observed benzene concentrations up to 14 µg m^{-3} above the expected value. Thereby, the role of toluene seems to be of particular significance, since its relative impact on benzene concentrations increased to 55%. The high levels of benzene were accompanied by high PM, NO_2, and TNMHC concentrations, as well as ozone levels below 70 µg m^{-3}. Regarding meteorological parameters, the medium- and the highly-supporting environment was related to air temperatures below 15 °C, low wind speeds (below 2 m s^{-1}), low planetary boundary layer height of 200 m, and high air humidity above 80%, all of which can be assigned to a cold period of the year. During autumn and winter months, these unfavorable meteorological conditions contribute to high concentrations of pollutants originating from fossil fuel burning for heating purposes. Additionally, reactions with photochemically produced hydroxyl radicals, which represent the principal mechanism of vapor-phase toluene and benzene atmospheric removal, are suppressed during the cold months, which results in the prolongation of pollutants' lifetimes from a few days in the summer season to several weeks in autumn and winter.

The second sub-environment refers to conditions with a lower impact on benzene concentrations (up to 8 µg m^{-3}) that govern high levels of benzene (>15 µg m^{-3}), m, p-xylene (>14 µg m^{-3}), and NO_2 (>40 µg m^{-3}), and above average TNMHC levels (t). Concurrently medium to high air and soil temperatures, air pressure, and humidity, as well as a toluene-to-benzene ratio in the range between 1 and 2, can be attributed to the site-specific and year-round continuous contribution of traffic (Figure 4) [87].

5.2. Particulate Matter (PM_1)

The aerosol fraction PM_1 and benzene are interrelated in four environments, which define an average of 16.2% observed concentrations and lead to the average increase of about 0.6 µg m^{-3} (Table 5) in benzene levels. In the first case, the benzene concentrations exhibit a decrease by 1 µg m^{-3}, in two of the identified environments benzene levels increase by 2.5 µg m^{-3}, while the interrelationship with PM_1 in the last case does not seem to affect its levels (Figure 5).

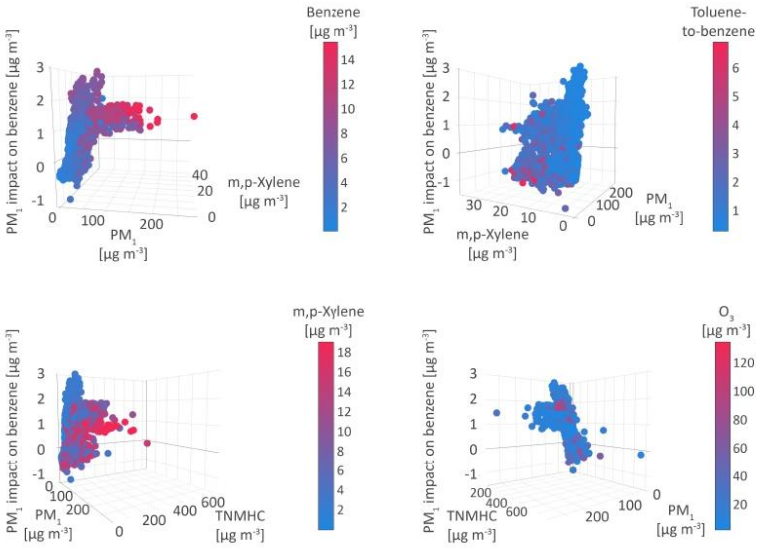


Figure 5. PM_1 impact on benzene.

In the first environment, the reduction of benzene concentrations by 1 µg m^{-3} is affected by a decrease in PM_1 concentrations to 20 µg m^{-3} and complemented by low concentrations of atmospheric aerosols $PM_{2.5}$ and PM_{10} (up to 30 µg m^{-3}) and high concentrations of O_3 (above 70 µg m^{-3}). The concentrations of toluene, TNMHC, and m, p-xylenes are observed in a wide range of values, which disables drawing conclusions on the relationship between particles and VOCs, however, the calculated toluene-to-benzene ratio above 2 suggests the dominant impact of the industrial evaporation processes in this environment [88]. The observed pollutant levels have occurred under low atmospheric pressure, and medium or higher temperatures, planetary boundary layer height, humidity and momentum flux intensity, i.e., the atmospheric conditions that enable vertical mixing, pollutant dispersion and transport. Previous research has shown that, despite the constant pollutant emissions, PM levels could fluctuate up to several times with the change in influential weather variables [89,90]. Some studies have reported elevated PM concentrations under calm weather, mild wind and low planetary boundary layer height, temperature and relative humidity, while findings associating high PM levels with high wind speed and low humidity, or increased precipitations are also available. These contrasting results can be explained by the fact that aerosol water largely impacts complex heterogeneous gas/liquid/solid partitioning of freshly emitted particles and precursor gases.

Medium concentrations of benzene registered in the second environment were independent (SHAP values are zero) of PM_1 levels ranging from 20 to 30 µg m^{-3}, and accompanied by moderate levels of $PM_{2.5}$, PM_{10}, and m, p-xylenes, and higher concentrations of toluene and NO_2. The prevailing conditions can be described by the average air humidity of 50%, air pressure of 1000 mbar, boundary layer height ranging from 300 to 500 m, and temperatures from 10 to 20 °C.

In the highly affecting environment, the increase of benzene concentrations by 2.5 µg m^{-3} was driven by the PM_1 levels ranging from 30 to 92 µg m^{-3}, under the impact of temperatures below 10 °C, planetary boundary layer heights below 400 m, and medium or higher air pressure and humidity. These meteorological conditions and the toluene-to-benzene ratio below 1 correspond to the cold part of the year when the burning of fossil fuels can be considered the major cause of low air quality.

An additional interrelation pushes benzene levels by 1.5 μg m^{-3} with the PM_1 concentrations exceeding 60 μg m^{-3}, in the fourth environment, defined by higher concentrations of TNMHC, m, p-xylenes, all fractions of atmospheric aerosols, NO and NO_2, and low O_3 levels. The atmospheric conditions, which can be attributed to the cold part of the year, including low wind speed, temperature, and planetary boundary layer height (up to 2 m s^{-1}, 10 °C and 400 m, respectively), medium and high air and soil humidity, and high air pressure, created the unfavorable environment for the production of secondary pollutants, which explains the lower concentrations of O_3, while slightly higher toluene-to-benzene ratio values equal or above 1 indicates the contribution of traffic emissions to high pollutant concentrations during autumn and winter season [91].

5.3. Temperature

Temperature is recognized as the third important parameter that shapes 15.8% of benzene levels, lowering its concentrations by about 0.5 μg m^{-3} on average. Its impact is complex but relatively symmetrical and monotonically decreasing with increasing temperature, with a pronounced positive effect at temperatures lower than 9 °C and a negative effect at temperatures higher than 14 °C (positive/negative effect refers to an increase/decrease of benzene concentrations), Figure 6.

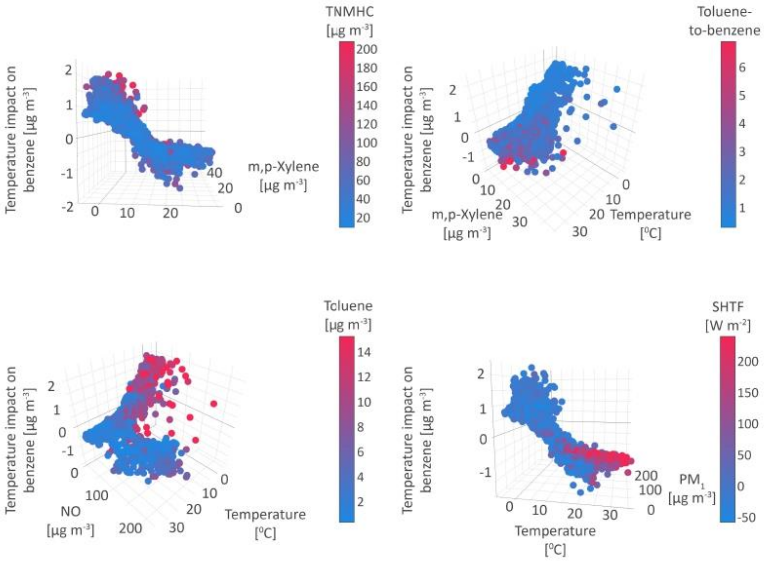


Figure 6. Temperature impact on benzene.

Within the range of lower temperatures, two environments have been identified. In the first case, the benzene concentration increase up to 1 μg m^{-3} is followed by low to medium concentrations of toluene (2 μg m^{-3} on average), m, p-xylenes (5 μg m^{-3} on average), nitrogen oxides (25 μg m^{-3} on average) and all atmospheric aerosol fractions, low concentrations of O_3, as well as high cloudiness and intensity of momentum flux, low insolation and air pressure, and a very low planetary boundary layer height of 350 m.

In the second environment, which can be assigned to fossil fuel burning for heating purposes during the cold season, benzene levels increase almost linearly with the drop in temperature by 2.3 μg m^{-3} on average. The increase is followed by high concentrations of aerosols, nitrogen oxides, and benzene (above 12 μg m^{-3} for the latter), extremely low concentrations of O_3 (below 20 μg m^{-3}), stable atmospheric conditions, low temperature, and planetary boundary layer, but high pressure and air humidity.

As shown in Figure 6, interrelations between benzene and temperature, which ranged from 9 to 14 °C, were apportioned in three subenvironments. In the first case, a positive impact of temperature was accompanied by an increase in benzene concentrations and lower levels of toluene (2.5 μg m^{-3} on average), medium levels of m, p-xylenes (5 μg m^{-3} on average) and NO_x (38 μg m^{-3} on average), and medium boundary layer heights. Contrary, in the second case, temperature impact reduces to a minimum while the environment is characterized by an increase in the pollutant concentrations, approximately 25%, 35%, and 52% for VOCs, NO_x, and all particle fractions, respectively. Under these conditions, the elevation of SHTF for 680%, downward short-wave radiation flux (DSWF) for 54%, and LHTF for 26% was observed, while soil moisture (SOLM) and low cloud cover (LCLD) declined for 7% and 28%, respectively. Ambiances with higher PM concentrations and gaseous pollutants have been associated with ambivalent impacts on the surface temperature while an increase in water vapor induces a rise in the shortwave cloud radiative forcing [92]. In addition, SOA formation from precursors such as particles, NO_x and dominantly the benzene homologs, toluene, and xylene, is enhanced in the presence of water vapor with NO_x being the most soluble species. Prevalent temperature values in described subarea are not sufficient to render photolysis of VOCs. In the third subarea, when temperature impact is negative, the ambiance is shaped by a decrease of benzene and even higher levels of pollutants (more than 50, 40, and 80% for VOCs, NO_x, and PM, respectively). Much of the anticipated energy balance concept in the land–atmosphere interactions rely on soil moisture as a key variable. The content of soil moisture depends on atmospheric conditions such as precipitation, radiation, and evaporation, which further alters surface turbulent and radiative heat fluxes. Some studies witnessed that low precipitation suppressed the availability of soil moisture causing a decrease in latent heating (LHTF) and an elevation of sensible heating at the surface (SHTF). These conditions, accompanied by increased temperature, affect atmospheric thermodynamics and the structure of PBL and make the atmosphere less suitable to maintain deep convection. Causality in the coupled land/surface–atmosphere system becomes more complicated in the presence of atmospheric pollutants and variations of local meteorological conditions. For example, suspended particles scatter shortwave radiation and trap longwave radiation to a different extent, which could modify surface temperature and heat fluxes [92].

In the environment with temperatures above 14 °C, benzene concentrations, lower up to 1.5 μg m^{-3} on average, are accompanied by higher concentrations of PM_1, NO, m, p-xylenes and TNMHC, as well as by low humidity, higher air and soil surface temperatures, and planetary boundary layer heights above 1200 m. In addition to this, the toluene-to-benzene ratio over 3, reflects the industrial activities at the regional chemical plant Azotara Pančevo, which manufactures nitrogen chemicals and mineral fertilizers, but also soil preparation, maturing, and other farm production processes in the surrounding agricultural land during the spring and summer season. Besides, meteorological conditions in the warm season are favorable for benzene removal. In the troposphere, photolysis transforms VOCs or they react with OH and NO_3 radicals and O_3. Concerning the reaction rates with the OH radical, which is a dominant loss of most VOCs, lifetimes of benzene, toluene, and xylenes in the air are up to 10 days, 2 days, and 6 hours, respectively, while their presence is up to several years regarding loss by NO_3 radical and O_3. In the presence of sunlight, the reaction between VOCs and •OH yields a peroxy ($•HO_2$) and an alkyl or substituted alkyl radical ($•ROO_2$). Such produced radicals further react with NO, converting it to NO_2, which photolysis forms O_3. As evident from the presented simplified chemical mechanisms, a photo equilibrium between NO, NO_2, O_3, and kinetic reactivity of radicals and VOCs lead to no net formation or loss of O_3. The VOC/NO_x, ratio impacts the production of O_3 as follows: (i) the occurrence of NO_x, sinks (NO_x-limited conditions) lowers the amount of formed O_3, and (ii) during VOC-limited conditions, a net formation or loss of OH radicals leads to an intensification or reduction of overall reactivity of all presented VOCs.

6. Conclusions

The air pollutant level dynamics in space and time is extremely intense, and thus air pollution represents a phenomenon of regional and global significance and a real challenge for scientific exploration. The behavior of polluting species in the atmosphere depends on the strength and other characteristics of emission sources, as well as on the atmospheric environment, which provides the conditions for their dispersion, transformation, and removal. Building on our previous experience in modeling environmental pollutant level dynamics, we have employed, coupled, and optimized advanced artificial intelligence modeling tools to investigate and quantify the interrelations between air pollutants and meteorological parameters, and to capture the types of environmental conditions under which air pollutant behavior exhibits certain regularities and consistency. For this study, we have used two-year air quality monitoring and meteorological data to portray types of the environment that shapes the observed benzene levels in the urban area surrounded by large industrial complexes, including plants for the manufacturing of artificial fertilizers, chemical industry, and oil refinery. Among eight algorithms that have been previously confirmed as well-known optimizers by successfully solving nondeterministic polynomial-hard challenges in the past, the firefly algorithm has been shown to achieve a superior level of performance and provide the best results for all performance indicators. As shown, the interrelations between benzene, and toluene and finest aerosol fraction are the most prominent, while the factors which shape the supportive environment for these relationships to take place include meteorological parameters temperature, volumetric soil moisture content, and momentum flux direction, as well as concentrations of total nonmethane hydrocarbons and total nitrogen oxides. Toluene, temperature, and the finest aerosol fraction were recognized to affect 35%, 16.2%, and 15.8% of benzene concentrations, respectively. The first two played role in decreasing benzene levels, while the aerosol–benzene interrelation led to an increase of benzene concentrations of about 0.6 $\mu g\ m^{-3}$. Since there is large space for improvements, considering the fact that the ML models have not been thoroughly tested for environmental datasets (challenges), as part of future research, it was planned to validate other ML models (baseline and tuned by metaheuristics) against the one used in this research and other environmental datasets.

Author Contributions: Conceptualization, A.S. and N.B.; methodology, A.S. and N.B.; software, L.J., N.B. and M.Z.; validation, G.J., M.P., N.B. and A.S.; investigation, N.B., L.J., M.Z., M.P., G.J., F.A., S.S. and A.S.; data curation, M.P.; writing—original draft preparation, N.B., M.P., G.J., F.A., S.S. and A.S.; writing—review and editing, A.S. and N.B.; visualization, A.S. and N.B.; supervision, A.S. and N.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia GRANT No. #6524105, AI—ATLAS.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BTEX	benzene, toluene, ethylbenzene, and xylene
SHAP	SHapley Additive exPlanations
SOA	secondary organic aerosols
TNMHC	total non-methane hydrocarbons
VOCs	volatile organic compounds
XGBoost	eXtreme Gradient Boosting

Meteorological paramater	Unit	Label
Pressure at surface	hPa	PRSS
Pressure reduced to mean sea level	hPa	MSLP
Accumulated precipitation (6 h accumulation)	m	TPP6
Momentum flux intensity (3- or 6-h average)	N m^{-2}	MOFI
Momentum flux direction (3- or 6-h average)	°	MOFD
Sensible heat net flux at surface (3- or 6-h average)	W m^{-2}	SHTF
Downward short wave radiation flux (3- or 6-h average)	W m^{-2}	DSWF
Relative Humidity at 2 m AGL	%	RH2M
Wind speed at 10 m AGL	m s^{-1}	WS
Wind direction at 10 m AGL	°	WD
Temperature at 2 m AGL	°C	TO2M
Total cloud cover (3- or 6-h average)	%	TCLD
Geopotential height	gpm *	SHGT
Convective available potential energy	J Kg^{-1}	CAPE
Convective inhibition	J Kg^{-1}	CINH
Standard lifted index	°C	LISD
Best 4-layer lifted index	°C	LIB4
Planetary boundary layer height	m	PBLH
Temperature at surface	°C	TMPS
Accumulated convective precipitation (6 h accumulation)	m	CPP6 **
Volumetric soil moisture content	frac.	SOLM
Categorial snow (yes = 1, no = 0) (3- or 6-h average)		CSNO
Categorial ice (yes = 1, no = 0) (3- or 6-h average)		CICE
Categorial freezing rain (yes = 1, no = 0) (3- or 6-h average)		CFZR
Categorial rain (yes = 1, no = 0) (3- or 6-h average)		CRAI
Latent heat net flux at surface (3- or 6-h average)	W/m^2	LHTF
Low cloud cover (3- or 6-h average)	%	LCLD
Middle cloud cover (3- or 6-h average)	%	MCLD
High cloud cover (3- or 6-h average)	%	HCLD

* geopotential meters
** Beginning with 00 UTC 15 July 2019, CPPA (total accumulation) instead of CPP6 (6-hour accumulation)

References

1. Faber, P.; Drewnick, F.; Borrmann, S. Aerosol particle and trace gas emissions from earthworks, road construction, and asphalt paving in Germany: Emission factors and influence on local air quality. *Atmos. Environ.* **2015**, *122*, 662–671. [CrossRef]
2. WHO. *Global Air Quality Guidelines Aim to Save Millions of Lives from Air Pollution*; WHO: Geneva, Switzerland, 2021. Available online: https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution (accessed on 1 December 2022).
3. UN. Global Perspective Human Stories. 2022. Available online: https://news.un.org/en/story/2022/04/1115492 (accessed on 1 November 2022).
4. Begou, P.; Kassomenos, P. One-year measurements of toxic benzene concentrations in the ambient air of Greece: An estimation of public health risk. *Atmos. Pollut. Res.* **2020**, *11*, 1829–1838. [CrossRef]
5. Sekar, A.; Varghese, G.K.; Varma, M.R. Analysis of benzene air quality standards, monitoring methods and concentrations in indoor and outdoor environment. *Heliyon* **2019**, *5*, e02918. [CrossRef] [PubMed]
6. Ji, Y.; Gao, F.; Wu, Z.; Li, L.; Li, D.; Zhang, H.; Zhang, Y.; Gao, J.; Bai, Y.; Li, H. A review of atmospheric benzene homologues in China: Characterization, health risk assessment, source identification and countermeasures. *J. Environ. Sci.* **2020**, *95*, 225–239. [CrossRef]
7. Cheng, X.; Chen, Q.; Jie Li, Y.; Zheng, Y.; Liao, K.; Huang, G. Highly oxygenated organic molecules produced by the oxidation of benzene and toluene in a wide range of OH exposure and NO x conditions. *Atmos. Chem. Phys.* **2021**, *21*, 12005–12019. [CrossRef]
8. Deng, Y.; Li, J.; Li, Y.; Wu, R.; Xie, S. Characteristics of volatile organic compounds, NO_2, and effects on ozone formation at a site with high ozone level in Chengdu. *J. Environ. Sci.* **2019**, *75*, 334–345. [CrossRef] [PubMed]

9. National Research Council. *Rethinking the Ozone Problem in Urban and Regional Air Pollution*; National Academies Press: Washington, DC, USA, 1992.
10. Li, J.; Deng, S.; Li, G.; Lu, Z.; Song, H.; Gao, J.; Sun, Z.; Xu, K. VOCs characteristics and their ozone and SOA formation potentials in autumn and winter at Weinan, China. *Environ. Res.* **2022**, *203*, 111821. [CrossRef]
11. Zhan, J.; Feng, Z.; Liu, P.; He, X.; He, Z.; Chen, T.; Wang, Y.; He, H.; Mu, Y.; Liu, Y. Ozone and SOA formation potential based on photochemical loss of VOCs during the Beijing summer. *Environ. Pollut.* **2021**, *285*, 117444. [CrossRef]
12. Whaley, C.H.; Galarneau, E.; Makar, P.A.; Moran, M.D.; Zhang, J. How much does traffic contribute to benzene and polycyclic aromatic hydrocarbon air pollution? Results from a high-resolution North American air quality model centred on Toronto, Canada. *Atmos. Chem. Phys.* **2020**, *20*, 2911–2925. [CrossRef]
13. Galán-Madruga, D.; García-Cambero, J.P. An optimized approach for estimating benzene in ambient air within an air quality monitoring network. *J. Environ. Sci.* **2022**, *111*, 164–174. [CrossRef]
14. Jephcote, C.; Mah, A. Regional inequalities in benzene exposures across the European petrochemical industry: A Bayesian multilevel modelling approach. *Environ. Int.* **2019**, *132*, 104812. [CrossRef]
15. Stojić, A.; Maletić, D.; Stojić, S.S.; Mijić, Z.; Šoštarić, A. Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. *Sci. Total. Environ.* **2015**, *521*, 19–26. [CrossRef]
16. Perišić, M.; Maletić, D.; Stojić, S.S.; Rajšić, S.; Stojić, A. Forecasting hourly particulate matter concentrations based on the advanced multivariate methods. *Int. J. Environ. Sci. Technol.* **2017**, *14*, 1047–1054. [CrossRef]
17. Stojić, A.; Jovanović, G.; Stanišić, S.; Romanić, S.H.; Šoštarić, A.; Udovičić, V.; Perišić, M.; Milićević, T. The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels. *Chemosphere* **2022**, *289*, 133154. [CrossRef] [PubMed]
18. García, M.V.; Aznarte, J.L. Shapley additive explanations for NO_2 forecasting. *Ecol. Inform.* **2020**, *56*, 101039. [CrossRef]
19. Dai, H.; Huang, G.; Zeng, H.; Zhou, F. PM2. 5 volatility prediction by XGBoost-MLP based on GARCH models. *J. Clean. Prod.* **2022**, *356*, 131898. [CrossRef]
20. Dai, H.; Huang, G.; Zeng, H.; Yu, R. Haze Risk Assessment Based on Improved PCA-MEE and ISPO-LightGBM Model. *Systems* **2022**, *10*, 263. [CrossRef]
21. Belotti, J.T.; Castanho, D.S.; Araujo, L.N.; da Silva, L.V.; Alves, T.A.; Tadano, Y.S.; Stevan, S.L., Jr.; Correa, F.C.; Siqueira, H.V. Air pollution epidemiology: A simplified Generalized Linear Model approach optimized by bio-inspired metaheuristics. *Environ. Res.* **2020**, *191*, 110106. [CrossRef]
22. Yonar, A.; Yonar, H. Modeling air pollution by integrating ANFIS and metaheuristic algorithms. *Model. Earth Syst. Environ.* **2022**, 1–11. [CrossRef]
23. Drewil, G.I.; Al-Bahadili, R.J. Air pollution prediction using LSTM deep learning and metaheuristics algorithms. *Meas. Sens.* **2022**, *24*, 100546. [CrossRef]
24. Stojić, A.; Stanić, N.; Vuković, G.; Stanišić, S.; Perišić, M.; Šoštarić, A.; Lazić, L. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Sci. Total. Environ.* **2019**, *653*, 140–147. [CrossRef] [PubMed]
25. Šoštarić, A.; Stojić, S.S.; Vuković, G.; Mijić, Z.; Stojić, A.; Gržetić, I. Rainwater capacities for BTEX scavenging from ambient air. *Atmos. Environ.* **2017**, *168*, 46–54. [CrossRef]
26. Stanišić, S.; Perišić, M.; Jovanović, G.; Maletić, D.; Vudragović, D.; Vranić, A.; Stojić, A. What Information on Volatile Organic Compounds Can Be Obtained from the Data of a Single Measurement Site Through the Use of Artificial Intelligence? In *Artificial Intelligence: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 207–225.
27. Stojić, A.; Mustać, B.; Jovanović, G.; Đinović Stojanović, J.; Perišić, M.; Stanišić, S.; Herceg Romanić, S. Patterns of PCB-138 Bioaccumulation in Small Pelagic Fish from the Eastern Mediterranean Sea Using Explainable Machine Learning Prediction. In *Artificial Intelligence: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 175–189.
28. Stojić, A.; Vuković, G.; Perišić, M.; Stanišić, S.; Šoštarić, A. Urban air pollution: An insight into its complex aspects. In *A Closer Look at Urban Areas*; Nova Science Publishers: New York, NY, USA, 2018.
29. Stegherr, H.; Heider, M.; Hähner, J. Classifying Metaheuristics: Towards a unified multi-level classification system. *Nat. Comput.* **2020**, *21*, 155–171. [CrossRef]
30. Emmerich, M.; Shir, O.M.; Wang, H. Evolution strategies. In *Handbook of Heuristics*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 89–119.
31. Fausto, F.; Reyna-Orta, A.; Cuevas, E.; Andrade, Á.G.; Perez-Cisneros, M. From ants to whales: Metaheuristics for all tastes. *Artif. Intell. Rev.* **2020**, *53*, 753–810. [CrossRef]
32. Beni, G. Swarm intelligence. In *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 791–818.
33. Abraham, A.; Guo, H.; Liu, H. Swarm intelligence: Foundations, perspectives and applications. In *Swarm Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 3–25.
34. Dorigo, M.; Birattari, M.; Stutzle, T. Ant colony optimization. *IEEE Comput. Intell. Mag.* **2006**, *1*, 28–39. [CrossRef]
35. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948.
36. Karaboga, D.; Basturk, B. On the performance of artificial bee colony (ABC) algorithm. *Appl. Soft Comput.* **2008**, *8*, 687–697. [CrossRef]

37. Yang, X.S. A new metaheuristic bat-inspired algorithm. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 65–74.
38. Yang, X.S.; Gandomi, A.H. Bat algorithm: A novel approach for global engineering optimization. *Eng. Comput.* **2012**, *29*, 464–483. [CrossRef]
39. Yang, X.S.; Slowik, A. Firefly algorithm. In *Swarm Intelligence Algorithms*; CRC Press: Boca Raton, FL, USA, 2020; pp. 163–174.
40. Mirjalili, S. SCA: A sine cosine algorithm for solving optimization problems. *Knowl.-Based Syst.* **2016**, *96*, 120–133. [CrossRef]
41. Abualigah, L.; Diabat, A.; Mirjalili, S.; Abd Elaziz, M.; Gandomi, A.H. The arithmetic optimization algorithm. *Comput. Methods Appl. Mech. Eng.* **2021**, *376*, 113609. [CrossRef]
42. Tanyildizi, E.; Demir, G. Golden sine algorithm: A novel math-inspired algorithm. *Adv. Electr. Comput. Eng.* **2017**, *17*, 71–78. [CrossRef]
43. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]
44. Zivkovic, M.; Bacanin, N.; Venkatachalam, K.; Nayyar, A.; Djordjevic, A.; Strumberger, I.; Al-Turjman, F. COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustain. Cities Soc.* **2021**, *66*, 102669. [CrossRef] [PubMed]
45. Zivkovic, M.; Venkatachalam, K.; Bacanin, N.; Djordjevic, A.; Antonijevic, M.; Strumberger, I.; Rashid, T.A. Hybrid Genetic Algorithm and Machine Learning Method for COVID-19 Cases Prediction. In Proceedings of the International Conference on Sustainable Expert Systems: ICSES 2020, Lalitpur, Nepal, 28–29 September 2020; Springer Nature: Berlin/Heidelberg, Germany, 2021; Volume 176, p. 169.
46. Bacanin, N.; Bezdan, T.; Tuba, E.; Strumberger, I.; Tuba, M.; Zivkovic, M. Task scheduling in cloud computing environment by grey wolf optimizer. In Proceedings of the 2019 27th Telecommunications Forum (TELFOR), Serbia, Belgrade, 26–27 November 2019; pp. 1–4.
47. Bezdan, T.; Zivkovic, M.; Tuba, E.; Strumberger, I.; Bacanin, N.; Tuba, M. Multi-objective Task Scheduling in Cloud Computing Environment by Hybridized Bat Algorithm. In Proceedings of the International Conference on Intelligent and Fuzzy Systems, Izmir, Turkey, 21–23 July 2020; pp. 718–725.
48. Bezdan, T.; Zivkovic, M.; Antonijevic, M.; Zivkovic, T.; Bacanin, N. Enhanced Flower Pollination Algorithm for Task Scheduling in Cloud Computing Environment. In *Machine Learning for Predictive Analysis*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 163–171.
49. Zivkovic, M.; Bezdan, T.; Strumberger, I.; Bacanin, N.; Venkatachalam, K. Improved Harris Hawks Optimization Algorithm for Workflow Scheduling Challenge in Cloud–Edge Environment. In *Computer Networks, Big Data and IoT*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 87–102.
50. Zivkovic, M.; Bacanin, N.; Tuba, E.; Strumberger, I.; Bezdan, T.; Tuba, M. Wireless Sensor Networks Life Time Optimization Based on the Improved Firefly Algorithm. In Proceedings of the 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 15–19 June 2020; pp. 1176–1181.
51. Zivkovic, M.; Bacanin, N.; Zivkovic, T.; Strumberger, I.; Tuba, E.; Tuba, M. Enhanced Grey Wolf Algorithm for Energy Efficient Wireless Sensor Networks. In Proceedings of the 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), Online, 26–27 May 2020; pp. 87–92.
52. Bacanin, N.; Tuba, E.; Zivkovic, M.; Strumberger, I.; Tuba, M. Whale Optimization Algorithm with Exploratory Move for Wireless Sensor Networks Localization. In Proceedings of the International Conference on Hybrid Intelligent Systems, Sehore, India, 10–12 December 2019; pp. 328–338.
53. Zivkovic, M.; Zivkovic, T.; Venkatachalam, K.; Bacanin, N. Enhanced Dragonfly Algorithm Adapted for Wireless Sensor Network Lifetime Optimization. In *Data Intelligence and Cognitive Informatics*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 803–817.
54. Bezdan, T.; Cvetnic, D.; Gajic, L.; Zivkovic, M.; Strumberger, I.; Bacanin, N. Feature Selection by Firefly Algorithm with Improved Initialization Strategy. In Proceedings of the 7th Conference on the Engineering of Computer Based Systems, Sad Novi Sad, Serbia, 26–27 May 2021; pp. 1–8.
55. Nadimi-Shahraki, M.H.; Zamani, H.; Mirjalili, S. Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study. *Comput. Biol. Med.* **2022**, *148*, 105858. [CrossRef] [PubMed]
56. Bezdan, T.; Zivkovic, M.; Tuba, E.; Strumberger, I.; Bacanin, N.; Tuba, M. Glioma Brain Tumor Grade Classification from MRI Using Convolutional Neural Networks Designed by Modified FA. In Proceedings of the International Conference on Intelligent and Fuzzy Systems, Izmir, Turkey, 21 July 2020; pp. 955–963.
57. Zivkovic, M.; Bacanin, N.; Antonijevic, M.; Nikolic, B.; Kvascev, G.; Marjanovic, M.; Savanovic, N. Hybrid CNN and XGBoost Model Tuned by Modified Arithmetic Optimization Algorithm for COVID-19 Early Diagnostics from X-ray Images. *Electronics* **2022**, *11*, 3798. [CrossRef]
58. Strumberger, I.; Tuba, E.; Zivkovic, M.; Bacanin, N.; Beko, M.; Tuba, M. Dynamic search tree growth algorithm for global optimization. In Proceedings of the Doctoral Conference on Computing, Electrical and Industrial Systems, Costa de Caparica, Portugal, 8–10 May 2019; pp. 143–153.
59. Jovanovic, D.; Antonijevic, M.; Stankovic, M.; Zivkovic, M.; Tanaskovic, M.; Bacanin, N. Tuning Machine Learning Models Using a Group Search Firefly Algorithm for Credit Card Fraud Detection. *Mathematics* **2022**, *10*, 2272. [CrossRef]

60. Petrovic, A.; Bacanin, N.; Zivkovic, M.; Marjanovic, M.; Antonijevic, M.; Strumberger, I. The AdaBoost Approach Tuned by Firefly Metaheuristics for Fraud Detection. In Proceedings of the 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 17–19 June 2022; pp. 834–839.
61. Bacanin, N.; Sarac, M.; Budimirovic, N.; Zivkovic, M.; AlZubi, A.A.; Bashir, A.K. Smart wireless health care system using graph LSTM pollution prediction and dragonfly node localization. *Sustain. Comput. Inform. Syst.* **2022**, *35*, 100711. [CrossRef]
62. Bacanin, N.; Zivkovic, M.; Stoean, C.; Antonijevic, M.; Janicijevic, S.; Sarac, M.; Strumberger, I. Application of Natural Language Processing and Machine Learning Boosted with Swarm Intelligence for Spam Email Filtering. *Mathematics* **2022**, *10*, 4173. [CrossRef]
63. Stankovic, M.; Antonijevic, M.; Bacanin, N.; Zivkovic, M.; Tanaskovic, M.; Jovanovic, D. Feature Selection by Hybrid Artificial Bee Colony Algorithm for Intrusion Detection. In Proceedings of the 2022 International Conference on Edge Computing and Applications (ICECAA), Coimbatore, India, 21–23 September 2022; pp. 500–505.
64. Milosevic, S.; Bezdan, T.; Zivkovic, M.; Bacanin, N.; Strumberger, I.; Tuba, M. Feed-Forward Neural Network Training by Hybrid Bat Algorithm. In Proceedings of the Modelling and Development of Intelligent Systems: 7th International Conference, MDIS 2020, Sibiu, Romania, 22–24 October 2020; Revised Selected Papers 7; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 52–66.
65. Gajic, L.; Cvetnic, D.; Zivkovic, M.; Bezdan, T.; Bacanin, N.; Milosevic, S. Multi-layer Perceptron Training Using Hybridized Bat Algorithm. In *Computational Vision and Bio-Inspired Computing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 689–705.
66. Bacanin, N.; Stoean, C.; Zivkovic, M.; Jovanovic, D.; Antonijevic, M.; Mladenovic, D. Multi-Swarm Algorithm for Extreme Learning Machine Optimization. *Sensors* **2022**, *22*, 4204. [CrossRef]
67. Jovanovic, L.; Jovanovic, D.; Bacanin, N.; Jovancai Stakic, A.; Antonijevic, M.; Magd, H.; Thirumalaisamy, R.; Zivkovic, M. Multi-Step Crude Oil Price Prediction Based on LSTM Approach Tuned by Salp Swarm Algorithm with Disputation Operator. *Sustainability* **2022**, *14*, 14616. [CrossRef]
68. Qiu, Y.; Zhou, J.; Khandelwal, M.; Yang, H.; Yang, P.; Li, C. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Eng. Comput.* **2021**, *38*, 4145–4162. [CrossRef]
69. Jiang, H.; He, Z.; Ye, G.; Zhang, H. Network intrusion detection based on PSO-XGBoost model. *IEEE Access* **2020**, *8*, 58392–58401. [CrossRef]
70. Yun, K.K.; Yoon, S.W.; Won, D. Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Syst. Appl.* **2021**, *186*, 115716. [CrossRef]
71. Zivkovic, M.; Tair, M.; Venkatachalam, K.; Bacanin, N.; Hubálovskỳ, Š.; Trojovskỳ, P. Novel hybrid firefly algorithm: An application to enhance XGBoost tuning for intrusion detection classification. *PeerJ Comput. Sci.* **2022**, *8*, e956. [CrossRef]
72. Zivkovic, M.; Jovanovic, L.; Ivanovic, M.; Bacanin, N.; Strumberger, I.; Joseph, P.M. XGBoost Hyperparameters Tuning by Fitness-Dependent Optimizer for Network Intrusion Detection. In *Communication and Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 947–962.
73. AlHosni, N.; Jovanovic, L.; Antonijevic, M.; Bukumira, M.; Zivkovic, M.; Strumberger, I.; Mani, J.P.; Bacanin, N. The XGBoost Model for Network Intrusion Detection Boosted by Enhanced Sine Cosine Algorithm. In Proceedings of the International Conference on Image Processing and Capsule Networks, Bangkok, Thailand, 20–21 May 2022; pp. 213–228.
74. Tair, M.; Bacanin, N.; Zivkovic, M.; Venkatachalam, K.; Strumberger, I. XGBoost Design by Multi-verse Optimiser: An Application for Network Intrusion Detection. In *Mobile Computing and Sustainable Informatics*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 1–16.
75. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777. Available online: https://papers.nips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (accessed on 17 November 2022).
76. Molnar, C. Interpretable Machine Learning. 2020. Available online: https://christophm.github.io/interpretable-ml-book/index.html (accessed on 17 November 2022).
77. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
78. Goldberg, D.E.; Richardson, J. Genetic algorithms with sharing for multimodal function optimization. In *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*; Lawrence Erlbaum: Hillsdale, NJ, USA, 1987; Volume 4149.
79. Goldberg, D.E.; Deb, K. A comparative analysis of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*; Elsevier: Amsterdam, The Netherlands, 1991; Volume 1, pp. 69–93.
80. Mirjalili, S. Genetic algorithm. In *Evolutionary Algorithms and Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 43–55.
81. Tuba, M.; Bacanin, N. Artificial Bee Colony Algorithm Hybridized with Firefly Algorithm for Cardinality Constrained Mean-Variance Portfolio Selection Problem. *Appl. Math. Inf. Sci.* **2014**, *8*, 2831–2844. [CrossRef]
82. Yang, X.S. Firefly algorithms for multimodal optimization. In Proceedings of the International Symposium on Stochastic Algorithms, Sapporo, Japan, 26–28 October 2009; pp. 169–178.
83. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]

84. Heidari, A.A.; Faris, H.; Aljarah, I.; Mirjalili, S.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]
85. Zeng, H.; Shao, B.; Bian, G.; Dai, H.; Zhou, F. A hybrid deep learning approach by integrating extreme gradient boosting-long short-term memory with generalized autoregressive conditional heteroscedasticity family models for natural gas load volatility prediction. *Energy Sci. Eng.* **2022**, *10*, 1998–2021. [CrossRef]
86. Kang, Y.; Choi, H.; Im, J.; Park, S.; Shin, M.; Song, C.K.; Kim, S. Estimation of surface-level NO_2 and O_3 concentrations using TROPOMI data and machine learning over East Asia. *Environ. Pollut.* **2021**, *288*, 117711. [CrossRef]
87. Ji, X.; Xu, K.; Liao, D.; Chen, G.; Liu, T.; Hong, Y.; Dong, S.; Choi, S.D.; Chen, J. Spatial-temporal Characteristics and Source Apportionment of Ambient VOCs in Southeast Mountain Area of China. *Aerosol Air Qual. Res.* **2022**, *22*, 220016. [CrossRef]
88. Ibragimova, O.P.; Omarova, A.; Bukenov, B.; Zhakupbekova, A.; Baimatova, N. Seasonal and Spatial Variation of volatile organic compounds in ambient air of Almaty city, Kazakhstan. *Atmosphere* **2021**, *12*, 1592. [CrossRef]
89. Shi, Z.; Huang, L.; Li, J.; Ying, Q.; Zhang, H.; Hu, J. Sensitivity analysis of the surface ozone and fine particulate matter to meteorological parameters in China. *Atmos. Chem. Phys.* **2020**, *20*, 13455–13466. [CrossRef]
90. Zhang, X.; Xiao, X.; Wang, F.; Brasseur, G.; Chen, S.; Wang, J.; Gao, M. Observed sensitivities of PM2. 5 and O_3 extremes to meteorological conditions in China and implications for the future. *Environ. Int.* **2022**, *168*, 107428. [CrossRef]
91. Guo, S.; Wang, Y.; Zhang, T.; Ma, Z.; Ye, C.; Lin, W.; Yang Zong, D.J.; Yang Zong, B.M. Volatile organic compounds in urban Lhasa: Variations, sources, and potential risks. *Front. Environ. Sci.* **2022**, *23*, 1337. [CrossRef]
92. Parida, B.R.; Bar, S.; Roberts, G.; Mandal, S.P.; Pandey, A.C.; Kumar, M.; Dash, J. Improvement in air quality and its impact on land surface temperature in major urban areas across India during the first lockdown of the pandemic. *Environ. Res.* **2021**, *199*, 111280. [CrossRef] [PubMed]

atmosphere

MDPI

Article

An AI-Based Framework for Characterizing the Atmospheric Fate of Air Pollutants Within Diverse Environmental Settings

Nataša Radić [1], Mirjana Perišić [2,3], Gordana Jovanović [2,3], Timea Bezdan [3], Svetlana Stanišić [3], Nenad Stanić [3] and Andreja Stojić [2,3,*]

[1] Academy of Applied Studies Politehnika, Katarine Ambrozić 3, 11000 Belgrade, Serbia; nbukumiric@politehnika.edu.rs
[2] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia; mirjana.perisic@ipb.ac.rs (M.P.); gordana.jovanovic@ipb.ac.rs (G.J.)
[3] Software and Information Engineering, Singidunum University, Danijelova 32, 11000 Belgrade, Serbia; tbezdan@singidunum.ac.rs (T.B.); sstanisic@singidunum.ac.rs (S.S.); nenad.stanic.08@singimail.rs (N.S.)
* Correspondence: andreja.stojic@ipb.ac.rs

Abstract: This study introduces a novel artificial intelligence (AI) modeling framework that combines machine learning algorithms optimized through metaheuristics with explainable AI to capture complex interactions among pollutant concentrations, meteorological data, and socio-economic indicators. Applied to a COVID-19-related dataset comprising 404 variables, with benzene concentrations as the target—measured using proton transfer reaction–mass spectrometry in Belgrade, Serbia—the framework demonstrated exceptional sensitivity in assessing the impact of complex environmental and societal changes during the pandemic. Explainable AI techniques, such as SHAP and SAGE, were employed to reveal the influence of each predictor, while the clustering of SHAP values identified distinct environmental settings that influenced benzene behavior. Three distinct settings were identified regarding benzene levels during the onset of the state of emergency. The first, involving local petroleum-related activities, biomass burning, chemical manufacturing, and traffic, led to a 15.7% reduction in benzene levels. The second, characterized by non-combustion processes, nocturnal chemistry, and the specific meteorological context, resulted in a 51.9% increase. The third, driven by local industrial processes, contributed to a modest 2.33% reduction. The study underscored the critical role of environmental settings in shaping air pollutant behavior, emphasizing the importance of integrating broader environmental contexts into models to gain a more comprehensive understanding of air pollutants and their dynamics.

Keywords: benzene; COVID-19; proton transfer reaction–mass spectrometry; machine learning; explainable artificial intelligence

Academic Editor: Changqing Lin

Received: 21 January 2025
Revised: 14 February 2025
Accepted: 15 February 2025
Published: 18 February 2025

Citation: Radić, N.; Perišić, M.; Jovanović, G.; Bezdan, T.; Stanišić, S.; Stanić, N.; Stojić, A. An AI-Based Framework for Characterizing the Atmospheric Fate of Air Pollutants Within Diverse Environmental Settings. *Atmosphere* **2025**, *16*, 231. https://doi.org/10.3390/atmos16020231

1. Introduction

Understanding the complex behaviour of air pollutants, such as benzene—one of the key volatile organic compounds (VOCs) impacting public health—is essential for effective environmental management and policy development [1]. This complexity arises from the dynamic interplay of various environmental, societal, and economic factors [2]. The ability to disentangle the contributions and significance of influencing factors is particularly insightful during globally impactful events, such as the COVID-19 pandemic, which induced profound changes in traffic, economic, industrial, and day-to-day activities at both global and local levels. These changes resulted in notable alterations in pollutant

emissions and environmental conditions, creating unique research opportunities to observe environmental dynamics that are rarely, if ever, available under normal circumstances [3,4].

A deeper understanding of atmospheric pollutant behavior requires advanced methodologies, particularly artificial intelligence (AI). AI, broadly defined as the use of computational systems that emulate human cognitive functions, such as learning, pattern recognition, and decision-making, has become ubiquitous in modern society. From personalized recommendation algorithms in streaming platforms and e-commerce to real-time traffic navigation systems, predictive healthcare diagnostics, and smart home devices, AI-driven technologies increasingly shape daily life by optimizing efficiency, personalization, and problem-solving. In scientific research, AI's capacity to process vast datasets, uncover hidden patterns, and model nonlinear relationships has revolutionized fields such as climate science, epidemiology, and environmental science. In this study, we leveraged data on VOC concentrations (measured by the proton transfer reaction–mass spectrometer, PTR–MS), criteria air pollutants, meteorological parameters, epidemiological data, socio-economic indicators (Oxford COVID-19 Government Response Tracker—OxCGRT), mobility reports from Google and Apple, and stock market indices, within a novel AI-based framework to examine how widespread societal shifts have influenced benzene concentrations during the pandemic in Belgrade, Serbia. By incorporating pandemic-related lagged variables, we gained nuanced insights, shedding light on both the immediate and delayed direct and indirect effects of global events on air quality.

Our modeling approach employed ensemble machine learning algorithms optimized through several metaheuristics to model benzene concentrations, allowing for the interrelation of complex nonlinear relationships among variables. The best-performing model was interpreted using Shapley additive global importance (SAGE) and Shapley additive explanation (SHAP) values [5,6], explainable artificial intelligence (XAI) methods that elucidated the contribution of each predictor and the conditions under which they were most influential [7–10]. Additionally, we used the hierarchical density-based spatial clustering of applications with noise algorithm (HDBSCAN, McInnes and Healy 2017) to cluster SHAP values, where each cluster characterizes a unique environmental setting [11].

The term environmental setting, as used herein, refers to a set of coexisting conditions and factors within the environment that shape the environmental fate and predictability of benzene. This setting is defined by a variety of natural and anthropogenic variables, including, but not limited to, meteorological conditions, temporal variations, and human activities influencing benzene-related processes and concentration dynamics. By incorporating not only pollutant concentrations but also meteorological factors and a diverse range of variables reflecting the societal aspects of human activities, this concept offers a more robust framework for understanding emissions, sinks, and many contextual effects, surpassing the limitations of conventional source apportionment techniques. This innovative approach underscores the critical importance of the environmental context for accurately interpreting the relationships between variables and pollutant variability. The context is crucial because the same variable—such as temperature or wind speed—can exert different effects on the target compound, or display varying relationships, whether linear or non-linear, depending on the ambient conditions. Although the relationships presented are statistical, they may also reflect underlying causal mechanisms.

This approach has been applied to explore benzene dynamics during the onset of the COVID-19 state of emergency, which, as far as we are aware, has not been analyzed with such comprehensive contextualization, incorporating a broad spectrum of variables and employing AI at this level of detail to capture the broader environmental dynamics.

2. Methodology

2.1. Mitigation Measures

The first confirmed case of SARS-CoV-2 in Serbia was reported on 6 March 2020. Initial measures included suspending international travel and restricting entry from high-risk countries. By 10 March, the government had recommended enhanced hygiene practices, mask usage, and social distancing. As cases rose to 55 by 15 March, the measures escalated, leading to the closure of schools, limitations on public gatherings, and reductions in public transport [12]. A curfew was implemented on 18 March, initially from 8:00 p.m. to 5:00 a.m. By late April, restrictions began to ease, curfew hours were shortened, and some sectors started reopening. The state of emergency was lifted on 6 May, allowing for the resumption of transport and the reopening of shopping centers under strict health protocols.

