Публикације категорије М10

Studies in Computational Intelligence

Volume 973

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

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

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

Preface

Artificial Intelligence (AI) has become an important part of our everyday life, and it will continue to grow in importance in the future. Therefore, managing further developments in AI theory, and its applications in complex systems, is a matter of urgency. The most important research is in the domains of the human brain and powerful computers endowed with strong software. These are studied both separately and together. As Alan Turing said, "What we want is a machine that can learn from experience." Machine Learning (ML) is a branch of artificial intelligence, and is the study of computer algorithms that allow computer programs to automatically improve through experience.

The present book is an up-to-date collection of 18 chapters, written by internationally recognized experts in the fields of AI and environmental research. It is related to the project ATLAS, https://ai.ipb.ac.rs/, and completed in collaboration with leading scientists in the fields of AI and environmental research.

The volume centers on two main subjects, and is therefore organized into two parts, theory and applications, although all chapters contain both approaches.

Part I of this book is devoted to theoretical issues which are mostly mathematically based.

The chapter "Mathematical Foundation of Artificial Intelligence" is devoted to some of the mathematical tools used in AI, which are largely related to decision-making. Among the main tools are general aggregation functions (operators) with some important special cases, such as triangular norms and copulas. A significant extension of classical mathematical analysis is pseudo-analysis, which is illustrated with a short overview of some of its important applications. General fuzzy measures and the corresponding integrals, including a number of many different generalizations of the Choquet integral, are also presented.

The chapter "Collection and Decomposition Integrals in Multicriteria-Decision Support" focuses on recently introduced collection and decomposition integrals, and covers the Choquet, the Shilkret and the PAN integrals, among others. It discusses some properties and contains several examples of these integrals, in particular ones extending the Lebesgue integral. Possible applications in multicriteria-decision support and domains containing imprecision and uncertainty are shown.

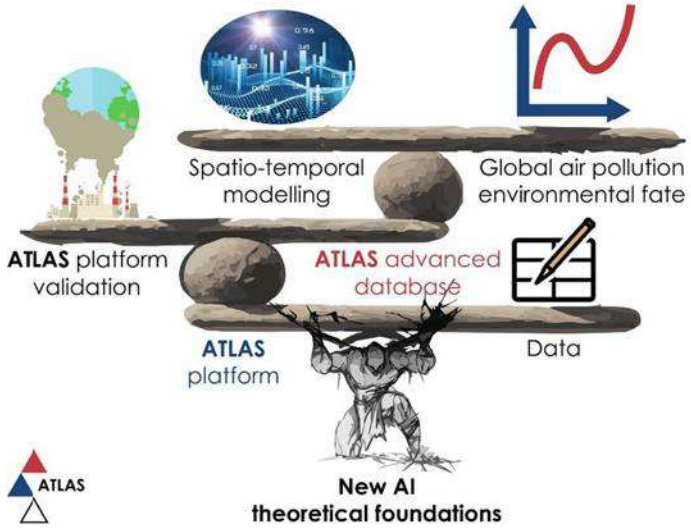


Fig. 1 ATLAS project and completed in collaboration with leading scientists

The chapter "A Refinement of the Jensen Type Inequality for the Pseudo-integral" presents new results related to the Jensen-type inequality for the pseudo-integral, using the Fubini type result for pseudo-integral. The fact that the pseudo-product of two pseudo measures reduces to repeated pseudo-integrals, enables obtaining a refinement of the Jensen-type inequality.

The chapter "Convolutional Neural Networks Hyperparameters Tuning" discusses image classification with convolution neural networks. In this early stage of CNN development, the common method of tuning CNN is by guessing and estimating, also known as the guestimating method. Since this is a hard optimization problem, there is a chance for optimization metaheuristic to be applied. One of the promising approaches is the application of swarm intelligence algorithms.

In the chapter "The Case for Quantifying Artificial General Intelligence with Entropy Semifields" it is supposed that Information Semifields are the underlying calculi that brains operate on, and it is postulated that strong artificial intelligence should try to imitate them. These semifields emerge from pseudo-calculus by the rational choices of generator functions that respect the classical postulates of information theory, and align with the Renyi entropy and divergence. The chapter provides relevant evidence from he fields of Neuroscience, Machine Learning, Cognitive Computation, Information Theory and Complex Systems Theory.

The chapter "Fuzzy Metrics and Its Applications in Image Processing" is devoted to fuzzy metrics and analyzes their application in image processing. Two applications of fuzzy metrics related to image processing are illustrated: Image filtering and Copy-move forgery detection in images. The aim was to improve the sharpness and the quality of the image, as measured by the image quality indices UIQI and CPBD. Fuzzy metric parameters that give the best image quality and sharpness are

determined experimentally. Copy-Move Forgery Detection (CMFD) is one of the methods used to detect forgeries.

The chapter "Aggregation Operators and Distributivity Equations" is devoted to the issue of distributivity of aggregation operators, which is crucial in many different areas such as decision-making theory and integration theory. The characterization of all pairs *(F; G)* of special aggregation operators that satisfy the distributivity law are given.

The chapter "The Use of Fuzzy Logic in Various Combinatorial Optimization Problems" presents how fuzzy logic can be used for modeling uncertainties in a number of combinatorial problems and improve their quality. The focus is on the Location Set Covering Problem (LSCP), the Maximal Covering Location Problem (MCLP) and the Minimal Covering Location Problem (MinCLP) as a special modification of MCLP. These problems are applicable in searching optimal places for desired and undesired facilities under given conditions. Firstly, small dimensional instances of problems will be solved with an exact algorithm and using the CPLEX optimizer tool. When the dimensions of the problem become too large for exact solutions, the instances will then be solved with the Particle Swarm Optimization (PSO) metaheuristic.

The chapter "An Improved BAT Algorithm for Solving Job Scheduling Problems in Hotels and Restaurants" is an improvement to the original Bat Algorithm (BA) to speed up the convergence has been made. Each solution consists of many parameters. To conduct a comprehensive comparative analysis between the original BA, the modified BA proposed in this chapter and other state-of-the-art bio143inspired metaheuristics, the performance of both approaches is evaluated on a standard set of 23 (unimodal, multimodal and fixed-dimension multimodal) benchmark functions. Subsequently, the modified BA was applied to solve real-world practical job scheduling problems in hotels and restaurants.

Part II of the book presents a number of important applications of AI.

In the chapter "Patterns of PCB-138 Bioaccumulation in Small Pelagic Fish from the Eastern Mediterranean Sea Using Explainable Machine Learning Prediction" is 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 nonlinear relationships between PCB-138 and the other examined variables that were explained by SHAP values. The ten fuzzy clusters of SHAP values that were obtained 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 doesn't result in higher bio-accumulation of PCB-138.

In the chapter "Patterns of PCB-138 Occurrence in the Breast Milk of Primiparae and Multiparae Using SHapley Additive exPlanations Analysis", a novel SHapley Additive exPlanations (SHAP) method was used to examine the key parameters that govern the distribution of PCB-138 in breast milk. According to the results, PCB-156, PCB-180, HCB, HCH and PCB-118 have a major impact, while PCB-28, PCB-52 and PCB-189 have a minor impact on PCB-138 distribution in breast milk.

Similar contaminant behaviors, which belong to both the indicator congener group (−28, −52, −180) and the toxicologically relevant PCBs (−118, −189), were also noted.

In the chapter "What Information on Volatile Organic Compounds Can Be Obtained from the Data of a Single Measurement Site Through the Use of Artificial Intelligence?", an innovative integrated methodology for the spatio-temporal characterization and concentration forecasts of various pollutants was used. Toxic, mutagenic and carcinogenic representatives of volatile organic species—benzene, toluene, ethylbenzene and xylene, commonly referred to as BTEX—were studied. Increasing air pollutant concentrations over the last few decades have been the focus of contemporary scientific research due to their adverse effects on public health, the environment and climate change. The methodology is based on receptor-oriented air circulation modeling and artificial intelligence, implemented through machine learning and explainable artificial intelligence methods. 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. XGBoost was efficient at forecasting BTEX levels, with low estimated errors (6–15%) significantly below the uncertainty obtained by conventional models for the evaluation of average annual pollutant concentrations.

The chapter "The Linear Fuzzy Space: Theory and Applications" covers the main theoretical results of the linear fuzzy space. It also analyzes its capabilities to model spatial and spatio-temporal systems using representative examples of real-world problems. It contains a brief, yet comprehensive, presentation of linear fuzzy space theory. The third section presents an analysis of fuzzy linear space theory concerning the modeling of spatial and temporal systems using two examples which present its applications in modeling application topology and time series.

The chapter "Image Fuzzy Segmentation Using Aggregated Distance Functions and Pixel Descriptors" is a review of recent research on image fuzzy segmentation using a fuzzy c-means clustering algorithm based on the distance function constructed by applying an aggregation function on the sequence of the initial distance functions and pixel descriptors. In image segmentation algorithms, distance functions compare pixels and represent a decision criterion for classifying the pixels into image segments. The segmentation criterion is determined based on the information fusion process, where the application of the appropriated aggregation function enables the adjustment of the segmentation criteria according to the intuitively expected decision.

The chapter "A Generative Model for the Creation of Large Synthetic Image Datasets Used for Distance Estimation" presents a generative model for the creation of large synthetic image datasets. The model is implemented as a 3D scene, representing an urban environment in Blender. The images are rendered using the Cycles rendering engine which allows for the creation of high fidelity data. The data was used for end-to-end training of a convolutional neural network, designed to estimate distances from stereoscopic images. The evaluation of the neural networks' performance showed that the generative model presented in this chapter is a viable tool for the generation of large image datasets for training predictive models.

The chapter "Appraisal of Apartments in Belgrade Using Hedonic Regression: Model Specification, Predictive Performance, Suitability for Mass Appraisal, and Comparison with Machine Learning Methods" presents a Bayesian hedonic regression model for the appraisal of apartments located within the metropolitan area of Belgrade, Serbia. Nested random effects are used to model the hierarchical structure present in the location identifiers, and thin-plate spline functions are used to capture non-linear effects. Major factors affecting prices include area in m2, floor number, the total number of floors in the building, the availability of an elevator and condition. The model achieves a Mean Average Percent Error (MAPE) of 13.06% in the validation set. Its predictive performance is compared to that of three popular Machine Learning (ML) methods, and its suitability for mass appraisal is examined.

Chapter "The Role of Chatbots in Foreign Language Learning: The Present Situation and the Future Outlook" is devoted to chatbots, computer programs developed to engage in conversations with humans. The most natural and potentially powerful application of chatbots is related to their fundamental nature—language practice. However, their role and outcomes within both formal and informal language learning is currently tangential the best. Existing research in the area has generally focused on chatbots' comprehensibility and the motivation they inspire in their users. In this chapter, an overview of chatbots for learning languages is provided, existing approaches are critically analyzed and major challenges for future research are given.

The chapter "Intelligent Interactive Technologies for Mental Health and Well-Being" presents intelligent technologies used to automate the assessment and evaluation of psychological treatments and mental well-being and functioning. The technologies include different types of robots, video games and conversational agents. The chapter critically analyzes existing solutions with future research implications.

The book is intended for researchers in AI, and researchers in environmental sciences as well as for Ph.D. students.

The editor is grateful to the authors for their excellent contributions. Many thanks go to proofreaders Profs. Jasna Petrović and Luke Deighton Chrisostomides for their effort to help the volume reach its present form. This book was supported by the Science Fund of the Republic of Serbia, #Grant No. 6524105, AI-ATLAS.

Belgrade, Serbia
February 2021

Endre Pap

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 $\times$ 0.25 mm, Rtx-5 film thickness of 0.25 μ m, and (2) 30 m $\times$ 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

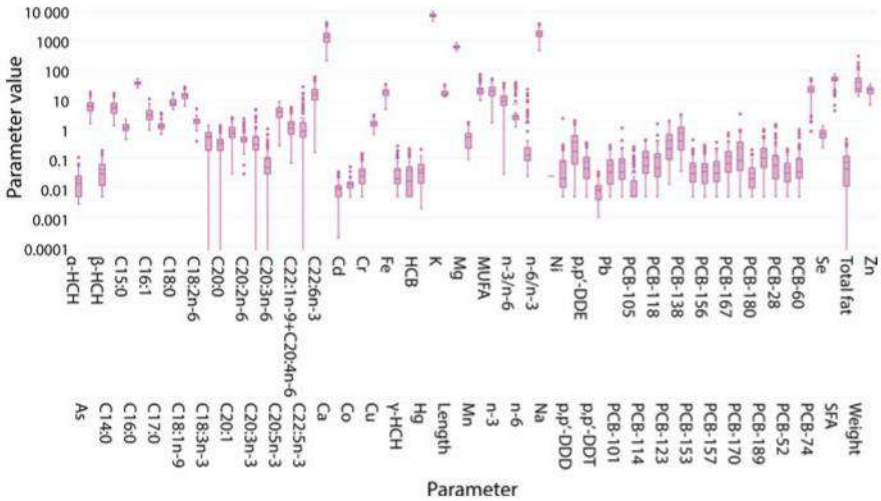


Fig. 1 Box plot

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

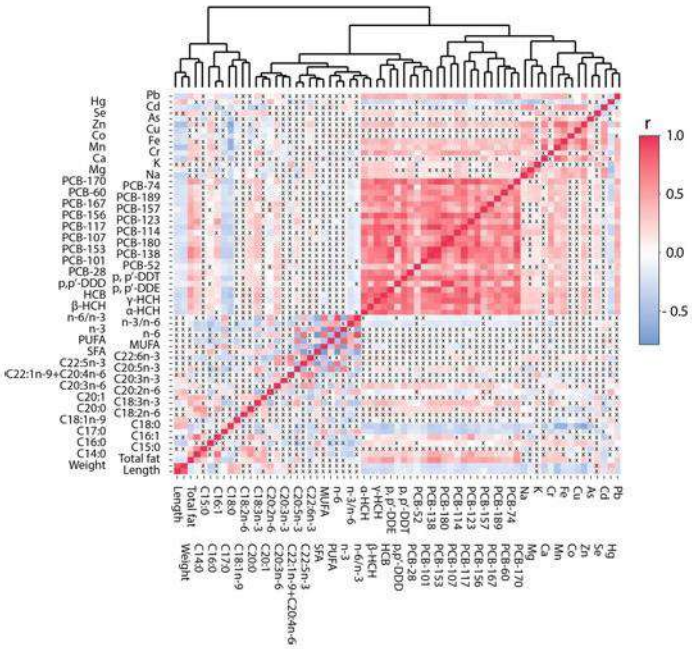


Fig. 2 Parameter value correlations matrix

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

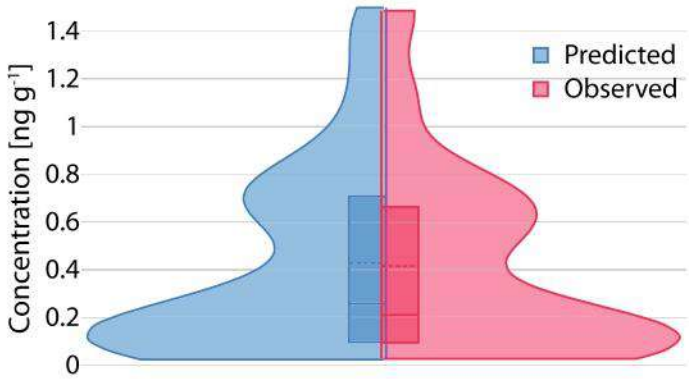


Fig. 3 Model evaluation

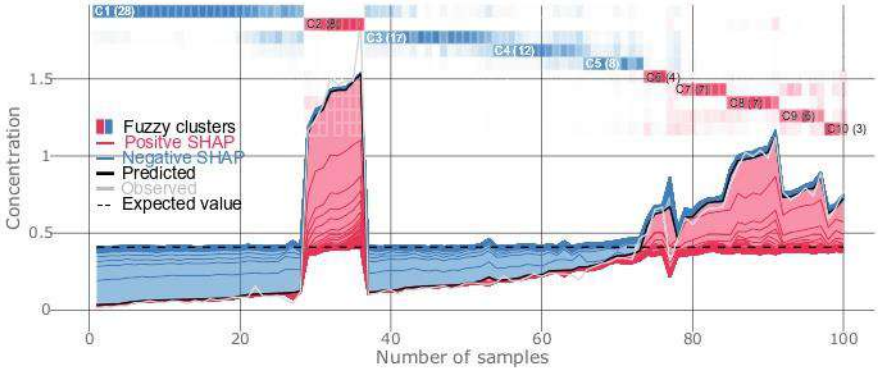


Fig. 4 SHAP value force plot

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:

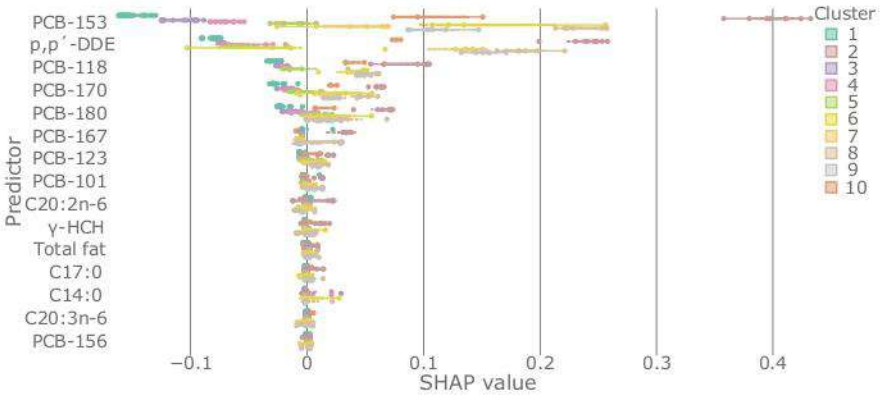


Fig. 5 SHAP value top ten clusters

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)

Patterns of PCB-138 Occurrence in the Breast Milk of Primiparae and Multiparae Using SHapley Additive exPlanations Analysis

Gordana Jovanović, Marijana Matek Sarić, Snježana Herceg Romanić, Svetlana Stanišić, Marija Mitrović Dankulov, Aleksandar Popović, and Mirjana Perišić

Abstract Breastfeeding provides numerous health benefits for newborns by meeting infants' nutritional needs and supporting associated immunological protection. Maternal milk is high in fat, and therefore, represents a very suitable medium for the bioaccumulation of lipophilic pollutants, such as organochlorine pesticides (OCPs) and polychlorinated biphenyls (PCBs). This makes breast milk the infant's primary source of postnatal exposure to persistent toxic xenobiotics. In this study, we applied a novel SHapley Additive exPlanations (SHAP) method to examine the key parameters that govern the distribution of PCB-138, an indicator of non-dioxin congeners, in the mother's milk. According to the accuracy metrics, the eXtreme Gradient Boost-

G. Jovanović (✉) · M. Perišić
Institute of Physics Belgrade, National Institute of the Republic of Serbia, Environment and Sustainable Development, Singidunum University, Belgrade, Serbia
e-mail: gordana.vukovic@ipb.ac.rs

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

M. Matek Sarić
Department of Health Studies, University of Zadar, Zadar, Croatia
e-mail: marsaric@unizd.hr

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

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

M. Mitrović Dankulov
Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia
e-mail: mitrovic@ipb.ac.rs

A. Popović
Faculty of Chemistry, University of Belgrade, Belgrade, Serbia
e-mail: apopovic@chem.bg.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_11

ing regression was employed successfully, with a predicted/observed relative error below 20% and a high correlation coefficient (r = 0.97), for modeling the relationships between PCB-138 and other non-dioxin congeners, the mother's age, and the number of births. According to the results, PCB-156, PCB-180, HCB, HCH and PCB-118 have a major impact, while PCB-28, PCB-52 and PCB-189 have a minor impact on PCB-138 distribution in breast milk. Similar contaminant behaviors, which belong to both the indicator congener group (−28, −52, −180) and the toxicologically relevant PCBs (−118, −189), were also noted. The SHAP conclusions were only partially consistent with the results of the correlation analysis suggesting that POPs exhibit non-linear dynamics and interrelations. Therefore, current knowledge on the 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

Persistent organic pollutants (POPs), including organochlorine pesticides (OCPs), polychlorinated biphenyls (PCBs), polychlorinated dibenzodioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs), have posed an environmental and health threat for over a century. Due to their common properties such as persistence, bioaccumulation, toxicity, and their ability to be transported through the air over long distances, the production and use of POPs was prohibited and/or restricted by the Stockholm Convention from 2001. Because of their persistence and lipophilic properties, they accumulate and biomagnify in the food chain, mainly in the lipid-rich tissues of humans and animals.

Research has shown that organochlorine compounds, such 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 increase mortality and the risk of cancer, cause allergies and hypersensitivity, damage the central and peripheral nervous system, cause reproductive disorders, and immune system disorders [5, 7]. Humans and wildlife are exposed to PCBs and OCPs in the polluted environment via air, water, sediment soil and food; though ingestion, inhalation, or skin contact. Previous studies suggest that more of these compounds are introduced into the body through food intake than by inhalation or through the skin. It is also known that children are prenatally exposed to organochlorine compounds that pass from the mother to the fetus through the placenta. Prenatal exposure to organochlorines has been linked to impaired fetal development, reduced cognitive and neurological progress and negative effects on the function of respiratory organs [3, 8].

Biological monitoring of POPs in the body requires invasive techniques involving the surgical extraction of adipose tissue. The alternative monitoring approach, and one of the most commonly used non-invasive methods for the assessment of human exposure, is the analysis of human milk's POPs content. Human milk, as the natural and superior food for newborns, meets infants' necessary nutritional requirements and provides numerous health benefits associated with immunological and psychological advancement. However, the high fat content of human milk makes it a suitable matrix for the accumulation of toxic lipophilic chemicals such as POPs, which is a result of the dietary habits and occupational exposure of the breastfeeding mother. Since 1987, the World Health Organization (WHO) has carried out six global surveys on POPs in human milk to evaluate in utero and lactational exposure to these compounds and identify the measures which should be taken to limit or minimize the impact of organochlorines on humans and the environment. Although the studies indicated that the presence of POPs in the environment has decreased, and consequently that their intake by humans has declined, evaluation of the risk factors and determinants of breastfed infants' exposure to POPs is still a challenge [26].

A number of literature sources have reported on POPs levels in breast milk regarding regional and global quantitative differences and time-trends and/or their dependence on parameters like the mother's age and weight loss, childbirth, dietary habits and occupation [6, 14, 20, 25]. Maternal age and parity are among the most frequently considered parameters. The concentration of organochlorines has been anticipated to increase in the fatty tissues of older mothers, whereas parity is expected to reduce their levels since lactation is recognized as the main pathway for the elimination of POPs from the body. Our previous research aimed to investigate 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 [10, 12, 13]. 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 an increasingly popular tool for predicting environmental phenomena [21, 24]. We aimed to obtain a detailed insight into the parameters, including organochlorine pollutants, the mother's age, and parity, that shape the PCB-138 distribution in the mother's milk and to offer a promising methodology for the evaluation of the observed interrelations. We chose PCB-138 since among the PCB6 group, it has been considered as an indicator PCB behavior patterns in various samples and 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 European food and feed regulations where dioxin-like PCBs resemble biochemical and toxicological characteristics of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). The present chapter is an extended version of the study presented at the International Scientific Conference on Information Technology and Data Related Research (Sinteza 2020), Singidunum University [23].

2 Materials and Methods

2.1 Sampling

Breast milk samples were collected from 150 healthy primiparae, secundiparae, and multiparae (having delivered three children), aged between 19 and 45 living in Zadar (Croatia). The sampling was performed three times between 2011 and 2016 and the second and third campaigns were part of the WHO survey on POP levels in breast milk. The samples were gathered when mature lactation was achieved, from 2 to 38 weeks postnatally. Milk was manually expressed into pre-cleaned glass bottles and stored at −20 °C until analysis.

All of the mothers were aware of the aim and relevance of the study and signed informed written consent forms to confirm their participation. The mothers completed an enclosed questionnaire and according to the results, the participants had no history of accidental or occupational exposure to the analyzed POPs. The collected samples and the personal and clinical data were only used for the purposes of the research. The study and the sampling were approved by the local Zadar County Health Centre Ethics Committee (01-745/2011 and 01-405/2014). More details about the sampling were given previously [10].

2.2 Chemical Analysis of PCBs and OCPs

The analytical procedure has been described in detail previously [13]. 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, the milk fat was dissolved in n-hexane and purified with sulfuric acid. Adsorption chromatography was additionally performed on a multilayer silica column using 4% diethyl-ether in n-hexane as a solvent. As a final step prior to the chromatography analysis, the organochlorine compounds were fractionated on carbon commercial tubes (ENVI-Carb SPE tubes, Supelco, the USA), eluted with n-hexane/toluene (99:1) mixture, reduced to dryness and dissolved in n-hexane.

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 analyzed. High-resolution gas chromatography with an electron capture detector was performed on a CLARUS 500 chromatograph using two capillary columns (Restek, Bellefonte, PA, USA) simultaneously: (1) 60 m × 0.25 mm, Rtx-5 film thickness of 0.25 μm, and (2) 30 m × 0.25 mm, Rtx-1701 film thickness of 0.25 μm. The column temperature was programmed as fol-

lows: (1) a gradient increase from 100 °C to 110 °C with incremental changes of 4 °C min^{-1}, (2) isothermally heating at 110 °C for 5 min, (3) a gradient increase to 240 °C with incremental changes of 15 °C min^{-1}, and (4) isothermally heating at 240 °C for 50 min. Nitrogen carrier gas was used with injector and detector temperatures of 250 °C and 270 °C, respectively.

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%. The reproducibility of the method expressed as a relative standard deviation was between 6% and 22%, and 7% and 24% for PCBs and OCPs respectively.

2.3 Data Analysis

The relationships between PCB-138 in breast milk and all other measured parameters were modeled using XGBoost regression. Details on the method are available elsewhere [4]. Briefly, XGBoost is a highly effective machine learning method based on a sequential tree-growing algorithm, where each decision tree tries to complement all others and correct for residuals in the predictions made by the previous trees. Due to its core advantages concerning computational efficiency, excellent predictive performance, competitive accuracy, highly optimized multicore, and tree complexity penalization that was not widely used in previous additive tree models, XGBoost has been successfully applied in various domains [9, 18, 19]. 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 a 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 the explainable artificial intelligence method SHapley Additive exPlanations. This method is capable of providing a straightforward and meaningful interpretation of machine learning model-derived decisions, which is now focused on user-readable logic rules [15]. Based on game theory and Shapley values that provide distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction), the SHAP values, as a measure of feature importance, represent the only possible locally accurate and globally consistent feature attribution values. SHAP overcomes the main drawback of attribution inconsistency and minimizes the possibility of underestimating the importance of a feature with a certain attribution value. In this study we used Python SHAP implementation (SHAP Python package). The captured attributed importance of a feature is visually presented as a SHAP summary plot.

To clearly indicate a change in the absolute SHAP value relative to other feature attributions, we obtained relative SHAP values [22]. Relative SHAP values, defined as the share of absolute SHAP in the total attributed importance of all features for the case, show the relative influence of a feature on the prediction. To test the stability of the obtained absolute and relative SHAP values, they were evaluated with a 50 times-replicated bootstrap method.

3 Results and Discussion

3.1 Pollutant Profiles

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 from the DDT group, exhibiting concentrations in the following ranges: 0.15–77.75 ng g^{-1} (average: 8.34 ng g^{-1}) for *p,p'*-DDE, 0.10–49.48 ng g^{-1} (average: 2.67 ng g^{-1}) for *p,p'*-DDD and 0.30–92.64 ng g^{-1} (average: 5.58 ng g^{-1}) for *p,p'*-DDT. The levels are comparable or even lower than the contamination reported in the investigations from Europe and the rest of the world [1, 2, 6]. The occurrence of *p,p'*-DDT in breast milk primarily depends on the mother's diet or direct occupational contact with this compound and therefore, high levels of *p,p'*-DDT may be associated with recent exposure to the xenobiotic. However, during active exposure, humans metabolically dechlorinate *p,p'*-DDT to *p,p'*-DDE or *p,p'*-DDD and increased values of *p,p*'-DDE indicate enhanced metabolic conversion. The *p,p'*-DDE compound is the most persistent metabolite with the longest half-life in the environment in comparison with its counterparts, but individual lev-

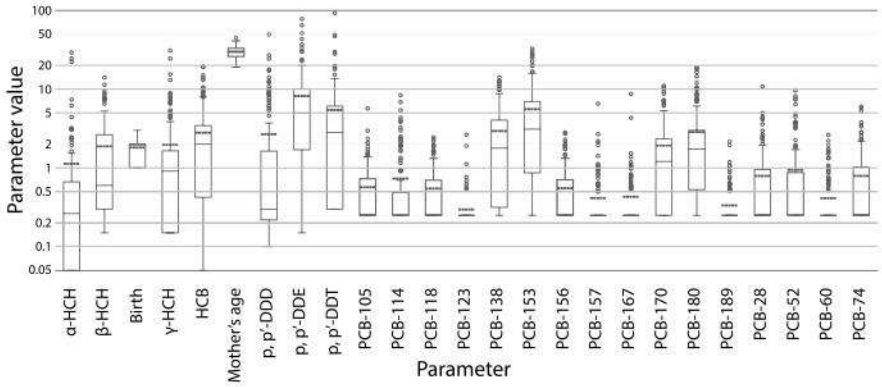


Fig. 1 Obtained parameter box-plot

els of *p,p'*-DDE in the mothers' body, which are the result of *p,p'*-DDT breakdown, are difficult to distinguish from direct ingestion of the already presented *p,p'*-DDE metabolite in the environment. To imply fresh inputs of *p,p'*-DDT in the organism as well as the rate of its transformation or excretion, some researchers suggest using the ratio *p,p'*-DDT/*p,p'*-DDE as an indicator [11]. The values of this marker were higher than 1 in approximately 30% of the samples collected in this study, indicating that one third of the mothers were recently exposed to *p,p'*-DDT via diet including home-produced free-range eggs [1] or that their metabolism is limited at reducing levels of this pollutant. In addition, since *p,p'*-DDT and *p,p'*-DDE bind to different hormone receptors, dissimilar biological effects are expected to occur in the mothers' body with different *p,p'*-DDT/*p,p'*-DDE ratios and consequently, to affect breastfed infants in another manner.

Among the other POPs, the most prevalent were by-products of lindane, γ-HCH and HCB, occurring in ranges from 0.10 to 19.31 ng g^{-1} (average: 2.86 ng g^{-1}) and from 0.31 to 30.85 ng g^{-1} (average: 2.01 ng g^{-1}), respectively. Among potential non-dietary (e.g., mothers age, nursing history, parity, residence) and dietary (e.g., fish consumption and its frequency, inclusion of milk products, intake of meat and eggs) predictors that could affect accumulation of HCH compounds and HCB in the mothers' organism, fatty fish and the reception of breastfeeding were suggested to be the most influential determinants [1].

The PCB contamination profile was dominated by PCB-153 (average: 5.76 ng g^{-1}) followed by other non-dioxin like congeners: PCB-138 (average: 3.04 ng g^{-1}), PCB-180 (average: 3.06 ng g^{-1}), and PCB-170 (average: 2.06 ng g^{-1}). This distribution represents a likely pattern of congener accumulation in the milk samples around the world reflecting the stronger resistance of heavy hexa- and hepta-chlorinated PCBs to biological degradation. However, individual concentrations are higher than those reported in the region [29], but lower than the congener levels around the world [1, 6, 16].

Significant linear correlation coefficients ($r \geq 0.90$) were found between the following pairs of 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). The results indicate that the listed pollutants have similar molecular structures and metabolic pathways, which is discussed below. However, we assumed that advanced methods, other than commonly applied correlation matrices, could be employed to deepen current understanding of PCB-138 behavior patterns in breast milk.

3.2 Interrelations of Pollutant Patterns

For investigating the non-linear relationships between PCB-138 and other congeners, the mother's age, and number of births, XGBoost regression analysis was used, with a predicted/observed relative error below 20% and a 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 behavior patterns

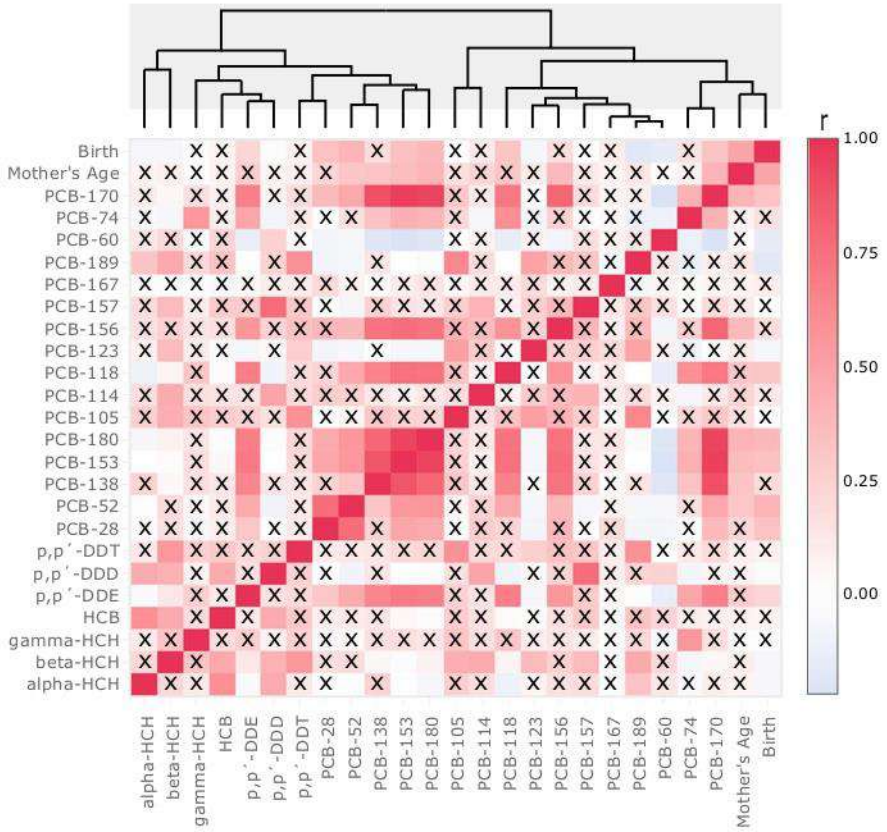


Fig. 2 Obtained parameter correlation matrix ('x'—p-value > 0.1)

in the examined milk samples were PCB-170 and PCB-153 (Fig. 4). The strongest influences were observed to be related to elevated concentrations of the listed pollutants, suggesting that "heavy" congeners substituted with chlorine atoms at 2' and 6 ortho positions are more prone to bioaccumulation in human milk compared to other PCBs. The position of the halogen substituting biphenyl ring of PCB-153 and PCB-170 provides the molecules with a structure and rigidity such that PCBs are able to pass from the blood to the breast milk. Other molecular descriptors including lipophilicity (octanol-water partitioning coefficients-logKow), molecular diameter and weight, the number of hydrogen-bond acceptors and the number of attached halogens appeared to be less determinative for PCB partitioning between blood and milk [27].

Several studies reported on the high serum/milk ratio of congeners with six or more chlorine substitutes suggesting that an increase in molar size, polarizability, and hydrophobicity cause a low rate of partition between blood serum and milk [17]. When considering compounds with the same number of attached chlorines, higher

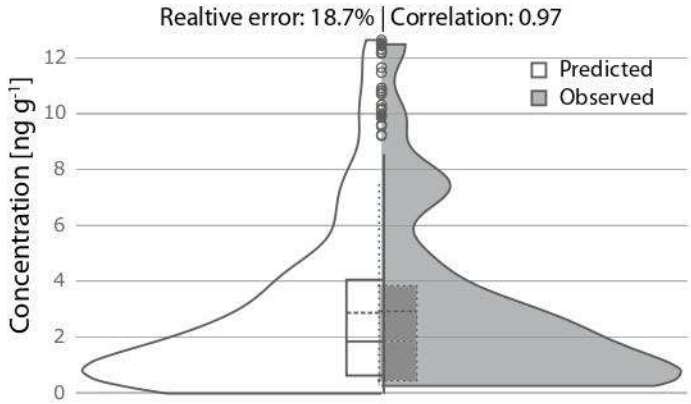


Fig. 3 XGBoost evaluation

mean serum/milk ratios were recorded for the toxicologically relevant dioxin-like PCBs compared to the indicator non-dioxin-like PCB, although a twofold difference between the lowest and highest serum/milk ratio of an individual congener was reported. The observations suggested a significant role is had by other chemical properties, but which one directly determines the serum/milk ratio is difficult to establish because of the collinearity of PCB molecules. The PCDD/Fs, with a molar volume comparable to PCBs and polybrominated diphenyl ethers, exhibited a higher serum/milk ratio that confirms a reduced transfer of molecules with more rigid structures and higher hydrophobicity from serum to milk [17]. The comparison of the PCB serum/milk ratio from different studies is limited by the different methodologies used for the ratio calculation, an inconsistent number of PCB serum-milk pairs included in the studies, and analytical uncertainties associated with the determination of congeners. Further research is needed to support the conclusion of our study and provide valid information on the complex characteristics that influence PCB partition kinetics between human serum and milk. In addition, data on the chemical characteristics alone are insufficient to predict the serum/milk partition of PCBs and descriptors such as metabolic breakdown through dechlorination or binding to different protein sites in serum and milk should be considered.

The SHAP analyses revealed less significant impacts of PCB-156, PCB-180 and PCB-118 on PCB-138 behavior patterns in milk samples. The pollutants belong to a non-dioxin-like/indicator congener group (−138, −153 and −180) and the toxicologically relevant PCBs (−118 and −156), which elicit aryl hydrocarbon receptor-mediated biochemical and toxic responses and resistance in the food chain. A minor negative dependency was observed between PCB-138 and elevated levels of low-chlorinated congeners (PCB-52 and PCB-28), which are more volatile and susceptible to metabolic breakdown and excretion, and organochlorine *p,p'*-DDT.

Numerous studies reported that POPs are eliminated from the body during breastfeeding, which results in the lower level of pollutants in the breast milk of multipara,

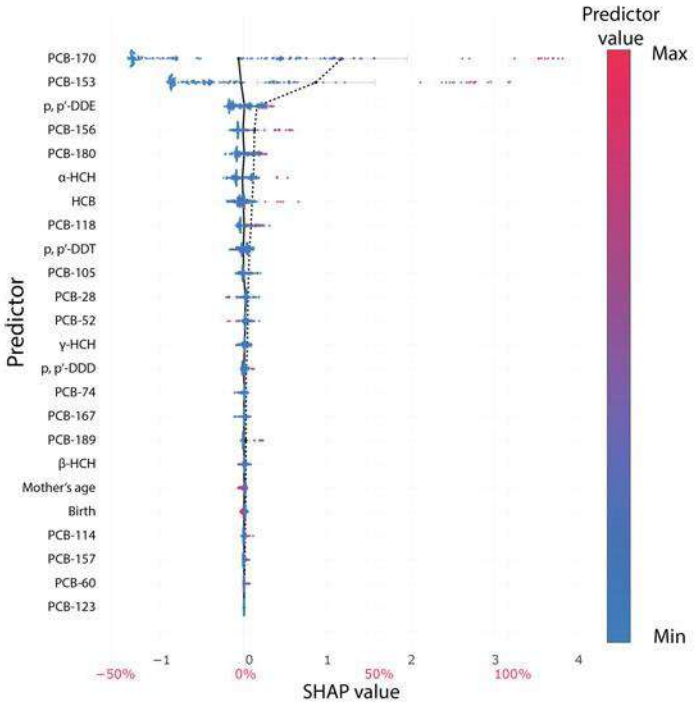


Fig. 4 PCB-138 SHAP summary plot

while the review of PCB changes during lactation emphasized non-consistent elimination rates of pollutants [16]. In addition, it is usually reported that the milk of older mothers contained higher concentrations of organochlorine xenobiotics. The positive associations may be related with the higher cumulative exposure of older mothers to POPs, and possibly a higher body mass index associated with increased fat content [1]. In the present study, the following age ranges were considered: 19–24 (N = 22), 25–30 (N = 53), 31–36 (N = 52) and 37–45 (N = 13; three mothers were above 40 years old). The pollutant levels were largely unaltered by the age of first-time mothers whereas their content increased in the breast milk of older multiparae. A slight decline in OCPs, the indicator and toxicologically relevant congeners, was shown in the milk of 29- and 30-year-old secundiparae, but the contaminant maximums were achieved in the samples of 35-year-old mothers. For the mothers with third child delivery, similar patterns were confirmed only for the toxicologically relevant congeners and PCB-74. However, as indicated by low SHAP values approaching zero, the mother's age and parity had no significant influences on the PCB-138 patterns (Fig. 4). Although the dataset is limited, it is comparable to the number of samples in other studies where a positive association between age, parity and POP levels was reported. For predicting the bioaccumulation of POPs in breast milk, infor-

mation on exposure routes including dietary habits, residence, working environment, and the mother's childhood nursing is required.

Ten clusters, representing groups of variable interrelations associated with certain PCB-138 levels, were identified through fuzzy clustering of the SHAP values (Fig. 5). Two of the clusters (C7 and C9) with the highest PCB-138 concentrations (up to 12 ng g^{-1}) were the most differentiated ones. Congeners PCB-153 and PCB-170 explain the higher PCB-138 concentrations in breast milk and their impact, which is characterized by relative SHAP values, and accounts for 69% and 75% in total of C7 and C9 respectively (relative error < 7.7%). Positive interrelations between PCB-138, and PCB-153 and PCB-170 were indicated by the SHAP (PCB-153: 0.4-2.8; PCB-170: 0.5–3.5 ng g^{-1}) and relative SHAP values (PCB-153: 19.8–44.7%; PCB-170: 23.8–44.7%). The higher and extreme values of PCB-138 are related to the higher and extreme values of PCB-153 (3.6–8.2 ng g^{-1}) and PCB-170 (7.4–22.9 ng g^{-1}), which are mostly above upper fence values. The impact of the congeners −170 and −153 increase as their concentration increases as well.

Fuzzy clustering identified three clusters (C1–C3) representing groups of variable interrelations associated with medium PCB-138 levels (relative error < 12%). Medium PCB-138 levels were associated with PCB-153 and PCB-170 concentrations above their mean values, 5.6 and 1.9 ng g^{-1} respectively. Apart from C2 and C3, the negative impact of PCB-153 on the PCB-138 pattern is evident in C1. Since the concentration of both congeners is low compared to other clusters, the outcome could be the result of differing intakes of PCB contaminants through differences in the mothers' diets. The C1 cluster is also characterized by the higher concentrations and influence of *p,p'*-DDT, α-HCH, and HCB as well as by the lower levels and impact *p,p'*-DDE. In some participants, fresh inputs of pollutants from the OCP group are evident, but some of the mothers could possess a limited capacity to metabolically transform and eliminate organochlorine xenobiotics as well. Overall, the result indicates a similar origin of OCPs and PCBs in the maternal body and/or similar metabolic pathways as discussed in detail in previous paragraphs.

The overall implication is that compounds such as DDT and lindane, which were previously thought to have exclusively agricultural uses, may have also been used in cities for insect control and these cities may be significant sources of the compounds in the atmosphere. As indicated by the relative SHAP values given in brackets, the occurrence of high (up to 7.7 ng g^{-1}) and medium (up to 5.3 ng $^{-1}$) PCB-138 concentrations are explained by the following pollutants: *p,p'*-DDE (C1: 1.9%, C2: 4.4%, C3: 7.5%, C7: 3.2%, and C9: 2.8%), PCB-118 (C2: 2.5%, C3: 5.3%, and C7: 1.9%), PCB-156 (C9: 4.8%), PCB-180 (C1: 3.8%, C2: 4.8%, C3: 3.9%, C7: 3%, and C9: 2.6%) and α-HCH (C1: 3.4%). In the clusters, the listed compounds exhibited the following concentrations (ng g^{-1}): *p,p'*-DDE (C1: 7.5, C2: 16.3, C3: 12.0, C7: 14.7, and C9: 30.4, which ranged from mean to above upper fence values), PCB-105 (C1: 0.7, 75th percentile value), PCB-118 (C2: 0.8, C3: 0.9, and C7: 1.3, which ranged from 75th percentile to upper fence values), PCB-156 (C1: 0.5, and C9: 1.8 as referred to above mean and upper fence values), PCB-180 (C1: 2.6, C2: 5.1, C3: 2.9, C7: 7.5 and C9: 12.8, above mean and upper fence values). The listed clusters are characterized by excessive concentrations of POPs which imply similar

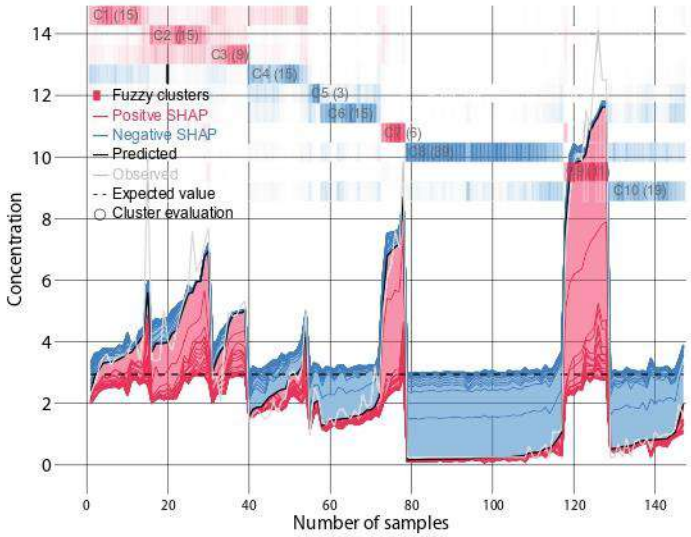


Fig. 5 PCB-138 Force plot

exposure to the pollutants because of fresh inputs or metabolic transformation and degradation.

Low PCB-138 concentrations (<4 ng g^{-1}) are distributed in five clusters (C4–C6, C8 and C10) (Fig. 5) and the same constituents as discussed above determine the clusters (relative error < 20%). The low concentrations of PCB-138 are related with the occurrence (ng g^{-1}) of: PCB-170 (C4: 1.7, C5: 1.2, C6: 1.0, C8: 0.3, and C10: 0.45), PCB-153 (C4: 3.6, C5: 3.1, C6: 3.0, C8: 0.4, and C10: 1.7), PCB-118 (C5: 0.3), PCB-180 (C5: 2.1 and C8: 0.4), PCB-156 (C5: 0.3 and C8: 0.3), *p,p'*-DDE (C5: 6.0, C6: 5.5, and C8: 1.2), α-HCH (C4: 2.0 and C6: 0.2), PCB-52 (C4: 0.56), and *p,p'*-DDD (C5: 4.8). The strength of the effects is described by the following relative SHAP values (%): PCB-170 (C4: 9.5, C5: −33, C6: −42, C8: −44, and C10: −44), PCB-153 (C4: −25, C5: −27, C6: −20, C8: −29, and C10: −25), PCB-118 (C5: −3.1), PCB-180 (C5: −3.3 and C8: −3.9), PCB-156 (C5: −4.1 and C8: −2.5), *p,p'*-DDE (C5: 2.0, C6: 1.3, and C8: −5.1), α-HCH (C4: 1.6 and C6: 1.8), PCB-52 (C4: 1.6), and *p,p'*-DDD (C5: 2.1). The results imply the congener uptake from different sources related to variations in the mothers' occupational exposure and dietary habits. The low positive SHAP relative values of *p,p'*-DDD and *p,p'*-DDE suggests intensified metabolic degradation of the POPs. Out of the extracted groups, the clusters characterized by excessive concentrations of POPs could be considered the most influential for the observed high levels of PCB-138 as indicated by the absolute (Fig. 6) and relative SHAP values (Fig. 7), which indicate the impact of each cluster on the model's output.

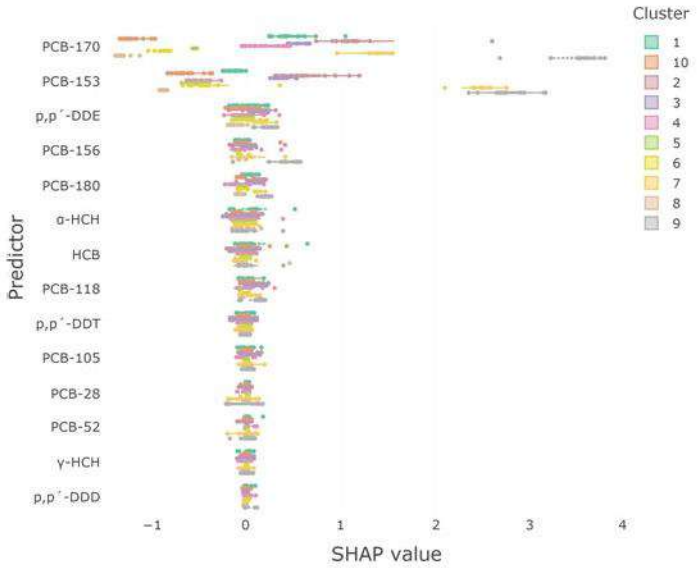


Fig. 6 Distribution of SHAP values within cluster for predictors explaining 90% of PCB-138 concentrations

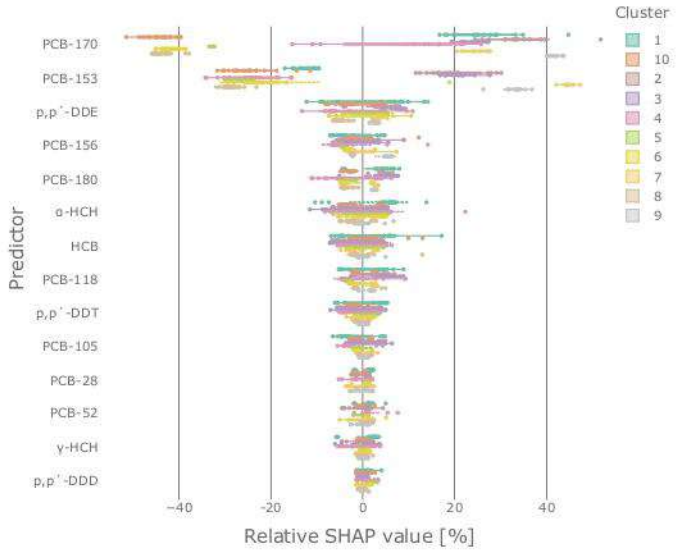


Fig. 7 Distribution of relative SHAP values within cluster for predictors explaining 90% of PCB-138 concentrations

4 Conclusion

The health burden of organochlorines in breast milk represents an issue of global concern because of the adverse impact of POPs on human health, particularly on sensitive sub-populations such as women and children. In this study, we presented a promising explainable artificial intelligence methodology (XGBoost and SHAP) with the aim of investigating organochlorine patterns as well as the dependence of PCB-138 on the mother's age and number of births. 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 an impact on PCB-138 distribution. No significant functional dependencies of PCB-138 patterns and maternal age and parity were observed, suggesting that the impacts of dietary habits and the health burden of POPs in residential and working environments should be investigated further in the future. The herein applied analyses could serve as a promising methodology for future epidemiological investigations and human health protection.

Acknowledgements The authors acknowledge the 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 and the bilateral scientific research project between the Republic of Serbia and Croatia 2019–2021 (No. 337-00-205/2019-09/22) "Persistent Organic Compounds in breast milk and their effects on the level of primary DNA damage in human cells". Also, this paper was financed by the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI—ATLAS.