2.2. Sampling Site

From 2 March to 15 May 2020, we measured VOCs and meteorological parameters at the Institute of Physics, Belgrade (44.86° N, 20.39° E). The site, situated 2 m above the roof and 10 m from ground level, ensured unobstructed airflow, minimizing potential bias from localized air currents. Located 8 km northwest of Belgrade's urban core and 2 km from central Zemun, the site is in a residential area along the Danube River. The area features small residential buildings with individual heating systems and is near a major bridge, 1 km northwest, impacting local air quality (Figure S1, Supplementary Materials).

2.3. Data

Detailed information about the experimental setup, measurement methods, calibration procedures, data preprocessing, and additional dataset information is available in the Supplementary Materials (Experimental Setup section; Figure S2 [13–16]). Briefly, we quantified concentrations of 249 protonated molecular masses of VOCs (21–270 amu) using a proton-transfer-reaction-quad mass spectrometer (PTR-quad-MS, Ionicon Analytic, Innsbruck, Austria). Our analysis incorporated external datasets, such as regulatory air pollution data, modeled meteorological variables, and information related to governmental measures, the pandemic, economic indicators, and human mobility patterns (Supplementary Materials, Data section, Table S1 [15–37] and Table S2). These datasets were sourced from the National Air Quality Network [37], the Global Data Assimilation System [38], and databases maintained by Oxford [39], Worldometer [40], Apple [41], and Google [42]. To assess delayed effects, we included time-lagged variables for pandemic containment measures, epidemiological data, and mobility metrics, offset by three days, and systematically indexed as lag1, lag2, and lag3.

2.4. Data Analysis

2.4.1. Bivariate Polar Plot

To identify potential spatial distribution of benzene emission sources, we employed a bivariate polar plot analysis [43] to examine the relationship between benzene concentrations and wind speed and direction. The analysis was conducted for periods before and during the state of emergency, with a focus on delineating differences in measured concentrations relative to wind characteristics, in order to assess the potential attenuation or intensification of specific sources and the emergence of new ones across different time frames. Subsequently, polar cluster analysis was applied to identify and group similar features and quantify the proportion of events associated with clusters.

2.4.2. Machine Learning

The data were analyzed using 7 regression tree ensemble machine learning algorithms that enhance prediction quality and robustness by combining multiple models into a single, more accurate model, and are particularly effective at reducing overfitting [44]. The ensembles include AdaBoost, ExtraTrees, Gradient Boosting, and Histogram Gradient Boosting (sklearn version 1.4), as well as CatBoost (1.2.7), LightGBM (4.5.0), and XGBoost (2.1.3). AdaBoost [45] increases focus on instances that were incorrectly predicted in previous rounds, thereby refining its performance continuously. CatBoost [46] optimizes gradient boosting by directly handling categorical features and reducing overfitting through ordered boosting. LightGBM [47] employs techniques like Gradient-based One-Side Sampling and Exclusive Feature Bundling to boost computational efficiency and feature management. XGBoost [48] constructs trees in parallel, supports multiple loss functions, and includes regularization to prevent overfitting. Both Gradient Boosting and Histogram Gradient Boosting, incorporated from the Python sklearn package [49], enhance modeling efficiency—Gradient Boosting builds models in a forward stage-wise fashion, while Histogram Gradient Boosting uses a binning method to speed up the training process and reduce memory usage.

To rigorously evaluate the performance of each model, we employed 5-fold cross-validation, ensuring comprehensive use of the dataset for both training and validation purposes, which significantly reduces the risk of overfitting. After training and evaluating the seven models, the top three performers were identified based on r-squared value and further refined to optimize performance using metaheuristics.

2.4.3. Metaheuristics

The effectiveness of machine learning models relies heavily on hyperparameter tuning, which is a nondeterministic polynomial-time (NP)-hard optimization challenge that is best addressed using metaheuristics. We incorporated six metaheuristic optimization algorithms to refine the hyperparameters of the top three models, thereby enhancing prediction accuracy. These algorithms included the Firefly Algorithm (FFA, [50]), Artificial Bee Colony (ABC, [51]), Harris Hawks Optimization (HHO, [52]), Sine Cosine Algorithm (SCA, [53]), Slime Mould Algorithm (SMA, [54]), and Quantum Search Algorithm (QSA, [55]) (mealpy version 3.0.1). These algorithms efficiently navigate the search space to identify near-optimal solutions. After determining the optimal hyperparameters for the three leading models, the best performing model was selected based on its r-squared value.

2.4.4. Explainable Artificial Intelligence

After identifying the best-performing model, we employed XAI methods SAGE and SHAP to ensure interpretability and transparency. SAGE assessed the global importance of each feature using an extension of the Shapley value from game theory, while SHAP values clarified each feature's contribution to individual predictions, allowing for comprehensive analysis across the dataset to identify trends in feature contributions.

We also calculated relative and normalized SHAP values and introduced a categorical system for interpreting SHAP values, referred to as inherent SHAP values. Relative SHAP values denoted the proportion of absolute SHAP within the overall attributed significance of all features for a single prediction, reflecting the relative impact of each feature [10]. Normalized SHAP values, normalized to the expected value, simplified the understanding of impact magnitude [11]. For inherent SHAP values, high negative impacts were defined as those below the mean of all negative SHAP values. Moderate negative impacts ranged from this mean to -10% of the 95th percentile of all absolute SHAP values, while minor impacts spanned from -10% to 10% of this 95th percentile. Moderate positive impacts

ranged from 10% of this percentile to the mean of positive SHAP values, and high positive impacts exceeded this mean. This classification system clarified the scale of each variable's effect regardless of its absolute magnitude compared to other predictors and highlighted its distinct contributions to the model's predictions. In this study, we utilized Python versions SAGE 0.2.4 and SHAP 0.46.0.

2.4.5. Cluster Analysis of Variable Impacts

To enhance the analytical depth of the study, we used dimensionality reduction and clustering techniques to analyze the relationships of variable local impacts. We utilized Uniform Manifold Approximation and Projection (UMAP, version 0.5.7, [56]) for dimensionality reduction, which preserves both local and global data structure, making it suitable for complex datasets. For clustering, we applied HDBSCAN [57] (version 0.8.40), which extends DBSCAN into a hierarchical clustering algorithm, dynamically selecting clusters based on data stability, allowing for the identification of clusters of varying densities. This combination allowed us to effectively identify and categorize localized impacts. Repeated framework executions consistently led to the identification of nine environmental settings, confirming the robustness of the clustering methodology.

According to the data and the obtained impacts, we have defined three ranges of normalized predictor levels: low (0–33%), medium (33–66%), and high (66–100%), for the variable absolute value, as well as three ranges of normalized impacts: low (1–5%), medium (5–15%), and high (15–100%).

3. Results and Discussion

3.1. Implementation of the State of Emergency

Prior to the state of emergency, benzene levels peaked at 2.26 ppb, with an average concentration of 0.48 ppb, driven by emission sources and weather conditions typical of the colder part of the year (Table S3, Supplementary Materials). In contrast, following the enactment of the emergency measures, the average benzene concentration was halved to 0.24 ppb, while peak levels dropped to 1.61 ppb.

Figure 1 visualizes the potential spatial distribution of benzene sources before and during the pandemic, enabling a direct comparison. The most intense sources, located in the N, NW, W, S, and SE, experienced significant reductions or were deactivated following the emergency measures. In the S and E areas, observed concentrations decreased by up to 0.8 ppb. However, new emission sources emerged in the N and SW areas.

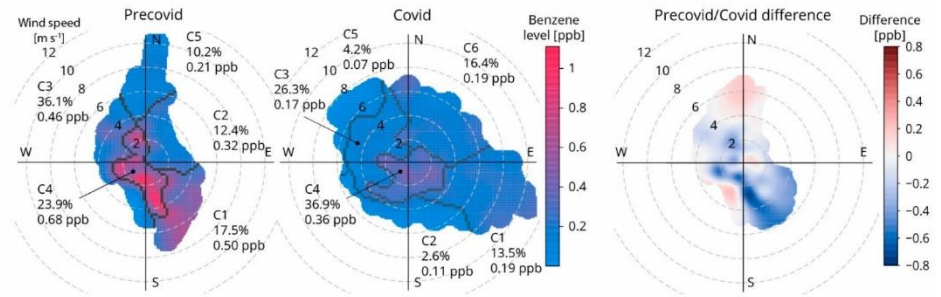


Figure 1. An indication of the spatial distribution (mean value and percentage of the total number of instances per cluster—C1–C6) and the difference in benzene levels before and during the pandemic.

The March state of emergency significantly impacted air quality, primarily driven by movement restrictions, institutional closures, reduced industrial activity, and decreased

traffic. These combined factors led to more stable benzene levels with smaller variations. Despite the sampling site being influenced by light to fresh breeze winds (2 to 10 m s^{-1}) from the north and east, regions typically associated with intense anthropogenic activities, no significant contribution to benzene concentrations was observed from these areas, likely due to the imposed governmental measures and subsequent changes in human behavior (Figure 2). Additionally, the onset of the state of emergency coincided with the transition into the warmer part of the year, suggesting that seasonal atmospheric changes and meteorological factors also contributed to the observed decrease in benzene concentrations. While lockdown measures were a dominant factor, the role of seasonal and meteorological variability cannot be overlooked. Future research will further refine this assessment by explicitly accounting for seasonal influences in a more comprehensive environmental analysis.

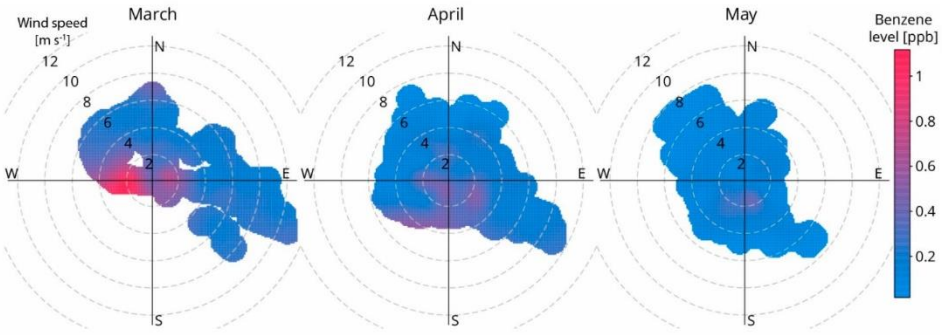


Figure 2. An indication of the monthly spatial distribution of benzene levels during the pandemic.

3.2. Modelling Results

The global explanation of the best-performing model (Gradient boosting model optimized with SCA metaheuristics, with MAE 0.034, max error 0.34, MSE 0.003, RMSE 0.056, MAPE 0.168, and R^2 0.94—MAE improvement of 15.1% compared to the non-optimized model) using SAGE revealed that 81 out of the 404 analyzed variables accounted for 95% of the global impact for benzene level prediction during the state of emergency (Table S4, Supplementary Materials). Notably, only 15 variables emerged as having a global impact greater than 1% (in descending order: m129, m41, m105, m97, m42, m45, m107, m95, m135, m62, m133, PM_{10}, m93, m163, and m73). Utilizing our clustering approach with SHAP explanations, we identified nine distinct groups of variable impacts, each representing a specific environmental setting (E0–E8). This method sheds light on how key factors sculpt the environmental context that affects benzene concentrations, particularly in emergency situations, like a pandemic, where changes in human activities play a significant role (Table 1). Additionally, a tenth group, E-1, remained unclustered, encompassing data instances likely to represent outliers or transitional regimes, rendering them unsuitable for categorization within a specific environmental setting.

The majority of data instances were grouped into E1 (38.6%), followed by E3 (10.5%), E7 (7.2%), and E6 (6.5%). The five least represented environmental settings accounted for 2.9% to 5.8% of the data, reflecting isolated or periodic events rather than consistent environmental regimes. Notably, three of these settings were associated with an increase in benzene levels relative to the expected value: E7 (0.37 ppb in average), E5 (0.26 ppb), and E2 (0.02 ppb). In contrast, the remaining settings demonstrated negative mean impacts from −0.12 to −0.01 ppb, indicating that the combinations of variables in these contexts are linked to environments characterized by lower benzene concentrations.

Table 1. Environmental setting statistics.

Environmental Setting	Mean Impact [ppb]	Mean Normalized Impact [%]	Mean Absolute Impact [ppb]	Population Percentage [%]	Dominant Inherent Impact	Dominant Inherent Impact Prevalence [%]
E-1	0.04	15.76	0.37	15.9		
E0	−0.12	−48.47	0.3	2.9	Moderate negative	36.9
E1	−0.09	−38.51	0.38	38.6	High negative	22.8
					Moderate negative	35.4
E2	0.02	8.89	0.27	4.0	Moderate negative	23.4
					Moderate positive	22.9
					High positive	29.0
E3	−0.04	−15.67	0.22	10.5	Moderate negative	29.2
					Moderate positive	27.1
E4	−0.01	−2.33	0.25	5.4	Moderate negative	24.4
					Moderate positive	26.5
E5	0.26	107.84	0.63	3.2	Moderate positive	23.7
					High positive	25.2
E6	−0.11	−45.15	0.32	6.5	Moderate negative	27.9
					Minor	21.3
					Moderate positive	24.6
E7	0.37	151.86	0.87	7.2	Moderate positive	29.6
					High positive	25.3
E8	−0.07	−27.26	0.32	5.8	Moderate negative	26.8
					Moderate positive	26.6

Based on the distribution of normalized variable levels and impacts across environmental settings (Figure 3), 27 variables, each with varying roles and magnitudes, are associated with the environmental fate of benzene. These include 18 VOCs (m41, m42, m45, m47, m62, m73, m75, m93, m95, m135, m61, m97, m105, m107, m129, m130, m133, and m163), two criteria air pollutants (PM_{10} and CO), three meteorological variables (total cloud cover—tcld, best four-layer lifted index—lib4, and temperature at 2 m—t02m), and four pandemic variables (confirmed deaths, total active cases lag 1, confirmed cases lag 2, and parks percentage change lag 3). Notably, m41, m42, m105, and m129 emerge as key determinants across all settings. We acknowledge that the ion signals at these masses could correspond to a variety of compounds. For instance, m41 could be associated with

alcohols, such as propanol, butanol, pentanol, and octanol, isoprene, acetonitrile, as well as 1-octen-3-ol and naphthalene—both identified at m129 (Table S1, Supplementary Materials). However, based on existing literature, we believe that the most likely compound at m41 is propyne. Similarly, the signals at m42, m105, and m129 are most likely attributable to acetonitrile, styrene, and naphthalene, respectively.

Figure 3. Normalized variable level and impact per environmental setting distribution [%]. The opacity was applied as a qualitative indicator of normalized variable level and impact magnitude. The legend implies the majority of positive (filled squares) and negative (open squares) impacts of all variables per setting.

Among the most influential variables, m45 (acetaldehyde) and m97 (cycloheptene) are key factors in nearly every environmental setting, with m107 (C8 aromatics) impacting all but two. Meanwhile, lib4, m62 (vinyl chloride), m95 (phenol or dimethyl sulfide), and m135 (C10 aromatics) play significant roles in all but three settings (Figure 3). The remaining variables offer valuable insights that can help trace the origins of benzene within each specific context, with special attention given to those that describe or reflect the immediate or delayed changes in human activities in response to the pandemic.

The environmental setting E3 initially aligned with the introduction of the state of emergency (Figure 4) while, as time progressed, the tightening of restrictions gave rise to periodic events attributed to E7 and regime E6. The subsequent relaxation of measures triggered the onset of environmental setting E1. The remaining settings represent isolated benzene pollution events that coincided with the development of specific environmental conditions. For the purpose of this study, we will focus on analyzing three environmental settings that marked the onset of the state of emergency: E3, E4, and E7.

3.2.1. Environmental Setting E3—Chemical Manufacturing, Combustion, and Petroleum-Related Emissions

During the first week of the state of emergency, the environmental setting identified as E3 became prominent, accounting for 10.5% of the total data instances (Figure 4). In this environmental content, observed benzene levels were, on average, 15.7% lower than the expected value (0.244 ppb; Table 1).

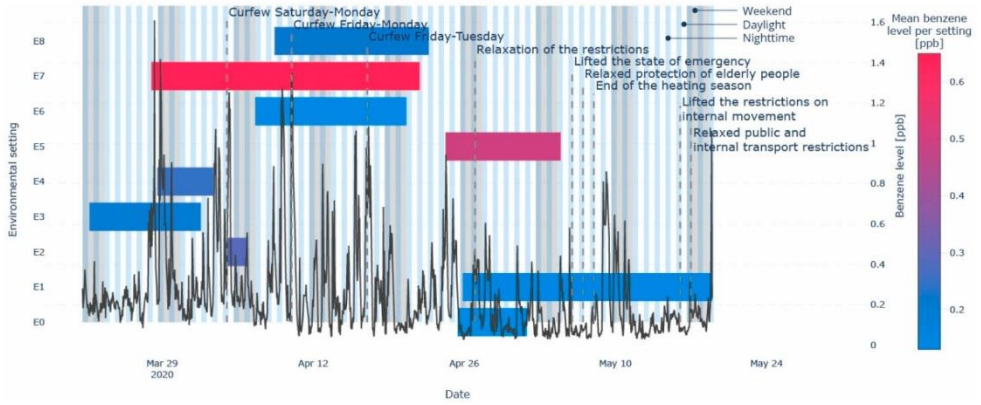


Figure 4. Benzene level and environmental settings time series. The main events during the pandemic are annotated. Weekend days are marked in gray, while daytime and nighttime are marked in white and light-blue.

The setting was primarily driven by 12 key variables, including a combination of VOCs, temperature (t02m), and lib4. As shown in Figure 5, the principal contributors among these were VOCs identified as cycloheptene, acetonitrile, styrene, and naphthalene, with relative impacts of 11.0%, −7.0%, −7.9%, and −11.9%, respectively, and medium mean normalized impacts on benzene level prediction of 8.5%, −5.2%, −6.0%, and −8.9% (Tables S1 and S5, Supplementary Materials). These compounds originate from petroleum refining, industrial emissions, chemical and plastic manufacturing, as well as the combustion of biomass, fossil fuels (vehicle exhaust, residential heating), coal, and wood [58,59], activities that are prevalent in the facilities surrounding the sampling site.

The positive association between benzene and cycloheptene (Group 1) suggests common sources that elevate benzene concentrations. In contrast, the negative association with styrene, and acetonitrile (Group 2), and even stronger negative association with naphthalene (Group 3), indicate that these compounds share sources that contribute to a reduction in benzene concentrations, with the sources of naphthalene likely being the weakest or most diminished among them. Industrial processes, including petroleum refining, storage and distribution, the production of synthetic rubber and plastics, and various combustion activities, like vehicle exhaust and biomass burning, are common sources of both cycloheptene and benzene. However, these processes are not necessarily associated with other compounds in Groups 2 and 3. The E3 setting was characterized by light to fresh breeze winds (averaging 5 m s^{-1}) from the north and especially the east, where the Pančevo Oil Refinery is located. Despite these conditions, the high levels of lib4 (averaging 12.5 °C) with minimal impact on benzene prediction indicate an atmosphere with significant vertical and some horizontal stability. This is likely to have limited vertical mixing, causing pollutants to remain near the surface. The wind speed was insufficient to disperse these pollutants effectively over a wide area, suggesting that the influence of distant sources, including the refinery, was insignificant, though other petroleum-related activities near the sampling site might still have been relevant. Given that the heating season was ongoing during the E3 period, local sources, particularly biomass burning for residential heating, are likely to have contributed to the observed increase in benzene concentrations. This is further supported by the observation that naphthalene levels and impacts were higher at night, as the extended time spent at home due to restrictions prolonged the heating period into the evening hours.

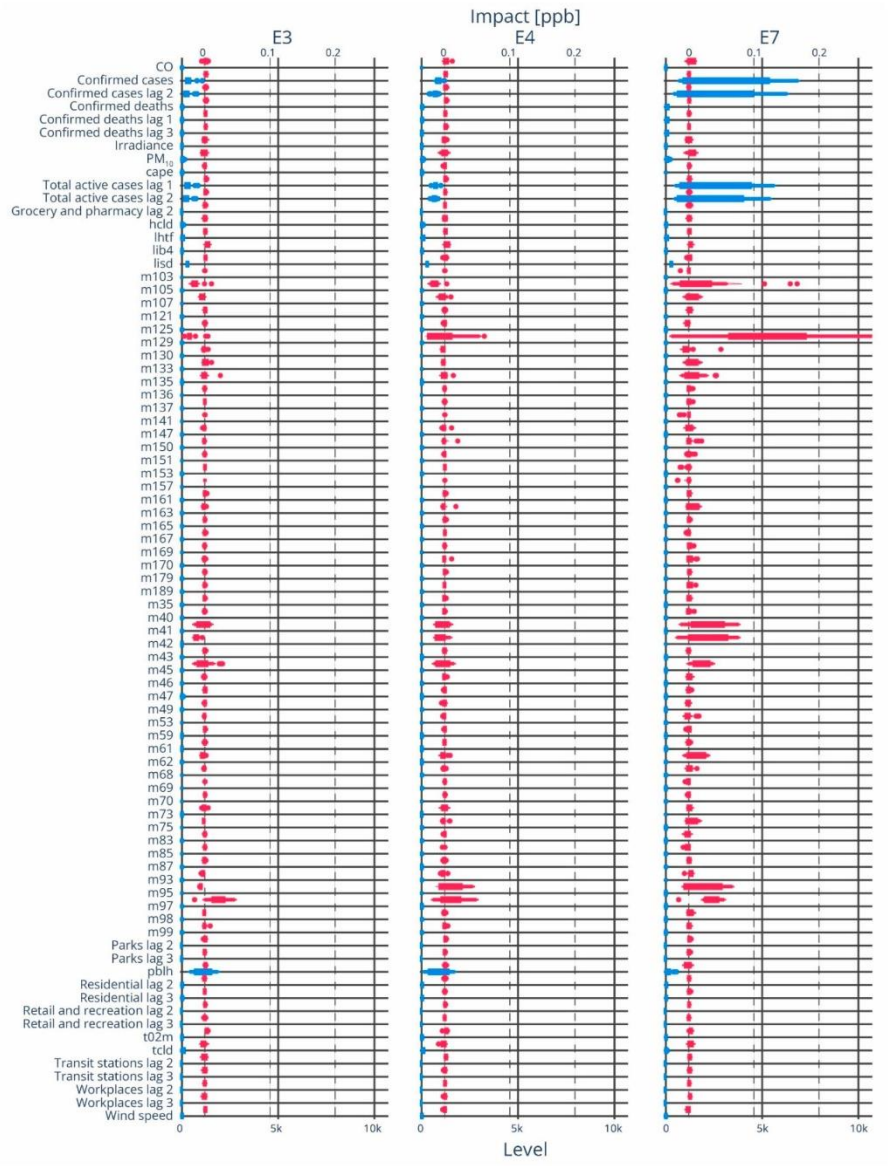


Figure 5. Distribution of the most important variable levels (blue) and impacts (red) for the environmental settings E3, E4, and E7.

In the case of Group 2, chemical manufacturing, particularly the production of acrylonitrile and polystyrene, emerged as a significant source, and the reduction in emissions from these activities ultimately led to a decrease in benzene levels. Regarding Group 3, the negative impact of naphthalene on benzene predictions, which is a byproduct of incomplete combustion of gasoline and diesel [60], was likely to have been attributable to reduced traffic activity near the measurement site. The onset of the state of emergency,

coinciding with the first registered COVID-19 cases, resulted in the closure of numerous institutions and schools, thereby significantly reducing vehicle usage. This reduction in traffic not only decreased emissions of key pollutants but also led to a lower presence of nitrogen oxides (NOx), which are precursors to both ozone and benzene formation. The primary compounds that contribute to benzene levels, including its precursors, are known to react with hydroxyl radicals, nitrate radicals and ozone, which leads to the formation of secondary pollutants, such as acetaldehyde. The observed further reductions in benzene levels appear to be linked to increased ozone concentrations, a phenomenon attributed to the reduction of aerosol and nitrogen oxide concentrations during the restrictions [61].

Finally, considering the other VOCs associated with this setting, their low levels resulted in low negative impacts on benzene prediction for propyne, acetaldehyde, C8 aromatics, phenol, dimethyl sulfide, vinyl chloride, and toluene (m93). This limited impact, coupled with the low concentrations, suggests a significant reduction in emissions from their typical sources.

At the end of the first week of the state of emergency, as the weekend began, benzene concentration patterns were influenced by the emergence of two additional environmental settings, E7 and E4 (Figure 4). Occurrences of E3 reappeared only in the middle of the second week of the state of emergency, but this setting faded before the second weekend, coinciding with the first implementation of the curfew.

3.2.2. Environmental Setting E7—Non-Combustion Emissions, Nocturnal Chemistry, and Meteorological Context

Within the environmental setting E7, which accounted for 7.2% of the data instances, benzene levels were notably elevated, with an average of 0.37 ppb. This setting captured high benzene events that periodically occurred and persisted until the final curfew, primarily during nighttime and early morning hours (10 p.m.–8 a.m.) towards the end of the workweek.

Benzene levels in this setting varied significantly, ranging from −8% to 421% relative to the expected value. These fluctuations were largely driven by 19 variables, including a mix of VOCs, CO, PM_{10}, and meteorological parameters (Figure 5). The most influential variables in predicting benzene levels were naphthalene (relative impact 21.4%), cycloheptene (7.8%), propyne (5.2%), and phenol/dimethyl sulfide (4.0%), contributing to increases of 50%, 14%, 13%, and 12%, respectively (Table S6, Supplementary Materials). Overall, the variables contributed highly positively to the prediction of benzene levels, except for C10 aromatics and lib4, which had moderate positive impacts, and styrene (m105), with a minor impact.

Among the compounds observed at high levels, naphthalene was the only one that made a significant contribution to the prediction of benzene levels, while cycloheptene and acetonitrile had moderate impacts, and CO had a minor impact (Table S6). These compounds reflect a mix of emissions from industrial activities and fossil fuels, as previously discussed, as well as potential nocturnal chemistry under specific atmospheric conditions [62]. This interpretation aligns with our findings, as the levels and corresponding impacts of styrene, cycloheptene, acetonitrile, PM_{10}, and CO increased during nighttime. In urban atmospheres, significant amounts of NO are oxidized to NO_2, which is then converted to the nitrate radical (NO_3) by ozone. NO_3, the primary nocturnal tropospheric oxidant, accumulates at night due to its rapid degradation through photolysis and reaction with NO during daylight. At night, NO_3-driven reactions significantly influence air pollution and regulate the lifespan of various trace gases, including VOCs, leading to the formation of complex oxidation products. The high concentration and reactivity of NO_3 make it responsible for the degradation of many unsaturated hydrocarbons during the night. For instance, styrene concentrations increased significantly within the stable nocturnal bound-

ary layer, resulting in heightened reactivity. Styrene thus serves as an effective indicator of NO_3 reactivity across different seasons and for estimating NO_3 reactivity toward other VOCs [62]. Conversely, the opposite trend was observed for propyne, C8 aromatics, C10 aromatics, fragmentation ions at m133, and lib4 (Table S6, Supplementary Materials).

Among the compounds exhibiting medium levels, propyne, styrene, acetaldehyde, and phenol/dimethyl sulfide recorded medium, while C8 aromatics, C10 aromatics, vinyl chloride, fragmentation ions at m133, PM_{10}, C12 aromatics, and hydroxy acetone showed low impacts on benzene prediction. The relatively low impacts of CO and PM_{10} suggested that combustion played a lesser role in this setting compared to other potential sources of benzene, a conclusion further supported by the periodic nature of this setting. Additionally, the analysis of traffic intensity during the weekdays revealed no surge in activity prior to curfews, making it unlikely that vehicle exhaust significantly contributed to the observed benzene levels.

The strong association of benzene and naphthalene, with their concentration consistently increasing during the workweek, then declining over the weekend, likely originating from non-combustion processes. These joint emissions can be attributed to various nearby facilities, including an oil refinery (NE and E), textile recycling (E and SE), paper industry (NW, W, SW, S, and SE), rubber and plastic production (W, NW, S, SE, E, and N), and furniture manufacturing (W, SW, S, SE, and E). The levels and impacts of VOCs generally decreased as the duration of this setting progressed, aligning with the reduction in anthropogenic activities due to increasing restrictions. Moreover, typical human activities intensified before each weekend curfew as people prepared for the impending lockdown, likely affecting the observed environmental conditions.

The meteorological context, shaped by the best four-layer lifted index, temperature, and total cloud cover, played an important role in defining this setting. Calm to light breeze winds (0–3.3 m s^{-1}) and medium levels of lib4, which exhibited a low impact on benzene prediction, suggest that, although dispersion and transport processes were present, they did not significantly affect benzene concentrations. While the observed medium-level temperatures (up to 13 °C) are likely to have accelerated chemical reactions, potentially increasing secondary pollutants like ozone and particulate matter [63], the nocturnal nature of this setting suggests that ozone formation was minimal during the early morning hours. Additionally, our results indicate that the recorded temperature levels had only a modest impact on benzene dynamics, averaging 1.3%.

Low levels of total cloud cover (tcld) had a low but notable positive impact on benzene levels through various mechanisms. At night, cloud cover traps pollutants closer to the surface by reducing the radiative cooling and preventing rapid heat loss, stabilizing the lower atmosphere and leading to pollutant accumulation near the ground [64]. During the day, tcld plays a critical role in modulating solar radiation, potentially lowering surface temperatures, which might reduce the volatilization rate of some VOCs or slow down the photochemical reactions necessary for the formation of certain secondary pollutants. The higher positive impact observed during the day compared to night suggests that cloud cover plays a more crucial role in shaping benzene prediction during daylight hours in this environmental setting.

3.2.3. Environmental Setting E4—Local Industrial Processes

Setting E4, accounting for 5.4% of data instances, emerged during the first weekend of the state of emergency and persisted until the following weekend as an isolated event (Figure 4). It contributed to a slight further reduction in observed benzene levels, averaging 2.33% below the expected value (Table 1).

The setting was primarily driven by eight variables, including a combination of VOCs and lib4 (Figure 5). The key contributors were cycloheptene (5.0%), naphthalene (−3.3%), acetonitrile (−3.4%), and styrene (−8.3%), with mean normalized impacts on benzene prediction of 4.4%, −1.1%, −2.4%, and −6.2%, respectively (Table S7, Supplementary Materials).

The differentiation between cycloheptene and benzene (Group 1), naphthalene, acetonitrile, and benzene (Group 2), and styrene and benzene (Group 3) primarily arises from specific industrial activities and combustion processes. Although Groups 1 and 2 share common sources, such as vehicle exhaust and industrial emissions, Group 2 is more specifically associated with chemical production activities, which play a less prominent role in the origin of the Group 1 compounds. Group 1 is closely associated with combustion processes, like vehicle exhaust, residential heating, and biomass burning, as well as industrial emissions from petroleum refining. Reduced commuting and increased time spent at home due to restrictive measures intensified emissions from residential heating and biomass burning. In contrast, Group 2 is more associated with chemical manufacturing processes and coal tar processing where acetonitrile and naphthalene are produced alongside benzene. Group 3 is distinct for its strong ties to plastic and resin production, where styrene is a key component, and its presence alongside benzene highlights these specific industrial activities. The light to moderate breeze (1.5–6 m s^{-1}) from the N and NW within this setting carried cleaner air from regions where the sources of Group 2 and 3 emissions were located, which were already operating at reduced capacity due to the state of emergency, further diminishing their impact and leading to benzene concentrations lower than those expected.

Similar to the E3, high levels of lib4 (mean 10.1 °C), which resulted in low impacts on benzene prediction suggest a stable atmosphere both vertically and, to some extent, horizontally. This stability is likely to have limited vertical mixing, with stratified air masses keeping pollutants close to the surface, thereby reducing the impact of more distant industrial facilities.

4. Conclusions

This study introduced a novel approach centered on "environmental settings", the combination of natural and human factors that govern air pollutant behavior. By integrating pollutant concentrations with contextual variables (e.g., meteorological conditions, temporal factors, and human activities), our AI framework offers a comprehensive view of benzene dynamics, surpassing the limitations of traditional source apportionment techniques.

Our methodology, combining advanced machine learning, metaheuristic optimization, explainable AI, and clustering, identified key drivers of benzene concentrations. Techniques like bivariate polar plots, polar clustering, SAGE, and SHAP uncovered both spatial patterns and model insights across diverse environmental settings. Applied to Belgrade's benzene data during the onset of COVID-19, the approach revealed how severely reduced transportation and industrial activities only modestly lowered benzene levels. This finding highlights the importance of considering a broad range of factors beyond direct emissions.

Overall, this framework provides actionable knowledge for air quality management, demonstrating how "environmental settings" can inform crisis responses and routine policies alike. By quantifying the maximum possible impact of altered activities and processes, decision-makers can better anticipate and mitigate changes in pollution levels. While a formal sensitivity analysis was not conducted, the stability of the results across iterations suggests that the selected number of clusters is representative. Future research will further refine this framework by incorporating an extended set of predictors to enhance environmental setting characterization. Additionally, future research will focus on uncertainty propagation and on assessing the stability of predictor importance rankings across multiple

model runs, and providing confidence intervals to ensure greater robustness and reliability of the results.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/atmos16020231/s1, Figure S1: Location of the study area in Belgrade and a photograph of the monitoring site (Source: Google Maps); Figure S2: Normalized sensitivities and the relative transmission curve; Table S1: Compounds detected by PTR-MS in ambient air; Table S2: Global data assimilation system data abbreviations; Table S3: Descriptive statistics for measured parameters before and after the introduction of the state of emergency; Table S4: SAGE statistics; Table S5: Environmental setting E3 statistics; Table S6: Environmental setting E7 statistics; Table S7: Environmental setting E4 statistics.

Author Contributions: N.R.: Funding acquisition, Investigation. M.P.: Conceptualization, Data Curation, Formal Analysis, Funding acquisition, Investigation, Writing—original draft, Writing—review and editing. G.J.: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Writing—original draft, Writing—review and editing. T.B.: Methodology, Project Administration, Resources, Software. S.S.: Funding acquisition, Investigation, Writing—original draft, Writing—review and editing. N.S.: Data curation. A.S.: Conceptualization, Data Curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge funding provided by the Institute of Physics Belgrade, through a grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, as well as by the Science Fund of the Republic of Serbia, Grant No. #7373, *Characterizing crises-caused air pollution alternations using an artificial intelligence-based framework—crAIRsis*.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zahed, M.A.; Salehi, S.; Khoei, M.A.; Esmaeili, P.; Mohajeri, L. Risk assessment of Benzene, Toluene, Ethyl benzene, and Xylene (BTEX) in the atmospheric air around the world: A review. *Toxicol. Vitr.* **2024**, *98*, 105825. [CrossRef]
2. Li, D.; Tao, X.; Song, X.; Liu, S.; Yuan, K.; Deng, F.; Guo, Y. Ambient volatile organic compounds concentration variation characteristics and source apportionment in Lanzhou, China during the COVID-19 lockdown. *Atmos. Pollut. Res.* **2024**, *15*, 102064. [CrossRef]
3. Cardito, A.; Carotenuto, M.; Amoruso, A.; Libralato, G.; Lofrano, G. Air quality trends and implications pre and post COVID-19 restrictions. *Sci. Total Environ.* **2023**, *879*, 162833. [CrossRef] [PubMed]
4. Ling, C.; Cui, L.; Li, R. Global impact of the COVID-19 lockdown on surface concentration and health risk of atmospheric benzene. *Atmos. Chem. Phys.* **2023**, *23*, 3311–3324. [CrossRef]
5. Covert, I.; Lundberg, S.M.; Lee, S.I. Understanding global feature contributions with additive importance measures. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17212–17223.
6. Lundberg, S. A unified approach to interpreting model predictions. *arXiv* **2017**, arXiv:1705.07874.
7. Bacanin, N.; Perisic, M.; Jovanovic, G.; Damaševičius, R.; Stanisic, S.; Simic, V.; Zivkovic, M.; Stojic, A. The explainable potential of coupling hybridized metaheuristics, XGBoost, and SHAP in revealing toluene behavior in the atmosphere. *Sci. Total Environ.* **2024**, *929*, 172195. [CrossRef] [PubMed]
8. Jovanovic, G.; Perisic, M.; Bacanin, N.; Zivkovic, M.; Stanisic, S.; Strumberger, I.; Alimpic, F.; Stojic, A. Potential of coupling metaheuristics-optimized-XGBoost and SHAP in revealing PAHs environmental fate. *Toxics* **2023**, *11*, 394. [CrossRef] [PubMed]
9. Jovanovic, L.; Jovanovic, G.; Perisic, M.; Alimpic, F.; Stanisic, S.; Bacanin, N.; Zivkovic, M.; Stojic, A. The explainable potential of coupling metaheuristics-optimized-xgboost and shap in revealing vocs' environmental fate. *Atmosphere* **2023**, *14*, 109. [CrossRef]

10. Stojić, A.; Jovanović, G.; Stanišić, S.; Romanić, S.H.; Šoštarić, A.; Udovičić, V.; Perišić, M.; Milićević, T. The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels. *Chemosphere* **2022**, *289*, 133154. [CrossRef] [PubMed]
11. Jovanović, G.; Perišić, M.; Bezdan, T.; Stanišić, S.; Radusin, K.; Popović, A.; Stojić, A. The PM 2.5-Bound Polycyclic Aromatic Hydrocarbon Behavior in Indoor and Outdoor Environments, Part III: Role of Environmental Settings in Elevating Indoor Concentrations of Benzo (a) pyrene. *Atmosphere* **2024**, *15*, 1520. [CrossRef]
12. Government of Serbia. 2020. Available online: https://www.srbija.gov.rs/vest/en/155727/serbia-lifts-state-of-emergency.php (accessed on 1 June 2023).
13. Lindinger, W.; Hansel, A.; Jordan, A. On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research. *Int. J. Mass Spectrom. Ion Processes* **1998**, *173*, 191–241. [CrossRef]
14. Stojić, A.; Maletić, D.; Stojić, S.S.; Mijić, Z.; Šoštarić, A. Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. *Sci. Total Environ.* **2015**, *521*, 19–26. [CrossRef] [PubMed]
15. Taipale, R.; Ruuskanen, T.M.; Rinne, J.; Kajos, M.K.; Hakola, H.; Pohja, T.; Kulmala, M. Quantitative long-term measurements of VOC concentrations by PTR-MS–measurement, calibration, and volume mixing ratio calculation methods. *Atmos. Chem. Phys.* **2008**, *8*, 6681–6698. [CrossRef]
16. Blake, R.S.; Monks, P.S.; Ellis, A.M. Proton-transfer reaction mass spectrometry. *Chem. Rev.* **2009**, *109*, 861–896. [CrossRef] [PubMed]
17. Buhr, K.; van Ruth, S.; Delahunty, C. Analysis of volatile flavour compounds by Proton Transfer Reaction-Mass Spectrometry: Fragmentation patterns and discrimination between isobaric and isomeric compounds. *Int. J. Mass Spectrom.* **2002**, *221*, 1–7. [CrossRef]
18. Dunne, E.; Galbally, I.E.; Lawson, S.; Patti, A. Interference in the PTR-MS measurement of acetonitrile at m/z 42 in polluted urban air—A study using switchable reagent ion PTR-MS. *Int. J. Mass Spectrom.* **2012**, *319*, 40–47. [CrossRef]
19. Zannoni, N.; Gros, V.; Lanza, M.; Sarda, R.; Bonsang, B.; Kalogridis, C.; Preunkert, S.; Legrand, M.; Jambert, C.; Boissard, C.; et al. OH reactivity and concentrations of biogenic volatile organic compounds in a Mediterranean forest of downy oak trees. *Atmos. Chem. Phys.* **2016**, *16*, 1619–1636. [CrossRef]
20. Nakajima, S.; Asakawa, T.; Sekiguchi, O.; Tajima, S.; Nibbering, N.M. The formation of protonated dimethyl ether from the metastable molecular ions of 1-methoxy-2-propanol, CH3OCH2CH (OH) CH3. *Eur. J. Mass Spectrom.* **2001**, *7*, 47–53. [CrossRef]
21. Wang, R.; Zhang, Y.; Jiao, L.; Zhao, X.; Gao, Z.; Dong, D. Proton-Transfer-Reaction Mass Spectrometry for Rapid Dynamic Measurement of Ethylene Oxide Volatilization from Medical Masks. *Atmosphere* **2024**, *15*, 114. [CrossRef]
22. Brown, P.; Watts, P.; Märk, T.D.; Mayhew, C.A. Proton transfer reaction mass spectrometry investigations on the effects of reduced electric field and reagent ion internal energy on product ion branching ratios for a series of saturated alcohols. *Int. J. Mass Spectrom.* **2010**, *294*, 103–111. [CrossRef]
23. Baasandorj, M.; Millet, D.B.; Hu, L.; Mitroo, D.; Williams, B.J. Measuring acetic and formic acid by proton-transfer-reaction mass spectrometry: Sensitivity, humidity dependence, and quantifying interferences. *Atmos. Meas. Tech.* **2015**, *8*, 1303–1321. [CrossRef]
24. Inomata, S.; Tanimoto, H.; Kato, S.; Suthawaree, J.; Kanaya, Y.; Pochanart, P.; Liu, Y.; Wang, Z. PTR-MS measurements of non-methane volatile organic compounds during an intensive field campaign at the summit of Mount Tai, China, in June 2006. *Atmos. Chem. Phys.* **2010**, *10*, 7085–7099. [CrossRef]
25. Pal, P.; Banat, F. Comparison of thermal degradation between fresh and industrial aqueous methyldiethanolamine with continuous injection of H2S/CO2 in high pressure reactor. *J. Nat. Gas Sci. Eng.* **2016**, *29*, 479–487. [CrossRef]
26. Awasthi, A.; Sinha, B.; Hakkim, H.; Mishra, S.; Mummidivarapu, V.; Singh, G.; Ghude, S.D.; Soni, V.K.; Nigam, N.; Sinha, V.; et al. Biomass burning sources control ambient particulate matter but traffic and industrial sources control VOCs and secondary pollutant formation during extreme pollution events in Delhi. *EGUsphere* **2024**, 1–35. [CrossRef]
27. Van Huffela, K.; Hansenb, M.J.; Feilbergb, A.; Liub, D.; Van Langenhovea, H. The power of online proton transfer reaction-mass spectrometry (PTR-MS) measurement of odorous emissions from a pig house. *Chem. Eng.* **2014**, *40*, 241–246.
28. Chen, L.; Yan, R.; Zhao, Y.; Sun, J.; Zhang, Y.; Li, H.; Zhao, D.; Wang, B.; Ye, X.; Sun, B. Characterization of the aroma release from retronasal cavity and flavor perception during baijiu consumption by Vocus-PTR-MS, GC× GC-MS, and TCATA analysis. *LWT* **2023**, *174*, 114430. [CrossRef]
29. De Gouw, J.; Warneke, C. Measurements of volatile organic compounds in the earth's atmosphere using proton-transfer-reaction mass spectrometry. *Mass Spectrom. Rev.* **2007**, *26*, 223–257. [CrossRef] [PubMed]
30. Perraud, V.; Meinardi, S.; Blake, D.R.; Finlayson-Pitts, B.J. Challenges associated with the sampling and analysis of organosulfur compounds in air using real-time PTR-ToF-MS and offline GC-FID. *Atmos. Meas. Tech.* **2016**, *9*, 1325–1340. [CrossRef]
31. Papurello, D.; Silvestri, S.; Tomasi, L.; Belcari, I.; Biasioli, F.; Santarelli, M. Natural gas trace compounds analysis with innovative systems: PTR-ToF-MS and FASTGC. *Energy Procedia* **2016**, *101*, 536–541. [CrossRef]

32. Mallette, N.D.; Knighton, W.B.; Strobel, G.A.; Carlson, R.P.; Peyton, B.M. Resolution of volatile fuel compound profiles from Ascocoryne sarcoides: A comparison by proton transfer reaction-mass spectrometry and solid phase microextraction gas chromatography-mass spectrometry. *AMB Express* **2012**, *2*, 1–13. [CrossRef] [PubMed]
33. Tasin, M.; Cappellin, L.; Biasioli, F. Fast direct injection mass-spectrometric characterization of stimuli for insect electrophysiology by proton transfer reaction-time of flight mass-spectrometry (PTR-ToF-MS). *Sensors* **2012**, *12*, 4091–4104. [CrossRef]
34. Bruns, E.A.; Slowik, J.G.; El Haddad, I.; Kilic, D.; Klein, F.; Dommen, J.; Temime-Roussel, B.; Marchand, N.; Baltensperger, U.; Prévôt, A.S. Characterization of gas-phase organics using proton transfer reaction time-of-flight mass spectrometry: Fresh and aged residential wood combustion emissions. *Atmos. Chem. Phys.* **2017**, *17*, 705–720. [CrossRef]
35. Westphal, F.; Junge, T. Ring positional differentiation of isomeric N-alkylated fluorocathinones by gas chromatography/tandem mass spectrometry. *Forensic Sci. Int.* **2012**, *223*, 97–105. [CrossRef]
36. Erickson, M.H.; Gueneron, M.; Jobson, B.T. Measuring long chain alkanes in diesel engine exhaust by thermal desorption PTR-MS. *Atmos. Meas. Tech. Discuss.* **2013**, *6*, 6005–6046. [CrossRef]
37. SEPA. 2022. Available online: http://amskv.sepa.gov.rs/mob/index.php?lng=en (accessed on 1 June 2023).
38. GDAS. 2022. Available online: https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00379 (accessed on 1 June 2023).
39. Oxford COVID-19 Government Response Tracker. OxCovid19 Database. 2022. Available online: https://covid19.oii.ox.ac.uk/database/ (accessed on 1 June 2023).
40. Worldometer. COVID-19 Coronavirus Pandemic Data [Data Set]. 2022. Available online: https://www.worldometers.info/coronavirus/ (accessed on 1 June 2023).
41. Apple Inc. COVID-19—Mobility Trends Reports. 2022. Available online: https://www.apple.com/covid19/mobility (accessed on 1 June 2023).
42. Google LLC. COVID-19 Community Mobility Reports. 2022. Available online: https://www.google.com/covid19/mobility/ (accessed on 1 June 2023).
43. Carslaw, D.C.; Beevers, S.D. Characterising and understanding emission sources using bivariate polar plots and k-means clustering. *Environ. Model. Softw.* **2013**, *40*, 325–329. [CrossRef]
44. Dietterich, T.G. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
45. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
46. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*. Available online: https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html (accessed on 12 December 2024).
47. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (accessed on 12 December 2024).
48. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
49. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
50. Yang, X.S. Firefly algorithms for multimodal optimization. In *International Symposium on Stochastic Algorithms*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 169–178.
51. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization*; Technical Report-tr06; Computer Engineering Department, Engineering Faculty, Erciyes University: Kayseri, Turkey, 2005; Volume 200, pp. 1–10.
52. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]
53. Mirjalili, S. SCA: A sine cosine algorithm for solving optimization problems. *Knowl.-Based Syst.* **2016**, *96*, 120–133. [CrossRef]
54. Li, S.; Chen, H.; Wang, M.; Heidari, A.A.; Mirjalili, S. Slime mould algorithm: A new method for stochastic optimization. *Future Gener. Comput. Syst.* **2020**, *111*, 300–323. [CrossRef]
55. Zhang, J.; Xiao, M.; Gao, L.; Pan, Q. Queuing search algorithm: A novel metaheuristic algorithm for solving engineering optimization problems. *Appl. Math. Model.* **2018**, *63*, 464–490. [CrossRef]
56. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
57. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density-based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [CrossRef]
58. Kong, L.; Zhou, L.; Chen, D.; Luo, L.; Xiao, K.; Chen, Y.; Liu, H.; Tan, Q.; Yang, F. Atmospheric oxidation capacity and secondary pollutant formation potentials based on photochemical loss of VOCs in a megacity of the Sichuan Basin, China. *Sci. Total Environ.* **2023**, *901*, 166259. [CrossRef] [PubMed]

59. Zhang, L.; Xu, T.; Wu, G.; Zhang, C.; Li, Y.; Wang, H.; Gong, D.; Li, Q.; Wang, B. Photochemical loss with consequential underestimation in active VOCs and corresponding secondary pollutions in a petrochemical refinery, China. *Sci. Total Environ.* **2024**, *918*, 170613. [CrossRef]
60. Marques, B.; Kostenidou, E.; Valiente, A.M.; Vansevenant, B.; Sarica, T.; Fine, L.; Temime-Roussel, B.; Tassel, P.; Perret, P.; Liu, Y.; et al. Detailed speciation of non-methane volatile organic compounds in exhaust emissions from diesel and gasoline Euro 5 vehicles using online and offline measurements. *Toxics* **2022**, *10*, 184. [CrossRef] [PubMed]
61. Nussbaumer, C.M.; Pozzer, A.; Tadic, I.; Röder, L.; Obersteiner, F.; Harder, H.; Lelieveld, J.; Fischer, H. Tropospheric ozone production and chemical regime analysis during the COVID-19 lockdown over Europe. *Atmos. Chem. Phys.* **2022**, *22*, 6151–6165. [CrossRef]
62. Hu, H.; Wang, H.; Lu, K.; Wang, J.; Zheng, Z.; Xu, X.; Zhai, T.; Chen, X.; Lu, X.; Fu, W.; et al. Variation and trend of nitrate radical reactivity towards volatile organic compounds in Beijing, China. *Atmos. Chem. Phys.* **2023**, *23*, 8211–8223. [CrossRef]
63. Hu, J.; Zhao, T.; Liu, J.; Cao, L.; Xia, J.; Wang, C.; Zhao, X.; Gao, Z.; Shu, Z.; Li, Y. Nocturnal surface radiation cooling modulated by cloud cover change reinforces PM2.5 accumulation: Observational study of heavy air pollution in the Sichuan Basin, Southwest China. *Sci. Total Environ.* **2021**, *794*, 148624. [CrossRef] [PubMed]
64. Luo, H.; Han, Y.; Lu, C.; Yang, J.; Wu, Y. Characteristics of surface solar radiation under different air pollution conditions over Nanjing, China: Observation and simulation. *Adv. Atmos. Sci.* **2019**, *36*, 1047–1059. [CrossRef]

Публикације категорије М30

Публикације након избора у претходно звање

Singidunum University

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Sinteza

Belgrade, 2024

sinteza.singidunum.ac.rs

Singidunum University

Sinteza

SINGIDUNUM UNIVERSITY INTERNATIONAL SCIENTIFIC CONFERENCE

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology, Computer Science, and Data Science- Sinteza 2024 has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia.

Belgrade
May 16, 2024.
sinteza.singidunum.ac.rs

Sinteza

Singidunum University International Scientific Conference

SCIENTIFIC COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Endre Pap, Singidunum University, Serbia
- Aleksandar Jevremović, Singidunum University, Serbia
- Andreja Stojić, Institute of Physics in Belgrade, Serbia
- Bratislav Milovanović, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Đorđe Obradović, Singidunum University, Serbia
- Goran Šimić, Military Academy, University of Defence, Serbia
- Gordana Dobrijević, Singidunum University, Serbia
- Gordana Jovanović, Institute of Physics in Belgrade, Serbia
- Jelena Filipović, Faculty of Philology, University of Belgrade, Serbia
- Miloš Antonijević, Singidunum University, Serbia
- Marina Marjanović Jakovljević, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Marko Šarac, Singidunum University, Serbia
- Milan Tair, Singidunum University, Serbia
- Milan Tuba, Singidunum University, Serbia
- Miljan Vučetić, Vlatacom Institute, Serbia
- Miloš Stojmenović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Mirjana Perišić, Institute of Physics in Belgrade, Serbia
- Miroslav Popović, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mlađan Jovanović, Singidunum University, Serbia
- Nebojša Bačanin Džakula, Singidunum University, Serbia
- Nemanja Stanišić, Singidunum University, Serbia
- Nina Dragićević, Singidunum University, Serbia
- Petar Spalević, Faculty of Technical Sciences in Kosovska Mitrovica, University of Priština, Serbia
- Predrag Popović, Vinča Institute, Serbia
- Sanja Filipović, Singidunum University, Serbia
- Saša Adamović, Singidunum University, Serbia
- Tijana Radojević, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Alexandru Nedelea, Stefan cel Mare University of Suceava, Romania
- Aurora Pedro Bueno, Department of Applied Economics, University of Valencia, Spain
- Deasún Ó Conchúir, Scatterwork GmbH, Ireland
- Diego Andina De la Fuente, Technical University of Madrid, Spain
- Dragana Vilić, Ekonomski fakultet, Univerzitet u Banja Luci
- Dušan Ličina, École polytechnique fédérale de Lausanne (EPFL), Switzerland
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Egons Lavendelis, Riga Technical University, Latvia
- Gordana Pesaković, Argosy University, USA
- Hong Qi, Dalian University of Technology, China
- Ivan Bajić, Simon Fraser University, Canada
- Ina Bikuviene, PhD Kauno kolegija- University of Applied Sciences, Kaunas (Lithuania)
- Jovica V. Milanović, University of Manchester, United Kingdom
- Li Liwen, Beijing Foreign Studies University, Beijing, PR China
- Lorenzo Fagiano, Polytechnico di Milano, Italy
- Luis de la Torre Cubillo, UNED, Dpt. Informática y Automática, Madrid, Spain
- Luis Hernández Gómez, Technical University of Madrid, Spain
- Maarten De Vos, University of Oxford, United Kingdom
- Martin Kampel, Institute of Visual Computing & Human-Centered Technology, TUW, Austria
- Mike Dawney, Middlesex University, United Kingdom
- Moe Win, Massachusetts Institute of Technology, USA
- Mohammed Ismail Elnaggar, The Ohio State University, USA
- Nataša Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nellie Swart, University of South Africa, Pretoria
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Nuno Manuel Garcia dos Santos, University Beira Interior, Portugal
- Özge Ercan, Faculty of Sport Sciences – Sinop University-Türkiye
- Riste Temjanovski, Goce Delčev University, Macedonia
- Roberta Grossi, Horizons University, France
- Simona Distinto, Department of Life and Environmental Sciences University of Cagliari, Italia
- Slobodan Luković, ALaRI, Switzerland
- Snezana Lawrence, Bath Spa University, United Kingdom
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Vassilis S. Moustakis, Technical University of Crete, Greece
- Violeta Grubliene, Klaipeda University, Lithuania
- Vladimir Terzija, University of Manchester, United Kingdom

REVIEWERS COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mlađan Jovanović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Milan Tair, Singidunum University, Serbia
- Jelena Gajić, Singidunum University, Serbia
- Ivana Brdar, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Srđan Marković, Singidunum University, Serbia
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Gordana Pešaković, Argosy University, USA
- Nataša Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Slobodan Luković, ALaRI, Switzerland
- Vassilis S. Moustakis,Technical University of Crete, Greece
- Jovica V. Milanović, University of Manchester, United Kingdom

ORGANIZING COMMITTEE

- Mladen Veinović
- Milan Tair
- Nebojša Bačanin Džakula
- Marko Tanasković
- Zora Konjović
- Valentina Gavranović
- Miroslav Popović
- Nina Dragićević
- Miodrag Živković
- Jelena Gavrilović
- Miloš Antonijević
- Mlađan Jovanović
- Srđan Marković
- Marina Marjanović
- Lazar Dražeta
- Aleksandar Mihajlović
- Nikola Savanović
- Miloš Mravik
- Predrag Obradović
- Marijana Mihajlović
- Miloš Višnjić
- Aleksa Vidaković
- Romario Stanković

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publisher: Singidunum University, 32 Danijelova Street, Belgrade
Editor-in-Chief: Milovan Stanišić, PhD
Prepress: Miloš Višnjić, Marijana Mihajlović
Design: Aleksandar Mihajlović
Year: 2024
Circulation: 400
Printed by: Caligraph, Belgrade ISBN: 978-86-7912-821-8

Contact us:
Singidunum University
32 Danijelova Street, 11010 Belgrade, Serbia
Phone No. +381 11 3093220, +381 11 3093290,
Fax. +381 11 3093294
E-mail: sinteza@singidunum.ac.rs
Web: sinteza.singidunum.ac.rs

ABOUT SINTEZA 2024

The 11th international scientific conference Sinteza was held on May 16, 2024 and organized in person at Singidunum University premises. The conference was dedicated to information technology, computer sciences, data science and their application in engineering systems, education, teaching foreign languages, sports, and Environmental and Sustainability Sciences. This year conference topics of particular interest have been related to artificial intelligence, machine learning and data research, and their application in solving real-world problems.

The conference again brought together researchers from the country and abroad. A total of 78 works were submitted, 63 of which were accepted. All accepted papers for the Sinteza 2024 conference are scientific papers, and have been reviewed accordingly. Additionally, all the accepted papers have passed detailed technical, language, and content reviews as well as the iThenticate check.

At the plenary, six keynote speakers from Iraq, Switzerland, Spain, North Macedonia, and Serbia presented their research, project work, and findings predominantly in information technology and artificial intelligence. Various topics, such as artificial intelligence, data privacy, IoT, and robotics were presented. After the plenary session, the conference continued with 6 parallel sessions: Computer Science and Artificial Intelligence, Information Technology, Data Science and Applications, Advanced Technologies and Applications, Management and Technology, and a special Student Session. Each parallel session was interactive and dynamic, allowing presenters to present their research papers, case studies, and innovative projects, and the conference participants to discuss relevant issues and receive feedback from experts in the field.

This year, for the first time at the conference, there was a special Tech Talks session that delved into the world of technological innovations and provided invaluable insights into the latest trends, emerging technologies, and disruptive ideas shaping the future of IT.

We want to thank the esteemed speakers at the plenary session, all conference participants, and the members of the Scientific Committee. We want to express our special gratitude to the colleagues from the Organizing Committee who technically prepared and supported the organization of the Sinteza 2024 conference.

Sincerely,

Sinteza 2024 Organising Committee

CONFERENCE SPEAKERS IN PLENARY SESSION

Conference Chairmen:

Mladen Veinović, PhD – *Singidunum University, Serbia*

Valentina Gavranović, PhD – *Singidunum University, Serbia*

Opening Speach:

Goranka Knežević, PhD – *Rector of Singidunum University, Serbia*

Nebojša Bačanin Džakula, PhD – *Vice-Rector for Scientific Research, Singidunum University, Serbia*

Keynote speakers:

Tarik Ahmed Rashid, PhD – *Acting Dean of the School of Science and Engineering, Director of the Centre for Artificial Intelligence and Innovation, University of Kurdistan Hewler, Hewler, Kurdistan Region, Iraq*

Dušan Ličina, PhD – *Director of the Human-Oriented Built Environment Lab (HOBEL). École polytechnique fédérale de Lausanne (EPFL), Switzerland*

Marko Tanasković, PhD – *Dean of the Faculty of Technical Sciences,Singidunum University, Serbia*

Francisco (Paco) Florez-Revuelta, PhD – *Professor, Coordinator of visuAAL Marie-Skłodowska-Curie ITN, Chair of GoodBrother COST Action*

Ivan Chorbev, PhD – *Professor at the Faculty of Computer Science and Engineering Ss. Cyril and Methodius University in Skopje, Republic of North Macedonia*

Božidar Belić, PhD – *Director of Service and Development at Crayon Serbia*

CONTENTS

COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE SESSION

CHAIRMAN: Miodrag Živković

INFORMATION TECHNOLOGY SESSION

CHAIRMAN: Mlađan Jovanović

DATA SCIENCE AND APPLICATIONS SESSION

CHAIRMAN: Marina Marjanović

ADVANCED TECHNOLOGIES AND APPLICATIONS SESSION

CHAIRMAN: Marko Tanasković

MANAGEMENT AND TECHNOLOGY SESSION

CHAIRMAN: Lazar Dražeta

ENVIRONMENTAL DATA SCIENCE SESSION

CHAIRMAN: Miroslav Popović

INFORMATION TECHNOLOGY IN TEACHING FOREIGN LANGUAGES SESSION

CHAIRMAN: Valentina Gavranović

INFORMATION TECHNOLOGY IN SPORTS SESSION

CHAIRMAN: Srđan Marković

STUDENT SESSION

CHAIRMAN: Timea Bezdan

ARTIFICIAL INTELLIGENCE-BASED FRAMEWORK FOR ANALYZING CRISES-CAUSED AIR POLLUTION

Timea Bezdan[1*],
[0000-0001-6938-6974]
Mirjana Perišić[1,2],
[0000-0002-8287-4136]
Gordana Jovanović[1,2],
[0000-0001-8657-423X]
Nebojša Bačanin-Džakula[1],
[0000-0002-2062-924X]
Andreja Stojić[1,2]
[0000-0002-5293-9533]

[1]Singidunum University,
Belgrade, Serbia

[2]Institute of Physics Belgrade,
Belgrade, Serbia

Abstract:

Understanding the impact of air pollution processes during a crisis is crucial due to the significant risk to human health and to ensure global sustainability. Addressing this issue, this study introduces a novel artificial intelligence-based framework designed to analyze air pollution alterations caused by crises. The framework utilizes seven machine-learning regression models for making predictions: AdaBoost, CatBoost, ExtraTrees, Gradient Boosting, Histogram Gradient Boosting, LightGBM, and XGBoost regressor. Cross-validation is employed to ensure the robustness of the models and to prevent overfitting. The framework includes different metaheuristics algorithms, such as the Firefly Algorithm, Artificial Bee Colony, Harris Hawks Optimization, Sine Cosine Algorithm, Slime Mould Algorithm, and Quantum Superposition Algorithm. The top three performing ensemble models are optimized with the selected metaheuristic algorithm to find the optimal set of hyperparameters and to improve the results. After the optimization process, the best model is selected and evaluated on the dataset, then for explainability, SHAP and SAGE analysis are applied to provide deeper insight into the factors that influence the best model's predictions. These techniques ensure that the models are not only making precise predictions but also transparent and interpretable, which allows informed decision-making. Finally, the obtained results are visualized interactively for easier analysis of underlying patterns. This study lays the groundwork for a more effective crisis management system to mitigate the adverse of human health and environmental outcomes associated with air pollution caused by crises.

Keywords:

Artificial Intelligence, Machine Learning, Explainable Artificial Intelligence, Metaheuristics, Air Pollution.

INTRODUCTION

Air quality is a critical environmental factor that greatly impacts human health and global environmental sustainability. The quality of the air is directly linked to various health issues, including respiratory diseases, cardiovascular conditions, and overall well-being [1]. During crises, such as the COVID-19 pandemic and war, understanding and analyzing air quality becomes even more important due to the alterations of pollution levels. While extensive research has been conducted on air pollution and its impacts, there remains a need for innovative methodologies that can effectively characterize and predict air quality alterations during crises.

Correspondence:
Timea Bezdan

e-mail:
tbezdan@singidunum.ac.rs

DOI: 10.15308/Sinteza-2024-281-287

Traditional monitoring and modeling approaches often fail to comprehensively capture air pollution patterns. This study introduces a novel artificial intelligence-based framework designed to characterize and predict air quality alterations during crises, the framework is named crAIRsis. The framework includes seven ensembles of advanced machine learning regression models and incorporates different metaheuristic optimization to precisely adjust the model parameters, enhancing the precision of the prediction. Additionally, the incorporation of explainable artificial intelligence (XAI) is crucial for supporting the output of a model [2]; the deployment of XAI techniques, such as SHapley Additive exPlanations (SHAP) and Shapley Additive Global importance (SAGE) analyses, allows users to understand better and to trust the result of the model, as well as helps in informed decision-making. This research presents significant advancement in environmental and crisis management.

The remainder of the paper is organized as follows: Section 2 describes the machine learning models incorporated into crAIRsis framework; Section 3 details the optimization and evaluation processes, as well as the application of XAI techniques; Section 4 outlines the workflow of the framework, and Section 6 concludes the work and gives potential directions for future work.

2. MACHINE LEARNING MODELS

This section briefly describes the regression ensemble machine learning algorithms employed in crAIRsis framework: AdaBoost, CatBoost, ExtraTrees, Gradient Boosting, Histogram Gradient Boosting, LightGBM, and XGBoost regressor, highlighting their strengths and uniqueness. In general, ensemble methods combine multiple machine learning models to create a single model, which results in improved quality of the prediction and robustness [3]. Additionally, ensemble methods effectively reduce overfitting, which is crucial for prediction reliability.

AdaBoost [4], short for Adaptive Boosting, is an ensemble technique that combines multiple weak learners to form a single strong model, the method assigns equal weights to the data points, and then iteratively adjusts the weights of instances based on their error, in every other iteration the instances with higher error have higher weights, which gives more importance and improves the prediction over time. This continuous refinement is particularly advantageous in the prediction of air pollution alterations, where predictions can be highly variable due to fluctuating environmental factors. CatBoost [5] is a relatively novel gradient boosting algorithm, the algorithm automatically handles categorical features, effectively reduces overfitting with a novel gradient-boosting scheme. The algorithm has two important innovations, the introduction of ordered boosting, and handling categorical features. Light Gradient Boosting Machine, in short LightGBM, is a high-performance gradient boosting method that uses tree-based learning algorithms [6]. This method is particularly efficient on large datasets because of the utilization of Gradient-based One-Side Sampling technique that is introduced by the authors of LightGBM. The second innovative techniuque in the paper is the Exclusive Feature Bundling, which allows handling large number of features. The algorithm is very efficient in terms of memory consumption and computational speed. Extreme Gradient Boosting (XGBoost) is well-known ensemble machine learning algorithm, with great performance on different problems, in terms of speed and quality of prediction. The algorithm supports several loss functions and enhancements to the basic gradient boosting algorithm, including regularization features to prevent overfitting. Additionally, the algorithm can manage missing data. XGBoost constructs trees in parallel, unlike other traditional Gradient Boosting Decision Tree (GBDT) methods [7] which builds trees sequentially. Gradient Boosting (GB) model is incorporated from the sklearn Python package [8], it is a powerful machine-learning technique that produces a prediction model in the form of an ensemble of weak prediction models. GB builds the model in forward stage-wise fashion and generalizes it by allowing optimization of an arbitrary differentiable loss function. Histogram-based Gradient Boosting is an advanced implementation of the gradient boosting method that uses histograms for decision tree learning, which speeds up the training process and reduces memory usage by discretizing the continuous feature values into bins and using these bins to construct the decision trees. This method is particularly useful for processing large and complex datasets.

The framework uses cross-validation to rigorously evaluate the performance of each model, and to ensure that each subset of the dataset is used both for training and validation [9]. This method helps in generalization to the dataset and minimizes the risk of overfitting. The cross-validation setup includes random shuffling of the data to prevent any biases that may influence the results due to the ordering of data points. Each of the seven ensemble models has unique strengths, making them suitable for inclusion in the crAIRsis framework. After the seven models are trained and evaluated on the dataset, the top three performing algorithms are selected for the given problem and further optimized to achieve better results and select the final best model based on the evaluation criteria.

3. MODEL OPTIMIZATION, EVALUATION, AND INTERPRETATION

The quality of machine learning models highly depends on the values of its hyperparameters. Hyperparameter tuning is an optimization process, and it belongs to NP-hard problems, where metaheuristics are shown to be successful [10], [11], [12], hence the crAIRsis framework, utilizes metaheuristic optimization algorithms for fine-tuning the best three models' hyperparameters and enhancing the quality of prediction. The following metaheuristic algorithms are implemented in crAIRsis framework to efficiently explore and exploit the search space and find near-optimal solutions: Firefly Algorithm (FFA) [13], Artificial Bee Colony (ABC) [14], Harris Hawks Optimization (HHO) [15], Sine Cosine Algorithm (SCA) [16], Slime Mould Algorithm (SMA) [17], and Quantum Search Algorithm (QSA) [18]. After optimizing the three best models and the set of optimal hyperparameters are identified by using the selected metaheuristic algorithms, the final best model is selected. For evaluating the performance of the models, different regression metrics are used, providing different insights, specifically: mean absolute error, mean squared error, mean absolute percentage error, R-squared, explained variance, and max error are used for evaluation purposes and the best model is selected based on the R-squared value.

In the domain of artificial intelligence, to create a trustworthy system and have human-understandable model, why specific decisions and actions made by the models are very important. Consequently, after selecting the best performing model, crAIRsis uses XAI techniques for interpretability, explainability, and transparency; specifically SAGE [19] and SHAP [20] XAI methods are used.

SAGE represents a global interpretability method that measures the importance of each feature in the dataset. The method extends the Shapley value concept from game theory to feature importance in machine learning. SHAP values explain the prediction of an instance by calculating the contribution of each feature to the prediction. By analyzing SHAP values across the entire dataset, we gain insights into the general behavior of the model, identifying patterns and trends in feature contributions. After obtaining all results, the dataset, the obtained results of the model and XAI are visualized in an interactive web application to make the AI system result analysis and interpretation more user-friendly.

4. DATASET AND FRAMEWORK WORKFLOW

The crAIRsis AI-based is a multiapproach framework. First, the data is collected from different reliable sources. Before modeling, data preprocessing is conducted to prepare the data for the framework. The selected datasets are input into the framework. The user selects the target or more targets and the metaheuristic algorithms for optimization. Initially, the framework automatically creates folders for saving all results, and then starts the training and optimization process. In this process, separate models are created for the period before the crisis, during the crises and after the crises, combined by each target and measurement site. Afterwards, the dataset split is carried out in 80:20 ratio, 80% for training and 20% for testing purposes. The preprocessed data for the given measurement site, period, and target is evaluated by the ensemble machine learning algorithms, using 5-fold cross-validation, then cross-validation prediction is carried out and the metrics are calculated and saved. Based on the evaluation metrics, the best three models are selected and optimized by one or more metaheuristic algorithms. In this process, the optimization history is saved, as well as the result of the three optimized methods. In the next step, the best model is selected, based on the R-squared value.

In the second stage, by using the best model, the crAIRsis framework works on the explainable part. The impact of each feature is analyzed by SAGE's marginal imputer and permutation estimator, and the calculated values of global impact, sensitivity, and their standard deviation, along with absolute and relative measures are saved. To understand the impact of individual features and their interaction on model prediction, the framework uses SHAP to compute the absolute, relative [21], and normalized impact and saves the results for further analysis. In this process, the main effect and the interaction values are also calculated and saved. For deeper understanding and interpretation, cluster analysis is performed using the SHAP values. Uniform Manifold Approximation and Projection (UMAP) [22] used for dimensionality reduction and clustering is performed by Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [23], which allows identifying groups of similar data points and outliers. The framework saves the dimensionality-reduced data, cluster probabilities, and detailed statistics. The crAIRsis flowchart is depicted in Figure 1 and an examples of the visualizations are presented in Figure 2-6.

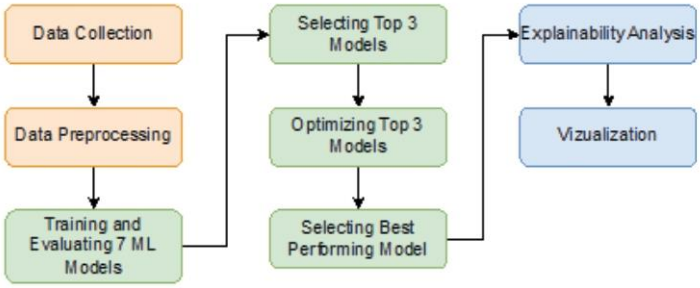


Figure 1. crAIRsis Framework Workflow.

Figure 2. crAIRsis Dashboard visualization.

Figure 3. crAIRsis Dashboard visualization – Time series plots.

Figure 4. crAIRsis Dashboard visualization – Model Evaluation.

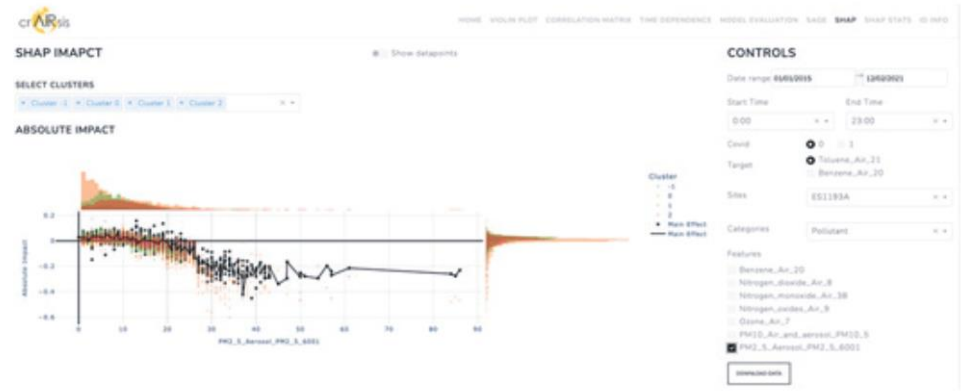


Figure 5. crAIRsis Dashboard visualization - SHAP.

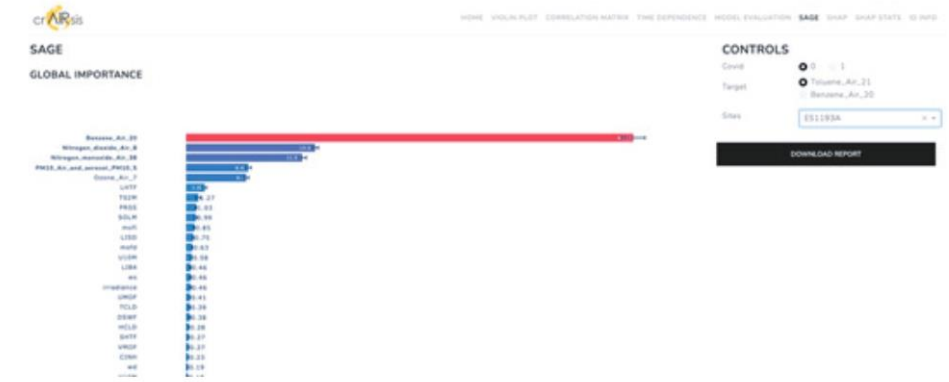


Figure 6. crAIRsis Dashboard visualization - SAGE.

5. CONCLUSION

This study introduces a robust artificial intelligence-based framework, crAIRsis for analyzing and predicting alterations in air pollution during crises. By integrating advanced ensamble machine learning models for regression, cross-validation and optimization techniques, the framework has demonstrated a high degree of reliability. Additionaly, the explainable AI techniques providea deep insights into the driving factors behind the models' actions and decision. The deployment of crAIRsis offers significant potential in making informed decisions during environmental crises. The interactive visualizations allow easier interpretations and analysis for practicioners.

As future work, the framework presents opportunities for expansion, such as implementing methods for classification problems, development, and implementation of hybrid metaheuristics for more efficient hyperparameter optimization. Additionally, there is a potential for integrating Large Language Models (LLMs) and automatize the interpretation. The current framework lays a solid foundation for analysing air pollution alterations during crises, the continued incorporation of novel AI methodologies promises further improvement.

6. ACKNOWLEDGEMENTS

This research was supported by the Science Fund of the Republic of Serbia, #7373, Characterizing crises-caused air pollution alternations using an artificial intelligence-based framework - crAIRsis

7. REFERENCES

[1] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and health impacts of air pollution: a review," *Front. Public Health*, vol. 8, p. 505570, 2020.

[2] A. B. Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020, doi: https://doi.org/10.1016/j.inffus.2019.12.012.

[3] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*, Springer, 2000, pp. 1–15.

[4] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997, doi: https://doi.org/10.1006/jcss.1997.1504.

[5] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.

[6] G. Ke et al., "Lightgbm: A highly efficient gradient boosting decision tree," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[7] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. Stat.*, pp. 1189–1232, 2001.

[8] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[9] S. Geisser, *Predictive inference*. Chapman and Hall/CRC, 2017.

[10] N. Bacanin, T. Bezdan, E. Tuba, I. Strumberger, and M. Tuba, "Optimizing Convolutional Neural Network Hyperparameters by Enhanced Swarm Intelligence Metaheuristics," *Algorithms*, vol. 13, no. 3, 2020, doi: 10.3390/a13030067.

[11] N. Bacanin et al., "Artificial Neural Networks Hidden Unit and Weight Connection Optimization by Quasi-Refection-Based Learning Artificial Bee Colony Algorithm," *IEEE Access*, vol. 9, pp. 169135–169155, 2021, doi: 10.1109/ACCESS.2021.3135201.

[12] N. Bacanin, T. Bezdan, M. Zivkovic, and A. Chhabra, "Weight Optimization in Artificial Neural Network Training by Improved Monarch Butterfly Algorithm," in *Mobile Computing and Sustainable Informatics*, S. Shakya, R. Bestak, R. Palanisamy, and K. A. Kamel, Eds., Singapore: Springer Nature Singapore, 2022, pp. 397–409.

[13] X.-S. Yang, "Firefly Algorithms for Multimodal Optimization," in *Stochastic Algorithms: Foundations and Applications*, O. Watanabe and T. Zeugmann, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 169–178.

[14] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Technical report-tr06, Erciyes university, engineering faculty, computer …, 2005.

[15] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Gener. Comput. Syst.*, vol. 97, pp. 849–872, 2019.

[16] S. Mirjalili, "SCA: A Sine Cosine Algorithm for solving optimization problems," *Knowl.-Based Syst.*, vol. 96, pp. 120–133, 2016, doi: https://doi.org/10.1016/j.knosys.2015.12.022.

[17] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Gener. Comput. Syst.*, vol. 111, pp. 300–323, 2020.

[18] J. Zhang, M. Xiao, L. Gao, and Q. Pan, "Queuing search algorithm: A novel metaheuristic algorithm for solving engineering optimization problems," *Appl. Math. Model.*, vol. 63, pp. 464–490, 2018, doi: https://doi.org/10.1016/j.apm.2018.06.036.

[19] I. Covert, S. M. Lundberg, and S.-I. Lee, "Understanding global feature contributions with additive importance measures," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 17212–17223, 2020.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[21] A. Stojić et al., "The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo [a] pyrene levels," *Chemosphere*, vol. 289, p. 133154, 2022.

[22] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *ArXiv Prepr. ArXiv180203426*, 2018.

[23] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering.," *J Open Source Softw*, vol. 2, no. 11, p. 205, 2017.

Singidunum University

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Sinteza

Belgrade, 2023

sinteza.singidunum.ac.rs

Singidunum University

Sinteza

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology and Data Related Research - Sinteza 2023 has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia.

Belgrade
May 27, 2023.
sinteza.singidunum.ac.rs

Sinteza

Singidunum University International Scientific Conference

SCIENTIFIC COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Endre Pap, Singidunum University, Serbia
- Aleksandar Jevremović, Singidunum University, Serbia
- Andreja Stojić, Institute of Physics in Belgrade, Serbia
- Bratislav Milovanović, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Đorđe Obradović, Singidunum University, Serbia
- Goran Šimić, Military Academy, University of Defence, Serbia
- Gordana Dobrijević, Singidunum University, Serbia
- Gordana Jovanović, Institute of Physics in Belgrade, Serbia
- Jelena Filipović, Faculty of Philology, University of Belgrade, Serbia
- Miloš Antonijević, Singidunum University, Serbia
- Marina Marjanović Jakovljević, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Marko Šarac, Singidunum University, Serbia
- Milan Milosavljević, Singidunum University, Serbia
- Milan Tuba, Singidunum University, Serbia
- Miloš Stojmenović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Mirjana Perišić, Institute of Physics in Belgrade, Serbia
- Miroslav Popović, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mladan Jovanović, Singidunum University, Serbia
- Nebojša Bačanin Džakula, Singidunum University, Serbia
- Nemanja Stanišić, Singidunum University, Serbia
- Nina Dragićević, Singidunum University, Serbia
- Petar Spalević, Faculty of Technical Sciences in Kosovska Mitrovica, University of Priština, Serbia
- Predrag Popović, Vinča Institute, Serbia
- Radosav Pušić, Faculty of Philology, University of Belgrade, Serbia
- Sanja Filipović, Singidunum University, Serbia
- Saša Adamović, Singidunum University, Serbia
- Tijana Radojević, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Alexandru Nedelea, Stefan cel Mare University of Suceava, Romania
- Aurora Pedro Bueno, Department of Applied Economics, University of Valencia, Spain
- Chen Yudong, Communication University of China, Beijing, People's Republic of China
- Deasún Ó Conchúir, Scatterwork GmbH, Ireland
- Diego Andina De la Fuente, Technical University of Madrid, Spain
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Egons Lavendelis, Riga Technical University, Latvia
- Gordana Pesaković, Argosy University, USA
- Hong Qi, Dalian University of Technology, China
- Ivan Bajić, Simon Fraser University, Canada
- Ina Bikuviene, PhD Kauno kolegija- University of Applied Sciences, Kaunas (Lithuania)
- Jesus Amador Valdés Díaz de Villegas, Iberoamericana University, Mexico
- Jovica V. Milanović, University of Manchester, United Kingdom
- Juan Ruiz Ramírez, University of Veracruz, Mexico
- Li Liwen, Beijing Foreign Studies University, Beijing, PR China
- Lorenzo Fagiano, Polytechnico di Milano, Italy
- Luis de la Torre Cubillo, UNED, Dpt. Informática y Automática, Madrid, Spain
- Luis Hernández Gómez, Technical University of Madrid, Spain
- Maarten De Vos, University of Oxford, United Kingdom
- María Magdalena Hernández Alarcón, University of Veracruz, Mexico
- Mike Dawney, Middlesex University, United Kingdom
- Moe Win, Massachusetts Institute of Technology, USA
- Mohammed Ismail Elnaggar, The Ohio State University, USA
- Nataša Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nellie Swart, University of South Africa, Pretoria
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Nuno Manuel Garcia dos Santos, University Beira Interior, Portugal
- Riste Temjanovski, Goce Delčev University, Macedonia
- Roberta Grossi, Horizons University, France
- Simona Distinto, Department of Life and Environmental Sciences University of Cagliari, Italia
- Slobodan Luković, ALaRI, Switzerland
- Snezana Lawrence, Bath Spa University, United Kingdom
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Vassilis S. Moustakis, Technical University of Crete, Greece
- Violeta Grubliene, Klaipeda University, Lithuania
- Vladimir Terzija, University of Manchester, United Kingdom
- Yipeng Liu, University of Electronic Science and Technology of China, China

REVIEWERS COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mladan Jovanović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Milan Tair, Singidunum University, Serbia
- Jelena Gajić, Singidunum University, Serbia
- Ivana Brdar, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Srđan Marković, Singidunum University, Serbia
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Gordana Pešaković, Argosy University, USA
- Nataša Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Slobodan Luković, ALaRI, Switzerland
- Vassilis S. Moustakis,Technical University of Crete, Greece
- Jovica V. Milanović, University of Manchester, United Kingdom

ORGANIZING COMMITTEE

- Milovan Stanišić
- Mladen Veinović
- Milan Tair
- Nebojša Bačanin Džakula
- Marko Tanasković
- Zora Konjović
- Valentina Gavranović
- Miroslav Popović
- Nina Dragićević
- Miodrag Živković
- Jelena Gavrilović
- Miloš Antonijević
- Mladan Jovanović
- Srđan Marković
- Marina Marjanović
- Aleksandar Mihajlović
- Nikola Savanović
- Miloš Mravik
- Predrag Obradović
- Marijana Mihajlović
- Miloš Višnjić
- Aleksa Vidaković
- Romario Stanković

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publisher: Singidunum University, 32 Danijelova Street, Belgrade
Editor-in-Chief: Milovan Stanišić, PhD
Prepress: Miloš Višnjić, Marijana Mihajlović
Design: Aleksandar Mihajlović
Year: 2023
Circulation: 85
Printed by: Caligraph, Belgrade ISBN: 978-86-7912-802-7

Contact us:
Singidunum University
32 Danijelova Street, 11010 Belgrade, Serbia
Phone No. +381 11 3093220, +381 11 3093290,
Fax. +381 11 3093294
E-mail: sinteza@singidunum.ac.rs
Web: sinteza.singidunum.ac.rs

ABOUT SINTEZA 2023

The 10th jubilee international scientific conference Sinteza was held on May 27, 2023, and it was organised in a hybrid mode - at Singidunum University premises and online. The conference was dedicated to information technology, computer sciences, data science and their application in engineering systems, education, teaching foreign languages, sports, marketing, ecology, medicine, and pharmacy. This year conference topics of particular interest have been related to artificial intelligence, machine learning and data research, and their application in solving real-world problems.

The conference once again brought together researchers from the country and abroad. A total of 63 works were submitted, 46 of which were accepted. All accepted papers have passed technical, language, and content reviews, as well as the iThenticate check. The authors came from 9 countries including Serbia.

At the plenary, six keynote speakers from the United States of America, Austria, the Russian Federation, Turkey, and Serbia presented their research and project work and findings that were predominantly in the field of information technology and artificial intelligence. Various topics, such as trends in artificial intelligence, the application of machine learning in different industries, and ethical issues related to artificial intelligence, were presented. After the plenary session, the conference continued with 6 parallel sessions: Computer Science and Artificial Intelligence, Information Technology, Advanced Technologies and Applications, Applied Information Technology, Information Technology in Sports and Information Technology in Foreign Language Teaching. The parallel sessions were interactive and dynamic, providing the presenters with the opportunity to present their research papers, case studies, and innovative projects, and the conference participants with the opportunity to discuss relevant issues and receive feedback from experts in their field.

We would like to thank the esteemed speakers at the plenary session, all conference participants, and the members of the Scientific Committee. We would like to express our special gratitude to the colleagues from the Organizing Committee who technically prepared and supported the organization of the Sinteza 2023 conference.

Sincerely,

Sinteza 2023 Organising Committee

CONFERENCE SPEAKERS IN PLENARY SESSION

Conference Chairmen:

Mladen Veinović, PhD – *Singidunum University, Serbia*
Valentina Gavranović, PhD – *Singidunum University, Serbia*

Opening Speach:

Goranka Knežević, PhD – *Rector of Singidunum University, Serbia*
Nebojša Bačanin Džakula, PhD – *Vice-Rector for Scientific Research, Singidunum University, Serbia*

Keynote speakers:

Milan Zdravković, PhD – *Faculty of Mechanical Engineering University of Niš, Serbia*
Blagoje Z. Đorđević, PhD – *Lawrence Livermore National Laboratory, California, United States*
Martin Kampel, PhD – *Institute of Visual Computing & Human-Centered Technology, TUW, Austria*
Alexander S. Geyda, PhD – *Russian presidential academy of national economy and public administration (RANEPA), St. Petersburg, Russian Federation*
Branislav Popović, PhD – *ECE University of Novi Sad, Serbia*
Özge Ercan, PhD – *Faculty of Sport Sciences – Sinop University-Türkiye*

CONTENTS

COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE SESSION

CHAIRMAN: Nebojša Bačanin Džakula

INFORMATION TECHNOLOGY SESSION

CHAIRMAN: Mlađan Jovanović

ADVANCED TECHNOLOGIES AND APPLICATIONS SESSION

CHAIRMAN: Marko Tanasković

APPLIED INFORMATION TECHNOLOGY SESSION

CHAIRMAN: Miodrag Živković

INFORMATION TECHNOLOGY IN SPORTS SESSION

CHAIRMAN: Srđan Marković

INFORMATION TECHNOLOGY IN TEACHING FOREIGN LANGUAGES SESSION

CHAIRMAN: Valentina Gavranović

REVEALING TOLUENE BEHAVIOUR IN THE ATMOSPHERE BASED ON COUPLING OF METAHEURISTICS, XGBOOST, AND SHAP

Gordana Jovanović[1,2*],
Mirjana Perišić[1,2],
Svetlana Stanišić[2],
Nebojša Bačanin-Džakula[2],
Andreja Stojić[1,2]

[1]Institute of Physics Belgrade,
Belgrade, Serbia

[2]University Singidunum,
Belgrade, Serbia

Abstract:

This study used an improved version of the reptilian search algorithm to investigate atmospheric patterns of toluene and its interactions with other polluting species under different environmental conditions. Toluene is a harmful aromatic hydrocarbon known for its role in the formation of secondary atmospheric pollutants. In this study, a two-year database of hourly pollutant concentrations, such as toluene, was analysed. The results were validated against other models using metaheuristic algorithms, and Shapley's additive explanations method was used to interpret them. The findings indicated a distinct correlation between toluene and m,p-xylene, and the study described the environmental conditions that influence their interactions. Overall, this research highlights the significance of using advanced analytical techniques to better understand the relationships between pollutants and their behaviour in different environmental conditions.

Keywords:

Machine Learning, Extreme Gradient Boosting, Metaheuristics, Explainable Artificial Intelligence, Volatile Organic Compounds.

INTRODUCTION

Air pollution involves intricate processes such as the dispersion, accumulation, or deposition of pollutants, which are affected by several factors, including pollutant interactions, unevenly distributed emission sources, measurement site characteristics, and meteorological conditions. To comprehend the behaviour of pollutants and their harmful effects on human health and the environment, data-driven research is essential. The complexity of air pollution-related processes necessitates an in-depth understanding of the underlying mechanisms, which can only be achieved through data-driven research.

Toluene is a mono-substituted aromatic hydrocarbon and its primary sources are traffic exhaust, cigarette smoke, and anthropogenic activities related to fuel, paint, adhesive, cleaner, polish, rubber, and lacquer production and use. Previous studies have shown that toluene concentrations range from 5 to 150 μg m^{-3} in urban locations, with extreme values near volatile pollutant sources [1]. Toluene is not prone to bioaccumulation and is rapidly absorbed, distributed throughout the body, and concentrated in vascularized organs, particularly the brain, due to its affinity for lipid-rich tissues [2].

Correspondence:
Gordana Jovanović

e-mail:
gordana.vukovic@ipb.ac.rs

DOI: 10.15308/Sinteza-2023-17-22

Toluene concentration is regularly monitored due to its toxicity on the nervous system and permanent brain damage observed in adhesive abusers [3]. Toluene is not labelled as carcinogenic, but in urban locations, anthropogenic benzene and toluene emission sources play a substantial role in ozone photochemistry and SOA-forming contribution, especially in low-NO_x regimes [4].

Our previous studies have tackled issues in analysing air pollution in complex urban environments, including the need for proper contextualization of data [5], [6], [7], [8], [9], and the use of statistical methods and artificial intelligence algorithms [10], [11] and [12]. In terms of data modelling, metaheuristic algorithms are commonly used to address nondeterministic polynomial (NP)-hard problems, particularly in machine learning hyperparameter optimization, due to their stochastic nature. In this study, we apply an enhanced variant of the reptile search algorithm (RSA) being hybridized with the firefly algorithm (FA) for resolving the shortcomings of the elementary RSA. The enhanced version of RSA metaheuristics is applied as an integral component of the machine learning framework to optimize the set of the XGBoost hyperparameters for toluene atmospheric fate research. The best-produced model is interpreted by applying Shapley Additive exPlanations (SHAP).

2. METHODOLOGY

2.1. DATA

For the analysis, we used the concentrations of inorganic gaseous pollutants (NO, NO_2, NO_x, O_3), particulate matter (PM_1, PM_{25}, and PM_{10}), and benzene, toluene, mp-xylene, and total non-methane hydrocarbons (TNMHC) obtained from the station of regulatory air quality monitoring Vatrogasni Dom in Pančevo, Serbia [13]. Additionally, meteorological parameters attained from the Global Data Assimilation System – GDAS1 [14], were used to complement the two-year (2019-2020) database of air pollutants. Hourly concentrations of organic pollutants (benzene, toluene, and m,p-xylene) and inorganic gaseous pollutants (NO, NO_2, NO_x, and O_3) were measured using referent sampling devices that adhere to European standards EN 14662-3, EN 14211, and EN 14625. The GRIMM EDM 180 measuring method was used to determine hourly concentrations of particulate matter, following the standards EN 12341 and EN 14907, while a gas chromatograph Syntech Spectras GC955 was employed for the concentrations of TNMHC. This device separates methane from other hydrocarbons and measures the concentration of both methane and other total non-methane hydrocarbons in the air.

2.2. STUDY AREA

Pančevo, with over 100,000 inhabitants, is situated on the left bank of the Danube, 20 km east and northeast of Belgrade, the largest Serbian metropolitan area. The sampling site (44°51′31″ N, 20°38′56″ E) is an urban background station located about 500 m south of the city centre at the regional fire station. The surrounding areas include residential areas to the east and northeast, a scrap metal sorting and storage centre, and a flour production factory. The E70, a European corridor with public transport and intensive vehicle flow, passes about 200 m in the S-SW direction from the sampling site. The confluence of the Tamiš and Danube rivers is located approximately 500 m in the SW direction. The South industrial zone of Pančevo, which includes three main factories: HTP Azotara, HTP Petrohemija, and Pančevo Oil Refinery, is situated two kilometres SE of the sampling station. The station is positioned in the dominant southeast direction of wind between the industrial zone and the city centre, according to the Air quality control program for the City of Pančevo and the Air Quality Plan for the City of Pančevo.

2.3. EXTREME GRADIENT BOOSTING – XGBOOST

XGBoost is a machine learning algorithm based on an ensemble of decision trees, where each tree is trained to correct the errors of the previous tree in the sequence. One of the key advantages of XGBoost is its ability to handle large datasets with high-dimensional features. It employs a regularization technique to prevent overfitting and can handle missing values in the data. The algorithm is highly customizable, allowing for the tuning of parameters such as learning rate, maximum depth, and number of trees to optimize performance. The details about XGBoost are provided elsewhere [15].

2.4. METAHEURISTICS

NP-hard challenges are a frequent occurrence that often requires the use of stochastic algorithms like metaheuristics because deterministic methods are impractical. Metaheuristic algorithms can be classified into various families based on the natural phenomena they imitate to guide the search process, such as evolution or insect behaviour [15]. The most significant families are nature-inspired methods (genetic algorithms and swarm intelligence), physical phenomenon-based methods (such as storms, gravity, and electromagnetism), algorithms that imitate human behaviour, and approaches based on mathematical laws.

Swarm intelligence is based on the coordinated and sophisticated behavioural patterns manifested by large groups of relatively modest units, such as insects or birds in swarms, while they hunt, feed, mate, or migrate [16]. These algorithms have proven highly efficient in solving various real-world NP-hard challenges. Well-known examples include particle swarm optimization (PSO) [17], ant colony optimization (ACO) [18], firefly algorithm (FA) [19] and bat algorithm (BA) [19]. More recently, highly efficient algorithms based on mathematical functions and their properties have emerged, such as the sine-cosine algorithm (SCA) [20] and arithmetic optimization algorithm (AOA) [21].

In this paper, we used a modified reptile search algorithm (RSA) inspired by crocodiles' hunting style [22] the RSA lacked sufficient exploitation power despite excellent exploration capability. We found the diversification-intensification trade-off balance biased towards exploration. We proposed integrating RSA with the FA to achieve a suitable balance between exploration and exploitation. The low-level hybrid approach combines both metaheuristics, with RSA at the start and FA during the search process to enhance the RSA's performance. The approach addressed RSA's weaknesses and improved its effectiveness in identifying optimal search regions.