References

1. Aerts, R., Van Overmeirea, 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.: Determinants of persistent organic pollutant (POP) concentrations in human breast milk of a cross-sectional sample of primiparous mothers in Belgium. Environ. Int. **131** (2019)
2. 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.: Country-specific chemical signatures of persistent organic pollutants (POPs) in breast milk of French, Danish and Finnish women. Environ. Pollut. **218**, 728–738 (2016)
3. Berghuis, S.A., Van Braeckel, K.N.J.A., Sauer, P.J.J., Bos, A.F.: Prenatal exposure to persistent organic pollutants and cognition and motor performance in adolescence. Environ. Int. **121**, 13–22 (2018)
4. 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)
5. De Cock, M., De Boer, M.R., Govarts, E., Iszatt, N., Palkovicova, L., Lamoree, M.H., Schoeters, G., Eggesbø, M., Trnovec, T., Legler, J., Van de Bor, M.: Thyroid-stimulating hormone levels in newborns and early life exposure to endocrine-disrupting chemicals: analysis of three European mother-child cohorts. Pediatr. Res. **82**, 429–437 (2017)
6. Fång, J., Nyberg, E., Winnberg, U., Bignert, A., Bergman, Å.: Spatial and temporal trends of the Stockholm convention POPs in mothers' milk—a global review. Environ. Sci. Pollut. Res. **22**, 8989–9041 (2015)

7. Fry, K., Power, M.C.: Persistent organic pollutants and mortality in the United States, NHANES 1999–2011. Environ. Health **16**, 105 (2017)
8. Govaerts, A., Verhaert, V., Covaci, A., Jaspers, V.L.B., Berg, O.K., Addo-Bediako, A., Jooste, A., Bervoets, L.: Distribution and bioaccumulation of POPs and mercury in the Ga-Selati River (South Africa) and the rivers Gudbrandsdalslågen and Rena (Norway). Environ. Int. **121**, 1319–1330 (2018)
9. Hartmann, J.: Classification using decision tree ensembles. Available at SSRN 3484009 (2019)
10. Jovanović, G., Herceg Romanić, S., Stoji'c, A., Klini'c, D., Matek Sari'c, M., Grzunov Letinić, J., Popović, A.: Introducing of modeling techniques in the research of POPs in breast milk—a pilot study. Ecotoxicol. Environ. Saf. **172**, 341–347 (2019)
11. Kezios, L.K., Xinhua, L., Cirillo, M.P., Cohn, A.B., Kalantzi, I.O., Yunzhu, W., Petreas, X.M., Park, J.S., Factor-Litvak, P.: Dichlorodiphenyltrichloroethane (DDT), DDT metabolites and pregnancy outcomes. Reprod. Toxicol. **35**, 156–164 (2013)
12. Klinić, D., Herceg Romanić, S., Brić Karaonji, I., Matek Sarić, M., Grzunov Letinić, J., Brajenović, N.: Organochlorine pesticides and PCBs (including dl-PCBs) in human milk samples collected from multiparae from Croatia and comparison with primiparae. Environ. Toxicol. Pharmacol. **45**, 74–79 (2016)
13. Klinić, D., Herceg Romanić, S., Matek Sarić, M., Grzunov, J., Dukić, B.: Polychlorinated biphenyls and organochlorine pesticides in human milk samples from two regions in Croatia. Environ. Toxicol. Pharmacol. **37**, 543–552 (2014)
14. Lignell, S., Winkvist, A., Bertz, F., Rasmussen, K.M., Glynn, A., Aune, M., Brekke, H.K.: Environmental organic pollutants in human milk before and after weight loss. Chemosphere **159**, 96–102 (2016)
15. Lundberg, S., Lee, S.: A unified approach to interpreting model predictions. Adv. Neural Inf. Process Syst. 4765–4774 (2017)
16. Man, Y.B., Chow, K.L., Guan, H.X., Janet, K.Y.C., Sheng, C.W., Wong, M.H.: 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. Environ. Pollut. **227**, 364–371 (2017)
17. Mannetje, A., Coakley, J., Mueller, F.J., Harden, F., Toms, L.-M., Douwes, J.: Partitioning of persistent organic pollutants (POPs) between human serum and breast milk: a literature review. Chemosphere **89**, 911–918 (2012)
18. Mitchell, R., Fank, E.: Accelerating the XGBoost algorithm using GPU computing. PeerJ Comput. Sci. **3**, e127 (2017)
19. Nielsen, D.: Tree boosting with XGBoost. NTNU Norwegian University of Science and Technology (2016)
20. Polder, A., Skaare, J.U., Skjerve, E., Løken, K.B., Eggesbø, M.: 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. **407**, 4584–4590 (2009)
21. Stanišić, S., Perišić, M., Jovanović, G., Milićević, T., Herceg Romanić, S., Jovanović, A., Šoštarić, A., Udoviić, V., Stojić, A.: The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part I: emission sources. Environ. Res. **193**, 110520 (2021)
22. Stojić, A., Jovanović, G., Stanišić, S., Herceg Romanić, S., Šoštarić, A., Udoviić, V., Perišić, M.: The PM2.5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: explainable prediction of benzo[a]pyrene levels. Submitted for publication (2021)
23. Stojić, A., Matek Sarić, M., Herceg Romanić, S.: SHapley additive explanations of indicator PCB-138 distribution in breast milk. In: Book of Proceedings of International Scientific Conference on Information Technology and Data Related Research (Sinteza 2020), Singidunum University, pp. 35–40 (2020)
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. **653**, 140–147 (2019)
25. United Nations Environment Programme.: 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. In: Conference of the Parties to the Stockholm Convention on Persistent Organic Pollutants, Sixth meeting, Geneva, 28 April 2010 (2013)

26. Van den Berg, M., Kypke, K., Kotz, A., Tritscher, A., Yong Lee, S., Magulova, K., Fiedler, H., Malisch, R.: WHO/UNEP global surveys of PCDDs, PCDFs, PCBs and DDTs in human milk and benefit-risk evaluation of breastfeeding. Arch. Toxicol. **91**, 83–96 (2017)
27. Vasios, G., Kosmidi, A., Kalantzi, O.I., Tsantili-Kakolidou, A., Kavantzas, N., Theocharis, S., Giaginis, C.: 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. **12**, 1273–1278 (2016)
28. Vigh, É., Colombo, A., Benfenati, E., Håkansson, H., Berglund, M., Bódis, J., Garai, J.: 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. Sci. Total Environ. **449**, 336–344 (2013)
29. Vukavić, T., Vojinović Miloradov, M., Mihajlović, I., Ristivojević, A.: Human milk POPs and neonatal risk trend from 1982 to 2009 in the same geographic region in Serbia. Environ. Int. **54**, 45–49 (2013)

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

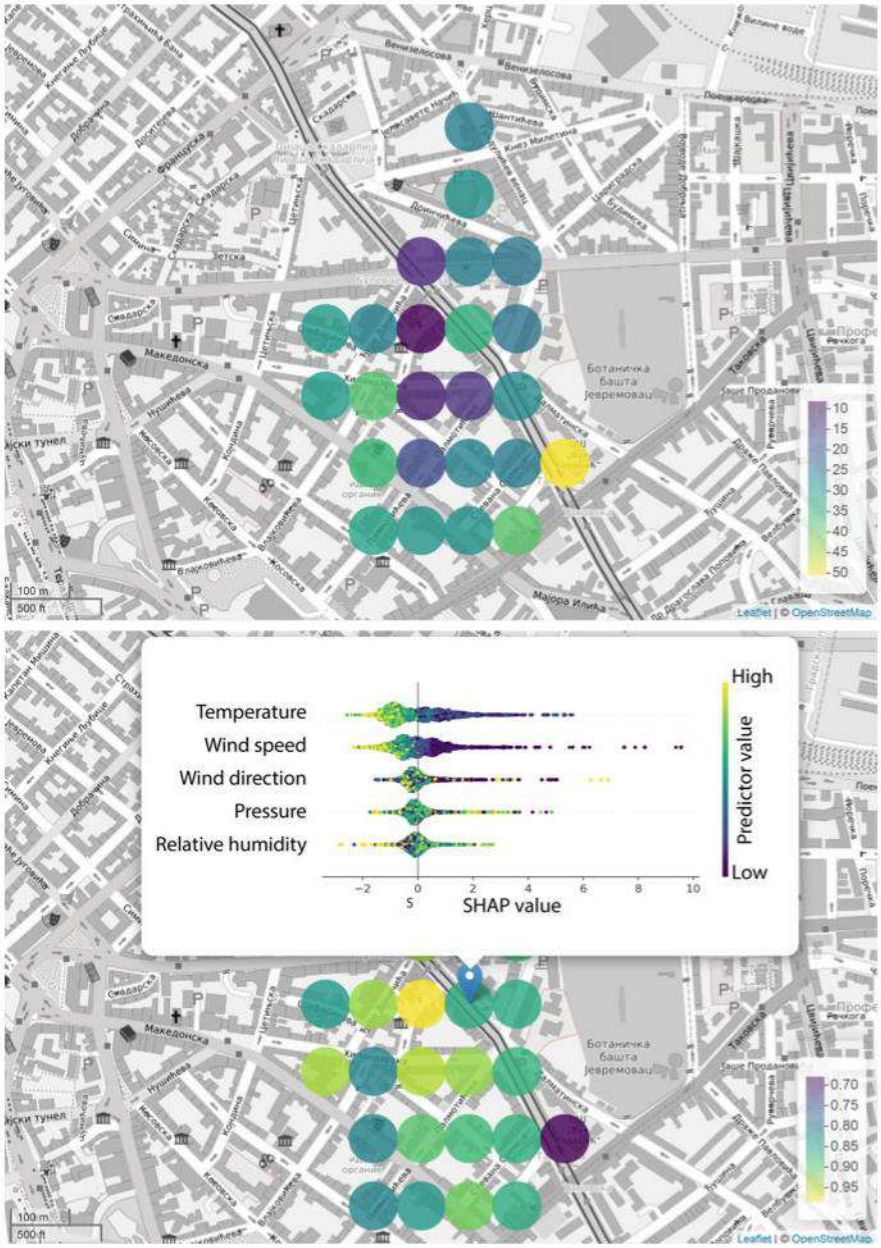


Fig. 1 Benzene forecast based on meteorological parameters—a relative error [%] (top) and SHAP values and predicted/observed correlation coefficients (bottom)

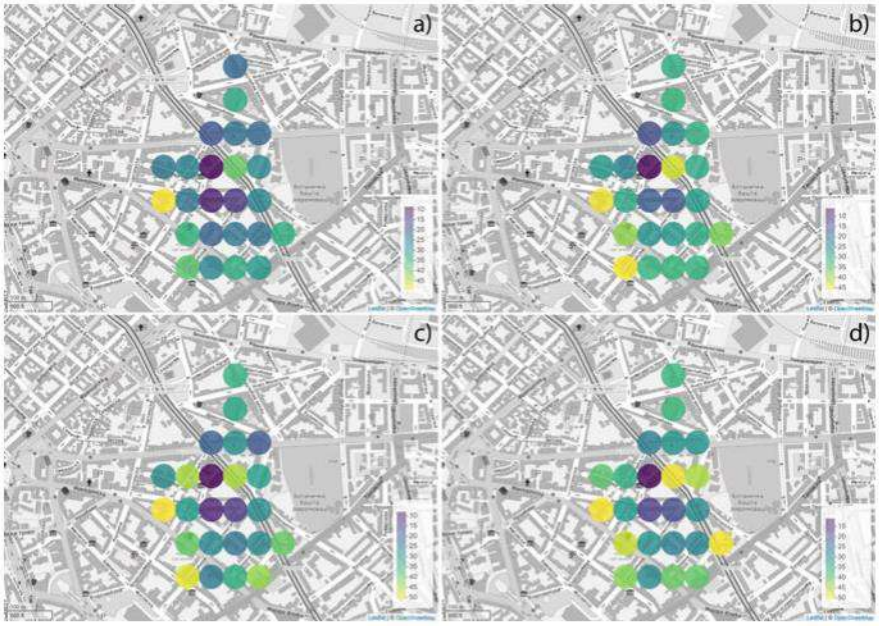


Fig. 2 Toluene (**a**), m,p-xylene (**b**), o-xylene (**c**), and ethylbenzene (**d**) relative error [%] forecasts based on meteorological parameters

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-

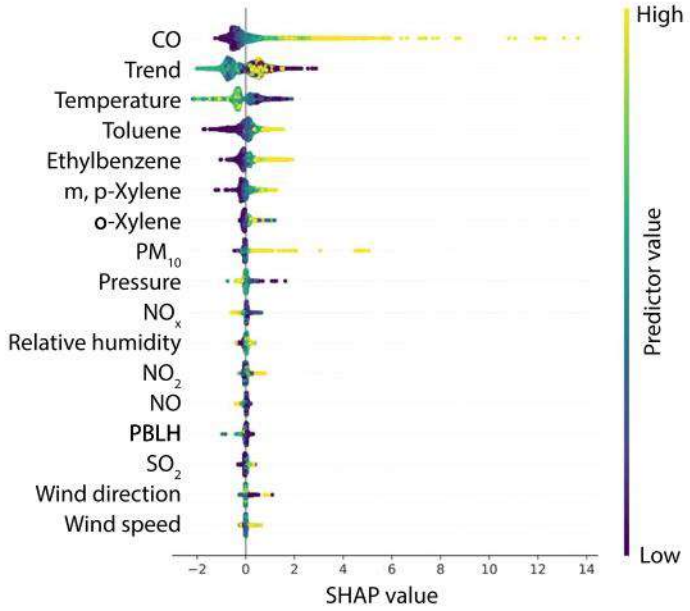


Fig. 3 SHAP summary plot for benzene

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.

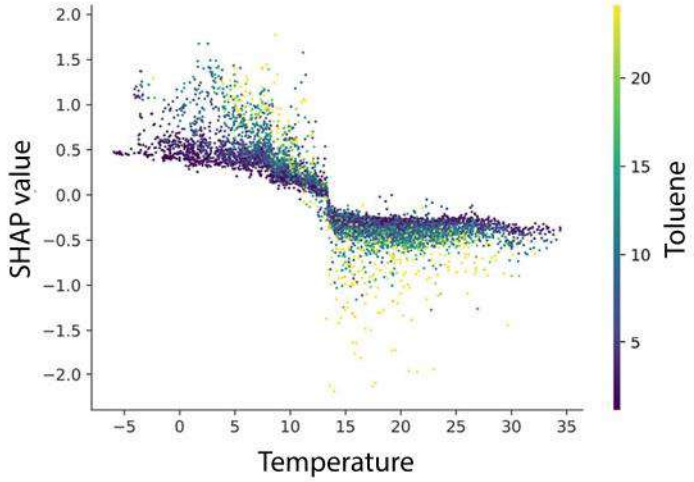


Fig. 4 Benzene SHAP dependency on temperature and toluene

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

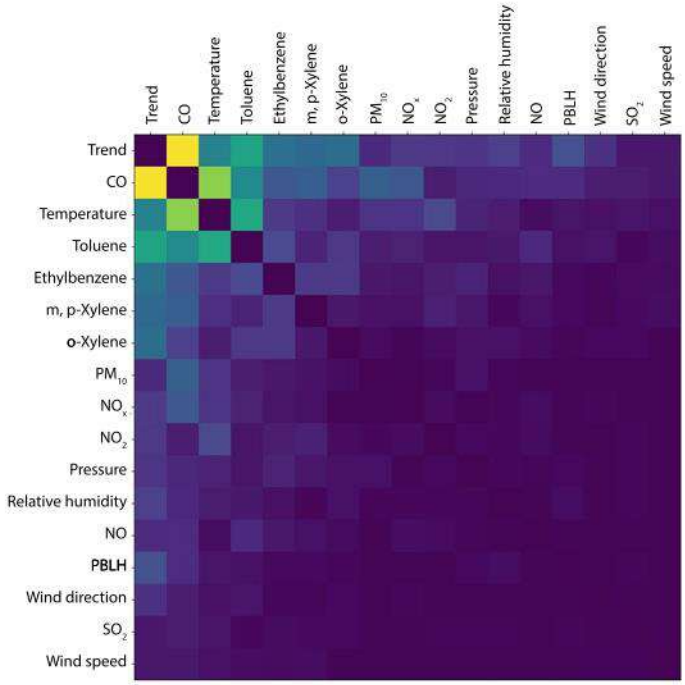


Fig. 5 SHAP interaction plot

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

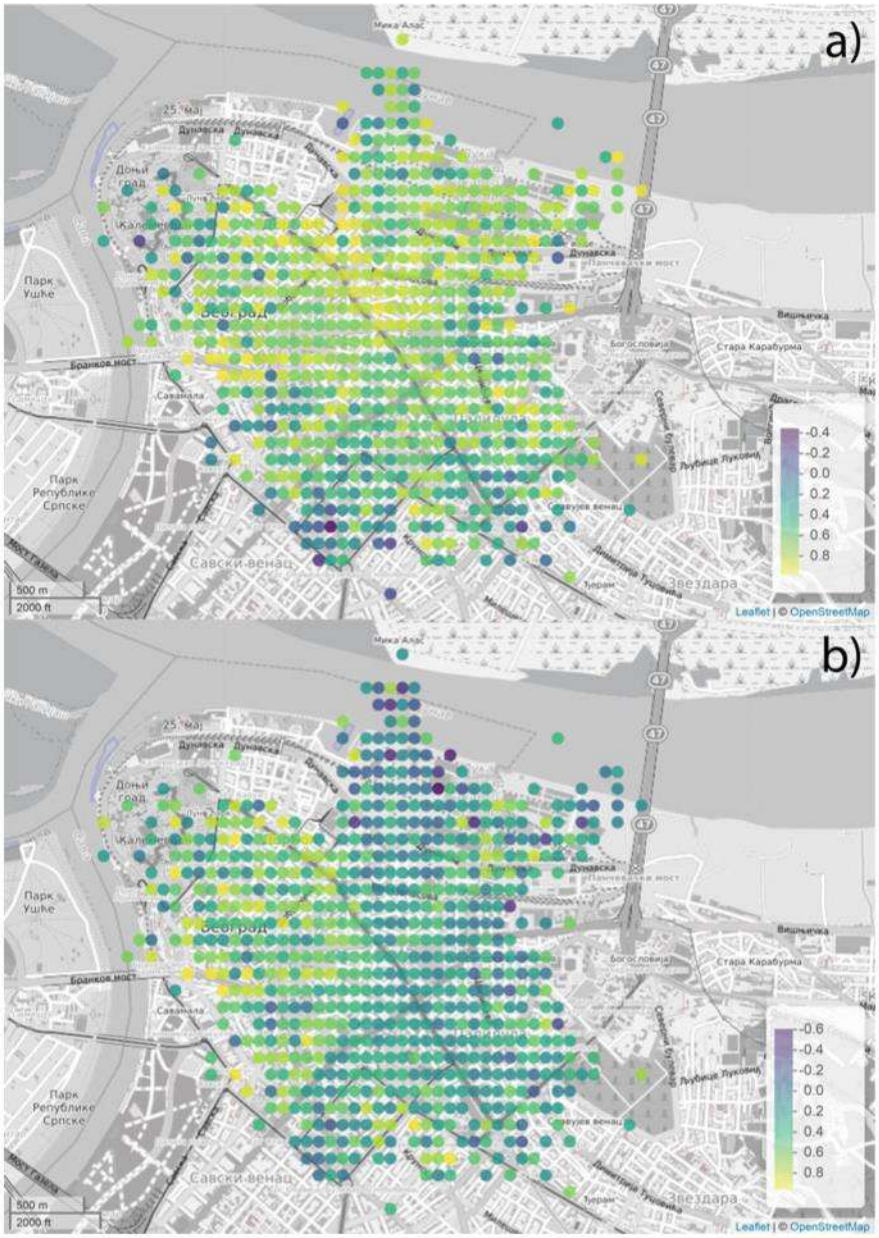


Fig. 7 Benzene correlation coefficients with CO (**a**) and NO_x (**b**)

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)

Quick Find

Use keywords to find the product you are looking for.
Advanced Search

What's New?

Fetal Alcohol Syndrome: Recognition, Differential Diagnosis and Long-Term Effects
$73.80

Shopping Cart

0 items

Information

Shipping & Returns
Privacy Notice
Conditions of Use
Contact Us

Notifications

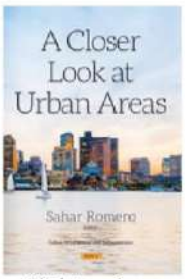
Notify me of updates to A Closer Look at Urban Areas

Tell A Friend

Tell someone you know about this product.

A Closer Look at Urban Areas

$82.00



Click to enlarge

Editors: Sahar Romero

Book Description:

Since the 'smart' component became one of the pillars of the European Union's development, many cities, particularly post-socialist ones, have perceived the smart city concept as a remedy for their problems in the economic, social or image-related spheres. *A Closer Look at Urban Areas* reviews and analyzes the implementation of the smart city concept applied in recent years in cities of the new Member States.

Due to the global phenomenon of urbanization, we are witnesses to cities' growth and the tremendous effects the growth process has on the environment and the health of populations inhabiting urban areas. Apart from numerous benefits such as opportunities for employment, entertainment, education and access to health care, the authors aim to address how urban lifestyles are also inextricably linked to numerous adverse health effects.

The increase in population, economic development, urbanization, industrialization and transport has also raised the issue of air pollution as one of the pressing concerns for contemporary society, primarily due to the harmful effects of elevated concentrations of polluting species on public health, environment and climate. Thus, the authors maintain that researching pollutant wet scavenging by atmospheric water is important for revealing the fate of air contaminants in atmospheric, terrestrial and aquatic systems. (Nova)

Special Focus Titles

01. Violent Communication and Bullying in Early Childhood Education
02. Cultural Considerations in Intervention with Women and Children Exposed to Intimate Partner Violence
03. Chronic Disease and Disability: The Pediatric Lung
04. Fruit and Vegetable Consumption and Health: New Research
05. Fire and the Sword: Understanding the Impact and Challenge of Organized Islamism, Volume 2

Table of Contents:

Series:
Urban Development and Infrastructure
Binding: ebook
Pub. Date: September 13, 2018
ISBN: 978-1-53614-450-5
Status: AV

Status Code	Description
AN	Announcing
FM	Formatting
PP	Page Proofs
FP	Final Production
EP	Editorial Production
PR	At Prepress
AP	At Press
AV	Available

Available Options:

Version: Download: Windows - English ▼

Add to Cart

Saturday 23 November, 2019

A Closer Look at Urban Areas

Sahar Romero
Editor

Urban Development and Infrastructure

NOVA

URBAN DEVELOPMENT AND INFRASTRUCTURE

A CLOSER LOOK AT URBAN AREAS

URBAN DEVELOPMENT AND INFRASTRUCTURE

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

URBAN DEVELOPMENT AND INFRASTRUCTURE

A CLOSER LOOK AT URBAN AREAS

SAHAR ROMERO
EDITOR

nova
science publishers
New York

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

Library of Congress Cataloging-in-Publication Data

ISBN: ;9:/3/75836/672/7"*gDqqm+

Published by Nova Science Publishers, Inc. † New York

CONTENTS

In: A Closer Look at Urban Areas
Editor: Sahar Romero
ISBN: 978-1-53614-449-9

Chapter 3

URBAN AIR POLLUTION: AN INSIGHT INTO ITS COMPLEX ASPECTS

Andreja Stojić[1,*], Gordana Vuković[1], Mirjana Perišić[1], Svetlana Stanišić[2] and Andrej Šoštarić[3]

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

ABSTRACT

The increase in population, economic development, urbanization, industrialization and transport, has raised the issue of air pollution as one of the pressing concerns for contemporary society, primarily because of the harmful effects of elevated concentrations of polluting species on public health, environment and climate system which are still insufficiently understood. Despite the perennial intensive theoretical and experimental research, identifying and parsing the environmental factors that affect complex ecosystems with numerous inputs and outputs in subtle or severe ways, remain a challenge. The main reason lies in the

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

fact that continuous burden on urban environment is governed not only by the increasingly intricate pollutant load, but also by pollution transport, dispersion and deposition processes, atmospheric chemistry, meteorological factors, topography, natural processes, etc. Moreover, containing not only gases, but also suspended particles and clouds, the atmosphere behaves as a colloidal medium. In this regard, researching the pollutant wet scavenging by atmospheric water is very important for revealing the environmental fate of air contaminants in atmospheric, terrestrial and aquatic systems.

The development of innovative measuring and modelling techniques, as well as the exploration of new scientific approaches, methods and technologies, are already underway with the aim of studying the interrelations of air pollution and its association with geophysical parameters. Over the last years, our research has been focused on the utilization and development of measurement methods capable of simultaneous real-time detection, monitoring and quantification of air pollution, as well as statistical and machine learning methods that can provide critical information regarding spatio-temporal pollutant variability and the non-linear nature of air pollution phenomena. This chapter presents the advances in machine learning application and fluctuation analysis aimed at capturing pollutant non-linear dynamics, non-stationarities and their relationships with relevant environmental factors; source apportionment methods for individual source characterization in broader areas; three-dimensional hybrid receptor modeling for the detection of pollution circulation patterns and its altitude distributions on various spatial scales; and environmental multiphase system analysis for the investigation of the mechanisms related to the pollutant particulate matter-water-gas transfer and distribution.

Keywords: air pollution, urban environment, environmental multiphase systems, machine learning, hybrid receptor models

INTRODUCTION

The environmental impacts of urbanization and climate change are probably two of the most pressing issues modern society is facing (WHO, 2016; UN, 2016). The unpredicted rate and the diversity of urbanization, including rapid geographical expansion and population growth, industry development, the increasing need for support infrastructure, and pollution and waste generation, lead to unprecedented change in the environment

which needs a deep understanding of its nature and of the responses it requires. For many key parameters, such as the variety of determinants, spatio-temporal dynamics and interconnections, urban systems are already moving far beyond the patterns which sufficient description would consider simple starting points. Micro-climates, green and built environment, the significant flows of people, energy, materials, resources, and pollutants, make the unlimited set of pathways for urban-ecosystem interactions which reflect the complexity of the urban environment.

Environmental pollution, encompassing air, water, soil, radioactive, light and thermal pollution, and noise, depends on complex interactions driven by a variety of forces embedded in numerous environmental factors. A broad spectrum of relationships being generated and operated on numerous levels, requires multiscale approach to understand the growing body of evidence on pollution impact on the environment, climate system, biodiversity and human health.

The recognition of the complexity of urban environmental factors moved analytical concepts towards a more sophisticated and synergistic modeling. An effort is made not only to identify key factors which shape the urban environment, but also to gain the convergence on the obtained facts, no matter if they are mutually complementary or in agreement.

In this respect, consideration of the air pollution issues from different viewpoints relies on the application and development of specialized methods and their assembling and hybridization for capturing pollutant statistical and spatio-temporal distributions, fluctuations, lag and cumulative effects, their interrelations and relationship with relevant environmental factors on various scales, etc. This chapter systematizes advances made in several case studies aimed at revealing some urban air pollution aspects of complexity. We place the focus on analytical tools capable of providing significant scientific information gain based on the characterization of the processes that generate pollution, govern its spatio-temporal dynamics and determine its environmental fate. Beside the application of statistical and machine learning methods for characterization of pollutant interrelations, non-linear dynamics, non-stationarities and singularities (Stojić et al., 2016, 2017a; Perišić et al., 2015, 2017a;

Todorović et al., 2016), this chapter presents their potential for air pollution forecasting (Stojić et al., 2015a; Perišić et al., 2017b), as well as for the identification of mechanisms underlying wet scavenging phenomena (Šoštarić et al., 2016; 2017). Furthermore, the source apportionment advances presented herein shift the focus from dominant sources to a large number of individual, locally specific emissions (Perišić et al., 2017a). Finally, in addition to generally accepted statistical methods for potential remote emission source identification, we present an improved air pollution transport analysis based on three-dimensional hybrid receptor models for the identification of volumetric pollution circulation patterns and pollutant altitude distributions on various spatial scales (Stojić and Stanišić Stojić, 2017).

METHODS

The data used for the analyses presented in this chapter were obtained from the Institute of Public Health Belgrade regular monitoring network and measurement campaigns conducted at sites located in Belgrade (Serbia) urban and semi-urban areas. The dataset comprises of volatile organic compound (VOC) measured at 39 molecular mases, inorganic gaseous pollutants (CO, SO_2, NO, NO_2, NO_x and tropospheric O_3), particulate matter (PM_{10} and $PM_{2.5}$) and its chemical constituents (As, Cd, Cr, Mn, Ni, Pb, BaP, Cl^-, NO_3^-, NH_4^+, SO_4^{2-}, Na^+, K^+, Mg^{2+} and Ca^{2+}), rainwater physico-chemical characteristics (rain temperature, turbidity, pH, UV extinction, total organic carbon, electric conductivity, F^-, Cl^-, NO_2^-, NO_3^-, SO_4^{2-}, Na^+, NH_4^+, K^+, Ca^{2+} and Mg^{2+}), rainwater concentrations of benzene, toluene, xylene and ethylbenzene (BTEX), and meteorological parameters (wind speed and direction, relative humidity, pressure and temperature).

Statistical analyses, including bivariate polar plot and bivariate cluster (k-means clustering, grouping similar conditions together) analysis, were performed with the statistical software environment R (Team, 2014), using the Openair package (Carslaw and Ropkins, 2012). Multifractal detrended

fluctuation analysis (MF-DFA) (Stojić et al., 2016) and multiscale multifractal analysis (MMA) (Stojić et al., 2017a) were used to investigate the presence of fractal behavior in complex $PM_{2.5}$ and PM_{10} time series. MMA is a generalization of the standard MF-DFA, which adds the dependence on scale, providing a broader analysis of the fluctuation properties, as well as more general and stable results (Gierałtowski et al., 2012).

Apportionment of dominant sources was obtained by the use of Unmix and Positive matrix factorization (PMF) (USEPA, 2007). The detailed descriptions of the models are presented elsewhere (Henry, 2003; Paatero, 1999). On the other hand, source apportionment of individual, locally specific emissions and information about pollutant relationships were obtained by means of the advanced bivariate polar plots coupled with pair-wise statistics applied to distinguish specific sources and to gain information about pollutant relations (Grange et al., 2016). The method includes weighted Pearson correlation, linear regression slope and Gaussian kernel to locally weight the statistical calculations on a wind speed-direction surface together with variable-scaling.

Time series decomposition into the additive components of the multiyear and seasonal trends, as well as the remainder was conducted using the Loess smoothing decomposition model (LSD) (Li et al., 2014), while hourly, daily, weekly and seasonal periodicity was analyzed by the use of Lomb-Scargle method (Ruf, 1999; Team, 2014).

Random forest (RF) and Guided regularized random forest (GRRF) were applied (Deng and Runger, 2013) for the selection of features that are most relevant for PM_{10} concentration (Stojić et al., 2016) and BTEX enrichment factor (EF) prediction (Šoštarić et al., 2017). Random forest consists of decision trees which every node represents a condition on a single variable designed to split the dataset in two parts so that similar responses end up in the same set. Variable importance, a measure based on which the optimal condition is chosen, measures how much each variable decreases the weighted impurity across the trees. GRFF uses the importance scores from a preliminary RF to guide the feature selection of regularized RF and has several advantages: it is robust and computationally

efficient, it can select compact feature subsets moderating the curse of dimensionality, it avoids the effort to analyze irrelevant or redundant features, and it has competitive accuracy performance.

Machine learning algorithms used for the environmental multiphase system analyses (the relationship between EF and features that are considered most relevant for EF prediction) are implemented in Weka 3.8 (Frank et al., 2005) and include: Alternating model tree which grows an alternating model tree by minimizing squared error (Frank et al., 2015); Conjunctive rule which induces a small as possible set of rules from data that captures all generalizable knowledge within that data (Cohen, 1995); Decision stump which represents one-level decision tree (Zhao and Zhang, 2008); Decision table, a majority classifier which evaluates feature subsets using best-first search (Witten and Frank, 2016); Elastic net, a generalization of the lasso method for linear regression which uses penalization methods that introduce additional constraints into the optimization of a predictive algorithm that bias the model toward lower complexity (Zou et al., 2006); Gaussian processes which use lazy learning and a measure of the similarity between points for prediction based on collection of random variables with consistent joint Gaussian distributions (Rasmussen, 2006); IBk and IBkLG which are a types of instance-based learning (k nearest neighbors – k-NN) (Aha et al., 1991; IBkLG, 2015); Isotonic regression which implements the method based on pair-adjacent violators approach by picking the attribute that results in the lowest squared error (Witten and Frank, 2016); K* which is an instance-based classifier predicting by entropy-based distance similarity function (Cleary and Trigg, 1995); Least median squares regression which is an utilization of the existing Weka linear regression (Rousseeuw and Leroy, 2005); Linear regression method which minimizes the sum of the squared difference between the observed and the predicted values creating a line for optimal data separation (Shi and Tsai, 2002); Locally weighted learning (LWL) which uses an instance-based algorithm to assign instance weights where a linear regression model is fit based on a weighting function centered on the instance for which a prediction is to be generated (Frank et al., 2002); M5P which is based on decision trees containing a multivariate

linear regression model at each node (Graczyk et al., 2009); M5 rules which generates a decision list using separate-and-conquer by building a model tree in each iteration and making the "best" leaf into a rule (Wang and Witten, 1997); Multilayer perceptron which is a feed forward neural network model with one or more hidden layers between input and output layer trained by the back-propagation algorithm that uses gradient descent to minimize error and adjust the weights to each connection between the hidden and output layer (Haykin, 1994); Pace regression which consists of a group of estimators that are either overall optimal or optimal under certain conditions, the method provably optimal under regularity conditions when the number of coefficients tends to infinity (Wang and Witten, 2002); Random forest which is an ensemble of unpruned trees, induced from bootstrap samples of the training data, using random feature selection in the tree induction process (Breiman, 2001; Zhao and Zhang, 2008); Random tree which constructs a tree that considers randomly chosen attributes at each node (Zhao and Zhang, 2008); Radial base function which is an artificial neural network which implements a normalized Gaussian radial basis function network and uses the k-means clustering algorithm to provide the basis functions and learns either a logistic regression (discrete class problems) or linear regression (numeric class problems) on top of that (Frank, 2014); REP tree which is a fast decision tree based on reduced-error pruning (with backfitting) (Kalmegh, 2015); Simple linear regression which picks the attribute that results in the lowest squared error; and SMOreg which is a Support vector machine implemented using the Sequential Minimal Optimization Regression (SMOreg) algorithm (Shevade et al., 1999).

Machine learning algorithms used for the VOC source contribution and PM_{10} concentration forecast are implemented in Toolkit for Multivariate Analysis (TMVA) (Hoecker et al., 2007) within the ROOT framework (Brun and Rademakers, 1997) and include (Stojić et al., 2015a): Boosted decision trees (BDT, BDTG); Artificial neural network Multilayer perceptron (MLP); MLP with Bayesian extension (MLPBNN); Support vector machine (SVM); Linear discriminant (LD) which provides an intermediate solution to the problem with the aim to solve relatively simple

or partially nonlinear problems.; Fisher discriminant (Fisher) which determines an optimal separating function in the multivariate space of all input variables analytically for the linear case; Multidimensional probability density estimator range search (PDE-RS) which is a variant of the k-nearest neighbor, Likelihood method and function discriminant analysis (FDA) which extends the linear discriminant to moderately non-linear correlations that are fit to the training data and Function discrimination analysis with genetic algorithm converger (FDA-GA).

Trajectories used for hybrid receptor modelling were calculated 72-h backward by the use of HYSPLIT model (Draxler and Rolph, 2014) for each hour UTC above the sampling sites at the half of the planetary boundary layer (PBL) height calculated from GDAS1 using MeteoInfo (Wang, 2014), as described in Stojić et al. (2016) and Stojić and Stanišić Stojić (2017).

RESULTS AND DISCUSSION

Spatio-Temporal Dynamics

In the atmosphere, air pollutants are prone to dynamic changes which do not occur as linear or single-compartment processes due to mutual interactions of numerous factors that drive their temporal fluctuations, such as meteorological parameters, type and intensity of emission sources, etc.

Temporal Variations and Periodicity

Atmospheric pollution levels exhibit variations on different time scales, from short-term diurnal fluctuations to long-term seasonal and annual deviations. Moreover, once emitted into the atmosphere, polluting species undergo a number of physical and chemical processes including atmospheric transport and deposition at a range of local and regional scales. At a city-scale, a variety of pollution sources, meteorological regime, urban microclimate and local topography, such as building distribution and street geometry, anthropogenic heat fluxes and planetary

boundary layer prominently affect the dispersion and dilution of air pollutants (Monks et al., 2009). While the studied urban areas in Belgrade (urban canyon avenue – UCA and urban boulevard – UB) are mostly affected by vehicle exhaust emissions, the suburban sites (urban industry – UI and rural industry – RI) are affected by emissions from fossil fuel burning for industry and heating purposes.

According to the results, spatio-temporal dynamics reveals annual cycle as the dominant mode of PM_{10} (Figure 1), NO and SO_2 (Figure 2) variability at each sampling location. The annual variations show pronounced winter maxima likely due to an increase of emissions from combustion processes and low PBL height, as well as the summer minima which could be attributed to the enhanced dispersion and dilution due to the expansion of daily PBL. At smaller temporal scales, a decrease of anthropogenic emissions during weekend reduces the amplitude of pollutant concentrations. Furthermore, the development and break-up of the nocturnal inversion layer, as well as the morning and afternoon rush-hour peaks shape diurnal variations, especially in an urban area.

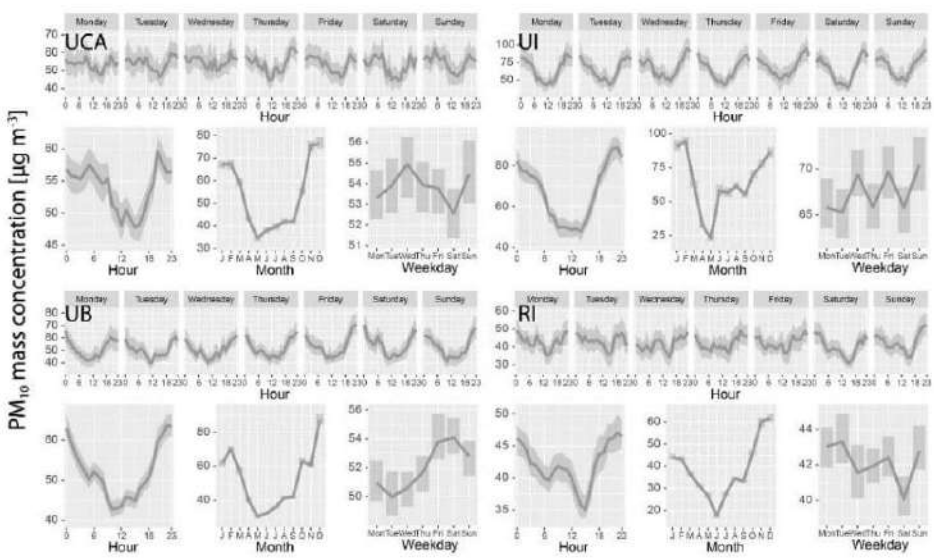


Figure 1. Daily, weekly and seasonal variations of PM_{10} mass concentrations.

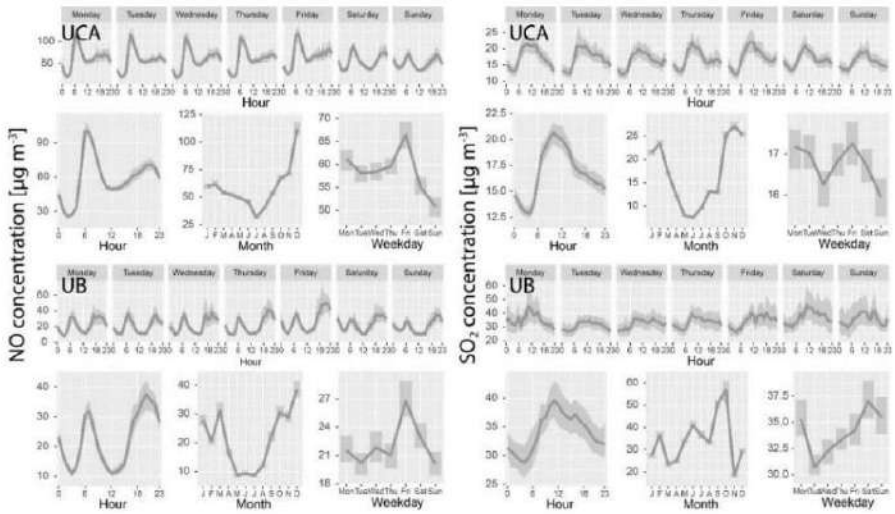


Figure 2. Daily, weekly and seasonal variations of NO and SO_2 concentrations.

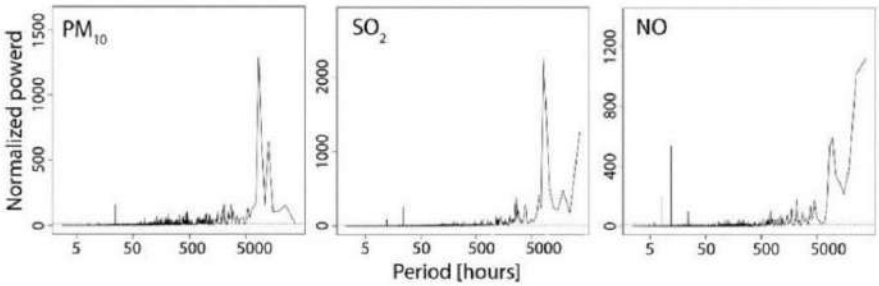


Figure 3. Lomb-Scargle periodogram.

Spectral analysis (Figure 3) indicates the importance of periodicity for the air pollution dynamics. At the scale of 12 and 24 h, 7 days and one year, the variability is mostly governed by meteorological conditions and anthropogenic emissions, while at the scales below 12 h, as in the case of NO, the additional peaks can be attributed to the local traffic-related emission sources.

Fluctuation Analysis

The constitutional characteristics of the pollutant time-series, i.e., their changes over time, could not be revealed by conventional approaches, but fractal (Mandelbrot, 1977) and Hurst rescaled analyses (Hurst, 1951;

Mandelbrot and Wallis, 1969) allow to quantify its structure via self-similarity and scale invariance. In brief, self-similarity and scale-invariance are the properties which characterize the pollutant long-term memory, i.e., fractal behavior, and the complexity of their concentrations. In other terms, the obtained results describe the effects of past patterns on future concentration or the extent to which time-series are nonrandom in behavior. The multifractal approach has been widely for the estimation of long-range correlations and forecasting of the PM_{10}, $PM_{2.5}$, NO_x, SO_2, and O_3 time-series (Windsor and Toumi, 2001; Dong et al., 2017; Plocoste et al., 2017; Stojić et al., 2017).

In our studies investigating PM_{10} and $PM_{2.5}$ time-series in Belgrade urban area (Stojić et al., 2017), the strength of multifractality was presented by the value of multifractal singularity, Q. The bigger the Q, the more the time-series is influenced by small or large fluctuations in negative or positive direction of Q interval while Q=0 reflects the absence of fluctuations and a monofractal behavior (He et al., 2017). Apart from Q, the following values of Hurst Exponent (H) were considered for searching for the time-series persistence level (Ihlen, 2012; Molino-Minero-Re et al., 2015):

- $H \geq 1.5$ represents Brown noise uncorrelated processes with infinite memory;
- H=1 is pink noise, the most well-known, firm and adaptable fractal phenomena with long memory;
- H>0.5 and H=0.5–1.5 illustrates random walk processes with long-range correlated and persistent structure;
- H=0.5 represents white noise, such uncorrelated random variables with no memory exhibited;
- H<0.5 reveals the anti-persistent increments and anti-correlated structure.

In addition, the span of multifractal singularity ($\Delta\alpha = \alpha_{max} - \alpha_{min}$) is widely used as an alternative way to study the strength of multifractality

(Stojić et al., 2016). The bigger $\Delta\alpha$ is, the stronger the multifractality degree is (He, 2017).

Both PM_{10} and $PM_{2.5}$ fluctuated nonrandom over time and were characterized by non-stationary signals with long-range dependent structure that reflects their multifractal behavior. It is shown that the source of multifractality, examined by PM time series randomization, originates both from nonlinear correlations and a fat-tailed probability distribution (Figure 4). More specifically, long-range correlated time series of PM_{10} are evident (H: 1−1.5), but their multifractality weakened over time and approached to the "healthy complex system" of pink noise (Figure 5).

As can be seen, PM_{10} patterns occasionally referred to a Brown noise (H>1.5) at temporal scales below 90 that could be associated with oscillation in emission sources that are sensitive to meteorological influences. In general, the PM_{10} persistence could be understood in terms of the steady temperature, humidity and general atmospheric circulation; the relationship between meteorological factors and PM_{10} is obvious since PM_{10} represents a mixture of microscopic solid or liquid suspended matters involved in inherent condensation and nucleation (He, 2017; Zhang et al., 2015). Furthermore, multifractal approach applied for studying PM_{10} seasonal dynamics indicated the coincidence in the maximum of autumn and winter spectra suggesting that similar sources and processes were underlying the concentrations observed in both periods (Stojić et al., 2016). Similarly, spring and summer spectra led to no significant differences in the origin of the observed PM_{10} loadings.

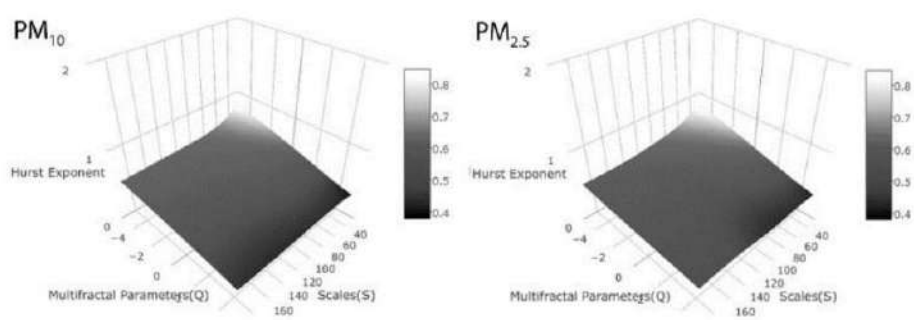


Figure 4. MMA derived Hurst surfaces for randomized PM time series.

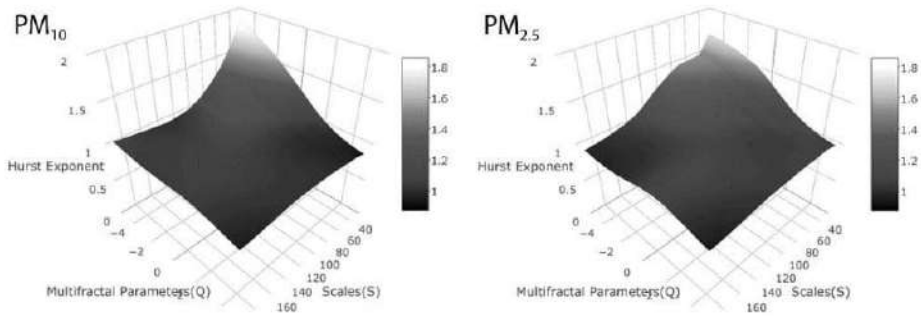


Figure 5. MMA derived particulate matter Hurst surfaces.

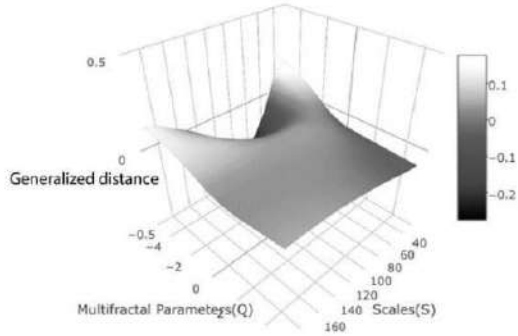


Figure 6. Generalized $PM_{10}/PM_{2.5}$ Hurst surface distance.

In accordance to PM_{10}, the $PM_{2.5}$ time series were persistent as shown by the local Hurst exponent in the interval between 1 and 1.5, but slightly affected by the concentrations occurring randomly. However, unlike PM_{10}, clear crossover was evident at the scales below 44, which corresponds to a period of about 2 days, denoting significant impact of random events that emerges at this scale. These indications support the findings of Xue et al. (2015) who reported dissimilarities among PM of different sizes in urban areas. More specifically, the authors recorded a strong multifractal nature of PM size below 5 μm, and a weak and/or monofractal stochastic behavior of PM>5 μm. In addition, generalized distance coefficient (0.069) between PM fractions Hurst surfaces exceeded the threshold value (0.065) that implies the $PM_{2.5}$ and PM_{10} time series to be considered statistically different (Figure 6); the difference is particularly pronounced in the area of small fluctuations and medium scales.

Contribution Decomposition

Concentration Decomposition

Discriminating the relative importance of background, local sources and transport processes and estimating their shares in pollutant concentrations can be considered as one of the key issues in air pollution analysis. We can assume that in urban areas, dominated by local emissions, transport and background jointly contribute to gradual variations of a concentration base level, whereas the superimposed pronounced peaks in pollutant time series occur as a result of local emissions (Figure 7).

Thereby, the differentiation between the contributions of local and remote emissions can be obtained by a two-step procedure, as described in our previous studies (Stojić et al., 2016; Stojić and Stanišić Stojić, 2017). In brief, excluding the contribution of local sources from the time series and obtaining a baseline can be performed by a number of functions available for baseline extraction (Kneen and Annegarn, 1996). Subsequently, Trajectory sector analysis (TSA) can be applied to the derived baseline to distinguish between the contributions of background and transport, and to obtain a transport time series, further used as an input for hybrid receptor modelling. In addition to TSA, background levels can be obtained by the use of 3D hybrid receptor models, which will be shown below.

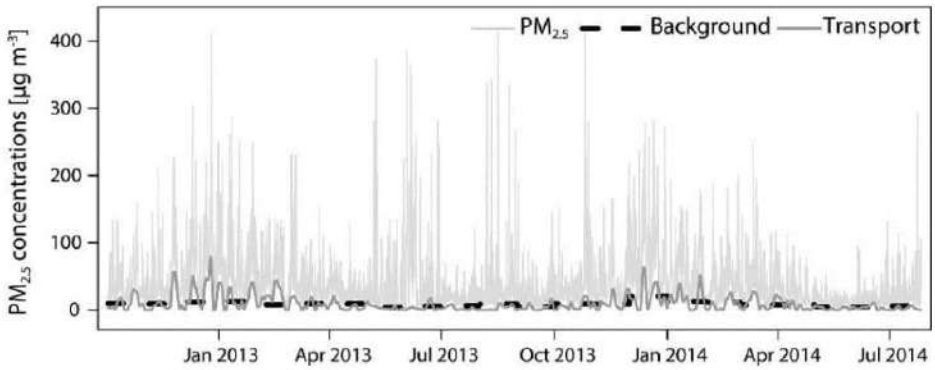


Figure 7. $PM_{2.5}$ concentration decomposition (Stojić and Stanišić Stojić, 2017).

Time Series Decomposition

Figure 8 illustrates the decomposition of PM_{10} temporal cycles indicating moderate impact of trend, significant impact of the seasonal and very large variance of the remainder components. In general, the remainder, expressed as irregularity, neither explained by the trend, nor by the seasonality, in the urban areas can be associated with anthropogenic processes which induce chemical and dynamical changes in the pollutant concentrations (Chehade et al., 2014).

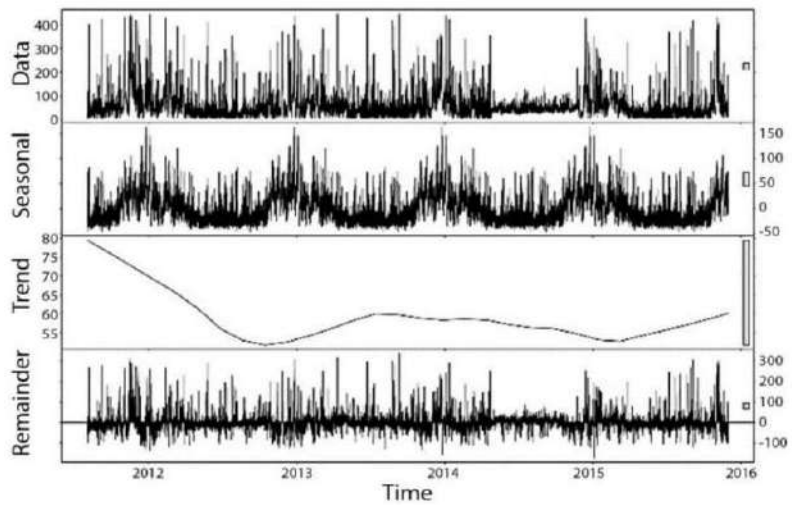


Figure 8. PM_{10} time series decomposition [$\mu g\ m^{-3}$].

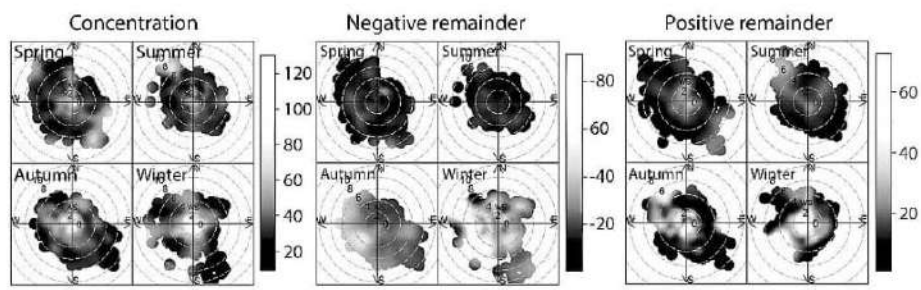


Figure 9. Bivariate polar plot of PM_{10} concentrations (left) and its remainder components: negative (middle) and positive (right) [$\mu g\ m^{-3}$].

To further examine their origin, the remainder shares can be subjected to the bivariate polar plot analysis, separately applied on positive and

negative values. As shown in Figure 9, the episodes of the highest variations in Belgrade mainly occur during the colder part of the year – the positive ones are locally originated, or related to SW winds, while the negative ones are caused by strong NW winds.

Environmental Impacts

Local Sources

Investigating the influences of the surrounding emission sources by the use of bivariate polar plot and bivariate cluster analyses (Figure 10) reveal the dominant average contributions of the sources associated with restricted pollutant dispersion in an urban area (UCA), and by direct exposure to the emissions from the proximate industrial activities (UI and RI). At urban sites such as UB, PM_{10} concentrations tended to elevate as a result of S and SE winds during the episodes of unstable atmospheric conditions (wind speed higher than 8 m s^{-1}). Higher concentrations occur under the impact of heating plant emissions and nearby intersections with intensive traffic. Moreover, seasonal polar plots show the highest concentrations during low wind speed periods and the colder part of the year, which is expected as a result of more pronounced local emission sources and reduced advection.

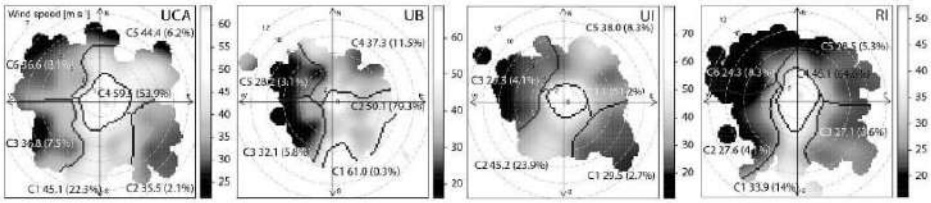


Figure 10. The relationship between PM_{10} concentrations and wind characteristics: bivariate plot (frequency [%] and average contributions [μg m^{-3}]) for the entire period in Belgrade.

Air Pollution Transport

Over the last decades, significant improvements in general hybrid receptor modeling have been made. Receptor-oriented models based on

conditional probability and residence time analysis became widely accepted not only for studying dynamical processes and pollutant circulation patterns in the atmosphere, but for investigation of spatial distribution of potential emission sources and for the assessment of their impact on the receptor site without using emission inventories (Brereton and Johnson, 2012; Bycenkiene et al., 2014; Sen et al., 2016; Li et al., 2017). Hybrid receptor models consider residence time of trajectory segment endpoints in a potential source area, accounting for the above-threshold pollutant concentrations at the receptor site, as in the case of Potential source contribution function (PSCF) (Ashbaugh et al., 1985), or accounting for concentration gradients, as in the case of Concentration weighted trajectory model (CWT) (Hsu et al., 2003).

The conventional approach of hybrid receptor modelling has several drawbacks basically related to the concentration and trajectory inclusion in the model based on their representativeness. In order to obtain more realistic transport analysis, the innovative three-dimensional approach was introduced in Stojić and Stanišić Stojić, 2017. The concept relies on concentration and backward trajectory preprocessing, which enables more accurate inclusion of pollution levels in the transport model, and accounting for meteorological factors that govern the vertical distribution of air pollution, respectively.

Namely, by accounting the concentrations greater than the arbitrary chosen value (e.g., mean) it is implicitly assumed that individual peaks in the pollutant time series arise mainly as a result of transport (Kassomenos et al., 2006; Grivas et al., 2008) which may be considered appropriate only for the background sites. However, suburban and urban areas are dominated by many local emission sources implying that pronounced peaks in pollutant time series originate from local emissions and are superimposed on gradual variations of a concentration baseline level, encompassing both transport and local background contributions. The concept of concentration preprocessing which provide transport time series is described in the previous section, and the trajectory preprocessing concept will be explained in brief herein.

The lowest part of the troposphere, PBL, directly responds to surface forcing (Stull, 1988), by trapping and dispersing air pollution emitted from the ground, so the near-surface pollutant concentrations could correspond to atmospheric concentrations only within the PBL. To account PBL height fluctuations, the three-dimensional improvement of the conventional PSCF and CWT models, as well as new hybrid receptor model, Concentration weighted boundary layer (CWBL), that uses a two-dimensional grid and a PBL height as a frame of reference, are presented.

The contribution of a specific emission source to pollutant concentration at the receptor site is considered to be directly related to air mass residence time over the grid cell where the source is located (Dimitriou and Kassomenos, 2014). The probability of event A_{ij} when n trajectory endpoints fall into the ijth cell is given as:

$$P[A_{ij}] = \frac{n_{ij}}{N} \tag{1}$$

where N is the total number of endpoints (Hopke et al., 1993). However, the probability of high concentration event B_{ij}, when only a subset of m_{ij} grid cell endpoints for which the corresponding trajectories reach the receptor site when the transported concentrations are higher than the criterion value, is given as:

$$P[B_{ij}] = \frac{m_{ij}}{N} \tag{2}$$

The two-dimensional PSCF value for the ijth cell can be defined as:

$$P_{ij} = \frac{P[B_{ij}]}{P[A_{ij}]} = \frac{m_{ij}}{n_{ij}} \tag{3}$$

where the cells with higher PSCF values have a higher probability of containing emission sources.

We define three-dimensional PSCF as the ratio of a subset of m_{ijk} endpoints in the kth volume v_{ijk} of the predefined height above ijth cell for

which the corresponding trajectories arrive at the receptor site when the contribution of transport to the measured concentrations are higher than the criterion value and the total number of endpoints in the *ij*th cell (n_{ij}):

$$P_{ijk} = \frac{P[B_{ijk}]}{P[A_{ij}]} = \frac{m_{ijk}}{n_{ij}} \tag{4}$$

The threshold for the selection of trajectories, which reflects the highest pollutant injection point, can be PBL height, which we consider appropriate for urban areas, or any altitude in the 3D grid map at each trajectory endpoint.