2.5. SHAPLEY ADDITIVE EXPLANATIONS

To gain insight into the decision-making process of a best-performing model, we employed the explainable artificial intelligence SHAP (SHapley Additive exPlanations) method [23]. SHAP allows for a meaningful and straightforward interpretation of the decisions derived from the model, without sacrificing accuracy or interpretability. It is based on a game-theory approach that calculates Shapley values as a feature importance measure, which provides an understanding of the impact of each feature on individual predictions.

The Shapley values represent fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). SHAP assigns an important measure to each feature as a measure of its contribution to a particular prediction and compares its impact to the model's prediction if that feature took some baseline value (mean). This provides valuable insights into the model's behaviour by overcoming the main drawback of inconsistency, minimizing the possibility of underestimating the importance of a feature with a specific attribution value, and capturing feature interaction effects. However, the main challenges of the method include the computation of Shapley values and the choice of background data, which can lead to uncertain or unintuitive feature attributions.

3. RESULTS

As shown by mean absolute and relative SHAP values (Table1), the concentrations of benzene, followed by m,p-xylene levels, appeared to be the major factors that shape the toluene dynamic in the air. Additionally, the toluene's environmental fate is affected by the concentrations of THNMC, PM1, and NOX, as well as meteorological parameters, including volumetric soil moisture content (SOLM) and the direction and intensity of momentum flux (MOFD and MOFI). In the present study, to demonstrate the potential of the applied methodology, we will focus on m,p-xylene as the main predictor.

Table 1 - SHAP values.

	Benzene	mpXylene	TNMHC	PM_1	NO_x	SOLM	MOFD	MOFI	NO_2	NO	PM_{10}	T02M
Absolute SHAP	1.28	0.85	0.25	0.11	0.09	0.07	0.05	0.04	0.03	0.03	0.03	0.03
Relative SHAP [%]	36.09	27.27	8.31	3.06	3.05	2.43	2.01	1.49	1.12	1.23	0.94	0.91

The results suggest that m,p-xylene concentrations of 0.85 μg m^{-3} on average govern the toluene dynamics in the air as shown by absolute SHAP values. The most positive impact of m,p-xylene on the toluene dynamics is accompanied by the increase of m,p-xylene levels (up to 5 μg m^{-3}) as well as by the lowest concentrations of the inorganic gases, volatile non-aromatics and particles. The interrelationships between benzene homologues, m,p-xylene and toluene, could be explained by their coexistence in ambient air. Although toluene contains one methyl group which can be placed at any position on the benzene ring and m,p-xylene has two methyl groups attached to the benzene ring, they share common emission sources. The most dominant sources originated from anthropogenic activity including the petrochemical industry, chemical production of organic solvents and evaporative emission from storage facilities, but also to a lesser extent combustion of fossil fuels for heating and traffic purposes. The low to moderate levels of other pollutants in environmental conditions are associated with a negative impact of m,p-xylene, resulting in a decrease in toluene levels of up to 2.7 μg m^{-3}, while the highest concentrations of other pollutants are mostly associated with a moderate positive and negative impact on toluene dynamic (from -1.6 to 2 μg m^{-3}). Positive impact suggests the occasional influence of common sources of toluene and NOX and THNMC whereas negative interrelations imply different sources and the possible different behaviour of the pollutants in the air.

When PM is present in the air, the photochemical oxidation of aromatics such as toluene contributes to SOA formation (95%) compared to volatile organics and alkenes [24]. Similarly, Zhan et al. [25] reported that aromatics dominantly lead to the production of ground-level O_3.

The increase in m,p-xylene levels is linearly correlated with PM_1, $PM_{2.5}$, TNMHC, and NO_2 concentrations, but not with benzene and NO (Figure 1-3). High levels of NO were observed when lower levels of m,p-xylene were recorded (Figure 2), whereas higher concentrations of m,p-xylene corresponded to lower values of benzene (Figure 3) indicating the impact of different emission sources surrounding the measuring site. Additionally, an area of moderate influence of m,p-xylene on toluene is observed, when m,p-xylene concentrations are in the interval from 10 to 25 μg m^{-3}, and NO records high values - above 100 μg m^{-3} (Figure 2).

The analysis shows that low air and soil temperatures, high relative humidity, low PBLH, and stable atmospheric conditions as indicated by the other meteorological parameters, accompany the highest levels of all analysed polluted species. The conditions could be associated with the cold part of the year (winter and autumn months) when unfavourable meteorological conditions together with intensified fossil fuel burning for heating purposes contribute to high concentrations of pollutants. In addition, toluene and m,p-xylene removal through photochemically induced reactions are suppressed during the cold periods.

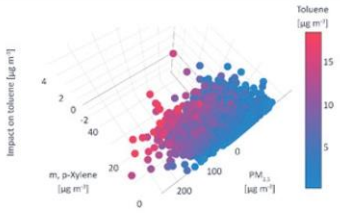


Figure 1 – Absolute m, p-xylene impact on toluene in the context of $PM_{2.5}$.

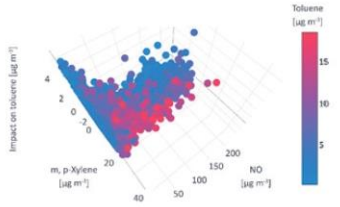


Figure 2 – Absolute m, p-xylene impact on toluene in the context of NO.

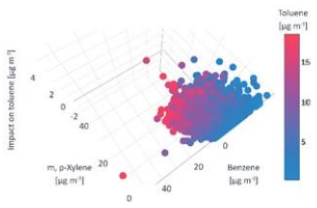


Figure 3 - Absolute m, p-xylene impact on toluene in the context of benzene.

4. CONCLUSION

Toluene is widespread in the atmosphere because it is used in many commercial products and when it is present at high concentrations, this pollutant poses serious adverse effects on human health. The behaviour of toluene in air is complex and its lifetime and abundance highly depend on environmental factors including meteorological conditions, the presence and intensity of emission sources and interactions with the other (co)polluting compounds. The present study using metaheuristics, XGBoost, and SHapley Additive exPlanations methods showed that concentrations of m,p-xylene mainly impact the dynamic of toluene in air. The positive interrelations between toluene and m,p-xylene could be linked with common emission sources and favourable values of temperature, humidity and PBLH.

5. ACKNOWLEDGEMENTS

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia GRANT No. #6524105, AI – ATLAS.

6. REFERENCES

[1] D. Murindababisha, Y. Abubakar, S. Yong, W. Chengjun and R. Yong, "Current progress on catalytic oxidation of toluene: a review," *Environmental Science and Pollution Research*, pp. 1-31, 2021.

[2] C. Davidson, J. Hannigan and S. Bowen, "Effects of inhaled combined Benzene, Toluene, Ethylbenzene, and Xylenes (BTEX): Toward an environmental exposure model," *Environmental toxicology and pharmacology*, vol. 81, p. 103518, 2021.

[3] N. P. Cheremisinoff and P. E. Rosenfeld, "Sources of air emissions from pulp and paper mills," *Handbook of pollution prevention and cleaner production*, vol. 2, pp. 179-259, 2010.

[4] G. Z. Whitten, H. Gookyoung, K. Yosuke and E. McDonald-Bul, "A new condensed toluene mechanism for Carbon Bond: CB05-TU," *Atmospheric Environment*, vol. 44, no. 40, pp. 5346-5355, 2010.

[5] L. Jovanovic, G. Jovanović, M. Perisic, F. Alimpic, S. Stanisic, N. Bacanin, M. Zivkovic and A. Stojic, "The Explainable Potential of Coupling Metaheuristics-Optimized-XGBoost and SHAP in Revealing VOCs' Environmental Fate," *Atmosphere*, vol. 14, no. 1, p. 109, 2023.

[6] A. Stojić, G. Vuković, M. Perišić, S. Stanišić and A. Šoštarić, "Urban air pollution: an insight into its complex aspects," in *A Closer Look at Urban Areas*, NY, USA, Nova Science Publishers, 2018.

[7] Š. Andrej, S. Stojić Stanišić, G. Vuković, Z. Mijić, A. Stojić and I. Gržetić, "Rainwater capacities for BTEX scavenging from ambient air," *Atmospheric Environment*, vol. 168, pp. 46-54, 2017.

[8] S. Andreja, N. Stanić, G. Vuković, S. Stanišić, M. Perišić and A. Šoštarić, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition," *Science of The Total Environment*, vol. 653, pp. 140-147, 2019.

[9] S. Stanišić, G. Jovanović, M. Perišić , S. Herceg-Romandić, T. Milićević and A. Stojić, "Explaining the Environmental Fate of PAHs in Indoor and Outdoor Environments by the Use of Artificial Intelligence," in *Polycyclic aromatic hydrocarbons Hauppauge*, NY, USA, Nov Science, 2022, pp. 1-36.

[10] A. Stojić, D. Maletić, S. Stojić Stanišić, Z. Mijić and A. Šoštarić, "Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods," *Sciente of the total environment,* pp. 19-26, 2015.

[11] M. Perišić, D. Maletić, S. Stojić Stanišić, S. Rajšić and A. Stojić, "Forecasting hourly particulate matter concentrations based on the advanced multivariate methods," *International Journal of Environmental Science and Technology*, vol. 14, no. 5, p. 1047–1054, 2017.

[12] S. Stanišić, M. Perišić, G. Jovanović, D. Maletić and D. Vudragović, "What Information on Volatile Organic Compounds Can Be Obtained from the Data of a SingleMeasurement Site Through the Use of Artificial Intelligence?," in *Artificial Intelligence: theory and Applications*, Springer, 2021, pp. 207-225.

[13] "SEPA," [Online]. Available: http://www.amskv.sepa.gov.rs/.

[14] "GDAS1," [Online]. Available: https://www.ready.noaa.gov/gdas1.php.

[15] C. Tianqi and C. Guestrin, "Xgboost: A scalable tree boosting system," in *The 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016.

[16] B. G., "Swarm intelligence," *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models,* pp. 791-818, 2020.

[17] J. Kennedy and Enneby, "Particle swarm optimization," in *ICNN'95-international conference on neural networks,* 1995.

[18] M. Dorigo, M. Birattari and T. Stutzle, "Ant colony optimization," *IEEE computational intelligence magazine,* vol. 1, no. 14, pp. 28-39, 2006.

[19] X. Yang, "Firefly algorithms for multimodal optimization," in *In Stochastic Algorithms: Foundations and Applications: 5th International Symposium, SAGA 2009*, Sapporo, Japan, 2009.

[20] S. Mirjalili, "SCA: a sine cosine algorithm for solving optimization problems," *Knowledge-based systems,* vol. 96, pp. 120-133, 2016.

[21] L. Abualigah, A. Diabat, C. Mirjalil and M. Abd Elaziz, "The arithmetic optimization algorithm," *Computer methods in applied mechanics and engineering*, vol. 376, p. 113609, 2021.

[22] L. Abualigah, M. Abd Elaziz, . P. Sumar, Z. Geem and Gandomi, "Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer," *Expert Systems with Applications*, vol. 191, p. 116158.

[23] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[24] J. Li, S. Deng, G. Li, Z. Lu, H. Song, J. Gao, Z. Sun and K. Xu, "VOCs characteristics and their ozone and SOA formation potentials in autumn and winter at Weinan, China," *Environmental Research*, vol. 203, p. 111821, 2022.

[25] J. Zhan, Z. Feng, P. Liu, X. He, Z. He, T. Chen, Y. Wang, H. He, Y. Mu and Y. Liu, "Ozone and SOA formation potential based on photochemical loss of VOCs during the Beijing summer," *Environmental Pollution*, vol. 285, p. 117444, 2021.

Singidunum University

Sinteza

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY AND DATA RELATED RESEARCH

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology and Data Related Research - Sinteza 2021 has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

Belgrade
Jun 25, 2021.
sinteza.singidunum.ac.rs

Sinteza

Singidunum University International Scientific Conference

SCIENTIFIC COMMITTEE

ORGANIZING COMMITTEE

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY AND DATA RELATED RESEARCH

Publisher: Singidunum University, 32 Danijelova Street, Belgrade
Editor-in-Chief: Milovan Stanišić, PhD
Prepress: Miloš Višnjić, Jovana Maričić
Design: Aleksandar Mihajlović
Year: 2021
Circulation: 80
Printed by: Caligraph, Belgrade
ISBN: 978-86-7912-755-6

Contact us:
Singidunum University
32 Danijelova Street, 11010 Belgrade, Serbia
Phone No. +381 11 3093220, +381 11 3093290,
Fax. +381 11 3093294
E-mail: sinteza@singidunum.ac.rs
Web: sinteza.singidunum.ac.rs

ABOUT SINTEZA 2021

International Scientific Conference SINTEZA provides an ideal platform for the exchange of information and dissemination of best practices, ideas and advancements in the state-of-the-art and technical improvements in the domain of Information Technology and Data Related Research.

Rapid advances in Information Technologies (IT) in recent decades have had a huge impact on numerous facets of everyday life and have created tremendous opportunities for economic, technological and social gains at a global scale. In particular, the advances in data-science, block-chain technology and optimization techniques are becoming the driving force behind many changes in both technology and business. Emergence of new technologies has caused widespread expansion of the internet of things. At the same time problems related to cyber security, security of communications, as well as the security in the cloud are becoming important topics.

New technologies and scientific breakthroughs have already altered the working and living environments making them safer, more convenient and more connected. These scientific advances are now also used for solving some of the most pressing problems our society is facing today, such as climate change and environmental issues.

The conference seeks submissions from academics, researchers, and industry professionals presenting novel research on all practical and theoretical aspects in the field of Information Technology and Data Related Research and their applications in a range of business, engineering, environmental and research fields. Traditionally held each year, conference features several prominent keynote speakers and presentations organized in thematic sessions covering topics such as computer science, information systems, IT security, applications of IT and data science in environmental engineering, education and sports. In addition, there is a special student session reserved for research work done by undergraduate students.

Sincerely,

Organising Committee of Sinteza 2021

CONFERENCE SPEAKERS IN PLENARY SESSION

Opening Keynote speakers:

Goranka Knežević, PhD – *Rector of Singidunum University, Serbia*
Nebojša Bačanin Džakula, PhD – *Vice-Rector for Scientific Research, Singidunum University, Serbia*

Keynote speakers:

Milan Gospić – *Country manager Serbia at Microsoft*
Vladimir Reković, PhD – *CERN, Particle Physics Researcher*
Miroslav Popović, PhD – *Singidunum University,Serbia and UC Berkeley, California*

CONTENTS

COMPUTER SCIENCE, COMPUTATIONAL METHODS, ALGORITHMS AND ARTIFICIAL INTELLIGENCE

ENVIRONMENTAL DATA SCIENCE SESSION

INFORMATION SECUITY AND ADVANCED ENGINEEING SYSTEMS SESSION

INFORMATION SYSTEMS, SOFTWARE DEVELOPMENT, INTERNET TECHNOLOGIES AND SOCIAL NETWORKING SESSION

INTERNET TECHNOLOGIES AND APPLICATIONS IN EDUCATION AND LANGUAGE LEARNING SESSION

IT APPLICATIONS IN SPORT SESSION

STUDENT SESSION

THE INFLUENCE OF COVID-19 LOCKDOWN ON BTEX LEVEL DISTRIBUTIONS IN BELGRADE

Nataša Bukumirić[1,2*],
Mirjana Perišić[2,3],
Svetlana Stanišić[2],
Gordana Jovanović[2,3],
Andreja Stojić[2,3]

[1]Academy of Technical Vocational study,
Belgrade, Serbia

[2]Singidunum University,
Belgrade, Serbia

[3]Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
Belgrade, Serbia

Abstract:
In this study, we have used the Standard Proton Transfer Reaction Mass Spectrometer (PTR-quad-MS) for online measurements of volatile organic compounds during the three-month campaign before, during, and after the state of emergency introduced as a preventive measure to the COVID-19 pandemic. The obtained data were analyzed by using correlations with hierarchical clustering, box plots, time variations, and bivariate polar plots with correlation and slope factor analysis, to provide better insight into the behavior and sources of the analyzed pollutants. As shown, pollutant concentrations have decreased only a week after the introduction of the curfew, and the benzene concentration dynamic was shown to be different compared to toluene, ethylbenzene, and xylenes behavior pattern.

Keywords:
Air Quality, BTEX, COVID-19, Lockdown, PTR-MS.

INTRODUCTION

In Spring 2020, the lockdown was implemented in many countries worldwide to prevent person-to-person SARS-CoV-2 virus transmission. During that period several studies have been performed in different countries to investigate the impact of prevention measures and restrictions on air quality.

The study of Jephcote et al. [1] registered a decline in monthly average traffic counts by 69%, which was reflected in the decrease of ozone, NO_2, and $PM_{2.5}$ concentrations by 7.6, 38.3, and 16.5%, respectively. However, it has been shown that traffic had a relatively modest contribution to air quality in the UK and meteorological conditions which were associated with the observed episodes of high particulate levels confirmed the importance of long-range transport and distant emission sources. The study of Mor et al. [2] aimed at investigating the relationships between 14 pollutant concentrations and meteorological factors during the four periods of lockdown, each of them lasting for 20 days, has confirmed the impact of local residential emission sources and regional atmospheric pollutant transport on local air quality.

Correspondence:
Nataša Bukumirić
e-mail:
nbukumiric@politehnika.edu.rs

DOI: 10.15308/Sinteza-2021-72-76

Sari and Esen [3] used data from 61 air quality monitoring stations in 31 cities to investigate the impact of restrictions on PM_{10} and SO_2 levels. Their results have shown that mean PM_{10} and SO_2 levels were decreased by 38.7% and 33.9%, and the observed effects of restrictions of both human and industrial activities on air quality were more pronounced than the effects of meteorological conditions.

As regards volatile organic compounds (VOCs), the study of Pakkattil et al. [4] examined the impact of lockdown on ground benzene, toluene, ethylbenzene, and xylenes (BTEX) levels in various metropolitan cities and according to the results, an enormous decline of 82% in BTEX concentrations was registered. However, despite the decline in BTEX levels and reduction of the ozone-forming BTEX potential, the corresponding decline in ozone concentrations was not observed. In the study of Kerimray et al. [5], the concentrations of PM2.5, NO2, SO2, CO, and O3 were decreased by 15 to 49%, however, the levels of benzene and toluene were 2-3 times higher than those registered during the previous years.

The pandemic-related measures and lockdown represented the sort of a real-world experiment that was used in many studies to derive important information and confirm conclusions that could enhance environmental policies and interventions in the future. In Serbia, preventive measures included restricted human mobility after 5 PM, during the weekend and on public holidays, except for medical personnel. Both human and industrial activities were minimized. In this study, we have investigated the impact of the most stringent introduced measures to air quality.

2. MATERIALS AND METHODS

The measurements of VOCs and meteorological parameters were conducted in an urban area of Belgrade, Serbia (44.86° N, 20.39° E). The measurement period (2nd March-2nd June 2020) covers two weeks before the introduction of the lockdown introduced as a response to the COVID-19 pandemic, nearly two months of curfew, and almost a month after the measures were lifted. Standard Proton Transfer Reaction Mass Spectrometer (PTR-quad-MS, Ionicon Analytik, GmbH, Austria) was used for online measurements of more than 230 masses [6], [7]. Meteorological parameters were measured using the Vaisala weather station. The calibration of PTR-MS measurements was done according to Taipale and coauthors [8] by using referent gases and a liquid calibration unit (Ionicon Analytik). The obtained data were analyzed by using correlations with hierarchical clustering, box plots, time variations, and bivariate polar plots with correlation and slope factor analysis [9]. Mobility trend reports were obtained from Google and Apple.

3. RESULTS AND DISCUSSION

Figure 1 shows BTEX concentrations and human activity change which were registered as a result of lockdown and curfew implemented for public safety and prevention of COVID-19 pandemic spread in Serbia.

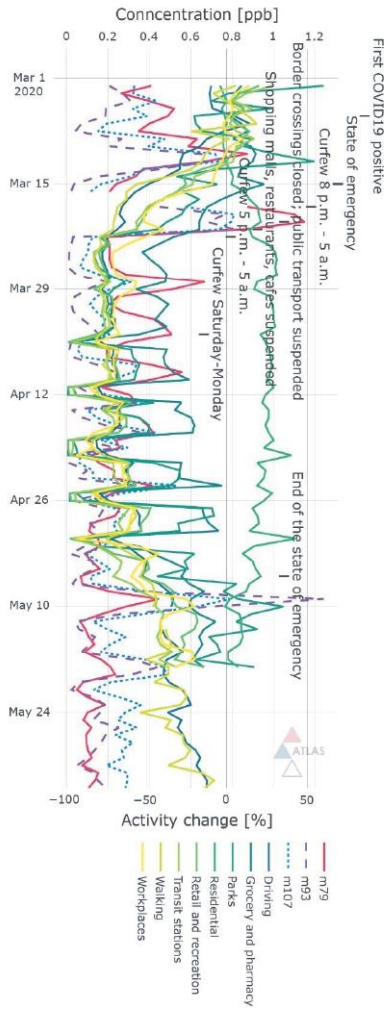


Figure 1 - BTEX concentration and human activity time series.

As can be seen, after the lockdown introduction, pollutant concentrations did not appear to reflect instant air quality change. It is worth noting that despite the abrupt cessation of human mobility and industrial activities, the beginning of the curfew period was characterized by an increase in BTEX concentrations. It cannot be excluded that the factors i.e., unfavorable meteorological conditions and the nature of the emission sources which govern the air quality in this part of the year could be responsible for the observed BTEX dynamics.

The first significant drop in BTEX concentrations was noticed a week later after the human mobility and industrial activities were minimized. The registered declines could be related to the first curfew periods that ranged, first for 9 hours (8 p.m.-5 a.m.), then 12 (5 p.m.-5 a.m.), and finally throughout all the weekend (Saturday-Monday). After several weeks of strict measures, a certain amount of human mobility was re-introduced, but BTEX levels continued to decline. The measurement campaign ended before the intensity of human activities returned to the common level.

In the period before the introduction of lockdown, the correlation analysis shows that the compounds registered at m/z 107 (ethylbenzene and total xylene) were in good correlation with compounds at protonated masses 93 (toluene) (r=0.93) and 79 (benzene) (r=0.9). This period was characterized by a good correlation between benzene and human activities such as spending time in retail and recreation (r=0.83) and transit stations (r=0.82), as well as between compounds registered at m/z 107 and activities in parks (r=0.86) (Figure 2). Among BTEX, the linear relationship was not observed only between benzene and toluene (r=0.73). During the lockdown, a strong correlation was observed between all compounds of the BTEX groups, with no significant correlations between BTEX levels and human mobility. After the lockdown, the relationships between all volatiles strengthened, but the correlations with human activities were not re-established.

Figure 3 represents the changes in mean BTEX concentrations during and after restrictions relative to the period before the state of emergency. As can be seen, the decrease in BTEX levels during the lockdown was in the range of 31 to 45%. The levels of volatiles increased after the human mobility and industrial activities were re-established, with exception of benzene which continued to decline up to 71% relative to the concentrations in the period before the pandemic. The box plot in Figure 4 also illustrates the decline in benzene concentrations with time. As shown, in the period after the lockdown, the 7^{5th} percentile of benzene concentrations was only around 0.15 ppb. In this period, the traffic intensity showed a stable increase, reaching the level that was only 15% lower than before the measures (Figure 1), which is not accompanied by an increase in benzene concentrations. This suggests that the contribution of traffic emissions to the total benzene levels was overestimated in the previous literature.

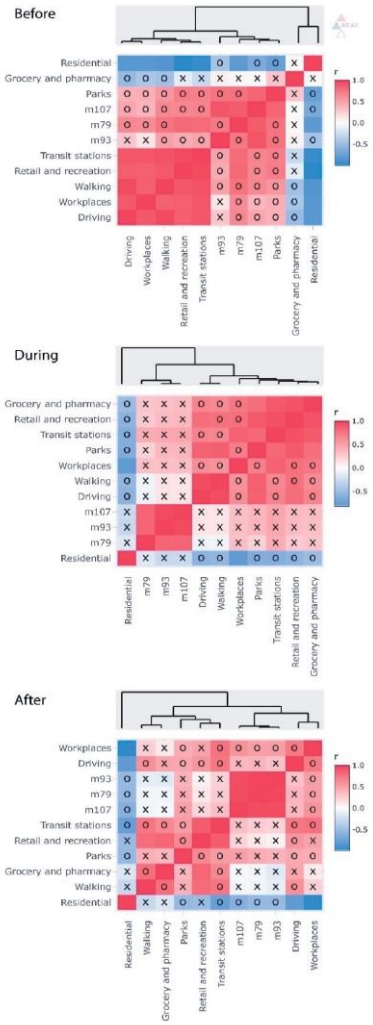


Figure 2 - Parameter value correlation matrix.

Figure 5 shows the dependence of the correlation and slope of toluene and benzene on wind parameters. The strong relationship between these compounds before restrictions indicates the dominant emission sources. The highest correlation (r=1) was recorded from all wind directions in the speed range from 1 m s-1 in the west to 8 m s-1 in the northeast from the measurement site.

The high ratio of toluene to benzene (T/B ratio > 2) suggests the existence of evaporative emissions (probably from industrial activities) being located in the north, northeast, south, and southwest.

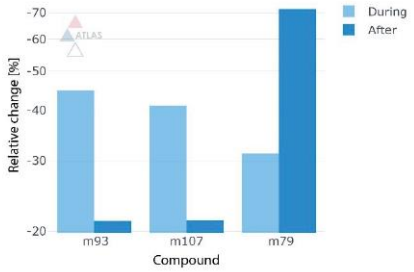


Figure 3 - Mean BTEX concentration difference during and after relative to the period before the state of emergency.

The dominance of the combustion process (T/B ratio < 2) was observed during the state of emergency, while after the lockdown period, the dominant evaporative emissions were restored, mainly in the southwest direction. This may indicate the reestablishment of the industrial activities and intense evaporations supported by higher temperatures in the period May-June.

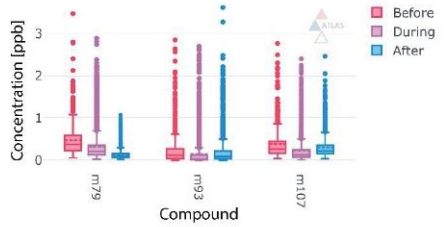


Figure 4 - BTEX box plots.

The daily and weekly variations in BTEX levels are shown in Figure 6. Daily variations (pronounced peaks early in the morning, late afternoon, and evening) indicated the expected distribution of VOC concentrations governed by meteorological parameters, photochemical processes, planetary boundary layer height evolution, and human activities in all three examined periods.

In contrast to the relatively stable daily dynamics, BTEX weekly variations changed over time. Before the lockdown, the highest BTEX levels were registered on Tuesday and Wednesday. During the lockdown, concentration peaks were displaced to Friday, while after the lockdown period, BTEX levels peaked on weekend.

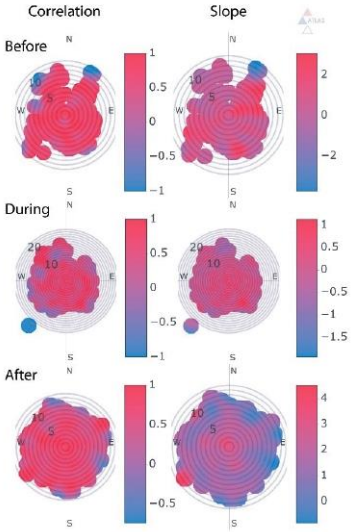


Figure 5 - Toluene and benzene correlation and slope dependency on wind parameters.

After the restrictive measures have ended, the pronounced BTEX peaks on weekends might be associated with travel, recreation activities, and staying outdoors, although based on the analysis of the time series (Figure 1), it is clear that increased human activities after the lockdown did not induce an increase in benzene concentrations, neither reestablishment of the correlations between BTEX compounds and human activity (Figure 2).

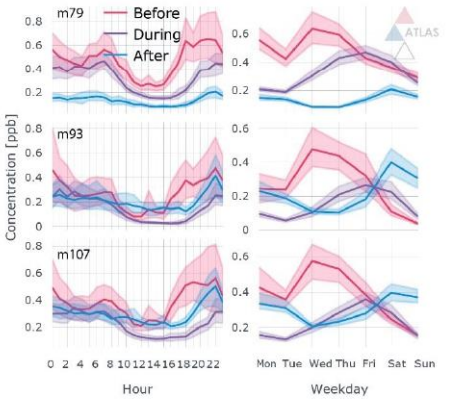


Figure 6 - BTEX diurnal and weekday variations.

As this type of analysis cannot indicate the main causes for the observed weekly variations in BTEX levels, it is necessary to approach more complex and precise analyzes.

4. CONCLUSION

While in the period before the introduction of the state of emergency and reduced human mobility, benzene and toluene levels exhibited no correlation, during and after the lockdown period strong correlations were observed between BTEX compounds, but no significant correlations between BTEX levels and human mobility were detected. An increase of human activities after the lockdown did not induce an increase in benzene concentrations, neither reestablishment of the correlations between BTEX compounds and human activities.

5. ACKNOWLEDGEMENTS

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia and the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI-ATLAS.

REFERENCES

[1] J. Calvin, A. Hansell, K. Adams i J. Gulliver, „Changes in air quality during COVID-19 'lockdown' in the United Kingdom," *Environmental Pollution*, t. 272, p. 116011, 2021.

[2] S. Mor, S. Kumar, T. Singh, S. Dogra, V. Pandey i K. Ravindra, „Impact of COVID-19 lockdown on air quality in Chandigarh, India: understanding the emission sources during controlled anthropogenic activities," *Chemosphere*, t. 263, p. 127978, 2021.

[3] M. F. Sari i F. Esen, „Effect of COVID-19 on PM10 and SO2 concentrations in Turkey," *Environmental Forensics*, pp. 1-10, 2021.

[4] A. Pakkattil, M. Muhsin i R. Varma, „COVID-19 lockdown: Effects on selected volatile organic compound (VOC) emissions over the major Indian metro cities," *Urban Climate*, t. 37, p. 100838, 2021.

[5] A. Kerimray, N. Baimatova, O. Ibragimova, B. Bukenov, B. Kenessov, P. Plotitsyn i F. Karaca, „Assessing air quality changes in large cities during COVID-19 lockdowns: The impacts of traffic-free urban conditions in Almaty, Kazakhstan," *Science of the Total Environment*, t. 730, p. 139179, 2020.

[6] W. Lindinger, A. Hansel and A. Jordan, "On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research," *International Journal of Mass Spectrometry and Ion Processes*, vol. 173, no. 3, pp. 191-241, 1998.

[7] A. Stojić, S. Stanišić Stojić, Z. Mijić, A. Šoštarić i S. Rajšić, „Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling," *Atmospheric Environment*, t. 106, pp. 71-79, 2015.

[8] R. Taipale, T. M. Ruuskanen, J. Rinne, M. K. Kajos, H. Hakola, T. Pohja i M. Kulmala, „Quantitative long-term measurements of VOC concentrations by PTR-MS–measurement, calibration, and volume mixing ratio calculation methods," *Atmospheric Chemistry and Physics*, t. 8, br. 22, pp. 6681-6698, 2008.

[9] S. Grange, A. Lewis i D. Carslaw, „Source apportionment advances using polar plots of bivariate correlation and regression statistics," *Atmospheric Environment*, t. 145, pp. 128-134, 2016.

METEOROLOGICAL FACTORS GOVERNING PARTICULATE MATTER DISTRIBUTION IN AN URBAN ENVIRONMENT

Mirjana Perišić[1,2],
Gordana Jovanović[1,2],
Svetlana Stanišić[2*],
Andrej Šoštarić[3],
Andreja Stojić[1,2]

[1]Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
Belgrade, Serbia

[2]Singidunum University,
Belgrade, Serbia

[3]Institute of Public Health Belgrade,
Belgrade, Serbia

Abstract:
In this study, the impact of meteorological factors on PM10 concentrations in the Belgrade urban area was investigated by using eXtreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP) attribution methods. As shown, XGBoost provided reliable PM10 predictions with relative errors in the range from approx. 19% to 26% and correlation coefficients higher than 0.95. The change in emission source intensity, momentum flux intensity, lifted index, humidity, and temperature, as well as concentrations of benzene, NO, NO_x and SO_2 were the most important variables that described the PM concentration dynamics in Belgrade urban area.

Keywords:
Particulate Matter, Meteorological Factors, Machine Learning, Explainable Artificial Intelligence.

INTRODUCTION

Suspended particulate matter refers to a complex mixture of compounds in a solid and liquid state, of organic and inorganic origin. Depending on the size, they are characterized as small/fine or $PM_{2.5}$ (with a diameter of up to 2.5 μm) and large/coarse fraction or PM_{10} (with a diameter of 2.5 μm to 10 μm). In the short run, the consequences of exposure to high concentrations of PM are irregular heartbeat and bronchial asthma exacerbation. In the long run, the adverse health effects include reduced lung capacity, increased risk of malignant diseases, increased susceptibility to systemic inflammation, as well as diabetes and its complications, exacerbation of chronical conditions, higher susceptibility to infectious viral or bacterial diseases, and increased risk of atherosclerosis and its consequences, heart attack and stroke. In addition to the impact on human health, PM has effects on the environment and other living beings. For instance, it has been widely recognized that PM contributes to the formation of acid rain, which changes the acidity of freshwater systems, reduces soil fertility, damages plant species and agricultural crops, threatens biodiversity and endangers world cultural heritage.

Correspondence:
Svetlana Stanišić

e-mail:
sstanisic@singidunum.ac.rs

DOI: 10.15308/Sinteza-2021-89-93

In this study, we used regression analysis by means of machine learning eXtreme Gradient Boosting method (XGBoost) for estimating the relationships between PM_{10} concentrations and a number of environmental parameters in Belgrade, Serbia [1]. The influence of meteorological factors on PM_{10} concentrations in the Belgrade urban area was investigated and explained by using SHapley Additive exPlanations (SHAP) attribution method [2]. The provided methodology has already been approved in several case studies [3], [4], [5].

2. MATERIALS AND METHODS

The ground-based data, including benzene, inorganic gaseous pollutants (SO_2, NO, NO_2, NO_x), were provided by the Institute of Public Health Belgrade, Serbia. Meteorological data were provided by the Global Data Assimilation System (GDAS1).

The relationships between PM_{10} and other environmental parameters were obtained by XGBoost. XGBoost is an ensemble method of supervised machine learning based on a sequential tree growing algorithm. Each decision tree aims to complement all the others and correct for residuals in the predictions made by the previous trees by iteratively reweighing the training data to improve regression performance. XGBoost uses a gradient descent algorithm to minimize loss when adding new models. The method includes many optimizations and enhancements. The dataset was split into training (80%) and validation (20%) sets. Hyperparameter tuning was implemented using a brute-force grid search and 10-fold stratified cross-validation. The best performing hyperparameter values were used for the final model.

SHapley Additive exPlanations (SHAP) is a method based on Shapley values, calculated as a measure of feature importance using a game-theory approach that provide an impact of features on individual predictions. SHAP values represent the only possible locally accurate and globally consistent feature attribution method.

In this paper, XGBoost and SHAP method implementations within the Python software environment were used.

3. RESULTS AND DISCUSSION

XGBoost provided reliable PM10 predictions with relative errors in the range from approx. 19% to 26% and correlation coefficients higher than 0.95 (Figure 1). The best performing model with the lowest relative error and the highest correlation coefficient was obtained for the monitoring station of rural/industrial type located in Ovča.

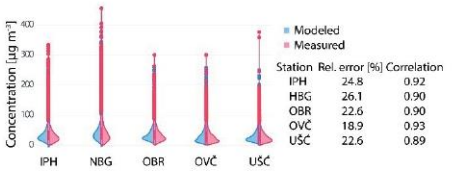


Figure 1 - XGBoost model evaluation.

On the other hand, the highest difference between modeled and measured values were detected at urban-type monitoring stations Novi Beograd and the Institute of Public Health Belgrade, both of which are exposed to traffic emissions. The modeling results were not satisfying for rural/industrial monitoring station located in Veliki Crljeni (relative error>30%, correlation coefficient<0.8), which implies that the PM10 level dynamic was mostly governed by variables other than available pollutant concentrations and meteorological parameters.

PM_{10} concentrations in Belgrade were predominantly determined by a variable that is defined as a trend of changing the intensity of emission sources (Figure 2).

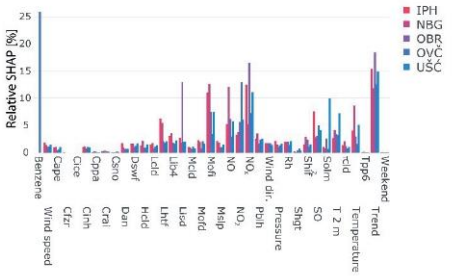


Figure 2 - Feature importance.

This variable appears to be the most important at three monitoring locations (Institute of Public Health Belgrade, Obrenovac, and Ušće), while at the stations New Belgrade and Ovča it was among the first three most significant.

Meteorological parameters including momentum flux intensity – Mofi, standard lifted index – Lisd, volumetric soil moisture content – Solm, and temperature, as well as pollutants such as benzene, NO, NO_x, and SO_2, were among the five most important variables that described the dynamics of suspended particulate matter in the territory of Belgrade area.

3.1. VOLUMETRIC SOIL MOISTURE CONTENT

In urban areas, suspended particulate matter, benzene, nitrogen oxides and SO_2 originate from common anthropogenic sources that include emissions from traffic and industrial activities, as well as the combustion of fossil fuels in thermal power plants, heating plants, and households. After the emission, the pollutants are subject to a variety of physical, chemical, and photochemical reactions. Suspended particulate matter, benzene, nitrogen oxides, and SO_2 participate in the formation of secondary atmospheric aerosols. A number of processes take place on the surface of suspended particulate matter, including gas-particles conversion, adsorption, desorption, absorption and gas dissolution, condensation of volatile compounds, as well as nucleation and coagulation. Under conditions of increased humidity in the presence of soot and inorganic oxides as catalysts (for example MgO_2 or Fe_2O_3), SO_2 will be adsorbed on the surface of suspended particulate matter to form a secondary sulfate aerosol. On the other hand, nitrogen oxides are less soluble in water compared to SO_2, so they will be less adsorbed on the surface of the particles (Figure 3). Nevertheless, in the conditions of high temperatures and intense solar activity, nitrogen oxides and volatile organic compounds such as benzene will rather participate in photochemical reactions with hydroxy, peroxy, and organic radicals in the air in which tropospheric ozone is formed.

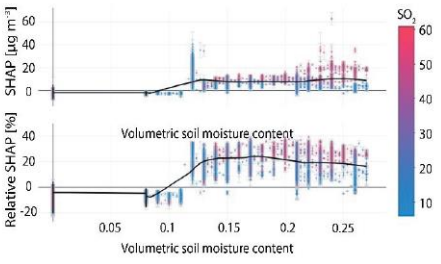


Figure 3 - PM_{10} SHAP dependency on volumetric soil moisture content and SO_2.

3.2. STANDARD LIFTED INDEX

The lifted index indicates the degree of atmospheric stability. The temperature in the atmosphere decreases with an increase in altitude, and the air that rises from the surface of the ground cools. However, when a temperature inversion occurs, air that rises to higher altitudes is warmer than the one near the ground level, which can lead to atmospheric instability. At all measuring points included in the analysis, there was a significant influence of maximum positive values of this parameter on PM_{10} concentrations (on average about 8 µg m^{-3}), which indicates that the dynamics and transformations of PM_{10} depended on atmospheric stability, Figure 4.

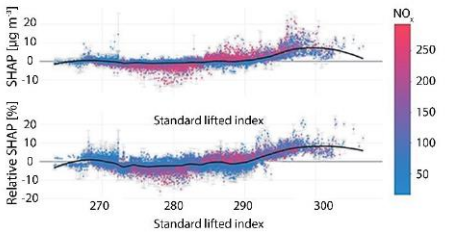


Figure 4 - PM_{10} SHAP dependency on standard lifted index and NO_x.

3.3. MOMENTUM FLUX INTENSITY

For the forecast and assessment of meteorological conditions, the momentum flux intensity is usually observed together with the wind speed. This parameter provides information important for understanding airflow in the vertical structure of the atmosphere. It can also be used to assess the stability of air mass flows in the planetary boundary layer and the occurrence of turbulent transmissions and vortices. Under stable meteorological conditions, the values of this parameter do not change significantly from the surface to the higher layers of the atmosphere and usually have lower values compared to the values measured in the case of turbulent movements. High SHAP values corresponding to increasing concentrations of suspended particulate matter up to several tens of µg m^{-3} at lower values of momentum flux intensity (<0.2) indicate a significant influence of vertical movements on the dynamics of PM_{10} and other pollutants

(NO) when they are present in high concentrations in the air (Figure 5). At all monitoring locations, a significant impact of this parameter is recorded at its lower values, which indicates that the stated relationships between pollutants are observed in conditions of stable meteorological conditions.

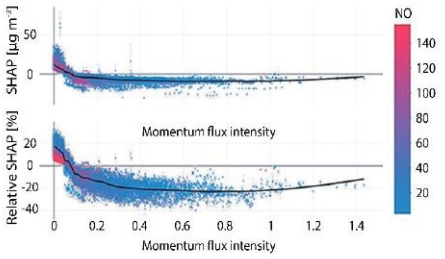


Figure 5 - PM_{10} SHAP dependency on momentum flux intensity and NO.

3.4. TEMPERATURES

The intensive combustion of fossil fuels for heating at temperatures below zero contributes to an increase in PM_{10} concentrations by an average of 10 µg m^{-3} (Figure 6). In the case of using fuels with high sulfur content, this increase can be as high as 20 µg m^{-3}. From only a few degrees above zero to about 25 °C, the effect of temperature on the suspended particulate matter is negligible, while during warmer weather, at temperatures above 25 °C, the resuspension of particles contributes to an increase in concentrations of about 4 µg m^{-3} on average.

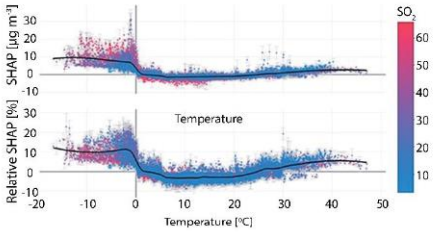


Figure 6 - PM_{10} SHAP dependency on temperature and SO_2.

3.5. ATMOSPHERIC PRESSURE

The effect of pressure on the concentrations of suspended particulate matter is relatively small and constant (Figure 7). Somewhat stronger impact on their level dynamics is recorded in the urban atmosphere being characterized by the presence of NO_2 higher concentrations. Low-pressure conditions can contribute to a reduction in PM concentrations of up to 3 µg m^{-3}.

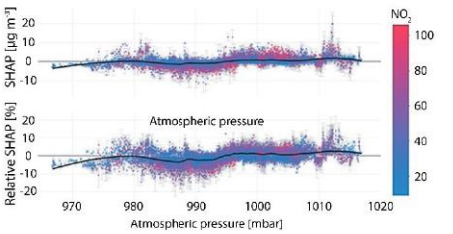


Figure 7 - PM_{10} SHAP dependency on atmospheric pressure and NO_2.

4. CONCLUSION

Based on the analysis of the dependence of suspended particulate matter concentrations on environmental factors (concentrations of SO_2, NO, NO_2, NO_x, and benzene, modeled meteorological parameters - GDAS base, trend, daily and weekend variations), the change in emission source intensity is singled out as a variable that dominantly determines the dynamics of PM_{10} concentration in Belgrade. This variable stands out as the most important one in three measuring points - Institute of Public Health Belgrade, Obrenovac, and Ušće. Also, meteorological parameters including momentum flux intensity, lifted index, humidity, and temperature, as well as concentrations of benzene, NO, NO_x, and SO_2 were among the five most important variables that described the PM concentration dynamics in Belgrade urban area.

5. ACKNOWLEDGEMENTS

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105,

AI-ATLAS, as well as the City of Belgrade, Department of Environmental protection of the city administration, Serbia, Air quality plan for the City of Belgrade.

REFERENCES

[1] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco California USA, 2016.

[2] S. Lundberg, G. Erion, H. Chen, A. DeGrave, J. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature machine intelligence*, vol. 2, no. 1, pp. 56-67, 2020.

[3] A. Stojić, D. Maletić, S. Stanišić Stojić, Z. Mijić and A. Šoštarić, "Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods," *Science of the Total Environment*, vol. 521, pp. 19-26, 2015.

[4] A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić and L. Lazić, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition," *Science of The Total Environment*, vol. 653, pp. 140-147, 2019.

[5] M. Perišić, D. Maletić, S. Stanišić Stojić, S. Rajšić and A. Stojić, "Forecasting hourly particulate matter concentrations based on the advanced multivariate methods," *Journal of Environmental Science and Technology*, vol. 14, no. 5, pp. 1047-1054, 2017.

RECEPTOR ORIENTED MODELING FOR REVEALING AIR POLLUTION EMISSION SOURCES AFFECTING AN URBAN AREA

Svetlana Stanišić[1]*,
Mirjana Perišić[1,2],
Gordana Jovanović[1,2],
Andreja Stojić[1,2]

[1]Singidunum University,
Belgrade, Serbia

[2]Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
Belgrade, Serbia

Abstract:
In this study, we have determined PM (particulate matter) emission sources and some of the criteria air pollutant transport contribution at various locations in the Belgrade area by applying advanced receptor-oriented models, as well as the pre-processing of concentrations and air back trajectories. As shown, the monitoring locations were most directly exposed to PM emissions from the nearest surrounding. Further, the background levels and air pollution transport mostly contributed to the observed SO_2 (70%) and NO_2 levels (45%).

Keywords:
Particulate Matter, Air Pollution Transport, Receptor Oriented Models.

INTRODUCTION

Low air quality represents a particular problem in urban areas due to overpopulation, a large number of emission sources, and topographic features which prevent the dispersion of pollution. The cities, in which around 85% of global economic activity takes place, currently contain 55% of the world's population, and it is expected that two-thirds of the world's population will live in metropolitan areas by 2050. The World Health Organization estimates that the highest number of deaths related to atmospheric pollution was registered as a consequence of ischemic cardiovascular diseases, heart attacks and strokes (80%), and chronic obstructive pulmonary disease (11%), while a significantly lower number of deaths occurred as a consequence of lung cancer (6%) and acute inflammation of the lower respiratory tract in children (3%). The health effects of air pollutants vary depending on the type of pollutant, i.e., size and composition of suspended particles, the concentration of species, and the length of exposure.

Correspondence:
Svetlana Stanišić

e-mail:
sstanisic@singidunum.ac.rs

DOI: 10.15308/Sinteza-2021-94-97

The concentrations of pollutants in the air on the territory of Belgrade area are a consequence of intensive emissions mainly from local anthropogenic sources, which can be related to the increase in population, in the number of motor vehicles, inadequate investment in the energy sector, and outdated technologies in the economic sector.

In terms of sources of pollutant emissions in the city, the following can be emphasized as significant: fossil fuel burning for energy production (heating plants, thermal power plants, boiler rooms, individual furnaces, i.e., around 300,000 individual chimneys), some industrial facilities, traffic, as well as small and medium production processes.

On the other hand, air circulation in complex topographic and meteorological conditions of the urban environment potentially leads to long retention or accumulation of pollution in certain locations, which further causes large differences in the exposure of the population in spatially close locations.

The aim of this paper is to determine emission sources of suspended particulate matter at various locations in Belgrade area by applying advanced receptor-oriented models, as well as the pre-processing of concentrations and air back trajectories.