The major drawback of PSCF is related to high values that may occur as a consequence of a small number of grid cell endpoints corresponding to poor air quality at the receptor site. The problem can be overcome by the use of arbitrary weighting functions based on point count multiplied by the local PSCF (Zeng and Hopke, 1989) or by weighting trajectory endpoints (CWT) obtained by averaging estimated transport contribution at the receptor site that corresponds to the trajectories passing across the grid cell (*i,j*):

$$CWT_{ij} = \frac{\sum_{l=1}^{L} C_l \tau_{ijl}}{\sum_{l=1}^{L} \tau_{ijl}} \tag{5}$$

As in the case of 3D PSCF, 3D CWT for each volume cell v_{ijk} can be defined as:

$$CWT_{ijk} = \frac{\sum_{l=1}^{L} C_l \tau_{ijlk}}{\sum_{l=1}^{L} \tau_{ijl}} \tag{6}$$

where C_l is the pollutant concentration corresponding to the arrival of back trajectory l; τ_{ijl} is the number of trajectory segment endpoints in a grid cell (*i,j*) for back trajectory l; τ_{ijlk} is the number of trajectory segment endpoints in *k*th volume cell v_{ijk} for back trajectory l; L is the total number of back trajectories. As a result, the sum of altitude CWT distribution for each *ij*th cell amounts to the conventional 2D CWT solution.

Figure 11 illustrates main source regions altitude distribution that contributed to the elevated concentrations of $PM_{2.5}$ in Belgrade, while Figure 12 illustrates the altitude distribution of transported $PM_{2.5}$ concentrations and VOC industrial emissions according to CWT analysis.

Both 3D PSCF and CWT can be employed only for identification of potential source regions defined by longitude, latitude and altitude, and not for the analysis of pollutant altitude distribution along the transport pathway. For this reason, Concentration weighted boundary layer hybrid receptor model, which uses a 2D grid and a PBL height as a frame of reference, is introduced to obtain the vertical profile of pollutant concentrations above the receptor site and each grid cell along the transport pathway. The method relies on the fact that pollutant altitude distribution does not exhibit significant variation within the largest part of the PBL, except for the very top and very bottom layer (Stull, 1988; Gan et al., 2011). The CWBL value at each 2D grid cell (Figure 13) is calculated by averaging the transport contribution to pollutant concentrations that correspond to all endpoints falling into the selected cell (i,j) within the corresponding PBL heights as:

$$CWBL_{ijh} = mean(C_{l|PBLH_{ijl}^{e} \geq h}^{e}) \tag{7}$$

where C_l^e is the concentration attributed to each endpoint e of trajectory l, and $PBLH_{ijl}^e$ refers to the PBL height at each endpoint in the moment when the air parcel passed grid cell (i, j). Figure 14 illustrates the results of CWBL-derived VOC industrial emissions altitude distribution as seen from the receptor site located in Belgrade urban area (UCA).

Moreover, CWBL is not limited to trajectory endpoint analysis – it provides pollutant altitude distribution above the receptor site (Figure 15). As can be seen, PM_{10} concentrations remained high in the ground layer of the troposphere and exhibited a rapid decline with height, reaching the minimum values for the highest PBL, which is in compliance with the aerosol altitude distribution described by Stull (1988), as well as with the findings of some recent empirical studies obtained by combining a charge-

coupled device (CCD) side-scatter lidar with simultaneous ground level measurements (Tao et al., 2016).

Altitude profiles obtained by the use of CWBL provide an insight into the complexity of several factors that govern pollutant spatio-temporal distribution. Namely, PM_{10} concentrations in the ground layer are directly influenced by spatial distribution of emission sources, primarily consisting of anthropogenic emissions (Bravo-Aranda et al., 2017).

On the other hand, concentrations and residence time of particles in the upper layers are additionally affected by meteorological conditions, topography and complex atmospheric reactions, which further leads to the formation of secondary aerosols.

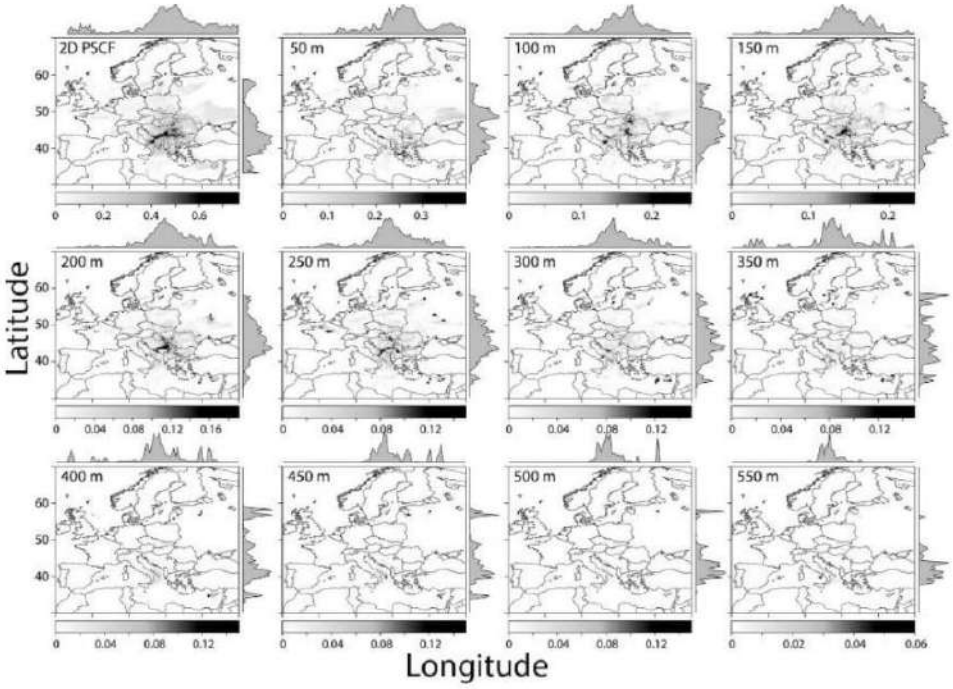


Figure 11. 2D PSCF and 3D PSCF derived maps indicating main source regions altitude distribution that contributed to the elevated concentrations of $PM_{2.5}$ in Belgrade (Stojić and Stanišić Stojić, 2017).

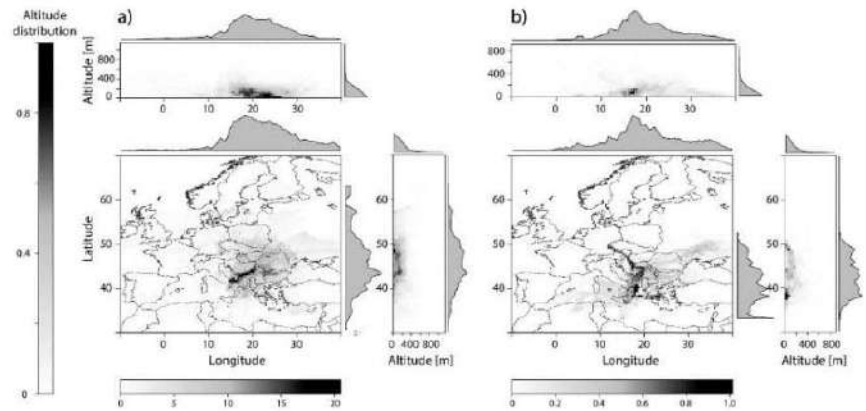


Figure 12. Two-dimensional CWT derived map for $PM_{2.5}$ [μg m^{-3}] (a), and VOC industrial emissions (average = 1) (b), and three-dimensional CWT longitudinal/latitudinal altitude distribution of pollutants during transport within the PBL (Stojić and Stanišić Stojić, 2017).

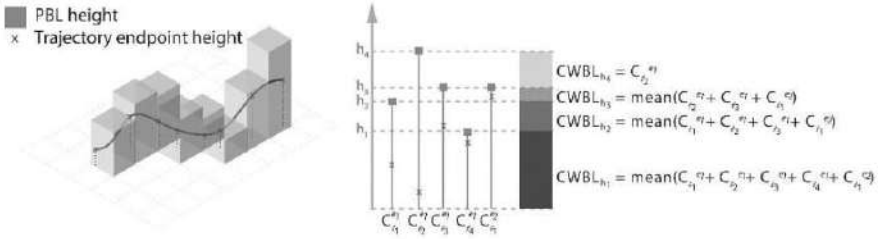


Figure 13. The concept of CWBL hybrid receptor model (left) and CWBL graphical illustration (right) (Stojić and Stanišić Stojić, 2017).

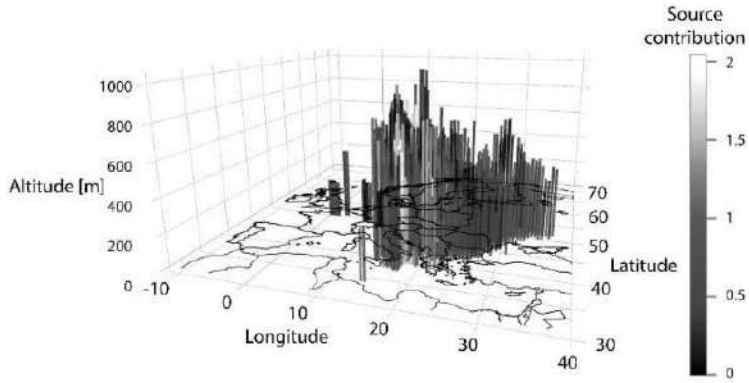


Figure 14. CWBL derived map for VOC industrial emissions (average = 1) representing its altitude distribution above the receptor site and above remote source regions as seen from Belgrade urban area (Stojić and Stanišić Stojić, 2017).

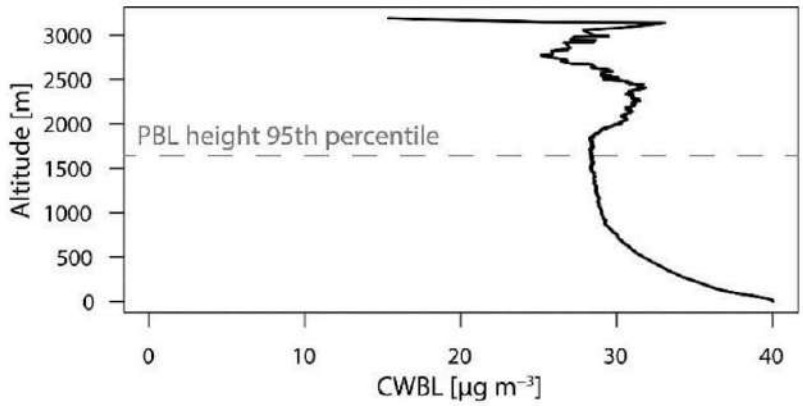


Figure 15. CWBL derived PM_{10} altitude distribution in Belgrade (Stojić et al., 2017).

Moreover, concentrations at greater heights could be representatives of regional background which cannot be measured directly (Wu et al., 2015; Han et al., 2015), so the contribution of the methods, like CWBL, which provide its assessment is of vital importance.

Environmental Multiphase Systems – Rainwater Scavenging

The relevance of rainwater in the air pollution studies is striking since it has a significant role in the spatial and temporal balancing of pollutants as well as their exchange between different environmental compartments through the process of wet deposition. Once the air pollutants are emitted into the atmosphere, they undergo atmospheric transport in particle-bound form or deposit via precipitations (rainwater, snow, fog and/or mist). The rainwater scavenges the pollutants present in the vapor/gaseous, aqueous and particulate phase involving two mechanisms: rainout (incloud scavenging) and washout (below-cloud scavenging) (Kajino and Aikawa, 2015). The rainout involves cloud condensation nuclei activation of aerosols above the cloud base, whereas washout is the collection of aerosols formed by falling hydrometeors, such as liquid precipitation in the form of rainwater. Our research regarding rainwater influences on air pollutants in urban areas focused on VOC, particularly BTEX (Šoštarić et al., 2016, 2017), which are xenobiotics widely recognized for their detrimental effects on human health and urban environment (Marć et al., 2015; Stojić et al., 2015b).

In natural environment, the incorporation of VOC in atmospheric water and their washout is a complex process, which is governed by a variety of factors. According to the available literature, the VOC uptake is primarily dependent on Henry's law constant (K_H) and consequently, on temperature and the "salting in/out" effects that are induced by different constituents of water such as dissolved salts, organic material and acids (Kampf et al., 2013; Kurtén et al., 2015; Okochi et al., 2005; Sander, 2015). The non-negligible aspects are: the VOC susceptibility to photochemical radical- and ozone-reactions as well as aerosol formation in the presence of suspended solids (Shen et al., 2013; Słomińska et al., 2014; Starokozhev et al., 2011); gas-water surface interactions including hydrogen bonding, surface adsorption and van der Waals forces (Furutaka and Ikawa, 2002; Goss, 2004; Roth et al., 2004); physico-chemical properties of the rainwater (Allou et al., 2011); as well as the sources of air masses (Mullaugh et al., 2015). To better understand the fate of BTEX in multiphase systems, our studies discussed BTEX partition and distribution in ultra-pure water and urban-environment rainwater samples by considering surface interactions (Šoštarić et al., 2016), physico-chemical properties of rainwater and meteorological parameters (Šoštarić et al., 2017). For this purpose, in-laboratory dynamic dilution system coupled with Proton Transfer Reaction Mass Spectrometer (PTR-MS), a tool for real-time measurements of VOC with high sensitivity, fast time response, and low detection limit (de Gouw et al., 2003), was used.

The following physico-chemical properties of BTEX were considered to be relevant for investigating the rainwater washout process: K_H, water solubility, octanol-water partitioning coefficient (K_{OW}), ionization potential and van der Waals surface area. Herein we will only elaborate on the parameters which were through pointed out in our experiments providing its brief theoretical background and main findings. In atmospheric chemistry, Henry's law constant is the most common way to describe the distribution of VOC in a multiphase system. There are many variants many variants of Henry's law constants, but all of them essentially represent the proportionality factor, which describes the relationship between the amount of gas dissolved in the aqueous phase and its partial pressure in the

gas phase (Sander, 2015). Therefore, the distribution equilibrium of a compound could be fundamentally characterized with a unit (8) or dimensionless (9) Henry solubility:

$$K_H = \frac{C_a}{p_g} \text{ [M (mol dm}^{-3}\text{) atm}^{-1}\text{]} \tag{8}$$

$$K_H = \frac{C_g}{C_a} \tag{9}$$

where C_a is the compound concentration in the aqueous solution (refers to concentrations of less than 5–50 g L^{-1} for a compound with a molecular weight of 100 g mol^{-1}) and P_g is its partial pressure in the gas mixture (Šoštarić et al., 2017), C_g and C_a are the appropriate molar concentrations in the gas and water phases (Görgényia et al., 2002). These simplified equations could be used when reporting about measurements at constant temperatures, but for the real-world samples, it is important to introduce K_H dependence on the temperature as expressed by the alternative form of the van't Hoff equation (Sander, 2015):

$$K_H T = K_H(298.15) exp\left\{\frac{-\Delta H}{R}\left(\frac{1}{298.15} - \frac{1}{T_r}\right)\right\} \tag{10}$$

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_r is the temperature of real rainwater sample, and R is the universal gas constant (8.314 J K^{-1} mol^{-1}).

In addition to the above-described functionalities, K_H plays an essential role in quantifying the EF of VOC together with the distribution coefficient D_{OBS}, which is defined as the ratio between aqueous concentrations C_a [nM] and gas phase mixing ratios *VMRi* [ppmV, ppbV] (Okochi et al., 2004):

$$EF = \frac{D_{OBS}}{K_H} \tag{11}$$

Our in-laboratory study indicated that D_{OBS} and EF values of BTEX significantly exceeded the value predicted by Henry's law in pure water at 25°C, independently of the pollutant water solubility, volatility and ionization potential (Šoštarić et al., 2016). Furthermore, adsorption is probably the principal phenomenon that explains the mechanisms of BTEX partitioning between the gaseous and the aqueous phase as suggested by highly positive correlation between calculated EF and parameters characterizing interfacial adsorption, such as K_{OW} and van der Waals surface. The presented laboratory experiment served as a step forward to the clarification of the phenomenon in urban-environment samples. Thereby we examined the functional dependency of BTEX enrichment factor on their gaseous concentrations, 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 meteorological parameters (temperature, relative humidity, pressure, wind speed, wind direction, and rainfall intensity/amount) during summer and autumn rainfall events (Šoštarić et al., 2017).

As assumed by Słomińska et al. (2014), BTEX concentrations in rainwater are expected to be low due to their small K_H values. However, in the Belgrade urban area, the BTEX levels, expressed by EF, are significantly above the values theoretically predicted by Henry's law constant. In addition, the BTEX enrichment was prominent during the cold part of the year that corresponds to the less intense photochemical removal and widely triggered emission sources during autumn (Figure 16). The K_H values and consequently EF for benzene, ethylbenzene and xylenes exhibited similar patterns across different altitudes; as the raindrop falls to the ground, EF decrease and *vice versa*. We note that the fluctuations between calculations using the average air and rainwater temperatures were in the range of ±20%, which suggest that the rainwater temperature measured at the ground level can serve as a good indicator of atmospheric conditions under which BTEX undergo changes. The findings support the available observations regarding the distribution of aromatic hydrocarbons

between air and rainwater (Okochi et al., 2004; Sato et al., 2006) as well as between air and urban dew (Okochi et al., 2005).

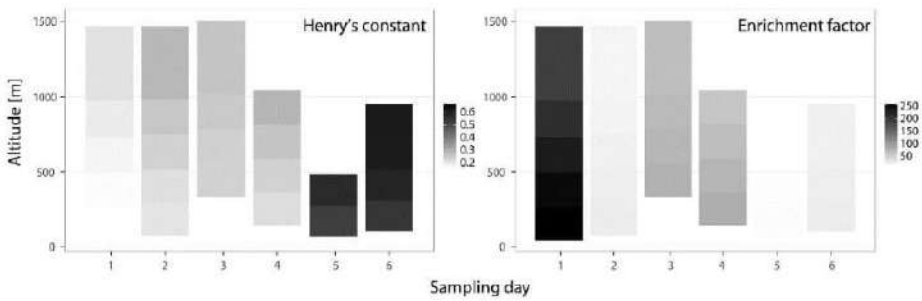


Figure 16. Toluene Henry's constant and EF altitude distributions (Šoštarić et al., 2017).

The enhanced BTEX enrichment in the rainwater samples proved to be not associated with its physicochemical properties since the significant correlation was only shown between the benzene concentrations, and F^- (-0.72) and NH_4^+ (0.83) ions. Furthermore, aromatic compounds in the gaseous form were apportioned to an Unmix-derived factor apart from the other chemical constituents with relatively high shares (99%, 44.5% and 52.2% for BTEX, respectively). Toluene, ethylbenzene and xylenes also exhibited moderate Unmix-shares together with SO_4^{2-} and NO_3^- anions, being recognized as the aerosol constituents. Beside gaseous oxides (SO_2 and NO_2), BTEX susceptibility to photo-oxidation with ozone and OH radicals along with NO_x and SO_2, is well known and leads to the formation of secondary organic aerosol (Shen et al., 2013). As enabled by on-line PTR-MS measurements (please see Šoštarić et al., 2016; 2017 for detailed explanation of the method performances), the extended BTEX exsufflation times for rainwater compared to pure water additionally suggested that not only the physico-chemical properties are affecting BTEX enhanced retention, but also the aerosol fraction have certain impact (Figure 17). The contributions of individual BTEX compounds to the aerosol formation should further be investigated as they behave differently in the atmosphere due to differences in the methyl chain substituent and the alkyl chain

length, e.g., benzene is less prone to the heterogeneous reactions compared toluene and xylenes (Słomińska et al., 2014).

Finally, the interfacial adsorption is assumed to be the major mechanism that governs BTEX washout from the atmosphere, and the process is more efficient for lower gaseous concentration of pollutants. Equal surface available for smaller number of molecules and the prolonged contact time between the two phases when wind-driven raindrops were falling under a certain angle appeared to be the main contributors to such result. More specific, details/examples of environmental factors that synergistically influence spatio-temporal BTEX distribution in the multiphase system, including ambient mixing ratios, physico-chemical properties of rainwater and meteorological data are given below.

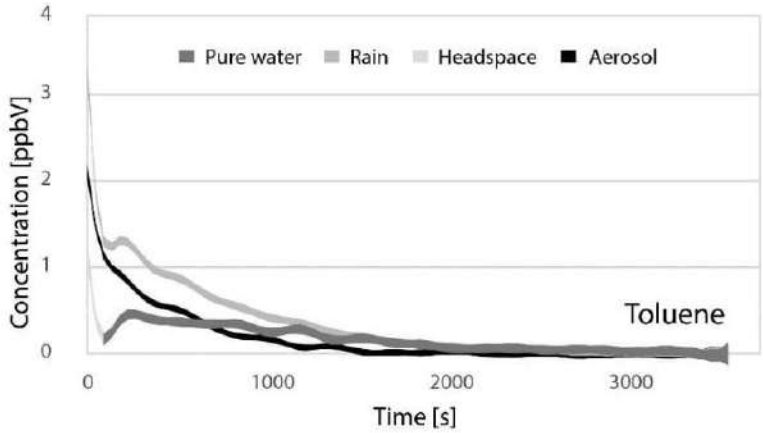


Figure 17. Toluene exsufflation from ultra-pure water and rainwater (Šoštarić et al., 2017).

Environmental Factor Interrelations

Air pollution system is a complex being influenced by multi-interrelations between various components that include numerous sources of air pollutants and the xenobiotics mutual interaction coupled with meteorological regime and atmosphere self-purification. In an open and dissipative system such as atmosphere, understanding the relevant factors

and their correlations on different spatio-temporal scales are the essential task for estimation of adverse health effects and air pollution control policy (Stanišić Stojić et al., 2016b, 2016c, 2016d, 2016e). Therefore, a more meticulous attention in statistical research has been paid to the analysis of combined effects of environmental and anthropogenic factors beginning with factor analysis, principle component analysis, latent class analysis that provides information on the correlation among the investigated variables, etc. However, all these methods cannot assess the interaction of different air pollutants and the corresponding combined effects, particularly for the potential non-linear exposure-response relationship (Tonga et al., 2018).

Recent advances in modeling enabled extracting information relevant for addressing complex air pollution issues (Stojić et al., 2015a; Perišić et al., 2017a; Perišić et al., 2017c; Šoštarić et al., 2017). Machine learning is widely used methodology where non-linear associations between a target variable and a potentially unlimited number of explanatory predictors can be revealed without explicit knowledge of underlying processes by "letting the data speak for itself" (McCabe et al., 2017). In the section below, we presented the way to explore non-linear interconnections between different air pollutants using GRRF method.

Variable Importance

In addition to the usual interpretation of the effects of exposure to a single pollutant, it is crucial to understand the relationships between multi-pollutant mixtures to which people are inevitably exposed (Braun et al., 2016). In order to demonstrate how the complexity of pollutant interrelation can be esplored, we applied correlation and advanced supervised learning algorithms (GRRF) on the dataset comprising PM_{10} concentrations, its chemical constituents (As, Cd, Cr, Mn, Ni, Pb, BaP, Cl^-, NO_3^-, NH_4^+, SO_4^{2-}, Na^+, K^+, Mg^{2+} and Ca^{2+}), and gaseous pollutants (CO, SO_2, NO, NO_2, NO_x, benzene, toluene, o- and m-, p-xylene) (Perišić et al., 2017a). According to the results which refer to the measurement site in the city center, the highest relative importance of As, Cd, BaP, CO and benzene as indicators of PM_{10} concentrations were observed, which is in

accordance with the highest Pearson's correlation coefficients obtained between PM_{10} and its constituents: BaP (0.83), As (0.81) and Cd (0.79), as well as for CO (0.56) and benzene (0.46). The results suggest that environmental burden in urban area is mainly associated with fossil fuel combustion and that the GRRF could be considered as reliable as traditional statistical methods for exploring the potential origin of pollutants. In this analysis, inconsistency between correlation and GRRF analysis was observed for toluene which had higher importance for PM_{10} prediction than nitrogen oxides, but its correlation coefficient was among the lowest (0.25), indicating strong non-linear relationship between the two.

Source Characterization

Characterization of Dominant Sources

Widely applied receptor models Unmix and PMF, used to estimate contribution of different pollution sources in a densely populated central urban zone with heavy and slow traffic (Stojić et al., 2015a; Stojić et al., 2016), start from the assumption that in a complex pollutant mixture, species emitted from the same source are statistically interrelated. Two different datasets were considered: the first consisting of VOC (29 compounds) and five inorganic gaseous pollutants (Stojić et al., 2015a), and the second consisting of PM_{10} concentrations, its chemical constituents, and inorganic gaseous pollutants (Stojić et al., 2016).

The results of Unmix and PMF were in agreement, both showing a six-factor solution as the most feasible. Estimated source contributions were the highest for the profiles attributed to vehicle exhaust and industrial activities, followed by profiles identified as aged plums, solid fuel burning and biomass emission. Traffic-related (TR) profiles were distinguished by the highest loadings of ethylbenzene and xylenes, whereas the presence of gaseous oxides, NOx, NO_2, NO and CO, indicated the association with combustion processes. The commonly used indicator of this source category, toluene to benzene (T/B) ratio higher than 2.2 for traffic-related

profiles was in a good agreement with the previously reported values for vehicle exhaust emissions (Lough et al., 2005). Also, the diurnal pattern of profile contributions reflects the variations in traffic density, with pronounced peaks in the morning and evening rush hours, and decreasing trend at the weekends (Figure 18, up). Bivariate cluster analysis for this source profile resolved dominant portion of locally generated emissions, which is expected in the urban areas with heavy, slow traffic and low ventilation (Figure 18, down).

Beside the identification of traffic emissions, both model solutions suggested gasoline evaporations, oil refinery emissions and petrochemical industry as sources with notable impacts on air quality in the studied urban area. Coupled with bivariate polar plot analysis, model results enabled the identification of the main regional industrial petrochemical complex, located only 13 km from the measurement site, as a potential source of pollutants. Regarding specific pollutants, certain shares of propylene, benzene, toluene, styrene, methacrolein, acetic acid and propionic acid produced by ethylene hydrocarboxylation characterized all profiles. As shown for the traffic emissions, minimum in daily contribution variations of all profiles was registered in morning hours, with small exceptions from the observed pattern in some cases. The minimum was followed by the increase during working hours. Furthermore, the profiles attributed to solid fuel burning was characterized by the highest contribution of SO_2, and low T/B ratios, as expected for biofuel, wood and coal burning (Johansson et al., 2004), while main contributions to aged plums were related to species with longer atmospheric lifetimes (Kwak and Baik, 2012).

According to the Unmix, four source profiles for particulate and inorganic gaseous pollutants were obtained as the most feasible model solution (Figure 19). The profile assigned to *solid fuel burning* with the highest is distinguished by the dominant shares of BaP and As. Occasional wood- and coal-burning for heating purposes during cold season in the vicinity of the receptor site was also reflected in daily variations of the profile contributions. The second profile assigned to *industrial emissions* comprised of significant shares of Cd and Pb originating from industrial activities (Pacyna, 1984). Typically for anthropogenic activities, the

significant weekday/weekend differences in contributions of this profile were also observed. The profile identified as *traffic and particle resuspension* comprised high portions of EC, gaseous oxides, and soot. The seasonal pattern of particle resuspension showed decreased contribution in the cold season due to higher precipitation and snow cover, while the highest contribution was registered in summer, when the soil is dry and loose. High shares of NH_4^+, SO_4^{2-} and NO_3^- ions were apportioned to the *secondary aerosols*, with highest profile contributions during the cold season, when the concentrations of their precursors (NO_x and SO_2) typically reach their annual maximum.

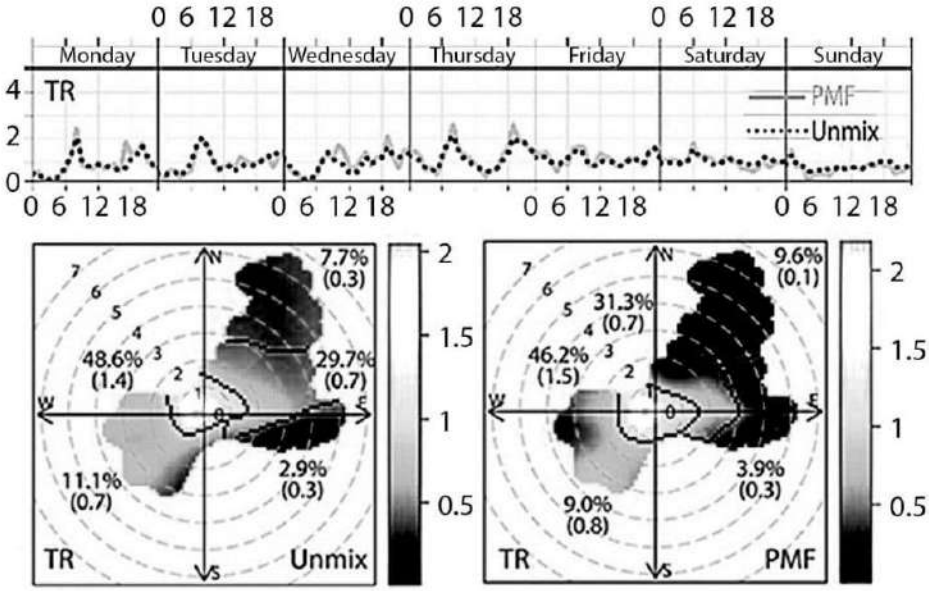


Figure 18. Diurnal dynamics (up) and relationship between source contributions and wind characteristics, together with bivariate cluster plot (percentage and average contribution) for traffic related profiles (Unmix and PMF) (down).

The presented receptor models applied at different databases have a good performance in the reproducing the dominant sources which affect air quality. In both cases, models extracted pollution associated with heating, traffic and transportation, and industrial activities as the main contributors affecting air quality in Belgrade urban area. As can be noted, additional

analysis including bivariate polar plot are required for reliable identification of air pollution sources in complex urban environment.

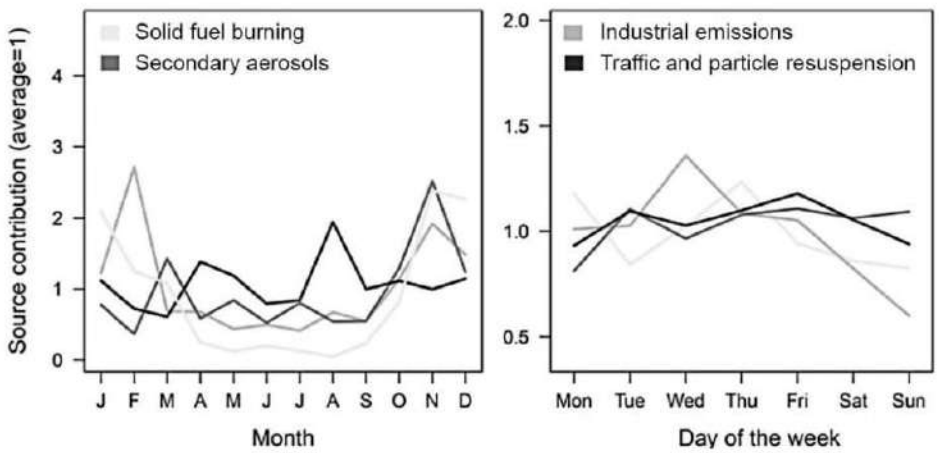


Figure 19. Monthly and weekly variations of Unmix-resolved source contributions.

Characterization of Individual Sources

In order to improve the source identification derived by receptor models and bivariate plot analysis, as well as to distinguish between individual potential emission sources based on the information of pollutant interrelationships, combined bivariate polar plot analysis and pair-wise statistics were applied. For this purpose, we used the dataset comprised of PM_{10} and gaseous pollutants (CO, benzene and toluene) concentrations observed at the sampling site in the city urban zone (Perišić et al., 2017a).

As presented in Figure 20, high correlations of PM_{10}, CO and benzene concentrations ($r \approx 0.7$) during the episodes of NW winds indicate towards several common sources of the pollutants in the vicinity of the sampling site. Moreover, the source composition obtained from slope diagrams of PM_{10} contribution to CO could be associated with various biomass combustion processes (traffic activities, heating plants and individual heating units in Belgrade urban area) (Yokelson et al., 2007).

The polar plot of the slope between PM_{10} and benzene demonstrates a similar surface pattern to the correlation polar plot (Figure 20), with maximum contributions during the periods of low wind speed. This clear

and consistent PM_{10} – benzene ratio of 1:12 in the NW quadrant can be interpreted as a contribution to pollutant concentrations dominated by local traffic sources (Ielpo et al., 2014). In addition to the sources in the vicinity of the sampling site, prominent sources of the PM and toluene are located in SW, S, NE and SE directions ($r > 0.8$, wind speed > 4 m s^{-1}). Unlike sources in S and W, characterized by PM_{10} to toluene ratio of 1:1 which could be assigned to mineral oil and gas refineries, the source located in the NE is characterized by the ratio of 1:6, which indicates the impact of chemical industry, particularly production of basic organic chemicals including aromatic hydrocarbons (EPER, 2006).

Described methodology, which includes commonly used pair-wise statistics and bivariate plots, illustrate how the information on pollutant sources and processes could be enhanced by overlapping pollutant concentrations and wind characteristics. This unequivocally provides a better understanding of the relevant factors and processes in a complex urban environment.

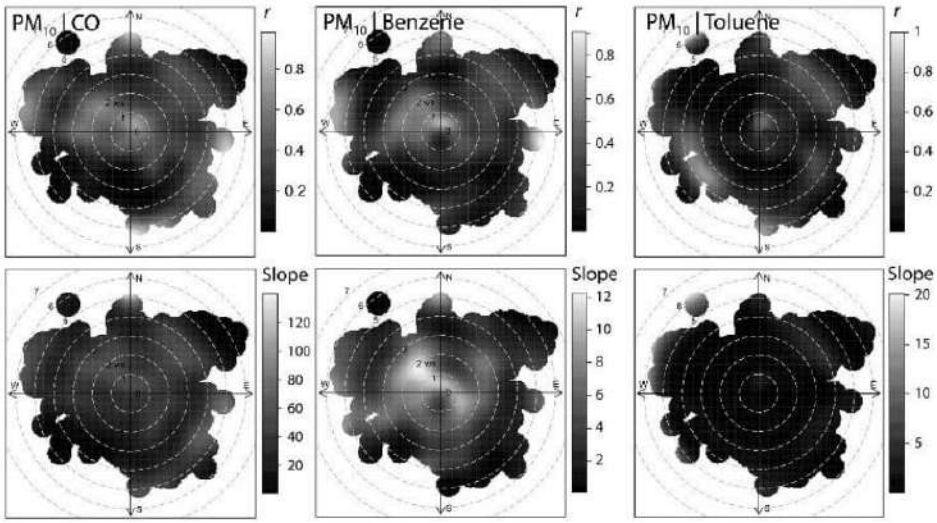


Figure 20. Bivariate polar source apportionment (Perišić et al., 2017a).

Regression Analysis, Prediction and Forecast

Nowadays, the forecasting of air pollutant concentrations in urban areas is of enormous importance given that air pollution has been a concern, particularly present in modern times due to the explosive development of urban infrastructures, industrial activities and population growth. Although forecast by itself cannot reduce the city's air pollution, a reliable and satisfactory forecast is essential as part of an air quality management system with respect to the issues of health alerts and public warnings, and as a supplement to the existing emission control programs and operational planning. Most of the research of air pollution forecasting has been initiated over the last 20 years, and it has dramatically evolved from diagnostics and empirical techniques to the latest advances in forecast resolution, more sophisticated data processing, online and statistical models along with the improvements in deterministic meteorology-chemistry approaches (Ryan, 2016). As a more economical alternative to computationally expensive and time-consuming deterministic models, the parametric or non-parametric statistical approaches to interpreting the underlying specific dependencies between pollutant concentrations and their routine prediction of accidental high concentration episodes have been investigated (Feng et al., 2015). From the numerous statistical methods reported in the literature, such as regression, artificial neural networks, fuzzy algorithms and models, multilayer perception etc., many techniques were successfully applied in forecasting air pollutants including SO_2, NO_2, CO_2, NO, CO, NO_x, $PM_{2.5}$, and PM_{10} (Bai et al., 2018; Coman et al., 2008; Domańska and Wojtylak, 2014; Hrust et al., 2009; Siwek and Osowski, 2012). In regards to the described research initiatives, our studies demonstrated successful application of various machine learning algorithms for forecasting the contributions of industry emissions and vehicle exhausts to VOC (Stojić et al., 2015a; Stanišić Stojić et al., 2016a) and PM_{10} (Perišić et al., 2017b) levels in the urban area relying on meteorological data and the concentrations of inorganic gases.

As noted previously, VOC are hazardous air pollutants that induce chronic toxicity even in small concentrations (Stojić et al. 2015b) and they

commonly originate from both biogenic and anthropogenic emissions that are dominated by petrochemical industry and incomplete fossil fuel oxidation (Stojić et al., 2015c, 2015d). The abundance and spatial distribution of VOCs are mainly driven by their atmospheric lifetimes, emission intensity and meteorological factors (Liu et al., 2012). Part of our VOC-related research aimed at enhancing knowledge and understanding of pollutant variability has been focused on the assessment of the impacts which traffic- and industry-related sources have on VOC levels in urban area (Stojić et al., 2015a). Please see Stojić et al. (2015b) for detailed explanation of the VOC measurements and the basic statistics for selected pollutants, which is beyond the scope of this chapter. We employed several machine learning algorithms for classification and regression to: *i*) differentiate pollution indicators of high and low significance for the source contribution dynamics, and *ii*) forecast source contribution dynamics of the petrochemical/chemical industry (PC) and vehicle exhaust emissions (VE) relying on meteorological parameters and the concentrations of inorganic gaseous pollutants (NO_x, NO, NO_2, SO_2 and CO).

For the prediction of potentially health-damaging events of both PC and VE source contributions, extracted by Unmix and PMF receptor models, two arbitrary limits were considered: the levels above 60% as the increased level of caution and those exceeding 75% as extremely high-alarm triggering values. Among the meteorological variables, for both pollution sources, wind speed and temperature along with pressure appeared to have significant influence on source contributions in the urban canyon avenue (UCA). However, the method performances were not irrespective of input variables and the receptor model type used for the pollution profile selection. For instance, BTD method exhibited the best accuracy for VE-Unmix prediction of higher values based only on meteorological data, but not on the concentrations of the inorganic gases, whereas poorer predictions were observed for the other sources. This variability in method performances suggests that input of additional variables, which were not considered in our study, could possibly improve the method predictions. Raised to the general level, the machine learning

method forecast is more precise when using Unmix-derived source contributions relying exclusively on the input data.

Discussion on the dependency of the source contributions on the examined meteorological data and inorganic gases concentrations, BDT, BDTG, MLP, LD showed the best regression performances (with absolute errors lower than 0.50), particularly when both meteorological data and inorganic gases were used as input variables (Stojić et al., 2015b). Similarly to the classification methods, regression forecasting was sensitive to the receptor model selected to derive pollution sources as well as input variables. In general, forecasting of PC-related contributions was more accurate when both meteorological data and gaseous concentrations were used, whereas higher absolute errors were recorded based on meteorological data exclusively. Nevertheless, in both cases less accurate predictions were still satisfying. In regards to the particular VOC, isoprene, MLP was selected as the best performing method with relative error of 6.94% and the correlation coefficient of 0.99% based on highly correlated VOC (1,3-butadiene, isoprene oxidation product at *m/z*71 and styrene), inorganic gases and meteorological data (Stanišić Stojić et al., 2016a). Contrary to this finding, the other machine learning methods, SMO and LR, exhibited the relative error of only 3% based on the meteorological data whereas the input of inorganic gases and other VOCs resulted in an increase in the relative error.

To conclude, the best regression methods, with relative errors starting at 6%, could provide satisfying forecasting of hourly source contributions of traffic-related VOC emissions as well as of industry-related sources, which are prone to the regional transport. Thereby, although highly dependent on the precise public meteorological forecast, the advanced machine learning methods represent a promising mean for the prediction of the episodes of dangerous pollutant concentrations.

Forecasting PM_{10} Concentrations

In the study of Perišić et al. (2017b), presented herein in brief, we evaluated the performances of machine learning methods for forecasting PM_{10} concentrations and prediction of related health-damaging events

based on 5-year (2011–2015) dataset comprising hourly concentrations of PM_{10}, SO_2 and meteorological data (atmospheric pressure, temperature, humidity, wind speed and direction). Four sampling sites: urban canyon avenue (UCA; urban traffic), urban boulevard (UB; urban traffic), urban industry (UI) and rural industry (RI), were chosen from the regulatory monitoring network of air quality across the city of Belgrade to reflect substantial differences in air quality due to specifics in topography, residential structure and prevalent source of pollution.

Out of the 12 examined regression machine learning methods, BDTG and MLP with absolute and relative errors in the range between 10.6 to 24 $\mu g\ m^{-3}$, and 25.2 to 37.9%, respectively, appeared to be the most satisfying to interpret the relationships between PM_{10} and the SO_2 and meteorological input data. However, the prediction quality was partially affected by the site characteristics such as microclimate conditions and topography as well as by the presence of emission sources. The weakest predictability of PM_{10} levels in urban canyon avenue (UCA) is probably caused by the complexity of urban environment where the wind flow not only facilitates pollutant suspension, but also attenuates its dispersion (Vardoulakis et al., 2003). Moreover, the emission sources in the central city zone are numerous and primarily refer to traffic congestions, as well as to intense combustion processes in local fireboxes and heating plants during autumn and winter seasons. Conversely, a strong influence of a single dominant pollution source, such as an industrial facility, and the regularity of prevalent emissions on a daily, weekly and seasonal basis, led to the most accurate machine learning air quality forecast at the industrial sites, UI and RI. In addition to the regression purposes, the certain methods (BDTG and MLP) enabled reliable classification of the PM_{10} levels which require a high degree of caution (50 $\mu g\ m^{-3}$).

The PM_{10} time series evaluated by the machine learning regression methods correlated very well with the observed concentrations (Figure 21) at all sampling sites (RI, UB, UCA and UI). This result implied that relevant input variables were selected for the forecasting process. However, beside BDTG and MLP, the other methods exhibited the significant PM_{10} forecast errors when being used for regression compared

to classification, at least based on the observed input variables that emphasize the need to investigate more appropriate variables to improve the method forecasting quality.

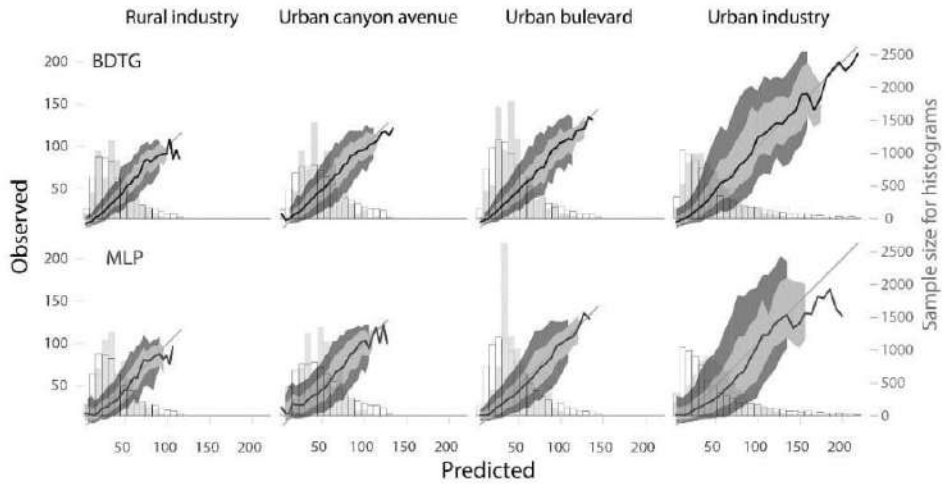


Figure 21. The comparison of the observed and best performing machine learning-predicted PM_{10} mass concentrations [μg m^{-3}] (Perišić et al., 2017b).

Finally, the presented errors for both VOC and PM forecasting in urban area suggest that the pollutant concentrations can be successfully predicted using the machine learning algorithms. What is of the utmost importance, the accurate forecasts could support the air quality management system assisting in health alerts for susceptible categories, operational planning along with reduction of regulatory monitoring expenses.

Enrichment Factor

Figure 22 illustrates an increase of EF for toluene, ethylbenzene and xylenes during summer as a function of the compound ambient gas mixing ratios (Šoštarić et al., 2017). The amplified enrichment might be associated with higher temperature during summer causing a decrease of the surface tension and favoring an interfacial adsorption (Bruant and Conklin, 2002), that has been proposed as the main mechanism for BTEX distribution in the aqueous phase (Šoštarić et al., 2016). More importantly, Figure 22

showed that higher rainwater enrichment by toluene occurred for low gaseous toluene concentrations, which is in accordance with previous studies (e.g., Sato et al., 2006).

As presented in Figure 23, higher enrichment of toluene, ethylbenzene and xylenes was mostly related to a higher wind speed at the sampling site (up to 30 m s^{-1}) and air masses coming from SW area, whereas the lowest rainwater enrichment was observed under relatively stable atmospheric conditions (wind speed<5 m s^{-1}) (Šoštarić et al., 2017). A possible explanation for this phenomenon is the prolonged contact time between the aqueous and the gaseous phases, during the period when strong wind-driven raindrops were falling obliquely. Similarly, bivariate plots revealed that gaseous organic and aerosol Unmix-derived factors were mostly associated with northern wind of moderate speed (<10 m s^{-1}), which clearly reflects their local origin.

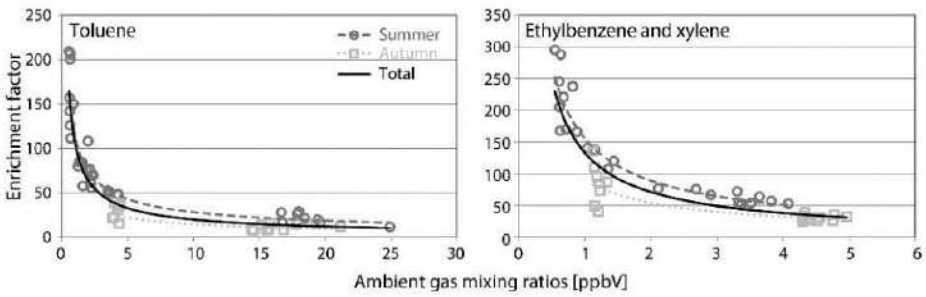


Figure 22. The relationship between T and EX enrichment factors and their gaseous concentrations (Šoštarić et al., 2017).

Feature selection uncovered the rainwater gaseous pollutant concentration and the presence of aerosol constituents to be more important for the prediction of toluene, ethylbenzene and xylenes enrichment than meteorological parameters such ambient temperature, wind speed, pressure and relative humidity. Furthermore, it became evident that BTEX concentrations associated with the ground level-emissions have higher impact on the pollutant partition to the aqueous phase than the polluted air masses coming from the greater atmospheric heights. Finally, out of the 24 examined machine learning regression methods, RF, IBk and

IBkLG provided the most accurate predictions of toluene, ethylbenzene and xylenes enrichment with relative errors of approximately 20% and correlation coefficients around 0.95 and 0.87.

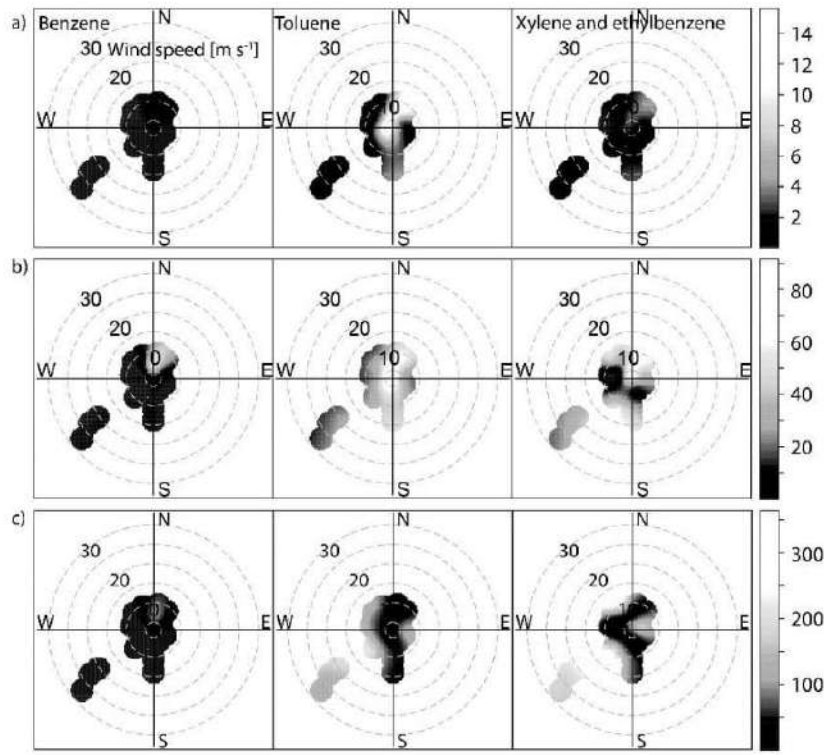


Figure 23. The relationship between BTEX air mixing ratios (ppb) (a), rain concentrations (nM) (b) and enrichment factor and wind characteristics (c) (Šoštarić et al., 2017).

CONCLUSION

Investigation and understanding of the coupled processes that govern the pollutant fate in a dissipative and complex system such as an atmosphere still remain a challenge for contemporary scientific community and environmental policy makers. Despite significant progress having been made, there are deep and possibly insurmountable levels of uncertainty in understanding the interconnectivity of the environmental factors, which

arises from uncertainty in air pollution modeling and understanding the extent of changes in environmental factors. This chapter strives to demonstrate how a joint application of statistical tools could enable a valuable interpretation of the aspects concerning urban air complexity, and enhance the current knowledge drawn with respect to: *i)* spatio-temporal dynamics of pollutants, *ii)* decomposition of background, local and remote sources and its contribution to the pollutant levels, *iii)* three dimensional hybrid receptor modeling for the detection of pollution circulation patterns and its altitude distributions on various spatial scales, *iv)* the mechanisms governing pollutant distribution between the air-water multiphase system, *v)* source apportionment for source characterization across wider geographical area, and *vi)* possibilities for the forecasting and prediction of the pollutant non-linear interrelations. Specifically, air quality management, which aspires to the ambitious goals to face less pollution and lower risks of premature death and other serious health effects, as well as to reduce environmental damages from air pollution could greatly benefit from complementary rather than competitive assembling and hybridization of the advanced methods presented herein. Beside commonly used methods of statistical analysis and source apportionment, the improved transport and multifractal analysis presented herein could reveal specific transport pathways of pollution and support relevant industry sectors to control the pollutant release and distribution. Given the fact that timely information on occurrence of dangerous air pollutant levels is very important for prevention of hazardous health effects and implementation of early warning systems, classification and regression machine learning methods capable of supporting forecasting system of dangerous pollutant episodes depending on the meteorological regime are of high relevance in the contemporary environmental science.

REFERENCES

Aha, D. W., Kibler, D., and Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1), 37-66.

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(17), 2991-2998.

Ashbaugh, L. L., Malm, W. C., and Sadeh, W. Z. (1985). A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmospheric Environment* (1967), 19(8), 1263-1270.

Bai, L., Wang, J., Ma, X., and Lu, H. (2018). Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health*, 15(4), 780.

Braun, J. M., Gennings, C., Hauser, R., and Webster, T. F. (2016). What can epidemiological studies tell us about the impact of chemical mixtures on human health? *Environmental Health Perspectives*, 124(1), A6.

Bravo-Aranda, J. A., de Arruda Moreira, G., Navas-Guzmán, F., Granados-Muñoz, M. J., Guerrero-Rascado, J. L., Pozo-Vázquez, D., ... and Arboledas, L. A. (2017). A new methodology for PBL height estimations based on lidar depolarization measurements: analysis and comparison against MWR and WRF model-based results. *Atmospheric Chemistry and Physics*, 17(11).

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Brereton, C. A., and Johnson, M. R. (2012). Identifying sources of fugitive emissions in industrial facilities using trajectory statistical methods. *Atmospheric Environment*, 51, 46-55.

Bruant, R. G., and Conklin, M. H. (2002). Adsorption of benzene and methyl-substituted benzenes at the vapor/water interface. 3. Finite binary-component VHOC adsorption. *The Journal of Physical Chemistry B*, 106(9), 2232-2239.

Brun, R., and Rademakers, F. (1997). ROOT—an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1-2), 81-86.

Byčenkienė, S., Plauškaitė, K., Dudoitis, V., and Ulevicius, V. (2014). Urban background levels of particle number concentration and sources in Vilnius, Lithuania. *Atmospheric Research*, 143, 279-292.

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

Chehade, W., Weber, M., and Burrows, J. P. (2014). Total ozone trends and variability during 1979–2012 from merged data sets of various satellites. *Atmospheric Chemistry and Physics*, 14(13), 7059-7074.

Cleary, J. G., and Trigg, L. E. (1995). K*: An instance-based learner using an entropic distance measure. In *Machine Learning Proceedings* 1995 (pp. 108-114).

Cohen, W. W. (1995). Fast effective rule induction. In *Machine Learning Proceedings* 1995 (pp. 115-123).

Coman, A., Ionescu, A., and Candau, Y. (2008). Hourly ozone prediction for a 24-h horizon using neural networks. *Environmental Modelling and Software*, 23(12), 1407-1421.

de Gouw, J., Warneke, C., Karl, T., Eerdekens, G., van der Veen, C., and Fall, R. (2003). Sensitivity and specificity of atmospheric trace gas detection by proton-transfer-reaction mass spectrometry. *International Journal of Mass Spectrometry*, 223, 365-382.

Deng, H., and Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition*, 46(12), 3483-3489.

Dimitriou, K., and Kassomenos, P. (2014). Decomposing the profile of PM in two low polluted German cities–mapping of air mass residence time, focusing on potential long range transport impacts. *Environmental Pollution*, 190, 91-100.

Domańska, D., and Wojtylak, M. (2014). Explorative forecasting of air pollution. *Atmospheric Environment*, 92, 19-30.

Dong, Q., Wang, Y., and Li, P. (2017). Multifractal behavior of an air pollutant time series and the relevance to the predictability. *Environmental Pollution*, 222, 444-457.

Draxler, R. R., and Rolph, G. D. (2014). *HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) model access via NOAA ARL*

READY. NOAA Air Resources Laboratory, Silver Spring. https://ready.arl.noaa.gov/HYSPLIT.php.

EPER. Regulation, E. P. (2006). *Regulation (EC) no. 166/2006 of the European Parliament and of the Council concerning the establishment of a European pollutant release and transfer register and amending Council Directives 91/689/EEC and 96/61*.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., and Wang, J. (2015). Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, 107, 118-128.

Frank, E., Hall, M., and Pfahringer, B. (2002, August). Locally weighted naive bayes. In *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence* (pp. 249-256). Morgan Kaufmann Publishers Inc.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I., and Trigg, L. (2005). "Weka." In *Data Mining and Knowledge Discovery Handbook,* edited by Maimon, Oded and Rokach, Lior, 1305-14. Springer US.

Frank, E. (2014). *Fully supervised training of Gaussian radial basis function networks in WEKA*.

Frank, E., Mayo, M., and Kramer, S. (2015, April). Alternating model trees. In *Proceedings of the 30th annual ACM symposium on applied computing* (pp. 871-878). ACM.

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(2), 751-755.

Gan, C. M., Gross, B., Wu, Y., and Moshary, F. (2011). Application of Remote Sensing Instrument in Air Quality Monitoring. In *Air Quality Monitoring, Assessment and Management*. InTech.