2. MATERIALS AND METHODS

The analysis of regional transport and the assessment of pollutant emission sources was conducted by using receptor-oriented models developed within the project "Mapping of sources of toxic, mutagenic, and carcinogenic volatile organic compounds in the city of Belgrade", funded by the Green Fund of the Ministry of Environmental Protection of Serbia. The description of the methods can be found elsewhere [1].

The analysis of the contribution of regional transport was done by using the method of concentration weighted boundary layer - CWBL [2]. The method provides data on the three-dimensional distribution of pollutants based on the measured concentrations at the receptor site (measurement site), the air mass transport path and the height of the planetary boundary layer along the transport path. Based on [3], [4], [5], [6], [7], and [8], using CWBL, it is possible to estimate the regional transport of pollutants within the planetary boundary layer by determining concentrations at higher altitudes above the Earth's surface. The description of the method is presented elsewhere [2].

3. RESULTS AND DISCUSSION

Within the analysis of air quality, it is crucial to separate the different contributions to the total measured concentrations at the selected monitoring site. One way to do it is to distinguish between the contribution of emissions from local sources in the immediate vicinity of the measuring location, the contribution of regional and long-range transport, and the share of air pollution background. As can be seen in the time series of PM_{10} and SO_2 concentrations (Figure 1), narrow and high peaks are superimposed on a wider and much lower base level. The peaks probably originate from the local emission in the immediate vicinity of the measuring point, whereas the baseline level can be assumed to originate from the transported air pollution and the background.

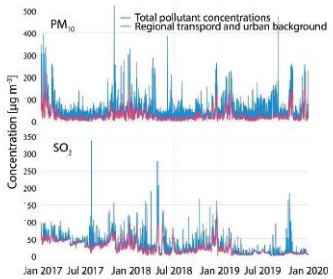


Figure 1 - An example of separating the contribution of emissions from local sources from regional transport and the background of the urban environment at the measuring point of the City Institute for Public Health Belgrade in the period from 2017 to 2019

The example shown in Figure 1 shows a different structure of PM_{10} and SO_2 concentration time series. Unlike PM_{10}, SO_2 concentrations are characterized by the less frequent occurrence of narrow peaks superimposed at the baseline level. This can be an indicator of the high contribution of background and/or regional transport to the total concentrations in the urban environment. The most important sources of SO_2 in urban areas are related to the combustion of fossil fuels for heating purposes. The position of stationary, point sources (chimneys) that are more distant and mostly distributed in a wider area without a direct impact on the monitoring station leads to the less pronounced pollutant concentration dynamics. In addition to the specificity of the emission sources, the position of the

monitoring station at the Institute for Public Health Belgrade in the canyon type street can also be the cause of high levels of urban background due to the retention and accumulation of air pollution.

The share of regional transport and background averaged at all monitoring locations of automatic monitoring (Figure 2) is the highest in the case of SO_2 when compared to all other analyzed pollutants (70%). The estimated contribution of regional transport and the background to the measured concentrations of suspended particles and nitrogen oxides is moderate and ranges from 45% to 55%. In the case of suspended particles, the existence of frequent short-term peaks in the time series (Figure 1) is an indicator of the dominance of local emission sources. The reason for this dynamics can be the direct exposure of the monitoring station to a certain type of emissions (mobile sources - traffic and transport, resuspension, and local economic activities), but also the processes of dry and wet deposition that contribute to faster removal of particles from the air. Of nitrogen oxides, it was estimated that the share of regional transport and air pollution background is the highest in the case of NO_2, which is a consequence of greater stability of the compound and therefore, the possibility of its transport over long distances, but also the formation of this compound as a secondary pollutant in the reactions of photochemical transformations in the atmosphere.

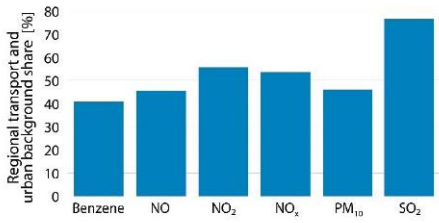


Figure 2 - The share of regional transport and the background of the urban environment with the measured concentrations of pollutants in Belgrade in the period from 2017 to 2019

By applying multireceptor-oriented models to PM10 concentrations measured at 6 automatic monitoring locations in the period from 2017 to 2019, the distribution of regional sources and sources located on the periphery of the agglomeration, which affect air quality in central urban area was obtained (Figure 3).

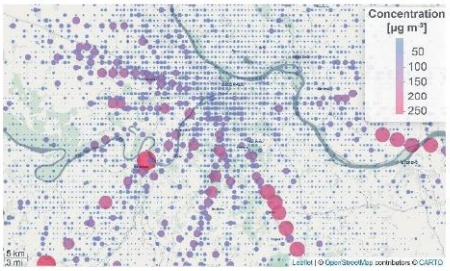


Figure 3 – The distribution of regional sources of PM_{10} emissions on the territory of Belgrade and neighboring municipalities in the period from 2017 to 2019

The results of the analysis show that the area of Belgrade is exposed to the impact of regional sources of PM_{10} emissions located south, southwest, and southeast of the city, as well as slightly less impact of sources located in the areas west and east of the analyzed area. Significant emission sources in the southwestern areas on the outskirts of the agglomeration can be associated with the thermal power plant "Nikola Tesla" near Obrenovac, as well as with somewhat more distant mining basins near Veliki Crljeni. Apart from that, a source in the southeastern area that has an impact on the urban zone of Belgrade can be attributed to the Vinča city landfill, whereas several identified sources on the left bank of the Danube, outside the agglomeration, can be linked to agricultural activities in Banat. Regional sources of suspended particles whose impact is estimated to be significant, and which are located southeast at a greater distance, can be connected with "Železara Smederevo", as well as with the thermal power plant and coal mine "Kostolac". In the western region of Belgrade, along the international highway E-70, sources of slightly lower intensity have been identified, which can be attributed to traffic activities. Also, it should be taken into consideration that a large number of facilities of small economic activities (production plants, processing and storage of goods) have been built in this area in recent years, whose emissions also contribute to air pollution. Figure 3 also shows PM_{10} emission sources located south of Belgrade, which most likely represent the contribution of more remote areas, or even part of the long-distance cross-border transport route.

4. CONCLUSION

Apart from the influence of local sources, the air quality in the area of Belgrade is affected by various distant sources of emissions. The impact of strong local sources was least noticed in the case of sulfur dioxide, while monitoring locations were most directly exposed to suspended particulate emissions from the immediate environment (mobile sources - traffic and transport, resuspension, and local economic activities). On the other hand, the share of background and transport of air pollution was the highest in the case of SO_2 - 70% (combustion of fossil fuels for heating and pollutant transport from remote power plants) and NO_2 - 45% (pollutant transport and formation in photochemical atmospheric transformations).

The analysis of the contribution of regional pollutant transport to the measured PM_{10} concentrations has shown a significant impact of sources located southeast ("Železara Smederevo" and thermal power complex "Kostolac") and southwest (thermal power plants "Nikola Tesla" and mining basin "Tamnava") from Belgrade. The contribution of somewhat weaker sources located in the western area of Belgrade, can be related to traffic activities along the international highway E-70 and economic activities in its surroundings. For detailed characterization of the identified emission sources, and thus the improvement of insufficiently updated emission inventories, it is necessary to include other pollutants in the analysis and to apply the most advanced artificial intelligence methods.

5. ACKNOWLEDGEMENTS

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI-ATLAS, as well as the City of Belgrade, Department of Environmental protection of the city administration, Serbia, Air quality plan for the City of Belgrade.

REFERENCES

[1] Institute of Physics Belgrade, Singidunum University i School of Leectrical Engineering, „Mapping of sources of toxic, mutagenic, and carcinogenic volatile organic compounds in the city of Belgrade," Green Fund, Ministry of Envrionmental Protection, Republic of Serbia, 1 11 2018. [Na mreži]. Available: http://bpm.ipb.ac.rs/. [Accessed 1 6 2021].

[2] A. Stojić and S. Stanišić Stojić, "The innovative concept of three-dimensional hybrid receptor modeling," *Atmospheric Environment*, vol. 164, pp. 216-223, 2017.

[3] R. Stull, „Mean boundary layer characteristics," u *An Introduction to Boundary Layer Meteorology*, Dordrecht, Springer, 1988, pp. 1-27.

[4] W. Hong, Y.-f. Zhang, S.-q. Han, J.-h. Wu, X.-h. Bi, G.-l. Shi, J. Wang, Q. Yao, Z.-y. Cai i Y.-c. Feng, „Vertical characteristics of PM2. 5 during the heating season in Tianjin, China," *Science of the Total Environment*, t. 523, pp. 152-160, 2015.

[5] S. Stanišić, M. Perišić i A. Stojić, „The use of innovative methodology for the characterization of benzene, toluene, ethylbenzene and xylene sources in the Belgrade area," u *Sinteza 2020, International scientific conference on information technology and data related research*, Belgrade, Serbia, 2020.

[6] A. Stojić i S. Stanišić Stojić, „Concentration weighted boundary layer hybrid receptor model for analyzing particulate matter altitude distribution," u *6th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management*, Belgrade, Serbia, 2017.

[7] M. Perišić, A. Stojić, G. Jovanović i S. Stanišić, „Receptor oriented modeling of urban particulate air pollution: source characterization and spatial distribution," u *7th International WeBIOPATR*, Belgrade, Serbia, 2019.

[8] S. Han, Y. Zhang, J. Wu, X. Zhang, Y. Tian, Y. Wang, J. Ding, W. Yan, X. Bi, G. Shi i Z. Cai, „Evaluation of regional background particulate matter concentration based on vertical distribution characteristics," *Atmospheric Chemistry and Physics*, t. 15, br. 19, pp. 11165-11177, 2015.

STRUCTURAL CHARACTERISTICS OF PARTICULATE MATTER TIME SERIES OBSERVED IN AN URBAN ENVIRONMENT

Gordana Jovanović[1,2],
Svetlana Stanišić[2*],
Mirjana Perišić[1,2],
Andreja Stojić[1,2]

[1]Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
Belgrade, Serbia

[2]Singidunum University,
Belgrade, Serbia

Abstract:
In this study, we used the fractal and multifractal analysis to explore the structural characteristics of PM10 time series, among which self-similarity and invariance can be considered particularly important. The eXtreme Gradient Boosting method was used to fill in the missing data for multiscale multifractal analysis. The analysis has revealed self-similarity in PM_{10} time series with a positively correlated structure which was stable over a study period. Small fluctuations of PM_{10} levels were observed as a result of variations in local emissions and meteorological conditions. The uncoordinated and uncorrelated intervals in concentration time series were observed as a consequence of occasional pollution events in the areas dominated by industrial activities or as a consequence of the remote emission source activity when wind direction and speed were favorable.

Keywords:
Particulate Matter, Time Series Analysis, Multiscale Multifractal Analysis.

INTRODUCTION

According to the estimate from the World Health Organization, air pollution caused 4.2 million cases of premature death worldwide in the year 2015, whereas the recent estimates indicate that the mortality rate due to exposure to high levels of air pollutants is significantly higher and accounts for 8.9 million. In addition, research has indicated that in case the trend of low air quality continues and the approach to environmental issues is not fundamentally changed, the numbers could be twice as high by 2050.

Environmental science is facing many problems in achieving its mission to guarantee sustainable future in an increasingly complex and rapidly changing overpopulated world. The continuous pollution burden on the environment is dependent not only on the increasing pollutant load, but also on many known processes such as pollution transport, dispersion and deposition, atmospheric chemistry, meteorological factors, solar and cosmic radiation, topography, etc., as well as those which are not even known yet.

Correspondence:
Svetlana Stanišić

e-mail:
sstanisic@singidunum.ac.rs

DOI: 10.15308/Sinteza-2021-98-101

The issues that prevent the environmental science to fulfil its' mission are related to (1) complexity, non linearity, interactivity, and cross-compartment interconnectivity of environmental phenomena, (2) insufficiency of data-driven knowledge, especially the knowledge derived as a result of global-scale and multi-compartment research, (3) asymmetric access to data, information, and knowledge, (4) lack of adequate infra-structures regarding environmental big data, (5) barriers and gaps to technological innovation access, (6) high pressure on human and institutional capacities regarding innovation, *etc.*

Particulate matter (PM) emitted from different both natural and anthropogenic emission sources can remain in the air for a few hours or days depending on local meteorological conditions, susceptibility to chemical and physical transformations, and factors that contribute to sedimentation and precipitation. Self-similarity and invariance are important features of pollutant concentration time-series. These structural characteristics of PM time series revealed by using fractal and multifractal analysis could be considered when assessing their behavior patterns in the present and predicting their behavior in the future [1], [2], [3], [4], [5], [6], [7]. These analyses assume that phenomena and dynamic behavior of the system do possess the property of self-similarity and that the features of the system on one scale resemble the ones on different scales [8], [9], [10].

The atmosphere of the urban environment contains up to several hundred types of particulate matter, some of which are toxic, mutagen and, carcinogen. Adequate consideration of air quality is significantly limited by relying on data on gaseous inorganic oxides, or the concentration data of only the coarse PM fraction (PM_{10}) and several of its constituents. European Union countries measure concentrations for as much as 40 pollutants, as well as numerous constituents of three PM fractions (PM_{10}, $PM_{2,5}$, and PM_1).

In this paper, we investigate the fractal behavior of PM_{10} time-series across Belgrade area by the use of multiscale multifractal analysis (MMA) with the aim to obtain a more comprehensive understanding of the particulate matter behavior and environmental fate.

2. MATERIALS AND METHODS

Analysis of the structural characteristics of PM time series (fluctuation, self-similarity, and invariance) has been performed using MMA. The analysis of the characteristic parameters of the MM-spectrum (Hurst exponent, multifractal parameter, and scale) provided information on particularities of air pollution dynamics at a given location. A detailed description of the method is to be found elsewhere [11].

Hurst exponent (H) is used to describe the self-similarity of fractal properties, i.e., time series of pollutants in the presented analysis [12], [13]. In general, if H<0.5, the correlation between the intervals in the time series is a negative one, the change that occurs in the next moment will be opposite compared to the previous one, and the system has a pronounced tendency to fluctuate. The processes characterized by H=0.5 are random, similar to Brownian motion, and there is no correlation between the increments in the time series. If 0.5<H<1.5 is valid, there is a positive correlation between the shifts in the next moment will show similar tendencies as the previous one and the time series possess the property of self-similarity. Self-similarity is more pronounced the closer H gets to 1. When H>1.5 the time series is characterized by uncoordinated and uncorrelated intervals. In addition to the Hurst exponent, a multifractal parameter with both negative and positive values can be used to assess the fractal characteristics. The higher the value of the parameter, the higher the degree of fluctuation, whereas the absence of fluctuations leads to the multifractal parameter value of 0 and represents monofractal behavior.

The eXtreme Gradient Boosting method was used to fill in the missing data for the MMA application. The study used method implementation within the Python software environment. A detailed description of the method is to be found elsewhere [14].

3. RESULTS AND DISCUSSION

Characterization of PM10 time series observed in Belgrade has been performed by using MMA. At almost all monitoring stations, the value of Hurst exponent between 0.70 and 1.5 indicates self-similar time series PM_{10} with a positively correlated structure that is stable over a long period, Figure 1.

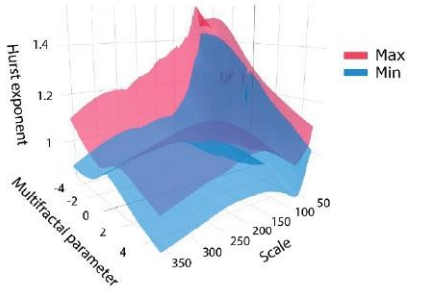


Figure 1 – PM_{10} Hurst exponent range in Belgrade 2017-2019.

Multifractal analysis of PM10 time series at monitoring stations at the Institute of Public Health of Belgrade and Obrenovac are presented in Figures 2 and 3.

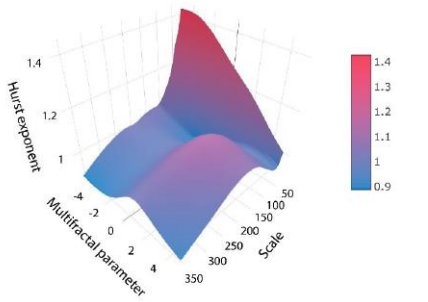


Figure 2 – Structural characteristics of time series of concentrations PM_{10} at monitoring place at the Institute of Public Health of Belgrade for the period from 2017 to 2019.

After reaching the maximums, values of Hurst exponent plummet to the value of 1 in the areas of small and large fluctuations and a time scale of up to 120 hours, thus indicating the most stable fractal nature of PM_{10} time series with a correlated structure over a prolonged period of time, i.e., the existence of the "long-term memory". This trend generally continues on scales from 150 to 350 hours with episodes of higher fluctuations (multifractal parameter = -0.5 – 1.8) for the period from 130 to 245 hours, in which the values of Hurst exponent do not exceed 1.13.

The PM10 concentration variability characterized by H values from 0.73 to 1.54 and values of multifractal parameter from -5 to 5 at the Obrenovac sampling site are shown at Figure 3. In the domain of lower fluctuations (multifractal parameter ≈-5), two peaks stand out: H>1.5 between 165 and 240 hours and H=1.3 on a time scale of up to 30 hours.

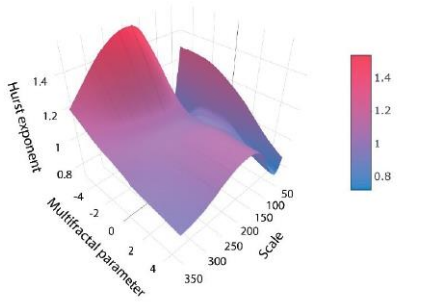


Figure 3 – Structural characteristics of time series of concentration PM_{10} at the monitoring station Obrenovac for the period from 2017 to 2019.

The time series characterized by H-values greater than 1.5 consists of uncoordinated and uncorrelated intervals which can be attributed to occasional emission intensification in the areas dominated by industrial activities or to impacts of remote emission sources when wind direction and speed were favorable. At the monitoring station Obrenovac, the most significant impacts can be attributed to emissions from the power plant "Nikola Tesla", as well as to works on construction sites of the A2 highway over the course of the observed period. Upon reaching the maximum values, steep slopes were observed in the domains of higher and lower fluctuations on the time scale of up to 350 (H up to 1.03) and 75 hours (H up to 0.72), respectively. The results indicated that the variability of PM time series decreased when the environment factors weaken and that there is a positive correlation between time intervals demonstrating similar dynamics.

4. CONCLUSION

There is an urgent need to stimulate new practices of interdisciplinary evidence-based research and innovation in which the research, design, development, deployment, and the use of advanced statistical, and numerical methods is anchored in environmental science. Further progress of environmental science and environmental pollution research will certainly depend on its integration with other scientific disciplines, among which high-performance computing seem to be of extreme significance. Moreover, the modern science requires infrastructure being data-based, efficient, real-time responsive, scalable, flexible, and robust enough to allow the understanding of the evolution of global pollution

impact and climate change in real-time and to anticipate future trends and challenges for the sake of global sustainability.

Application of MMA indicated self-similarity in PM10 time series with a positively correlated structure which was stable over a longer period of time at almost all monitoring stations. The results from AMS Institute of Public Health of Belgrade and AMS Obrenovac were taken as representative for further interpretation. At the AMS Institute of Public Health of Belgrade, the most pronounced variations were recorded in the domain of low fluctuations and on small scales of up to 45 hours. Fluctuations of PM_{10} on small scales were impacted by the intensity of local emissions and meteorological conditions governing the processes of condensation and nucleation, as well as physicochemical transformations and formation of secondary aerosols. Upon reaching the above stated maximum, a stable and positively correlated fractal nature of PM_{10} time series over a longer period of time in the domain of lower and higher fluctuations was observed.

At the Obrenovac measurement site, variability of PM concentrations in the domain of lower fluctuations was evidenced, on a scale between 165 and 240 hours. The PM time series consisted of uncoordinated and uncorrelated intervals, as a consequence of occasional pollution events in the areas dominated by industrial activities or as a consequence of the impact of remote emission sources when wind direction and speed were favorable. Also, the results indicate that variability of PM time series decreases when the impact of environment factors weakens, and that there is a positive correlation between time intervals indicating similar dynamics.

5. ACKNOWLEDGEMENTS

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI-ATLAS, as well as the City of Belgrade, Department of Environmental protection of the city administration, Serbia, Air quality plan for the City of Belgrade.

REFERENCES

[1] A. Chelani, "Long-memory property in air pollutant concentrations," *Atmospheric Research*, vol. 171, pp. 1-4, 2016.

[2] Q. Dong, Y. Wang and L. Peizhi, "Multifractal behavior of an air pollutant time series and the relevance to the predictability," *Environmental Pollution*, vol. 222, pp. 444-457, 2017.

[3] T. Plocoste, R. Calif and S. Jacoby-Koaly, "Temporal multiscaling characteristics of particulate matter PM10 and ground-level ozone O3 concentrations in Caribbean region," *Atmospheric Environment*, vol. 169, pp. 22-35, 2017.

[4] T. Stadnitski, "Measuring fractality," *Frontiers in physiology*, vol. 3, p. 127, 2012.

[5] A. Stojić, S. Stanišić Stojić, I. Reljin, M. Čabarkapa, A. Šoštarić, M. Perišić and Z. Mijić, "Comprehensive analysis of PM 10 in Belgrade urban area on the basis of long-term measurements," *Environmental Science and Pollution Research*, vol. 23, no. 11, pp. 10722-10732, 2016.

[6] A. Stojić, G. Vuković, S. Stanišić, V. Ćućuz, D. Trifunović, V. Udovičić and A. Šoštarić, "Multifractality of isoprene temporal dynamics in outdoor and indoor university environment," in *8th International PTR-MS Conference,* Innsbruck, Austria, 2019.

[7] A. Stojić, S. Stanišić Stojić, M. Perišić and Z. Mijić, "Multiscale multifractal analysis of nonlinearity in particulate matter time series," in *6th International WeBIOPATR Workshop & Conference Particulate Matter: Research and Management,* Belgrade, Serbia, 2017.

[8] H. E. Hurst, "Long-term storage capacity of reservoirs," *Transactions of the American society of civil engineers,* vol. 116, no. 1, pp. 770-799, 1951.

[9] B. Mandelbrot, The fractal geometry of nature, vol. 1, New York: WH freeman, 1982.

[10] B. Reljin and I. Reljin, "Fraktalna i multifraktalna analiza signala," in *Telfor 2001*, Belgrade, Serbia, 2001.

[11] J. Gierałtowski, J. Żebrowski and R. Baranowski, "Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia," *Physical Review E*, vol. 85, no. 2, p. 021915, 2012.

[12] E. A. F. Ihlen, "Introduction to multifractal detrended fluctuation analysis in Matlab," *Frontiers in physiology*, vol. 3, p. 141, 2012.

[13] E. Molino-Minero-Re, F. García-Nocetti and H. Benítez-Pérez, "Application of a time-scale local hurst exponent analysis to time series," *Digital Signal Processing*, vol. 37, pp. 92-99, 2015.

[14] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *22nd acm sigkdd international conference on knowledge discovery and data mining,* San Francisco California USA, 2016.

EVOLUTION OF INDUSTRY-RELATED VOLATILE ORGANIC COMPOUND LEVELS AFFECTED BY COVID-19 LOCKDOWN IN BELGRADE

Filip Alimpić[1]*,
Mirjana Perišić[1,2],
Svetlana Stanišić[1],
Gordana Jovanović[1,2],
Andreja Stojić[1,2]

[1]Singidunum University,
Belgrade, Serbia

[2]Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
Belgrade, Serbia

Abstract:
In this study, we have evaluated the impacts of emergency state and curfew period on the industry-related volatile organic compound concentrations in Belgrade, Serbia. Pollutant concentrations were registered during the three-month period by using Standard Proton Transfer Reaction Quadrupole Mass Spectrometer (PTR-MS) and data analyses included correlation analysis with hierarchical clustering, probability density functions, and bivariate polar plots. As shown, all compounds, except those registered at protonated mass m/z 121, exhibited a significant drop in concentrations only a week after curfew was introduced. The behavior of analyzed compounds suggests that the VOC concentrations are more affected by industrial than traffic emissions.

Keywords:
Air Quality, Volatile Organic Compounds, COVID-19, Lockdown, PTR-MS.

INTRODUCTION

The COVID-19 pandemic in Spring 2020 had a major impact on human behavior, which resulted in significant changes in air quality worldwide and reported benefits to the natural environment. In the period that followed, several studies have used this real-world experiment to enhance our understanding of air pollution and its sources.

The study of Berman and Ebisu has shown statistically significant NO_2 declines of 25.5%, as well as a somewhat smaller decrease of $PM_{2.5}$ levels in urban counties and counties where early non-essential business closures were introduced [1]. The study of Querol et al. investigated air quality changes across 11 metropolises in Spain [2]. Their results emphasized the importance of the massive use of public transport that was reduced because of the fear of infection. While NO_2 levels fell below 50% of the WHO annual air quality guidelines, $PM_{2.5}$ levels were reduced less than expected due to fact that traffic was not the major factor contributing to high PM levels, but also due to the increased contributions from biomass burning or meteorological conditions favoring secondary aerosol formation.

Correspondence:
Filip Alimpić

e-mail:
alimpic.filip@outlook.com

DOI: 10.15308/Sinteza-2021-107-111

In compliance with this, the study of Briz-Redón et al. has shown that the 4-week lockdown had a significant impact on reducing the atmospheric levels of NO_2, as well as CO, SO_2, and PM_{10} in some cities, but the levels of O_3 were increased [3]. The study of Chen et al. has concluded that the interventions adopted to limit the COVID-19 outbreak have resulted in improvements in air quality and associated health benefits in non-COVID-19 deaths, which could have outnumbered the confirmed deaths attributable to COVID-19 in China [4]. In this study, we have evaluated the impact of 3-month preventive measures and curfew on air quality in Belgrade (Serbia) based on industry-related volatile organic compounds (VOCs).

2. MATERIALS AND METHODS

The measurements of VOCs and meteorological parameters were conducted in Belgrade urban area (44.86° N, 20.39° E) in the period from 2nd March to 2nd June 2020. They covered two weeks before the implementation of the state of emergency introduced as a response to the COVID-19 pandemic and lasted almost one month after the measures were lifted. Standard Proton Transfer Reaction Quadrupole Mass Spectrometer (PTR-MS, Ionicon Analytik, GmbH, Austria) was used for online measurements of 21 to 270 amu mass range [5], [6], while Vaisala weather station was used for measuring meteorological parameters. Calibration of PTR-MS measurements was done according to Taipale and coauthors [7] by using referent gases and a liquid calibration unit (Ionicon Analytik). Data analyses included correlation analysis with hierarchical clustering, probability density functions, and bivariate polar plots [8]. Mobility data was obtained from Google and Apple.

Figure 1 shows industry-related VOC concentrations and human activity change which accompanies the start of emergency measures caused by the COVID-19 pandemic in Serbia. It can be seen that the evolution of concentrations did not indicate an immediate change in air quality with introducing a state of emergency.

Although the decrease in the intensity of human mobility and industrial activities started with the appearance of the first COVID-positive cases, this period was also accompanied by an increase in the concentrations of all measured compounds. For all compounds (except compounds with protonated mass m/z 121) a significant drop in concentrations was observed only a week later, when the curfew was enforced for quite some time, first for 9 hours, then for 12 hours, and finally throughout the weekends.

Starting from the second half of April, a gradual intensification of human activities and a decrease in the stay-at-home campaign could be observed (intensive adherence to extremely restrictive measures seemed to have lasted only 2, at most 3 weeks), but VOCs concentrations continued to fall. By the end of the measurement campaign, human activities had not returned to the level before the introduction of a state of emergency, especially mobility (walking and driving).

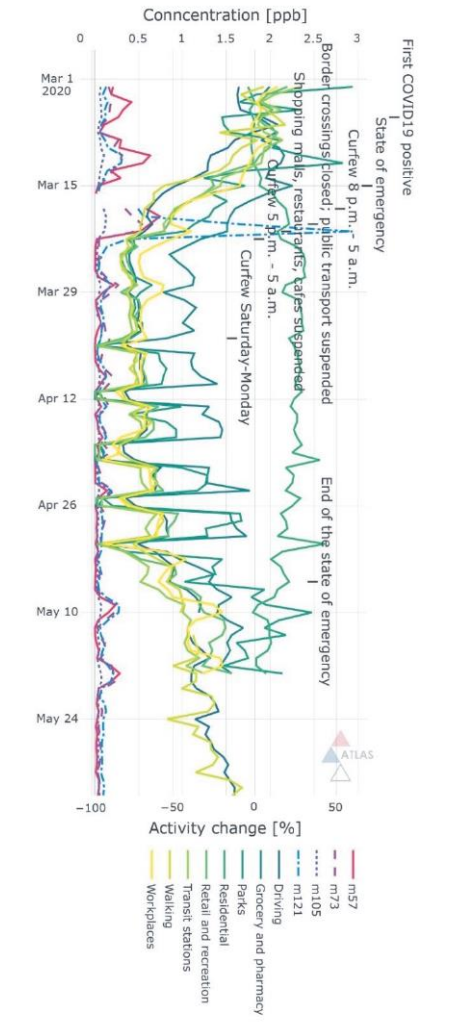


Figure 1 - Industry-related VOC concentration and human activity time series.

It should be noted that there were a few spikes of concentrations registered at all four masses in the period just before and during the lockdown. For compounds registered at m/z 57, the spike almost 4 times the mean concentration was observed a day before the start of the lockdown. For compounds registered at m/z 73, the spike almost 3 times the mean concentration occurred six days after the beginning of the lockdown. Compounds registered at m/z 105 had an intensity spike of 2 times the mean concentration in the event that has occurred 19 days after the start of the lockdown. The largest spike was observed for compounds registered at m/z 121, which raised almost 13 times more than the average readings, and it occurred at the same time as the spike of compounds measured at m/z 73 (six days after the start of the lockdown). General changes in compound concentrations during and after lockdown can be assigned to low industrial activity, while the origin of spikes demands further investigation.

In the period before the introduction of the emergency state, the correlations between all analyzed VOCs (r-values were in the range from

3. RESULTS AND DISCUSSION

0.81 to 0.95), as well as the interconnection of compounds detected at m/z 73 and human activities such as walking, spending time in retail, recreation and transit were observed (Figure 2). The correlation between compounds registered at m/z 105 and 121 was the strongest (r=0.95), while the weakest correlations were calculated between compounds registered at m/z 73 and other VOCs (lowest being between compounds registered at m/z 73 and 121 at r=0.81). During the state of emergency, the interconnection of all VOCs persisted, excluding compounds registered at m/z 121, as well as the correlations between the observed concentrations and human activities. No correlations between VOCs and human mobility were observed in the period following the lockdown, although the correlations between volatiles, including compounds registered at m/z 121, were re-established.

Figure 3 shows the relative changes in mean VOC concentrations during and after the lockdown compared to the pre-introduction period. During the state of emergency, the concentrations of all compounds (except compounds registered at m/z 121) dropped in the range from 30 to 73% compared to the period before its introduction. After the state of emergency was lifted, concentrations continued to fall (35-80%).

The concentrations of compounds registered at m/z 121 recorded an increase of over 36% during the state of emergency, while in the period after the lockdown the level was 20% lower compared to the period before the introduction of the emergency measures.

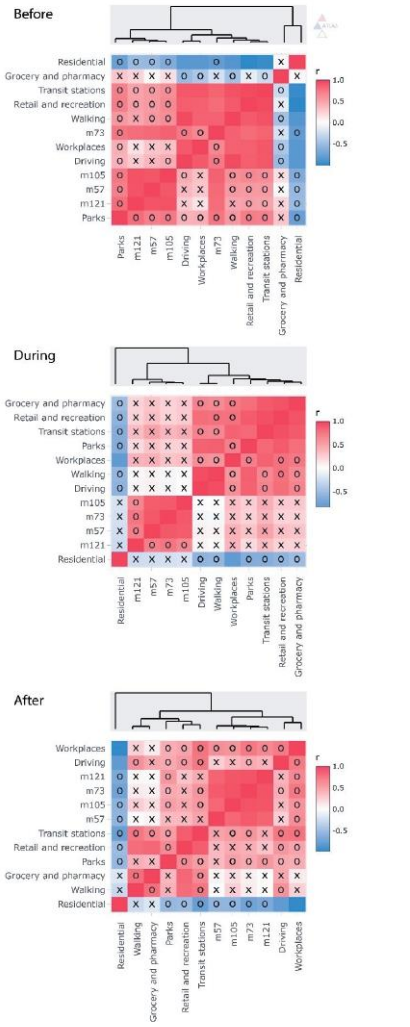


Figure 2 - Parameter value correlation matrix.

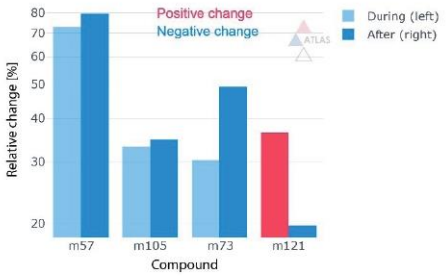


Figure 3 - Mean VOC concentration difference during and after relative to the period before the state of emergency.

The probability distribution functions (PDFs) showed unimodality with pronounced peaks during all three periods of the measurement campaign (Figure 4). The PDFs of m/z 57, 73, and 105 suggested lowering the intensity of emissions of dominant sources during the pre-lockdown period. The compounds that have been detected at m/z 121 had a similar unimodal shape of PDF during every period of the measurement campaign.

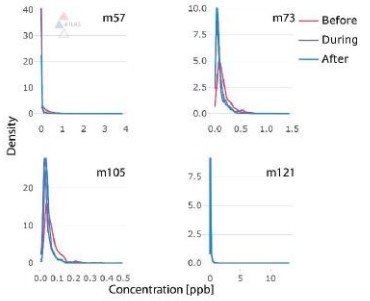


Figure 4 - VOCs density plots.

The dependence of concentrations on wind direction and speed before the restriction period indicated that the common sources for the analyzed compounds were located in the eastern, western, and southwestern directions from the monitoring site (Figure 5). Additionally, certain emission sources of the compounds registered at m/z 73, 105, and 121 were revealed northeast. The highest concentrations of m/z 57 were recorded for the wind speeds of 3 to 7 m s 1, coming from the western and eastern direction which suggests that the most intense emission sources were distant ones. The compounds registered at m/z 73, 105, and 121 had similar behavior patterns. High concentrations for wind speeds ranging from 4 to 6 m s 1 also indicate the influence of remote sources.

With the introduction of emergency measures, a homogenization of pollution in the ground layers of the atmosphere took place, which was reflected in the relatively uniform distribution of concentrations of all analyzed volatiles regarding wind direction. The reduction of contribution of distant sources and the dominance of local ones was observed.

After the lockdown period, VOC concentrations remained low, with notable activation of the sources of compounds registered at m/z 73 and 105 in the western and southwestern areas.

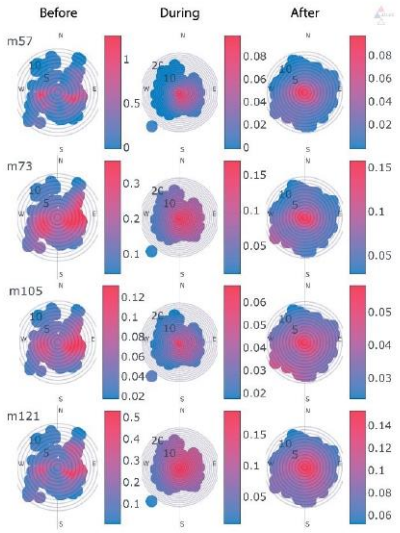


Figure 5 - Industry-related VOC dependency on wind parameters.

4. CONCLUSION

As can be concluded, an immediate change in air quality with introducing a state of emergency was not observed. A few spikes of concentrations of industry-related volatile organic compounds were registered in the period just before and during the lockdown, while the general decrease in pollutant concentrations during and after lockdown can be assigned to low industrial activity.

In the period before and during the lockdown, the correlations between all analyzed VOCs, with exception of compounds registered at m/z 121, and human activities were observed, while after the lockdown no correlations between VOCs and human mobility were detected. The probability distribution functions showed the unimodal distribution of the concentrations with pronounced peaks during all three periods of the measurement campaign. During and after the lockdown, the reduction of contribution of distant sources and the dominance of local ones was observed. As can be concluded, the industry appears to be the major source of analyzed volatiles.

5. ACKNOWLEDGEMENTS

The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia and the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI-ATLAS.

REFERENCES

[1] J. Berman i K. Ebisu, „Changes in US air pollution during the COVID-19 pandemic," *Science of the Total Environment,* t. 739, p. 139864, 2020.

[2] X. Querol, J. Massagué, A. Alastuey, T. Moreno, G. Gangoiti, E. Mantilla i J. Duéguez, „Lessons from the COVID-19 air pollution decrease in Spain: Now what?," *Science of The Total Environment,* t. 779, p. 146380, 2021.

[3] Á. Briz-Redón, C. Belenguer-Sapiña i Á. Serrano-Aroca, „Changes in air pollution during COVID-19 lockdown in Spain: a multi-city study," *Journal of environmental sciences,* t. 101, pp. 16-26, 2021.

[4] C. Kai, M. Wang, C. Huang, P. Kinney i P. Anastas, „Air pollution reduction and mortality benefit during the COVID-19 outbreak in China," *The Lancet Planetary Health*, t. 4, br. 6, pp. e210-e212, 2020.

[5] W. Lindinger, A. Hansel and A. Jordan, "On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research," *International Journal of Mass Spectrometry and Ion Processes*, vol. 173, no. 3, pp. 191-241, 1998.

[6] A. Stojić, S. Stanišić Stojić, A. Šoštarić, L. Ilić, Z. Mijić i S. Rajšić, „Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling," *Environmental Science and Pollution Research*, t. 22, br. 17, pp. 13137-13152, 2015.

[7] R. Taipale, T. M. Ruuskanen, J. Rinne, M. K. Kajos, H. Hakola, T. Pohja i M. Kulmala, „Quantitative long-term measurements of VOC concentrations by PTR-MS–measurement, calibration, and volume mixing ratio calculation methods," *Atmospheric Chemistry and Physics,* t. 8, br. 22, pp. 6681-6698, 2008.

[8] S. Grange, A. Lewis i D. Carslaw, „Source apportionment advances using polar plots of bivariate correlation and regression statistics," *Atmospheric Environment*, t. 145, pp. 128-134, 2016.

Singidunum University

Sinteza

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY AND DATA RELATED RESEARCH

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology and Data Related Research - Sinteza 2020 has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

Belgrade
October 17, 2020.
sinteza.singidunum.ac.rs

We have noted that these fatty acids were not the most abundant in the studied fish species, while the dominance of other FAs was recorded in the following order: palmitic > oleic > docosahexaenoic > stearic > myristic > eicosapentaenoic > palmitoleic > linoleic acid. Saturated acids such as C14:0 and C17:0 dominantly occured as esterified polar phospholipids and therefore, SFAs and associated contaminants are more bioavailable than FAs presented in non-esterified free form. Eicosapentaenoic acid (EPA, 20:5n-3) and docosahexaenoic acid (DHA, 22:6n-3), are widely known as the most nutritionally relevant ω-3 PUFAs in oily blue fish, such as mackerel, sardines and anchovies. However, they appear to have no impact on the PCB-138 accumulation. Chub mackerel has been known to contain high percentage of the PUFAs in the form of free FAs, which are less efficiently absorbed than other lipid classes and easily eliminated from human intestines where they are dispersed into mixed micelles and bound to soluble lipid-binding proteins [11]. In addition, PUFAs are more susceptible to oxidative degradation when found as free FAs.

Four clusters (C1 and C3–C5, ≈70%) dominantly showed negative correlations with PCB-138 concentrations below 0.5 ng g^{-1}. The same parameters as discussed above shaped the clusters and negative relationships imply that low concentrations of SFAs do not add to PCB-138 uptake, or uptake of POPs from different sources, that are found to have variable impact, depending on the sampling time, seasonal and other fishery zone-related factors.

The plot in *Fig. 3.* shows the impact which each cluster has on output of the model. High levels of the cluster (C2, C6-C10) constituents, primarily PCB-153, *p,p'*-DDE, PCB-118, PCB-170, PCB-180 and myristic acid, has a high and positive impact on the PCB-138 patterns, as shown by right-oriented long distribution tail. Lower concentrations of these variables (< 0.1 ng g^{-1}) are negatively correlated with the target variable, as indicated by negative relative SHAP value.

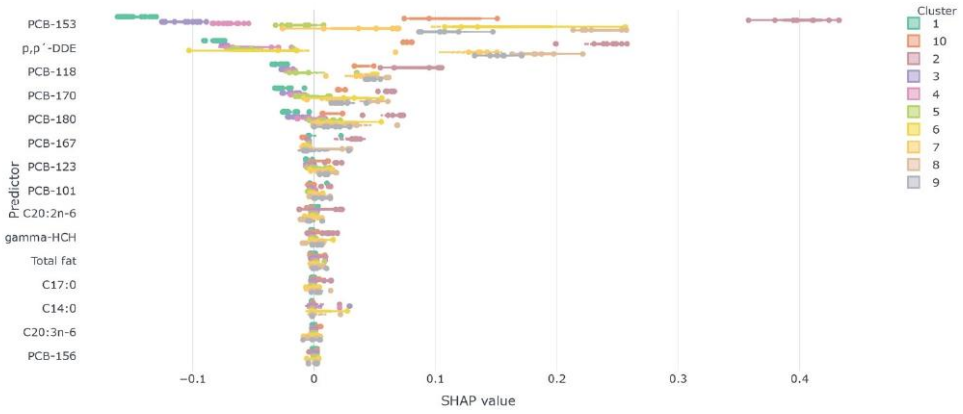


Fig. 3. PCB-138 SHAP value cluster evaluation.

4. CONCLUSION

Small edible fish species could be considered as both nutritionally valuable food source and sources of hazardous organochlorine pollutants, which negatively affect human health. Significant indications on relationships between POPs uptake and fatty acid in fish tissue has been reported worldwide. In this study, we presented a promising methodology, an explainable artificial intelligence methods (XGBoost and SHAP), which was used with the aim of gaining better understanding of specific interrelations between fatty acid content and contaminants in consumable marine fish species. Out of 18 fatty acids, two saturated (myristic and margaric) and two ω-3 and 6 (eicosadienoic and dihomo-γ-linolenic) acids were identified as crucial for the bioaccumulation of PCB-138 in sardine, anchovy and mackerel species. However, nutritionally beneficial EPA and DHA are

PCB-118; PCB-138–PCB-153; PCB-153–PCB-118; PCB-153–PCB-170 and PCB-156–PCB-180. The correlations between the POPs pairs: γ-HCH–PCB-170; γ-HCH–PCB-156; γ-HCH–PCB-118; γ-HCH–PCB-114; γ-HCH–PCB-180; *p,p*'-DDE–PCB-153; *p,p*'-DDT–PCB-170; *p,p*'-DDT–PCB-156; *p,p*'-DDT–PCB-118; *p,p*'-DDT–PCB-153; PCB-138–PCB-170; PCB-180–PCB-170; PCB-180–PCB-156; PCB-105–PCB-180; PCB-105–PCB-156 and PCB-105–PCB-170 were in the range from 0.80 to 0.90. As can be seen, the listed species with similar chemical structure and common origin display similar behavior patterns, as discussed below. We assumed that other methods apart from commonly applied correlation matrices could be further employed to describe the associations between POPs and FAs or inorganic contaminants in more details.

In this study, XGBoost has been successfully employed for the investigation of non-linear relationships between PCB-138 levels and key variables that shape its' behavior pattern in marine fish tissue. The predicted/observed calculated relative error was 11.2%, while the r^2 were 0.99 (*Fig. 1*).

Fuzzy clustering of SHAP values extracted ten groups (relative error < 10%) of similar variables that shape the dynamics of PCB-138 (*Fig. 2.* As labelled by red color, the constituents of six clusters, C2 and C6–C10, strongly positively impacted PCB-138 concentrations higher than 0.5 ng g^{-1} and up to 1.5 ng g^{-1}. Dominant influences are attributed to the higher levels of myristic acid (> 7%) and the following compounds: *p,p*'-DDE, PCB-101, PCB-118, PCB-123, PCB-153, and PCB-180, as well as to lower margaric content (≈ 1%). Although indicator congeners (-28, -52, -101, -138, -153, and -180), which are often classified as non-dioxin-like PCBs, dominated over dioxin-like PCBs, both classes were of significance when evaluating PCB-138 patterns. The relationships between PCB congeners and their presence at higher concentration levels are due to their structure referring to molecule rigidity and the number of attached halogen atoms, and consequently, perseverance and prolonged half-lives in marine environment.

Out of 18 investigated FAs, saturated myristic (C14:0) and margaric (C17:0) acid mostly contributed to the PCB-138 bioaccumulation followed by nutritionally beneficial eicosadienoic (C20:2n-6) and dihomo-γ-linolenic acid (C20:2n-3).

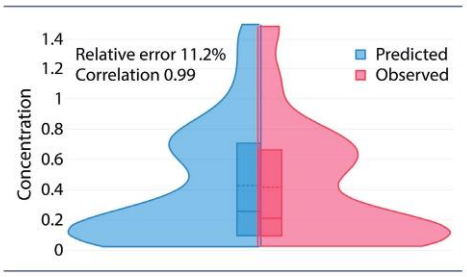


Fig. 1. XGBoost evaluation statistics.

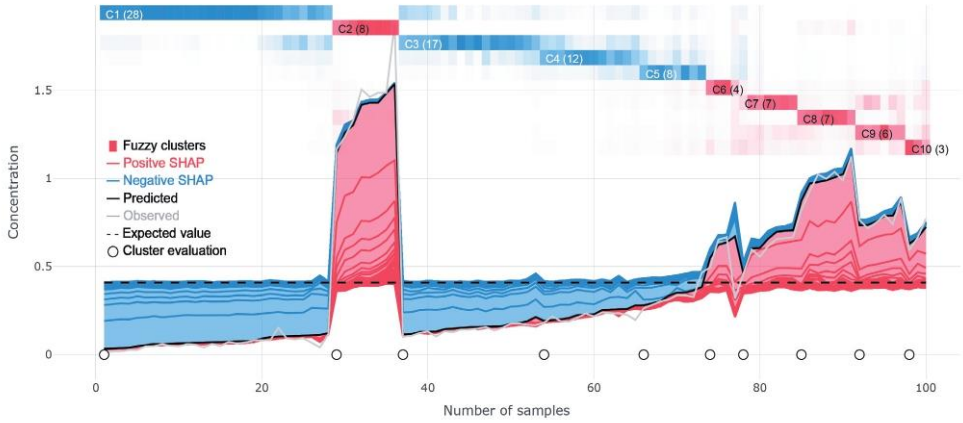


Fig. 2. PCB-138 SHAP force plot.

To check for the accuracy of the analysis, the certified reference material NIST SRM 1577c (bovine liver, Gaithersburg, MD, USA) was analyzed in the same manner as the fish samples. For all elements, the obtained results were within the satisfactory range of the certified values

Fatty acids

The concentrations of 6 SFAs (myristic, pentadecylic, palmitic, palmitoleic, margaric, and stearic acid), 3 MUFAs (oleic, paullinic and arachidonic acid) and 9 PUFAs, ω -6 and ω-3 families (linoleic, α-linolenic, icosadienoic, dihomo-γ-linolenic, eicosatrienoic, arachidonic, eicosapentaenoic, docosapentaenoic and docosahexaenoic) were determined according to procedure given in detail previously [5]. Prior analysis, the samples were partially thawed at +4 °C. Accelerated solvent extraction (ASE 200, Dionex, Sunnyvale, CA) using a mixture of n-hexane and iso-propanol (60:40 v/v) were applied for determination of total lipids for fatty acid content The samples were analysed as FAME (fatty acis methyl esters) by gas–liquid chromatography (Shimdzu, Japan). Flame ionisation detector (GC/FID) and fused silica cianopropyl HP-88 column were used. The chromatographic peaks were identified and quantified using Supelco 37 Component mix standard and internal standard (heneicosanoic acid methyl ester), respectively.