Gierałtowski, J., Żebrowski, J. J., and Baranowski, R. (2012). Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia. *Physical Review E*, 85(2), 021915.

Görgényi, M., Dewulf, J., and Van Langenhove, H. (2002). Temperature dependence of Henry's law constant in an extended temperature range. *Chemosphere*, 48(7), 757-762.

Goss, K. U. (2004). The air/surface adsorption equilibrium of organic compounds under ambient conditions. Critical Reviews in *Environmental Science and Technology*, 34(4), 339-389.

Graczyk, M., Lasota, T., and Trawiński, B. (2009, October). Comparative analysis of premises valuation models using KEEL, RapidMiner, and WEKA. In *International conference on computational collective intelligence* (pp. 800-812). Springer, Berlin, Heidelberg.

Grange, S. K., Lewis, A. C., and Carslaw, D. C. (2016). Source apportionment advances using polar plots of bivariate correlation and regression statistics. *Atmospheric Environment*, 145, 128-134.

Grivas, G., Chaloulakou, A., and Kassomenos, P. (2008). An overview of the PM_{10} pollution problem, in the Metropolitan Area of Athens, Greece. Assessment of controlling factors and potential impact of long range transport. *Science of the Total Environment*, 389(1), 165-177.

Han, L., Zhou, W., and Li, W. (2015). Increasing impact of urban fine particles (PM 2.5) on areas surrounding Chinese cities. *Scientific reports, 5, 12467.*

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.

He, H. D. (2017). Multifractal analysis of interactive patterns between meteorological factors and pollutants in urban and rural areas. *Atmospheric Environment*, 149, 47-54.

Henry, R. C. (2003). Multivariate receptor modeling by N-dimensional edge detection. *Chemometrics and intelligent laboratory systems*, 65(2), 179-189.

Hoecker, A., Speckmayer, P., Stelzer, J., Therhaag, J., Von Toerne, E., Voss, H., ... and Dannheim, D. (2015). TMVA users guide—toolkit for multivariate data analysis, *PoSACAT 040.*

Hopke, P. K., Gao, N., and Cheng, M. D. (1993). Combining chemical and meteorological data to infer source areas of airborne pollutants. *Chemometrics and Intelligent Laboratory Systems*, 19(2), 187-199.

Hrust, L., Klaić, Z. B., Križan, J., Antonić, O., and Hercog, P. (2009). Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. *Atmospheric Environment*, 43(35), 5588-5596.

Hsu, Y. K., Holsen, T. M., and Hopke, P. K. (2003). Comparison of hybrid receptor models to locate PCB sources in Chicago. *Atmospheric Environment*, 37(4), 545-562.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Journal of Transportation Engineering. ASCE/American Society of Civil Engineers*, 116, 770-799.

IBkLG. 2015. *Instance Based-kNN Log and Gaussian*. Computer software. Available at https://github.com/sheshas/IBkLG.

Ielpo, P., Paolillo, V., de Gennaro, G., and Dambruoso, P. R. (2014). PM 10 and gaseous pollutants trends from air quality monitoring networks in Bari province: principal component analysis and absolute principal component scores on a two years and half data set. *Chemistry Central Journal*, 8(1), 14.

Ihlen, E. A. F. E. (2012). Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology*, 3, 141.

Johansson, L. S., Leckner, B., Gustavsson, L., Cooper, D., Tullin, C., and Potter, A. (2004). Emission characteristics of modern and old-type residential boilers fired with wood logs and wood pellets. *Atmospheric Environment*, 38(25), 4183-4195.

Kajino, M., and Aikawa, M. (2015). A model validation study of the washout/rainout contribution of sulfate and nitrate in wet deposition compared with precipitation chemistry data in Japan. *Atmospheric Environment*, 117, 124-134.

Kalmegh, S. (2015). Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news. *International Journal of Innovative Research in Science, Engineering and Technology*, 2(2), 438-446.

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(9), 4236-4244.

Kassomenos, P., Karakitsios, S., and Papaloukas, C. (2006). Estimation of daily traffic emissions in a South-European urban agglomeration during a workday. Evaluation of several "what if" scenarios. *Science of the Total Environment*, 370(2-3), 480-490.

Kneen, M. A., and Annegarn, H. J. (1996). Algorithm for fitting XRF, SEM and PIXE X-ray spectra backgrounds. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms,* 109, 209-213.

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(19), 4509-4514.

Kwak, K. H., and Baik, J. J. (2012). A CFD modeling study of the impacts of NOx and VOC emissions on reactive pollutant dispersion in and above a street canyon. *Atmospheric Environment*, 46, 71-80.

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, M., Zhang, Q., Kurokawa, J. I., Woo, J. H., He, K., Lu, Z., ... and Cheng, Y. (2017). MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP. *Atmospheric Chemistry and Physics*, 17(2): 935-63.

Liu, S., Wang, B., He, J., Tang, X., Luo, W., and Wang, C. (2012). Source fingerprints of volatile organic compounds emitted from a municipal solid waste incineration power plant in Guangzhou, China. *Procedia Environmental Sciences*, 12, 106-115.

Lough, G. C., Schauer, J. J., Lonneman, W. A., and Allen, M. K. (2005). Summer and winter nonmethane hydrocarbon emissions from on-road motor vehicles in the Midwestern United States. *Journal of the Air and Waste Management Association*, 55(5), 629-646.

Mandelbrot, B. B., and Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5(5), 967-988.

Mandelbrot, B. B. (1982). *The fractal geometry of nature (Vol. 1)*. New York: WH freeman.

Marć, M., Bielawska, M., Wardencki, W., Namieśnik, J., and Zabiegała, B. (2015). The influence of meteorological conditions and anthropogenic activities on the seasonal fluctuations of BTEX in the urban air of the Hanseatic city of Gdansk, Poland. *Environmental Science and Pollution Research*, 22(15), 11940-11954.

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(7), 3879-3914.

Molino-Minero-Re, E., García-Nocetti, F., and Benítez-Pérez, H. (2015). Application of a time-scale local Hurst exponent analysis to time series. *Digital Signal Processing*, 37, 92-99.

Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., Akimoto, H., ... and Blake, N. (2009). Atmospheric composition change–global and regional air quality. *Atmospheric Environment*, 43(33), 5268-5350.

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.

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(33), 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(26), 4403-4414.

Paatero, P. (1999). The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel

factor analysis model. *Journal of Computational and Graphical Statistics*, 8(4), 854-888.

Pacyna, J. M. (1984). Estimation of the atmospheric emissions of trace elements from anthropogenic sources in Europe. *Atmospheric Environment* (1967), 18(1), 41-50.

Perišić, M., Stojić, A., Stojić, S. S., Šoštarić, A., Mijić, Z., and Rajšić, S. (2015). Estimation of required PM10 emission source reduction on the basis of a 10-year period data. *Air Quality, Atmosphere and Health*, 8(4), 379-389.

Perišić, M., Vuković, G., Mijić, Z., Šoštarić, A., and Stojić, A. (2017). Relative importance of gaseous pollutants and aerosol constituents for identification of PM_{10} sources of variability. Paper presented at *The sixth international WeBIOPATR workshop and conference*, Serbia, Belgrade, September 6-8.

Perišić, M., Maletić, D., Stojić, S. S., Rajšić, S., and 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.

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 and Health*, 10(1), 93-103.*and*

Plocoste, T., Calif, R., and Jacoby-Koaly, S. (2017). Temporal multiscaling characteristics of particulate matter PM_{10} and ground-level ozone O_3 concentrations in Caribbean region. *Atmospheric Environment*, 169, 22-35.

Rasmussen, C. E., and Williams, C. K. (2006). *Gaussian process for machine learning*. MIT press.

Roth, C. M., Goss, K. U., and Schwarzenbach, R. P. (2004). Sorption of diverse organic vapors to snow. *Environmental science and technology*, 38(15), 4078-4084.

Rousseeuw, P. J., and Leroy, A. M. (2005). *Robust regression and outlier detection* (Vol. 589). John wiley and sons.

Ruf, T. (1999). The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. *Biological Rhythm Research*, 30(2), 178-201.

Ryan, W. F. (2016). The air quality forecast rote: Recent changes and future challenges. *Journal of the Air and Waste Management Association*, 66(6), 576-596.

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

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(8), 1231-1233.

Sen, A., Ahammed, Y. N., Banerjee, T., Chatterjee, A., Choudhuri, A. K., Das, T., ... and Mandal, T. K. (2016). Spatial variability in ambient atmospheric fine and coarse mode aerosols over Indo-Gangetic plains, India and adjoining oceans during the onset of summer monsoons, 2014. *Atmospheric Pollution Research*, 7(3), 521-532.

Shen, X., Zhao, Y., Chen, Z., and Huang, D. (2013). Heterogeneous reactions of volatile organic compounds in the atmosphere. *Atmospheric environment*, 68, 297-314.

Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., and Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE transactions on neural networks*, 11(5), 1188-1193.

Shi, P., and Tsai, C. L. (2002). Regression model selection—a residual likelihood approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 237-252.

Siwek, K., and Osowski, S. (2012). Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors. *Engineering Applications of Artificial Intelligence*, 25(6), 1246-1258.

Słomińska, M., Konieczka, P., and Namieśnik, J. (2014). The fate of BTEX compounds in ambient air. *Critical Reviews in Environmental Science and Technology*, 44(5), 455-472.

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.

Stanišić Stojić, S., Stojić, A., and Perišić, M. (2016). Relationship between isoprene, related gaseous pollutants and meteorological factors in an urban area. Paper presented at the *13th International Conference on Fundamental and Applied Aspects of Physical Chemistry*, Serbia, Belgrade, September 26-30.

Stojić, S. S., Stanišić, N., and 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.

Stanišić Stojić, S., Stanišić, N., and 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* edited by Burns, Gerald L. Nova Science Publishers, Hauppauge, New York, USA.

Stanišić Stojić, S., Stanišić, N., and Džamić, V. (2016). Seasonal mortality variations of cardiovascular, respiratory and malignant diseases in the city of Belgrade. *Stanovništvo*, 54(1), 83-104.

Stojić, S. S., Stanišić, N., Stojić, A., and 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.

Starokozhev, E., Sieg, K., Fries, E., and Püttmann, W. (2011). Investigation of partitioning mechanism for volatile organic compounds in a multiphase system. *Chemosphere*, 82(10), 1482-1488.

Stojić, A., Maletić, D., Stojić, S. S., Mijić, Z., and 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.

Stojić, A., Stojić, S. S., Mijić, Z., Šoštarić, A., and Rajšić, S. (2015). Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling. *Atmospheric Environment*, 106, 71-79.

Stojić, A., Stojić, S. S., Šoštarić, A., Ilić, L., Mijić, Z., and 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.

Stojić, A., Stojić, S. S., Mijić, Z., Ilić, L., Tomašević, M., Todorović, M., and Perišić, M. (2015). Comprehensive analysis of VOC emission sources in Belgrade urban area. *Urban and built environments: sustainable development, health implications and challenges*. Nova Science, New York, ISBN, 978-1.

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(11), 10722-10732.

Stojić, A., Stanišić Stojić, S., Perišić, M., and Mijić, Z. (2017). Multiscale multifractal analysis of nonlinearity in particulate matter time series. Paper presented at *The sixth international WeBIOPATR workshop and conference*, Serbia, Belgrade, September 6-8.

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

Stull, R. B. (2012). *An introduction to boundary layer meteorology (Vol. 13)*. Springer Science and Business Media.

Tao, W. K., Wu, D., Lang, S., Chern, J. D., Peters-Lidard, C., Fridlind, A., and Matsui, T. (2016). High-resolution NU-WRF simulations of a deep convective-precipitation system during MC3E: Further improvements and comparisons between Goddard microphysics schemes and observations. *Journal of Geophysical Research: Atmospheres*, 121(3), 1278-1305.

Team, R. C. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2013.

Todorovic, M. N., Perisic, M. D., Kuzmanoski, M. M., Stojic, A. M., Sostarić, A. I., Mijic, Z. R., and Rajsic, S. F. (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.

Tong, Y., Luo, K., Li, R., Pei, L., Li, A., Yang, M., and Xu, Q. (2018). Association between multi-pollutant mixtures pollution and daily cardiovascular mortality: An exploration of exposure-response relationship. *Atmospheric Environment*, 186, 136-143.

UN. 2016. *The World's Cities in 2016 - Data Booklet*. http://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf. Accessed April 2018.

USEPA. Henry, R. C., and Norris, G. A. (2007). *EPA UNMIX 6.0 fundamentals and user guide*. National Exposure Research Laboratory, US Environmental Protection Agency, Research Triangle Park, NC.

Vardoulakis, S., Fisher, B. E., Pericleous, K., and Gonzalez-Flesca, N. (2003). Modelling air quality in street canyons: a review. *Atmospheric environment*, 37(2), 155-182.

Wang, Y., and Witten, I. H. (1996). *Induction of model trees for predicting continuous classes*.

Wang, Y., and Witten, I. H. (2002). Modeling for optimal probability prediction. In *Proceedings of the Nineteenth International Conference on Machine Learning*, Sydney, Australia, July (pp. 650-657).

Wang, Y. Q. (2014). MeteoInfo: GIS software for meteorological data visualization and analysis. *Meteorological Applications*, 21(2), 360-368.

Windsor, H. L., and Toumi, R. (2001). Scaling and persistence of UK pollution. *Atmospheric Environment*, 35(27), 4545-4556.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

WHO. World Health Organization. (2016). *Ambient air pollution: A global assessment of exposure and burden of disease*.

Wu, H., Zhang, Y. F., Han, S. Q., Wu, J. H., Bi, X. H., Shi, G. L., ... and Feng, Y. C. (2015). Vertical characteristics of $PM_{2.5}$ during the heating

season in Tianjin, China. *Science of the Total Environment*, 523, 152-160.

Xue, Y., Pan, W., Lu, W. Z., and He, H. D. (2015). Multifractal nature of particulate matters (PMs) in Hong Kong urban air. *Science of the Total Environment*, 532, 744-751.

Yokelson, R. J., Urbanski, S. P., Atlas, E. L., Toohey, D. W., Alvarado, E. C., Crounse, J. D., Wennberg, P. O., Fisher, M. E., Wold, C. E., Campos, T. L., Adachi, K., Buseck, P. R., and Hao, W. M. (2007). Emissions from forest fires near Mexico City, *Atmospheric Chemistry and Physics* 7:5569-5584.

Zeng, Y., and Hopke, P. K. (1989). A study of the sources of acid precipitation in Ontario, Canada. *Atmospheric Environment* (1967), 23(7), 1499-1509.

Zhang, C., Ni, Z., and Ni, L. (2015). Multifractal detrended cross-correlation analysis between $PM_{2.5}$ and meteorological factors. *Physica A: Statistical Mechanics and its Applications*, 438, 114-123.

Zhao, Y., and Zhang, Y. (2008). Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41(12), 1955-1959.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2), 265-286.

Публикације категорије М20

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 (Sostarić 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)
Shlf	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
[1]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
[2]Acenophthylene [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
[1]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
[2]Acenophthene [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
[1]Anthracene [ng m^{-3}]	0.029	0.043	0.010	0.020	0.008	0.005	0.286	0.281	3.500	16.369	0.003	0.034	0.005	0.005	0.039	0.089	0.01
[2]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
[1]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
[2]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
[1]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
[2]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
[1]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
[2]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
[1]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
[2]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
[1]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	0.196	0.530	1.864	0.01
[2]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
[1]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	0.060	0.141	0.543	1.866	0.01
[2]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
[1]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
[2]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
[2]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	/
[1]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
[2]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
[1]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
[2]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	0.067	0.173	0.609	2.602	0.01
[2]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	/
[1]Cl$^-$ [μg m^{-3}]	1.216	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
[2]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
[1]CO [mg m^{-3}]	0.300	0.089	0.277	0.292	0.066	0.163	0.552	0.389	0.854	0.067	0.010	0.086	0.187	0.244	0.329	0.490	0.1
[2]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
[2]Cnai	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	/
[1]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
[a]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
[b]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
[c]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
[d]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	/
[e]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
[f]Fluoranthene [ng m^{-3}]	0.303	0.289	0.290	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
[g]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
[h]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
[a]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	/
[b]Hours [h]	4.050	3.328	5.075	3.958	4.571	0	9.917	9.917	−0.044	−1.571	0.367	6.779	0	0.050	6.029	8.571	0.02
[c]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
[d]Indeno(1,2,3-cd) pyrene [ng m^{-3}]	0.544	0.593	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
[e]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	/
[f]Lhf [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	54.632	26.739	47.589	102.421	150.402	/
[g]Lib4 [K]	3.491	3.261	2.958	3.290	3.425	−2.279	10.531	12.810	0.469	−0.791	0.379	4.903	−0.629	0.776	5.579	9.467	/
[h]Lsd [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	/
[a]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	/
[b]Mo [ng m^{-3}]	3.415	2.021	3.043	3.111	1.148	1.000	11.816	10.816	1.531	4.271	0.235	1.556	1.000	2.398	3.949	8.076	2
[c]Mn [ng m^{-3}]	3.525	1.479	3.102	3.318	0.845	1.000	9.188	8.188	1.458	2.392	0.172	1.478	2.106	2.580	4.058	6.322	2
[d]Mcld [°]	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	/
[e]Mcld [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	0.280	/
[f]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	/
[a]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
[b]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
[a]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
[b]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
[a]Ni [ng m^{-3}]	7.926	8.196	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
[b]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
[a]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
[b]NO_3^- [µg m^{-3}]	4.423	3.896	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
[a]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
[a]Pblh [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	/
[b]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
[a]People	185.973	206.425	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
[a]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
[b]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
[a]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
[a]$PM_{2.5}$ [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
[b]$PM_{2.5}$ [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
[a]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	997.060	999.670	1
[b]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	/
[a]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}]	0.336	0.313	0.216	0.282	0.190	0.011	1.426	1.415	1.560	1.635	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	56.605	78.674	86.306	/
*Rh [%]	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	49.399	75.273	86.138	0.1
*Rn Bq m^{-3}]	74.525	24.616	67.894	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	/
*SO_2 [μg m^{-3}]	1.849	1.453	1.294	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
*SO_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	4.554	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	96.435	/
*Temp [°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
*Temp [°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	310.856	/
*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

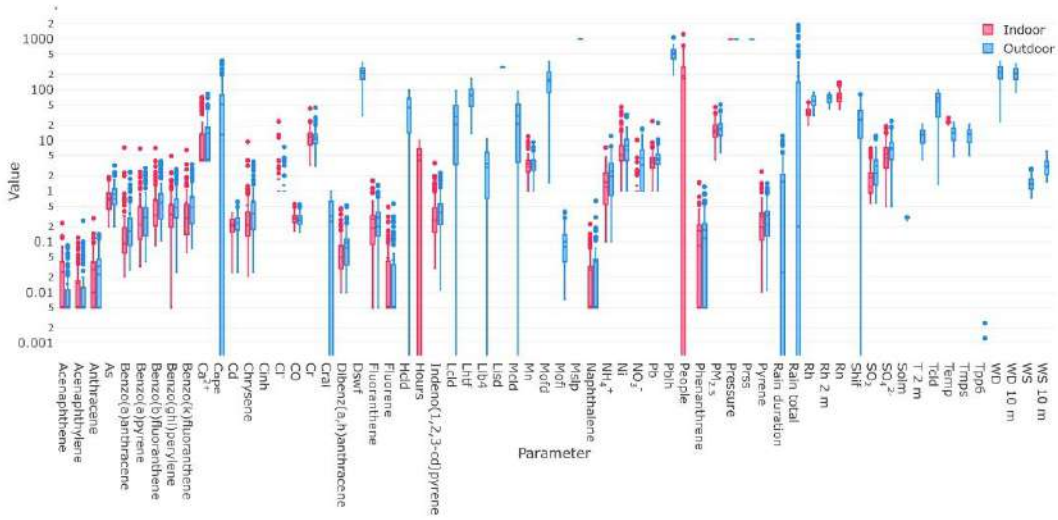


Fig. 1. The box plot of all indoor and outdoor measured parameters.

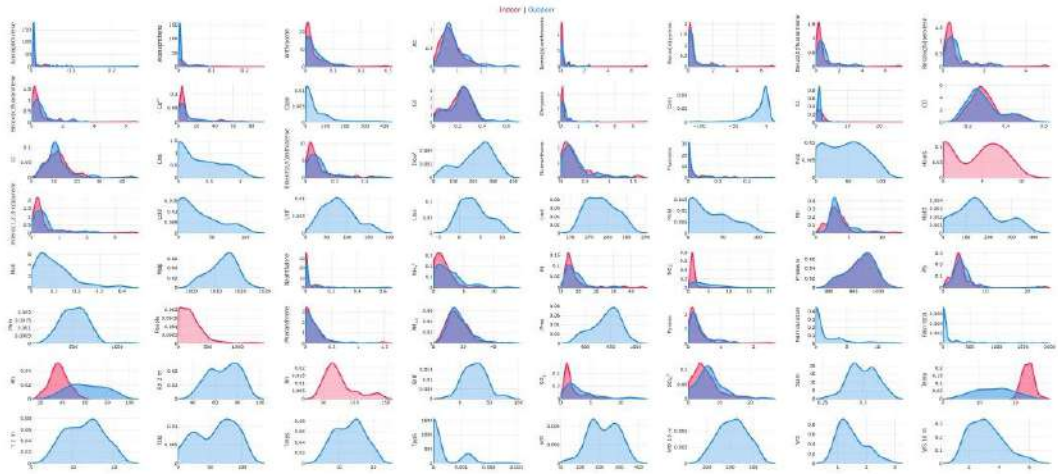


Fig. 2. The probabilty density function for indoor and outdoor measured parameters.

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

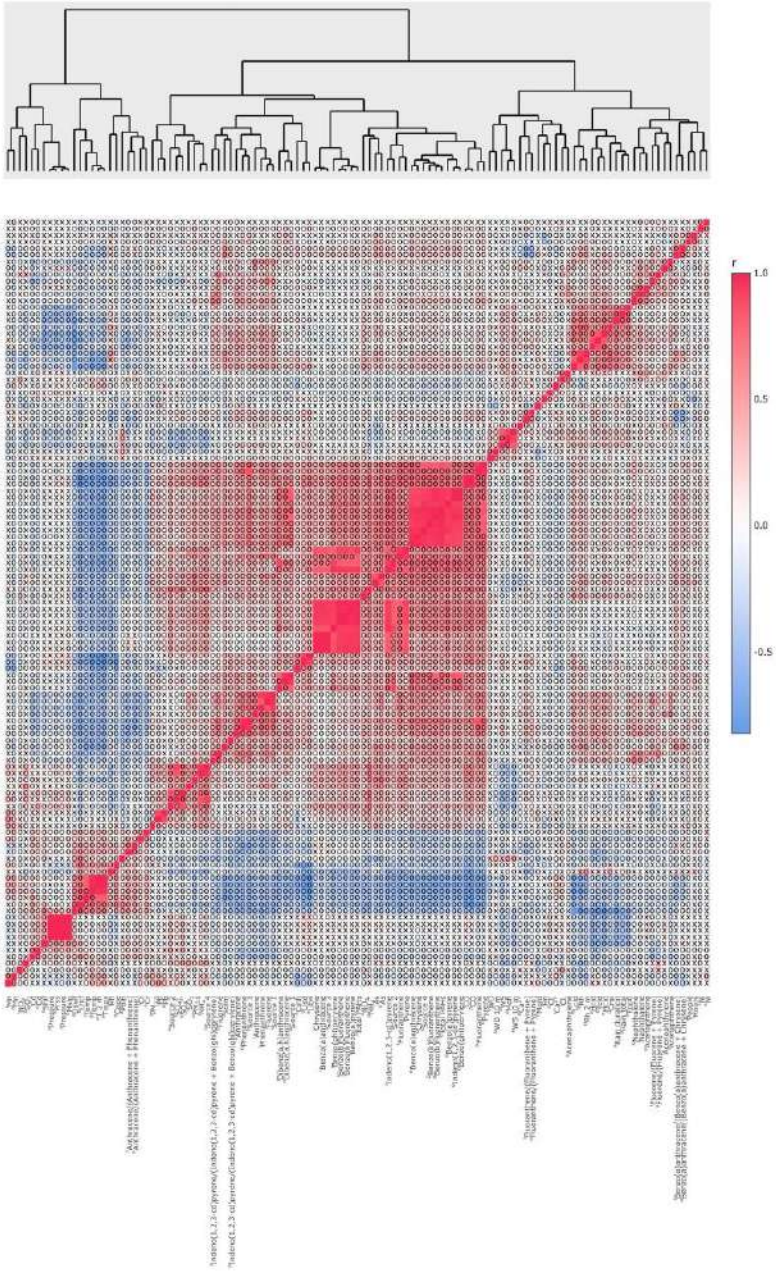


Fig. 3. The correlation matrix for indoor and outdoor measured parameter values.

Table 3
Unmix-resolved source profiles [%].

Species	Indoor				Outdoor			
	Source 1 *PE	Source 2 *CPC	Source 3 *PP	Source 4 *I/TE	Source 1 °CC	Source 2 °TE/PE	Source 3 °PE	Source 4 °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, ° – 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 (°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 (°TE/PE). The contribution of the outdoor profile 3 attributed to petrogenic emissions (°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 (°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	0.146	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	0.896	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	0.060	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 (83–98%) 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., Gonneman, 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., Cui, 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., Strukil, Z.S., Zero, 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.

Šostarić, A., Stojić, S.S., Vuković, G., Mijić, Z., Stojić, A., Grzetić, 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., Cabarkapa, 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.

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

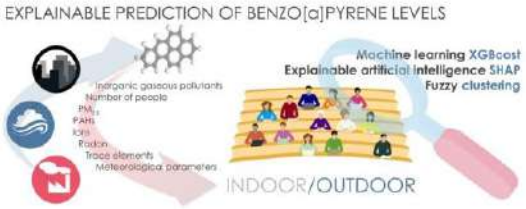
Andreja Stojić [a,b], Gordana Jovanović [a,b], Svetlana Stanišić [b,*], Snježana Herceg Romanić [c], Andrej Šoštarić [d], Vladimir Udovičić [a], Mirjana Perišić [a,b], Tijana Milićević [a]

[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

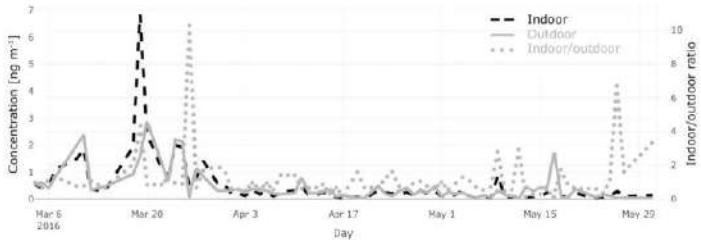


Fig. 1. Benzo(*a*)pyrene concentrations and indoor/outdoor ratio.

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,

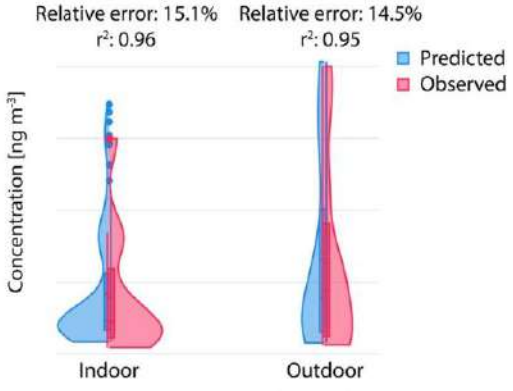


Fig. 2. XGBoost model evaluation.

which are normally distributed in a gas phase.

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%).

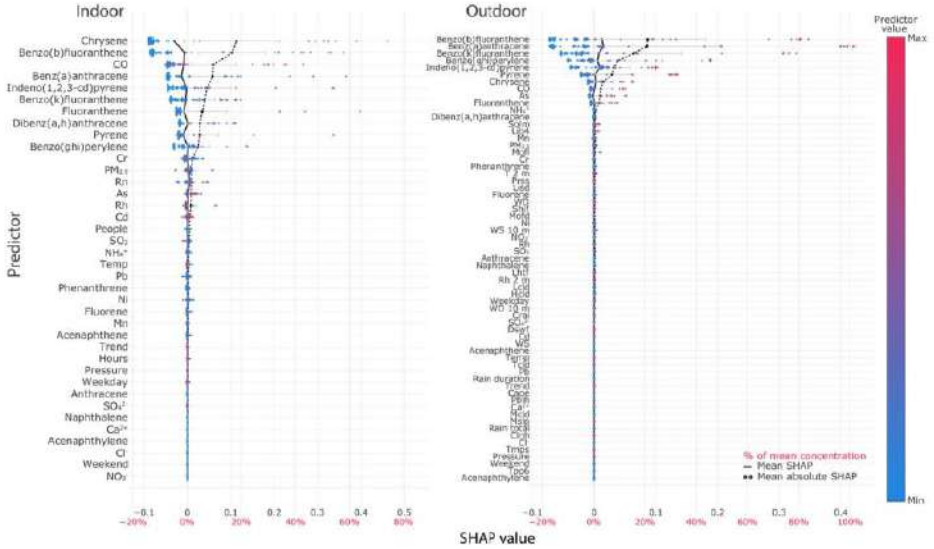


Fig. 3. Indoor (left) and outdoor (right) benzo(*a*)pyrene SHAP summary plots.

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

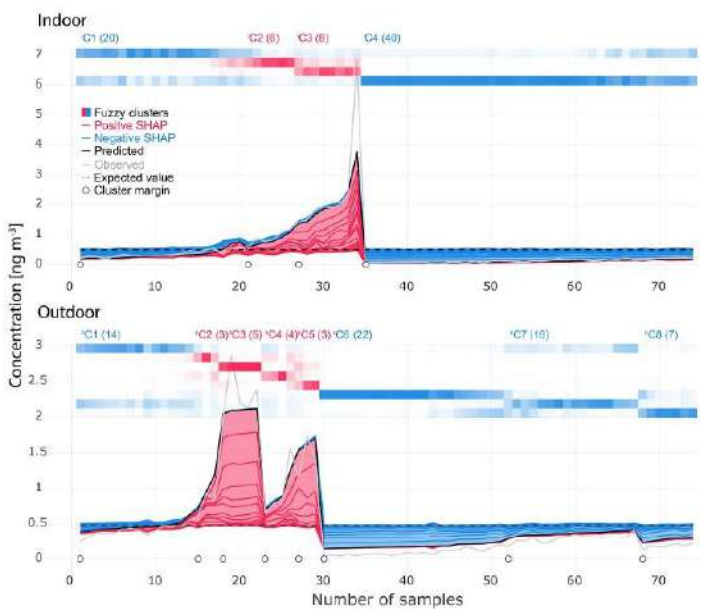


Fig. 4. Indoor (above) and outdoor (below) benzo(*a*)pyrene SHAP force plots.

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

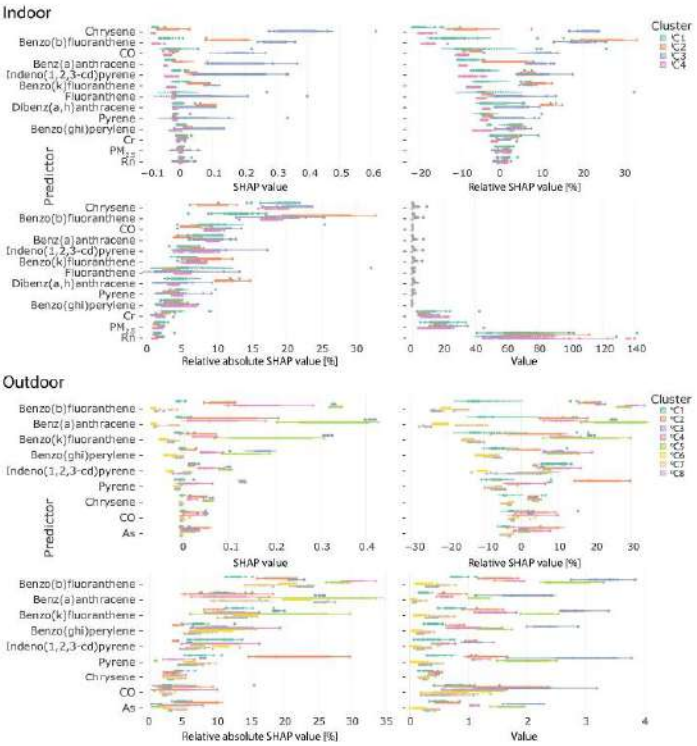


Fig. 5. Indoor (above) and outdoor (below) benzo(*a*)pyrene SHAP force plot cluster statistics.

and CO which is negative for ^{L}C2 negative, and positive for ^{L}C3. We did not observe the cause of differentiation within the features used in this study.

During the days attributed to clusters °C2, °C3, °C4, and °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 °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 °C6 (22 days in total in the period 26th March - 31st May), °C7 (16 days in the period 30th March - 24th May) and °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 °C6 (absolute error 0.09 and relative error 97.8%) and °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 Milicević: 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 (l 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.

Blaszczyk, 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 European 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/ssenl42122012/. (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%201490 2-2005_6294.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/lanconc/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., Kovacić, 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., Garlazzo, 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., Lozzati, 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., Sostarić, 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.

Science of the Total Environment 653 (2019) 140–147

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

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

Andreja Stojić [a,*], Nenad Stanić [b], Gordana Vuković [a], Svetlana Stanišić [b], Mirjana Perišić [a], Andrej Šoštarić [c], Lazar Lazić [d]

[a] Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11000 Belgrade, Serbia
[b] Singidunum University, Danijelova 32, 11000 Belgrade, Serbia
[c] Institute of Public Health Belgrade, Despota Stefana 54, 11000 Belgrade, Serbia
[d] Faculty of Physics, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia

HIGHLIGHTS

- XGBoost and SHAP methods were applied to investigate TEX air – rainwater partition.
- XGBoost relative errors were below 20% when evaluating variable relationships.
- Air TEX concentrations, rainwater and air temperature govern TEX distribution.
- Ion rainwater concentrations and wind speed occasionally impact TEX transfer.

GRAPHICAL ABSTRACT

Figure: FEATURE IMPACT ON RAINWATER TOLUENE

ARTICLE INFO

Article history:
Received 3 September 2018
Received in revised form 25 October 2018
Accepted 27 October 2018
Available online 30 October 2018

Editor: Jianmin Chen

Keywords:
Multiphase system
Wet deposition
BTEX
Machine learning
XGBoost
SHAP

ABSTRACT

Current research suggests that, apart from photochemical reactions, toluene, ethylbenzene and xylene (TEX) removal from ambient air might be affected by atmospheric precipitation, depending on the concentrations and water solubility of the compounds, Henry's law, physico-chemical properties of the water, as well as the frequency and intensity of precipitation events. Nevertheless, existing knowledge of the role that wet deposition plays in biogeochemical cycles of volatile species remains insufficient, and this topic requires more scientific effort to be explored and understood. In this study, we employed the eXtreme Gradient Boosting tree ensemble for revealing TEX transfer from ambient air to rainwater, and applied a novel SHapley Additive exPlanations feature attribution framework to examine the relevance of the monitored parameters and identify key factors that govern wet deposition of TEX. According to the results, main impacts, including ambient air TEX concentrations, and rainwater and air temperatures, and occasional, but less important impacts, including wind speed, air pressure, turbidity, and total organic carbon, NO_3^-, Cl^- and K^+ rainwater concentration, shaped TEX partition between gaseous and aqueous phases during rain events.

1. Introduction

The behavior and distribution of volatile organic compounds (VOCs) in ambient air and atmospheric precipitation has attracted considerable

* Corresponding author.
E-mail address: andreja.stojic@ipb.ac.rs (A. Stojić).

https://doi.org/10.1016/j.scitotenv.2018.10.368
0048-9697/

scientific attention over recent years, due to the abundance and environmental effects of VOCs, as well as their evidenced toxic and carcinogenic nature (Stojić et al., 2015). Theoretically, the partition of volatile species between the gaseous and aqueous phases in the atmosphere might be described by Henry's law and an enrichment factor (EF) (Kampf et al., 2013; Kurtén et al., 2015; Okochi et al., 2005; Sander, 2015). However, some empirical findings suggest the process of wet deposition appears to be more complex and also dependent on other factors, including physico-chemical properties of atmospheric water and surrounding air, as well as on the rainwater dissolved aerosols containing VOCs (Šoštarić et al., 2017). Therefore, the important topic of air pollutant deposition requires more scientific effort to be explored and understood.

Over recent years, the application of artificial intelligence implemented in machine learning has become evident in the field of environmental science, and supported by the great availability of high-dimensional data. Nevertheless, tremendous advances in prediction of air pollutant distribution and behavior in the environment remained somewhat limited and underexploited due to the fact that complex models' performances and interpretability are in apparent conflict (Alvarez-Melis and Jaakkola, 2018a, 2018b). While the best explanation of a simple model is the model formulation itself, understanding and correctly interpreting models parameterized with a large number of variables, such as random forest or an extreme gradient boosting machine (XGBoost), can be challenging (Staniak and Biecek, 2018). Likewise, a single simple decision tree is easy to interpret, but ensembles of hundreds and thousands of trees, which would have superior predictive capabilities, are not interpretable (Tan et al., 2016).

Although some authors consider the tendency to explain complex models with a single point-wise approach too optimistic (Alvarez-Melis and Jaakkola, 2018a, 2018b), simple approximations provide valuable insight into understanding causality, features that drive the model's prediction, the prediction's accuracy and finally, provide opportunities for robust validation procedures and model improvement (Fong and Vedaldi, 2017). In other words, interpretation methods have been developed to shed light on scientific problems where human intuition and domain knowledge are often limited (Montavon et al., 2017). The goal of current interpretation methods which are able to mimic the behavior of machine learning algorithm is not to explain the logical concept behind the black box, but to provide feasible reasons for the choice of a particular instance (Guidotti et al., 2018). As shown by Lundberg et al. (2018), current interpretation frameworks lead to inconsistent results and often contradictory explanations for machine learning algorithms. Therefore, Lundberg and Lee (2017a) recently developed SHAP (Shapley Additive exPlanations), an additive feature attribution method, which they showed has a unique solution in the class of explanation models aimed at post-hoc interpreting machine learning methods, and which is more aligned with human intuition. Unlike approaches that provide a specific global predictor, the SHAP framework provides an explanation of the tree ensemble's overall behavior in the form of particular feature contributions. As well as other methods for interpreting machine learning predictions, SHAP is becoming increasingly popular as a tool in predicting natural and social phenomena. For instance, the recent study of Janizek et al. (2018) combined SHAP and an XGBoost tree-based approach to predict and explain the synergy of novel drug combinations for precision cancer treatment.

Our previous studies were aimed at exploring the contribution of atmospheric precipitation on benzene, toluene, ethylbenzene and xylene (BTEX) removal from ambient air and mechanisms of BTEX partition between aqueous and gaseous phases through field experiments and laboratory simulations (Šoštarić et al., 2016, 2017). In this study, we used the SHAP algorithm to obtain a more detailed insight into the factors that govern toluene, ethylbenzene and xylene (TEX) environmental distribution and transfer from ambient air to rainwater. Beside the accuracy and consistency of the results achieved by SHAP analysis, the present study reveals in which respect the predictors affected the investigated variables.

2. Materials and methods

A detailed description of the sampling and measurement campaign, as well as the obtained dataset used in this study, is given in our previous paper (Šoštarić et al., 2017). In brief, TEX concentrations in air and rainwater were measured simultaneously during several rain events in the summer and autumn of 2015 in a suburban residential area in Belgrade, Serbia, where there were a number of local TEX emission sources, mainly solvent-related or winter-active fireboxes. The measurements were obtained by proton transfer reaction mass spectrometry (Standard PTR-quad-MS, Ionicon Analytic GmbH, Innsbruck, Austria). Beside TEX concentrations, the dataset contained rainwater physico-chemical parameters including major inorganic anions (F^-, Cl^-, SO_4^{2-}, NO_2^- and NO_3^-), dissolved cations (Na^+, NH_4^+, K^+, Ca^{2+} and Mg^{2+}), total organic carbon (TOC), electrical conductivity (EC), UV extinction, turbidity (NTU), pH, rainwater temperature and meteorological parameters (rain intensity, wind speed and direction, pressure, humidity and temperature).

Regression analysis by means of XGBoost was implemented for estimating the relationships between TEX rainwater concentrations and EFs on one hand, and TEX ambient air concentrations, and physico-chemical and meteorological parameters, on the other (Chen and Guestrin, 2016). Gradient boosting is a technique implemented in a complex prediction model by iterative combinations of ensembles of weak prediction models into a single strong learner. Regarding decision trees, gradient boosting builds a sequential series of smaller trees, where each tree tries to complement each other and correct for the residuals in the predictions made by all previous trees (Sheridan et al., 2016). XGBoost is a general-purpose supervised machine learning method achieving high accuracy in a wide range of practical applications, usually outperforming random forests, support vector machines and deep learning neural networks. In this study we used Python (Python Software Foundation) XGBoost implementation (XGBoost Python Package). The dataset was split into stratified 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 values were used for the final model.

The ability to accurately interpret a model's prediction supports deeper understanding of the process being modeled. A widely accepted interpretation approach is the Single tree approximation, which induces a single tree as a comprehensive global predictor that approximates the concept represented by the algorithm, covers the highest possible number of correct training examples, and minimizes the error of the remaining examples. The Decision Rules is another commonly used understandable method for extracting and refining the set of rules, and which capture only the most significant clauses without duplicating information from the trained models. Furthermore, the Feature Importance and Saliency Mask methods are designed to identify the smallest patch which exposes most clearly the central properties of the dataset and removal of which would affect the error of the model significantly (Dabkowski and Gal, 2017). There are also other currently employable interpretation methods, the description of which is beyond the scope of this brief summary.

Nowadays, a variety of statistical tools including tree-based modeling packages can be implemented to compute a measure of feature importance and provide information regarding the features that govern a model's prediction. Unlike the conventional attribution methods, such as those implemented in gradient boosting machines and random forests, which are not individualized for each prediction and because of this, are prone to inconsistency, the SHAP method offers uniquely consistent and locally accurate attribution values (Lundberg et al., 2018). Based on unification and additive attribution algorithms, SHAP values attribute to each feature the change in the expected model prediction

when conditioning on that feature (Lundberg and Lee, 2017a). SHAP overcomes the main drawback of inconsistency, suppressing the possibility of underestimating the importance of a feature with a certain attribution value.

The SHAP runs from exponential to $O(TLD^2)$ for unbalanced trees and $O(TL \log^2 L)$ 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). The idea of the polynomial time algorithm for SHAP values instead of an exponential time algorithm is to recursively keep track of what proportion of all possible subsets flow down into each tree leaf. The exponential reduction in complexity provides alternatives to traditional partial dependence and feature importance plots (Friedman et al., 2001), termed SHAP dependence and SHAP summary plots, respectively. Namely, since they are individualized feature attributions unique to every prediction, SHAP values enable better capture of interaction effects. Unlike partial dependence plots, which represent the dependency of a model on a subset of features with all other features fixed, SHAP dependence plots capture a feature's attributed importance, and changes as the feature's value varies. Moreover, contrary to standard partial dependence plots which only produce lines, SHAP dependence plots capture interaction effects in the model, representing them as vertical dispersion. The combination of SHAP dependence plots and SHAP interaction and SHAP main effect values (representing the impact of a feature after all interaction effects have been removed) can reveal global interaction patterns which could not be identified otherwise. In this paper we used supervised clustering based on SHAP feature attributions, SHAP summary plots and partial SHAP dependency and interaction plots to explore the interaction effects between relevant factors.

3. Results and discussion

The role of different forms of atmospheric water in TEX's environmental fate is still an issue of scientific research. In theory, the capacity of rainwater to scavenge TEX could be determined by a distribution coefficient, Henry's law constant, and the ratio of these two factors (defined as EF, Supplementary material). However, existing studies dealing with wet deposition of volatile species and their transfer from gaseous to aqueous phases are not only scarce, but produced contradictory evidence and conclusions (Okochi et al., 2005; Sato et al., 2006; Allou et al., 2011).

In general, BTEX concentrations in water are expected to be low due to the small Henry's law constant values of aromatic compounds. However, Delzer et al. (1996) were among the first to show that, in spite of their low solubility, toluene and total xylenes were the most frequently detected volatiles in urban stormwater. In the study that compared shallow groundwater BTEX concentrations with the values estimated on the basis of their atmospheric concentrations, assuming Henry's law constant at 15 °C, the authors reported that the observed BTEX concentrations in water samples were higher than expected (Baehr et al., 1999). Furthermore, those authors claimed that BTEX presence in shallow groundwater is directly related to their atmospheric concentrations only at estimated aqueous levels below 0.1 μg L^{-1}. Some years later, Okochi et al. (2004, 2005) also reported the concentrations of volatile species in rainwater and dew exceeded the values predicted by Henry's law. In addition to this, the laboratory study of Sato et al. (2006) demonstrated that dissolved organic compounds, including humic acids, might act as co-solvents, enhancing VOC transfer to atmospheric water. Finally, our recent findings (Šoštarić et al., 2016, 2017) confirmed that levels of BTEX in both deionized water in laboratory simulations and rainwater exceeded theoretically calculated values of BTEX transfer between phases. Moreover, our studies contributed to the existing knowledge by elaborating how meteorological conditions govern BTEX partition, as well as by exploring the relationships between the wet deposition process, the main inorganic constituents of rainwater, and rainwater's physico-chemical properties (pH, EC, NTU, UV extinction and TOC).

In contrast to this, Mullaugh et al. (2015) propose the wet deposition process is of less significance for the removal of aromatic hydrocarbons from the lower layers of the atmosphere due to the high volatility, hydrophobic nature and relatively short atmospheric lifetimes of the compounds. Furthermore, the authors identified that rainfall intensities and amounts were not controlling factors for BTEX concentrations in the rainwater, as the results presented herein have also confirmed.

An XGBoost tree-based method was successfully employed for exploring nonlinear relationships between rainwater TEX concentrations and EFs based on TEX concentrations in ambient air, rainwater physico-chemical properties and meteorological conditions during rain events. As can be seen in Table 1, the predicted/observed calculated relative errors were below 20%, while the correlation coefficients exceeded 0.93. The gain XGBoost F score, as one of the indicators of feature importance (Fig. 1), suggested that the ambient air TEX concentrations appeared to be far the most important predictors of toluene deposition in rainwater. Additionally, XGBoost weight scores showed that TOC and wind direction might also affect toluene rainwater concentrations. In contrast to this, the cover feature attribution score showed that a number of other factors, including wind speed, rain intensity, air and rainwater temperature, TEX ambient air concentrations, and concentrations of rainwater constituents, such as Na^+, SO_4^{2-}, K^+, Ca^{2+} and Cl^- could be influential attributes for rainwater scavenging of toluene.

Similarly, inconsistent results and highly variable XGBoost F scores were registered for ethylbenzene and xylenes (Fig. 1, Supplementary material). As noted by Lundberg et al. (2018), the commonly employed interpretations of XGBoost methods provide contradicting conclusions on the relevance of examined features for prediction of TEX rainwater concentrations and EFs.

According to the results, the mean SHAP value distribution (Figs. 2, S1 and S2, Supplementary material) and supervised clustering analysis (Fig. 3) demonstrated the way ambient air TEX concentrations, together with rainwater and air temperatures, governed the process of TEX wet deposition and rainwater enrichment. These were also impacted by contributions from TOC, wind speed, air pressure, NTU and NO_3^-, NO_2^-, Na^+, Cl^- and K^+ concentrations. As indicated by the long-tailed distribution to the right (Fig. 2), increased toluene concentrations in rainwater were mainly associated with high ambient air TEX concentrations (ranging from 14.4 to 24.9 ng g^{-1} and from 3.4 to 5.0 ng g^{-1} for toluene and ethylbenzene/xylenes, respectively) and NO_3^- concentrations (Fig. 3). The relationship between toluene and nitrates was particularly evident (represented by significant red feature cluster attributions) when the nitrate ion concentrations exceeded 2 mEq L^{-1}, as well as during NW/N wind episodes (Fig. 3), probably being associated with the arrival of polluted air masses which resulted in higher NO_3^- concentrations in rainwater. Moreover, these species could be emitted from the common pollution sources, such as fossil fuel burning and traffic exhaust. Under stable atmospheric conditions with wind speed below 5 m s^{-1}, NO_3^- concentrations dropped below 2 mEq L^{-1}, and the relationship of NO_3^- to toluene transfer to rainwater weakened (as indicated by blue cluster attributions; Fig. 3). On the other hand, ethylbenzene and xylenes transfer to rainwater

Table 1
XGBoost performance.

Variable	Predicted/observed		
	Correlation	Absolute error	Relative error [%]
Rainwater toluene [nM]	0.94	9.73	17
Rainwater ethylbenzene/xylenes [nM]	0.94	5.80	11
Toluene enrichment factor	0.97	7.59	17
Ethylbenzene/xylenes enrichment factor	0.93	16.57	20

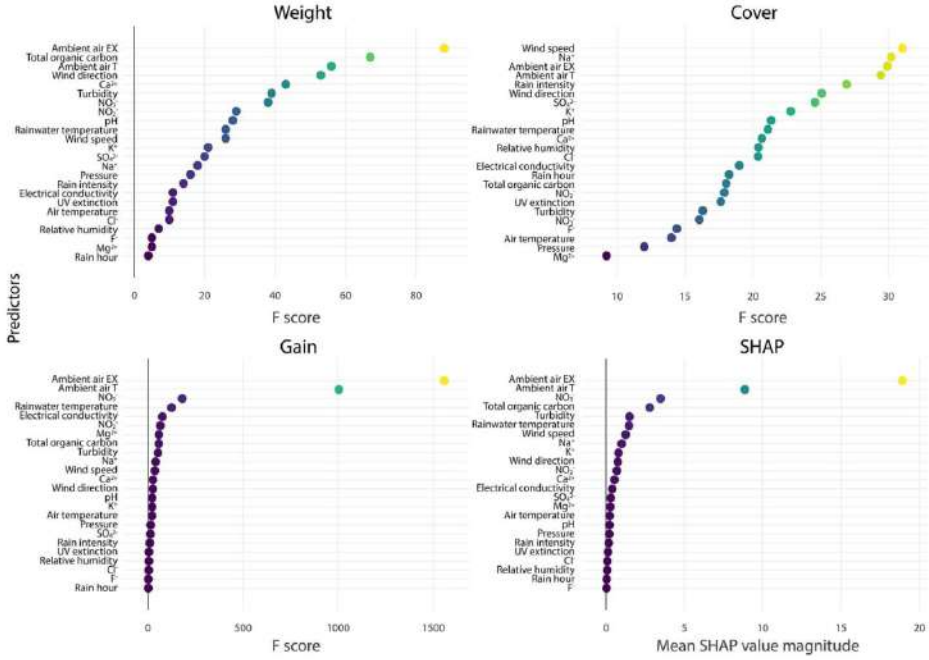


Fig. 1. Importance score comparisons for rainwater toluene concentrations.

was additionally affected by meteorological conditions (rainwater temperature, wind speed and direction, as well as air pressure in a few rainfall events) and TOC (Fig. S2). Unsurprisingly, increased air and rainwater temperatures, which cause the solubility of gases to decrease, as well as high wind speed and air pressure negatively affected ethylbenzene/xylenes transfer to rainwater (Figs. 2 and 3). Namely, high wind speed leads to enhanced dispersion of volatile species and short-lasting contact between gaseous and aqueous phases during rain events.

The fact that other features including Na^+, K^+, Cl^- and NO_2^- concentrations, as well as NTU and UV extinction were arranged in non-gradually decreasing order of importance to the left suggests that they could also occasionally exhibit some impact on TEX rainwater

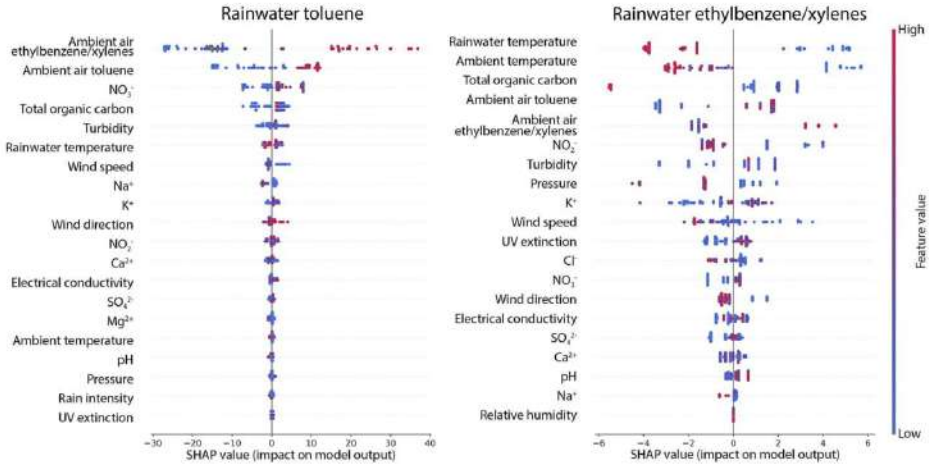


Fig. 2. SHAP summary plots of toluene and ethylbenzene/xylene concentrations in rainwater.

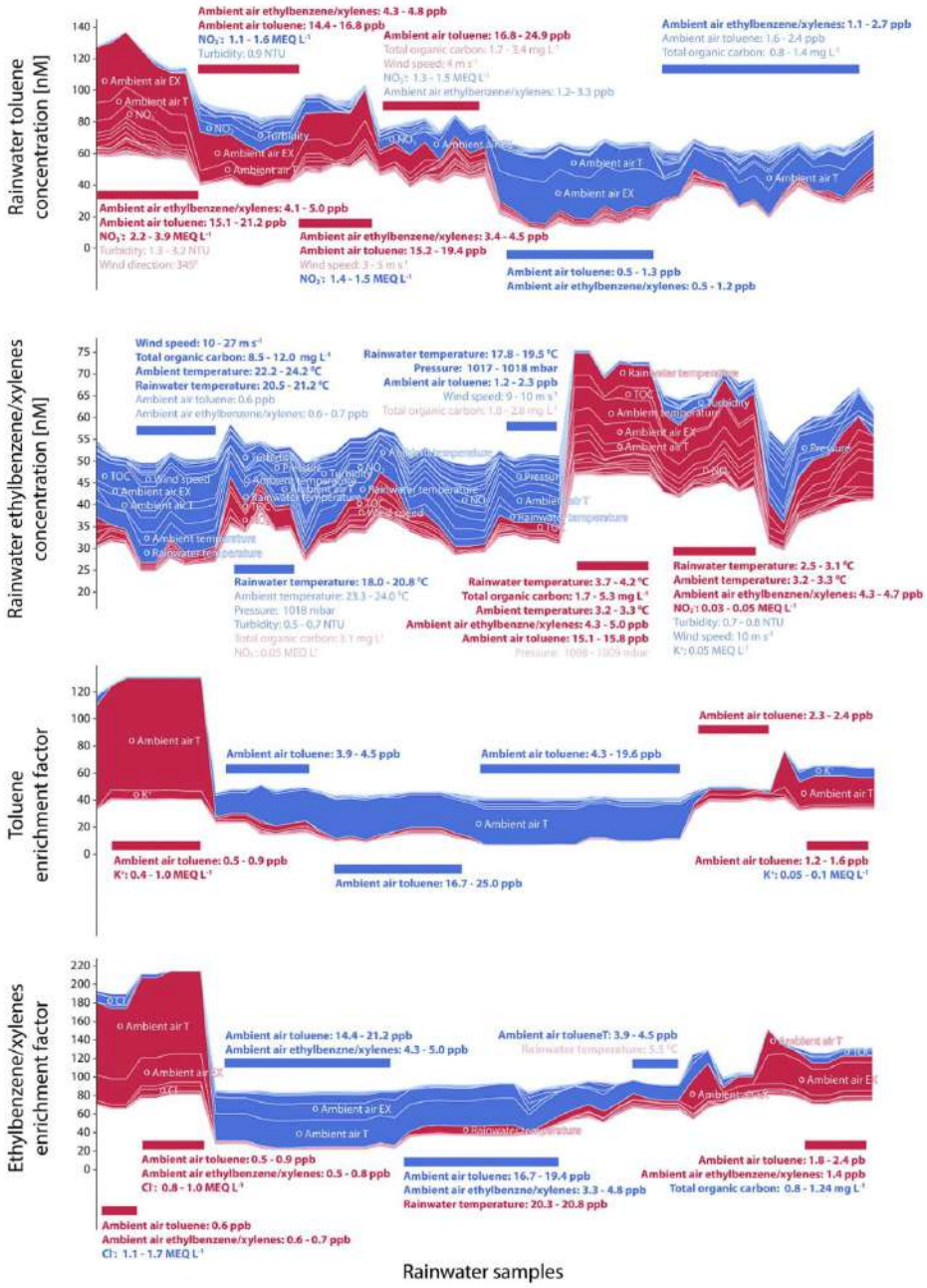


Fig. 3. SHAP supervised clustering: rainwater toluene and ethylbenzene/xylene concentrations and enrichment factors (features present in all the samples in a cluster are represented in bold).

concentrations (Figs. 2 and 3, S4–S40). Theoretically, the impact of Na^+, K^+ and Cl^- presence on gas behavior in aqueous solution can be explained by the salting out effect. This effect refers to the observation that the solubility of gases in a single or mixed electrolyte solution is decreased compared to that in pure water under the same conditions (Sander, 2015). As shown by Allou et al. (2011), Henry's constants for formaldehyde and benzaldehyde in 35 g L^{-1} NaCl salt solution were 27–66% and 12–21% lower, respectively, than the corresponding values in deionized water. While our results mostly comply with this theoretical background, in a few cases, increases of Na^+, K^+ and Cl^- concentrations in rainwater were associated with higher TEX transfer to rainwater, and correspondingly, higher Henry's constant values (Figs. S4–S40).