Data analysis

The relationships between PCB-138 and all other measured parameters were modeled by the XGBoost regression. The details on the method are given elsewhere [6]. In this study, we used Python XGBoost implementation. The dataset was split into training (80%) and validation (20%) sets. Hyperparameter tuning was implemented using a brute-force grid search and 10-fold stratified cross-validation. The best performing hyperparameter values were used for the final model.

The explainability of the produced XGBoost model that operates with high-dimensional input data in a non-linear fashion was obtained by using explainable artificial intelligence method SHapley Additive exPlanations (SHAP) [7]. Based on the game theory, the Shapley explanations represent the only possible locally accurate and globally consistent feature attribution values. In this study, we applied the fuzzy clustering of absolute SHAP attributions to identify and characterize the relations among the measured parameters responsible for PCB-138 behavior.

2. RESULTS AND DISCUSSION

Pollutant toxicological profile

Small edible pelagic fishes live short at the bottom of the marine food chain (plankton<sardine species, anchovy<mackerel species), which are expected to bioaccumulate low levels of environmental contaminants. The transfer of accumulated hazards to the human body via fish intake is generally consider to be low. As shown by study results, inorganic compounds, macro-elements and toxic heavy metals, were prominently more abundant in the fish tissue than organic xenobiotics, OCPs and PCBs.

Biological effects of inorganic elements depend on the processes including absorption, accumulation, elimination and biotransformation into less or more toxic metabolites. Several minutes from absorption in the gut, the elements reach internal organs such as heart, liver, kidney and brain, while their penetration to muscles and adipose tissue occurs more slowly, up to several hours [8]. However, when they are present in low concentrations in the aquatic surrounding, their accumulation is less frequent as well. In the studied fish species, the concentrations of macro- (K > Na > Ca > Mg) and micro-elements (Fe > Zn > Cu) were within the normal physiological range, below the levels that could potentially cause pathological changes in tissues and organs, because of the absence of dominant well-known sources of pollution. In addition, the levels of the following elements: Hg, Pb and Cd, were lower than the maximum thresholds prescribed by the existing EU regulation [9].

Among POPs, *p,p'*-DDE, PCB-153, PCB-138 and PCB-180 were the most dominant. Although it is to be expected that larger, long-lived species, such as chub mackerel and horse mackerel, that are at higher trophic level, will uptake more organic contaminants than sardine species and/or anchovy, the values of the toxicological parameters such as the total quantity of indicator PCBs and WHO-dioxin-like PCBs toxic equivalents were below threshold concentrations of 75 ng g^{-1} w.w. and 6.5 pg g^{-1} [10]. As can be concluded, the examined fish speies appeared to be safe for human diet with regards to the presence of toxic chemicals.

PCB-138 patterns

As represented by Pearson's correlation analysis, significant linear correlation coefficients (r>0.90) were found between the following pairs of the investigated variables: *p,p'*-DDE–PCB-118; *p,p'*-DDE–PCB-138; *p,p'*-DDD–PCB-105; *p,p'*-DDD–PCB-180; PCB-138–

or forbidden in most countries. POPs bind tightly to particles in soil and sediment, which can act as a secondary sources of contamination for environmental media (water, air and living organisms).

At a global level, the oceans are the final sink/destination of POPs. The oceans act as a secondary source of contamination because POPs are slowly degraded and bioaccumulated in marine organisms, which are at the bottom of the food chain and represent a source of chemical hazards in human nutrition. The assessment of environmental exposure to marine toxins is based on POPs concentration data in the samples of water, plants and food. Small pelagic oily fish are highly recommended nutrient source worldwide due to their content of protein, minerals and healthy fats including omega-3 (ω-3) and 6 ω-6) polyunsaturated fatty acids (PUFAs) [1]. Benefits of the fatty acid consumption are associated with normal growth and development, the prevention of cardiovascular and inflammatory diseases, as well as cognitive decline and dementia. However, these FAs represent a very suitable matrix for bioaccumulation of highly lipophilic xenobiotics such as POPs. Although FAs profile and organochlorines content have been evaluated in numerous marine fish species worldwide, there are few data on their interrelations.

In this study, the presence of OCPs and PCBs has been investigated in small pelagic edible fish species: sardine *Sardina pilchardus* (Walbaum, 1792), anchovy *Engraulis encrasicolus* (Linnaeus, 1758), round sardinella *Sardinella aurita* (Valenciennes, 1847), chub mackerel *Scomber japonicus* (Houttuyn, 1782) and horse mackerel *Trachurus trachurus* (Linnaeus, 1758). We applied eXtreme Gradient Boosting (XGBoost), SHapley Additive exPlanations (SHAP), and SHAP value fuzzy clustering aiming to obtain a detailed insight into the distribution of indicator congener PCB-138 in the fish species. The impacts of the following factors including the level of OCPs, PCBs, saturated fatty acids (SFAs), monounsaturated fatty acids (MUFAs), PUFAs and heavy metals were evaluated by SHAP since the method offers uniquely consistent and locally accurate solutions that have been confirmed in the previous investigations of environmental phenomena [2].

2. MATERIAL AND METHODS

Sampling

Fish samples were collected along the eastern Croatian Adriatic Sea during 2014 and 2016. Details about sampling were previously described [3]. Total of 107 fish samples were collected in various fisheries coastal (A, E, and F) and off-coast (B and C) zones. Approximately 50 specimens was sampled randomly using purse seine catches (mesh size: 8 mm/bar length) totaling 107 pooled samples from a fillet of specimen.

Chemical analyses

POPs

Chemical analysis of POPs was previously described elsewhere [3]. In brief, seven OCPs (HCBα-, β-, and γ-HCH, *p,p'*-DDT, *p,p'*-DDE, and *p,p'*-DDD), six indicator PCB congeners (PCB-28, PCB-52, PCB-101, PCB-138, PCB-153, and PCB-180), eight mono ortho congeners (PCB-105, PCB-114, PCB-118, PCB-123, PCB-156, PCB-157, PCB-167, PCB-189) and () PCB-60, PCB-74 and PCB-170] were analysed. High-resolution gas chromatography with electron capture detector (s) was applied for the compound identification and details are described previously [3].

The recoveries for the PCBs was in the range between 75% to 89% while relative standard deviation (RSD) between 1% to 11% was obtained. The recoveries for OCPs were in the range from 76% to 86%, with RSD from 1% to 11%. for both PCBs and OCPs, the determination limits were 0.01 ng g–1 of fresh weight.

Elements

Sixteen macro- and micro-elements including toxic metals (Na, Mg, K, Ca, As, Cd, Co, Cr, Cu, Mn, Fe, Hg, Ni, Zn, Pb and Se) were analyzed. Homogenized fish samples (0.5 g) were basted by 5 mL of nitric acid (67% TraceMetal grade, Fisher Scientific, Bishop, UK) and 1.5 mL of hydrogen peroxide (30% analytical grade, Sigma-Aldrich, St. Louis, MA, USA). Afterwards, microwave digestion was tperformed as previously described [4].Analysis of the elements was performed by inductively coupled plasma mass spectrometry (ICP-MS), (iCap Q mass spectrometer, Thermo Scientific, Bremen, Germany). The most abundant isotopes were used for quantification.

For five-point calibration, solutions of Fe, Zn, Cu, Mn, Se, Cr, Co, Ni, Na, K, Mg and Ca were prepared in the concentration range of 0.2–2.0 mg L^{-1}. The concentration of calibration-solution forCd, Hg and As was in the range of between 0.2 and 2.0 μg L^{-1} and for Pb in the range between 2.0 and 20.0 μg L^{-1}.

ARTIFICIAL INTELLIGENCE ATLAS SESSION

EXPLAINABLE MACHINE LEARNING PREDICTION OF PCB-138 BEHAVIOR PATTERNS IN EDIBLE FISH FROM CROATIAN ADRIATIC

Andreja Stojić[1,2*],
Bosiljka Mustać[3],
Gordana Jovanović[1,2]

[1]Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia
[2]Environment and Sustainable Development, Singidunum University, Belgrade, Serbia
[3]Department of ecology, agronomy and aquaculture, University of Zadar, Zadar, Croatia

Abstract:
Fish consumption, especially consumption of oily marine species, is globally increasing since it has been recommended by dieticians due to high content of polyunsaturated ω-3 and ω-6 (PUFAs) fatty acids in fish tissue. Health benefits of PUFA ingestion coincide with the risk of intake of hazardous lipophilic persistent pollutants including organochlorine pesticides (OCPs) and related polychlorinated biphenyls (PCBs). We examined the impacts of 18 fatty acids (FAs) and 36 toxic organic and inorganic contaminants on the behavior patterns of indicator congener PCB-138 in marine fish using eXtreme Gradient Boosting (XGBoost), SHapley Additive exPlanations (SHAP), and SHAP value fuzzy clustering. XGBoost indicated non-linear relationships between investigated variables that surpasses indications suggested by commonly applied correlation matrices. Ten extracted fuzzy clusters of SHAP values revealed that higher intake of saturated myristic-C14:0 and margaric-C17:0 acids followed by intake of nutritionally beneficial eicosadienoic acid (C20:2n-6) mostly contributed to the PCB-138 bioaccumulation. Important impacts on PCB-138 behavior patterns were also registered for chemically allied indicator congeners (-153 and -180) and organochlorines' metabolite *p,p'*-DDE. Less prominent were the associations between target congener and the most toxic dioxin-like PCBs.

Keywords:
persistent organic pollutants (POPs), (omega-3-6) fatty acids, heavy metals, Shapley Additive exPlanations (SHAP), fuzzy methods.

1. INTRODUCTION

Anthropogenic activities have led to several global, regional and local environmental issues related to air, water and soil pollution, gradual decrease of the stratospheric ozone layer, decrease in biodiversity, etc. Organic and inorganic contaminant emissions and dispersion caused by human activities refer to different classes of polychlorinated biphenyls (PCBs), polycyclic aromatic pollutants, trace metalsand natural radioactivity. Organochlorine pesticides (OCPs) and PCBs are well-known as persistent organic pollutants (POPs). They are long-lived contaminants and possess numerous adverse effects on living organisms including humans and animals. Because of that their production and use is limited

Correspondence:
Andreja Stojić

e-mail:
andreja.stojic@ipb.ac.rs

DOI: 10.15308/Sinteza-2020-23-28

assumed to have no impacts on the contaminant uptake. The content of macro-elements and heavy metals is not related to PCB-138 chemodynamics in fish tissue, while influences of *p,p'*-DDE and both indicator and toxicologically congeners (-101, -118, -123, -153, and -180) were evident. Finally, methods have been successfully verified as a reliable means for examination of the relationships between POPs and FAs that overpowers commonly employed statistical approaches.

ACKNOWLEDGMENT

The authors acknowledge funding provided by the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS.

REFERENCES

[1] Food and Agriculture Organization of the United Nations (FAO), "The state of world fisheries and aquaculture worldwide 2018 - Meeting the sustainable development goals" 2018, pp. 1-227, ISBN 978-92-5-130562-1.

[2] A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić, and L. Lazić, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition." Sci. Total Environ., 2019, pp. 140-147.

[3] G. Vuković, S. Herceg Romanić, Ž. Babić, B. Mustać, M. Štrbac, I. Deljanin, D. Antanasijević, "Persistent organic pollutants (POPs) in edible fish species from different fishing zones of Croatian Adriatic." Marine Pollut. Bull. 2018, pp. 71-80.

[4] J.M. Đinović-Stojanović, D.M. Nikolić, D.V. Vranić, J.A. Babić, M.P. Milijašević, L. Pezo, S. Janković "Zinc and magnesium in different types of meat and meat products from the Serbian market." J. Food Compost. Anal., 2017, pp. 50-54.

[5] A. Špirić, D. Trbović, D. Vranić, J. Đinović, R. Petronijević, V. Matekalo-Sverak, "Statistical evaluation of fatty acid profile and cholesterol content in fish (common carp) lipids obtained by different sample preparation procedures.", Anal. Chim. Acta, 2010, pp. 66-71.

[6] T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system", In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016, pp. 785-794.

[7] S. Lundberg, S. Lee, "A unified approach to interpreting model predictions". In Adv. Neural Inf. Process. Syst. 2017, pp. 4765-4774.

[8] E. Brucka-Jastrzêbska, D. Kawczuga, M. Rajkowska, M. Protasowick, "Levels of microelements (Cu, Zn, Fe) and macroelements (Mg, Ca) in freshwater fish." J. Elementol., 2009, pp. 437-447.

[9] European Commission, 2006. Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs 20.12.2006 Off. J. Eur. Union, pp. L 364, 5-24.

[10] European Commission, 2011. Commission Regulation (EU) No 1259/2011 of 2 December 2011 amending Regulation (EC) No 1881/2006 as regards maximum levels for dioxins, dioxin-like PCBs and non-dioxin-like PCBs in foodstuffs. Off. J. Eur. Union, pp. L 320, 18-23.

[11] M. Á. Rincón-Cervera, V. González-Barriga, R. Valenzuela, S. López-Arana, J. Romero, A. Valenzuela, "Profile and distribution of fatty acids in edible parts of commonly consumed marine fishes in Chile." Food Chem, 2019, pp. 123-129.

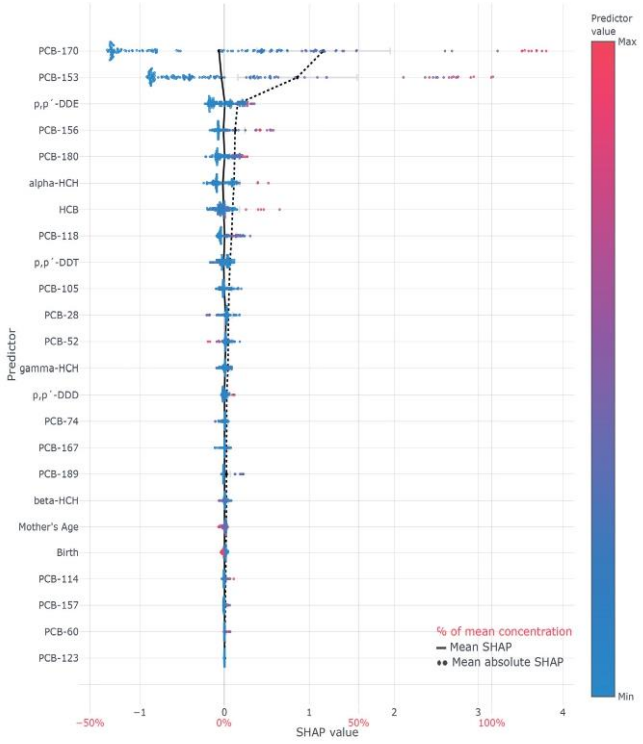


Fig. 4. PCB-138 SHAP summary.

ACKNOWLEDGMENT

The authors acknowledge funding provided by the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS.

REFERENCES

[1] J. Fång, E. Nyberg, U. Winnberg, A. Bignert, and Å. Bergman, "Spatial and temporal trends of the Stockholm Convention POPs in mothers' milk — a global review." Environ. Sci. Pollut. Res., 2015, pp. 8989–9041.

[2] United Nations Environment Programme, 2013. Results of the global survey on concentrations in human milk of persistent organic pollutants by the United Nations Environment Programme and the World Health Organization, Conference of the Parties to the Stockholm Convention on Persistent Organic Pollutants, Sixth meeting, Geneva, 28 April–10 May

[3] S. Lignell, A. Winkvist, F. Bertz, K.M. Rasmussen, A. Glynn, M. Aune, and H.K. Brekke "Environmental organic pollutants in human milk before and after weight loss." Chemosphere, 2016, pp. 96–102.

[4] A. Polder, J.U. Skaare, E. Skjerve, K.B. Løken, and M. Eggesbø "Levels of chlorinated pesticides and polychlorinated biphenyls in Norwegian breast milk (2002–2006), and factors that may predict the level of contamination." Sci. Total Environ., 2009, pp. 4584–4590.

[5] D. Klinčić, S. Herceg Romanić, M. Matek Sarić, J. Grzunov, and B. Dukić, "Polychlorinated biphenyls and organochlorine pesticides in human milk samples from two regions in Croatia." Environ. Toxicol. Pharmacol, 2014, pp. 543-552.

[6] D. Klinčić, S. Herceg Romanić, I. Brčić Karačonji, M. Matek Sarić, J. Grzunov Letinić, and N. Brajenović "Organochlorine pesticides and PCBs (including dl-PCBs) in human milk samples collected from multiparae from Croatia and comparison with primiparae." Environ. Toxicol. Pharmacol, 2016, pp. 74-79.

The results indicated that the listed pollutants have similar molecular structure and metabolic pathways, but we assumed that the advanced methods apart from commonly applied correlation matrices could be employed to deepen the current understanding of PCB-138 behaviour patterns in breast milk.

For investigating the non-linear relationships between PCB-138 and other congeners, mother's age, and number of births, the XGBoost regression analysis was successfully employed, with a relative error below 20% and high correlation coefficient (r=0.97) (*Fig. 3*). As shown by the highest positive (up to 4) and negative (up to -1.5) SHAP values, the most important variables that shaped PCB-138 behaviour pattern in the examined milk samples were PCB-170 and PCB-153 (*Fig. 4*).

The strongest influences were observed to be related to the elevated concentrations of the listed pollutants suggesting that mono-chlorine congeners are more prone to bioaccumulation in human milk compared to other PCBs.

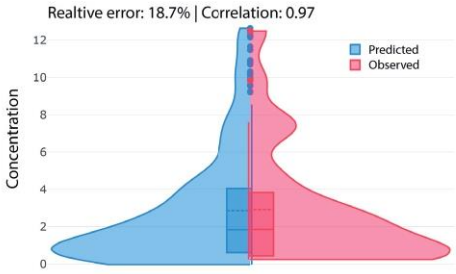


Fig. 3. XGBoost evaluation.

Position of the halogen substituting biphenyl ring of PCB-153, PCB-170 and PCB-153 provide such structure and rigidity of the molecules that facilitate PCBs' ability to pass from blood to breast milk. The other molecular properties including lipophilicity, molecular diameter and weight, and the number of attached halogens appeared to be less determinative for the PCB partitioning between blood and milk [9].

The SHAP analyses revealed less important impacts of PCB-156, PCB-180 and PCB-118 on the PCB-138 behaviour patterns in milk samples. The pollutants belong to non-dioxin-like/indicator congener group (-138,–153 and –180) and toxicologically relevant PCBs (–118 and –156), which elicit aryl hydrocarbon receptor-mediated biochemical and toxic responses and resist in the food chain. Minor negative functional dependency was observed between PCB-153 and the elevated levels of low-chlorinated congeners (PCB-52 and PCB-28), which are more volatile and susceptible to metabolic breakdowns and excretion.

Numerous studies reported that POPs are eliminated from the body during the breastfeeding, which results in the lower levels of the pollutants in the breast milk of multipara. In addition, it has been usually reported that the milk of elder mothers contained higher concentrations of organochlorine xenobiotics [10]. As indicated by low SHAP values approaching zero, in this study, no significant influences of the mother's age and parity on the PCB-153 patterns was recorded (*Fig. 3*). For prediction of the POPs bioaccumulation in breast milk, the exposure routes including dietary habits, residential and working environment as well as the mother's childhood nursing history should deserve more attention in future research.

4. CONCLUSION

The health burden of organochlorines in breast milk represents an issue of global concern because of the POPs adverse impacts on human health, particularly sensitive sub-populations such as women and children. In this study, we presented promising explainable artificial intelligence methodology (XGBoost and SHAP) with the aim to investigate the organochlorine patterns as well as the PCB-138 dependence on the mother's age and child delivery. According to the results, similarly structured molecules, which belong to both non-dioxin-like/indicator congener group (–28, –52, –180) and toxicologically relevant PCBs (–118, –189, -156) as well as *p,p'*-DDT metabolite (*p,p'*-DDE) have impacts on PCB-138 distribution. No significant functional dependencies of the PCB-138 patterns and the maternal age and parity was observed suggesting that the impacts of dietary habits and health burden POPs in the residential and working environment should be more investigated in the future. The herein applied analyses could serve as a promising methodology for future epidemiological investigations and human health protection.

was implemented using a brute-force grid search and 10-fold stratified cross-validation. The best performing hyperparameter values were used for the final model.

The explainability of the produced XGBoost model that operates with high-dimensional input data in a non-linear fashion was obtained by using explainable artificial intelligence method SHapley Additive exPlanations (SHAP) [5]. Based on the game theory, the Shapley explanations represent the only possible locally accurate and globally consistent feature attribution values. The captured attributed importance of a feature is visually presented as SHAP summary plot.

3. RESULTS AND DISCUSSION

The concentrations of organochlorines found in the milk samples of primiparae, secundiparae and multiparae aged between 19 and 45 (with an average age of 30 years) are given in *Fig. 1*. The most abundant pollutants were the by-products of lindane (γ-HCH and HCB), the organochlorine pesticide metabolites (*p,p'*-DDE and *p,p'*-DDD), and heavy hexa- and hepta-chloro non-dioxin like congeners (PCB-138, PCB-153, PCB-180 and PCB-170).

Significant linear correlation coefficients (r ≥0.90) were found between the following pairs of the investigated variables: PCB-170–PCB-138; PCB-170–PCB-153; PCB-170–PCB-180; PCB-153–PCB-180 and PCB-153–PCB-138 (*Fig. 2*).

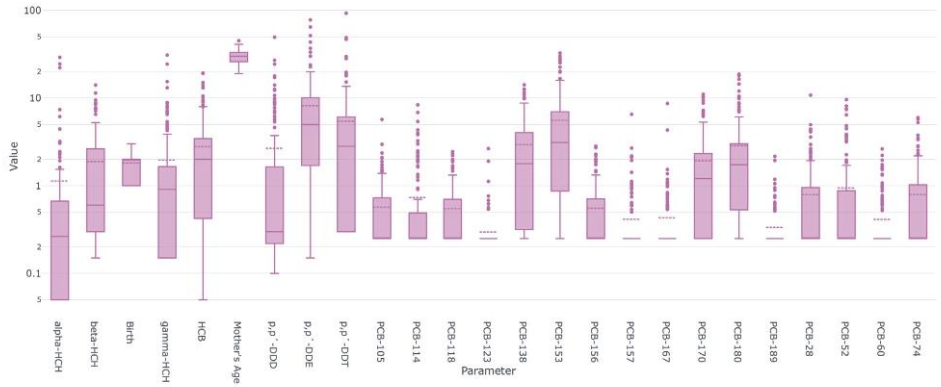


Fig. 1. Measured parameter box plots.

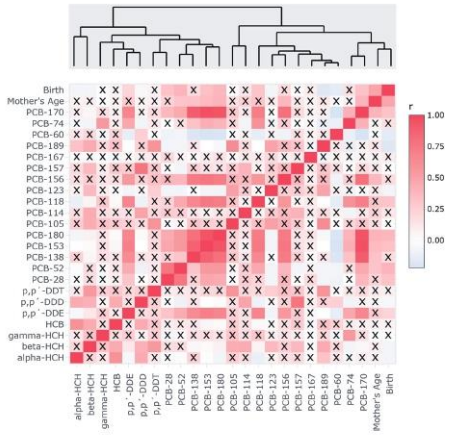


Fig. 2. Obtained parameter correlation matrix.

Research has shown that organochlorine compounds, as endocrine modulators, affect human health. The toxic potential of individual compounds is different, but the overall toxic effect is additive. According to epidemiological and laboratory studies, depending on the compound, these toxins can affect the development of cancer, cause allergy and hypersensitivity, as well as central and peripheral nervous system damage, reproductive disorders and immune system disorders. Humans and wildlife are exposed to PCBs and organochlorine pesticides in the polluted environment via air, water, sediment soil and food, by ingestion, inhalation or skin contact. Previous studies suggest that more of these compounds is introduced into the body by food intake than by inhalation or through the skin. It is also known that unborn children are exposed to organochlorine compounds that pass through the placenta. Biological monitoring of these compounds in the body requires invasive techniques referring to the surgical extraction of adipose tissue. The alternative monitoring approach and one of the mostly used non-invasive methods for the assessment of human exposure is the analysis of human milk POPs content.

A number of literature sources reported on POPs interrelations in the breast milk or/and their dependence on parameters like mother's age and weight loss, childbirth, dietary habits and occupation [1, 2, 3, 4]. Our previous investigations were aimed at investigating the presence of OCPs and PCBs in human milk as well as their mutual interrelations and associations with the mother's age and parity using machine learning (ML) algorithms [5, 6, 7]. In this study, we applied the additive feature attribution method – Shapley Additive exPlanations (SHAP), which offers uniquely consistent and locally accurate attribution values in comparison with the conventional attribution methods and has become increasingly popular as a tool for predicting environmental phenomena [8]. We aimed to obtain a detailed insight into the parameters including organochlorine pollutants, mother's age and parity, that shape the PCB-138 distribution in the mother's milk. We chose this congener since it has been considered among PCB_6 group as indicator for the PCB behavior patterns in various samples and it has been also taken as the most suitable target compound for evaluating non-dioxin-like PCBs. The classification of non-dioxin- and dioxin-like congeners is used in the European food and feed regulation where dioxin-like PCBs resemble biochemical and toxicological characteristics of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD).

2. MATERIAL AND METHODS

Sampling

Breast milk samples were collected from 152 healthy primiparae, secundiparae and multiparae (third child delivery), aged between 19 and 45, living in Zadar (Croatia). The mothers had no history of accidental or occupational exposure to the analyzed POPs. More details about sampling were described previously [7].

Chemical analysis of PCBs and OCPs

The analytical procedure was described in details previously [5]. Briefly, two subsamples of each unfrozen milk sample (5 g) were extracted twice with a mixture of chloroform and methanol and dried under a nitrogen flow. Subsequently, milk fat was dissolved in n-hexane and purification and fractionation were performed.

Seven OCPs [hexachlorobenzene (HCB), hexachlorocyclohexane isomers (α-, β-, and γ-HCH), 1,1,1-trichloro-2,2-di(4-chlorophenyl)ethane (*p,p'*-DDT), 1,1-dichloro-2,2-di(4-chlorophenyl)ethylene(*p,p'*-DDE), 1,1-dichloro-2,2-di(4-chlorophenyl)ethane (*p,p'*-DDD)], 17 PCB congeners [PCB-28, PCB-52, PCB-101, PCB-138, PCB-153, PCB-180 (six indicator congeners), PCB-105, PCB-114, PCB-118, PCB-123, PCB-156, PCB-157, PCB-167, PCB-189 (eight mono ortho congeners), PCB-60, PCB-74 and PCB-170] were analysed. High-resolution gas chromatography with electron capture detector (s) was performed on a CLARUS 500 chromatograph using two capillary columns (Restek, Bellefonte, PA, USA) simultaneously: (1) 60m×0.25 mm, Rtx-5 film thickness of 0.25 μm, and (2) 30m×0.25 mm, Rtx-1701 film thickness of 0.25 μm. The LODs for the analyzed compounds were 0.5 ng g^{-1} milk fat for PCB congeners, 0.1 ng g^{-1} milk fat for α-HCH and HCB, 0.2 ng g^{-1} milk fat for *p,p´*-DDE, 0.3 ng g^{-1} milk fat for β-HCH, γ-HCH and *p,p´*-DDD, and 0.6 ng g^{-1} milk fat for *p,p´*-DDT. The average recoveries for PCBs ranged between 58% and 86% and for organochlorine pesticides between 59% and 92%. Method reproducibility expressed as relative standard deviation was between 6% and 22%, and 7% and 24% for PCBs and OCPs, respectively.

Data analysis

The relationships between PCB-138 in breast milk and all other measured parameters were modeled by using XGBoost regression. The details on the method are given elsewhere [4]. In this study, we used Python XGBoost implementation. The dataset was split into training (80%) and validation (20%) sets. Hyperparameter tuning

SHAPLEY ADDITIVE EXPLANATIONS OF INDICATOR PCB-138 DISTRIBUTION IN BREAST MILK

Andreja Stojić[1, 2*],
Marijana Matek Sarić[3],
Snježana Herceg Romanić[4]

[1]Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
University of Belgrade,
Belgrade, Serbia

[2]Environment and Sustainable Development,
Singidunum University, Belgrade,
Serbia

[3]Department of Health Studies,
University of Zadar,
Zadar, Croatia

[4]Institute for Medical Research and
Occupational Health,
Zagreb, Croatia

Abstract:
Breastfeeding provides numerous health benefits for newborns by meeting the infants' nutritional needs and supporting associated immunological protection. Maternal milk is high in fat, and therefore, represents a very suitable media for the bioaccumulation of lipophilic pollutants such as organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs). This makes breast milk the primary source of the infant's postnatal exposure to persistent toxic xenobiotics. In this study, we applied a novel SHapley Additive exPlanations (SHAP) method to investigate the key parameters that govern distribution of PCB-138, an indicator of non-dioxin congeners, in the mother's milk. According to the accuracy metrics, the eXtreme Gradient Boosting regression was successfully employed, with relative error below 20% and high correlation coefficient (r=0.97), for finding the relationships between PCB-138 and other non-dioxin congeners, mother's age, and number of births. According to the results, PCB-156, PCB-180, HCB, HCH and PCB-118 have the major, while PCB-28, PCB-52 and PCB-189 have the minor impact on PCB-138 distribution in breast milk. Similar behavior of contaminants, which belong to the both indicator congener group (−28, −52, −180) and toxicologically relevant PCBs (−118, −189), was recognized. The SHAP conclusions were only partially consistent with the results of correlation analysis suggesting that POPs exhibit non-linear dynamics and interrelations. Therefore, the current knowledge on contamination of complex biomatrices would benefit from further detailed analyses of pollutant intermittent relationships.

Keywords:
human biomonitoring, persistent organic pollutants (POPs), organochlorine pesticides (OCPs), polychlorinated biphenyls (PCBs), Shapley Additive exPlanations (SHAP).

1. INTRODUCTION

Organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs) belong to a group of compounds known as Persistent Organic Pollutants (POPs). Due to their common properties such as persistence, bioaccumulation, toxicity, and ability to be transported via air over long distances, the production and use of POPs are eliminated and/or restricted by the Stockholm Convention. Additionally, due to their persistence and lipophilic properties, they tend to accumulate in the fatty tissues of humans and animals.

Correspondence:
Andreja Stojić

e-mail:
andreja.stojic@ipb.ac.rs

DOI: 10.15308/Sinteza-2020-35-40

[7] G. Jovanović, S. Herceg Romanić., A. Stojić, D. Klinčić, M. Matek Sarić, J. Grzunov Letinić, and A. Popović, "Introducing of modeling techniques in the research of POPs in breast milk – A pilot study." Ecotoxicol. Environ. Saf., 2019, pp. 341-347.

[8] A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić, and L. Lazić, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition." Sci. Total Environ., 2019, pp. 140-147.

[9] G. Vasios, A. Kosmidi, O.-I. Kalantzi, A. Tsantili-Kakolidou, N. Kavantzas, S. Theocharis, and C. Giaginis "Simple physicochemical properties related with lipophilicity, polarity, molecular size and ionization status exert significant impact on the transfer of drugs and chemicals into human breast milk." Expert Opin. Drug Metab. Toxicol., 2016, pp. 1273–1278.

[10] R. Aerts, I. Van Overmeirea, A. Colles, M. Andjelković, G. Malarvannan, G. Poma, E. Den Hond, E. Van de Mieroop, M.-C. Dewolf, F. Charlet, A. Van Nieuwenhuyse, J. Van Loco, A. Covaci, "Determinants of persistent organic pollutant (POP) concentrations in human breast milk of a cross-sectional sample of primiparous mothers in Belgium", Environ. Int., 2019, pp. 104979.

THE USE OF INNOVATIVE METHODOLOGY FOR THE CHARACTERIZATION OF BENZENE, TOLUENE, ETHYLBENZENE AND XYLENE SOURCES IN THE BELGRADE AREA

Svetlana Stanišić[1*],
Mirjana Perišić[1,2],
Andreja Stojić[1,2]

[1]Singidunum University,
Belgrade, Serbia
[2]Environmental Laboratory, Institute of Physics Belgrade,
National Institute of the Republic of Serbia,
University of Belgrade,
Belgrade, Serbia

Abstract:
The growth of urban population, economic development, urbanization and transport have a strong impact on environmental pollution. The increase in air pollutant concentrations over the last few decades has been in focus of contemporary science and research mainly for its adverse effects on public health, environment and climate change. In this paper, we are using the innovative integrated methodology for spatio-temporal air pollution modeling, based on receptor-oriented air circulation modeling and artificial intelligence implemented through machine learning methods for detailed characterization of toxic, mutagenic and carcinogenic representatives of volatile organic species – benzene, toluene, ethylbenzene and xylene, in the Belgrade area. Also, the study evaluates the possibilities of spatio-temporal forecast based on the integrated methodology. The results suggest that temperature and wind speed represent the main parameters which govern the spatio-temporal distribution of benzene, while the impact of other factors shows significant variations depending on the characteristics of receptor location.

Keywords:
Keywords – BTEX, artificial intelligence, machine learning, volatile organic compounds.

1. INTRODUCTION

Ambient air pollution accounts for an estimated 4.2 million deaths per year due to cardiovascular, malignant and chronic respiratory diseases [1]. Around 91% of the world's population lives in places where air pollution levels exceed World Health Organization limits [2]. Holgate (2017) emphasizes that 40,000 excess deaths in the UK annually can be associated with low air quality, and society would be much more aware of its significance if this mortality was the consequence of drinking polluted water [3].

Among the air pollutants that are of interest for current and future research due to their detrimental effects on both human health and the environment are volatile organic compounds (VOCs), a heterogeneous

Correspondence:
Svetlana Stanišić

e-mail:
sstanisic@singidunum.ac.rs

DOI: 10.15308/Sinteza-2020-41-45

group of organic species with boiling points <250 °C. Their representatives are benzene, toluene, ethylbenzene and xylene, commonly referred to as BTEX. Over the last few decades in developed countries, reducing the levels of BTEX is still challenging [4], due to their enormous chemical diversity and abundance, their numerous emission sources, their complex atmospheric chemistry, insufficient funds for establishment and maintenance of monitoring networks, and the fact that abatement programs might have negative impacts on economic output.

The health effects of BTEX are diverse. For instance, the research has shown that long-term exposure to benzene increases the risk of developing malignant blood disorders, while long-term exposure to toluene causes renal tubular acidosis [5]. Furthermore, the studies have shown that after reduction of benzene, styrene, and tetrachloroethylene concentrations in industrial and urban areas, lifetime cancer risk decreased by one order of magnitude [6]. Populations in highly industrialized areas, socioeconomically deprived, as well as children, pregnant women and elderly people, appear to be more susceptible to pollution-related morbidity and mortality [7]. Apart from their impact on human health, BTEX and other VOCs are associated with climate change and increases in the oxidation capability of the atmosphere [8]. Not only volatile species directly and indirectly contribute to climate change, but their emissions and fates are expected to be influenced and increased by the forthcoming global warming.

Despite the fact that the big shifts in development and integration of different approaches in the area of environmental science have been made recently, spatio-temporal air pollution modeling remained a challenge. Two main approaches are typically employed to forecast air quality and to identify the factors that govern certain pollutant concentrations. The first approach relies on atmospheric diffusion models, while the second refers to statistical models that capture the essential relationships between the variables [9]. Thereby, multidimensionality and size of data sets, as well as the complexity of air pollutant processes and interactions, set too high requirements for conventional statistical methods. For this reason, methods of machine learning, a subfield of artificial intelligence that enables automatized big data analysis and development of learning algorithms, have been introduced into environmental science and research. In this paper, we used the innovative and integrated methodology for spatio-temporal air pollution modeling, based on artificial intelligence and implemented through machine learning methods for detailed characterization of dominant and particular sources of BTEX in a wider region surrounding receptor site that was not covered by regular monitoring. The presented methodology has the potential to provide the basis for establishment of unique and sustainable system for air pollution source identification and enhanced air pollution data coverage that doesn't require additional investments in monitoring equipment. In long term, results of such an approach would provide a solid basis for establishing the sustainable system aimed at improved air pollution management and control.

2. METHODOLOGY

Machine learning algorithms are based on the extraction of patterns and selection of specific attributes from a large number of data, while eliminating irrelevant information. By identification of prediction most important attributes, machine learning methods acquire knowledge and define substantial relationships that exist between input and output parameters by placing a special focus on the data aspect that is most useful for efficient forecasting. The fact that methods based on decision trees, such as Gradient Boosting and Random Forest, have been shown to provide inconsistent attribute contributions, has led to the development of SHAP (SHapley Additive exPlanation), a method that estimates the contribution of each instance of an attribute, which further enables interpretation of the model outputs [10, 11].

EXtreme Gradient Boosting (XGBoost) can be assigned as ensemble method of supervised machine learning which combines the results of more than one decision tree approaches. The main feature of the XGBoost method is focus on obtaining more precise prediction, compared to the one that could be provided by applying a single constitutive decision algorithm. XGBoost method is based on boosting technique that sequentially defines smaller series of decision trees for classifying input data into two or more attribute-defined classes. Each consecutive decision tree is trained through iterations by taking into account the registered errors of previous classification.

3. RESULTS AND DISCUSSION

The variations of particular meteorological parameter affect the changes in other related parameters, which makes it difficult to distinguish between their particular impact on air pollution phenomena. Thus, the impact of meteorological factors is not observed as isolated impact of a single parameter and its variations, but rather as an impact of a certain weather type. In addition to this, a number of other factors can contribute to final impact of meteorological conditions, including the distribution of pollutant emission sources, local topography, street geometry and distribution of all elements and surfaces that can be of significance for air flow regime, pollutant dispersion conditions, their transport pathways and thus, the spatio-temporal variability of their levels. For instance, Liao et al. (2017) have identified ten typical air circulation types within one of the most polluted areas of China and explored their synergetic contribution with topography to local air quality [12]. Ning et al. (2019) have shown that air pollution forecast can be significantly impacted by the complexity of terrain areas because topographic features can to a certain extent limit pollutant dispersion under different weather conditions [13]. In addition to this, the consistency of meteorological conditions significantly affects the extent of volatile pollutant dispersion. For instance, previous study which has been dealing with the accidental benzene release risk assessment in an urban area using an atmospheric dispersion model has shown that benzene spreads over a much larger area during the nighttime due to a stable boundary layer, while during the daytime the enhanced vertical mixing results in limited dispersion of pollutant over the study area [14].

The results of this study suggest that low temperatures and weak to moderate wind represent the main parameters which govern the spatio-temporal distribution of benzene, while the impact of other factors shows significant variations depending on the characteristics of receptor location (Fig. 1). In a similar manner, the presented figures demonstrating TEX distribution can be interpreted (Fig. 2). As can be seen, the figures also include the relative errors and correlations between the observed and estimated BTEX concentrations, which can contribute to better XGBoost method performance evaluation.

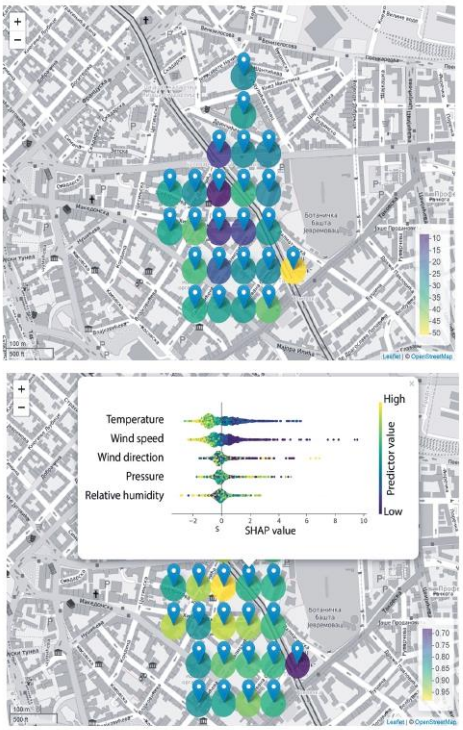


Fig. 1. Benzene forecast based on meteorological parameters – relative error [%] (above) and SHAP values and predicted/observed correlation coefficients (below).

Based on high correlation coefficients (>0,80) that were obtained for most of the analyzed data, it can be concluded that XGBoost can be rated as successful and efficient method for air pollution forecasting in the urban area. It should be emphasized that the estimated method errors are significantly lower than uncertainty (50%) which has been requested for evaluation of average annual benzene concentrations obtained by conventional modeling.

The results have also shown high correlations (r>0,70) between toluene and nitrate oxide concentrations, which can be considered as indicators of fossil fuel burning, which suggests the common origin of these pollutants in all locations being covered by the conducted analysis, except the old city area and Kalemegdan (Fig. 3), where the toluene concentrations most probably reflected the oxygenated air masses in some narrow canyon-type streets. Furthermore, high correlations (r>0,70) between benzene and inorganic oxides (NO_x, CO, SO_2) in western city region suggest the detrimental impact of remote air pollution sources, such as thermal plant Nikola Tesla A and B in Obrenovac (Fig. 4).

Fig. 2. Toluene (a), m,p-xylene (b), o-xylene (c), and ethylbenzene (d) relative error [%] forecasts based on meteorological parameters.

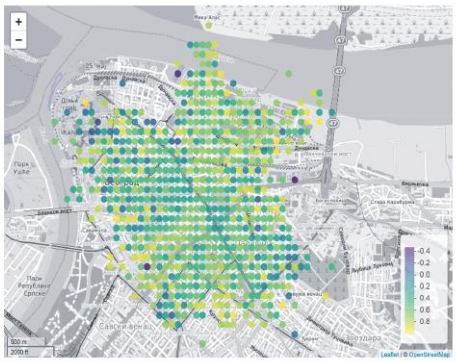


Fig. 3. Toluene and NO correlation coefficient.

Relatively low correlations between benzene and inorganic oxides in the northern and eastern city area suggest that benzene in this urban region can be related to evaporations and emissions coming from petrochemical industry, Oil refinery Pančevo and chemical industry Petrohemija.

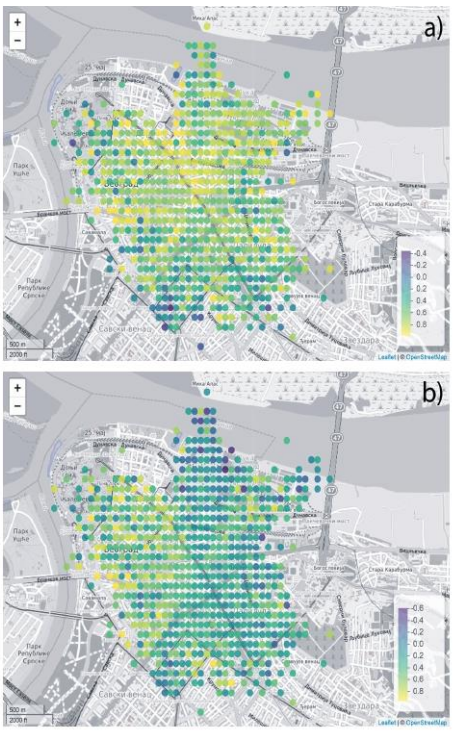


Fig. 4. Benzene correlation coefficients with CO (above) and NOx (below).

4. CONCLUSION

As can be concluded, we have demonstrated the use of efficient methods for spatio-temporal BTEX concentration modeling in the Belgrade area, based on receptor-oriented air circulation modeling and artificial intelligence implemented through machine learning and explainable artificial intelligence methods. The estimated method errors were shown to be lower than the requested uncertainty for conventional modeling. According to the results, temperature and wind speed represented the main parameters which governed the spatio-temporal distribution of benzene. In addition to this, the correlations between different air pollutant concentrations were considered for determination of their origin in all locations covered by the conducted analysis.

ACKNOWLEDGMENT

The authors acknowledge funding provided by the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS.

REFERENCES

[1] S. Rajagopalan, A-K. Sadeer, D. R. Brook, "Air pollution and cardiovascular disease: JACC state-of-the-art review," *J. Am. Coll. Cardiol.* vol. 72, pp. 2054-2070, 2018.

[2] C. Y. Wright, D. A. Millar, "A global statement for air pollution and health," *Clean Air J,* vol. 29, pp. 1-2, 2019.

[3] S. T. Holgate, "Every breath we take: the lifelong impact of air pollution'–a call for action," *Clin. Med*, vol. 17, pp. 8, 2017.

[4] M. J. Milazzo, J. M. Gohlke, D. L. Gallagher, A. A. Scott, B. F. Zaitchik, L. C. Marr, "Potential for city parks to reduce exposure to BTEX in air," *Environ. Sci. Process. Impacts,* vol. 21, pp. 40-50, 2019.

[5] E. J. Werder, L. S. Engel, A. Blair, R. K. Kwok, J. A. McGrath, D. P. Sandler, "Blood BTEX levels and neurologic symptoms in Gulf states residents," *Environ. res,* 175, pp. 100-107, 2019.

[6] J. E. C. Lerner, T. Kohajda, M. E. Aguilar, L. A. Massolo, E. Y. Sánchez, A. A. Porta, P. Opitz, G. Wichmann, O. Herbarth, A. Mueller, "Improvement of health risk factors after reduction of VOC concentrations in industrial and urban areas," *Environ. Sci. Poll. Res,* vol. 21, pp. 9676-9688, 2014.

[7] S. Bose, G. B. Diette, "Health disparities related to environmental air quality," *In Health disparities in respiratory medicine*, pp. 41-58, Humana Press, Cham, 2016.

[8] P. Campbell, Y. Zhang, F. Yan, Z. Lu, D. Streets, "Impacts of transportation sector emissions on future US air quality in a changing climate. Part II: Air quality projections and the interplay between emissions and climate change," *Environ. Poll*, vol. 238, pp. 918-930, 2018.

[9] H. B. Ly, L. M. Le, L. V. Phi, V. H. Phan, V. Q. Tran, B. T. Pham, T. T. Le, S. Derrible, "Development of an AI model to measure traffic air pollution from multisensor and weather data," *Sensors*, vol. 19, pp. 4941, 2019.

[10] M. V. García, J. L. Aznarte, "Shapley additive explanations for NO_2 forecasting," *Ecol. Inform,* vol. 56, pp. 101039, 2020.

[11] A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić, L. Lazić, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition" Sci Total Environ, vol. 653, pp.140-147, 2019.

[12] Z. Liao, M. Gao, J. Sun, S. Fan, "The impact of synoptic circulation on air quality and pollution-related human health in the Yangtze River Delta region," *Sci. Tot. Environ,* vol. 607, pp. 838-846, 2017.

[13] G. Ning, S. H. L. Yim, S. Wang, B. Duan, C. Nie, X. Yang, J. Wang, K. Shang, K, "Synergistic effects of synoptic weather patterns and topography on air quality: a case of the Sichuan Basin of China," *Climate Dynamics*, vol. 53, pp. 6729-6744, 2019.

[14] S. C. Truong, M. I. Lee, G. Kim, D. Kim, J. H. Park, S. D. Choi, G. H. Cho, "Accidental benzene release risk assessment in an urban area using an atmospheric dispersion model," *Atmos. Environ*, vol. 144, pp. 146-159, 2016.