Diverse properties of the compounds themselves control the distribution of toluene and ethylbenzene/xylenes between environmental compartments. These properties include differences in solubility, volatility and other physicochemical properties, and the fact that ethylbenzene/xylenes are more chemically inert than toluene due to longer alkyl chain and methyl substitutions (Odermatt, 1994). Thereby, the diversity of properties is reflected in both the diversity of identified features that control TEX distribution and SHAP values which are an order of magnitude higher for toluene than for ethylbenzene/xylenes, as can be seen in Fig. 2.

Besides exhibiting an inverse relationship with TEX solubility, rainwater temperature decline is accompanied with less intense molecular vibrations, increasing the tendency of water molecules to form hydrogen bonds, so lowering the concentration of H^+ ions, and therefore, increasing the pH. Furthermore, this basification is associated with lower acid-base dissociation of chemical species that produce SO_4^{2-}, NO_3^- and NO_2^- and which originate from the gaseous oxides SO_2 and NO_x, and/or their acidic products (Wang et al., 2015). In compliance with this, during several rainfall events with neutral to alkaline rainwater (Fig. S5, S6, S25 and S26), TEX transfer to rainwater was favored, particularly under stable atmospheric conditions (wind speed <6 m s^{-1}) when limited wind-induced dispersion of particles and/or gas enabled prolonged contact between gaseous and aqueous phases. It is also worth mentioning that TEX scavenging by rainwater was more pronounced immediately after the rain began, *i.e.* during the first 1–2 h of a rain event (Fig. S6 and S26).

Intense UV extinction can be considered a good indicator of conjugated systems, which can refer to the presence of monoaromatic volatile species. However, the almost negligible associations between the TEX concentrations and UV extinction, as well as TOC, suggest the UV factor could be associated with the occurrence of VOC in rainwater in general, rather than particularly with toluene and/or ethylbenzene/xylenes (Figs. S5, S9, S21, S29). Similar conclusions can be drawn for turbidity and electrical conductivity. For example, the highest TEX rainwater concentrations were associated with lower NTU values (1 to 3), *i.e.* the highest TEX transfer from ambient air to rainwater occurred in cases when rainwater contained negligible amounts of particulate matter that theoretically should enhance TEX transfer (Figs. 3 and S5, S9, S21, S25). Therefore, the observed relationship between NTU, EC and TEX deposition in rainwater is rather indirect, probably caused by the presence of ions and/or suspended particles in rainwater and the aforementioned salting out effect (Figs. S12, S15, S37, S39). What is common to all parameters (TOC, UV, EC and NTU) is that their maximum values were associated with high wind speed (20–30 m s^{-1}) and wind direction below 250°, during the intensive summer rainfall events, probably because intense wind enhances air transport and suspension of organic and inorganic particles in rainwater. Otherwise, in the majority of observations, W, NW and N winds predominated, which corresponded to lower feature values (Figs. S13, S17, S36, S38) and presumably smaller impacts of local pollutant sources that has been confirmed in our previous study (Šoštarić et al., 2017).

As shown by SHAP interaction plots and zero SHAP values (Figs. S18, S40, S72, S73, S76), the TEX deposition process in rainwater is independent of mutual interactions between the physico-chemical parameters of rainwater. On the other hand, the interactions between ambient air toluene and ethylbenzene/xylene concentrations or meteorological conditions were noticeable for toluene transfer to rainwater, although with significantly lower impact on the model output in the SHAP dependence plot, as indicated by the lower SHAP values. The conjoint interdependencies of the examined features (SHAP values: ambient air ethylbenzene/xylenes−NO_3^- (−4 to 6); ambient air ethylbenzene/xylenes−NTU (−3 to 3); ambient air toluene−TOC (−2 to 2); ambient air toluene−wind direction (−0.5 to 1.5); ambient air ethylbenzene/xylenes−EC (−0.5 to 1.5); ambient air ethylbenzene/xylenes−NO_2^- (−1.5 to 1.5); ambient air toluene−ambient air ethylbenzene/xylenes (−3 to 1); ambient air ethylbenzene/xylenes− rainwater temperature (−3 to 1); ambient air toluene−NO_3^- (−1 to 1)) appeared to have a slight/moderate impact on toluene rainwater concentrations, whereas the rest were irrelevant for the toluene wet deposition process (Figs. 4, and S18, S72). We also speculate that the interconnections among the following feature 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 interest for further investigations of ethylbenzene/xylenes transfer to rainwater (Figs. S40, S73).

The results revealed that toluene enrichment factor (EF_T) variations were governed primarily by the dependencies of ambient air toluene concentrations on the other variables (SHAP = 80, Figs. S41–S44), together with the significantly lower impact of rainwater constituents and their associations with K^+ (6), pH (1), NO_2^- (0.10), Cl^- (0.5), TOC (0.4), rainwater ethylbenzene/xylenes (0.4) and toluene (0.2) concentrations, as well as with meteorological parameters: rainwater (2) and ambient air (0.75) temperature, and air pressure (4) (Figs. S45–S53). The conjoint interdependences of the noted variables (SHAP values: ambient air toluene–K^+: up to 6; ambient air toluene−rainwater temperature: up to 1; ambient air toluene−air pressure: up to 1) appeared to have a slight influence on EF_T values, whereas other interrelations could be considered irrelevant due to their zero SHAP values (Figs. S54, S74, S77).

It is important to add that high SHAP vales characterizing mainly lower air and rainwater toluene concentrations were accompanied by increases in TOC, NTU values and concentrations of polar species (Figs. S42 and S43). These polar species are heterogeneous products of fossil fuel burning and traffic emission compounds (K^+ and NO_2^-), but they are also precursors of secondary organic aerosols (NO_2^-). These findings could be expected since K^+ and NO_2^- ions are more soluble than aromatic hydrocarbons, which lack the capability of forming strong hydrogen bonds with polar water molecules. Secondly, TOC and NTU are indicators of particulate matter presence in rainwater, and these solid particles are able to modify gaseous-aqueous TEX partition by creating new solid-liquid interfaces. In these three-phase systems, suspended particles, which can originate from urban aerosols, diesel soot or mineral dust, are able to reduce the gaseous and aqueous levels of toluene and ethylbenzene/xylenes by 10% to 20%, respectively (Starokozhev et al., 2011). In addition, the reduced amounts of TEX in rainwater could be a consequence of heterogeneous reactions of NO_2 which are facilitated by particulate cations, and result in large amount of particulate nitrite (Wang et al., 2015).

The SHAP results, which shape the prediction of ethylbenzene/xylene enrichment factor (EF_{EX}) in rainwater depending on TOC (2), UV extinction (0.4), and NTU (0.3), were complementary to those obtained for EF_T, and illustrate the ethylbenzene and xylene partition in the three-phase system. Regarding the meteorological conditions, the most influential were rainwater temperature and air pressure, whereas

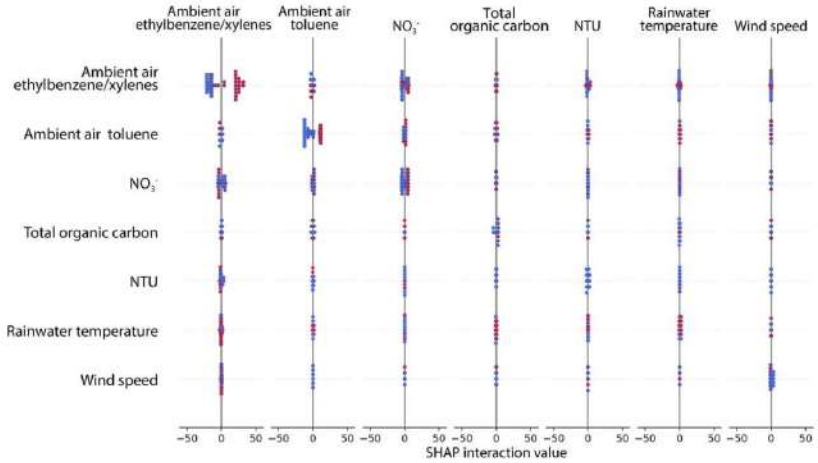


Fig. 4. SHAP interaction summary plot of the most important features for rainwater toluene prediction.

air temperature and wind characteristics were found to be insignificant for describing EF_{EX} in rainwater. As discussed in detail for EF_T, we noted that interrelationships among pressure and the other variables, as well as 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 that would require new measurements and larger datasets (Figs. S71, S75, S78).

4. Conclusions

Elevated ambient concentrations of toluene, ethylbenzene and xylene as a result of anthropogenic activities in urban areas are an issue of scientific and practical concern due to their impacts on both the environment and human health. Due to the fact that advances in machine learning have resulted in numerous applications of complex algorithms for predicting environmental processes and concentrations of pollutant species in environmental samples, it is critical researchers gain insight into the way such algorithms arrive at their predictions. Recently, several methods for interpretable approximations of sophisticated models have been developed, with the focus on producing *a posteriori* explanations and introducing simpler formulations able to capture the input-output behavior and preserve certain key features. In this study, the extreme gradient boosting tree-based method was successfully employed (with relative errors lower than 20%) for predicting TEX concentrations and corresponding enrichment factors in rainwater, based on their concentrations in ambient air and rainwater, 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 enabled insight into the main factors that govern deposition of TEX in rainwater, while overcoming the usually inconsistent and contradictory interpretations derived from commonly applied machine learning algorithms. As the modeling showed, ambient air TEX concentrations, and rainwater and air temperatures were the main features that shaped TEX partition between gaseous and aqueous phases during rain events. 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 proved to be negligible. Moreover, the current knowledge on TEX deposition in rainwater would benefit from further in-depth exploration of the interaction effects between the examined features.

Acknowledgments

This paper was realized as part of projects No III43007, No III41011 and OI176013, which were financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia No III43007, No III41011 and OI176013 for the period 2011-18, and was supported by the Institute of Public Health Belgrade, Serbia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2018.10.368.

References

Allou, L., El Maimouni, L., Le Calvé, S., 2011. Henry's law constant measurements for formaldehyde and benzaldehyde as a function of temperature and water composition. Atmos. Environ. 45 (17), 2991–2998. https://doi.org/10.1016/j.atmosenv.2010.05.044.
Alvarez-Melis, D., Jaakkola, T.S., 2018a. Towards robust interpretability with self-explaining neural networks. arXiv:1806.07538 (arXiv preprint).
Alvarez-Melis, D., Jaakkola, T.S., 2018b. On the robustness of interpretability methods. arXiv:1806.08049 (arXiv preprint).
Baehr, A.L., Stackelberg, P.E., Baker, R.J., 1999. Evaluation of the atmosphere as a source of volatile organic compounds in shallow groundwater. Water Resour. Res. 35 (1), 127–136. https://doi.org/10.1029/1998WR900030.
Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: August (Ed.), Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794 https://doi.org/10.1145/2939672.2939785.
Dabkowski, P., Gal, Y., 2017. Real time image saliency for black box classifiers. Advances in Neural Information Processing Systems , pp. 6967–6976. arXiv:1705.07857v1.
Delzer, G.C., Zogorski, J.S., Lopes, T.J., 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, pp. 96–4145 https://doi.org/10.3133/wri964145.
Fong, R.C., Vedaldi, A., 2017. Interpretable explanations of black boxes by meaningful perturbation. arXiv:1704.03296 (arXiv preprint).
Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Vol. 1 (10). Springer series in statistics, New York, NY, USA 978-0-387-84858-7.
Guidotti, R., Monreale, A., Turini, F., Pedreschi, D., Giannotti, F., 2018. A survey of methods for explaining black box models. arXiv:1802.01933 (arXiv preprint).
Janizek, J.D., Celik, S., Lee, S.I., 2018. Explainable Machine Learning Prediction of Synergistic Drug Combinations for Precision Cancer Medicine. bioRxiv, p. 331769.
Kampf, J.C., Waxman, M.E., Slowik, G.J., Dommen, J., Pfaffenberger, L., Praplan, P.A., Prévôt, S.H.A., Baltensperger, U., Hoffmann, T., Volkamer, R., 2013. Effective Henry's law

partitioning and the salting constant of glyoxal in aerosols containing sulfate. Environ. Sci. Technol. 47, 4236–4244. https://doi.org/10.1021/es400083d.

Kurtén, T., Elm, J., Prisle, L.N., Mikkelsen, V.K., Kampf, J.C., Waxman, M.E., Volkamer, R., 2015. Computational study of the effect of glyoxal-sulfate clustering on the Henry's law coefficient of glyoxal. J. Phys. Chem. A 119, 4509–4514. https://doi.org/10.1021/jp510304c.

Lundberg, S.M., Lee, S.I., 2017a. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, pp. 4765–4774. arXiv:1705.07874v2.

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).

Montavon, G., Samek, W., Müller, K.R., 2017. Methods for interpreting and understanding deep neural networks. Digital Signal Process. 73, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011.

Mullaugh, K.M., Hamilton, J.M., Avery, G.B., Felix, J.D., Mead, R.N., Willey, J.D., Kieber, R.J., 2015. Temporal and spatial variability of trace volatile organic compounds in rainwater. Chemosphere 134, 203–209. https://doi.org/10.1016/j.chemosphere.2015.04.027.

Odermatt, J.R., 1994. Natural chromatographic separation of benzene, toluene, ethylbenzene and xylenes (BTEX compounds) in a gasoline contaminated ground water aquifer. Org. Geochem. 21 (10–11), 1141–1150. https://doi.org/10.1016/0146-6380(94)90076-0.

Okochi, H., Sugimoto, D., Igawa, M., 2004. The enhanced dissolution of some chlorinated hydrocarbons and monocyclic aromatic hydrocarbons in rainwater collected in Yokohama, Japan. Atmos. Environ. 38 (26), 4403–4414. https://doi.org/10.1016/j.atmosenv.2004.03.053.

Okochi, H., Kataniwa, M., Sugimoto, D., Igawa, M., 2005. Enhanced dissolution of volatile organic compounds into urban dew water collected in Yokohama, Japan. Atmos. Environ. 39 (33), 6027–6036. https://doi.org/10.1016/j.atmosenv.2005.05.025.

Python Software Foundation, d. Python language reference, version 3.6. Available at. http://www.python.org, Accessed date: June 2018.

Sander, R., 2015. Compilation of Henry's law constants (version 4.0) for water as solvent. Atmos. Chem. Phys. 15 (8), 4399–4981. https://doi.org/10.5194/acp-15-4399-2015.

Sato, E., Matsumoto, K., Okochi, H., Igawa, M., 2006. Scavenging effect of precipitation on volatile organic compounds in ambient atmosphere. Bull. Chem. Soc. Jpn. 79 (8), 1231–1233. https://doi.org/10.1246/bcsj.79.1231.

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.

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.

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.

Staniak, M., Biecek, P., 2018. Explanations of model predictions with live and breakdown packages. arXiv:1804.01955 (arXiv preprint).

Starokozhev, E., Sieg, K., Fries, E., Püttmann, W., 2011. Investigation of partitioning mechanism for volatile organic compounds in a multiphase system. Chemosphere 82, 1482–1488. https://doi.org/10.1016/j.chemosphere.2010.11.033.

Stojić, A., Stojić, S.S., Šoštarič, A., Ilič, L., Mijić, Z., Rajšič, S., 2015. Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling. Environ. Sci. Pollut. Res. 22 (17), 13137–13152. https://doi.org/10.1007/s11356-015-4540-5.

Tan, H.F., Hooker, G., Wells, M.T., 2016. Tree space prototypes: another look at making tree ensembles interpretable. arXiv:1611.07115 (arXiv preprint).

Wang, L., Wen, L., Xu, C., Chen, J., Wang, X., Yang, L., ... Zhang, Q., 2015. HONO and its potential source particulate nitrite at an urban site in North China during the cold season. Sci. Total Environ. 538, 93–101. https://doi.org/10.1016/j.scitotenv.2015.08.032.

XGBoost Python Package, d. . Available at. https://github.com/dmlc/xgboost/tree/master/python-package, Accessed date: June 2018.

plants

MDPI

Article

Antagonistic Interaction between Phosphinothricin and *Nepeta rtanjensis* Essential Oil Affected Ammonium Metabolism and Antioxidant Defense of Arabidopsis Grown *In Vitro*

Slavica Dmitrović [1,*], Milan Dragićević [1], Jelena Savić [1], Milica Milutinović [1], Suzana Živković [1], Vuk Maksimović [2], Dragana Matekalo [1], Mirjana Perišić [3] and Danijela Mišić [1,*]

[1] Institute for Biological Research "Siniša Stanković"—National Institute of Republic of Serbia, University of Belgrade, Bulevar despota Stefana 142, 11060 Belgrade, Serbia; mdragicevic@ibiss.bg.ac.rs (M.D.); savic.jelena@ibiss.bg.ac.rs (J.S.); milica.milutinovic@ibiss.bg.ac.rs (M.M.); suzy@ibiss.bg.ac.rs (S.Ž.); dragana.bozic@ibiss.bg.ac.rs (D.M.)

[2] Institute for Multidisciplinary Research, University of Belgrade, Kneza Višeslava 1, 11030 Belgrade, Serbia; maxivuk@imsi.rs

[3] Institute of Physics Belgrade—National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia; mirjana@ipb.ac.rs

* Correspondence: smile@ibiss.bg.ac.rs (S.D.); dmisic@ibiss.bg.ac.rs (D.M.); Tel.: +381112078385 (D.M.)

Abstract: Phosphinothricin (PPT) is one of the most widely used herbicides. PTT targets glutamine synthetase (GS) activity in plants, and its phytotoxicity is ascribed to ammonium accumulation and reactive oxygen species bursts, which drives rapid lipid peroxidation of cell membranes. In agricultural fields, PPT is extensively sprayed on plant foliage; however, a portion of the herbicide reaches the soil. According to the present study, PPT absorbed via roots can be phytotoxic to Arabidopsis, inducing more adverse effects in roots than in shoots. Alterations in plant physiology caused by 10 days exposure to herbicide via roots are reflected through growth suppression, reduced chlorophyll content, perturbations in the sugar and organic acid metabolism, modifications in the activities and abundances of GS, catalase, peroxidase, and superoxide dismutase. Antagonistic interaction of *Nepeta rtanjensis* essential oil (*Nr*EO) and PPT, emphasizes the existence of complex control mechanisms at the transcriptional and posttranslational level, which result in the mitigation of PPT-induced ammonium toxicity and in providing more efficient antioxidant defense of plants. Simultaneous application of the two agents in the field cannot be recommended; however, *Nr*EO might be considered as the PPT post-treatment for reducing harmful effects of herbicide residues in the soil on non-target plants.

Keywords: BASTA; phosphinothricin; *Nepeta*; essential oil; glutamine synthetase; ammonium toxicity; antioxidant defense; Arabidopsis; antagonism

check for updates

Citation: Dmitrović, S.; Dragićević, M.; Savić, J.; Milutinović, M.; Živković, S.; Maksimović, V.; Matekalo, D.; Perišić, M.; Mišić, D. Antagonistic Interaction between Phosphinothricin and *Nepeta rtanjensis* Essential Oil Affected Ammonium Metabolism and Antioxidant Defense of Arabidopsis Grown *In Vitro*. *Plants* **2021**, *10*, 142. https://doi.org/10.3390/plants10010142

Received: 21 December 2020
Accepted: 9 January 2021
Published: 12 January 2021

1. Introduction

Herbicides remain the primary tool for implementing weed management to maintain high yields of economically important crops. Phosphinothricin (PPT), also known as glufosinate (commonly used in the form of glufosinate-ammonium), is the major active ingredient in many non-selective herbicide formulations, including BASTA® (BASF SE, Germany; previously Bayer Crop Science AG, Germany). It acts by inhibiting glutamine synthetase (GS) activity in plants [1–5], thus disabling the utilization of ammonium [2,6,7]. Accumulation of ammonium is highly phytotoxic, it provokes growth inhibition, leaf chlorosis and root atrophy [4,8–11], and induces oxidative stress, which is accompanied by the production of reactive oxygen species (ROS) [12,13] and increased activities of antioxidative enzymes [14]. Takano et al. [15] proposed that GS inhibition provokes the disruption of photorespiration and light reactions of photosynthesis, which leads to the photoreduction of molecular oxygen, and subsequent generation of ROS [16]. Actually, the equilibrium between the production and scavenging of ROS is disturbed when GS is inhibited [15].

Takano et al. [15,17] further suggested that the production of ROS, which drives the harmful lipid peroxidation of cell membranes and rapid cell death, is the major cause of rapid PPT toxicity. PPT provokes detrimental effects on the overall plant physiology and biochemistry, causing plant cell death within a few hours of treatment.

Under low concentrations of PPT, its uptake by plant leaves is probably driven by an active transporter, and it is suggested that PPT and glutamine may compete for the same transporter [15]. Under common field conditions, absorption mostly occurs by cell-to-cell diffusion due to high PPT concentrations [15]. After PPT is absorbed, it is translocated by the apoplast in the xylem, which is dependent on the transpiration rate [18]. During its foliar application under field conditions or when the affected leaves fall off on the soil before harvesting, a part of the herbicide reaches the soil. Phosphinothricin does not remain in the soil for long because it is rapidly degraded by the soil microorganisms via oxidation, transamination, and acetylation reactions [16]. According to estimations, the half-life of PPT and its residues varies from 1 to 25 days [16,19–22], depending on the type of soil and environmental conditions. In sterile soils, the half-life of PPT is longer. However, PPT is susceptible to leaching from the soil surface [19], but it is degraded before it reaches the lower soil layers and aquatic ecosystems [23]. As PPT is highly soluble in water it might also be absorbed by plants via roots [24]. The absorption and translocation mechanisms, as well as the mode of action of PTT entering the plant via roots is not well understood, but it could contribute to the overall toxicity of the herbicide.

Mixtures containing PPT and other herbicides offer opportunities to improve the efficacy of PPT. For example, synergistic effects of PPT and 2,4-D or PTT and dicamba are well known [25–27]. However, in some cases, antagonistic effects were described. PPT efficacy was reduced in mixtures with monosodium methyl arsenate [28], or with the essential oil (EO) of *Nepeta rtanjensis* [4], which is an interesting bioherbicide for weeds. This EO, rich in iridoid monoterpenoids nepetalactones, was phytotoxic for *Ambrosia artemisifollia* [29], *Stellaria media* (L.) Vill, *Rumex crispus* L. [30] and *Arabidopsis thaliana* L. Heynh [4]. Joint foliar application of BASTA and *Nepeta rtanjensis* EO (*Nr*EO), resulted in the preservation of GS activity in Arabidopsis along with the maintenance of sub-toxic and/or sub-lethal ammonium concentrations in tissues [4]. In search for more explanations of the described phenomenon, we hypothesized that the PPT present in the Arabidopsis root surrounding also displays herbicidal effects, which can be mitigated by *Nr*EO. We further hypothesized that antagonistic interaction between the two agents involves not only GS activity preservation, but also the perturbations in the antioxidant system of Arabidopsis. In order to test the postulated hypotheses, we prepared an *in vitro* experimental setup that enabled PPT supply through the culture medium, and exposure of Arabidopsis to *Nr*EO volatiles via the atmosphere of culture vessels.

2. Results

The *in vitro* experimental setup enabled the 10-day exposure of Arabidopsis roots to BASTA (B5 and B10), while *Nr*EO (2*Nr*EO and 4*Nr*EO) components were present in the atmosphere within the glass vessels (Figure 1A). We examined the changes induced by BASTA and *Nr*EO in shoots and roots individually, in a dose-dependent manner. Organic volatiles present in the atmosphere of the glass vessels, released from the surface of the filter paper moistened with *Nr*EO (2*Nr*EO and 4*Nr*EO), were quantified using PTR-MS (Figure 1B). The amount of nepetalactones (*cis,trans*- and *trans,cis*-nepetalactone) with $[M+1]^+$ at m/z 167, and of total monoterpenoids with the protonated masses at m/z 137 and m/z 153 were traced. Compounds with $[M+1]^+$ at m/z 205 were also analyzed. All these compounds were previously identified in *Nr*EO by GC–MS and GC–FID analyses [29,31–33]. Concentrations of *Nr*EO (2% and 4%) at the beginning of the experiment resulted in ~1700 and ~3400 ppbV of nepetalactones in the atmosphere of glass vessels, respectively. However, the nepetalactone concentration in the atmosphere of the glass vessels was severely decreased after 10 days: the concentration of nepetalactones in treatments with 2*Nr*EO was around 18 ppbV, while these compounds reached 29 ppbV in treatments with

4*Nr*EO (Figure 1B). As expected, nepetalactone was not detected in the control group of plants nor in the treatments with BASTA alone (Figure 1B). The same trend was observed for other traced compounds. Interestingly, 10 days after the beginning of the experiment, nepetalactones (with the m/z 167) were the second most abundant group of analyzed terpenoids. More abundant were monoterpenoids that exhibited $[M+1]^+$ at m/z 137 (Figure 1B), which might be, at least partially, the result of the accumulation of compounds with the same masses released from the surface of the Arabidopsis leaves.

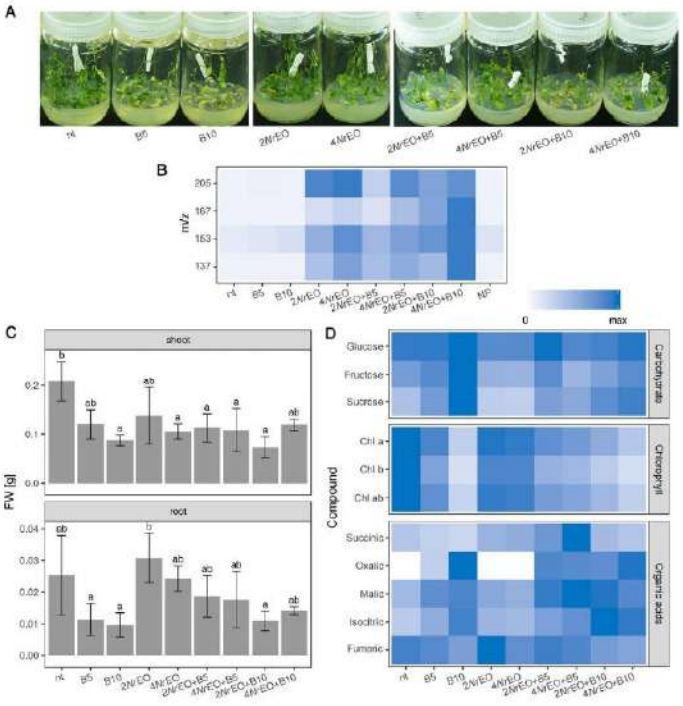


Figure 1. The effects of herbicide BASTA (B5 and B10, with 5 and 10 mg L^{-1} glufosinate ammonium, respectively), and *N. rtanjensis* essential oil volatiles (2*Nr*EO and 4*Nr*EO with 2% and 4% *Nr*EO, respectively) and their combined application on Arabidopsis plants. (**A**) *in vitro* grown Arabidopsis exposed to different combinations of BASTA and/or *Nr*EO, as well as non-treated (nt) plants. BASTA was supplied through culture medium, while *Nr*EO volatiles were present in the atmosphere of the glass jars. (**B**) Values of the relative abundances (ppbV, parts-per-billion-volume) of organic volatiles present in the atmosphere of the glass vessels after 10 days of the treatment, revealed by the PTR-MS measurements, are presented relatively as a heat-map: $[H+1]^-$ at m/z 167—*cis*,*trans*-nepetalactone + *trans*,*cis*-nepetalactone; $[M+1]^-$ at m/z 137 and m/z 153—total monoterpenoids (m/z 137—α- and β-pinene, m/z 153—α-campholenal, neral and geranial); $[H+1]^-$ at m/z 205—total sesquiterpenoids (γ- and δ- cadinene, *cis*- and *trans*-caryophyllene, and α-humulene). (**C**) Arabidopsis shoot fresh weight (FW) and root FW. Values are presented as means ± SD. Significant differences according to Tukey's HSD post-hoc test at $p < 0.05$ are indicated with a compact letter display. (**D**) Content of chlorophylls (Chl a, Chl b, Chl a+b), organic acids (succinic, oxalic, malic, isocitric and fumaric acid) and soluble sugars (glucose, fructose and sucrose) in shoots. Maximal values on the colour scales represent maximal values recorded for each parameter. Abbreviations: nt- non treated plants; NP- culture vessels containing no plants; B- BASTA; *Nr*EO- *Nepeta rtanjensis* essential oil.

2.1. Growth and Metabolism of Arabidopsis as Influenced by BASTA and/or NrEO

BASTA-exposed Arabidopsis was characterized by the highly significant reduced fresh weight (FW) of shoots and roots (Supplementary Materials Table S1), and this decrease was dose-dependent in shoots (Figure 1C). However, FW of Arabidopsis shoots and roots were not significantly affected by *Nr*EO treatments, while the simultaneous application of the two agents (BASTA/*Nr*EO interaction) had a significant effect on shoot FW (Supplementary Materials Table S1). Post-hoc analysis suggested significant growth reduction of Arabidopsis shoots by *Nr*EO, which was also dose-dependent and more pronounced at higher *Nr*EO concentrations (Figure 1C). Simultaneous application of the two agents did not induce additional FW reduction compared to separate treatments with each factor, and in the case of 4*Nr*EO+B10 treatment, a slight but not significant increase in shoot FW was observed compared to separate 4*Nr*EO and B10 treatments (Figure 1C). Chlorophyll (Chl) content was the highest in non-treated Arabidopsis shoots, where the Chl a+b concentration reached ~450 μg g^{-1} FW. Exposure of plants to BASTA via-roots resulted in visible discoloration of *in vitro* grown Arabidopsis shoots (Figure 1A), which was most likely accompanied by chlorophyll (Chl) degradation or its reduced synthesis. The reduction of Chl content was dose-dependent and more pronounced at higher BASTA concentrations, where around 125 μg g^{-1} FW of Chl a+b was recorded (Figure 1D). The observed BASTA effect was statistically significant ($p < 0.001$) (Supplementary Materials Table S1). However, *N. rtanjensis* EO (both 2*Nr*EO and 4*Nr*EO treatments) had no significant effect on Chl content in Arabidopsis shoots (Supplementary Materials Table S1). Simultaneous application of BASTA and *Nr*EO induced a discoloration effect similar to that observed in BASTA treatments. The exception was the treatment with 4*Nr*EO+B10, where plants exhibited a slightly lower reduction in Chl a, Chl b and Chl a+b content, in comparison to those treated with B10 alone. Non-treated Arabidopsis were characterized by the presence of high amounts of isocitric (~3.5 μg mg^{-1} FW) and fumaric acid (~2.9 μg mg^{-1} FW). Malic (~1.9 μg mg^{-1} FW), succinic (~1.6 μg mg^{-1} FW) and oxalic acid (traces) were present in significantly lower amounts. Our results revealed that BASTA treatment altered the amount and ratio of organic acids in Arabidopsis shoots. It induced the increment in the content of oxalic, malic and isocitric acid, while succinic and fumaric acid content in Arabidopsis shoots decreased (Figure 1D). These trends were especially obvious in B10 treatments. However, application of *Nr*EO significantly increased the content of isocitric, malic, and succinic acid. The contents of oxalic, isocitric, malic and fumaric acids in Arabidopsis shoots simultaneously treated with BASTA and *Nr*EO, followed the trends observed in treatments with BASTA alone. Succinic acid content in shoots increased significantly following joint B5 and *Nr*EO (both 2% and 4%) treatments, as compared to B5-treated plants. ANOVA indicated significant effects of BASTA on oxalic, isocitric, malic, succinic ($p < 0.001$) and fumaric acid ($p < 0.05$) content in Arabidopsis shoots, while *Nr*EO significantly affected succinic ($p < 0.001$), malic and isocitric acid ($p < 0.05$) content (Supplementary Materials Table S1). Significant interaction ($p < 0.01$) of the two agents was only recorded for succinic acid (Supplementary Materials Table S1). The main sugars identified in Arabidopsis shoots were sucrose (Suc, ~387 mg 100 mg^{-1} FW), glucose (Glu, ~138 mg 100 mg^{-1} FW) and fructose (Fru, ~126 mg 100 mg^{-1} FW). The content of Suc increased upon treatments with BASTA, in a dose-dependent manner (Figure 1D). At B10 treatments, the Suc concentration reached 911 mg 100 mg^{-1} FW. Nevertheless, application of *Nr*EO resulted in a slight decrease in Fru (~98 mg 100 mg^{-1} FW) and Suc (~300 mg 100 mg^{-1} FW) content in Arabidopsis shoots. As growth was prevented, carbohydrate accumulation in shoots could be attributed to the lack of their utilization. Joint action of BASTA and *Nr*EO induced a slight decrease in Fru and Suc content, when compared to BASTA-treated plants. ANOVA statistical analysis indicated the significant influence of BASTA and *Nr*EO on Fru and Suc content in Arabidopsis shoots, while the concomitant effect of the two agents was not significant (Supplementary Materials Table S1).

2.2. Ammonium Metabolism in Arabidopsis Shoots and Roots as Influenced by BASTA and/or NrEO

The activity of GS was generally higher in Arabidopsis roots than in shoots (Figure 2A). While the application of BASTA significantly reduced GS activity in both shoots and roots, *Nr*EO showed no significant effect. Simultaneous application of BASTA and *Nr*EO slightly mitigated the inhibitory effect of BASTA on GS activity in Arabidopsis roots; but still, this effect was not statistically significant according to Tukey's HSD post-hoc test, except for 2*Nr*EO+B10 treatment. In terms of GS activity, analyses of variance indicated a significant effect of BASTA in both shoots and roots, while *Nr*EO was only effective in roots (Supplementary Materials Table S1). Native-PAGE zymograms of Arabidopsis proteins from shoots and roots (Figure 2B), stained for GS activity, displayed multiple GS activity bands. In shoots, the activities and mobility of individual isoforms varied in response to BASTA and *Nr*EO treatments. In the non-treated plants and in plants treated with *Nr*EO, GS isoforms in shoots with higher mobility were more active in comparison to low-mobility isoforms. Application of B5 and B10, and simultaneous application of BASTA and *Nr*EO increased the activity of some low-mobility isoforms in shoots (Figure 2B). BASTA increased the activity of highly mobile GS isoforms in roots, while *Nr*EO decreased their activity (Figure 2B). Immunoblotting analysis of GS proteins in shoots and roots of *in vitro* grown Arabidopsis (Figure 2C) revealed two distinct GS bands, which were assigned to GS1 (40 kDa) and GS2 (44 kDa) according to literature data [34]. GS1 proteins were proved to be more abundant than GS2 proteins in both shoots and roots. However, GS1 proteins were more abundant in roots, than in shoots. Following BASTA treatments, the amount of GS proteins in shoots and roots increased, and this increment was more pronounced at higher BASTA concentrations (B10) in roots (Figure 2B). The application of *Nr*EO decreased the amount of GS1 proteins in Arabidopsis roots, while in shoots their amount was unchanged compared to already low amounts in the control. Interestingly, simultaneous application of *Nr*EO and BASTA in shoots, led to increased amount of GS1 proteins compared to non-treated plants; however, this increment was lower than in treatments with BASTA only, with the exception for 2*Nr*EO+B10 treatment. The results indicated an obvious suppression of BASTA-induced changes in GS1 abundance. This phenomenon was evident in both shoots and roots. Abundance of GS2 was also slightly increased in roots upon BASTA treatments, but it was not affected when BASTA was applied simultaneously with *Nr*EO (Figure 2C).

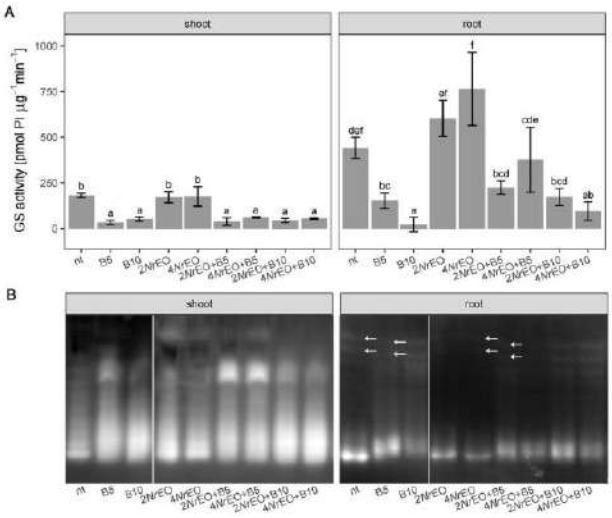


Figure 2. *Cont.*

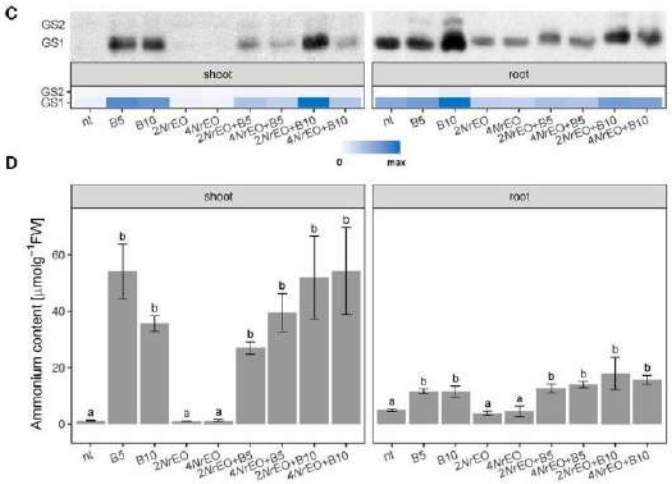


Figure 2. Dose-dependent effects of BASTA (B5 and B10, with 5 and 10 mg L^{-1}, respectively), *N. rtanjensis* essential oil (2*Nr*EO and 4*Nr*EO with 2% and 4% *Nr*EO, respectively), and their mixtures (*Nr*EO + BASTA) on *in vitro* grown Arabidopsis plants. **(A)** Total GS activity in shoots and roots. Values are presented as means ± SD, and significant differences according to Tukey's HSD *post-hoc* test at $p < 0.05$ are indicated with a compact letter display. **(B)** Distribution of GS isoforms activity from shoot and root extracts, stained after native PAGE (50 µg total protein per well). Arrows indicate a mobility shift of GS bands. **(C)** Immunoblots conducted with specific GS antibodies. Heat-maps show relative abundances of GS proteins. Maximal values on the color scales represent maximal values recorded for each immuno blot, independently. **(D)** Ammonium content in shoots and roots. Values are presented as means ± SD. Significant differences according to Tukey's HSD post-hoc test at $p < 0.05$ are indicated with a compact letter display.

The reduction in GS activity in BASTA-treated Arabidopsis shoots and roots were accompanied by a significant increase in ammonium accumulation (Figure 2A,D, Table S1). Volatile organic compounds (VOCs) of *Nr*EO present in the glass vessels showed no effect on ammonium accumulation in shoots and roots, while the BASTA/*Nr*EO interaction was significant ($p < 0.05$) in shoots (Supplementary Materials Table S1). The effect of the interaction can be observed since the simultaneous application of B5 and *Nr*EO resulted in a reduced ammonium accumulation in shoots, when compared to B5-treated Arabidopsis (Figure 2D), but this reduction was not statistically significant according to Tukey's HSD test.

Expression of *GS1* and *GS2* genes in Arabidopsis shoots and roots was analyzed. *in vitro* grown Arabidopsis accumulated different levels of *GS1* and *GS2* gene transcripts, i.e., transcripts of *GS1* genes were more abundant than those of *GS2*, in both *in vitro*-grown shoots and roots (Supplementary Materials Figure S1). Among the *GS1* genes, transcripts of *GLN1;2* were the most abundant in shoots and roots, and were followed by *GLN1;3*, *GLN1;1*, and *GLN1;4*. The expression of plastid *GLN2* isoform was low in both shoots and roots (Supplementary Materials Figure S1).

The expression patterns of all analyzed GS genes in Arabidopsis treated with BASTA, *Nr*EO, *Nr*EO+BASTA, and in the control group of plants, displayed rather opposite trends in shoots and roots (Figure 3). Both BASTA treatments (B5 and B10) induced a slight increase in *GLN1;2*, *GLN1;3*, and *GLN1;4* expressions in shoots, while the expression of *GLN2* was slightly decreased. Treatment with B10 significantly enhanced the expression of *GLN1;1* in shoots. Changes observed for *GS1* genes were dose-dependent, and were particularly pronounced upon B10 treatments (Figure 3). However, BASTA treatment clearly down-regulated all five tested genes in roots (Figure 3), with *GLN1;1*, *GLN1;2* and *GLN1;3* being significantly affected at higher BASTA concentrations (B10). The application of 2*Nr*EO or 4*Nr*EO induced no significant changes in the expression of *GLN1;1*,

GLN1;2, *GLN1;3*, *GLN1;4*, and *GLN2* in shoots of Arabidopsis, while in roots it resulted in slightly decreased expressions of all five genes; these changes were statistically significant for *GLN1;1* and *GLN1;3* at 2*Nr*EO treatment (Figure 3). Simultaneous treatment with BASTA and *Nr*EO followed the trends observed in BASTA treatments for all analyzed genes in Arabidopsis shoots. However, simultaneous application of *Nr*EO (especially of 4*Nr*EO) with BASTA, reduced the inhibitory effect of BASTA on the expression of *GLN1;1*, *GLN1;2*, *GLN1;3*, *GLN1;4* and *GLN2* in roots (Figure 3, Supplementary Materials Table S1). Therefore, BASTA and *Nr*EO interaction was not statistically significant for the expression of five analyzed genes in Arabidopsis shoots, while in roots it significantly affected the expression of all GS genes according to the analysis of variance (Supplementary Materials Table S1).

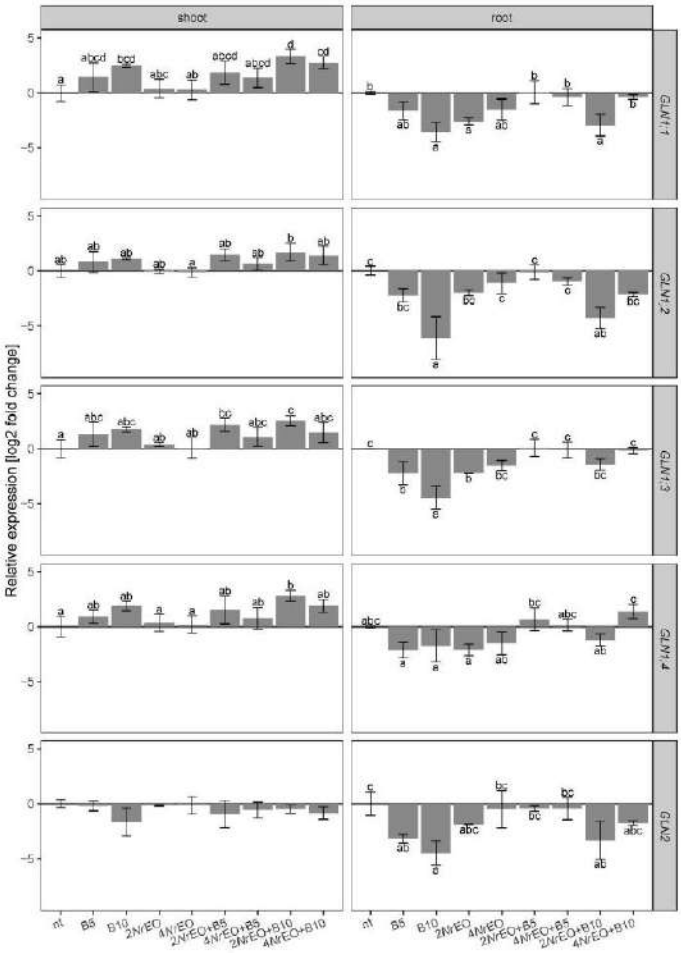


Figure 3. The effects of BASTA applied at two concentrations B5 (5 mg L^{-1}) and B10 (10 mg L^{-1}), *N. rtanjensis* essential oil applied at two concentrations 2% (2*Nr*EO) and 4% (4*Nr*EO), and of their combinations on the relative expression of GS coding genes in Arabidopsis shoots and roots, measured 10 days after the beginning of the treatment, grown *in vitro*. Values are presented as means ± SD. Letters above the bars denote significant differences according to Tukey's HSD post-hoc test at $p < 0.05$. The expression of *GLN1;5* was not detected.

2.3. Antioxidant Enzymes Activity in Arabidopsis Shoots and Roots as Influenced by BASTA and/or NrEO

The activities and abundance of antioxidant enzymes, catalase (CAT), peroxidase (POX) and superoxide dismutase (SOD), were used as biochemical markers of the stress resistance capacity of *in vitro* grown Arabidopsis to BASTA (B5 and B10) and/or *Nepeta rtanjensis* essential oil (2*Nr*EO and 4*Nr*EO).

Three CAT isoforms were detected in Arabidopsis shoot and root samples (Figure 4A,B). A strong decrease of CAT activity was recorded in shoots and roots of Arabidopsis treated with *Nr*EO (2*Nr*EO and 4*Nr*EO) for 10 days, when compared with non-treated plants (nt) (Figure 4A,B). Treatment with *Nr*EO reduced the activity of all three CAT isoforms, and this effect was more pronounced in shoots (Figure 4A) than in roots (Figure 4B). BASTA treatments (B5 and B10) only induced a slight decrease in CAT activity in shoots (Figure 4A). In the roots the activity of CAT was increased (Figure 4B), probably because of the induced activity of the low mobility CAT1 isoform, which is the major contributor to the overall activity of this enzyme. Joint application of BASTA and *Nr*EO resulted in the decline of CAT activity in shoots, with the exception of 4*Nr*EO+B5 treatment, which induced no changes. Increase in the activity of high mobility CAT3 isoform and decrease in the activity of CAT1 were recorded on B10 treatments with 2*Nr*EO and 4NrEO (Figure 4A), when compared to the B10 treatment. In roots, both *Nr*EO+B10 treatments increased the CAT activity, especially of CAT1 isoform, when compared to the non-treated plants (Figure 4B). Changes in the activities of CATs in Arabidopsis shoots and roots influenced by BASTA and/or *Nr*EO treatments, are followed by changes in the abundance of this enzyme in plant organs (Figure 4C,D). Treatments with BASTA increased the abundance of CATs in roots, while *Nr*EO decreased the abundance of this enzyme in both organs. Abundance of CAT in Arabidopsis shoots and roots exposed to *Nr*EO+B5 action was decreased when compared to B5-treated plants. In roots, on treatments with B10 and 2*Nr*EO or 4*Nr*EO, the abundance of enzyme was increased, when compared to B10 treatment (Figure 4D).

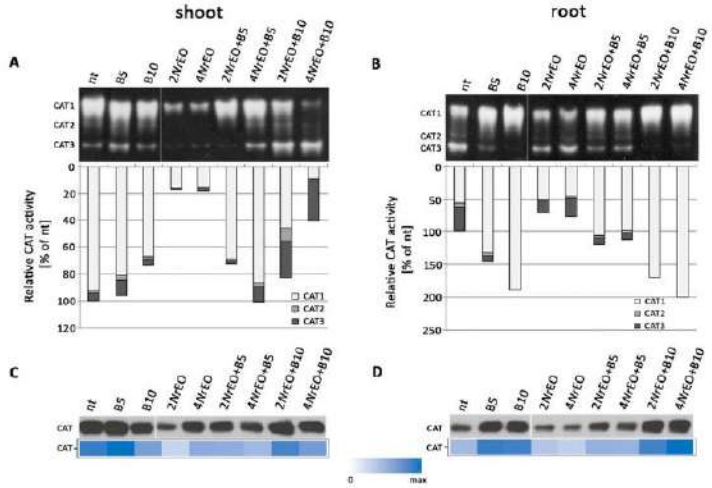


Figure 4. Catalases (CAT) activity in Arabidopsis grown *in vitro*, measured 10 days after treatment with BASTA applied at two concentrations B5 (5 mg L^{-1}) and B10 (10 mg L^{-1}), *N. rtanjensis* essential oil applied at two concentrations 2% (2*Nr*EO) and 4% (4*Nr*EO), and of their combinations. (**A**,**B**) CAT activity stain after native-PAGE separation (5 µg per lane of total protein), in shoots and roots, respectively. The detected activities were measured densitometrically and presented as relative CAT activity (% of control) in comparison to non-treated plants (nt). (**C**,**D**) CAT immunoblot (20 µg of soluble proteins per line) in shoots and roots, respectively. Heat-maps show relative abundances of CAT proteins. Maximal values on the color scales represent maximal values recorded for each immunoblot.

Native-PAGE gels stained for POX activity, after 10 days of BASTA and *Nr*EO treatments, revealed changes in Arabidopsis shoots and roots. In shoots, POX activity (especially that of the high mobility POX3 isoform) was increased by the application of B10, alone or in combination with 2*Nr*EO or 4*Nr*EO (Figure 5A). When applied separately, 2*Nr*EO induced around a 20% decrease in POX activity in shoots. In roots, B5 and B10 treatments induced significant increase in total POX activity, while 2*Nr*EO and 4*Nr*EO treatments reduced the activity of this enzyme (Figure 5B). Both POX1 and POX2 isoforms were equally affected. Exposure of Arabidopsis to 2*Nr*EO+B5 and 4*Nr*EO+B5 treatments decreased the POX activity in roots, when compared to B5 treatment. Treatment with 4*Nr*EO+B10 slightly decreased the activity of the enzyme, when compared to B10-treated plants (Figure 5B). Treatments with 2*Nr*EO+B10 and 4*Nr*EO+B10 significantly increased POX activity, when compared to the non-treated plants (Figure 5B). Based on immunoblot analysis eight isoforms in shoots (POXs) and seven in roots (POXr) can be distinguished (Figure 5C,D). In shoots, the most abundant isoforms were POX4s and POX6s, while in roots POX6r, POX7r and POX3r prevailed. The highest total POX abundance in shoots was recorded on treatments with BASTA (B5 and B10), while the 2*Nr*EO treatment was the most efficient in reducing the abundance of this enzyme (Figure 5C). This is especially visible for the most abundant POX isoforms in shoots (POX4s and POX6s). It was demonstrated that application of *Nr*EO (both 2*Nr*EO and 4*Nr*EO) simultaneously with BASTA (B5 and B10) reduced the abundance of POX isoforms, in comparison to appropriate controls (B5 and B10, respectively) (Figure 5C). In roots, on the other hand, BASTA treatments reduced the abundance of the majority of POX isoforms, while *Nr*EO treatments increased the abundance of this enzyme (Figure 5D). The exception is the abundance of POX7r isoform, which was increased following B10 application. Joint application of BASTA and *Nr*EO generally reduced the abundance of POX isoforms in roots when compared to appropriate BASTA treatments.

Native-PAGE in-gels assays for SOD activity revealed that Arabidopsis exposed to BASTA (B5 and B10) and/or *Nr*EO (2*Nr*EO and 4*Nr*EO) treatments for 10 days, demonstrated unchanged activity in all shoot samples (Figure 6A), whereas the SOD activity was decreased following 2*Nr*EO and 2*Nr*EO+B5 treatments in roots (Figure 6E). Three SOD isoforms were detected in shoot samples, and five in roots. Based on different sensitivity to corresponding inhibitors, Mn-, Fe-, and CuZn-SOD isoforms were demonstrated in Arabidopsis shoots and roots (Figure 6B,F). In Arabidopsis shoots, CuZn-SOD isoform was the major contributor to the overall SOD activity, and was followed by Mn-SOD and Fe-SOD isoforms. In roots, one Fe-SOD and one CuZn-SOD were recorded, as well as three Mn-SOD isoforms (Figure 6E,F). The abundance of Mn-SOD and Fe-SOD isoforms in Arabidopsis shoots and roots was evaluated using immunodetection assay (Figure 6C,D,G,H). All treatments, except 2*Nr*EO, induced a decrease in Mn-SOD abundance in shoots, in comparison to non-treated plants (Figure 6C). The opposite model of action was observed in root samples, since all treatments, except 2*Nr*EO and 2*Nr*EO+B5, increased Mn-SOD abundance in comparison to non-treated plants (Figure 6G). The increase was the most pronounced on B5 treatment. Interestingly, the application of *Nr*EO (2*Nr*EO and 4*Nr*EO) simultaneously with BASTA (B5 and B10) in shoots mitigated the inhibitory effects of herbicide on Mn-SOD abundance (Figure 6C). In roots, *Nr*EO mitigated the stimulatory effect of B5 on Mn-SOD abundance (Figure 6G). As for the Fe-SOD isoform abundance, it was increased on all treatments, with the exception of B10 treatment, where changes were not observed (Figure 6D). In roots, a similar trend was observed as in shoots (Figure 6H), and FeSOD abundance was especially increased on B5 treatment. The results of immunodetection assay revealed significant amounts of Fe-SOD protein in both shoots and roots (Figure 6D,H).