WeBIOPATR 2023

The Ninth International WEBIOPATR
Workshop & Conference
Particulate Matter: Research and Management

Abstracts of Keynote Invited Lectures and Contributed Papers

Milena Jovašević-Stojanović,
Alena Bartoňová,
Danka Stojanović and Simon Smith, Eds

VInča Insitute of Nuclear Sciences
National Institute of the Republic of Serbia, University of Belgrade
Vinča, Belgrade 2023

ABSTRACTS OF KEYNOTE INVITED LECTURES AND CONTRIBUTED PAPERS

The Ninth WeBIOPATR Workshop & Conference

Particulate Matter: Research and Management

WeBIOPATR 2023

28th November to 1st December 2023

Vinča, Belgrade, Serbia

Editors

Milena Jovašević-Stojanović

Alena Bartoňová

Danka Stojanović and Simon Smith

Publisher

Vinča Institute of Nuclear Sciences

Prof. Dr Snežana Pajović, Director

P.O.Box 522

11001 Belgrade, Serbia

Printed by

Vinča Institute of Nuclear Sciences

Number of copies: 100

ISBN-978-86-7306-177-1

www.vin.bg.ac.rs/

SCIENTIFIC COMMITTEE

Aleksandar Jovović, Serbia

Alena Bartoňová, Norway

Antonije Onjia, Serbia

David Broday, Israel

Dikaia Saraga, Athens, Greece

Griša Močnik, Slovenia

Ivan Gržetić, Serbia

María Cruz, Spain

Milena Jovašević-Stojanović, Serbia

Miloš Davidović, Serbia

Saverio de Vito, Italy

Selahattin Incecik, Turkey

Slobodan Ničković, Serbia

Simone Barreira Morais, Portugal

Zoran Mijić, Serbia

Zoran Ristovski, Australia

Zorana Jovanović-Andersen, Denmark

Renata Kovačević, Serbia

ORGANIZING COMMITTEE

Aleksandra Stanković, Serbia

Alena Bartoňová, Norway

Andrej Šoštarić, Serbia

Anka Cvetković, Serbia

Bojana Petrović, Serbia

Bojan Radović, Serbia

Branislava Matić, Serbia

Lidija Marić-Tanasković, Serbia

Ivan Lazović, Serbia

Danka Stojanović (Secretary), Serbia

Duška Kleut (Secretary), Serbia

Marija Živković, Serbia

Maja Jovanović, Serbia

Milena Jovašević-Stojanović, Serbia

Miloš Davidović, Serbia

Mira Aničić Urošević, Serbia

Mirjana Perišić, Serbia

Nenad Živković, Serbia

Tihomir Popović, Serbia

Uzahir Ramadani (Secretary), Serbia

Vesna Slepčević, Serbia

CONFERENCE TOPICS

1. **Atmospheric Particulate Matter - Physical and Chemical Properties**
    i. Sources and formation of particulate matter
    ii. Particulate matter composition and levels outdoors and indoors
    iii. Environmental modeling
    iv. Nanoparticles in the environment
2. **Particulate Matter and Health**
    i. Exposure to particulate matter
    ii. Health aspects of atmospheric particulate matter
    iii. Full chain approach
    iv. COVID-19 and particulate matter
3. **Particulate Matter and Regulatory Issues**
    i. Issues related to monitoring of particulate matter
    ii. Legislative aspects
    iii. Abatement strategies

Organizers

Vinča Institute of Nuclear Sciences, Serbia

Environment and Climate Research Institute NILU, Norway

The 9th WeBIOPATR Workshop and Conference,
Particulate Matter: Research and Management, WEBIOPATR 2023
is supported by:

EC H2020 Framework Program for Research and Innovation, area "Spreading excellence and widening participation", VIDIS project (2020-2023) coordinated by Vinca Institute of Nuclear Sciences, Grant agreement number 952433.

EC Horizon Europe Framework Program for Research and Innovation, area "Spreading excellence and widening participation", WeBaSOOP project (2022-2025) coordinated by Vinca Institute of Nuclear Sciences, Grant agreement number 101060170

Ministry of Science, Technological Development and Innovation, Republic of Serbia

TABLE OF CONTENTS

11.12 THE IMPACT OF THE COVID-19 PANDEMIC ON PARTICULATE AIR POLLUTION IN THE REPUBLIC OF SERBIA

M. Perišić (1,2), A. Stojić (1,2), G. Jovanović (1,2)

(1) Institute of Physics Belgrade, a National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia, (2) University Singidunum, Belgrade, Serbia

mirjana.perisic@ipb.ac.rs

In recent years, the global community has been confronted with the profound challenges presented by the COVID-19 pandemic, which originated in late 2019 in China and swiftly escalated into a global crisis in 2020. To mitigate the spread of the virus, numerous countries implemented a range of measures, including restrictions on movement, the adoption of remote work arrangements, the transition to virtual education, and the curtailment of economic activities. From an environmental perspective, this worldwide health crisis, despite inflicting substantial negative impacts on public health and global well-being, yielded a favourable outcome in the reduction of pollutant emissions. Research outcomes consistently indicate an overall improvement in air quality during the lockdown period when compared to pre-lockdown, post-lockdown, and historical reference periods. Specifically, in the countries with higher levels of air pollution witnessed notable reductions in NO2 and PM concentrations. In contrast, the concentrations of O3 primarily saw an increase, while the levels of SO2 and CO exhibited greater variability (Rodríguez-Urrego and Rodríguez-Urrego, 2020; Kumari and Toshniwal, 2022).

The Republic of Serbia, like many other nations, declared a state of emergency during the pandemic, implementing curfews and various regulations aimed at curtailing citizen activities. The principal objective of this study was to evaluate the influence of the state of emergency and the accompanying regulatory measures on air quality in the Republic of Serbia. This evaluation was conducted through the analysis of pollutant concentrations obtained from monitoring stations within the National network of automatic stations for air quality monitoring (SEPA, 2023). While suspended particulate matter was chosen as the primary pollutant of interest, concentrations of other pollutants were also included in separate analyses.

Out of the 35 monitoring stations contributing data to the European Monitoring Network (EEA, 2023) in the Republic of Serbia, only five stations provided data for suspended particles during both the pre-COVID and COVID periods, with one station measuring both PM10 and PM2.5 concentrations. These selected monitoring points were predominantly located in urban areas, particularly in Belgrade and Niš, with the station at Kamenički Vis near Niš serving as the background location. The analysis spanned two eight-week periods: the first commencing with the declaration of the state of emergency on March 15, 2020, and concluding on May 6, 2020, while the second involved data collected during the same eight-week period from the preceding three years (i.e., from March 15 to May 6 in 2017, 2018, and 2019).

We used a sophisticated data analysis methodology encompassing statistical techniques, predictive machine learning models, and explainable artificial intelligence (Jovanovic et al, 2023; Stojić et al, 2022) to explain the complex relations between particle concentrations, the concentrations of other pollutants, meteorological variables, and anthropogenic activities. This approach is designed to foster a comprehensive understanding of the mechanisms that underlie the initiation of air pollution.

REFERENCES

EEA, 2023. https://www.eea.europa.eu/themes/air/country-fact-sheets/2022-country-fact-sheets/serbia-air-pollution-country

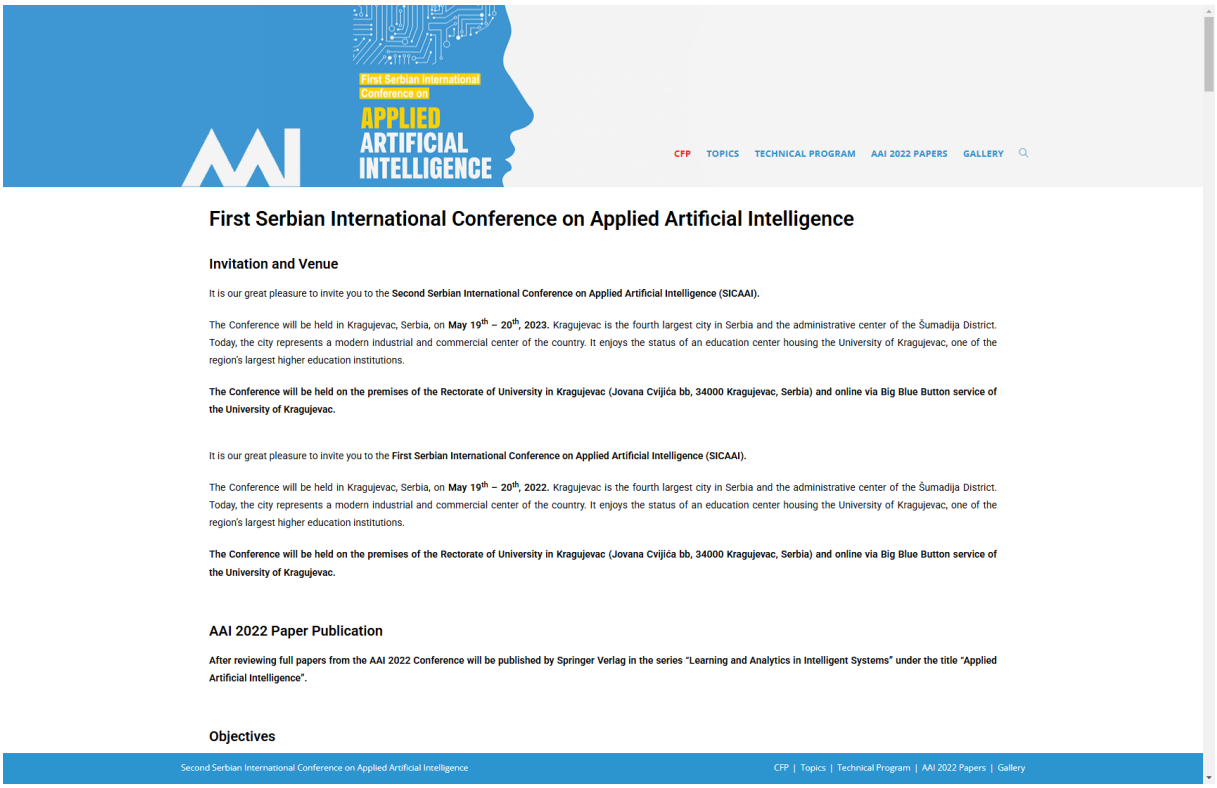
Jovanovic, et al, 2023. The explainable potential of coupling metaheuristics-optimized-xgboost and shap in revealing VOCs' environmental fate. Atmosphere 14.1

Kumari, P., & Toshniwal, D, 2022. Impact of lockdown measures during COVID-19 on airquality–A case study of India. International Journal of Environmental Health Research, 32(3), 503-510.

Rodríguez-Urrego, D., & Rodríguez-Urrego, L, 2020. Air quality during the COVID-19: PM2.5 analysis in the 50 most polluted capital cities in the world. Environmental Pollution, 266, 115042.

Stojić, A., Jovanović, G., Stanišić, S., Romanić, S. H., Šoštarić, A., Udovičić, V, et al, 2022. The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: Explainable prediction of benzo[a]pyrene levels. Chemosphere, 289, 133154.

SEPA, 2023. http://www.amskv.sepa.gov.rs/?lng=en


First Serbian International
Conference on
APPLIED
ARTIFICIAL
INTELLIGENCE
CFP
TOPICS
TECHNICAL PROGRAM
AAI 2022 PAPERS
GALLERY
First Serbian International Conference on Applied Artificial Intelligence
Invitation and Venue
It is our great pleasure to invite you to the Second Serbian International Conference on Applied Artificial Intelligence (SICAAI).
The Conference will be held in Kragujevac, Serbia, on May 19th – 20th, 2023. Kragujevac is the fourth largest city in Serbia and the administrative center of the Šumadija District. Today, the city represents a modern industrial and commercial center of the country. It enjoys the status of an education center housing the University of Kragujevac, one of the region's largest higher education institutions.
The Conference will be held on the premises of the Rectorate of University in Kragujevac (Jovana Cvijića bb, 34000 Kragujevac, Serbia) and online via Big Blue Button service of the University of Kragujevac.
It is our great pleasure to invite you to the First Serbian International Conference on Applied Artificial Intelligence (SICAAI).
The Conference will be held in Kragujevac, Serbia, on May 19th – 20th, 2022. Kragujevac is the fourth largest city in Serbia and the administrative center of the Šumadija District. Today, the city represents a modern industrial and commercial center of the country. It enjoys the status of an education center housing the University of Kragujevac, one of the region's largest higher education institutions.
The Conference will be held on the premises of the Rectorate of University in Kragujevac (Jovana Cvijića bb, 34000 Kragujevac, Serbia) and online via Big Blue Button service of the University of Kragujevac.
AAI 2022 Paper Publication
After reviewing full papers from the AAI 2022 Conference will be published by Springer Verlag in the series "Learning and Analytics in Intelligent Systems" under the title "Applied Artificial Intelligence".
Objectives
Second Serbian International Conference on Applied Artificial Intelligence
CFP | Topics | Technical Program | AAI 2022 Papers | Gallery

Session T.3A: 15:30-17:00

Applied AI in Industry – (part II)

Chair: Miloš Ivanović

T.3A.1 – Recurrent Neural Networks for the Prediction of Seat-to-head Transfer Functions – *Igor Saveljić, Slavica Mačužić Saveljić, Branko Arsić, Nenad Filipović*

T.3A.2 – Energy Management Platform Based on Automated Machine Learning – *Boško Laković, Nikola Andrijević, Lazar Krstić, Branka Jovanović, Dušan Stefanović, Nikola Bojović, Miloš Ivanović, Boban Stojanović*

T.3A.3 – Physisc Informed Neural Networks for 1D Flood Routing – *Filip Bojović, Miloš Milašinović, Branka Jovanović, Lazar Krstić, Boban Stojanović, Miloš Ivanović, Dušan Prodanović, Nenad Jaćimović, Nikola Milivojević*

T.3A.4 – Fast Self Learning Spectral Clustering – *Aleksandar Trokicić, Branimir Todorović*

T.3A.5 – Modeling 2D Steady Flow through a Porous Medium with Free Surface Using Physics-Informed Neural Networks – *Branka Jovanović, Miloš Ivanović, Boban Stojanović, Filip Bojović, Nikola Andrijević*

T.3A.6 – Artificial Intelligence in Revealing Air Pollution-related Processes – *Andreja Stojić, Mirjana Perišić, Gordana Jovanović, Svetlana Stanišić*

Session T.2B: 15:30-17:00

Mini-symposium V – Supercomputing for Artificial Intelligence – (part II)

Chair: Miloš Kotlar

T.2B.1 – Hybrid Supercomputing Architectures for Artificial Intelligence: Analysis of Potentials – *Jovan Popović, Vladisav Jelisavčić, Nenad Korolija*

T.2B.2 – Detection of Anomalies using Metadata in Automated Machine Learning Systems – *Miloš Kotlar*

T.2B.3 – Applications of Machine Learning in Query Execution of Database Systems – *Zoran Babović, Filip Hadžić*

T.2B.4 – Compensating the Lack of Big Data in Construction Industry Using Expert Knowledge: A Case Study – *Zoran Stojadinović*

T.2B.5 – Axial Strength Prediction of Square CFST Columns Based on The Ann Model – *Filip Đorđević, Svetlana Kostić*

T.2B.6 – Artificial Intelligence in Agriculture – *Miloš Kovačević, Petar Bursać, Branislav Bajat, Milan Kilibarda*

17:00 - 17:30	Presentation: **Winners of the "Best Technological Innovation in 2021" award, Farmbot d.o.o. Novi Sad** Coffee Break

Session T.4A: 17:30-18:15

ARTIFICIAL INTELLIGENCE IN REVEALING AIR POLLUTION-RELATED PROCESSES

Andreja M. Stojić[1,2], Mirjana D. Perišić[1,2], Gordana P. Jovanović[1,2], Svetlana M. Stanišić[2]

[1] Institute of Physics, Pregrevica 118, 11080 Belgrade, Serbia
e-mail: andreja.stojic@ipb.ac.rs
[2] Singidunum University, Danijelova 32, 11000 Belgrade, Serbia
e-mail: mirjana.perisic@ipb.ac.rs, gordana.vukovic@ipb.ac.rs, sstanisic@singidunum.ac.rs

Abstract:

In this study, we present one of the results of the Air Quality Plan for the Agglomeration of the City of Belgrade for the Period 2021-2031 project, supported by the City Administration of Belgrade. We used machine learning regression by means of the eXtreme Gradient Boosting (XGBoost) to interrelate PM_{10} concentrations and a number of environmental parameters in Belgrade, Serbia. The influence of meteorological factors affecting PM_{10} concentrations was investigated and explained using the SHapley Additive exPlanations (SHAP) attribution method. The ambient temperature is one of the most important parameters responsible for altering the PM_{10} concentrations in the range from -4 to 20 $\mu g\ m^{-3}$ relative to the average level.

Keywords: air pollution, particulate matter, machine learning, explainable artificial intelligence

1. Introduction

The modern world is facing many environmental issues, with all environmental compartments being affected. According to recent estimates, the mortality rate due to exposure to high levels of air pollution accounts for 8.9 million deaths annually [1]. Containing several hundred types of chemical species, some of which are toxic, mutagenic, and carcinogenic, particulate matter (PM) pollution rises as a significant problem in urban areas. The environmental fate of PM is governed by a diversity of emission sources, meteorological factors, or topographic features, as well as their mutual interrelations.

The understanding of environmental pollution-related processes is yet to be enhanced based on data-driven research. The enormous potential for the enhancement lies in the effective interrogation of environmental data using artificial intelligence, advanced statistical analysis, and numerical modelling, as well as modelling hybridization. This paper elaborates on the concept required for an in-depth understanding of environmental pollution from the perspective of contextual data analysis and the ATLAS Project (*Artificial Intelligence Theoretical Foundations for Advanced Spatio-Temporal Modelling of Data and Processes*) supported by the Science Fund of the Republic of Serbia [2]. ATLAS aims to shift the methodology and current approaches to modelling spatio-temporal data and processes related to the global fate of air pollutants. The shift will enhance the understanding of the global environmental fate of air pollutants and lead to more thoughtful environmental protection practices, policies, and strategies.

2. Methodology

In this study, we used the air pollution data (PM_{10}, benzene, SO_2, NO, NO_2, and NOx) obtained from air quality stations within the Institute of Public Health Belgrade network and meteorological parameters obtained from ARLs Global Data Assimilation System (GDAS1). The eXtreme Gradient Boosting (XGBoost) regression machine learning method was used to investigate the relation between PM and other air pollutants, and meteorological parameter dynamics in the urban area of Belgrade, Serbia. We used the SHapley Additive exPlanations (SHAP) explainable artificial intelligence method to investigate and interpret the governing factors in shaping PM_{10} levels. The method uses Shapley values, calculated as a measure of feature importance based on a game-theory approach, that provide an impact of features on individual predictions [3]. These values are considered as fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). The main advantage of the approach is that SHAP represents the only possible locally accurate and globally consistent feature attribution method. We used Python XGBoost and SHAP implementations, and the TreeExplainer which reduces the complexity of exact Shapley value computation from exponential to low-order polynomial time by leveraging the internal structure of tree-based models [4]. The relative SHAP values were calculated according to Stojić et al. 2022 [5]. The stabilities of the obtained SHAP values were evaluated the bootstrap method.

3. Results

The main focus of the Air Quality Plan for the Agglomeration of the City of Belgrade for the period 2021-2031 project, supported by the City Administration of Belgrade [6], was time-resolved, contextual, in-depth, and synergetic modelling of particulate matter pollution, primarily based on machine learning, explainable artificial intelligence, and numerical modelling within the ATLAS software platform.

The most important variables which explain PM level dynamics in the urban area of Belgrade include meteorological variables, such as momentum flux intensity, standard lifted index, volumetric soil moisture content, and temperature, as well as the concentrations of benzene, NO, NO_x, and SO_2. Figure 1 represents the example of the relation between the temperature and PM_{10} concentrations in the context of SO_2 concentrations.

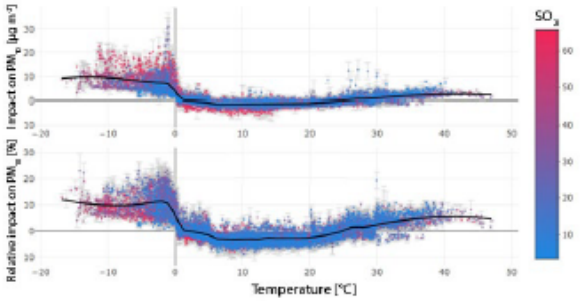


Fig. 1. The influence of temperature on PM_{10} concentrations in the context of SO_2 concentrations in Belgrade urban area

At temperatures below zero, an increase in PM_{10} concentrations by an average of 10 μg m^3 was observed. This may be the consequence of the impact of intensive combustion of fossil fuels for heating. In the ambient of high SO_2 concentrations, this increase may be as high as 20 μg m^{-3} which may be associated with the usage of fuels with high sulphur content. In the range from a few degrees above zero to about 25°C, the effect of temperature on

suspended particles is the least significant. During warmer weather, at temperatures above 25°C, the impact of temperature is related to an increase of PM concentrations to about 4 μg m^{-3} and may be related to the processes of particulates resuspension.

4. Conclusions

The understanding of environmental pollution-related processes must be enhanced by data-driven research. The concept supporting ATLAS can help to harmonize environmental research via facilitating access to environmental data, data analysis, exploration, and exploitation of the results. It can increase efficiency, creativity, and productivity of research and, at the same time, scale-up data analysis, support transdisciplinary, and lead to more thoughtful environmental protection practices, policies, and strategies. The inclusion of advanced technologies, such as artificial intelligence, in solving problems both locally and globally, clearly indicates benefits, as shown in the Air Quality Plan for the Agglomeration of the City of Belgrade project.

Acknowledgement: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI-ATLAS, as well as the City of Belgrade, Department of Environmental protection of the city administration, Serbia, Air quality plan for the City of Belgrade.

References

[1] Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C. A. et al, 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proceedings of the National Academy of Sciences, 115(38), 9592-9597.

[2] Science Fund of the Republic of Serbia - http://fondzanauku.gov.rs/atlas/?lang=en

[3] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 2522-5839.

[4] Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Tot. Environ. 653, 140-147.

[5] Stojić, A., Jovanović, G., Stanišić, S., Romanić, S.H., Šoštarić, A., Udovičić, V., Perišić, M. and Milićević, T., 2022. The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: explainable prediction of benzo [a] pyrene levels. Chemosphere, 289, p.133154.

[6] ATLAS Project - https://ai.ipb.ac.rs/

[7] AQP, 2021. Air quality plan for the City of Belgrade (In Serbian) – https://www.beograd.rs/images/data/7a0a3b18c076a6bfb21688ca7d314015_4302538578.pdf

WeBIOPATR 2021

The Eighth International WEBIOPATR
Workshop & Conference
Particulate Matter: Research and Management

Abstracts of Keynote Invited Lectures and Contributed Papers

Milena Jovašević-Stojanović,

Alena Bartoňová,

Miloš Davidović and Simon Smith, Eds

Vinča Insitute of Nuclear Sciences

Vinča, Belgrade 2021

ABSTRACTS OF KEYNOTE INVITED LECTURES AND CONTRIBUTED PAPERS

The Eighth WeBIOPATR Workshop & Conference

Particulate Matter: Research and Management

WeBIOPATR 2021

29th November to 1st December 2021

Vinča, Belgrade, Serbia

Editors

Milena Jovašević-Stojanović

Alena Bartoňová

Miloš Davidović

Simon Smith

Publisher

Vinča Institute of Nuclear Sciences

Prof. Dr Snežana Pajović, Director

P.O.Box 522

11001 Belgrade, Serbia

Printed by

Vinča Institute of Nuclear Sciences

ISBN 978-86-7306-164-1

www.vin.bg.ac.rs/

SCIENTIFIC COMMITTEE

Aleksandar Jovović, Serbia
Alena Bartoňová, Norway
Antonije Onjia, Serbia
David Broday, Israel
Dikaia Saraga, Greece
Griša Močnik, Slovenia
Ivan Gržetić, Serbia
María Cruz Minguillón, Spain
Milena Jovašević-Stojanović, Serbia
Miloš Davidović, Serbia
Saverio de Vito, Italy
Selahattin Incecik, Turkey
Slobodan Ničković, Serbia
Simone Barreira Morais, Portugal
Zoran Mijić, Serbia
Zoran Ristovski, Australia
Zorana Jovanović-Andersen, Denmark

ORGANIZING COMMITTEE

Aleksandra Stanković, Serbia
Alena Bartoňová, Norway
Andrej Šoštarić, Serbia
Anka Cvetković, Serbia
Biljana Filipović, Serbia
Branislava Matić, Serbia
Lidija Marić-Tanasković, Serbia
Uzahir Ramadani, Serbia
Ivan Lazović, Serbia
Sonja Dmitrašinović (Secretary), Serbia
Marija Živković (Secretary), Serbia
Milena Jovašević-Stojanović, Serbia
Miloš Davidović, Serbia
Mira Aničić Urošević, Serbia
Mirjana Perišić, Serbia
Nenad Živković, Serbia
Tihomir Popović, Serbia
Vesna Slepčević, Serbia
Viša Tasić, Serbia

CONFERENCE TOPICS

1. **Atmospheric Particulate Matter - Physical and Chemical Properties**
    i. Sources and formation of particulate matter
    ii. Particulate matter composition and levels outdoors and indoors
    iii. Environmental modeling
    iv. Nanoparticles in the environment
2. **Particulate Matter and Health**
    i. Exposure to particulate matter
    ii. Health aspects of atmospheric particulate matter
    iii. Full chain approach
    iv. COVID-19 and particulate matter
3. **Particulate Matter and Regulatory Issues**
    i. Issues related to monitoring of particulate matter
    ii. Legislative aspects
    iii. Abatement strategies

Organizers

Vinča Institute of Nuclear Sciences, Serbia

Public Health Institute of Belgrade, Serbia

NILU Norwegian Institute for Air Research, Norway

The 8th WeBIOPATR Workshop and Conference,

Particulate Matter: Research and Management, WEBIOPATR 2021

is supported by:

EC H2020 Framework Program for Research and Innovation, area "Spreading excellence and widening participation", VIDIS project (2020-2023) coordinated by Vinca Institute of Nuclear Sciences, Grant agreement number 952433.

Ministry of Education, Science and Technological Development of Republic of Serbia

TABLE OF CONTENTS

10.4 KEY FACTORS GOVERNING PARTICULATE MATTER ENVIRONMENTAL FATE IN AN URBAN ENVIRONMENT

Gordana Jovanović (1,2), Svetlana Stanišić (2), Mirjana Perišić (1,2), Andrej Šoštarić (3), Andreja Stojić (1,2)

(1) Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 118 Pregrevica Street, 11000 Belgrade, Serbia, (2) Singidunum University, 32 Danijelova Street, 11000 Belgrade, Serbia, (3) Institute of Public Health Belgrade, 54a Despota Stefana Street, 11000 Belgrade, Serbia

andreja.stojic@ipb.ac.rs

According to recent estimates, the mortality rate due to exposure to high levels of air pollution accounts for 8.9 million deaths annually (Burnett et al, 2018). Containing several hundred types of chemical species, some of which are toxic, mutagenic, and carcinogenic, particulate matter (PM) pollution rises as a significant problem in urban areas. The fate of PM is governed by a diversity of emission sources, meteorological factors, or topographic features, as well as their mutual interrelations.

In this study, we used the eXtreme Gradient Boosting (XGBoost) regression machine learning method to investigate the relation between PM and other air pollutants, and meteorological parameter dynamics in the urban area of Belgrade, Serbia. The air pollution data (PM_{10}, benzene, SO2, NO, NO_2, and NOx) was obtained from six air quality stations within the Institute of Public Health Belgrade network, while the meteorological parameters were obtained from ARLs Global Data Assimilation System (GDAS1).

We used the SHapley Additive exPlanations (SHAP) explainable artificial intelligence method to investigate and interpret the governing factors in shaping PM_{10} levels. The method uses Shapley values, calculated as a measure of feature importance based on a game-theory approach, that provide an impact of features on individual predictions (Lundberg et al, 2020). These values are considered as fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). The main advantage of the approach is that SHAP represents the only possible locally accurate and globally consistent feature attribution method. We used Python XGBoost and SHAP implementations, and the TreeExplainer method which reduces the complexity of exact Shapley value computation from exponential to low-order polynomial time by leveraging the internal structure of tree-based models (Stojić et al, 2019). The stabilities of the obtained SHAP values were evaluated by 50 times replicated bootstrap method.

As shown, the most important variables which describe PM level dynamics in the urban area of Belgrade include meteorological variables – momentum flux intensity, standard lifted index, volumetric soil moisture content and temperature, as well as the concentrations of benzene, NO, NOx, and SO_2.

Acknowledgements: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI-ATLAS, as well as the City of Belgrade, Department of Environmental protection of the city administration, Serbia, Air quality plan for the City of Belgrade.

REFERENCES

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C. A. et al, 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proceedings of the National Academy of Sciences, 115(38), 9592-9597.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 2522-5839.

Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Tot. Environ. 653, 140-147.

Innovation Center of Faculty of Mechanical Engineering

Faculty of Mechanical Engineering, University of Belgrade

Center for Business Trainings

„International Conference of Experimental and Numerical Investigations and New Technologies“

Sponsored by:

MINISTRY OF EDUCATION, SCIENCE AND TECHNICAL DEVELOPMENT

OF THE REPUBLIC OF SERBIA

Programme

and

The Book of Abstracts

29 June – 02 July 2021

Zlatibor, Serbia

„International Conference of Experimental and Numerical Investigations and New Technologies“

CNN TECH 2021

29 June – 02 July 2021

Hotel Mona, Miladina Pecinara 26, Zlatibor, Serbia

http://cnntechno.com

Programme

and

The Book of Abstracts

Organised by:

Innovation Center of Faculty of Mechanical Engineering

Faculty of Mechanical Engineering, University of Belgrade

Center for Business Trainings

Sponsored by:

Ministry of Education, Science and Technical development of the Republic of Serbia

Title:	International Conference of Experimental and Numerical Investigations and New Technologies – **CNN TECH 2021** **PROGRAMME AND THE BOOK OF ABSTRACTS**
Publisher:	Innovation Center of Faculty of Mechanical Engineering Kraljice Marije 16, 11120 Belgrade 35 tel: (+381 11) 3302-346, fax 3370364 e-mail: cnntechno@gmail.com web site: http://cnntechno.com, http://www.inovacionicentar.rs
Editors:	Dr Goran Mladenovic, Associate Professor Dr Martina Balac, Senior Scientific Researcher Dr Aleksandra Dragicevic, Scientific Researcher
Technical editor	Dr Goran Mladenovic, Associate Professor
Cover page:	Dr Goran Mladenovic, Associate Professor
Printed in:	Innovation Center of Faculty of Mechanical Engineering Kraljice Marije 16 11120 Belgrade 35 tel: (+381 11) 3302-346
Circulation:	100 copies. The end of printing: June 2021.

ISBN: 978-86-6060-077-8

"International Conference of Experimental and Numerical Investigations and New Technologies"

CNN TECH 2021

SCIENTIFIC COMMITTEE:

Milos Milosevic, Serbia (chairman)
Nenad Mitrovic, Serbia (co-chairman)
Aleksandar Sedmak, Serbia
Hloch Sergej, Slovakia
Drazan Kozak, Croatia
Nenad Gubeljak Slovenia
Monka Peter, Slovakia
Snezana Kirin, Serbia
Samardzic Ivan, Croatia
Martina Balac, Serbia
Mládková Ludmila, Czech Republic
Johanyák Zsolt Csaba, Hungary
Igor Svetel, Serbia
Aleksandra Mitrovic, Serbia
Valentin Birdeanu, Romania
Danilo Nikolic, Montenegro
Goran Mladenovic, Serbia
Bajic Darko, Montenegro
Tasko Manski, Srbija
Luis Reis, Portugal
Zarko Miskovic, Serbia
Tozan Hakan, Turkey
Nikola Momcilovic, Serbia
Traussnigg Udo, Austria
Gordana Bakic, Serbia
Katarina Colic, Serbia
Peter Horňak, Slovakia
Róbert Huňady, Slovakia
Martin Hagara, Slovakia
Jovan Tanaskovic, Serbia
Aleksa Milovanovic, Serbia
Marija Durkovic, Serbia
Tsanka Dikova, Bulgaria
Ján Danko, Slovakia
Ognjen Pekovic, Serbia
Jelena Svorcan, Serbia
Suzana Filipovic, Serbia
Darko Kosanovic, Serbia
Nebojsa Manic, Serbia
Zorana Golubovic, Serbia
Vera Pavlovic, Serbia

ORGANIZING COMMITTEE:

Nenad Mitrovic (chairman)
Milos Milosevic (co-chairman)
Aleksandar Sedmak
Martina Balac
Vesna Miletic
Igor Svetel
Goran Mladenovic
Aleksandra Mitrovic
Aleksandra Dragicevic
Zarko Miskovic
Katarina Colic
Milan Travica
Dragana Perovic
Aleksandra Joksimovic
Beti Kostadinovska Dimitrovska
Tsanka Dikova
Isaak Trajkovic
Toni Ivanov
Snezana Kirin
Igor Stankovic
Ivana Vasovic Maksimovic
Nina Obradovic
Andreja Stojic
Ivana Jevtic

enterprise europe network
Business Support on Your Doorstep

CONTENTS

THE POTENTIAL FOR FORECASTING THE PARTICULATE MATTER LEVELS IN COMPLEX URBAN ENVIRONMENT

Mirjana Perisic [1,2], Andreja Stojic[1,2*], Gordana Jovanovic[1,2], Andrej Sostaric[3], Dimitrije Maletic[1], Dusan Vudragovic[1], Svetlana Stanisic[2]

[1]Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 118 Pregrevica Street, 11000, Belgrade, Serbia

[2]Singidunum University, 32 Danijelova Street, Belgrade, 11000, Serbia

[3]Institute of Public Health Belgrade, 54 Despota Stefana Street, 11000, Belgrade, Serbia

*Corresponding author e-mail: andreja.stojic@ipb.ac.rs

Abstract

In this study, we employed regression analysis by means of machine learning eXtreme Gradient Boosting method for estimating the relationships between particulate matter (PM_{10}) concentrations and a number of parameters including benzene, inorganic gaseous pollutants (SO_2, NO, NO_2, NOx), Global Data Assimilation System-modeled (GDAS1) meteorological parameters, as well as daily and weekend PM variations in Belgrade, Serbia. The data was provided by the Institute of Public Health Belgrade, Serbia. The successful and reliable predictions were provided with relative errors in the range from approx. 19% to 26% and correlation coefficients higher than 0.95. The lowest relative error and the highest correlation coefficient were obtained for monitoring station of rural/industrial type located in Ovca, while the highest difference between modeled and measured values were detected at urban-type monitoring stations Novi Beograd and Institute of Public Health Belgrade, both of which are exposed to traffic emissions. The modeling results were not satisfying for rural/industrial monitoring station located in Veliki Crljeni (relative error>30%, corr. coefficient<0,8), which implies that the dynamic of PM_{10} emissions at the selected monitoring site were not governed by the available data on pollutant concentrations and meteorological parameters.

Keywords

Particulate matter, air pollution forecast, machine learning

Acknowledgement

We acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science, and Technological Development of the Republic of Serbia. This research was supported by the Science Fund of the Republic of Serbia, #GRANT No. 65241005, AI-ATLAS.

THE IMPACT OF HUMIDITY AND TEMPERATURE ON PARTICULATE MATTER ENVIRONMENTAL FATE

Andreja Stojic[1,2*], Gordana Jovanovic[1,2], Svetlana Stanisic[2], Andrej Sostaric[3], Ana Vranic[1], Marija Mitrovic Dankulov[1], Mirjana Perisic[1,2]

[1]Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 118 Pregrevica Street, 11000, Belgrade, Serbia

[2]Singidunum University, 32 Danijelova Street, Belgrade, 11000, Serbia

[3]Institute of Public Health Belgrade, 54 Despota Stefana Street, 11000, Belgrade, Serbia

*Corresponding author e-mail: andreja.stojic@ipb.ac.rs

Abstract

In urban environments, particulate matter, benzene, NOx, and SO_2 originate from common anthropogenic sources including traffic and industrial emissions, as well as fossil fuel burning for the purpose of heat and electricity production. In this study, the influence of relative humidity and temperature on PM_{10} concentrations in the Belgrade urban area was investigated using SHapley Additive exPlanations (SHAP) attribution method. In the presence of higher humidity or moisture in the re-suspended particles, and in the presence of soot and inorganic oxides as catalyzers (MgO_2 or Fe_2O_3), SO_2 will be adsorbed on the particle surface resulting in the secondary aerosol formation. Unlike SO_2, NOx tends to be less water-soluble, which makes them less prone to adsorption to the particle surface. In the presence of sunlight and hot weather, NOx and volatile organic compounds will more often participate in photochemical reactions with hydroxy, peroxy, and organic radicals in the air, resulting in the formation of tropospheric ozone.

The impact of intensive fossil fuel combustion for heating purposes contributes to an increase in PM_{10} concentrations by an average of 10 μg m^{-3}. In the case of using fuels with high sulfur content, this increase can be as high as 20 μg m^{-3}. With the temperature in the range from 0 to 25°C, the effect of temperature on suspended particles is negligible, while during warmer weather, at temperatures exceeding 25°C, the resuspension of particles contributes to an increase in particle concentrations to about 4 μg m^{-3} on average.

Keywords

atmospheric aerosols, meteorological factors, artificial intelligence

Acknowledgement

We acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science, and Technological Development of the Republic of Serbia. This research was supported by the Science Fund of the Republic of Serbia, #GRANT No. 65241005, AI-ATLAS
.

THE IMPACT OF GASEOUS POLLUTANTS ON PARTICULATE MATTER DISTRIBUTION

Svetlana Stanisic[2], Mirjana Perisic[1,2], Andreja Stojic[1,2*], Andrej Sostaric[3], Dusan Vudragovic[1], Dimitrije Maletic[1], Gordana Jovanovic[1,2]

[1]Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 118 Pregrevica Street, 11000, Belgrade, Serbia

[2]Singidunum University, 32 Danijelova Street, Belgrade, 11000, Serbia

[3]Institute of Public Health Belgrade, 54 Despota Stefana Street, 11000, Belgrade, Serbia

*Corresponding author e-mail: andreja.stojic@ipb.ac.rs

Abstract

In this study, we used eXtreme Gradient Boosting machine learning and SHapley Additive exPlanations (SHAP) explainable artificial intelligence methods to examine the relationships between PM_{10} and other air pollutant concentrations in the urban area of Belgrade. The data was provided by the Institute of Public Health Belgrade, Serbia. As shown, the most significant relative impact of benzene levels (50%) on the increase of PM_{10} concentrations up to several tens of μg m^{-3} was recorded at the occasions when benzene concentrations exceeded 5 μg m^{-3} and the concentrations of NO_2 were low (combustion of fossil fuels). The same effect was less pronounced at higher NO_2 concentrations. Taking into consideration the relative impact of SO_2 on PM_{10} levels and the observed relationship between NO and PM_{10}, four dominant environment types that describe the PM level dynamics were distinguished. In the first-type environment, the decrease of PM_{10} levels noticed for SO_2 levels below 50 μg m^{-3} and the dominance of sources with a significant share of NO (> 120 μg m^{-3}) were attributed to traffic emissions. The ambiance recognized as type 2 with no effect on PM levels is characterized by low gaseous oxide concentrations. The third and the fourth type of environment are characterized by SO_2 values exceeding 50 μg m^{-3} and their significant impact on the increase of PM_{10} concentrations.

Keywords

particulate matter, inorganic gaseous pollutants, explainable artificial intelligence

Acknowledgement

We acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science, and Technological Development of the Republic of Serbia. This research was supported by the Science Fund of the Republic of Serbia, #GRANT No. 65241005, AI-ATLAS.

SPATIO-TEMPORAL ANALYSIS OF MOBILITY PATTERNS IN THE CITY OF BELGRADE

Nikola Stupar[1], Ana Vranic[1], Andreja Stojic[1,2], Gordana Vukovic[1,2], Dusan Vudragovic[1], Dimitrije Maletic[1], Marija Mitrovic Dankulov[1*]

[1] Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia

[2]Singidunum University, 32 Danijelova Street, Belgrade, 11000, Serbia

Corresponding author e-mail: mitrovic@ipb.ac.rs

Abstract

Information about human mobility is essential for sustainable development planning. It is crucial for urban and transportation planning and the role of mobility in reducing contributions of human activities in air pollution and fighting climate change. While *human* mobility patterns are typically stable over space and time, they can dramatically change due to critical events, such as earthquakes and epidemics. We still lack a detailed description and understanding of how these critical events influence mobility patterns. In this work, we combine artificial intelligence tools with tools and methods from complex network theory to study human mobility patterns in the City of Belgrade during the COVID-19. We use data about mobility in the City of Belgrade between May 2020 and January 2021 from the Facebook Data for Good dataset. The City of Belgrade is divided into 201 cells with information on the mobility before and during the crises within and between the cells. The time resolution of the data is 8 hours. We use the k-mean clustering technique to cluster the data and find five clusters with different average mobility patterns. All mobility time series have trends and cycles, but they differ between clusters. While residential and suburban area clusters have the peaks of activity during the working days, clusters including central municipalities have the peak of activity during Saturday afternoon. Our results show that different city areas react differently to COVID19 information and that this aspect has to be considered in air pollution and crisis management planning.

Keywords

mobility patterns, time series analysis, k-means clustering

Acknowledgement

We acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science, and Technological Development of the Republic of Serbia. This research was supported by the Science Fund of the Republic of Serbia, #GRANT No. 65241005, AI-ATLAS. Numerical simulations were run on the PARADOX-IV supercomputing facility at the Scientific Computing Laboratory, National Center of Excellence for the Study of Complex Systems, Institute of Physics Belgrade.

ENVIRONMENTAL FACTORS GOVERNING PARTICULATE MATTER DISTRIBUTION IN AN URBAN ENVIRONMENT

Gordana Jovanovic[1,2], Svetlana Stanisic[2], Mirjana Perisic[1,2], Andrej Sostaric[3], Marija Mitrovic Dankulov[1], Ana Vranic[1], Andreja Stojic[1,2*]

[1]Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, 118 Pregrevica Street, 11000, Belgrade, Serbia

[2]Singidunum University, 32 Danijelova Street, Belgrade, 11000, Serbia

[3]Institute of Public Health Belgrade, 54 Despota Stefana Street, 11000, Belgrade, Serbia

Corresponding author e-mail: andreja.stojic@ipb.ac.rs

Abstract

In this study, we employed SHapley Additive exPlanations (SHAP) attribution method to investigate the PM_{10} concentrations in Belgrade (Serbia) and interpret the behavior of regression machine learning eXtreme Gradient Boosting method obtained on benzene, inorganic gaseous pollutants, Global Data Assimilation System-modeled meteorological parameters, as well as daily and weekend PM_{10} variations. The data was provided by the Institute of Public Health Belgrade, Serbia. As it was concluded, PM_{10} concentrations were dominantly defined by a variable assigned to emission source intensity variations. The impact of the emission sources on registered PM_{10} concentrations was not continual, but rather showed variations of up to 50%, when compared to impacts of other analyzed factors. The most important variables which describe PM level dynamics in the urban area of Belgrade include meteorological variables momentum flux intensity, standard lifted index, volumetric soil moisture content and temperature, as well as concentrations of benzene, NO, NOx, and SO_2.

Keywords

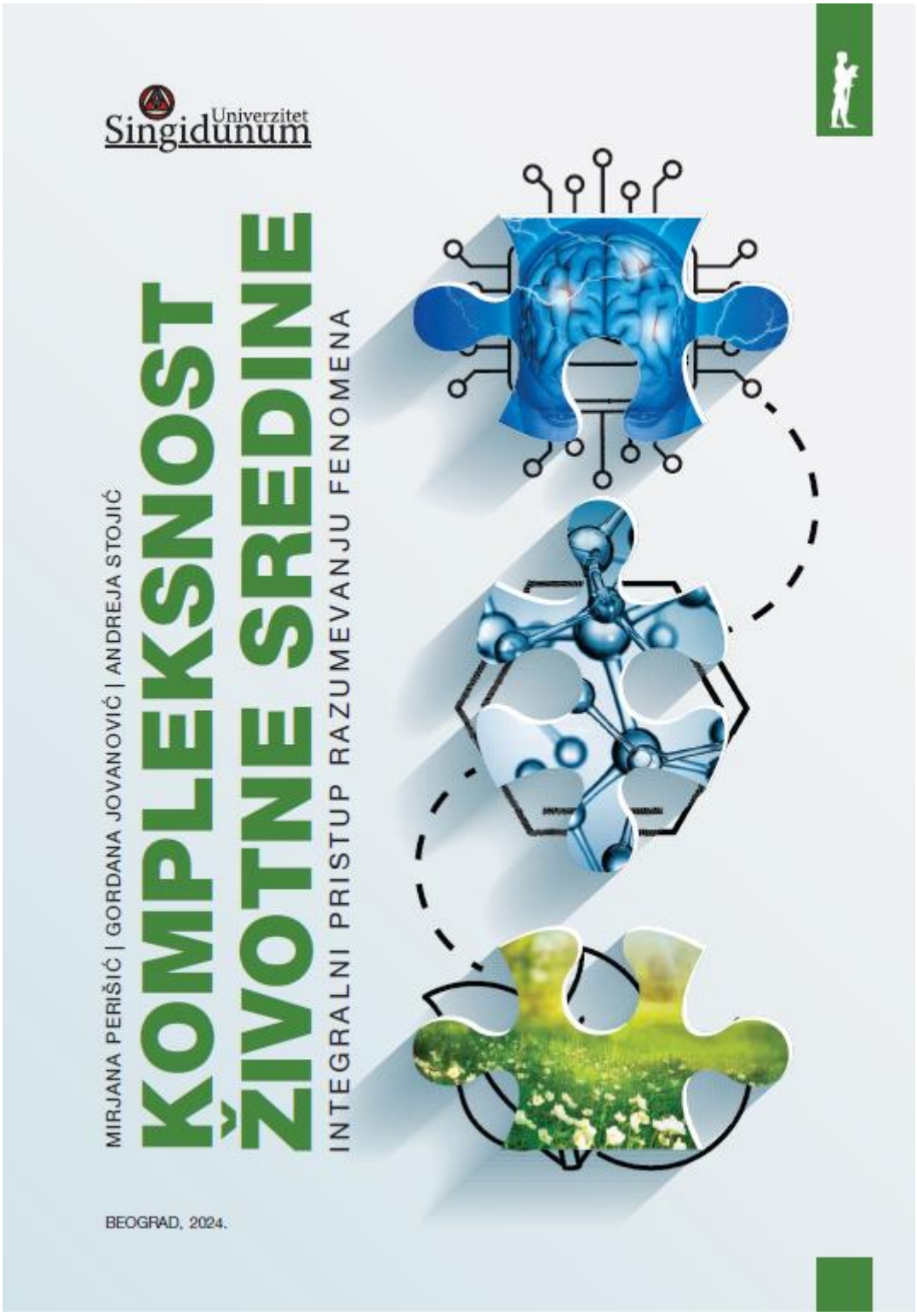
criteria air pollutants, machine learning, explainable artificial intelligence

Acknowledgement

We acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science, and Technological Development of the Republic of Serbia. This research was supported by the Science Fund of the Republic of Serbia, #GRANT No. 65241005, AI-ATLAS.

Публикације категорије М40

Публикације након избора у претходно звање



UNIVERZITET SINGIDUNUM
ŽIVOTNA SREDINA I ODRŽIVI RAZVOJ

Mirjana D. Perišić
Gordana P. Jovanović
Andreja M. Stojić

KOMPLEKSNOST ŽIVOTNE SREDINE:
integralni pristup razumevanju fenomena

Prvo izdanje

Beograd, 2024.

KOMPLEKSNOST ŽIVOTNE SREDINE: Integralni pristup razumevanju fenomena

Autori:
dr Mirjana D. Perišić
dr Gordana P. Jovanović
dr Andreja M. Stojić

Recenzenti:
Milena Rikalović, vanredni profesor, Univerzitet Singidunum
Svetlana Stanišić, redovni profesor, Univerzitet Singidunum

Izdavač:
UNIVERZITET SINGIDUNUM
Beograd, Danijelova 32
www.singidunum.ac.rs

Za izdavača:
dr Milovan Stanišić

Priprema za štampu:
Miloš Višnjić

Dizajn korica:
Aleksandar Mihajlović, MA

Godina izdanja:
2024.

Tiraž:
300

Štamparija:
Caligraph, Beograd

ISBN: 978-86-7912-826-3

Sadržaj

Prof. dr Andreja Stojić

Andreja Stojić doktorirao je na Fizičkom fakultetu Univerziteta u Beogradu u oblasti atomske i molekularne fizike i njene primene u nauci o životnoj sredini. Zaposlen je u Laboratoriji za fiziku životne sredine Instituta za fiziku u Beogradu i na studijskom programu Životna sredina i održivi razvoj Univerziteta Singidunum. Osnovne oblasti istraživanja obuhvataju nauku o podacima i nauku o životnoj sredini, uz fokus na primeni veštačke inteligencije u domenu atmosferske hemije i fizike. Učestvovao je u više nacionalnih i medunarodnih naučnih i primenjenih projekata, uključujući H2020 i projekte Fonda za nauku Republike Srbije u oblastima životne sredine i nauke o podacima. Koautor je više publikacija indeksiranih u relevantnim naučnim bazama podataka. Član grupe Neverne bebe.

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

502.131.1

ПЕРИШИЋ, Мирјана, 1979-
Kompleksnost životne sredine : integralni pristup razumevanju fenomena / Mirjana D. Perišić, Gordana P. Jovanović, Andreja M. Stojić. - 1. izd. - Beograd : Univerzitet Singidunum, 2024 (Beograd : Caligraph). - VIII, 473 str. : ilustr. ; 24 cm

Slike autora. - Tiraž 300. - Biografije autora: str. 470-472. - Bibliografija: str. 448-457.

ISBN 978-86-7912-826-3

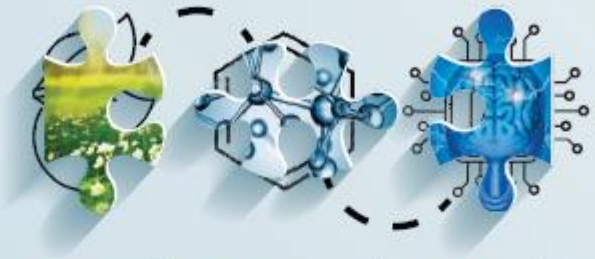
1. Јовановић, Гордана, 1988- [autor] 2. Стојић, Андреја, 1976- [autor]

а) Животна средина -- Одрживи развој

COBISS.SR-ID 160199945

www.singidunum.ac.rs

9 788679 128263



MIRJANA PERIŠIĆ | GORDANA JOVANOVIĆ | ANDREJA STOJIĆ

KOMPLEKSNOST ŽIVOTNE SREDINE

INTEGRALNI PRISTUP RAZUMEVANJU FENOMENA

Rezultat modernog načina života i odluka koje donosimo na ličnom i civilizacijskom planu nepovratno narušava stanje životne sredine, uzrokujući prevremenu smrt više od 9 miliona ljudi godišnje. To je kao da svake godine sa lica Zemlje nestane država veća od Srbije. Svakog sata više od hiljadu ljudi, svake četiri sekunde jedan čovek. Prema poslednjem izveštaju Svetske zdravstvene organizacije, 99% stanovništva udiše vazduh lošeg kvaliteta. Degradirano je blizu 30% zemljišta što utiče na kvalitet života svakog trećeg čoveka na planeti. Problem sa pijaćom vodom ima svaki četvrti čovek.

Očigledno je da savremeno društvo većinski ne razume ozbiljnost i kompleksnost problema, ne doživljava da ga se tiču, niti shvata šta je neophodno učiniti kako bi se trenutno stanje unapredilo. Zamislimo situaciju u kojoj lečenje nije u rukama lekara? Suprotno potrebi, ispostavlja se da na celokupno stanje životne sredine dominantno utiču kapital i donosioci odluka. Novac. Stručnost, sposobnost i angažovanje. Zakoni, regulative, primene i kontrola. Organizacija, interakcija, komunikacija i saradnja. Ponašanje, etika, odgovornost i svest. Savest.

Ova knjiga skicira jedan od koncepata dolaska do znanja utemeljenog u podacima u oblasti životne sredine: kako podatke pretvoriti u činjenice, činjenice u znanje, znanje u uvid i na kraju uvid u mudrost koja će pomoći u razumevanju sveta u kome živimo, ali i donošenju boljih odluka, odluka u skladu sa civilizacijskim trenutkom u kom se nalazimo.

Univerzitet Singidunum

Учешће на пројектима након избора у претходно звање

Прилог 3

Списак чланова тима Пројекта, запослених у НИО Учесницама пројекта
(Руководилац пројекта и чланови пројектног тима)

Члан пројектног тима својим потписом потврђује да је упознат са садржином овог Уговора, као и правима, обавезама и одговорностима члана пројектног тима у реализацији Пројекта, и да је сагласан да се његови/њени лични подаци обрађују у сврху извршавања и надзора овог Уговора, као и праћења и анализе програма.

Руководилац пројекта:	НИО – послодавац:	Потпис Руководиоца:
Проф. др Милан Туба	US	

	Члан пројектног тима:	НИО – послодавац:	Потпис члана пројектног тима:
1.	Небојша Бачанин-Џакула	US	
2.	Ева Туба	US	
3.	Марија Врањеш Милосављевић	IPB	
4.	Ненад Врањеш	IPB	
5.	Андреја Стојић	IPB	
6.	Гордана Јовановић	IPB	
7.	Мирјана Перишић	IPB	
8.	Немања Станишић	US	
9.	Светлана Станишић	US	
10.	Ендре Пап	US	
11.	Зора Коњовић	US	
12.	Ђорђе Обрадовић	US	
13.	Тимеа Бездан	US	
14.	Иван Радосављевић	US	
15.	Младен Видовић	US	
16.	Вељко Максимовић	IPB	

23

S3.5. Integrative framework including:

S3.5.1. *Interoperability architecture design and implementation*. (1) GeoDCAT-based applicative profile for data sets interoperability design and implementation (***TM15***, *TM 14*), and (2) ONNX-based applicative profile for pre-trained models' interoperability design and implementation (***TM12***, *TM11, TM13, TM14, TM15*)

S3.5.2. *Software architecture design and implementation*. (1) software architecture design relying upon Spark engine and JADE persistency framework (***TM12***, *TM 11, TM14, TM15*), and (2) Public cloud implementation of the designed software architecture (*TM14, TM15*)

S3.5.3. Security architecture design and implementation. (1) Security architecture design based on WASP, CSA Cloud Controls Matrix, and Spark. (***TM12***, *TM 11, TM14, TM15*), and (2) Implementation of the designed architecture (***TM14***, *TM15*).

S3.6. **Preparation of scientific publications**. Four journal and several conference papers submitted for publication. (*All WP3 members*).

Deliverables of the work package

D3.1 **Data preparation layer technical report**. *Month – 6*

D3.2 **Machine learning layer technical report**. *Month – 15*

D3.3 **Results explanation layer technical report**. *Month – 18*

D3.4 **Results presentation layer technical report**. *Month – 27*

D3.5. **Integrative framework technical report**. *Month – 20*

D3.6 **Published scientific journal and conference papers**. *Month – 18, 24, 30, 36*

Work package number	4	Work package title	Pandemic and war crisis case studies
Responsible SRO	IPB		
WP Coordinator – team member's ID	TM5		
Team member ID	PI, TM1, TM2, TM3, TM4, TM5, TM6, TM7, TM8, TM9, TM13, TM14, TM16		

Objectives

The central objective of WP4 is to assess and characterize the resulting AQ changes caused by crisis-related factors within two case studies, each containing a (1) pilot study covering RS and (2) global study covering Europe. Achieving the objective will provide an advanced insight into environmental conditions which shape AP in RS and Europe based on synergetic, multi-disciplinary analyses of the obtained crAIRsis modeling results.

Description of work and role of the team members (the task leader is presented in bold)

S4.1.1 **COVID-19 pandemic pilot study**. (1) Explore the data available for RS for the period from the first COVID positive to the beginning of the war in Ukraine (***TM6***, *TM7, TM9, TM14, TM16*), (2) apply the crAIRsis modeling to three relevant periods: pre-covid (to the first COVID positive case), "covid shock" (from the first infected, during the intensive application of restrictive measures – approximately two months), and covid (from the first loosening of the measures) (***TM1***, *TM2, TM3, TM4, TM5, TM9, TM13*), (3) analyze and compare modeling results to assess the direct and indirect impacts of pandemic-related factors – governmental measures, lockdowns, human activity, and economic changes. (***TM5***, *PI, TM1, TM6, TM7, TM8, TM9, TM16*)

S4.1.2 **COVID-19 pandemic global study**. Based on the COVID-19 pandemic pilot study experience and knowledge: (1) explore the data available for Europe for the period from the first COVID positive to the beginning of the war in Ukraine (***TM7***, *TM6, TM9, TM16*), (2) apply the crAIRsis modeling to pre-covid, "covid shock", and covid periods (***TM1***, *TM2, TM3, TM4, TM5, TM9, TM13*), (3) analyze and compare modeling results to assess region-specific, policy-specific, and society-specific pandemic-related factors' direct and indirect impacts. (***TM6***, *TM5, TM7, TM8, TM9, TM16*)

S4.1.3 **War crisis pilot study**. (1) Explore the data available for RS from the beginning of the war in Ukraine (***TM7***, *TM6, TM9, TM14, TM16*), (2) apply the crAIRsis modeling (***TM2***, *TM1, TM3, TM4, TM9, TM5, TM13*), (3) analyze and compare modeling results to assess the direct and indirect impacts of war crisis-related factors – governmental measures and human activity, energy, and economic changes. (***TM5***, *PI, TM1, TM6, TM7, TM8, TM9, TM16*)

S4.1.4 **War crisis global study**. Based on the War crisis pilot study experience and knowledge: (1) explore the data available for Europe from the beginning of the war in Ukraine (***TM6***, *TM7, TM9, TM16*), (2) apply the crAIRsis modeling (***TM2***, *TM1, TM3, TM4, TM9, TM5, TM13*), (3) analyze and compare modeling results to assess region-specific, policy-specific, and society-specific the direct and indirect impacts of war crisis-related factors. (***TM7***, *TM5, TM6, TM8, TM9, TM16*)

S4.2.1 **Preparation of findings for dissemination to the general public, policymakers, and other non-scientific stakeholders**. (*All WP4 members*)

S4.2.2. **Preparation of scientific publications**. Four journal and several conference papers related to the activities within S4.1.1-S4.1.4 submitted for publication. (*All WP4 members*).

Deliverables of the work package

D4.1 **Technical report**. Materials for dissemination to the non-scientific community and relevant stakeholders. *Month – 12, 18, 24, 30, 36*

D4.2 **Published scientific journal and conference papers**. *Month – 15, 18, 21, 24,27, 30, 33, 36*

3. Implementation Plan

3.1. Credentials of PI and members of Project team

Table 3.1. Members of the Project team

ID	Name and family name	SRO	Person-months	Effective person-months
PI	Milan Tuba	US	36	8.1
TM1	Nebojša Bačanin Džakula	US	36	7.2
TM2	Eva Tuba	US	36	5.4
TM3	Marija Vranješ Milosavljević	IPB	36	7.2
TM4	Nenad Vranješ	IPB	30	5.4
TM5	Andreja Stojić	IPB	36	7.2
TM6	Gordana Jovanović	IPB	36	5.4
TM7	Mirjana Perišić	IPB	36	5.4
TM8	Svetlana Stanišić	US	30	3.6
TM9	Nemanja Stanišić	US	18	3.6
TM10	Endre Pap	US	25	3.6
TM11	Zora Konjović	US	36	3.6
TM12	Đorđe Obradović	US	30	3.6
TM13	Timea Bezdan	US	36	4.8
TM14	Ivan Radosavljević	US	33	3.3
TM15	Mladen Vidović	US	33	3.3
TM16	Filip Alimpić	IPB	36	4
			Total Person-months: 559	Total Effective person-months: 84.7

The **crAIRsis Principal Investigator** is **Milan Tuba**. His research interest includes AI, ML, non-deterministic optimization algorithms, image processing, and networks. During the last 30 years, he was an investigator in projects founded by the Ministry of Education, Science and Technological Development of Yugoslavia and Serbia. In both **career and single-year impact lists**, **Stanford University** listed Prof. Tuba in the **World's Top 2%** of Scientists in 2020 and 2021. He is the author or co-author of more than **250 scientific papers** and editor, coeditor, or member of the editorial board or scientific committee of many scientific journals and conferences.

The crAIRsis team members' scientific expertise is presented in Table 3.1.1.

Table 3.1.1 The team members' scientific crAIRsis-related expertise.

DISCIPLINE	PI	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM8	TM9	TM10	TM11	TM12	TM13	TM14	TM15	TM16	EC
Metaheuristics optimization	1	1	1											1				1
Machine learning	1	1	1	1	1	1				1		1	1	1				1
Explainable artificial intelligence						1	1	1	1	1	1							
Environmental science						1	1	1	1	1							1	
Software engineering												1	1	1	1	1		
Data fusion		1		1	1										1	1	1	1

Briefly:

- **TM1, TM2,** and **EC** (external collaborator – international researcher) with extensive experience in AI, ML, and metaheuristics optimization. **TM1** and **TM2** are members of prestigious **Stanford University** list among **2% best world researchers** for 2020 and 2021. According to the Alper-Doger Scientific Index **TM1** is ranked as the **3rd best researcher** in Serbia in **computer science**.
- **TM3** and **TM4**, with extensive experience in AI and nuclear physics participating in many national and international scientific projects dedicated to high energy physics (**CERN**), are the authors of almost **1400 publications each (h-index 108 and 113, citations 58,533 and 66,214 – SCOPUS**; **h-index 223 and 179, citations 226,814 and 187,084 – Google Scholar**), respectively. Their activities are related to O1-3, WP2-3;
- **TM5**, **TM6**, **TM7**, **TM8**, and **TM9** collaborated to explore AP and environmental and health issues using advanced statistics, ML, and XAI and participated in several national and international (**EU FP6 and H2020**) projects. Their activities are related to O1-3, WP1-4;
- **TM10**, one of the **pioneers of the theory of non-additive measures and the corresponding integrals**, is the author of more than **400 scientific papers** (**h-index 34, citations 3,422 – SCOPUS and h-index 49, citations 15,626 – Google Scholar**). His activities are related to O3, WP3;
- **TM11** and **TM12** have extensive experience in the **development of software products** and several **international enterprise software products**, as well as the **management of complex spatial infrastructural projects**. Their activities are related to O1-4, WP2-3;

Наслов Предлога пројекта: Теоријске основе вештачке интелигенције за напредно моделирање просторно-временских података и процеса (Artificial Intelligence Theoretical Foundations for Advanced Spatio-Temporal Modelling of Data and Processes)
Акроним пројекта: **ATLAS**
Акредитоване научноистраживачке организације (НИО) у којој је запослен Руководилац предложеног Пројекта, односно чланови пројектног тима током реализације Пројекта:
Универзитет Сингидунум, Данијелова 32, Београд, и
Институт за физику у Београду, Прегревица 118, Београд.

УНИВЕРЗИТЕТ СИНГИДУНУМ
Бр. 712
29.01.20
БЕОГРАД, Данијелова 32

ФОНДУ ЗА НАУКУ РЕПУБЛИКЕ СРБИЈЕ

ЗАЈЕДНИЧКА ИЗЈАВА СВИХ УЧЕСНИКА

Овом изјавом потврђујемо да смо упознати и сагласни са свим елементима и садржајем Предлога пројекта **„Теоријске основе вештачке интелигенције за напредно моделирање просторно-временских података и процеса** (Artificial Intelligence Theoretical Foundations for Advanced Spatio-Temporal Modelling of Data and Processes - **ATLAS**)", да смо сагласни са условима Програма за развој пројеката из области вештачке интелигенције (Програм), као и да прихватамо сарадњу са доле потписаним члановима пројектног тима у случају да Предлог пројекта буде одобрен за финансирање.

Изјављујемо да смо сагласни да Предлог пројекта буде пријављен за финансирање кроз Програм.

Овлашћена лица свих НИО које учествују у Предлогу пројекта

Име и презиме	НИО (пун назив, скраћени назив/акроним)	Потпис са печатом
Милован Станишић, председник	Универзитет Сингидунум (US)	
Александар Богојевић, директор	Институт за физику у Београду (IPB)	

Руководилац и чланови тима Предлога пројекта

Име и презиме	Звање	НИО	Потпис
Професор Ендре Пап, руководилац пројекта	Професор емеритус	US	
Зора Коњовић, члан пројектног тима П1	Редовни професор	US	
Димитрије Малетић, члан пројектног тима П2	Виши научни сарадник	IPB	
Душан Вудраговић, члан пројектног тима П3	Истраживач сарадник	IPB	
Андреја Стојић, члан пројектног тима П4	Научни сарадник	IPB	

0801-10212
29.01.2020

ФОНД ЗА НАУКУ РЕПУБЛИКЕ СРБИЈЕ Београд III

УНИВЕРЗИТЕТ У БЕОГРАДУ
ИНСТИТУТ ЗА ФИЗИКУ | БЕОГРАД
ИНСТИТУТ ОД НАЦИОНАЛНОГ ЗНАЧАЈА ЗА РЕПУБЛИКУ СРБИЈУ
Прегревица 118, 11080 Земун - Београд, Република Србија
Телефон: +381 11 3713000, Факс: +381 11 3162190, www.ipb.ac.rs
ПИБ: 100105980, Матични број: 07018029, Текући рачун: 205-66984-23

Фонду за науку Републике Србије

[redacted]

Место и датум:

Београд, 30.11.2020.

др Андреја Стојић

Координатор активности на пројекту АТЛАС
у Институту за физику у Београду

Рецензије радова у међународним часописима

Review History Report

Andreja Stojić

From: 15 September 2020 To: 8 February 2025

All dates in GMT

Total journals reviewed for:	**5**
Total reviews completed:	**10**

Atmospheric Pollution Research	2
Environmental Pollution	2
Environmental Research	1
Future Generation Computer Systems	1
Science of the Total Environment	4

1

REVIEW CONFIRMATION CERTIFICATE

We are pleased to confirm that

Andreja M. StojiĆ

has reviewed 21 papers for the following MDPI journals in the period 2020–2023:

Tropical Medicine and Infectious Disease, Applied Sciences, Sustainability, Atmosphere, Electronics, Remote Sensing, Computers, Systems, International Journal of Environmental Research and Public Health

S. Tochev

Stefan Tochev, Chief Executive Officer
8 February 2025

MDPI

MDPI is a publisher of open access, international, academic journals. We rely on active researchers, highly qualified in their field to provide review reports and support the editorial process. The criteria for selection of reviewers include: holding a doctoral degree or having an equivalent amount of research experience; a national or international reputation in the relevant field; and having made a significant contribution to the field, evidenced by peer-reviewed publications.

Активност у научним и научно-стручним друштвима

atmosphere

an Open Access Journal by MDPI

Indexed in: Ei Compendex | CITESCORE 3.7 | IMPACT FACTOR 3.110

CERTIFICATE OF SERVICE

AS

Topical Advisory Panel Member of *Atmosphere*

Dr. Andreja StojiĆ

Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade, Serbia

MDPI Academic Open Access Publishing since 1996

Basel, December 2022

Dr. Shu-Kun Lin
Publisher & President


atmosphere
an Open Access Journal by MDPI
Indexed in: Ei Compendex
CITESCORE 3.7
IMPACT FACTOR 3.110
CERTIFICATE OF SERVICE
AS
Guest Editor of Special Issue
"Is Our Future up in the Air? Odorous Volatile Organic Compounds (VOCs) and Greenhouse Gas Emissions"
Dr. Andreja StojiĆ
Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade, Serbia
MDPI
Academic Open Access Publishing
since 1996
Basel, December 2022
Dr. Shu-Kun Lin
Publisher & President

Singidunum University

Sinteza

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY AND DATA RELATED RESEARCH

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology and Data Related Research - Sinteza 2021 has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

Belgrade
Jun 25, 2021.
sinteza.singidunum.ac.rs

Sinteza

Singidunum University | International Scientific Conference

SCIENTIFIC COMMITTEE

ORGANIZING COMMITTEE

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY AND DATA RELATED RESEARCH

Publisher: Singidunum University, 32 Danijelova Street, Belgrade
Editor-in-Chief: Milovan Stanišić, PhD
Prepress: Miloš Višnjić, Jovana Maričić
Design: Aleksandar Mihajlović
Year: 2021
Circulation: 80
Printed by: Caligraph, Belgrade
ISBN: 978-86-7912-755-6

Contact us:
Singidunum University
32 Danijelova Street, 11010 Belgrade, Serbia
Phone No. +381 11 3093220, +381 11 3093290,
Fax. +381 11 3093294
E-mail: sinteza@singidunum.ac.rs
Web: sinteza.singidunum.ac.rs

Singidunum University

Sinteza

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY AND DATA RELATED RESEARCH

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology and Data Related Research - Sinteza 2022 has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

Belgrade
April 16, 2022.
sinteza.singidunum.ac.rs

Sinteza

Singidunum University | Formations | Scientific Conference

SCIENTIFIC COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Endre Pap, Singidunum University, Serbia
- Aleksandar Jevremović, Singidunum University, Serbia
- Andreja Stojić, Institute of Physics in Belgrade, Serbia
- Bratislav Milovanović, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Đorđe Obradović, Singidunum University, Serbia
- Goran Šimić, Military Academy, University of Defence, Serbia
- Gordana Dobrijević, Singidunum University, Serbia
- Gordana Jovanović, Institute of Physics in Belgrade, Serbia
- Ivan Ćuk, Singidunum University, Serbia
- Jelena Filipović, Faculty of Philology, University of Belgrade, Serbia
- Miloš Antonijević, Singidunum University, Serbia
- Marina Marjanović Jakovljević, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Marko Šarac, Singidunum University, Serbia
- Milan Milosavljević, Singidunum University, Serbia
- Milan Tair, Singidunum University, Serbia
- Milan Tuba, Singidunum University, Serbia
- Miloš Stojmenović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Mirjana Perišić, Institute of Physics in Belgrade, Serbia
- Miroslav Popović, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mladan Jovanović, Singidunum University, Serbia
- Nebojša Bačanin Džakula, Singidunum University, Serbia
- Nemanja Stanišić, Singidunum University, Serbia
- Nina Dragićević, Singidunum University, Serbia
- Petar Spalević, Faculty of Technical Sciences in Kosovska Mitrovica, University of Priština, Serbia
- Predrag Popović, Vinča Institute, Serbia
- Radosav Pušić, Faculty of Philology, University of Belgrade, Serbia
- Sanja Filipović, Singidunum University, Serbia
- Saša Adamović, Singidunum University, Serbia
- Tijana Radojević, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Alexandru Nedelea, Stefan cel Mare University of Suceava, Romania
- Aurora Pedro Bueno, Department of Applied Economics, University of Valencia, Spain
- Chen Yudong, Communication University of China, Beijing, People's Republic of China
- Deasún Ó Conchúir, Scatterwork GmbH, Ireland
- Diego Andina De la Fuente, Technical University of Madrid, Spain
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Egons Lavendelis, Riga Technical University, Latvia
- Gordana Pesaković, Argosy University, USA
- Hong Qi, Dalian University of Technology, China
- Ivan Bajić, Simon Fraser University, Canada
- Ina Bikuviene, PhD Kauno kolegija- University of Applied Sciences, Kaunas (Lithuania)
- Jesus Amador Valdés Díaz de Villegas, Iberoamericana University, Mexico
- Jovica V. Milanović, University of Manchester, United Kingdom
- Juan Ruiz Ramirez, University of Veracruz, Mexico
- Li Liwen, Beijing Foreign Studies University, Beijing, PR China
- Lorenzo Fagiano, Polytechnico di Milano, Italy
- Luis de la Torre Cubillo, UNED, Dpt. Informática y Automática, Madrid, Spain
- Luis Hernández Gómez, Technical University of Madrid, Spain
- Maarten De Vos, University of Oxford, United Kingdom
- María Magdalena Hernández Alarcón, University of Veracruz, Mexico
- Mike Dawney, Middlesex University, United Kingdom
- Moe Win, Massachusetts Institute of Technology, USA
- Mohammed Ismail Elnaggar, The Ohio State University, USA
- Nataša Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nellie Swart, University of South Africa, Pretoria
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Nuno Manuel Garcia dos Santos, University Beira Interior, Portugal
- Riste Temjanovski, Goce Delčev University, Macedonia
- Roberta Grossi, Horizons University, France
- Simona Distinto, Department of Life and Environmental Sciences University of Cagliari, Italia
- Slobodan Luković, ALaRI, Switzerland
- Snezana Lawrence, Bath Spa University, United Kingdom
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Vassilis S. Moustakis, Technical University of Crete, Greece
- Violeta Grubliene, Klaipeda University, Lithuania
- Vladimir Terzija, University of Manchester, United Kingdom
- Yipeng Liu, University of Electronic Science and Technology of China, China

ORGANIZING COMMITTEE

- Milovan Stanišić
- Nebojša Bačanin Džakula
- Marko Tanasković
- Mladen Veinović
- Valentina Gavranović
- Miroslav Popović
- Nina Dragićević
- Miodrag Živković
- Ivan Ćuk
- Miloš Antonijević
- Mladan Jovanović
- Srđan Marković
- Milan Tair
- Aleksandar Mihajlović
- Petar Jakić
- Uroš Arnaut
- Miloš Mravik
- Jelena Gavrilović
- Predrag Obradović
- Jovana Maričić
- Miloš Višnjić

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY AND DATA RELATED RESEARCH

Publisher: Singidunum University, 32 Danijelova Street, Belgrade
Editor-in-Chief: Milovan Stanišić, PhD
Prepress: Miloš Višnjić, Jovana Maričić
Design: Aleksandar Mihajlović
Year: 2022
Circulation: 75
Printed by: Caligraph, Belgrade
ISBN: 978-86-7912-800-3

Contact us:
Singidunum University
32 Danijelova Street, 11010 Belgrade, Serbia
Phone No. +381 11 3093220, +381 11 3093290,
Fax. +381 11 3093294
E-mail: sinteza@singidunum.ac.rs
Web: sinteza.singidunum.ac.rs

Singidunum University

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Sinteza

Belgrade, 2023

sinteza.singidunum.ac.rs

Singidunum University

Sinteza

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology and Data Related Research - Sinteza 2023 has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia.

Belgrade
May 27, 2023.
sinteza.singidunum.ac.rs

Sinteza

Singidunum University | International Scientific Conference

SCIENTIFIC COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Endre Pap, Singidunum University, Serbia
- Aleksandar Jevremović, Singidunum University, Serbia
- Andreja Stojić, Institute of Physics in Belgrade, Serbia
- Bratislav Milovanović, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Đorđe Obradović, Singidunum University, Serbia
- Goran Šimić, Military Academy, University of Defence, Serbia
- Gordana Dobrijević, Singidunum University, Serbia
- Gordana Jovanović, Institute of Physics in Belgrade, Serbia
- Jelena Filipović, Faculty of Philology, University of Belgrade, Serbia
- Miloš Antonijević, Singidunum University, Serbia
- Marina Marjanović Jakovljević, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Marko Šarac, Singidunum University, Serbia
- Milan Milosavljević, Singidunum University, Serbia
- Milan Tuba, Singidunum University, Serbia
- Miloš Stojmenović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Mirjana Perišić, Institute of Physics in Belgrade, Serbia
- Miroslav Popović, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mladan Jovanović, Singidunum University, Serbia
- Nebojša Bačanin Džakula, Singidunum University, Serbia
- Nemanja Stanišić, Singidunum University, Serbia
- Nina Dragićević, Singidunum University, Serbia
- Petar Spalević, Faculty of Technical Sciences in Kosovska Mitrovica, University of Priština, Serbia
- Predrag Popović, Vinča Institute, Serbia
- Radosav Pušić, Faculty of Philology, University of Belgrade, Serbia
- Sanja Filipović, Singidunum University, Serbia
- Saša Adamović, Singidunum University, Serbia
- Tijana Radojević, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Alexandru Nedelea, Stefan cel Mare University of Suceava, Romania
- Aurora Pedro Bueno, Department of Applied Economics, University of Valencia, Spain
- Chen Yudong, Communication University of China, Beijing, People's Republic of China
- Deasún Ó Conchúir, Scatterwork GmbH, Ireland
- Diego Andina De la Fuente, Technical University of Madrid, Spain
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Egons Lavendelis, Riga Technical University, Latvia
- Gordana Pesaković, Argosy University, USA
- Hong Qi, Dalian University of Technology, China
- Ivan Bajić, Simon Fraser University, Canada
- Ina Bikuviene, PhD Kauno kolegija- University of Applied Sciences, Kaunas (Lithuania)
- Jesus Amador Valdés Díaz de Villegas, Iberoamericana University, Mexico
- Jovica V. Milanović, University of Manchester, United Kingdom
- Juan Ruiz Ramirez, University of Veracruz, Mexico
- Li Liwen, Beijing Foreign Studies University, Beijing, PR China
- Lorenzo Fagiano, Polytechnico di Milano, Italy
- Luis de la Torre Cubillo, UNED, Dpt. Informática y Automática, Madrid, Spain
- Luis Hernández Gómez, Technical University of Madrid, Spain
- Maarten De Vos, University of Oxford, United Kingdom
- María Magdalena Hernández Alarcón, University of Veracruz, Mexico
- Mike Daweny, Middlesex University, United Kingdom
- Moe Win, Massachusetts Institute of Technology, USA
- Mohammed Ismail Elnaggar, The Ohio State University, USA
- Natala Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nellie Swart, University of South Africa, Pretoria
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Nuno Manuel Garcia dos Santos, University Beira Interior, Portugal
- Riste Temjanovski, Goce Delčev University, Macedonia
- Roberta Grossi, Horizons University, France
- Simona Distinto, Department of Life and Environmental Sciences University of Cagliari, Italia
- Slobodan Luković, ALaRI, Switzerland
- Snezana Lawrence, Bath Spa University, United Kingdom
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Vassilis S. Moustakis, Technical University of Crete, Greece
- Violeta Grubliene, Klaipeda University, Lithuania
- Vladimir Terzija, University of Manchester, United Kingdom
- Yipeng Liu, University of Electronic Science and Technology of China, China

REVIEWERS COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mladan Jovanović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Milan Tair, Singidunum University, Serbia
- Jelena Gajić, Singidunum University, Serbia
- Ivana Brdar, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Srđan Marković, Singidunum University, Serbia
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Gordana Pelaković, Argosy University, USA
- Natala Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Slobodan Luković, ALaRI, Switzerland
- Vassilis S. Moustakis,Technical University of Crete, Greece
- Jovica V. Milanović, University of Manchester, United Kingdom

ORGANIZING COMMITTEE

- Milovan Stanišić
- Mladen Veinović
- Milan Tair
- Nebojša Bačanin Džakula
- Marko Tanasković
- Zora Konjović
- Valentina Gavranović
- Miroslav Popović
- Nina Dragićević
- Miodrag Živković
- Jelena Gavrilović
- Miloš Antonijević
- Mladan Jovanović
- Srđan Marković
- Marina Marjanović
- Aleksandar Mihajlović
- Nikola Savanović
- Miloš Mravik
- Predrag Obradović
- Marijana Mihajlović
- Miloš Višnjić
- Aleksa Vidaković
- Romario Stanković

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publisher: Singidunum University, 32 Danijelova Street, Belgrade
Editor-in-Chief: Milovan Stanišić, PhD
Prepress: Miloš Višnjić, Marijana Mihajlović
Design: Aleksandar Mihajlović
Year: 2023
Circulation: 85
Printed by: Caligraph, Belgrade ISBN: 978-86-7912-802-7

Contact us:
Singidunum University
32 Danijelova Street, 11010 Belgrade, Serbia
Phone No. +381 11 3093220, +381 11 3093290,
Fax. +381 11 3093294
E-mail: sinteza@singidunum.ac.rs
Web: sinteza.singidunum.ac.rs

Singidunum University

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Sinteza

Belgrade, 2024

sinteza.singidunum.ac.rs

Singidunum University

Sinteza

SINGIDUNUM UNIVERSITY INTERNATIONAL SCIENTIFIC CONFERENCE

BOOK OF PROCEEDINGS

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publishing of Conference Proceedings of the International Scientific Conference on Information Technology, Computer Science, and Data Science- Sinteza 2024 has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia.

Belgrade
May 16, 2024.
sinteza.singidunum.ac.rs

Sinteza

SCIENTIFIC COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Endre Pap, Singidunum University, Serbia
- Aleksandar Jevremović, Singidunum University, Serbia
- Andreja Stojić, Institute of Physics in Belgrade, Serbia
- Bratislav Milovanović, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Đorđe Obradović, Singidunum University, Serbia
- Goran Šimić, Military Academy, University of Defence, Serbia
- Gordana Dobrijević, Singidunum University, Serbia
- Gordana Jovanović, Institute of Physics in Belgrade, Serbia
- Jelena Filipović, Faculty of Philology, University of Belgrade, Serbia
- Miloš Antonijević, Singidunum University, Serbia
- Marina Marjanović Jakovljević, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Marko Šarac, Singidunum University, Serbia
- Milan Tair, Singidunum University, Serbia
- Milan Tuba, Singidunum University, Serbia
- Miljan Vučetić, Vlatacom Institute, Serbia
- Miloš Stojmenović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Mirjana Perišić, Institute of Physics in Belgrade, Serbia
- Miroslav Popović, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mladan Jovanović, Singidunum University, Serbia
- Nebojša Bačanin Džakula, Singidunum University, Serbia
- Nemanja Stanišić, Singidunum University, Serbia
- Nina Dragićević, Singidunum University, Serbia
- Petar Spalević, Faculty of Technical Sciences in Kosovska Mitrovica, University of Priština, Serbia
- Predrag Popović, Vinča Institute, Serbia
- Sanja Filipović, Singidunum University, Serbia
- Saša Adamović, Singidunum University, Serbia
- Tijana Radojević, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Alexandru Nedelea, Stefan cel Mare University of Suceava, Romania
- Aurora Pedro Bueno, Department of Applied Economics, University of Valencia, Spain
- Deasún Ó Conchúir, Scatterwork GmbH, Ireland
- Diego Andina De la Fuente, Technical University of Madrid, Spain
- Dragana Vilić, Ekonomski fakultet, Univerzitet u Banja Luci
- Dušan Ličina, École polytechnique fédérale de Lausanne (EPFL), Switzerland
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Egons Lavendelis, Riga Technical University, Latvia
- Gordana Pesaković, Argosy University, USA
- Hong Qi, Dalian University of Technology, China
- Ivan Bajić, Simon Fraser University, Canada
- Ina Bikuršene, PhD Kauno kolegija- University of Applied Sciences, Kaunas (Lithuania)
- Jovica V. Milanović, University of Manchester, United Kingdom
- Li Liwen, Beijing Foreign Studies University, Beijing, PR China
- Lorenzo Fagiano, Polytechnico di Milano, Italy
- Luis de la Torre Cubillo, UNED, Dpt. Informática y Automática, Madrid, Spain
- Luis Hernández Gómez, Technical University of Madrid, Spain
- Maarten De Vos, University of Oxford, United Kingdom
- Martin Kampel, Institute of Visual Computing & Human-Centered Technology, TUW, Austria
- Mike Dawney, Middlesex University, United Kingdom
- Moe Win, Massachusetts Institute of Technology, USA
- Mohammed Ismail Elnaggar, The Ohio State University, USA
- Nataša Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nellie Swart, University of South Africa, Pretoria
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Nuno Manuel Garcia dos Santos, University Beira Interior, Portugal
- Özge Ercan, Faculty of Sport Sciences – Sinop University-Türkiye
- Riste Temjanovski, Goce Delčev University, Macedonia
- Roberta Grossi, Horizons University, France
- Simona Distinto, Department of Life and Environmental Sciences University of Cagliari, Italia
- Slobodan Luković, ALaRI, Switzerland
- Snezana Lawrence, Bath Spa University, United Kingdom
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Vassilis S. Moustakis, Technical University of Crete, Greece
- Violeta Grublienė, Klaipeda University, Lithuania
- Vladimir Terzija, University of Manchester, United Kingdom

REVIEWERS COMMITTEE

- Milovan Stanišić, Singidunum University, Serbia
- Mladen Veinović, Singidunum University, Serbia
- Mladan Jovanović, Singidunum University, Serbia
- Miodrag Živković, Singidunum University, Serbia
- Dragan Cvetković, Singidunum University, Serbia
- Marko Tanasković, Singidunum University, Serbia
- Zora Konjović, Singidunum University, Serbia
- Milan Tair, Singidunum University, Serbia
- Jelena Gajić, Singidunum University, Serbia
- Ivana Brdar, Singidunum University, Serbia
- Valentina Gavranović, Singidunum University, Serbia
- Srdan Marković, Singidunum University, Serbia
- Duško Lukač, Rheinische Fachhochschule Köln – University of Applied Sciences, Germany
- Gordana Pekaković, Argosy University, USA
- Nataša Vilić, Filozofski fakultet, Univerzitet u Banja Luci, BiH
- Nuno Gonçalo Coelho Costa Pombo, University Beira Interior, Portugal
- Stanimir Sadinov, Technical University of Gabrovo, Bulgaria
- Slobodan Luković, ALaRI, Switzerland
- Vassilis S. Moustakis,Technical University of Crete, Greece
- Jovica V. Milanović, University of Manchester, United Kingdom

ORGANIZING COMMITTEE

- Mladen Veinović
- Milan Tair
- Nebojša Bačanin Džakula
- Marko Tanasković
- Zora Konjović
- Valentina Gavranović
- Miroslav Popović
- Nina Dragićević
- Miodrag Živković
- Jelena Gavrilović
- Miloš Antonijević
- Mladan Jovanović
- Srdan Marković
- Marina Marjanović
- Lazar Dražeta
- Aleksandar Mihajlović
- Nikola Savanović
- Miloš Mravik
- Predrag Obradović
- Marijana Mihajlović
- Miloš Višnjić
- Aleksa Vidaković
- Romario Stanković

INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

Publisher: Singidunum University, 32 Danijelova Street, Belgrade
Editor-in-Chief: Milovan Stanišić, PhD
Prepress: Miloš Višnjić, Marijana Mihajlović
Design: Aleksandar Mihajlović
Year: 2024
Circulation: 400
Printed by: Caligraph, Belgrade ISBN: 978-86-7912-821-8

Contact us:
Singidunum University
32 Danijelova Street, 11010 Belgrade, Serbia
Phone No. +381 11 3093220, +381 11 3093290,
Fax. +381 11 3093294
E-mail: sinteza@singidunum.ac.rs
Web: sinteza.singidunum.ac.rs

„CNN Tech
International Conference of Experimental and Numerical Investigations and New Technologies "

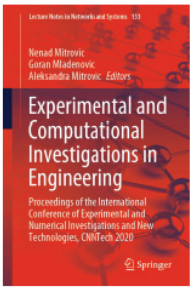
CNN TECH

Main Menu

- Home
- Program
- About Zlatibor
- Topic
- Abstract or full paper submission
- Important Dates
- Documents
- Committees
- Direct flight
- Accomodation
- Registration
- Regional Innovation Forum
- Contact Us

Important Updates

Abstract or full paper submission: until **15.03.2021.**



Committees

Scientific Committee

Miloš Milošević, Serbia (chairman)
Nenad Mitrović, Serbia (co-chairman)
Aleksandar Sedmak, Serbia
Hloch Sergej, Slovakia
Dražan Kozak, Croatia
Nenad Gubeljak Slovenia
Monka Peter, Slovakia
Snežana Kirin, Serbia
Samardžić Ivan, Croatia
Martina Balać, Serbia
Mládková Ludmila, Czech Republic
Johanyák Zsolt Csaba, Hungary
Igor Svetel, Serbia
Aleksandra Mitrović, Serbia
Valentin Birdeanu, Romania
Danilo Nikolić, Montenegro
Goran Mladenović, Serbia
Bajić Darko, Montenegro
Tasko Manski, Srbija
Luis Reis, Portugal
Žarko Mišković, Serbia
Tozan Hakan, Turkey
Nikola Momčilović, Serbia
Traussnigg Udo, Austria
Gordana Bakić, Serbia
Katarina Čolić, Serbia
Peter Horňak, Slovakia
Róbert Huňady, Slovakia
Martin Hagara, Slovakia
Jovan Tanaskovic, Serbia
Aleksa Milovanovic, Serbia
Marija Đurković, Serbia
Tsanka Dikova, Bulgaria
Ján Danko, Slovakia
Ognjen Peković, Serbia
Jelena Svorcan, Serbia
Suzana Filipović, Serbia
Darko Kosanović, Serbia

Organizing Committee

Nenad Mitrović (chairman)
Miloš Milošević (co-chairman)
Aleksandar Sedmak
Martina Balać
Vesna Miletić
Igor Svetel
Goran Mladenović
Aleksandra Mitrović
Aleksandra Dragicević
Žarko Mišković
Katarina Čolić
Milan Travica
Dragana Perovic
Aleksandra Joksimovic
Beti Kostadinovska Dimitrovska
Tsanka Dikova
Isaak Trajković
Toni Ivanov
Snežana Kirin
Igor Stanković
Ivana Vasović Maksimović
Nina Obradović
Andreja Stojić

Call for
Docume
Full pa
Conse
Abstr
Book of
Book of
Book of
Book of
Gallery
CNN TE
Support

UNESCO - етика вештачке интелигенције - консултације

From Saša Lazović Date 2020-07-14 09:22

Поштоване колеге,

Предложени сте за учешће у консултацијама у оквиру израде препоруке о етици вештачке интелигенције. Уколико сте сагласни са номинацијом, молим да нам данас доставите попуњену табелу из прилога.

У прилогу се такође налази позив *UNESCO-a* са свим потребним информацијама. Консултације ће се одржати од 27. до 29. јула.

Молим да потврдите пријем ове поруке. Хвала.

Срдачан поздрав,
Саша

др Саша Лазовић,
Помоћник министра за технолошки развој, трансфер технологија и иновациони систем

Министарство просвете, науке и технолошког развоја
Немањина 22-26, 11000 Београд
Тел: + 381 11 3640 232
e-mail: sasa.lazovic@mpn.gov.rs

Ангажованост у формирању научних кадрова након избора у претходно звање

Универзитет
Сингидунум

УНИВЕРЗИТЕТ СИНГИДУНУМ
Бр. 150
11.12.2025 год.
БЕОГРАД, Данијелова 32

ПОТВРДА

Овим се потврђује да је **Андреја (Милорад) Стојић**, рођен 03.01.1976. године у Јагодини, у звању научни сарадник, радно ангажован уговором о допунском раду на Универзитету Сингидунум, где је и изабран, у звању ванредног професора.

У току рада на Универзитету Сингидунум, на мастер академским студијама, на студијском програму Животна средина и одрживи развој проф. др Андреја Стојић је био ментор студенту Филипу Алимпићу број индекса: 2020/740037, који је одбранио завршни мастер рад на тему: „ПРОМЕНЕ У КОНЦЕНТРАЦИЈАМА ИСПАРЉИВИХ ОРГАНСКИХ ЈЕДИЊЕЊА ПОРЕКЛОМ ИЗ ИНДУСТРИЈЕ ПОД УТИЦАЈЕМ ПАНДЕМИЈЕ КОВИД-19 У БЕОГРАДУ“ дана 18.10.2021. године.

Такође, на докторским академским студијама, на студијском програму Животна средина и одрживи развој, проф. др Андреја Стејић је Одлуком Департмана за последипломске студије Универзитета Сингидунум број: 4-122-1/2021 од 25.05.2021. године, именован за ментора докторандкињи Наташи Радић (раније Букумирић) број индекса: 2019/485095 приликом израде докторске дисертације: „Процена утицаја измењених активности људи током пандемије COVID-19 на расподеле испарљивих једињења у Београду“.

Потврда се издаје на захтев именованог ради избора у научно звање и у друге сврхе се не може користити.

РЕКТОР
Проф. др Горанка Кнежевић

Ул. Данијелова 32, 11000 Београд, Србија
Тел. +381 11 30 93 220, 30 93 209; 30 93 229; Факс. +381 11 30 93 294
office@singidunum.ac.rs
www.singidunum.ac.rs

Универзитет
Сингидунум

На основу члана 73., 74. и 75. Закона о високом образовању ("Сл. гласник РС", бр. 88/2017 73/2018 27/2018 – др. Закон, 67/2019, 6/2020 – др.закон, 11/2021, 67/2021 и 67/2021 – др. закон, 76/2023), члана 41. став 1. тачка 21. и чл. 92. и 93. Статута, а у складу са Правилником о условима и поступку избора наставника и сарадника на УНИВЕРЗИТЕТУ СИНГИДУНУМ, (у даљем тексту Универзитет), Сенат Универзитета, на седници одржаној 26.9.2024. године, донео је

УНИВЕРЗИТЕТ СИНГИДУНУМ
Бр. 1150
26. 09. 2024
БЕОГРАД, Данијелова 32

О Д Л У К У
О ИЗБОРУ НАСТАВНИКА УНИВЕРЗИТЕТА

1. **Др Андреја Стојић** бира се у звање ванредног професора за ужу научну област „Науке о заштити животне средине".

2. Закључити Уговор о раду, у складу са општим актима Универзитета.

3. Одлуку доставити: именованом, финансијској служби, правној служби за досије наставника и архиви Универзитета.

О б р а з л о ж е њ е

Универзитет Сингидунум расписао је Конкурс на интернет страници Универзитета дана 5.8.2024. године за избор у звање ванредног професора за ужу научну област „Науке о заштити животне средине".

Комисија за припрему Извештаја у саставу проф. др Светлана Станишић, редовни професор Универзитет Сингидунум, председник Комисије и проф. др Небојша Бачанин Џакула, редовни професор Универзитета Сингидунум и др Ненад Вранеш, научни саветник Института за физику, чланови Комисије, предложила је након проучене документације да се именовани изабере у звање ванредног професора за ужу научну област „Науке о заштити животне средине".

Сенат Универзитета размотрио је поднету документацију, Извештај о професионалној, стручној и научној делатности кандидата и оценио је да резултати досадашњег рада у свему квалификују именованог кандидата за избор у звање ванредног професора.

На основу Извештаја и предлога Комисије, утврђено је да др Андреја Стојић у свему испуњава услове прописане Законом и Општим актима Универзитета за избор у звање ванредног професора, за ужу научну област „Науке о заштити животне средине", па је донета Одлука као у изреци.

ПРЕДСЕДНИК СЕНАТА
РЕКТОР
Проф. др Горанка Кнежевић

Ул. Данијелова 32, 11000 Београд, Србија
Тел. +381 11 30 93 220, 30 93 209; 30 93 229; Факс. +381 11 30 93 294
office@singidunum.ac.rs
www.singidunum.ac.rs