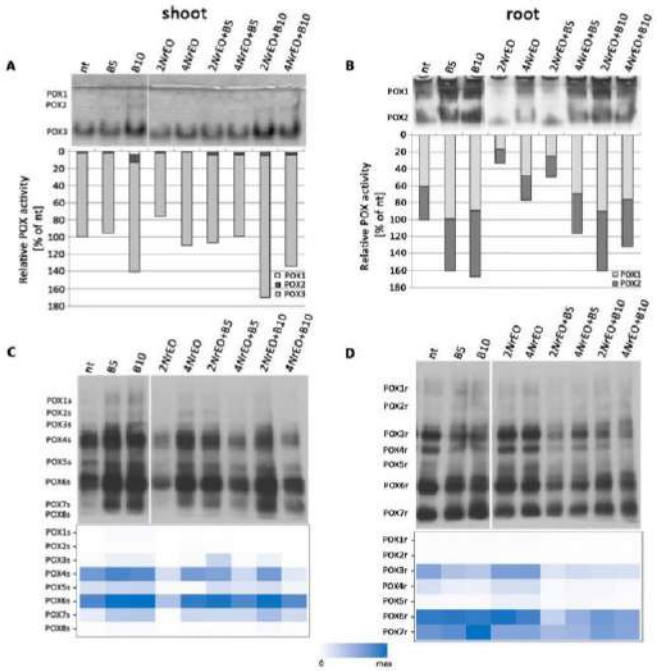


Figure 5. Peroxidases (POX) activity in Arabidopsis grown *in vitro*, measured 10 days after treatment with BASTA applied at two concentrations B5 (5 mg L^{-1}) and B10 (10 mg L^{-1}), *N. rtanjensis* essential oil applied at two concentrations 2% (2*Nr*EO) and 4% (4*Nr*EO), and of their combinations. Total soluble proteins were extracted from shoots and roots and loaded (20 µg per lane) on 10% gel, separated by native-PAGE and assayed for POX activity. **(A,B)** The detected activities were measured densitometrically and presented as relative POX activity (% of control) in comparison to non-treated plants (nt). **(C,D)** For SDS-PAGE 20 µg of soluble proteins per line from shoots and roots were loaded on 10% gel, transferred to nitrocellulose membrane and immunoblotted using primary antibodies antisheep horseradish POX. Corresponding heat-maps show relative abundances of POX proteins. Maximal values on the colour scales represent maximal values recorded for each immunoblot, independently. Abbreviations: s—shoot, r—root.

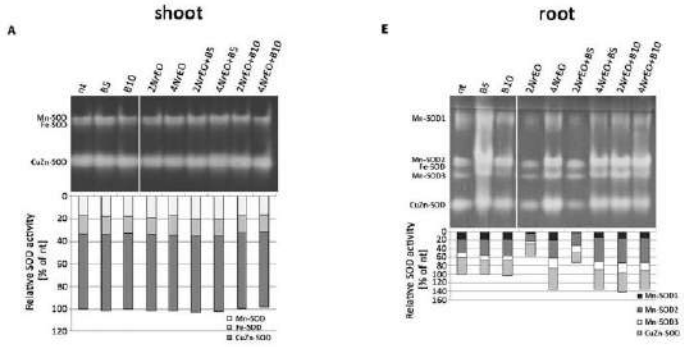


Figure 6. *Cont.*

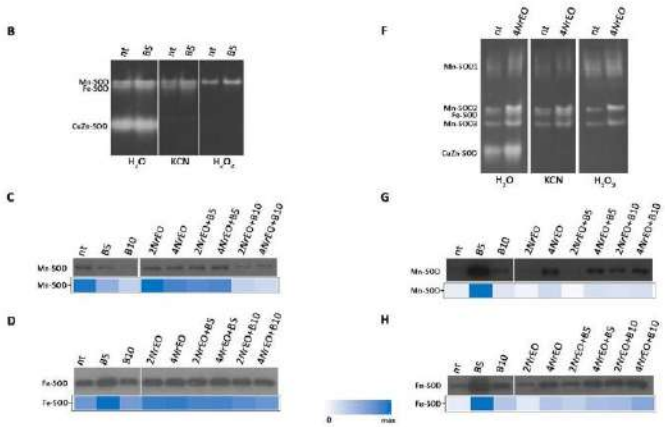


Figure 6. Superoxide dismutase (SOD) activity in Arabidopsis shoots and roots grown *in vitro*, measured 10 days after treatment with BASTA applied at two concentrations B5 (5 mg L^{-1}) and B10 (10 mg L^{-1}), *N. rtanjensis* essential oil applied at two concentrations 2% (2*Nr*EO) and 4% (4*Nr*EO), and of their combinations. Total soluble proteins were extracted from shoots and roots and loaded (5 µg per lane) on 10% gel, separated by native-PAGE and assayed for SOD activity. (**A**,**E**) The detected activities were measured densitometrically and presented as relative SOD activity (% of control) in comparison to non-treated plants (nt). (**B**,**F**) Identification of SOD isoforms based on differential sensitivity to hydrogen peroxide and potassium cyanide. (**C**–**H**) For SDS-PAGE 20 µg of soluble proteins per line from shoots/roots were loaded on 10% gels, transferred to nitrocellulose membranes and immunoblotted using primary antibodies antirabbit MnSOD (**C**,**G**) and antirabbit chloroplastic FeSOD (**D**,**H**). Corresponding heat-maps show relative abundances of SOD proteins. Maximal values on the color scales represent maximal values recorded for each immunoblot, independently.

3. Discussion

Phosphinothricin can be absorbed via roots, and further efficiently transported through the xylem sap to the above-ground parts of plants [24]. Information regarding PPT root uptake are scarce, therefore our intention was to describe the effects of PPT absorbed via roots on physiology and metabolism of Arabidopsis, and thus to get an insight into its contribution to the overall herbicidal action. *in vitro* experimental setup enabled the direct exposure of roots to BASTA incorporated in the sterile culture medium, where metabolizing by microorganisms is excluded. We further aimed to investigate whether the phenomenon of *Nr*EO-mediated mitigation of the BASTA-induced ammonium toxicity in Arabidopsis, observed during their simultaneous foliar application [4], can also be achieved when plants are exposed to herbicide via roots. It was possible to expose Arabidopsis to the atmosphere enriched with volatiles of *N. rtanjensis* EO, and to trace their amount by appropriate contemporary methodology, such as PTR-MS. Nepetalactone amount in the atmosphere of *in vitro* culture vessels was obviously dependent on the applied concentration of *Nr*EO, and it decreased gradually over 10 days of in vitro cultivation. As previously suggested by Dmitrović et al. [29], this decrease might result from VOCs transitions between liquid and gas phases, and their degradation, transformation, precipitation, or leakage from the glass vessels to the ambient air. Either way, the amount of this monoterpenoid was sufficient to induce nepetalactone-expected effects on Arabidopsis, when applied alone or in combination with BASTA.

Apparent BASTA-induced effects, similar to those induced by foliar application of herbicide [4], were visible within the present experimental setup, which indicated efficient PPT translocation from roots to the above ground plant parts. Arabidopsis displayed clear ammonium toxicity symptoms in the form of leaf chlorosis and reduced plant growth,

accompanied by decreased GS activity and increased tissue ammonium content in shoots and roots. Li et al. [11] demonstrated that shoot tissues of Arabidopsis are hypersensitive to ammonium exposure, compared to roots that store ammonium in vacuoles, thus avoiding ammonium toxicity. This is in accordance with our findings that the reduction in GS activity was more pronounced in Arabidopsis roots, where a less pronounced increase in ammonium content was recorded when compared to the shoots. The most severe BASTA-phenotype was observed following B10 treatments. Leaf chlorosis was accompanied by a decline in Chl a, Chl b and Chl a+b content, as well as by perturbation in the content of other traced metabolites (soluble sugars and organic acids). Treatments with *Nr*EO, however, induced expected changes in Arabidopsis growth, chlorophyll content, sugars and organic acid content. As in our previous study that adopted foliar *Nr*EO application [4], Arabidopsis shoots exposed to 2*Nr*EO and 4*Nr*EO via atmosphere displayed growth reduction and a decrease in soluble sugars content, but no significant changes in organic acids content. Applied in a similar way, this EO has previously been reported to induce growth inhibition of ragweed (*Ambrosia artemisiifolia* L.) shoots [29], and garden cress seedlings (*Lepidium sativum*) [35]. Regardless of the method of application (foliar or via-atmosphere), *Nr*EO induced a decrease in Chl a, Chl b and Chl a+b content in Arabidopsis [4], but also in ragweed shoots [29]. The result of BASTA and *Nr*EO joint application was the reduced growth of Arabidopsis shoots when compared to the non-treated plants, while roots were not affected. Generally, the recorded values of changes in chlorophyll and organic acids content in Arabidopsis shoots following *Nr*EO+BASTA treatments, are the mean values of the two agents applied independently.

As the PPT mechanism of action is based on its inhibitory effects on GS activity, we traced the effects of BASTA on GS activity and GS-coding genes expression, as well as on the content of ammonium in Arabidopsis shoots and roots. There are two classes of GS isoforms in plants: cytosolic GS1 and plastidic GS2. GS1 isoforms are located in the cytoplasm and are encoded by a multigene family (*GLN1;1*, *GLN1;2*, *GLN1;3*, *GLN1;4* and *GLN1;5*), while GS2 is encoded by the *GLN2* gene in Arabidopsis [34,36]. Opposite to the greenhouse-grown Arabidopsis, where *GLN2* expression in leaves was higher compared do *GS1* genes [4,37], transcript levels of *GS1* (especially *GLN1;2*) were more abundant compared to *GLN2* in plants grown under in vitro conditions. This suggested that GS1 is the major isoform responsible for ammonium assimilation in both shoots and roots of Arabidopsis grown *in vitro*. Taking into account in vitro mixotrophic growth conditions, which include external carbohydrate supply through culture medium and lowered photosynthetic rate, it was not surprising that the expression of plastidic *GLN2* isoform was low in both shoots and roots. However, it is well documented that expression levels of *GLN1;1* and *GLN1;2* in Arabidopsis are modulated in response to abiotic stresses, including ammonium toxicity [4,37], low and/or high N [38], and salt stress [39]. Ammonium accumulation induced by BASTA treatment within the present study induced the expression of *GS1* (*GLN1;1*, *GLN1;2*, *GLN1;3*, and *GLN1;4*) in Arabidopsis shoots, which was positively correlated with the amount of GS1 proteins, but negatively correlated with the activity of GS. This indicated, on the one hand, the existence of transcriptional regulation of GS expression by BASTA, probably via accumulated ammonium, but also the potent inhibition of GS activity by the herbicide. Such a mode of action was also recorded in Arabidopsis leaves during foliar application of herbicide [4]. As previously suggested, PPT irreversibly, but not covalently, binds to the active site of GS and inhibits its activity [1,40,41]. The result of reduced GS activity is the accumulation of phytotoxic ammonium in tissues, which alters the overall physiology of Arabidopsis. The activity and abundance of GS are generally higher in Arabidopsis roots, than in shoots, and BASTA-induced reduction in GS activity was more pronounced in roots. However, Arabidopsis shoots experienced more extensive ammonia accumulation than roots. Roots have efficient mechanisms for ammonium detoxification, which, as previously suggested [11], might include its storage in vacuoles.

The application of *NrEO* induced no significant changes in the expression of *GLN1;1*, *GLN1;2*, *GLN1;3*, *GLN1;4*, and *GLN2* in Arabidopsis shoots, on GS1 and GS2 abundance and

activity, as well as on ammonium content. This is in accordance with our previous study, when foliar application of *Nr*EO was adopted [4]. Furthermore, no obvious differences between BASTA and *Nr*EO+BASTA treated plants were visible at the level of *GS1* and *GS2* gene expression in shoots. However, GS abundance in shoots was higher in BASTA treatments than in *Nr*EO+BASTA treatments, with the exception of 2*Nr*EO+B10 treatment. The reason for such a discrepancy might be the induced expression of *GLN* genes earlier during the treatment which results in the increased amount of GS proteins and subsequent reduction of gene expression to the control values, which is visible after 10 days of treatments. This presumption is in accordance with our previous study, which revealed that changes in GS genes expression are more pronounced 1 day after the beginning of BASTA and/or *Nr*EO treatments than after 10 days [4]. In roots, the situation was even more complex. ANOVA indicated that the exposure of plants to *Nr*EO induced significant down-regulation of all analyzed genes. Post-hoc analysis indicated that *GLN1;1* and *GLN1;3* were significantly down-regulated following 2*Nr*EO in comparison to non-treated plants. In contrast to shoots, all *GLN1* genes except *GLN1;4*, as well as *GLN2* were down-regulated by BASTA, and the inhibitory effect was dose-dependent. Simultaneous application of BASTA and *Nr*EO severely reduced the inhibitory effect of BASTA on GS gene expression in roots, indicating the regulation on transcriptional level. Such gene expression profiles were not reflected at the protein level through GS1 and GS2 abundance in roots, which was increased in BASTA treatments. In roots, *Nr*EO-induced decrease in GS1 abundance was recorded. Simultaneous application of BASTA (B5 and B10) and *Nr*EO (2*Nr*EO and 4*Nr*EO) reduced GS1 amount in roots. Again, a strong suppression of BASTA-induced effects, following its simultaneous application with *Nr*EO, was observed 10 days after the treatment. However, the possibility that the increase in *GS* expression occurs earlier in treatments with BASTA, which is followed by the increase in GS abundance visible at 10DAT, should not be neglected. After that, the expression of GS probably decreases.

Ammonia accumulation in response to GS inhibition is often considered to be the driver of phosphinothricin toxicity. Along with the inhibition of GS, BASTA action also leads to oxidative stress in plants, which is most probably a secondary effect of the altered metabolic pathways [42]. It has recently been suggested that glufosinate is toxic to plants, not because of ammonia accumulation or carbon assimilation inhibition, but the production of ROS, which drive the lipid peroxidation of the cell membranes and cell death [16,17]. Plants have developed mechanisms to cope with oxidative stress induced by ROS accumulation, which include enzymatic and non-enzymatic antioxidants [43]. Thus, our intention was to investigate alterations in the antioxidant system of Arabidopsis grown in vitro, induced by the application of BASTA and/or *Nr*EO. BASTA (B5 and B10) and *Nr*EO (2*Nr*EO and 4*Nr*EO), when applied independently for 10 days, altered the activity and abundance of CAT, POX and SOD in Arabidopsis shoots and roots. Induced changes were more pronounced in roots than in shoots. BASTA induced significant increases in both CAT and POX activity and abundance in roots, while SOD activity was not affected. The increase in CAT and POX activities indicated that application of glufosinate induced the increasing ROS level in Arabidopsis roots, and that enzyme activities increased correspondingly in order to eliminate excess ROS. In shoots, B10 treatment induced decrease of CAT activity and abundance, while POX activity was increased. Superoxide dismutase activity in Arabidopsis shoots was not significantly affected by BASTA treatments. While the total SOD activity in shoots and roots remains unchanged in response to applied BASTA and *Nr*EO treatments, the contribution of individual isoforms to the overall activity is variable, depending on the BASTA treatment. Mn-SOD abundance in shoots was decreased following BASTA treatment, while in roots, the abundance of this isoform increased when B5 was applied. In Arabidopsis shoots and roots B5 treatment increased the abundance of Fe-SOD isoform. It has been recently suggested that glufosinate affects the balance between ROS generation and scavenging in *Amaranthus palmeri*, and induces the increase in CAT, POX and SOD activities in an attempt to quench the nascent ROS burst [15]. SOD participates in the removal of $O_2\cdot^-$ by conversion to H_2O_2 and O_2, and its activity reflects the readiness of plants to

scavenge cellular free radicals. CAT and POX can effectively remove intracellular H_2O_2 and translate it into H_2O and O_2. The increased activity and abundance of CAT and POX following BASTA treatment indicated that these two enzymes are the major responsible for ROS quenching. Similarly, activities of antioxidant enzymes (CAT, POX, SOD) increased in the leaves of maize seedlings treated by PPT [44]. Savič et al. [45] demonstrated that BASTA induced an increase in POX activity (about 42%) and appearance of new POX isoforms in *Lotus corniculatus* plants. In *Chlorella vulgaris* PPT increased CAT and SOD activity about 2.9 times, four days after application compared to control [46].

Interestingly, the effect of *Nr*EO on the activity of antioxidant enzymes in Arabidopsis shoots was opposite to those recorded for BASTA. Decrease in CAT activity in both Arabidopsis shoots and roots was followed by a decrease in CAT abundance in these organs. The same trend, which was more pronounced in roots, was observed for POX activity. These results are in accordance with previous studies, which described *Nr*EO-induced alterations in the antioxidative defense system of ragweed (*Ambrosia artemisifolia*) shoots, characterized by increased POX activity and decreased CAT and SOD activities [29]. Furthermore, reference [35] showed that nepetalactone-rich EOs decreased POX, CAT, and SOD (Fe-SOD and CuZn-SOD) activities in cress seedlings. The decrease in POX, CAT, and SOD activities in Arabidopsis roots was recorded for 2*Nr*EO treatment, which might be the result of decreased ROS formation and further studies are necessary to confirm this hypothesis. However, SOD activity was increased in Arabidopsis roots treated with higher concentrations of *Nr*EO (4*Nr*EO), which suggests that this enzyme is active in scavenging $O_2^{\cdot -}$ in roots. It is well documented that components of essential oils display phytotoxic effects by generating ROS, and inducing oxidative stress [29,47,48]. Singh et al. [47] demonstrated that the content of H_2O_2 was increased, as well as the activity of CAT, POX and SOD, in *Cassia occidentalis* roots after α-pinene application. α-Pinene induced oxidative stress in *C. occidentalis* roots, which was visible through the disruption of membrane integrity, lipid peroxidation, H_2O_2 generation, proline accumulation, and increased activities of SOD, ascorbate peroxidase (APX), guaiacol peroxidase (GPX), CAT, and glutathione reductase (GR).

Simultaneous application of *Nr*EO and BASTA, two agents inducing opposite effects on the Arabidopsis antioxidant system, especially in roots, mitigated the effects of the two agents applied independently. This was evident for CAT and POX activities, which were significantly lower on *Nr*EO+B5 treatments (both 2*Nr*EO and 4*Nr*EO) than on B5 treatments, and higher than on *Nr*EO treatments. However, simultaneous application of BASTA and *Nr*EO significantly increased SOD activity in Arabidopsis roots, which suggests that this enzyme is more active in scavenging O_2^{-} than on BASTA treatments. *Nr*EO preserves the GS activity and maintains subtoxic ammonia levels in Arabidopsis leaves during simultaneous BASTA and *Nr*EO application [4], and it most likely acts by inducing SOD activity in roots and thus more efficient ROS scavenging, which contributes to mitigating the phytotoxic effects of BASTA. Our further work, involving sophisticated methods and tools (e.g., ROS tissue localization, analysis of other antioxidant enzymes activity at the protein and gene expression level, metabolomics, tracing PPT uptake and metabolism in Arabidopsis), will help us to comprehensively explain the phenomenon of antagonistic interaction between PPT and *Nr*EO. In summary, *Nr*EO has elicited considerable agronomic interest, and future efforts directed towards unraveling the mechanism by which these two agents interact will contribute to the possible utilization of *Nr*EO as an eco-friendly bioherbicide and an agent for mitigation of the effects of PPT residues in the soil on non-target plants.

4. Materials and Methods

4.1. Chemicals and Reagents

Commercial herbicide BASTA® (containing 150 g L^{-1} of active ingredient glufosinate-ammonium) was purchased from BASF SE (Germany). Solvents for HPLC–MS analyses (acetonitrile, acetic and formic acids) were of LC–MS grade (Fisher Scientific, Loughborough, UK).

Methanol for metabolites extraction and preparation of EO dilutions (HPLC grade) was purchased from AppliChem (Cheshire, CT, USA). Ultrapure water was generated by deionization (Millipore, Billerica, MA, USA). Standards for sugars and organic acids determination were purchased from Sigma-Aldrich (Sigma Co., St. Louis, MO, USA).

4.2. Preparation of Essential Oils

Nepeta rtanjensis Diklić & Milojević plants were cultivated at the experimental field of the Institute for Biological Research "Siniša Stanković"—National Institute of the Republic of Serbia, University of Belgrade, Serbia. Aerial parts of flowering plants were harvested, air-dried and used for the isolation of EO by hydrodistillation, as previously described by Skorić et al. [31]. To obtain 2% (*v*/*v*) and 4% (*v*/*v*) final EO concentrations, *N. rtanjensis* EO (*Nr*EO) was diluted in 99.8% methanol.

4.3. Joint Effect of BASTA via Roots and NrEO on A. thaliana Plants—In Vitro Phytotoxic Assay

Seeds of *Arabidopsis thaliana* (L.) Heynh, accession Col−8 (N60000), were obtained from the Nottingham Arabidopsis Stock Centre (http://arabidopsis.info). Arabidopsis seeds were surface sterilized in 20% solution of commercial bleach in ethanol (0.8% active chlorine) for 1 min and subsequently rinsed 5 times in sterile deionized water. Seeds were transferred into 9 cm Petri dishes containing 20 mL of basal medium (BM), a modified Murashige and Skoog [49] medium, supplemented with half-strength macro elements, 20 g L^{-1} sucrose and 7 g L^{-1} agar. The pH of the culture medium was adjusted to 5.8 before sterilization by autoclaving at 114 °C for 25 min. Seeds were stratified for 3 days at 4 °C in the dark. Petri dishes were subsequently transferred to a growth chamber at 25 ± 2 °C, under 16/8 h light/dark regime with a photon flux density of 70 μmol m^{-2} s^{-1}. After 11 days, Arabidopsis seedlings were transferred into 250 mL Erlenmeyer flasks containing 100 mL BM, and grown under the same light regime.

After two weeks, five well-developed seedlings were placed in 350 mL glass jars containing 75 mL BM, with or without dissolved BASTA at final concentrations of active ingredient (glufosinate ammonium) 5 mg L^{-1} (B5) or 10 mg L^{-1} (B10). Fifty μL of diluted *Nr*EO (2 or 4% solution in methanol) was applied to the filter paper (2.5 × 2.5 cm) which was subsequently rolled up, placed on a sterilized metal holder, and stuck into the culture medium with care not to contact the plants. Volatile components of the EO evaporated from the filter paper surface into the atmosphere of glass jar, where Arabidopsis seedlings were cultivated. For the negative control, filter paper was moistened with 50 μL of 99.8% methanol. Thus, 9 experimental groups were prepared: (1) non-treated plants (nt, plants exposed to methanol only), (2) methanol + BASTA 5 mg L^{-1} (B5), (3) methanol + BASTA 10 mg L^{-1} (B10), (4) 2% EO (2*Nr*EO), (5) 4% EO (4*Nr*EO), (6) 2% EO + 5 mg L^{-1} BASTA (2*Nr*EO+B5), (7) 4% EO + 5 mg L^{-1} BASTA (4*Nr*EO+B5), (8) 2% EO + 10 mg L^{-1} BASTA (2*Nr*EO+B10), and (9) 4% EO + 10 mg L^{-1} BASTA (4*Nr*EO+B10). All glass jars were closed with polycarbonate caps, and placed in a growth chamber at 25 ± 2 °C, under 16/8 h light/dark regime, with a photon flux density of 70 μmol m^{-2} s^{-1}. After ten days, *A. thaliana* shoots and roots were separately pooled, weighed, frozen in liquid nitrogen (LN), and stored at −80 °C until further use. Plant material for protein analysis (Native-PAGE electrophoresis and GS activity determination) was weighed and immediately used in assays. The results were obtained using three biological replicates. Each biological replicate represented a pool of tissues collected from five plants grown in the single glass jar.

4.4. Determination of Volatile Organic Compounds (VOCs) Concentration in the Atmosphere of Culture Vessels

Concentrations of *Nr*EO VOCs in the atmosphere of glass vessels used for Arabidopsis cultivation were measured by Proton Transfer Reaction Mass Spectrometer (Standard PTR-quad-MS, Ionicon Analytik, GmbH, Innsbruck, Austria). Concentrations of VOCs were recorded at the beginning of the experiment, and after 10 days of treatments. Analysis was targeted towards compounds displaying molecular ions $[M+1]^+$ at m/z 137, 153, 167, and 205 (0.1 s dwell time), and were obtained in 1.8 s cycles. Drift tube parameters

included: pressure 2.11–2.13 mbar; temperature 60 °C; voltage 600 V; E/N parameter 145 Td; and reaction time 90 µs. The count rate of H_3O^+/H_2O was 2.1–14.1% of the count rate of H_3O^+ ions, which was in the range $2.1 \times 10^6 - 6.1 \times 10^6$ counts s^{-1}. The calibration was done according to Taipale et al. [50], using TO−15 Supelco gas mixture (m/z 57, m/z 79, m/z 93, m/z 107, and m/z 121), diluted with ASGU 370-p HORIBA system zero air to five concentrations ranging from 0.5 to 100 parts-per-billion (ppb). For the calculation of transmission coefficients above m/z 170, the logarithmic extrapolation was used. Calculated normalized sensitivities were 2.23, 1.77, 1.52, and 1.33 npcs ppb^{-1} for m/z 137, 153, 167, and 205, respectively. Nepetalactones (*cis,trans*- and *trans,cis*-nepetalactone) showed $[M+1]^+$ at m/z 167. Compounds with $[M+1]^+$ at m/z 137 most likely corresponded to monoterpenoids α- and β-pinene, while α-campholenal, neral and geranial showed $[M+1]^+$ at m/z 153. Sesquiterpenoids showing protonated mass in PTR analysis of $[M+1]^+$ at m/z 205, corresponding to γ-cadinene, δ-cadinene, *cis*- and *trans*-caryophyllene, and α-humulene, were also analyzed. Measurements were performed in triplicate for each of the treatments. Values are presented as parts-per-billion-volume (ppbV).

4.5. Metabolic Profiling of Sugars, Organic Acids and Chlorophyll

HPLC analyses of soluble sugars (sucrose-Suc, fructose-Fru, and glucose-Glu) and organic acids (oxalic, succinic, malic, fumaric and isocitric acid) were performed as described earlier by Dmitrović et al. [4]. Values recorded for soluble sugars are calculated as mg 100 mg^{-1} FW, and for organic acids are µg mg^{-1} FW. Chlorophylls (Chl) were extracted from Arabidopsis shoots following the modified method of Porra et al. [51] and described in detail by Dmitrović et al. [4]. The content of Chl a, Chl b, and Chl a+b is calculated as µg g^{-1} FW.

4.6. Quantitative RT-PCR Analysis of GS-Coding Genes Expression

RNA extraction and the subsequent cDNA preparation were performed as described in Dmitrović et al. [4]. Quantitative RT-PCR (qRT-PCR) was conducted by using primers for *GLN1;1*, *GLN1;2*, *GLN1;3*, *GLN1;4* and *GLN2*, as described before [4]. *18S rRNA* was used as a reference gene. Relative quantification of gene expression was performed, as described by Livak and Schmittgen [52]. The results are presented as log2 fold change in expression compared to control treatments.

4.7. Determination of GS Activity and Ammonium Content

Protein extraction, native polyacrylamide gel electrophoresis (Native-PAGE) separation and activity staining of GS (EC 6.3.1.2) isoforms was performed as described by Dragićević et al. [53]. Total GS activity in protein extracts was assayed as described by Nikolić et al. [5]. Total ammonium content in leaves was determined by the phenol-hypochlorite method described in detail by Dmitrović at al. [4].

4.8. Determination of CAT, POX and SOD Activities

Antioxidant enzymes extraction and the activity determination were carried out as described by Dmitrović at al. [29]. Detailed procedures for Native-PAGE used for CAT (EC 1.11.1.6), POX (EC 1.11. 1.7) and SOD (EC 1.15.1.1) activity determination were described by Dmitrović at al. [29], with some modifications. POX activity was visualized after incubation of gels in staining solution containing 10% 4-chloro-α-naphthol and 0.03% H_2O_2 in 50 mM K-phosphate buffer pH 6.5.

4.9. SDS-PAGE Electrophoresis and Immunoblotting

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and subsequent transfer of proteins onto nitrocellulose membranes were performed, as reported by Dmitrović et al. [29]. Immunoblot analysis was conducted using rabbit polyclonal antibodies GLN1 GLN2 (AS08295), CAT (AS09501), MnSOD (AS09524), FeSOD (AS06125), or sheep polyclonal antibodies for POX (AS09548), all purchased from Agrisera, Sweden.

Primary antibodies were used in dilutions of 1:2500 (GS), 1:1000 (CAT), 1:2000 (POX), 1:1000 (MnSOD), and 1:1000 (FeSOD). In the case of GS, CAT, Mn-SOD and Fe-SOD goat anti-rabbit IgG-peroxidase conjugated (A0545, Sigma-Aldrich, USA) secondary antibodies were used, while goat anti-mouse IgG HRP conjugated (AS111772, Agrisera, Sweden) antibodies were applied for POX. Goat anti-rabbit IgG-peroxidase antibodies were diluted to 1:20,000 (*v:v*), and goat anti-mouse IgG HRP were diluted to 1:5000 (*v:v*). After intensive washing in PBS buffer, protein signals were visualized using an enhanced chemiluminescence detection system (ECL). Densitometric analysis of band intensities was performed using ImageJ 1.32j software (W. Rasband, National Institute of Health, Bethesda, MD, USA). The obtained signal intensities were normalized to the highest value, and the results are presented as the relative abundances.

To confirm equal loading in immunoblots, membranes were incubated for 2 h in primary Actin-11 antibodies (Anti Mouse monoclonal IgG2b lyophilized, AS10702; Agrisera Antibodies, Sweden) diluted to 1:1000 (*v:v*) and goat anti-mouse IgG horse radish peroxidase conjugated secondary antibodies (AS11 1772, Agrisera Antibodies, Sweden) diluted to 1:50,000 (*v:v*) using the procedure described above. The obtained signal intensities for GS were normalized to the actin values, and the obtained results were normalized to the highest value and presented as relative abundances.

4.10. Statistical Analyses

Statistical analyses were performed using R [54]. The data were subjected to "sequential" sum of squares (type I SS ANOVA) factorial ANOVA by testing the main effect of BASTA, followed by the main effect of *Nr*EO after the main effect of BASTA, followed by the interaction effect *Nr*EO+BASTA after the main effects. For the analyses of qRT-PCR data log2-fold change was used as a response variable in factorial ANOVA. For all other models, the homoscedasticity and normality of the residuals were checked graphically, and if these assumptions were violated, the data was transformed prior to the statistical analyses using Box-Cox power transformation [55] incorporated in the R library MASS [56]. ANOVA was followed by Tukey's post-hoc test at the $p < 0.05$ significance level using the R library emmeans [57].

Supplementary Materials: The following are available online at https://www.mdpi.com/2223-7747/10/1/142/s1, Figure S1: Copy number of *GLN1* (*GLN1;1*, *GLN1;2*, *GLN1;3* and *GLN1;4*) and *GLN2* transcripts in shoots and roots of non-treated Arabidopsis grown under *in vitro* conditions, Table S1: Results of factorial ANOVA on quantities obtained in Arabidopsis shoots and roots, measured 10 days after *in vitro* treatment with BASTA, *N. rtanjensis* essential oil, and of their combinations. The asterisks denote the level of statistical significance: * < 0.05, ** < 0.01 and *** < 0.001. B—BASTA, *Nr*EO—*Nepeta rtanjensis* essential oil.

Author Contributions: Conceptualization, D.M. (Danijela Mišić), S.D., M.D. and J.S.; Methodology, S.D., M.D., J.S., M.M., S.Ž., D.M. (Dragana Matekalo), M.P., V.M.; Validation, D.M. (Danijela Mišić), S.D., M.D. and J.S.; Investigation, S.D., M.D., J.S., S.Ž., M.M., D.M. (Dragana Matekalo), M.P., V.M.; Formal Analysis, M.D., S.D. and D.M. (Danijela Mišić), V.M., M.P.; Data Curation, D.M. (Danijela Mišić), S.D. and M.D.; Writing—Original Draft Preparation, S.D., M.M. and D.M. (Danijela Mišić); Writing—Review & Editing, M.D., J.S., S.Ž., D.M. (Dragana Matekalo), M.P., V.M.; Visualization, M.D., S.D., and D.M. (Danijela Mišić); Supervision, D.M. (Danijela Mišić); Funding Acquisition, D.M. (Danijela Mišić). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Serbian Ministry of Education, Science and Technological Development of the Republic of Serbia, grant numbers OI173024, 451-03-68/2020-14/200007, 451-03-68/2020-14/200053, and through the grant to the Institute of Physics Belgrade. The APC was funded by Serbian Ministry of Education, Science and Technological Development of the Republic of Serbia, grant number 451-03-68/2020-14/200007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Manderscheid, R.; Wild, A. Studies on the mechanism of inhibition by phosphinothricin of glutamine synthetase isolated from Triticum aestivum L. *J. Plant Physiol.* **1986**, *123*, 135–142. [CrossRef]
2. Lydon, J.; Duke, S.O.; Singh, B.K. Inhibitors of glutamine biosynthesis. In *Plant Amino Acids: Biochemistry and Biotechnology*; BK Singh: New York, NY, USA, 1999; pp. 445–464.
3. Wohlleben, W.; Arnold, W.; Broer, I.; Hillemann, D.; Strauch, E.; Punier, A. Nucleotide sequence of the phosphinothricin N-acetyltransferase gene from Streptomyces viridochromogenes Tü494 and its expression in Nicotiana tabacum. *Gene* **1988**, *70*, 25–37. [CrossRef]
4. Dmitrović, S.; Dragićević, M.; Savić, J.; Milutinović, M.; Živković, S.; Maksimović, V.; Matekalo, D.; Mišić, D. Nepetalactone-rich essential oil mitigates phosphinothricin-induced ammonium toxicity in *Arabidopsis thaliana* (L.) Heynh. *J. Plant Physiol.* **2019**, *237*, 87–94. [CrossRef] [PubMed]
5. Nikolić, R.; Zdravković-Korać, S.; Ninković, S.; Dragićević, M.; Miljuš-Đ, J.; Banović, B.; Bohanec, B.; Savić, J.; Mitić, N. Fertile transgenic Lotus corniculatus resistant to the non-selective herbicide phosphinothricin. *Ann. Appl. Biol.* **2013**, *163*, 475–493. [CrossRef]
6. Eisenberg, D.; Gill, H.S.; Pfluegl, G.M.U.; Rotstein, S.H. Structure-function relationships of glutamine synthetase. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2000**, *1477*, 122–145. [CrossRef]
7. Bernard, S.M.; Habash, D.Z. The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New Phytol.* **2009**, *182*, 608–620. [CrossRef]
8. Britto, D.T.; Kronzucker, H.J. NH4+ toxicity in higher plants: A critical review. *J. Plant. Physiol.* **2002**, *159*, 567–584. [CrossRef]
9. Bittsánszky, A.; Pilinszky, K.; Gyulai, G.; Komives, T. Overcoming ammonium toxicity. *Plant. Sci.* **2015**, *231*, 184–190. [CrossRef]
10. Daniel-Vedele, F.; Filleur, S.; Caboche, M. Nitrate transport: A key step in nitrate assimilation. *Curr. Opin. Plant. Biol.* **1998**, *1*, 235–239. [CrossRef]
11. Li, B.; Li, G.; Kronzucker, H.J.; Baluška, F.; Shi, W. Ammonium stress in Arabidopsis: Signaling, genetic loci, and physiological targets. *Trends Plant. Sci.* **2014**, *19*, 107–114. [CrossRef]
12. Nimptsch, J.; Pflugmacher, S. Ammonia triggers the promotion of oxidative stress in the aquatic macrophyte Myriophyllum mattogrossense. *Chemosphere* **2007**, *66*, 708–714. [CrossRef] [PubMed]
13. Skopelitis, D.S.; Paranychianakis, N.V.; Paschalidis, K.A.; Pliakonis, E.D.; Delis, I.D.; Yakoumakis, D.I.; Kouvarakis, A.; Papadakis, A.K.; Stephanou, E.G.; Roubelakis-Angelakis, K.A. Abiotic stress generates ROS that signal expression of anionic glutamate dehydrogenases to form glutamate for proline synthesis in tobacco and grapevine. *Plant. Cell* **2006**, *18*, 2767–2781. [CrossRef] [PubMed]
14. Patterson, K.; Cakmak, T.; Cooper, A.; Lager, I.D.A.; Rasmusson, A.G.; Escobar, M.A. Distinct signalling pathways and transcriptome response signatures differentiate ammonium-and nitrate-supplied plants. *Plant. Cell Environ.* **2010**, *33*, 1486–1501. [CrossRef]
15. Takano, H.K.; Beffa, R.; Preston, C.; Westra, P.; Dayan, F.E. A novel insight into the mode of action of glufosinate: How reactive oxygen species are formed. *Photosynth. Res.* **2020**, *144*, 361–372. [CrossRef] [PubMed]
16. Takano, H.K.; Dayan, F.E. Glufosinate-ammonium: A review of the current state of knowledge. *Pest Manag. Sci.* **2020**, *76*, 3911–3925. [CrossRef] [PubMed]
17. Takano, H.K.; Beffa, R.; Preston, C.; Westra, P.; Dayan, F.E. Reactive oxygen species trigger the fast action of glufosinate. *Planta* **2019**, *249*, 1837–1849. [CrossRef]
18. Shelp, B.J.; Swanton, C.J.; Hall, J.C. Glufosinate (phosphinothricin) mobility in young soybean shoots. *J. Plant. Physiol.* **1992**, *139*, 626–628. [CrossRef]
19. Behrendt, H.; Matthies, M.; Gildemeister, H.; Görlitz, G. Leaching and transformation of glufosinate-ammonium and its main metabolite in a layered soil column. *Environ. Toxicol. Chem. Int. J.* **1990**, *9*, 541–549. [CrossRef]
20. Gallina, M.A.; Stephenson, G.R. Dissipation of [14C] glufosinate ammonium in two Ontario soils. *J. Agric. Food Chem.* **1992**, *40*, 165–168. [CrossRef]
21. Ismail, B.S.; Ahmad, A.R. Attenuation of the herbicidal activities of glufosinate-ammonium and imazapyr in two soils. *Agric. Ecosyst. Environ.* **1994**, *47*, 279–285. [CrossRef]
22. Jia, G.; Xu, J.; Long, X.; Ge, S.; Chen, L.; Hu, D.; Zhang, Y. Enantioselective degradation and chiral stability of glufosinate in soil and water samples and formation of 3-methylphosphinicopropionic acid and N-acetyl-glufosinate metabolites. *J. Agric. Food Chem.* **2019**, *67*, 11312–11321. [CrossRef] [PubMed]
23. Dorn, E.; Görlitz, G.; Heusel, R.; Stumpf, K. Verhalten von Glufosinat-ammonium in der Umwelt–Abbau im und Einflußauf das Ökosystem. *Z. PflKrankh. PflSchutz Sonderh* **1992**, *13*, 459–468.
24. Takano, H.K.; Beffa, R.; Preston, C.; Westra, P.; Dayan, F.E. Physiological factors affecting uptake and translocation of glufosinate. *J. Agric. Food Chem.* **2020**, *68*, 3026–3032. [CrossRef] [PubMed]

25. Craigmyle, B.D.; Ellis, J.M.; Bradley, K.W. Influence of weed height and glufosinate plus 2,4-D combinations on weed control in soybean with resistance to 2,4-D. *Weed Technol.* **2013**. [CrossRef]
26. Merchant, R.M.; Sosnoskie, L.M.; Culpepper, A.S.; Steckel, L.E.; York, A.C.; Braxton, L.B.; Ford, J.C. Weed response to 2,4-D, 2,4-DB, and dicamba applied alone or with glufosinate. *J. Cotton Sci.* **2013**, *17*, 212–218.
27. Ganie, Z.A.; Jhala, A.J. Interaction of 2,4-D or dicamba with glufosinate for control of glyphosate-resistant giant ragweed (*Ambrosia trifida* L.) in glufosinate-resistant maize (*Zea mays* L.). *Front. Plant Sci.* **2017**. [CrossRef]
28. Koger, C.H.; Burke, I.C.; Miller, D.K.; Kendig, J.A.; Reddy, K.N.; Wilcut, J.W. MSMA antagonizes glyphosate and glufosinate efficacy on broadleaf and grass weeds. *Weed Technol.* **2007**, *21*, 159–165. [CrossRef]
29. Dmitrović, S.; Perišić, M.; Stojić, A.; Živković, S.; Boljević, J.; Nestorović Živković, J.; Aničić, N.; Ristić, M.; Mišić, D. Essential oils of two Nepeta species inhibit growth and induce oxidative stress in ragweed (*Ambrosia artemisiifolia* L.) shootsin vitro. *Acta Physiol. Plant.* **2015**. [CrossRef]
30. Nestorović Živković, J.; Dmitrović, S.; Jovanović, V.; Živković, S.; Božić, D.; Aničić, N.; Mišić, D. Allelopathic potential of essential oil of Nepeta rtanjansis. *Allelopath. J.* **2016**, *37*, 207–219.
31. Skorić, M.; Gligorijević, N.; Čavić, M.; Todorović, S.; Janković, R.; Ristić, M.; Mišić, D.; Radulović, S. Cytotoxic activity of Nepeta rtanjensis Diklić & Milojević essential oil and its mode of action. *Ind. Crop. Prod.* **2017**, *100*, 163–170. [CrossRef]
32. Sparks, J.T.; Bohbot, J.D.; Ristić, M.; Mišić, D.; Skorić, M.; Mattoo, A.; Dickens, J.C. Chemosensory responses to the repellent Nepeta essential oil and its major component nepetalactone by Aedes aegypti (Diptera: Culicidae), a vector of Zika virus. *J. Med. Entomol.* **2017**, *54*, 957–963. [CrossRef] [PubMed]
33. Aničić, N.; Matekalo, D.; Skorić, M.; Pećinar, I.; Brkušanin, M.; Nestorović Živković, J.; Dmitrović, S.; Dajić Stevanović, Z.; Schulz, H.; Mišić, D. Trichome-specific and developmentally regulated biosynthesis of nepetalactones in leaves of cultivated Nepeta rtanjensis plants. *Ind. Crops Prod.* **2018**, *117*, 347–358. [CrossRef]
34. Ishiyama, K.; Inoue, E.; Watanabe-Takahashi, A.; Obara, M.; Yamaya, T.; Takahashi, H. Kinetic properties and ammonium-dependent regulation of cytosolic isoenzymes of glutamine synthetase in arabidopsis. *J. Biol. Chem.* **2004**, *279*, 16598–16605. [CrossRef] [PubMed]
35. Nestorović Živković, J. Antioxidative, Antimicrobial and Allelopathic Effects of Three Endemic Nepeta Species (Lamiaceae). Ph.D. Thesis, University of Belgrade, Belgrade, Serbia, 2013.
36. Hirel, B.; Gadal, P. Glutamine synthetase in rice: A comparative study of the enzymes from roots and leaves. *Plant Physiol.* **1980**, *66*, 619–623. [CrossRef] [PubMed]
37. Guan, M.; de Bang, T.C.; Pedersen, C.; Schjoerring, J.K. Cytosolic glutamine synthetase Gln1;2 is the main isozyme contributing to GS1 activity and can be up-regulated to relieve ammonium toxicity. *Plant. Physiol.* **2016**, *171*, 1921–1933. [CrossRef]
38. Lothier, J.; Gaufichon, L.; Sormani, R.; Lemaître, T.; Azzopardi, M.; Morin, H.; Chardon, F.; Reisdorf-Cren, M.; Avice, J.C.; Masclaux-Daubresse, C. The cytosolic glutamine synthetase GLN1;2 plays a role in the control of plant growth and ammonium homeostasis in Arabidopsis rosettes when nitrate supply is not limiting. *J. Exp. Bot.* **2011**, *62*, 1375–1390. [CrossRef]
39. Debouba, M.; Dguimi, H.M.; Ghorbel, M.; Gouia, H.; Suzuki, A. Expression pattern of genes encoding nitrate and ammonium assimilating enzymes in Arabidopsis thaliana exposed to short term NaCl stress. *J. Plant. Physiol.* **2013**. [CrossRef]
40. Logusch, E.W.; Walker, D.M.; McDonald, J.F.; Franz, J.E. Inhibition of plant glutamine synthetases by substituted phosphinothricins. *Plant. Physiol.* **1991**, *95*, 1057–1062. [CrossRef]
41. Forlani, G.; Obojska, A.; Berlicki, Ł.; Kafarski, P. Phosphinothricin analogues as inhibitors of plant glutamine synthetases. *J. Agric. Food Chem.* **2006**, *54*, 796–802. [CrossRef]
42. Ahsan, N.; Lee, D.G.; Alam, I.; Kim, P.J.; Lee, J.J.; Ahn, Y.O.; Kwak, S.S.; Lee, I.J.; Bahk, J.D.; Kang, K.Y.; et al. Comparative proteomic study of arsenic-induced differentially expressed proteins in rice roots reveals glutathione plays a central role during as stress. *Proteomics* **2008**, *8*, 3561–3576. [CrossRef]
43. Gunes, A.; Inal, A.; Alpaslan, M.; Eraslan, F.; Bagci, E.G.; Cicek, N. Salicylic acid induced changes on some physiological parameters symptomatic for oxidative stress and mineral nutrition in maize (*Zea mays* L.) grown under salinity. *J. Plant. Physiol.* **2007**, *164*, 728–736. [CrossRef] [PubMed]
44. Zhang, Q.; Cui, Q.; Yue, S.; Lu, Z.; Zhao, M. Enantioselective effect of glufosinate on the growth of maize seedlings. *Environ. Sci. Pollut. Res.* **2019**, *26*, 171–178. [CrossRef] [PubMed]
45. Savić, J.; Platiša, J.; Dragićević, M.; Nikolić, R.; Mitić, N.; Cingel, A.; Vinterhalter, B. The activity of peroxidases and superoxide dismutases in transgenic phosphinothricin-resistant Lotus corniculatus shoots. *Arch. Biol. Sci.* **2010**, *62*, 1063–1070. [CrossRef]
46. Qian, H.; Chen, W.; Sun, L.; Jin, Y.; Liu, W.; Fu, Z. Inhibitory effects of paraquat on photosynthesis and the response to oxidative stress in Chlorella vulgaris. *Ecotoxicology* **2009**, *18*, 537–543. [CrossRef]
47. Singh, H.P.; Batish, D.R.; Kaur, S.; Arora, K.; Kohli, R.K. α-Pinene inhibits growth and induces oxidative stress in roots. *Ann. Bot.* **2006**, *98*, 1261–1269. [CrossRef]
48. Mutlu, S.; Atici, Ö.; Esim, N.; Mete, E. Essential oils of catmint (Nepeta meyeri Benth.) induce oxidative stress in early seedlings of various weed species. *Acta Physiol. Plant.* **2011**, *33*, 943–951. [CrossRef]
49. Murashige, T.; Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **1962**, *15*, 473–497. [CrossRef]

50. Taipale, R.; Ruuskanen, T.M.; Rinne, J.; Kajos, M.K.; Hakola, H.; Pohja, T.; Kulmala, M. Technical note: 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]
51. Porra, R.J.; Thompson, W.A.; Kriedemann, P.E. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: Verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **1989**, *975*, 384–394. [CrossRef]
52. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef]
53. Dragićević, M.; Todorović, S.; Bogdanović, M.; Filipović, B.; Mišić, D.; Simonović, A. Knockout mutants as a tool to identify the subunit composition of Arabidopsis glutamine synthetase isoforms. *Plant. Physiol. Biochem.* **2014**. [CrossRef] [PubMed]
54. R Core Team. *R: A Language and Environment for Statistical Computing*; Version R 3.5.0; R Foundation for Statistical Computing: Vienna, Austria, 2017; ISBN 3-900051-07-0. Available online: https://www.R-project.org (accessed on 1 May 2018).
55. Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B (Methodol.)* **1964**, *26*, 211–252. [CrossRef]
56. Venables, W.N.; Ripley, B.D. Tree-based methods. In *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; pp. 251–269.
57. Lenth, R.; Buerkner, P.; Herve, M.; LOve, J.; Riebl, H.; Singmann, H. emmeans: Estimated Marginal Means, Aka Least-Squares Means. R Package Version 1.2.2 2018. Available online: https://CRAN.R-project.org/package=emmeans (accessed on 1 July 2018).

biology

MDPI

Article

Rehydration Process in Rustyback Fern (*Asplenium ceterach* L.): Profiling of Volatile Organic Compounds

Suzana Živković [1,*], Marijana Skorić [1,*], Mihailo Ristić [2,†], Biljana Filipović [1], Milica Milutinović [1], Mirjana Perišić [3] and Nevena Puač [3]

[1] Institute for Biological Research "Siniša Stanković"—National Institute of the Republic of Serbia, University of Belgrade, Bulevar despota Stefana 142, 11060 Belgrade, Serbia; biljana.nikolic@ibiss.bg.ac.rs (B.F.); milica.milutinovic@ibiss.bg.ac.rs (M.M.)

[2] Institute for Medicinal Plant Research "Dr Josif Pančić", Tadeuša Košćuška 1, 11000 Belgrade, Serbia

[3] Institute of Physics, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia; mirjana@ipb.ac.rs (M.P.); nevena@ipb.ac.rs (N.P.)

* Correspondence: suzy@ibiss.bg.ac.rs (S.Ž.); mdevic@ibiss.bg.ac.rs (M.S.)

† We are sad to report that the author Mihailo Ristić is deceased.

Simple Summary: Severe environmental changes, such as drought, can delay growth, the development of plants, and induce injury to their tissues. However, a group of land plant species, called resurrection or desiccation-tolerant plants, is able to lose 95% of their cellular water and still remain viable for long periods, resuming full metabolic activity upon rehydration. Recovery from near-complete water loss is complex and requires the coordination of physical and chemical processes in the resurrection plants. Under stress conditions plants also synthesize and release a wide variety of volatile organic compounds with diverse biological and ecological functions. The rehydration process in resurrection rustyback fern (*Asplenium ceterach*) resulted in complete plant recovery within 72 h, accompanied by high emission of volatiles, mainly belonging to the group of fatty acid derivatives. These findings could have significant implications from biotechnological and ecological perspectives since the rustyback fern has been recently recognized as a valuable source of bioactive compounds.

Abstract: When exposed to stressful conditions, plants produce numerous volatile organic compounds (VOCs) that have different biological and environmental functions. VOCs emitted during the rehydration process by the fronds of desiccation tolerant fern *Asplenium ceterach* L. were investigated. Headspace GC–MS analysis revealed that the volatiles profile of rustyback fern is mainly composed of fatty acid derivatives: isomeric heptadienals (over 25%) and decadienals (over 20%), other linear aldehydes, alcohols, and related compounds. Aerial parts of the rustyback fern do not contain monoterpene-type, sesquiterpene-type, and diterpene-type hydrocarbons or corresponding terpenoids. Online detection of VOCs using proton-transfer reaction mass spectrometry (PTR–MS) showed a significant increase in emission intensity of dominant volatiles during the first hours of the rehydration process. Twelve hours after re-watering, emission of detected volatiles had returned to the basal levels that corresponded to hydrated plants. During the early phase of rehydration malondialdehyde (MDA) content in fronds, as an indicator of membrane damage, decreased rapidly which implies that lipoxygenase activity is not stimulated during the recovery process of rustyback fern.

Keywords: *Asplenium ceterach*; rehydration; volatile organic compounds; lipid peroxidation

check for updates

Citation: Živković, S.; Skorić, M.; Ristić, M.; Filipović, B.; Milutinović, M.; Perišić, M.; Puač, N. Rehydration Process in Rustyback Fern (*Asplenium ceterach* L.): Profiling of Volatile Organic Compounds. *Biology* **2021**, *10*, 574. https://doi.org/10.3390/biology10070574

Academic Editor: Jeffrey G. Duckett

Received: 25 May 2021
Accepted: 17 June 2021
Published: 23 June 2021

1. Introduction

Certain plant species, termed desiccation tolerant or resurrection plants, have evolved the remarkable ability to withstand extreme dehydration (to just 10% of their water content or less) and resume normal metabolic and physiological activity after rehydration of vegetative tissues without cell damage. Such rapid and ecologically beneficial changes allow a plant to survive equilibrium with 0% air humidity until water becomes available.

Upon re-watering, dried resurrection plants quickly revive and become fully photosynthetically active within 24 h [1]. Furthermore, drying and rehydration processes cause only limited damage to resurrection plant tissues, due to a number of morphological, physiological, biochemical and genetically different mechanisms, developed not only for diminishing damages suffered during severe water loss, but also during rehydration [2]. Controlled regulation of physical and metabolic processes enables the minimization of the stress associated with desiccation and allows full recovery once the plant is rehydrated. These adaptations apparently separate resurrection plants from desiccation-sensitive plant species [3]. The desiccated state is correlated with multiple obstacles at the cellular level, such as photo-oxidative stress caused by reactive oxygen species, the metabolic requirements of resurrection, and the mechanical stress of cell and tissue deformation [4,5].

Under stress conditions plants synthesize and release a wide variety of volatile organic compounds (VOCs) with diverse biological and ecological functions. Plant volatiles are usually complex mixtures of diverse organic compounds, including saturated and unsaturated hydrocarbons, esters, aldehydes, ketones, amines, oxides and sulfur compounds, derived from various biochemical pathways [6]. Fatty acid derivatives, including C_6 green leaf volatiles and their esters, may derive from enzymatic or non-enzymatic reactions [7]. The first response of plants to any environmental changes is closely related to cell membrane structures. Fast and nonspecific response of the membrane is based on the transformation of the cell membrane structural components to signaling compounds. Polyunsaturated fatty acids (PUFAs) are incorporated in cell membranes and together with enzyme lipid peroxidase (LOX) through a series of chemical reactions that give rise to a great variety of products which represents the "nonspecific biological signals" and do not require preceding activation of genes. They are produced as responses to environmental stresses and/or stimuli without requiring any specific gene expression, or long downstream signaling cascades to evoke them [8,9]. Maintenance of membrane integrity is of critical importance to ensure survival upon cellular dehydration. Desiccation stress can result in lipid destruction and membrane damage due to free radical production. It is well known that products of lipid peroxidation (LP) such as malondialdehyde (MDA) are often used as a marker for oxidative stress in plants [10]. The majority of studies on resurrection plants are focused on their strategies to cope with desiccation damage during dehydration, and less attention has been given to the process of rehydration, although the mechanisms for preventing and/or repairing cell damage upon rehydration are of great importance for the desiccation tolerance of resurrection plants [11].

Rustyback fern (*Asplenium ceterach* L.) belongs to the resurrection species and is widespread in Western and Central Europe, including the Mediterranean region. The adult fern (sporophyte) is a perennial herbaceous plant with leathery fronds (species name originates from the dark brown and densely scale-covered lower surface of fronds), which grows in limestone rock crevices and stone walls and may survive long dry periods between wet spells, passing quickly from anabiosis to full biological activity. Although *A. ceterach* belongs to a group of poikilohydric fernsand could sustain different rates of desiccation, recovering uninjured from complete dryness [12], this desiccation tolerant plant still prefers shaded sites with excellent drainageand suitable humid conditions (especially high air humidity) in its habitat, while becoming quiescent when water is unavailable.

Detailed phytochemical analysis of *A. ceterach* and other species from the family Aspleniaceae have been performed recently [13–17]. Froissard et al. [13] reported lipid derivatives as important volatile compounds in *A. ceterach*. The experiments presented here were aimed to study VOCs emission from the rustyback fern sporophyte following transition from dormant (desiccated) to an active, rehydrated state. This was achieved by using coupled Headspace GC–MS and online PTR–MS analysis of the VOCs. Since the first response of plants to any environmental changes is closely related to cell membrane structures, we have also postulated that the rehydration process in *A. ceterach* would disrupt maintenance of membrane integrity. Therefore, we have measured the changes in the lipid

status of the fern fronds during rehydration as evidence for the involvement of LOX in the plant cell membrane damage and/or formation of VOCs.

2. Materials and Methods

2.1. Plant Material

Dormant (desiccated) mature sporophytes of rustyback fern (*Asplenium ceterach* L.) were collected in East Serbia near the Monastery Gornjak (44°15′52.66″ N 21°32′40.21″ E) and stored in paper bags at room temperature until use. Species was authenticated by the authors and the corresponding voucher specimens have been deposited at the Department of Plant Physiology, Institute for Biological Research "Siniša Stanković"—National Institute of the Republic of Serbia, University of Belgrade, Serbia.

2.2. Headspace GC–MS Analysis of Volatiles

Headspace GC–MS technique was used to identify and quantify volatile organic compounds (VOCs) present in the injected headspace sample. Rustyback fern fronds (dry and fresh) were placed in a closed sampling vessel and deionized water was added. Samples were heated at 80 °C and the vapor in the vessel was sampled for analysis by using a heated gas-tight syringe (Agilent Technologies, Santa Clara, CA, USA). The GC–MS analyses were performed on a Hewlett Packard G1800C-GCD Series II apparatus equipped with a HP-5MS column (30 m × 0.25 mm, 0.25 μm film thickness). Carrier gas was helium (1 mL·min^{-1}) and the transfer line was heated to 260 °C. The mass spectra were acquired in EI mode (70 eV) in the *m/z* range of 40–400. Identification of individual components was accomplished by comparison of retention times with standard substances and by matching mass spectral data with those held in the Wiley 275 and NIST libraries of mass spectra. Confirmation was performed by using AMDIS software (AMDIS ver.2.1., National Institute of Standards and Technology-NIST, Standard Reference Data Program, Gaithersburg, MD, USA) and the available literature data [18].

2.3. Online PTR–MS Measurements of Volatiles

Desiccated rustyback fern sporophytes similar in size were selected and placed into 350 mL glass jars two hours prior to the experiment and allowed to adapt. Rehydration was initiated by adding 50 mL of deionized water into each jar. Measurements of the ambient level of the VOCs in the atmosphere of glass vessels during the rehydration process were conducted continuously during the next 24 h by using Standard Proton Transfer Reaction Quadrupole Mass Spectrometer (PTR–MS, IoniconAnalytik, GmbH, Innsbruck, Austria). The identification and quantification of VOCs by PTR–MS was based on protonated parent ion masses (molecular mass plus one atomic mass unit), yielded in proton transfer reaction with H_3O^+ ions, and the relative abundance of ions was obtained by the use of quadrupole mass spectrometer and secondary electron multiplier. Detailed descriptions of the PTR–MS technology can be found elsewhere [19,20]. In this investigation masses in the range from *m/z* 21 to *m/z* 300 (including five control parameters *m/z* 21, *m/z* 25, *m/z* 30, *m/z* 32, and *m/z* 37) were measured, with dwell time of 200 ms and for the time period of 24 h. Drift tube parameters included: pressure in range from 2.13 to 2.15 mbar; temperature 60 °C; voltage 600 V; E/N parameter 145 Td and reaction time 90 μs. The count rate of H_3O^+/H_2O was 2.1 to 14.1% of the count rate of H_3O^+ ions, which was in the range 6.1×10^6 to 9.7×10^6 counts^{-c}. Measurements were conducted on five independent plants. In order to avoid the possible deviations in experimental data caused by the changes in water content of plant tissue during rehydration, all results are normalized to the dry mass values (DW).

2.4. Determination of Relative Water Content

Plant material for measurements of relative water content (RWC) and LP was maintained in the same conditions as described for PTR–MS experiments and separately prepared from three independent plants each at different points of rehydration. RWC was

determined in detached fronds at specific time intervals during rehydration process (every two hours) and calculated by using the equation.

$$\text{RWC (\%)} = [(\text{FW}-\text{DW})/(\text{TW}-\text{DW})] \times 100 \quad (1)$$

FW-fresh weight of fronds was measured immediately after sampling and then the fronds were left in distilled water for the next 24 h at 25 °C in the dark (TW-turgid weight). DW-dry weight was determined after lyophilization of fronds for 24 h (Lyovac GT2, SRK-Systemtechnik GmbH, Riedstadt, Germany).

2.5. Measurement of Lipid Peroxidation

LP was determined by MDA or thiobarbituric acid-reactive-substances (TBARS) assay as described by Hodges et al. [21] with some modifications. Briefly, 0.2g of fern fronds were homogenized in 5 mL of 80% ethanol and then centrifuged at 12,000× *g* for 20 min (Thermo Scientific Heraeus Biofuge Stratos, Heraeus Holding GmbH, Hanau, Germany). A 1-mL aliquot of sample extract was mixed with 1 mL of either (i)−TBA solution comprised of 20% (*w/v*) trichloroacetic acid and 0.01% butylated hydroxytoluene or (ii)+TBA solution containing the above plus 0.65% TBA. Samples were then mixed vigorously, heated at 95 °C in water bath for 25 min, and then quickly cooled on ice. After centrifugation at 12,000× *g* for 20 min, the absorbance of the supernatant was recorded at 440 nm, 532 nm and 660 nm (HP Agilent 8453 spectrophotometer, Agilent Technologies, Waldbronn, Germany). MDA equivalents were calculated by using the following equations.

$$[(A_{532+TBA} - 3A_{600+TBA}) - (A_{532-TBA} - A_{600-TBA})] = A \quad (2)$$

$$[(A_{440+TBA} - A_{600+TBA}) \times 0.0571] = B \quad (3)$$

$$\text{MDA equivalents (nmol mL}^{-1}) = [(A - B)/157000] \times 10^6 \quad (4)$$

Results were expressed as MDA equivalents per g of dry weight (nmolg^{-1} DW). Three biological replicates of each treatment were used for evaluation.

2.6. Statistical Analysis

Statistical analyses were performed by using STATGRAPHICS software, v. 4.2 (Statgraphics Technologies, Inc., The Plains, VA, USA). The data were subjected to the analysis of variance (ANOVA), and the comparisons between the mean values of treatments were made by the least significant difference (LSD) test calculated at the confidence level of $p < 0.05$. Linear regression analysis was performed; correlation coefficients (R) and coefficients of determination (R^2) were calculated by using Statisticav. 10 software (StatSoft. Inc. 2011, Tulsa, OK, USA).

3. Results and Discussion

3.1. Relative Water Content Analysis

Resurrection plants have the remarkable ability to survive extreme loss of water (desiccation) while staying dormant for a long period of time. Once the water becomes available again, the rehydration process starts and plants revive quickly and regains full metabolic activity in a few hours or days. In order to tolerate desiccation, resurrection plants must be able to limit the possible damage associated with drying in order to maintain physiological integrity during dehydration and to activate specific mechanisms upon rehydration in order to repair the damage caused during desiccation and subsequent rehydration [22]. Desiccation tolerant rustyback fern (*A. ceterach*) is capable of shifting from an active metabolic to anabiotic state, and vice versa, several times during its life cycle without permanent structural and/or functional damage [22]. Morphological changes in the form of frond folding and in-curling represent one of the main responses to desiccation stress. In the present study dormant sporophytes of rustyback fern (RWC 6%) were rehydrated in glass vessels until they regained full turgid state (RWC 82%) after 72 h of

rehydration (Figure 1). The addition of water to desiccated ferns triggered the resurrection process and resulted in a gradual increase in RWC, accompanied by the unfolding of fronds. The rustyback fern sporophytes regained 40% of their RWC after 12 h of re-watering. The rehydration speed in the first 12 h was ~3% per hour and then the process slowed down slightly (between 12 h and 20 h of re-watering). One more boost in rehydration speed was observed in the last 4 h of the 24-h cycleand, afterwards, the process slowed down significantly (~0.3% per hour).

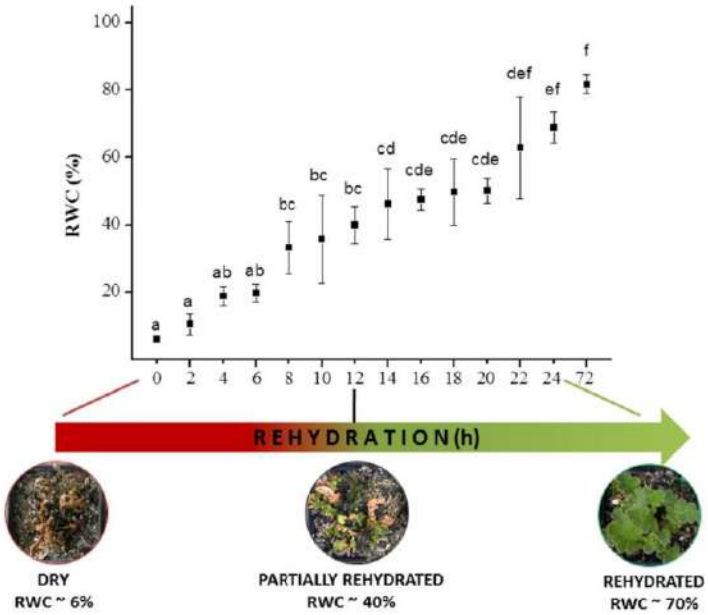


Figure 1. Rehydration response of rustyback fern (*A. ceterach*). Relative water content (RWC%) was determined during 72 h at the indicated time intervals and presented as the mean value of three independent measurements ± SE. Values with the same letter are not significantly different at the $p < 0.05$ level according to the LSD test. Three distinct phases of rehydration are shown at the bottom and correspond to dry plants at the beginning of experiment (0 h) and partly rehydrated (after 12 h) and rehydrated plants (24 h following rehydration).

It should be noted that the recovery time of dry resurrection plants during rehydration varies widely, depending greatly on plant size and age, methods of rehydration, drying rate, desiccation extent, and duration before rehydration [23,24].Indeed, during the measurements of experimental parameters presented here, light conditions have changed with the day/night rhythm and, therefore, when the jars were illuminated, the temperatures inside were slightly higher than in the surroundings (up to 5 °C). Moreover, due to water evaporation and transpiration of the plants during the experiment, the humidity inside the jars was quite high (reaching almost 100%), which most probably affected the course of the experiment and behavior of the plants. Nevertheless, these light/temperature variations are in accordance with the frequent environmental changes in the natural habitat of *A. ceterach*, especially during the rainy season.

Similar results were obtained for resurrection plants from the family Gesneriaceae, such as *Haberlea rhodopensis* and/or species within the genus *Ramonda*, of which the latter often shares the same natural habitat with *A. ceterach* [25]. Slow water uptake during the initial hours of rehydration process could be considered as an adaptive defense mechanism

for avoiding cellular damages by rapid water uptake upon rehydration [2]. Rakić et al. [26] stated that, at the beginning of rehydration, desiccation recovering plants go through an unstable and vulnerable short period, but then regain complete stability after 6 days of rehydration. Completely dried resurrection plants *Boea hygrometrica* restored their RWC very quickly upon rehydration, reaching 50% and 95% RWC within only 12 h and 24 h, respectively [27].

3.2. Emission of Volatiles during Rehydration

Plants emit a wide variety of VOCs, mostly lipophilic compounds with low molecular weight, which could easily cross cellular membranes and be released into the surrounding environment [28]. According to their biosynthetic origin, VOCs can be divided into several major groups: terpenoids, phenylpropanoids/benzenoids, fatty acid derivatives and amino acid derivatives [29]. VOCs are engaged in plant growth and protection and their emissions are strongly dependent on the environmental conditions and developmental stages of the plant tissue. Plants exposed to abiotic and biotic stresses emit numerous compounds from almost all vegetative parts. Among the volatiles emitted from stressed plants, the enhancement of LOX pathway volatiles, which is a mixture of various C_6 aldehydes and alcohols and their derivatives, represents atypical reaction. Many of these compounds are synthesized from the degradation of cellular structures and are used as indicators of cellular wall degradation or membrane denaturation [30]. The accessibility of Headspace GC–MS sampling together with PTR–MS analysis enabled research and examination of the changes in temporal and spatial emissions of VOCs [30–32]. Within the 40 compounds detected in rustyback fern during rehydration by Headspace GC–MS analysis, a pool of 30 volatiles was identified which represents 94.6% of the total VOCs composition (Figure 2, Table 1).

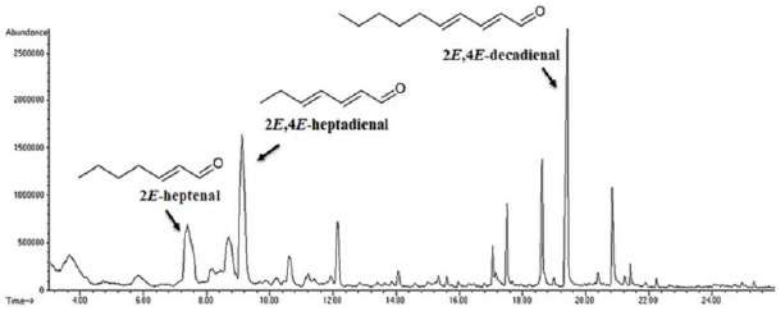


Figure 2. Headspace GC–MS chromatogram of VOCs detected in rustyback fern (*A. ceterach*). Peaks corresponding to the most abundant compounds are labeled.

The volatile pattern of *A. ceterach* was dominated by isomeric heptadienals (>25%): (*E*,*E*)-2,4-heptadienal (17.66%) and (*E*,*Z*)-2,4-heptadienal (8.43%), followed by decadienals (>20%): (*E*,*E*)-2,4-decadienal (15.03%) and (*E*,*Z*)-2,4-decadienal (5.68%). Significant amounts of other fatty acid derivatives (alcohols and aldehydes) were also found, i.e., (2*E*)-heptenal (11.54%; intense green fatty odor), (*E*)-2-undecenal (4.51%; fruity waxy odor), 1-hexen-3-ol (4.48%; ethereal rum-like odor), n-nonanal (4.24%; hay odor), 3-octen-1-ol (3.05%; fatty fruity odor), (*E*)-2-decenal (2.81%; waxy fatty odor), 1-Octen-3-ol (2.37%; sweet, mushroom like odor), and n-heptanal (2.32%; bitter odor). On the other hand, aerial parts of rustyback fern showed a markedly lack of green odor C_6-compounds, as well as monoterpene-type, sesquiterpene-type, and diterpene-type of hydrocarbons and/or corresponding terpenoids. Similarly, Froissard et al. [13] reported that VOCs profile of *A. ceterach* is dominated (77.4%) by lipid derivatives, mainly (*E*)-2-decenal, nonanal and (*E*)-2-heptenal, while significantly lower content of shikimic compounds (21.3%) and carotenoid derivatives (0.8%) was detected. The 1-octen-3-ol has also been reported from many mushrooms [33,34], ferns [35,36], and angiosperms [37]. (2*E*)-decenal, a natural plant and mushroom VOC

with a plastic fatty odor, was also previously detected in some ferns (*Adiantum capillus veneris*, *Blechnumspicant*, and *Asplenium trichomanes*) [35,36]. In addition, (2*E*)-decenal was an abundant component of the strong "stink bug" scent together with (2*E*,4*Z*)-decadienal and (2*E*,4*E*)-decadienal [38]. These VOCs were also found in *Equisetum ramosissimum* and *E. scirpioides* together with the (2*E*)-heptenal, while (2*E*,4*Z*)-heptadienal and (2*E*,4*E*)-heptadienal were detected in *E. scirpioides* and *E. hyemale* [39]. It has been reported that n-heptanal, 2,4-decadienal, and 2,4-heptadienal exhibit fishy odors [40] and represent the main metabolic products of some algal species [41].

2,4-decadienal has been known, for a long time, as a product of deteriorated fat compounds [42] and as a widespread volatile constituent of dry fruits [43]. Similar compounds are produced by higher flowering plants and are believed to have a significant role in plant defense by acting as chemical attractants, alarm signals against herbivore attack, and/or protective compounds [44]. In addition, these polyunsaturated aldehydes have been reported to interfere with the reproductive success of some marine invertebrates [45]. One must bear in mind that 2,4-unsaturated aldehydes are chemically highly reactive. The generation of 2,4-decadienal represents the fast response of cells to the changes in their membrane composition and, unlike other signaling compounds (ethylene, superoxide, jasmonic acid, salicylic acid, etc.), does not require a preceding activation of gene expression. Therefore, the products of the oxidative membrane LP constitute "biological signals", which produce nonspecific responses to a large variety of environmental stresses [46].

PTR–MS has become a commonly utilized technique for the analysis of trace amounts of VOCs and it offers many advantages over other conventional analytical methods. This is an online and non-invasive method with high sensitivity (parts per trillion/parts per billion concentrations) for plant VOCs assessment in real time at high throughput [31,47,48]. The main drawback of this technique is related to providing information limited to protonated molecular mass, which is not a specific indicator of chemical identity. Furthermore, the identification of the compounds is further complicated by the overlapping spectra of different VOCs species. In this study a chromatographic step (Headspace GC–MS characterization of volatiles in the atmosphere of glass jars), enabled the selection and identification of specific VOCs in rustyback fern fronds, and subsequently their dynamics during the process of rehydration was monitored by PTR–MS. PTR–MS and GC–MS, as complementary techniques for the analysis of volatiles which enable rapid quantification of selected substances [32].

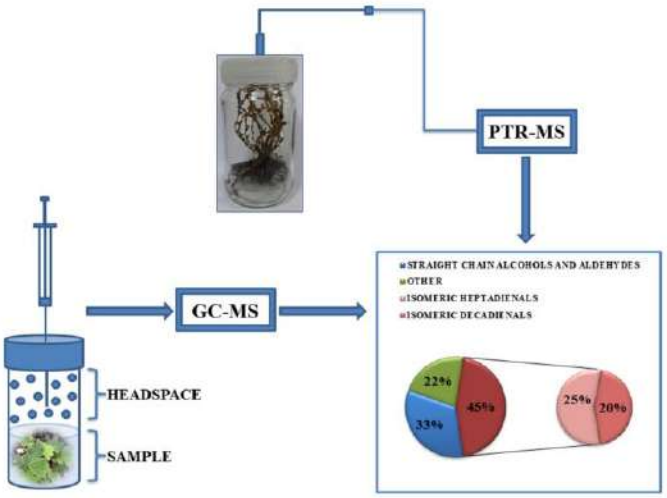
PTR–MS real-time detection and analysis of VOCs emission from the rehydrating rustyback fern sporophytes was performed by using desiccated (dormant) plants. The measurements started with the initiation of the rehydration process; that is, when the water was added to the plants (Scheme 1).

PTR–MS online detection of VOCs (as was previously shown by Headspace GC–MS analysis) revealed that the rehydration process enhanced the isomeric dienals emission rates. There was a striking signal at *m/z* 81 that was assigned to both 2,4-heptadienals and 2,4-dicadienals, which was apparently elevated at the onset of rehydration process, reaching 59.88 ppbV during the first hour and then declining progressively with the further increments of RWC (Figure 3a). The emission intensity of other fragments belonging to the isomeric alkadienals, with *m/z* 53 and *m/z* 110, followed the same trend and reached 21.58 or 13.01 ppbV, respectively. Furthermore, intensity levels of selected fragments were significantly correlated with *m/z* 81 emission ($R^2 = 0.99588$ and $R^2 = 0.7781$, respectively). Approximately 12 h after the onset of experiments, when the recorded RWC of the plant tissue was 40% (Figure 1), the concentration of isomeric dienals decreased and returned to the basal levels. Similar trends were also observed for the other targeted compounds.

Table 1. Volatile organic compounds (VOCs) composition of rustyback fern (*A. ceterach*) as revealed by Headspace GC–MS analysis.

No.	Constituent	KIE	KIL	t_R	% (m/m) [1]
1	Hex-1-ene-3-ol	829.5	832	3.66	4.48
2	n-Heptanal	903.8	901	5.88	2.32
3	(2*E*)-Heptenal	954.6	947	7.40	11.54
4	1-Octen-3-ol	980.0	974	8.16	2.37
5	3-Octanone	985.1	979	8.31	0.90
6	2-Pentyl furan	988.8	984	8.42	1.13
7	(2*E*,4*Z*)-Heptadienal	1000.1	988	8.76	8.43
8	(2*E*,4*E*)-Heptadienal	1010.7	1005	9.13	17.66
9	(2*E*)-Octen-1-al	1043.1	1034	10.23	0.96
10	Oct-3-ene-1-ol	1054.2	1044	10.61	3.05
11	n-Octanol	1071.2	1063	11.19	1.32
12	(2*E*)-Octen-1-ol	1078.5	1067	11.44	0.71
13	n-Undecane	1092.4	1100	11.93	1.40
14	n-Nonanal	1099.1	1100	12.15	4.24
15	(2*E*)-Nonen-1-al	1154.2	1157	14.07	0.89
16	n-Dodecane	1190.9	1200	15.35	0.58
17	n-Decanal	1198.3	1201	15.61	0.41
18	g-Octalactone	1241.2	1250	17.05	1.55
19	n.i.	1244.2	/	17.15	1.06
20	(2*E*)-Decenal	1254.8	1260	17.51	2.81
21	(2*E*,4*Z*)-Decadienal	1287.4	1292	18.62	5.68
22	n-Undecanal	1299.0	1305	19.00	0.42
23	(2*E*,4*E*)-Decadienal	1311.6	1315	19.42	15.03
24	n.i.	1342.1	/	20.40	0.70
25	(2*E*)-Undecenal	1355.8	1357	20.85	4.51
26	n.i.	1367.8	/	21.23	0.39
27	n.i.	1373.3	/	21.41	0.79
28	n-Dodecanal	1398.8	1408	22.24	0.30
29	n-Tridecanal	1499.5	1509	25.33	0.21
30	(6*Z*)-Pentadecen-2-one	1655.5	1667	29.79	0.16
31	(2*E*)-Tridecenol acetate	1701.1	1703	31.07	0.18
32	Hexadecanoic acid	1965.8	1959	37.75	0.30
33	n.i.	2062.5	/	40.02	0.23
34	Linoleic acid	2146.9	2132	41.89	0.77
35	n.i.	2153.2	/	42.02	0.32
36	n.i.	2282.9	/	44.79	0.28
37	n.i.	2455.6	/	48.25	0.43
38	n.i.	2460.5	/	48.35	0.66
39	Squalene	2808.6	2814	54.74	0.31
40	n.i.	3062.7	/	59.96	0.52
	Number of detected constituents			40	100%
	Number of identified constituents			30	94.61%

[1] KIE—Kovats (retention) index experimentally determined (AMDIS, uncorrected); KIL—Kovats (retention) index—literature data [17]; t_R—retention time (min); n.i.—not identified; % (m/m)—percentage of component emitted (mass on mass).



Scheme 1. A number of peaks corresponding to the protonated masses of the individual VOCs may be observed, including a smaller contribution of VOCs' fragments and clusters. Nonetheless, based on the GC–MS analysis, peaks which correspond to *m/z* 41, 43, 53, 70, 81, 83, 110 and 128 were of main interest.

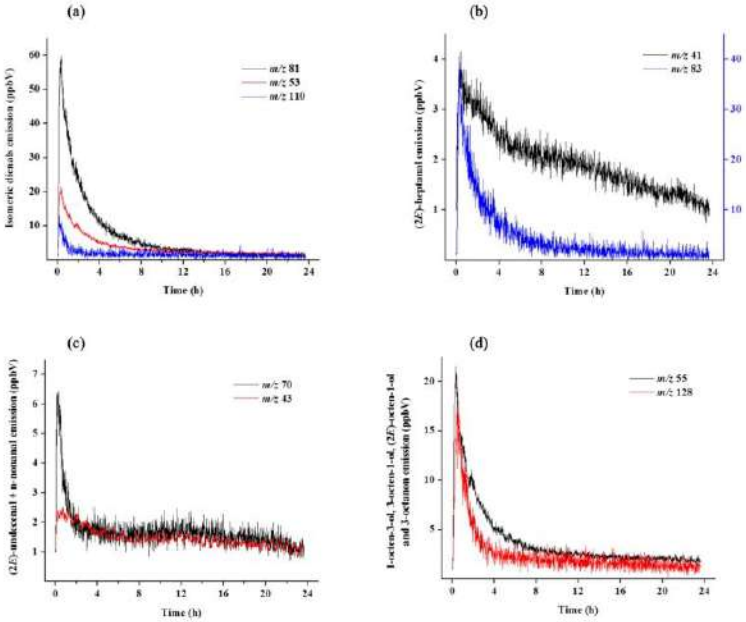


Figure 3. Time course of VOCs emission from rustyback fern (*A. ceterach*) fronds during the process of rehydration detected on-line by the PTR–MS method. Different colors indicate different ions corresponding to distinct volatiles: (**a**) Isomeric dienals, *m/z* 81 + *m/z* 53 + *m/z* 110; (**b**) (2*E*)-heptenal, *m/z* 83 + *m/z* 41; (**c**) (2*E*)-undecenal + n-nonanal, *m/z* 70 + *m/z* 43; (**d**) 1-octen-3-ol + 3-octen-1-ol + (2*E*)-octen-1-ol + 3-octanon, *m/z* 55 + *m/z* 128. Data shown represent mean sequence of plants from five independent experiments normalized to DW values.

The signals at *m/z* 83 and *m/z* 41 were attributed to 2*E*-heptenal, which belongs to the group of monounsaturated fatty aldehydes. The amount of compound with signal at *m/z* 83 emitted by rustyback fern fronds was increasing over the rehydration process with a correlation between distinct fragments of $R^2 = 0.81391$. The highest intensity of the *m/z* 83 signal was detected within the first hour of rehydration (up to 40.57 ppbV), with an almost complete absence of the signal 12 h after the start of measurement. Still, *m/z* 41 fragment showed a 10-fold lower emission rate (4.15 ppbV) compared to *m/z* 83, but remained significantly abundant towards the end of the experiment (Figure 3b). Emission of *m/z* 70 (6.44 ppbV) and *m/z* 43 (2.54 ppbV) was attributed to (2*E*)-undecenal and saturated fatty aldehyde n-nonanal, respectively (Figure 3c).

A wide variety of VOCs with the Mr 128 was recorded in rustyback fern sporophytes during the rehydration process, such as 1-octen-3-ol, 3-octen-1-ol, (2*E*)-octen-1-ol, and 3-octanon. Emissions of *m/z* 128 and *m/z* 55, selected as fragments present in all listed compounds, were 19.48 ppbV and 21.58 ppbV, respectively, with the significant correlation value of $R^2 = 0.93138$ (Figure 3d). The emission levels of ions corresponding to all the above mentioned compounds were increasing over the experimental time, with the maximum values reached during the first two hours of re-watering, as has already been outlined for other detected volatiles. However, the possible contribution of hydronium water cluster H_3O^+ $(H_2O)_2$ to the *m/z* 55 fragment intensity should not be neglected [49].

(*E*)-2-heptenal is a well-known and important volatile aldehyde formed during linoleic acid oxidation. Lee and Min [50] reported that, among other compounds, (*E*)-2-heptenal, 1-octen-3-ol, 2,4-heptadienal, 2-octen-1-ol, 2,4-octadienal, and 2,4-decadienal were identified as volatile oxidation compounds (VOOs) in chlorophyll mixed with linoleic acid under light storage. In addition, (*E*)-2-heptenal was detected as the main VOOs in olive oil exposed to light irradiation [51], while grape, flax, and black cumin seed oils had high content of (*E*)-2-heptenal, 2,4-hexadienal, and 2,4-heptadienal throughout the photo-oxidation conditions [52]. (2*E*)-undecenal occurs naturally in coriander leaves [53] and red pepper fruits [54]. Nonanal could be found in a large quantity in different plant species, such as tomato, canola, soybean, etc., and it has been reported to show antimicrobial activity against various bacterial and fungal pathogens [55]. Nonanal, together with 1-hexanal, *cis*-3-hexenol, and methyl jasmonate, belongs to the fatty acid derivative class of plant VOCs, which arise from C_{18} unsaturated fatty acids, linoleic or linolenic [29].1-octen-3-ol was mostly found to be released by pathogenic or endophytic fungi [56]. However, recent studies have shown that this compound is also emitted from a number of legume species [57]. Consequently, nonanal, 3-octanone, and 1-octen-3-ol were designated as main compounds responsible for the off-odor of infected strawberry fruit [58].

3.3. Estimation of Lipid Peroxidase Activity

The determination of membrane lipid peroxidation is often based on the measurements of MDA, as a product of unsaturated fatty acid peroxidation thatis commonly used as a biomarker of oxidative lipid injuries [59,60]. Lower MDA content indicates less oxidative damage, and was described as an indicator of more efficient stress tolerance. LP level, expressed as MDA content, in desiccated fronds of *A. ceterach* was ~80 nmol g^{-1} DW and sharply declined during the initial six hours of rehydration (Figure 4). Between 8 h and 24 h of rehydration, the MDA level gradually decreased to ~5 nmolg^{-1} DW, reaching the control values.

Desiccation stress in plants can result in lipid destruction and membrane damage due to free radical production. In plants under abiotic stress (e.g., drought), the enhanced formation of ROS was noted, which could easily oxidize PUFAs. PUFAs are lipid components in plant membranes, and their reaction with ROS triggers serial LP [59]. Peroxidation of lipid membrane is considered as the main damaging effect of ROS. Since plant cell structures rich in PUFAs (cell plasma membranes and chloroplast) are also exposed to the generation of ROS, there is a high risk of LP in these plant organelles. Higher MDA levels at the beginning of rehydration could indicate that oxidative damage of *A. ceterach* fronds

has occurred during desiccation process to some extent. Niinemets et al. [61] reported the emission burst of LOX volatiles upon re-watering in desiccation tolerant epiphytic filmy ferns from the family Hymenophyllaceae, suggesting a fast enhancement of ROS production. Similarly, LP level, expressed in terms of lipid hydroperoxide (LOOH) content, increases as the fronds of resurrection fern *Pleopeltis polypodioides* dehydrates, and rapidly decreases after rehydration [62]. The decrease in lipid peroxidation during rehydration was detected in fronds of *Selaginella brachystachya* [63], *S. tamariscina* [64] and *S. bryopteris* [65]. Similar patterns of change of LP were also revealed during the recovery of angiosperm *Haberlea rhodopensis* [66]. Conversely, in angiosperm *Paraisometrum mileense*, MDA levels were maintained close to the control values in both completely dehydrated and well rehydrated leaves [67].

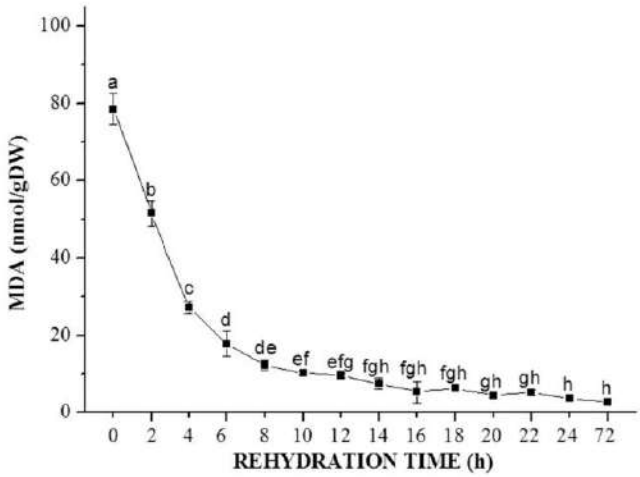


Figure 4. Determination of lipid peroxidation in fronds of rustyback fern (*A. ceterach*) during the process of rehydration. Results of MDA assay are presented as the means of three independent measurements ± SE. Values with the same letter are not significantly different at the $p < 0.05$ level according to the LSD test.

Results from studies on desiccation tolerance mechanism in resurrection fern *Adiantum raddianum* indicated that significant damage was induced on membrane integrity due to desiccation stress [68]. Furthermore, desiccation appeared to cause some damage in the cellular membranes of *Adiantum latifolium*, but this process was reversed and/or repaired upon rehydration to control values, suggesting that the antioxidant system was efficient in scavenging ROS, thus helping the plant to recover from the oxidative stress [69]. These findings, together with our results, support the thesis that much stronger volatile responses might be expected in homoiochlorophyllous species that maintain pigments and photosynthetic activity through desiccation. Although elevated MDA level in desiccated fronds of *A. ceterach* might indicate the oxidative degradation of PUFA residues of the membrane lipids and putative disturbed membrane integrity, we can speculate that repair mechanisms of this resurrection fern permitted reversible changes in the peroxidation of the membrane lipids and the rapid regaining of membrane configuration upon rehydration. Our results demonstrate the complete recovery of *A. ceterach* sporophytes after losing more than 90% of cellular water content.

Relatively high production of VOCs that are potentially associated with the considerable amounts of polyunsaturated aldehydes in rustyback fern sporophytes could have significant implications from biotechnological and ecological perspectives, since these plants could be taken into account as a valuable source for these bioactive compounds.

Further research on the lipid metabolism in rustyback fern would provide better insight into the mechanisms that protect the integrity of membrane lipids of the cell, including thylakoid lipids. Considering that restoration of vital functions and metabolic activities represent an integral part of the entire phenomenon of desiccation tolerance, this could highlight the importance of lipids and membrane-protecting compounds in *A. ceterach* during dehydration/rehydration process, and it will be the main focus of our future research.

4. Conclusions

Desiccation tolerant rustyback fern (*Asplenium ceterach*) revive upon watering and restores physiological activity within the 72 h. Rehydration was characterized with different metabolic changes which led to the complete recovery from a dry state, including increased emission of volatiles mainly belonging to the fatty acid derivatives. Lipoxygenase activity was not stimulated during the rehydration according to the decreased level of MDA as an indicator of membrane damage. Further research on lipid metabolism in rustyback fern would provide better insight into the mechanisms that protect the integrity of the cell membranes during dehydration/rehydration cycle.

Author Contributions: Conceptualization, S.Ž., M.S., M.P. and N.P.; investigation, S.Ž., M.S., M.R., B.F., M.M., M.P. and N.P.; writing—original draft preparation, S.Ž. and M.R.; writing—review and editing, S.Ž., M.S., M.M., B.F., M.P. and N.P.; visualization, S.Ž., M.S., and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Serbian Ministry of Education, Science and Technological Development of the Republic of Serbia, grant numbers III41011, OI173024, 451-03-68/2020-14/200007, 451-03-68/2020-14/200053, and through the grant to the Institute of Physics Belgrade. The APC was funded by the Serbian Ministry of Education, Science and Technological Development of the Republic of Serbia, grant number 451-03-68/2020-14/200007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bernacchia, G.; Salamini, F.; Bartels, D. Molecular characterization of the rehydration process in the resurrection plant *Craterostigma plantagineum*. *Plant Physiol.* **1996**, *111*, 1043–1050. [CrossRef]
2. Georgieva, K.; Mihailova, G.; Velitchkova, M.; Popova, A. Recovery of photosynthetic activity of resurrection plant *Haberlea rhodopensis* from drought- and freezing-induced desiccation. *Photosynthetica* **2020**, *58*, 911–921. [CrossRef]
3. Beckett, M.; Loreto, F.; Velikova, V.; Brunetti, C.; Di Fernando, M.; Tattini, M.; Calfapietra, C.; Farrant, J.M. Photosynthetic limitations and volatile and non-volatile isoprenoids in the poikilochlorophyllous resurrection plant *Xerophyta humilis* during dehydration and rehydration. *Plant Cell Environ.* **2012**, *35*, 2061–2074. [CrossRef] [PubMed]
4. Farrant, J.M.; Brandt, W.; Lindsey, G.G. An overview of mechanisms of desiccation tolerance in selected angiosperm resurrection plants. *Plant Stress* **2007**, *1*, 72–84.
5. Oliver, M.J.; Farrant, J.M.; Hilhorst, H.W.; Mundree, S.; Williams, B.; Bewley, J.D. Desiccation tolerance: Avoiding cellular damage during drying and rehydration. *Annu. Rev. Plant Biol.* **2020**, *71*, 435–460. [CrossRef] [PubMed]
6. Rosenkranz, M.; Schnitzler, J.-P. Plant Volatiles. In *eLS*; John Wiley & Sons, Ltd.: Chichester, UK, 2016. [CrossRef]
7. Jardine, K.; Abrell, L.; Kurc, S.A.; Huxman, T.; Ortega, J.; Guenther, A. Volatile organic compound emissions from *Larrea tridentata* (creosotebush). *Atmos. Chem. Phys.* **2010**, *10*, 12191–12206. [CrossRef]
8. Spiteller, G. The relationship between changes in the cell wall, lipid peroxidation, proliferation, senescence and cell death. *Physiol. Plant.* **2003**, *119*, 5–18. [CrossRef]
9. Foyer, C. Oxygen metabolism and electron transport in photosynthesis. In *The Molecular Biology of Free Radical Scavenging System*; Scandalios, J., Ed.; Cold Spring Harbor Laboratory Press: New York, NY, USA, 1997; pp. 587–621.
10. Garg, N.; Manchanda, G. ROS generation in plants: Boon or bane? *Plant. Biosyst.* **2009**, *143*, 81–96. [CrossRef]
11. Cooper, K.; Farrant, J.M. Recovery of the resurrection plant *Craterostigmawilmsii* fromdesiccation: Protection versus repair. *J. Exp. Bot.* **2002**, *53*, 1805–1813. [CrossRef]

12. Schwab, K.B.; Schreiber, U.; Heber, U. Response of photosynthesis and respiration of resurrection plants to desiccation and rehydration. *Planta* **1989**, *177*, 217–227. [CrossRef]
13. Froissard, D.; Rapior, S.; Bessière, J.-M.; Buatois, B.; Fruchier, A. Asplenioideae Species as a Reservoir of Volatile Organic Compounds with Potential Therapeutic Properties. *Nat. Prod. Commun.* **2015**, *10*, 1079–1083. [CrossRef]
14. Hammami, S.; Snène, A.; El Mokni, R.; Faidi, K.; Falconieri, D.; Dhaouadi, H.; Piras, A.; Mighri, Z.; Porcedda, S. Essential Oil Constituents and Antioxidant Activity of *Asplenium* Ferns. *J. Chromatogr. Sci.* **2016**, *54*, 1–5. [CrossRef]
15. Živković, S.; Skorić, M.; Šiler, B.; Dmitrović, S.; Filipović, B.; Nikolić, T.; Mišić, D. Phytochemical characterization and antioxidant potential of rustyback fern (*Asplenium ceterach* L.). *Lek. Sirovine* **2017**, *37*, 15–20. [CrossRef]
16. Tomou, E.-M.; Skaltsa, H. Phytochemical Investigation of the Fern *Asplenium ceterach* (Aspleniaceae). *Nat. Prod. Commun.* **2018**, *13*, 849–850.
17. Živković, S.; Milutinović, M.; Maksimović, V.; Ćirić, A.; Ivanov, M.; Božunović, J.; Banjanac, T.; Mišić, D. Antioxidant and antimicrobial activity of two *Asplenium* species. *S. Afr. J. Bot.* **2020**, *132*, 180–187. [CrossRef]
18. Adams, R. *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*, 4th ed.; Allured Publishing Corporation: Carol Stream, IL, USA, 2007.
19. 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 Process.* **1998**, *173*, 191–241. [CrossRef]
20. Ellis, A.M.; Mayhew, C.A. *Proton Transfer Reaction Mass Spectrometry: Principles and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
21. Hodges, M.D.; De Long, J.M.; Forney, C.F.; Prange, R.K. Improving the thiobarbituric acid-reactive-substances assay for estimating lipid peroxidation in plant tissues containing anthocyanin and other interfering compounds. *Planta* **1999**, *207*, 604–611. [CrossRef]
22. Bewley, J.D.; Krochko, J.E. Desiccation-tolerance. In *Encyclopaedia of Plant Physiology, Physiological Ecology II*; Lange, O.L., Nobel, P.S., Osmond, C.B., Ziegler, H., Eds.; Springer: Berlin/Heidelberg, Germany, 1982; Volume 12/B, pp. 325–378.
23. Farrant, J.M.; Cooper, K.; Kruger, L.A.; Sherwin, H.W. The effect of drying rate on the survival of three desiccation-tolerant angiosperm species. *Ann. Bot.* **1999**, *84*, 371–379. [CrossRef]
24. Farrant, J.M. Mechanisms of Desiccation Tolerance in Angiosperm Resurrection Plants. In *Plant Desiccation Tolerance*; Jenks, M., Wood, A., Eds.; Blackwell Publishing: Wallingford, IA, USA, 2007; pp. 51–90.
25. Lazarević, M.; Rakić, T.; Šinžar-Sekulić, J. Morphological differences between the flowers of *Ramonda serbica*, *R. nathaliae* and their hybrid. *Bot. Serb.* **2014**, *38*, 91–98.
26. Rakić, T.; Lazarević, M.; Jovanović, Ž.S.; Radović, S.; Šiljak-Yakovlev, S.; Stevanović, B.; Stevanović, V. Resurrection plants of the genus *Ramonda*: Prospective survival strategies–unlock further capacity of adaptation, or embark on the path of evolution? *Front. Plant Sci.* **2014**, *4*, 550. [CrossRef]
27. Tan, T.; Sun, Y.; Luo, S.; Zhang, C.; Zhou, H.; Lin, H. Efficient modulation of photosynthetic apparatus confers desiccation tolerance in the resurrection plant *Boea hygrometrica*. *Plant Cell Physiol.* **2017**, *58*, 1976–1990. [CrossRef] [PubMed]
28. Pichersky, E.; Noel, J.P.; Dudareva, N. Biosynthesis of Plant Volatiles: Nature's Diversity and Ingenuity. *Science* **2006**, *311*, 808–811. [CrossRef] [PubMed]
29. Dudareva, N.; Klempien, A.; Muhlemann, J.K.; Kaplan, I. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.* **2013**, *198*, 16–32. [CrossRef] [PubMed]
30. Loreto, F.; Schnitzler, J. Abiotic stresses and induced BVOCs. *Trends Plant Sci.* **2010**, *15*, 154–166. [CrossRef]
31. Tholl, D.; Boland, W.; Hansel, A.; Loreto, F.; Röse, U.S.R.; Schnitzler, J.P. Practical approaches to plant volatile analysis. *Plant J.* **2006**, *45*, 540–560. [CrossRef]
32. Majchrzak, T.; Wojnowski, W.; Lubinska-Szczygeł, M.; Różańska, A.; Namieśnik, J.; Dymerski, T. PTR-MS and GC-MS as complementary techniques for analysis of volatiles: A tutorial review. *Anal. Chim. Acta* **2018**, *1035*, 1–13. [CrossRef]
33. Rapior, S.; Breheret, S.; Talou, T.; Pélissier, Y.; Bessière, J.M. The anise-like odor of *Clitocybeodora*, *Lentinelluscochleatus* and *Agaricusessettei*. *Mycologia* **2002**, *94*, 373–376. [CrossRef]
34. Rapior, S.; Breheret, S.; Talou, T.; Pélissier, Y.; Milhau, M.; Bessière, J.M. Volatile components of fresh *Agrocybe aegerita* and *Tricholoma sulfureum*. *Cryptogam. Mycol.* **1998**, *19*, 15–23.
35. Fons, F.; Froissard, D.; Bessière, J.M.; Buatois, B.; Rapior, S. Biodiversity of volatile organic compounds from five French ferns. *Nat. Prod. Commun.* **2010**, *5*, 1655–1658. [CrossRef]
36. Froissard, D.; Fons, F.; Bessière, J.M.; Buatois, B.; Rapior, S. Volatiles of French ferns and "fougère" scent in perfumery. *Nat. Prod. Commun.* **2011**, *6*, 1723–1726. [CrossRef]
37. Fons, F.; Rapior, S.; Gargadennec, A.; Andary, C.; Bessière, J.M. Volatile components of *Plantago lanceolata* (Plantaginaceae). *Acta Bot. Gall.* **1998**, *145*, 265–269. [CrossRef]
38. Ho, H.Y.; Kou, R.; Tseng, H.K. Semiochemicals from the predatory stink bug *Eocanthecona furcellata* (Wolff): Components of metathoracic gland, dorsal abdominal gland, and sternal gland secretions. *J. Chem. Ecol.* **2003**, *29*, 2101–2114. [CrossRef]
39. Fons, F.; Froissard, D.; Bessière, J.-M.; Fruchier, A.; Buatois, B.; Rapior, S. Volatile Composition of Six Horsetails: Prospects and Perspectives. *Nat. Prod. Commun.* **2013**, *8*, 509–512. [CrossRef]
40. Watson, S.B. Aquatic taste and odor: A primary signal of drinking-water integrity. *J. Toxicol. Environ. Health Part A* **2004**, *67*, 1779–1795. [CrossRef]

41. Zhao, Y.; Jianwei, Y.; Ming, S.; Wei, A.; Min, Y. A fishy odor episode in a north China reservoir: Occurrence, origin, and possible odor causing compounds. *J. Environ. Sci.* **2013**, *25*, 2361–2366. [CrossRef]
42. Grosch, W. Lipid Degradation Products and Flavour. In *Food Flavours*; Morton, I.D., Mac Leod, A.J., Eds.; Elsevier: Amsterdam, The Netherlands, 1982; pp. 325–398.
43. Brunke, E.J.; Hammerschmidt, F.J.; Schmaus, G.; Akgul, A. The essential oil of *Rhuscoriaria* L. fruits. *Flavour Fragr. J.* **1993**, *8*, 209–214. [CrossRef]
44. Andreou, A.; Brodhun, F.; Feussner, I. Biosynthesis of oxylipins in nonmammals. *Prog. Lipid Res.* **2009**, *48*, 148–170. [CrossRef]
45. Ruocco, N.; Albarano, L.; Esposito, R.; Zupo, V.; Costantini, M.; Ianora, A. Multiple Roles of Diatom-Derived Oxylipins within Marine Environments and Their Potential Biotechnological Applications. *Mar. Drugs* **2020**, *18*, 342. [CrossRef]
46. Bhattacharjee, S. The language of reactive oxygen species signaling in plants. *J. Bot.* **2012**, *2012*, 985298. [CrossRef]
47. Hansel, A.; Jordan, A.; Holzinger, R.; Prazeller, P.; Vogel, W.; Lindinger, W. Proton transfer reaction mass spectrometry: On-line trace gas analysis at the ppb level. *Int. J. Mass Spectrom. Ion Process* **1995**, *149–150*, 609–619. [CrossRef]
48. Lindinger, W.; Jordan, A. Proton-transfer-reaction mass spectrometry (PTR–MS): On-line monitoring of volatile organic compounds at pptV levels. *Chem. Soc. Rev.* **1998**, *27*, 347. [CrossRef]
49. Lacko, M.; Piel, F.; Mauracher, A.; Spanel, P. Chemical ionization of glyoxal and formaldehyde with H_3O^+ ions using SIFT-MS under variable system humidity. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10170–10178. [CrossRef] [PubMed]
50. Lee, J.; Min, D.B. Analysis of Volatile Compounds from Chlorophyll Photosensitized Linoleic Acid by Headspace Solid-phase Microextraction (HS-SPME). *Food Sci. Biotechnol.* **2010**, *19*, 611–616. [CrossRef]
51. Kanavouras, A.; Hernandez-Münoz, P.; Coutelieris, F.; Selke, S. Oxidation-Derived Flavor Compounds as Quality Indicators for Packaged Olive Oil. *J. Am. Oil Chem. Soc.* **2004**, *81*, 251–257. [CrossRef]
52. Kiralan, M.; Çalik, G.; Kiralan, S.; Ramadan, M.F. Monitoring stability and volatile oxidation compounds of cold-pressed flax seed, grape seed and black cumin seed oils upon photo-oxidation. *Grasas Aceites* **2019**, *70*, e295. [CrossRef]
53. Potter, T.L.; Fagerson, I.S. Composition of coriander leaf volatiles. *J. Agric. Food Chem.* **1990**, *38*, 2054–2056. [CrossRef]
54. Gogus, F.; Ozel, M.Z.; Keskin, H.; KocakYanık, D.; Lewis, A.C. Volatiles of fresh and commercial sweet red pepper pastes: Processing methods and microwave assisted extraction. *Int. J. Food Prop.* **2015**, *18*, 1625–1634. [CrossRef]
55. Zhang, J.-H.; Sun, H.; Chen, S.; Zeng, L.; Wang, T.-T. Antifungal activity, mechanism studies on α-Phellandrene and Nonanal against *Penicilliumcyclopium*. *Bot. Stud.* **2017**, *58*, 13. [CrossRef]
56. Eberl, F.; Hammerbacher, A.; Gershenzon, J.; Unsicker, S.B. Leaf rust infection reduces herbivore-induced volatile emission in black poplar and attracts a generalist herbivore. *New Phytol.* **2018**, *220*, 760–772. [CrossRef]
57. Kigathi, R.N.; Weisser, W.; Reichelt, M.; Gershenzon, J.; Unsicker, S.B. Plant volatile emission depends on the species composition of the neighboring plant community. *BMC Plant Biol.* **2019**, *19*, 58. [CrossRef]
58. Jelen, H.H.; Krawczyk, J.; Larsen, T.O.; Jarosz, A.; Gołebniak, B. Main compounds responsible for off-odour of strawberries infected by *Phytophthoracactorum*. *Lett. Appl. Microbiol.* **2005**, *40*, 255–259. [CrossRef]
59. Alché, J.d.D. A concise appraisal of lipid oxidation and lipoxidation in higher plants. *Redox Biol.* **2019**, *23*, 101136. [CrossRef]
60. Kong, W.; Liu, F.; Zhang, C.; Zhang, J.; Feng, H. Non-destructive determination of Malondialdehyde (MDA) distribution in oilseed rape leaves by laboratory scale NIR hyperspectral imaging. *Sci. Rep.* **2016**, *6*, 35393. [CrossRef]
61. Niinemets, Ü.; Bravo, L.A.; Copolovici, L. Changes in photosynthetic rate and stress volatile emissions through desiccation-rehydration cycles in desiccation-tolerant epiphytic filmy ferns (*Hymenophyllaceae*). *Plant Cell Environ.* **2018**, *41*, 1605–1617. [CrossRef]
62. John, S.P.; Hasenstein, K.H. Biochemical responses of the desiccation-tolerant resurrection fern *Pleopeltis polypodioides* to dehydration and rehydration. *J. Plant Physiol.* **2018**, *228*, 12–18. [CrossRef]
63. NeeragundaShivaraj, Y.; Plancot, B.; Ramdani, Y.; Gügi, B.; Kambalagere, Y.; Jogaiah, S.; Driouich, A.; RamasandraGovind, S. Physiological and biochemical responses involved in vegetative desiccation tolerance of resurrection plant *Selaginella brachystachya*. *3 Biotech* **2021**, *11*, 135. [CrossRef]
64. Wang, Y.L.; Gao, S.S.; He, X.Y.; Li, Y.; Li, P.Y.; Zhang, Y.; Chen, W. Growth, secondary metabolites and enzyme activity responses of two edible fern species to drought stress and rehydration in Northeast China. *Agronomy* **2019**, *9*, 137. [CrossRef]
65. Pandey, V.; Ranjan, S.; Deeba, F.; Pandey, A.K.; Singh, R.; Shirke, P.A.; Pathre, U.V. Desiccation-induced physiological and biochemical changes in resurrection plant, *Selaginellabryopteris*. *J. Plant Physiol.* **2010**, *167*, 1351–1359. [CrossRef]
66. Djilianov, D.; Ivanov, S.; Moyankova, D.; Miteva, L.; Kirova, E.; Alexieva, V.; Joudi, M.; Peshev, D.; Van den Ende, W. Sugar ratios, glutathione redox status and phenols in the resurrection species *Haberlea rhodopensis* and the closely related non-resurrection species *Chirita eberhardtii*. *Plant Biol.* **2011**, *13*, 767–776. [CrossRef]
67. Li, A.; Wang, D.; Yu, B.; Yu, X.; Li, W. Maintenance or collapse: Responses of extraplastidic membrane lipid composition to desiccation in the resurrection plant *Paraisometrum mileense*. *PLoS ONE* **2014**, *9*, e103430. [CrossRef]
68. Banupriya, T.G.; Ramyashree, C.; Akash, D.; Yathish, N.S.; Sharthchandra, R.G. Studies on the mechanism of desiccation tolerance in the resurrection fern *Adiantumraddianum*. *J. Appl. Biol. Biotechnol.* **2020**, *8*, 6–14.
69. Lubaina, A.S.; Brijithlal, N.D.; Murugan, K. Unraveling desiccation and rehydration tolerance mechanism in the fern, *Adiantumlatifolium*. *Biosci. Biotechnol. Res. Commun.* **2016**, *9*, 672–679.

Int. J. Environ. Sci. Technol.
DOI 10.1007/s13762-016-1208-8

ORIGINAL PAPER

Forecasting hourly particulate matter concentrations based on the advanced multivariate methods

M. Perišić[1] · D. Maletić[1] · S. S. Stojić[2] · S. Rajšić[1] · A. Stojić[1]

Received: 22 April 2016 / Revised: 11 July 2016 / Accepted: 19 November 2016

Abstract In this study, several multivariate methods were used for forecasting hourly PM_{10} concentrations at four locations based on SO_2 and meteorological data from the previous period. According to the results, boosted decision trees and multi-layer perceptrons yielded the best predictions. The forecasting performances were similar for all examined locations, despite the additional PM_{10} spatio-temporal analysis showed that the sites were affected by different emission sources, topographic and microclimatic conditions. The best prediction of PM_{10} concentrations was obtained for industrial sites, probably due to the simplicity and regularity of dominant pollutant emissions on a daily basis. Conversely, somewhat weaker forecast accuracy was achieved at urban canyon avenue, which can be attributed to the specific urban morphology and most diverse emission sources. In conclusion to this, the integration of advanced multivariate methods in air quality forecasting systems could enhance accuracy and provide the basis for efficient decision-making in environmental regulatory management.

Editorial responsibility: An-Lei Wei.

Electronic supplementary material The online version of this article (doi:10.1007/s13762-016-1208-8) contains supplementary material, which is available to authorized users.

✉ M. Perišić
mirjana.perisic@ipb.ac.rs

[1] Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia

[2] Singidunum University, Danijelova 32, 11010 Belgrade, Serbia

Keywords Air quality · Environmental pollution · Regulatory management · Supervised learning algorithms

Introduction

Over the last century, changes in emission sources, methane concentrations and climate have affected atmospheric composition and led to the significant increase in the levels of particulate matter (PM) and gaseous pollutants, particularly in developing countries (Fang et al. 2013). According to recent estimates, about 3.5 million cardiopulmonary deaths annually and globally can be attributed to exposure to anthropogenic $PM_{2.5}$, and the projections are that this number could double by 2050 (Lelieveld et al. 2015). In addition to stringent abatement measures, the accurate and reliable prediction of air pollutant episodes and establishment of an early public warning system is of vital importance for the increase in life expectancy and reduction of health care expenditures.

Despite the fact that significant progress has been made through integration of different scientific approaches, modeling of air pollution data remains a challenge, due to complexity and non-linear nature of atmospheric phenomena and processes (Pai et al. 2013). The variety of techniques and tools described in the literature for air quality forecasting covers simple empirical approaches, statistical approaches including artificial neural networks and fuzzy logic methods, and physically-based approaches including deterministic methods and ensemble and probabilistic methods (Zhang et al. 2012). The deterministic approach mostly refers to meteorological and chemical transport models, such as sophisticated Community Air Quality Modelling System (CMAQ) for prediction of air quality index at locations with no real-time measurements.

Published online: 03 December 2016

The chemical transport models were first used in Germany for air quality forecasting purposes, and soon many other developed countries became aware of the benefits of such implementation and launched the centralized air quality forecasting systems based on different tools, from simple empirical to online-coupled meteorology and chemistry models. While deterministic models don't require a large quantity of observational data, they do demand sufficient knowledge and understanding of pollutant emission sources, transport and atmospheric reactions and transformations under the planetary boundary layer (Feng et al. 2015). Since crucial knowledge in this area is often limited and some processes are too complex to be presented within a model, deterministic models are computationally expensive and time-consuming for routine predictions and often employ approximations and simplifications that lead to strong biases and inaccuracy, thus making the forecasts useless for timely management of critical situations (Cobourn 2010; Russo and Soares 2014).

Over the last decade, the parametric or non-parametric statistical approaches have been proposed as a more economical alternative for discovering the underlying site-specific dependencies between pollutant concentrations and potential predictors (Feng et al. 2015). The most commonly examined were artificial neural networks, based on artificial neurons or nodes capable of learning relationships between the routinely-measured pollutant data and selected predictors through embedded functions and data from the previous period (Fernando et al. 2012). Unlike deterministic models, artificial neural networks provide more accurate air quality forecasts, whereas their major disadvantages are associated with "black box" nature and poor generalization performance (Moustris et al. 2013). Furthermore, both statistical and deterministic approaches show satisfactory or good performance in forecasting concentrations closer to average values, whereas the prediction of extreme pollution events is more challenging.

As summarized by Zhang et al. (2012), the integration of advanced statistical methods in future air quality forecasting systems could considerably reduce forecasting biases and further enhance accuracy. In our previous study, MVA methods were successfully applied for forecasting the contributions of industry and vehicle exhaust to volatile organic compound (VOC) levels in the urban area, with smallest relative forecast error of only 6% (Stojić et al. 2015a). In this study, we compared the performance of twelve advanced multivariate (MVA) methods for PM_{10} forecasting relying on meteorological data and SO_2 concentrations. The analysis was based on a multi-year dataset collected at four different locations, affected by traffic or industry emissions. The herein employed MVA classification and regression methods belong to the supervised learning algorithms designed within Toolkit for Multivariate Analysis (TMVA; Hoecker et al. 2007) within the ROOT framework (Brun and Rademakers 1997), for extracting the maximum available information from the extensive data in high-energy physics.

Materials and methods

The analyzed dataset comprising 5-year (2011–2015) hourly concentrations of PM_{10}, SO_2 and meteorological data (atmospheric pressure, temperature, humidity, wind speed and direction), was obtained from the automatic monitoring stations within the Institute of Public Health network, at four different sites (Fig. 1, Supplementary Material). In the urban area, mostly affected by vehicle-exhaust emissions, measurements were conducted at the Institute of Public Health and New Belgrade, the sites characterized as being urban canyon avenue (UCA) and urban boulevard (UB), respectively, due to their topographic configuration. In the area influenced by emissions from fossil fuel burning for industry and heating operations, the data were collected in Obrenovac and Grabovac, the sites corresponding to urban industry (UI) and rural industry (RI), respectively. The measurements at industrial sites were incomplete due to severe floods that affected the area in 2014. The concentrations of PM_{10} and SO_2 were measured by means of referent beta-ray attenuation (Thermo FH 62-IR) sampler and referent sampling device Horiba APSA 360, respectively. The meteorological data were obtained by using Lufft WS500-UMB Smart Weather Sensor. The accuracy and precision of detection methods are provided in Stojic et al. (2016).

The analyses of daily, weekly, seasonal and annual dynamics, trend (Pretty 2015) and periodicity were performed by means of Openair (Carslaw and Ropkins 2012) and Lomb (Ruf 1999) packages within the Statistical Software Environment R (Team 2012). The relationships between pollutant concentrations and wind characteristics were investigated by the use of bivariate polar plot and bivariate cluster analyses within the Openair package. The contribution of local emission sources, background and transport to the observed PM_{10} pollution was analyzed using the 72-h air mass back trajectories and trajectory sector analysis (TSA) as described in Stojić et al. (2016).

The following MVA methods were used for PM_{10} forecasting: Boosted decision trees (BDT, BDTG, BDTMitFisher), Artificial Neural Network Multilayer

Perceptron (MLP), MLP with Bayesian Extension (MLPBNN), Support Vector Machine (SVM), k-nearest neighbor (KNN), Linear Discriminant (LD), Boosted Fisher Discriminant (BoostedFisher), Multidimensional Probability Density Estimator Range Search Method (PDERS), Predictive Learning via Rule Ensembles (RuleFit) and Function Discriminant Analysis (FDA). All methods were used for both classification and regression. The five-year dataset was divided into two equal subsets, each consisting of PM_{10} concentrations and input data (meteorological and SO_2). One subset was used for method trainings, either to differentiate between high and low importance indicators for PM_{10} concentrations (classification), or to determine an approximation of the underlying functional behavior defining PM_{10} concentrations (regression). The other subset was utilized for method performance testing.

Results and discussion

Previous studies aimed at investigating the origin and spatio-temporal distribution of different pollutant species converge on the conclusion that poor air quality presents an important health risk factor in Belgrade area (Perišic et al. 2015; Stojić et al. 2015b). In the previous years, the mean annual PM_{10} concentrations in Belgrade area were in the range from 39.74 to 62.32 μg m^3, whereas the exceedances of the proposed air quality guideline value of 50 μg m^3 were registered during 20.5–42.2% of total number of days (Stanišić Stojić et al. 2016).

Specifics of measurement sites

In order to examine the MVA forecasting performances, PM_{10} observational data from four measurement sites affected by different emission sources were collected and analyzed (Fig. 1, Supplementary Material). The two locations defined as urban were affected by traffic emissions throughout the year. However, specific microclimatic conditions associated with contrasting urban morphology between UCA and UB plays an important role in spatial distribution of particles. The presence of tall buildings along both sides of the canyon avenue induces a complex wind flow that does not enhance the pollutant dispersion due to terrain configuration, but it facilitates suspension, particularly fine PM fraction (Vardoulakis et al. 2003). Furthermore, frequent congestions in the canyon avenue compared to free flowing traffic in the wide boulevard contributed to higher PM_{10} concentrations at UCA throughout the year, with the exception of winter season, when the air quality at UB was additionally affected by fuel burning from the neighboring heating plant.

The herein presented industrial locations were affected either by fuel burning emissions only (RI), or by emissions from both industrial activities and vehicle exhaust (UI). Within the range of 15–20 km in NW/N and SE/S direction around the two industrial sites, the strong emission sources including three thermal power plants, four open-pit mines of high-sulfur lignite and several coal ash disposal sites are located.

As can be seen, the highest mean PM_{10} concentration for the entire period was registered at UI (Table 1, Supplementary Material), which was partly driven by extreme pollutant loadings in 2012 (Fig. 2, Supplementary Material). It should be noted that the PM_{10} variations at two industrial locations exhibited similar pattern, only with less significant deviations at rural site, which points to the prevalence of the same emission sources.

Daily mean PM_{10} exceedances (>50 μg m^{-3}) were commonly observed, whereas the episodes of extreme pollutant levels were registered only at UI (Fig. 3, Supplementary Material). The winter PM_{10} concentrations were considerably higher at all examined locations, which can be partly attributed to heating operations, but also to lower planetary boundary layer (PBL) height in winter season. Unsurprisingly, the lowest PM_{10} levels for the entire period were observed at rural site, particularly during spring and summer season, with the values of 29.15 and 32.09 μg m^{-3} being registered, respectively. Conversely, the highest concentrations in warm season were measured at UCA, the only site predominately affected by traffic. The differences between the summer and winter concentrations were relatively small at UCA and RI, whereas the inter-seasonal variations at two other sites exposed to the emissions from two strong sources were almost two times higher.

In Fig. 4, Supplementary Material, daily, weekly and seasonal PM_{10} variations are displayed. Accordingly, the lowest concentrations were registered in May and June, probably due to intense precipitations. The particle resuspension processes and atmospheric photochemical reactions in dry summer months starting from July, led to the rising pollutant levels, particularly at industrial sites in the vicinity of ash disposals. The accumulation of particles during working days was followed by a significant decrease at the weekend at two locations dominated by vehicle exhaust emissions, whereas the weekday/weekend difference was not observed at UI and RI sites. As regards diurnal PM_{10} variations, the same pattern was detected at all locations: daytime levels tended to be low with the exception of

morning and afternoon rush hours, whereas the pronounced increase in nighttime concentrations could be attributed to stable atmospheric conditions and shallow PBL.

According to bivariate and cluster analysis, the average contributions of the surrounding emission sources were dominant at all locations (Fig. 1), particularly at UCA (59.5 μg m^{-3}), due to limited pollutant dispersion, and UI (73.1 μg m^{-3}), which has been directly exposed to emissions from the thermal plant which produces more than 50% of electricity for the Serbian market. The UCA is located in the central city area and thus, the polluted air masses were observed to come from all directions, whereas at UB, the impact of heating plant emissions from S and intersections with intensive traffic coming from E can be noted. In the case of industrial locations, local sources appeared to be particularly significant during the heating season, whereas in spring and summer, both UI and RI were affected by emissions from ash disposals and lignite mining sites in NW/N and SE/S. The dynamics of cluster contributions on a daily, weekly and seasonal basis are shown in Fig. 5, Supplementary Material. As can be seen, local emissions, corresponding to cluster 4 at industrial sites, exhibited extremely regular daily variations, which suggests the prominent role of anthropogenic sources. The rush hour peaks were noticeable only in the variations of locally-emitted PM_{10} concentrations at UCA (cluster 4), since the site has been dominated by traffic emissions.

The analysis was also performed to determine the impact of local emissions, transported pollution and background on the air quality at examined locations. According to TSA results, the estimated share of background was highest at rural site (48%), whereas the contribution of local production was the most significant factor (43%) for PM_{10} concentrations at UI, as previously shown by bivariate and cluster analysis.

Upon the presented analysis, we have reached the conclusion that the selected locations are substantially different in terms of air quality and factors closely associated with it, including micro-climatic conditions, topographic features and proximity of strong sources. This was considered a prerequisite for examining the dependency between the efficiency of MVA methods for air quality forecasting and site characteristics.

Classification MVA methods

As previously mentioned, the 5-year dataset, including PM_{10} and SO_2 concentrations, and meteorological data, was divided into two subsets equal in size, used for training and testing of MVA methods, respectively. In order to account for seasonal, *i.e.* weekday/weekend variations, two new variables were introduced for classification purposes: Yearreal is a quotient of the ordinal number of a day and total number of days per year, while Weekreal represents

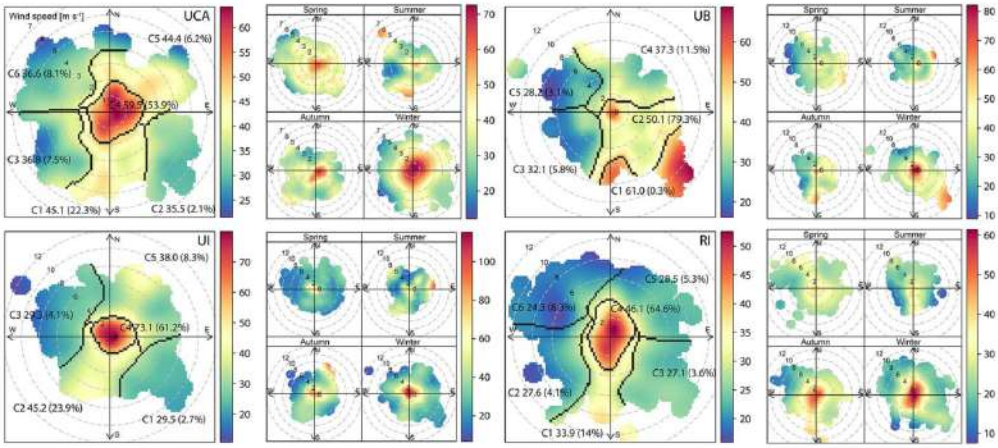


Fig. 1 The relationship between PM_{10} concentrations and wind characteristics: bivariate cluster plot [frequency (%) and average contributions (μg m^{-3})] for the entire period (*left*) and seasonal variations (μg m^{-3}) (*right*)

the quotient of the ordinal number of a day and number 7. Correlation and mutual information of input variables and the observed PM_{10} mass concentrations for all sampling sites are presented in Table 1.

For the purposes of classification, the PM_{10} levels above 50 μg m^{-3} are considered to require the increased level of caution, whereas those exceeding 100 μg m^{-3} are considered extremely high—alarm triggering values, both of which are chosen as arbitrary limits. The estimation of classification method performances by using the Receiver Operating Characteristic (ROC) curve is presented in Fig. 2. The highest separation between background and predicted PM_{10} concentrations was observed when PM_{10} classifier value of 100 μg m^{-3} was taken into account (Fig. 3), whereas somewhat poorer results were obtained for 50 μg m^{-3}, which suggests that including additional meteorological or pollutant variables as input data might further enhance classification performance.

The comparison of the results by evaluating signal and background efficiencies revealed that certain MVA methods are capable of classifying the PM_{10} levels which are considered to require a high degree of caution (Table 2, left). The results showed that BDTG and MLP exhibit the best results for all examined locations. Signal and background separation was most efficiently performed for RI and UB, and to a somewhat lower extent for UCA.

Table 1 Correlation (C) and mutual information (MI) of input variables (P, pressure; T, temperature; Rh, relative humidity; ws, wind speed; Yearreal, day of year; Weekreal, day of week) and measured PM_{10} concentrations at all sampling sites

Variable	UCA		UB		UI		RI	
	C	MI	C	MI	C	MI	C	MI
P	0.18	1.31	0.26	0.97	0.20	1.49	0.29	1.26
T	0.21	1.40	0.30	1.21	0.28	1.69	0.22	1.39
Rh	0.24	1.47	0.24	1.29	0.22	1.86	0.19	1.60
ws	0.29	1.39	0.25	0.82	0.26	1.57	0.32	1.18
SO_2	0.25	1.63	0.09	1.39	0.20	1.87	0.32	1.59
Yearreal	0.04	1.49	0.05	1.31	0.09	1.86	0.12	1.53
Weekreal	0.02	0.12	0.03	0.11	0.02	0.18	0.02	0.14

Regression MVA methods

Regression MVA methods were applied to interpret the relationships between pollutant concentrations and the examined input data. Similar to classification methods, BDTG and MLP exhibited the most satisfying performances with absolute and relative errors presented in Table 2, right. The MVA method performance was best for PM_{10} loadings at industrial sites, around 25%, while the forecast quality could be clearly seen at RI location, Fig. 4. It can be assumed that more accurate air quality forecasts can be

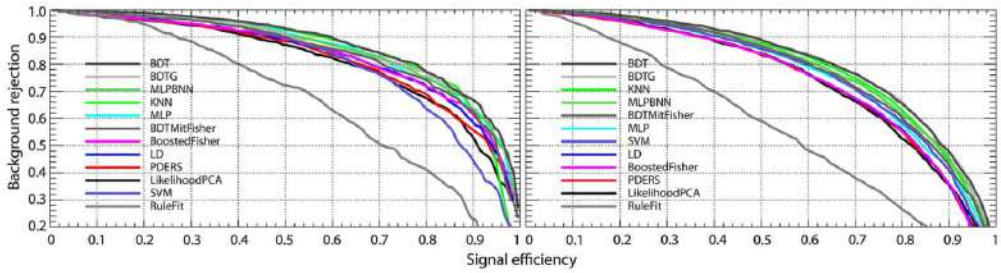


Fig. 2 ROC curves for MVA classification methods with PM_{10} classifier value of 100 μg m^{-3} (*left*) and 50 μg m^{-3} (*right*) for all sampling sites

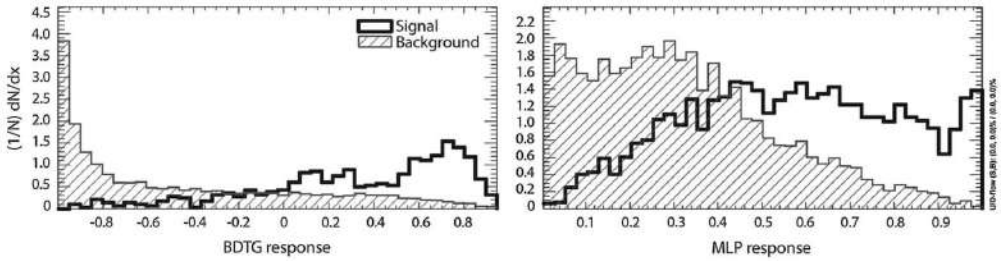


Fig. 3 MVA classification method response for PM_{10} classifier value of 100 μg m^{-3} (*left*) and 50 μg m^{-3} (*right*) for UCA site

Table 2 The comparison of best performing methods for ROC, separation and significance values for all measurement sites (left) and absolute (μg m^{-3}) and relative (%) errors of the best performing regression methods (right)

Sampling site	Method	Classification			Regression	
		ROC	Separation	Significance	Absolute error	Relative error
UCA	BDTG	0.806	0.282	0.883	17.2	29.6
	MLP	0.772	0.226	0.755	21.8	37.5
UB	BDTG	0.868	0.408	1.12	13.9	26.8
	MLP	0.841	0.352	1.015	17.4	33.5
UI	BDTG	0.855	0.379	1.059	15.6	24.6
	MLP	0.826	0.323	0.956	24.0	37.9
RI	BDTG	0.867	0.412	1.172	10.6	25.2
	MLP	0.837	0.345	0.962	15.1	36.0

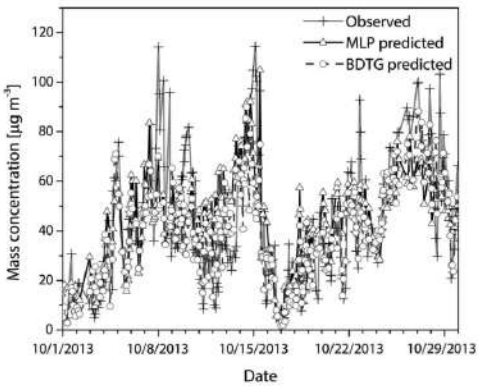


Fig. 4 The comparison of time series of the observed and best performing MVA-predicted PM_{10} concentrations (μg m^{-3}) at RI site

achieved at the locations such as RI, which are affected by less significant number of emission sources. Furthermore, the simplicity and regularity of dominant pollutant emissions on a daily, weekly and seasonal basis, as registered at UI location, as well as minor deviations from the commonly observed pollutant loadings, which is particularly evident for air quality forecasting at rural site, are probably the additional factors associated with forecast accuracy.

Conversely, the weakest MVA method performance was derived for PM_{10} concentrations at UCA, probably because the urban morphology of the canyon avenue represents the additional factor modifying the pollutant levels in a less predictable manner. Furthermore, the emission sources in the central city zone are diverse and primarily refer to traffic congestions and intense atmospheric reactions that take place in stagnant conditions of the canyon street. Moreover, they also relate to local fireboxes in residential area where lignite is burned

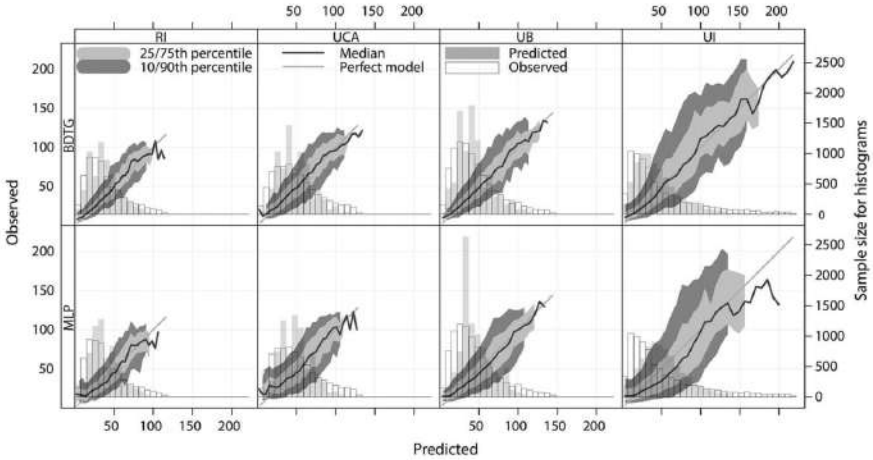


Fig. 5 The comparison of the observed and best performing MVA-predicted PM_{10} mass concentrations (μg m^{-3})

during autumn and winter season and local manufactures that are associated with pollutant emissions highly variable in time and intensity.

As can be seen in Fig. 5, the PM_{10} time series evaluated by means of MVA regression methods correlated very well with the observed concentrations at all sampling sites. Mutual information obtained for BDTG-predicted and the observed PM_{10} mass concentrations were 0.71, 0.7. 0.65 and 0.64 for RI, UB, UCA and UI, respectively. This suggests that significant input variables were used for the forecasting process. In addition, it could be noted that their distributions are relatively well.

Although the other MVA methods employed in the present study generated similar results when being used for classification, they generated the significant PM_{10} forecast errors when being used for regression, at least based on the observed input variables. The herein presented errors are mostly in compliance with the findings of our previous study, aimed at forecasting the contributions from traffic and industry to the observed VOC concentrations in the urban area, which suggests that both PM and VOC, as important air quality indicators, can be predicted using the MVA methods.

Conclusion

In this study, the performances of MVA methods for forecasting PM_{10} concentrations and prediction of related health-damaging events were evaluated on the basis of datasets from traffic- and industry-affected locations with substantial differences in air quality, which has also been verified through additional analyses. The results of both classification and regression methods were rather promising, particularly considering the fact that the presented forecast accuracy referred to hourly concentrations. The quality of the prediction might be partly dependent on microclimatic conditions, topographic characteristics, presence of strong emission sources and other site characteristics, as well as on the input data. All that implies that the selection of additional or different variables could enhance the method forecasting performances. The importance of accurate air quality forecasts as part of the management system is reflected in the potential applications, including health alerts for susceptible categories, operational planning, as well as amendment of pollutant time-series and reduction of regular monitoring expenditures.

Acknowledgements This study was carried out as part of the Project No. III43007, III41011 and OI171002, financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia for the period 2011–2016.

References

Brun R, Rademakers F (1997) ROOT—an object oriented data analysis framework. Nucl Instrum Methods A 389:81–86

Carslaw DC, Ropkins K (2012) Openair-an R package for air quality data analysis. Environ Model Softw 27:52–61

Cobourn WG (2010) An enhanced $PM_{2.5}$ air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmos Environ 44:3015–3023

Fang Y, Mauzerall DL, Liu J, Fiore AM, Horowitz LW (2013) Impacts of 21st century climate change on global air pollution-related premature mortality. Clim Change 121:239–253

Feng X, Li Q, Zhu Y, Hou J, Jin L, Wang J (2015) Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation. Atmos Environ 107:118–128

Fernando HJS, Mammarella MC, Grandoni G, Fedele P, Di Marco R, Dimitrova R, Hyde P (2012) Forecasting PM_{10} in metropolitan areas: efficacy of neural networks. Environ Pollut 163:62–67

Hoecker A, Speckmayer P, Stelzer J, Therhaag J, Von Toerne E, Voss H, Backes M, Carli T, Cohen O, Christov A, Dannheim D, Danilowski K, Henrot-Versille S, Jachowski M, Kraszewski K, Krasznahorkay A, Kruk M, Mahalalel Y, Ospanov R, Prudent X, Robert A, Schouten D, Tegenfeldt F, Voigt A, Voss K, Wolter M, Zemla A (2007) TMVA users guide—toolkit for multivariate data analysis, PoSACAT 040. http://arxiv.org/abs/physics/0703039. Accessed 14 June 2015

Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A (2015) The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525:367–371

Moustris KP, Larissi IK, Nastos PT, Koukouletsos KV, Paliatsos AG (2013) Development and application of artificial neural network modeling in forecasting PM_{10} levels in a Mediterranean city. Water Air Soil Pollut 224:1–11

Pai TY, Hanaki K, Chiou RJ (2013) Forecasting hourly roadside particulate matter in Taipei County of Taiwan based on first-order and one-variable grey model. CLEAN Soil Air Water 41:737–742

Perišić M, Stojić A, Stojić SS, Šoštarić A, Mijić Z, Rajšić S (2015) Estimation of required PM_{10} emission source reduction on the basis of a 10-year period data. Air Qual Atmos Health 8:379–389

Pretty R (2015) TheilSen {openair} tests for trends using Theil–Sen estimates. http://www.inside-r.org/packages/cran/openair/docs/TheilSen. Accessed 12 Dec 2015

Ruf T (1999) The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. Biol Rhythm Res 30:178–201

Russo A, Soares AO (2014) Hybrid model for urban air pollution forecasting: a stochastic spatio-temporal approach. Math Geosci 46:75–93

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. J Toxicol Environ Health A 79:17–27

Stojić A, Maletić D, Stojić SS, Mijić Z, Šoštarić A (2015a) Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. Sci Total Environ 521:19–26

Stojić A, Stojić SS, Mijić Z, Šoštarić A, Rajšić S (2015b) Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling. Atmos Environ 106:71–79

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. Environ Sci Pollut Res. doi:10.1007/s11356-016-6266-4

Team RC (2012) R: a language and environment for statistical computing. http://cran.case.edu/web/packages/dplR/vignettes/timeseries-dplR.pdf. Accessed 10 June 2015

Vardoulakis S, Fisher BE, Pericleous K, Gonzalez-Flesca N (2003) Modelling air quality in street canyons: a review. Atmos Environ 37:155–182

Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A (2012) Real-time air quality forecasting, part I: history, techniques, and current status. Atmos Environ 60:632–655

Публикације категорије М30

Innovation Center of Faculty of Mechanical Engineering

Faculty of Mechanical Engineering, University of Belgrade

Center for Business Trainings

CNN TECH

„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

ACKNOWLEDGEMENT

The organizing committee of the 5th International Conference of Experimental and Numerical Investigations and New Technologies – CNN TECH 2021 wishes to sincerely thank all the institutions and individuals who by means of personal engagement and constructive action helped organizing this conference.

We particularly wish to thank our sponsor, **The Ministry of Education, Science and Technological development**, Government of the Republic of Serbia.

We are also grateful to companies, **3D Republic**, **Shimatzu**, **Trokuttest, IMW Institute** and **Inter Cert** who have significantly contributed to the organization and realization of the conference.

PREFACE

Dear Friends and Colleagues, Welcome to CNN Tech 2021 Conference and the fabulous mountain of Zlatibor!

With 90 papers (17 by international authors) and contributions by authors from 12 different countries, International Conference of Experimental and Numerical Investigations and New Technologies CNN Tech 2021 successfully sets the high level for the future conferences. Participation of a large number of domestic and international authors, as well as the diversity of topics, justifies our efforts to organize this conference and contribute to exchange of knowledge, research results and experience of industry experts, research institutions and faculties which all share a common interest in the field in experimental and numerical investigations.

This year CNN Tech 2021 focuses on the following topics:

- Mechanical Engineering,
- Engineering Materials,
- Chemical and Process Engineering,
- Experimental Techniques,
- Numerical Methods,
- New Technologies,
- Clear sky,
- Sustainable Design and New Technologies,
- Advanced Materials and Technology,
- Artificial intelligence and
- Student session.

Apart from a plenty of interesting lectures, the participants will have a chance to lighten up and communicate in friendly and relaxed settings.

Organizing committee of CNN Tech 2021 would like to express gratitude to Ministry of Education, Science and Technological development for financial support of the Conference.

On behalf of the Innovation center of Faculty of Mechanical Engineering, Faculty of Mechanical Engineering and Center for Business Trainings, we wish this to be splendid CNN Tech conference filled with many memorable moments.

PROGRAMME AND ORGANIZING COMMITTEE

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.

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.

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

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.

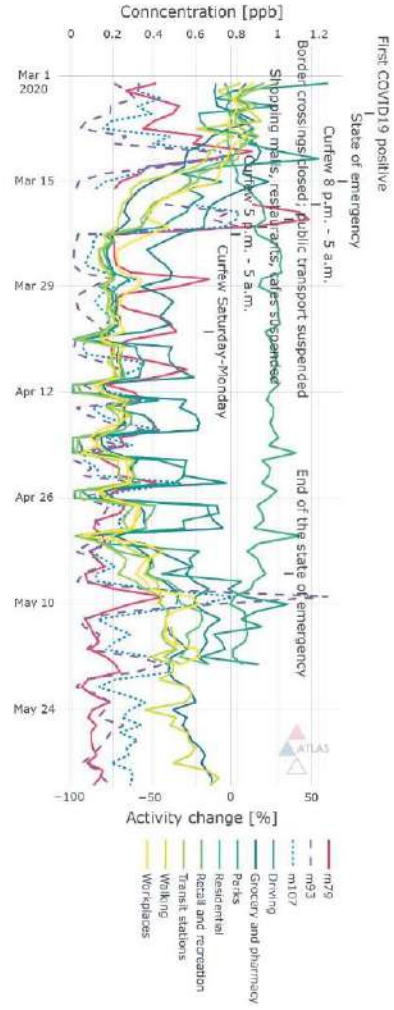


Figure 1 - BTEX concentration and human activity time series.

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 7th 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.

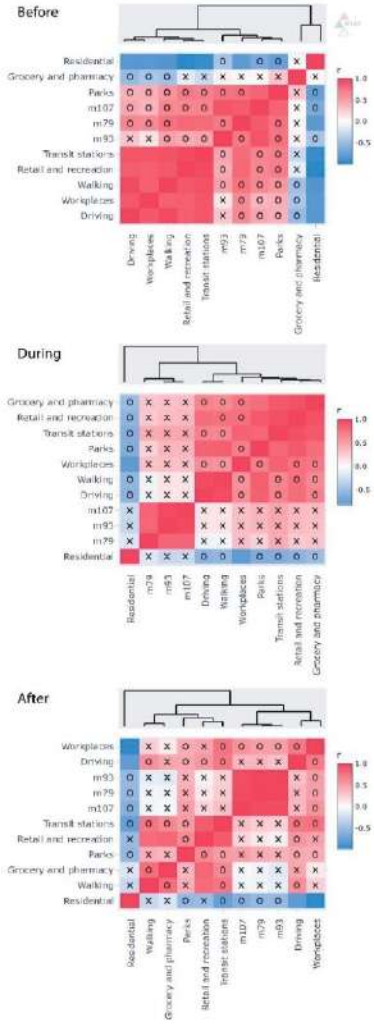


Figure 2 - Parameter value correlation matrix.

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.

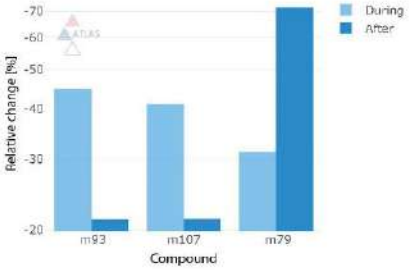


Figure 3 - Mean BTEX concentration difference during and after relative to the period before the state of emergency.

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.

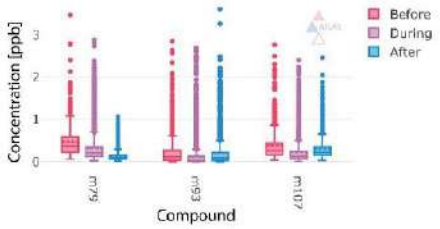


Figure 4 - BTEX box plots.

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.

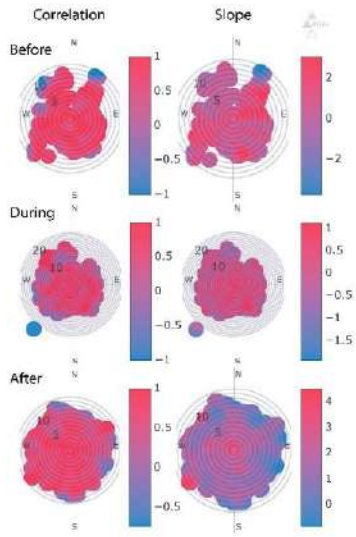


Figure 5 - Toluene and benzene correlation and slope dependency on wind parameters.

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).

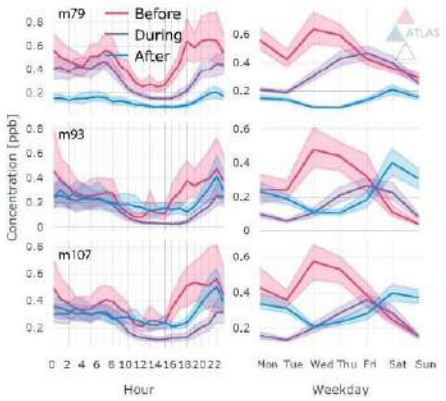


Figure 6 - BTEX diurnal and weekday variations.

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.

ENVIRONMENTAL DATA SCIENCE SESSION

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.

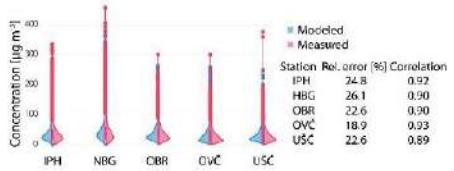


Figure 1 - XGBoost model evaluation.

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).

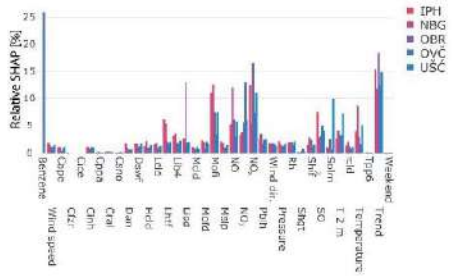


Figure 2 - Feature importance.

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.

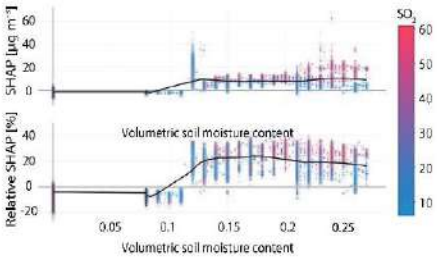


Figure 3 - PM_{10} SHAP dependency on volumetric soil moisture content and SO_2.

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.

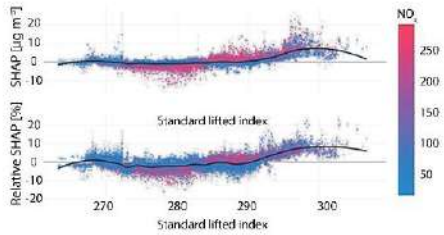


Figure 4 - PM_{10} SHAP dependency on standard lifted index and NO_x.

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.

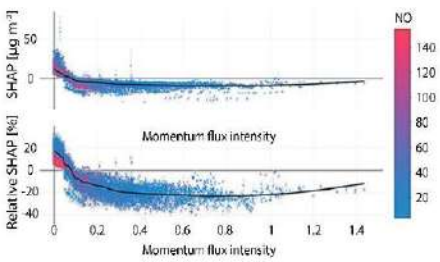


Figure 5 - PM_{10} SHAP dependency on momentum flux intensity and NO.

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.

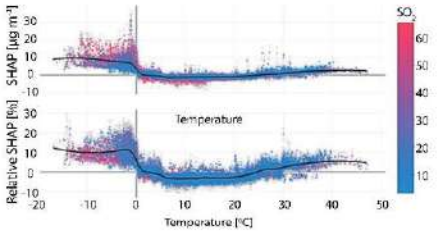


Figure 6 - PM_{10} SHAP dependency on temperature and SO_2.

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}.

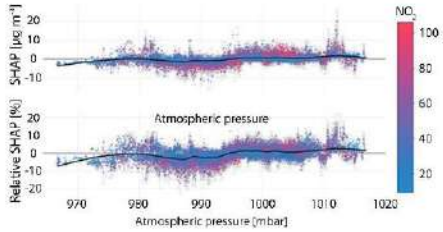


Figure 7 - PM_{10} SHAP dependency on atmospheric pressure and NO_2.

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.

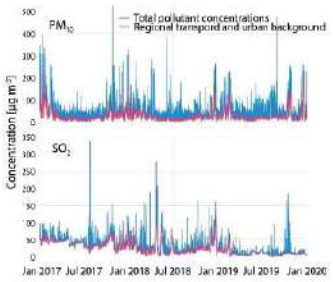


Figure 1 - An example of separating the contribution of emissions from local sources from regional transport and the background of the urban environment at the measuring point of the City Institute for Public Health Belgrade in the period from 2017 to 2019

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.

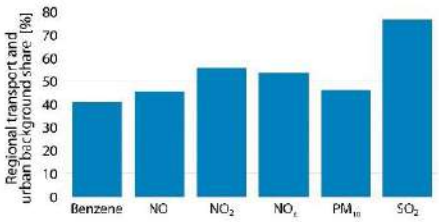


Figure 2 - The share of regional transport and the background of the urban environment with the measured concentrations of pollutants in Belgrade in the period from 2017 to 2019

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).

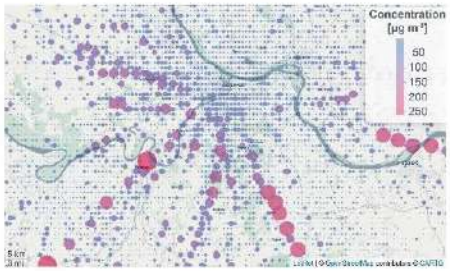


Figure 3 – The distribution of regional sources of PM_{10} emissions on the territory of Belgrade and neighboring municipalities in the period from 2017 to 2019

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 infrastructures 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.

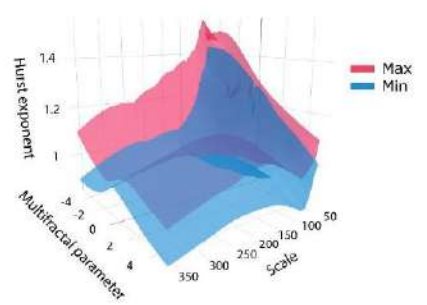


Figure 1 – PM_{10} Hurst exponent range in Belgrade 2017-2019.

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.

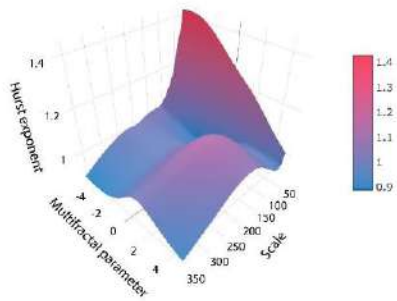


Figure 2 – Structural characteristics of time series of concentrations PM_{10} at monitoring place at the Institute of Public Health of Belgrade for the period from 2017 to 2019.

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.

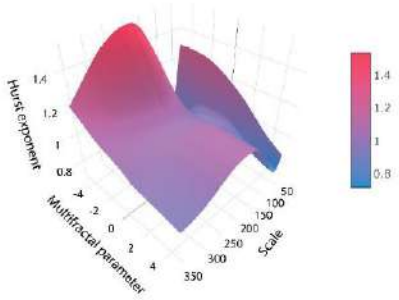


Figure 3 – Structural characteristics of time series of concentration PM_{10} at the monitoring station Obrenovac for the period from 2017 to 2019.

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 WeBI-OPATR 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.

ENVIRONMENTAL DATA SCIENCE SESSION

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).

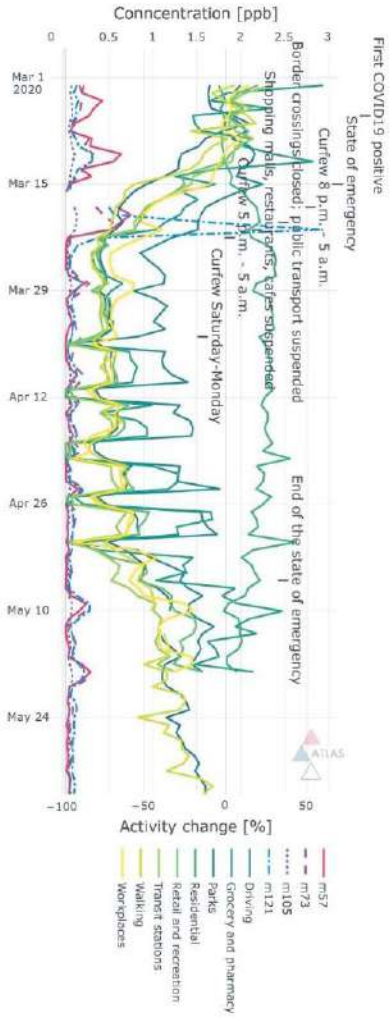


Figure 1 - Industry-related VOC concentration and human activity time series.

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.

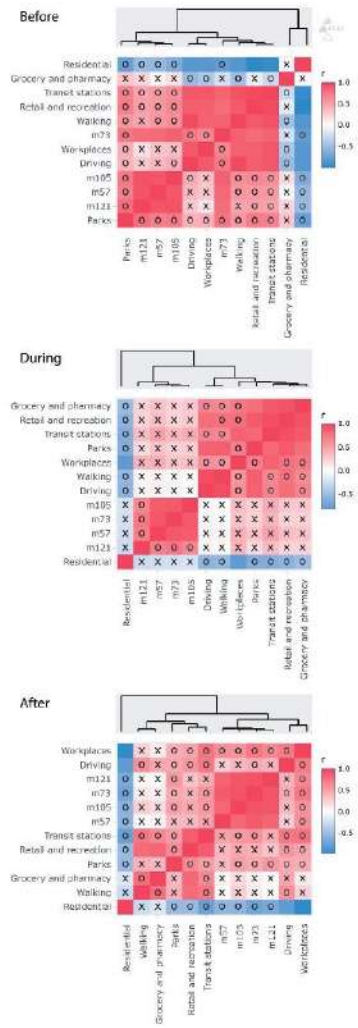


Figure 2 - Parameter value correlation matrix.

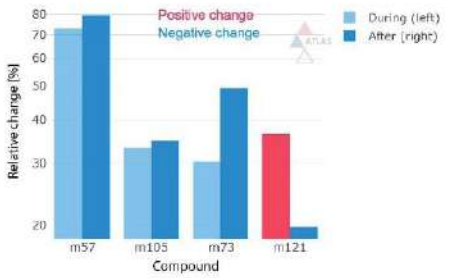


Figure 3 - Mean VOC concentration difference during and after relative to the period before the state of emergency.

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.

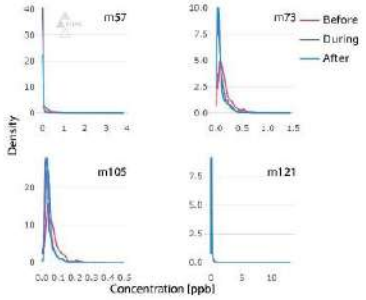


Figure 4 - VOCs density plots.

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.

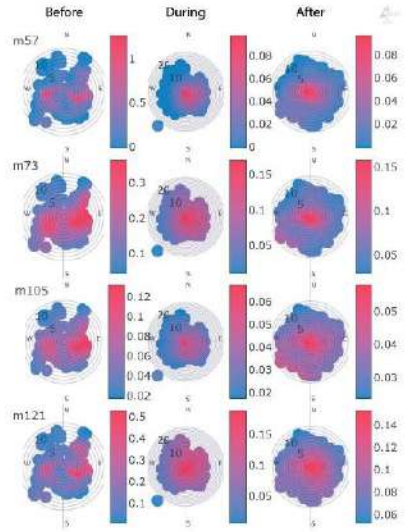


Figure 5 - Industry-related VOC dependency on wind parameters.

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.

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 Institute 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

Number of copies

150

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, University of Belgrade, National Institute of the Republic of Serbia, 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, WEBIOPATR2021

is supported by:

Virtual centre for Distributed atmospheric Sensing for reduction of pollution pressures

EC H2020 Framework Program for Research and Innovation,
area "Spreading excellence and widening participation",
VIDIS project (2020-2023) coordinated by Vinča Institute of Nuclear Sciences,
Grant agreement number 952433.

Ministry of Education, Science and Technological Development of the Republic of Serbia

PREFACE

Dear Colleagues,

Welcome to the 8th WeBIOPATR Conference, to be held at the premises of the Vinca Institute of Nuclear Sciences, Serbia, 29.11.–1.12.2021, as a combination of online and face-to-face event.

The International Workshop and Conference, Particulate Matter: Research and Management – WeBIOPATR is a biennial event held in Serbia since 2007. The conference addresses air quality in general and particulate matter specifically. Atmospheric particulate matter arises both from primary emissions and from secondary formation in the atmosphere. It is one of the least well understood local and regional air pollutants, has complex implications for climate change, and is perhaps the pollutant with the highest health relevance. It also poses many challenges to monitoring.

By WeBIOPATR, we aim to link the research communities with relevance to particulate matter with the practitioners of air quality management on all administrative levels, in order to facilitate professional dialogue and uptake of newest research into practice. The workshops usually draw an audience of about 70 and attract media attention in Serbia. It enjoys support of the responsible authorities, Ministry of Education, Science and Technological Development, Ministry of Health, Ministry of Environment, and the Serbian Environmental Agency whose sponsorship is indispensable and gratefully acknowledged. We also enjoy support of international bodies such as the WHO.

The 1st WeBIOPATR Workshop was held in Beograd, 20.-22. May 2007, associated with a project funded by the Research Council of Norway. The 2nd workshop was held in Mećavnik, Serbia, 28.8.-1.9.2009. WeBIOPATR2011 was held in Beograd 14.-17.11.2011 and for the first time, included a dedicated student workshop. WeBIOPATR2013 was held in Beograd 2.-4.10. 2013. It covered the traditional PM research and management issues, discussions on how to encourage citizens to contribute to environmental governance, and how to develop participatory sensing methods. WeBIOPATR2015 was held in Beograd 14.-16.10. 2015. Dedicated sessions were devoted to sensor technologies for air quality monitoring, utilizing information and input from the EU FP7 funded project CITI-SENSE (http://co.citi-sense.eu) and the EU COST action EuNetAir (www.eunetair.it). WeBIOPATR2017, the 6th conference, was held in Beograd 6.-8.9. 2017, with a wider than before Western Balkan participation. The 7th WeBIOPATR2019 was held 1.-4.10. 2019 at the Mechanical Faculty, University of Belgrade. It has attracted a record of over 50 contributions, and brought together scientists from 12 countries, documenting that the issues of atmospheric pollution, with their wide implications for climate change, human health and ecosystem services, are no less important today. This year's event will be with similar number of contributions that have been accepted.

In the past two years, all our lives were affected by the COVID-19 pandemic. We have adapted our ways of life and work – and now we hope that the new format of the conference

will be a success, for the participants physically present as well as for those who will participate online.

We are very grateful to our unrelenting national and international partners for their financial and scientific support for this event. In addition, WeBIOPATR2021 is supported by the VIDIS project, https://vidis-project.org/, funded by the EC H2020 Framework Programme for Research and Innovation, area "Spreading excellence and widening participation". VIDIS (2020-2023) is coordinated by Vinca Institute, Grant agreement number 952433.

Welcome to Vinca and online and have a stimulating and productive time!

Milena Jovašević-Stojanović and Alena Bartoňová

TABLE OF CONTENTS

3.2 THE HYBRID COMPUTATIONAL APPROACH IN REVEALING PARTICULATE MATTER RELATED PROCESSES

M. Perišić (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

mirjana.perisic@ipb.ac.rs

The modern world is facing many environmental issues, with all environmental compartments being affected. Air pollution accounts for an estimated 4.2 million deaths and 103.1 million disability-adjusted life years per year, while around 91% of the world's population lives in places where air pollution levels exceed the World Health Organization limits (WHO, 2021).

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 big data by using artificial intelligence, advanced statistical analysis, and numerical modelling, as well as modelling hybridization.

The talk will elaborate on the concept required for an in-depth understanding of environmental pollution from the perspective of contextual data analysis and the ATLAS Project supported by the Science Fund of the Republic of Serbia. 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. Also, ATLAS aims to harmonize environmental research via facilitating access to environmental data, data analysis, exploration, and exploitation of the results. This will 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 examples of the talk will cover the results produced within the Project Air Quality Plan for the Agglomeration of the City of Belgrade for the period 2021 – 2031, supported by the City Administration of Belgrade (AQP, 2021). The main focus will be on 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.

ACKNOWLEDGEMENTS

The author acknowledges the 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

Air Quality Plan for the Agglomeration of the City of Belgrade for the period 2021 – 2031 (in the Serbian language) https://www.beograd.rs/images/data/7a0a3b18c076a6bfb21688ca7d314015_4302538578.pdf

World Health Organization, 2021. WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: executive summary.

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.

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

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: 2020
Circulation: 10
Printed by: Caligraph, Belgrade
ISBN: 978-86-7912-735-8

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 2020

International Scientific SINTEZA (SYNTHESIS) Conference 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, its application and related innovations, as well as Data Related Research.

Rapid advances in Information Technologies over the previous decades have had a huge impact on numerous facets of everyday life and have created tremendous opportunities for economic, technological and social gains on 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. New technologies and scientific break-throughs have altered working and living environments making them safer, more convenient and more connected.

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, as well as their applications in a range of engineering and other fields of research.

Specific topics for this year are focused on artificial intelligence, machine learning, data research and analysis, as well as their application in solving practical real-life problems.

Due to the Covid-19 imposed pandemic, the conference has been held online this year, supported by the application of the Microsoft Teams Live Event software. The experiences we have gained are of precius value and could be extremely useful in the future – not only for conference organisers, but also for session moderators, speakers and all the participants.

Sincerely,
Organising Committee of Sinteza 2020

BENZENE SOURCE APPORTIONMENT USING BIVARIATE CORRELATION AND REGRESSION ANALYSES

Mirjana Perišić[1,2*],
Gordana Jovanović[1,2],
Ana Vranić[1],
Svetlana Stanišić[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

Abstract:

The aim of this study was to identify and characterize the individual sources of benzene in an urban area of Belgrade based on bivariate polar plot concentration, correlation, and regression analyses. The presented benzene behavior modeling relied on a weighted Pearson correlation coefficient, linear regression slope, and Gaussian kernel locally weighted by wind speed direction surface. The data, including the concentrations of volatile organic compounds (VOCs), inorganic gaseous pollutants and meteorological parameters, were obtained from a measurement campaign conducted at the Singidunum University (Belgrade, Serbia) during and after the heating season 2016. The results indicate the dominance of benzene, toluene, and NO local emission sources, as well as the significant impact of remote NO_2 sources located in the SW and SE. The strong interrelations between VOCs indicate the common origin of these compounds. High toluene to benzene ratio (>2) was almost independent of wind speed and direction, indicating that the entire area was severely exposed to fresh vehicular emissions. The absence of relationships between benzene and fossil fuel combustion gaseous pollutants including NO_x and O_3 suggests that evaporations from small chemical industrial complex situated in the S direction from the study site might be the main benzene and toluene emission source in the area. As regards inorganic gaseous pollutants, the relationship between benzene and NO in the N and NE might be related to the intensive anthropogenic activities in the central urban area of Belgrade and petrochemical industry in Pančevo. As presented herein, a combination of bivariate polar plot concentration, correlation, and regression analyses offers unique insight into the individual sources of air pollutants and their concentration dynamics.

Keywords:
benzene, volatile organic compounds, bivariate polar plot analysis.

1. INTRODUCTION

In the recent years, volatile organic compounds (VOCs) became an important environmental issue because of their harmful impacts on human health and the environment. They originate from numerous natural and anthropogenic sources and are involved in a wide range of chemical reactions in the atmosphere including the formation of secondary

Correspondence:
Mirjana Perišić

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

DOI: 10.15308/Sinteza-2020-8-13

organic aerosols and tropospheric ozone. The most abundant VOCs in the atmosphere are hydrocarbons, organic alcohols, halogenated organic compounds, and sulfur compounds [1]. Four aromatic hydrocarbons - benzene, toluene, ethylbenzene, and xylene, commonly known as a BTEX group, are found to be abundant in the urban atmosphere and are considered the representatives of volatile organics [2, 3]. Natural sources of BTEX include crude oil evaporations and emissions from volcanoes and forest fires, while the primary anthropogenic emissions of BTEX compounds are related to motor vehicles, cigarette smoke, petroleum products, and the production and use of paints, lacquers, thinners, rubber, cosmetics, and pharmaceutical products [4].

Among the BTEX, benzene draws special attention because it is considered to be the most toxic and detrimental compound [5]. The International Agency for Research on Cancer (IARC), classified benzene as carcinogenic compound to humans (Group 1), while ethylbenzene has assigned as probably carcinogenic (Group 2). There is no evidence of the toluene and xylenes carcinogenicity in humans (Group 3), but toluene is more soluble in lipids than benzene, and therefore toluene exposure can lead to adverse neurological effects [6].

The research aimed at the investigation of VOC emission sources has been intensified over the last two decades, and many of these studies have been focused on the identification of sources by estimating the ratios of BTEX concentrations in the ambient air and calculating their correlations [7, 8]. Toluene to benzene ratio is commonly used to distinguish the impact of traffic and non-traffic emission sources, while ethylbenzene to benzene ratio, as well as xylenes to benzene ratio are often applied as indicators of photochemical reaction intensity [9].

This study considered the relationships and ratios between benzene, and toluene and inorganic pollutants, in the context of simultaneously measured meteorological parameters, wind speed and wind direction. The bivariate polar plot method in a combination with pairwise statistics was used to indicate the most significant emission sources of benzene and investigate their characteristics in an urban area.

2. MATERIAL AND METHODS

For the purpose of this study, the measurements were conducted over the three-months period, at the Singidunum University building in the urban area of Belgrade (Serbia) during and after the heating season of 2016. The dataset comprised the concentrations of air pollutants and meteorological parameters [10].

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 Road Institute of Belgrade, a building company and beverage factory stockroom are located in the nearest vicinity. In addition, confectionery factory, footwear factory, and several small-scale chemical plants are located 600 m in the NW and S direction, respectively. Approximately 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.

During the campaign, besides the other pollutants including polycyclic aromatic hydrocarbons, gases and metals, the concentrations of benzene, toluene, nitrogen oxides (NO, NO_2, and NO_x), ozone (O_3), and meteorological parameters, including wind characteristics, were collected. The VOCs concentrations were measured in real time using a proton transfer reaction mass spectrometer (Standard PTR-MS, Ionicon Analytik, GmbH, Austria). A detailed description of the PTR-MS method is given elsewhere [11], while calibration procedure was done according to Taipale et al. [12]. Inorganic gaseous pollutant measurements were conducted by using Horiba 370 series devices which enabled continual pollutant concentration monitoring with a 2 minute-resolution data. The APNA-370 device was used for NO_x concentration measurements by a combination of dual cross-flow modulation type chemiluminescence principle and the referential calculation method according to SRPS EN 14211:2013 standard. Continuous monitoring of ozone concentrations was performed by the cross-flow modulated ultraviolet absorption method using APOA-370 device according to standard SRPS EN 14625:2013. Meteorological data were obtained by using Vaisala weather station (Weather Transmitter WXT530 Series).

The statistical data analyses included bivariate polar plots, and improved analytical method which combines bivariate polar plots with pair-wise statistics that provide information on how two pollutants are related to one another. The pair-wise statistics implemented include the weighted Pearson correlation and slope from two linear regression methods. More details about this method can be found in Grange et al. [13, 14].

The analyses were performed with the statistical software environment R [15], by the use of the Openair package [16].

3. RESULTS AND DISCUSSION

As indicated by the elevated pollutant concentrations at low wind speeds (< 1 m s^{-1}), there is a dominance of local traffic emissions of benzene, toluene, and NO in the nearest vicinity of the sampling site (*Fig. 1*).

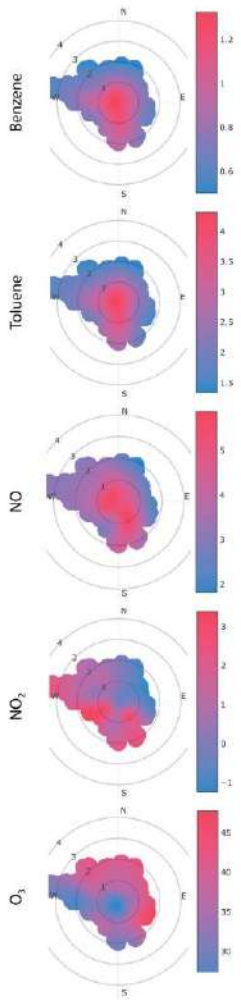


Fig. 1. Bivariate polar plot of benzene, toluene, NO, NO2, NO_2, and O3.

The highest concentrations (> 3 μg m^{-3}) of NO_2 were recorded during S and SW wind episodes which caused the air pollution transport from distant sources including the powerplants "Nikola Tesla" and "Kolubara" situated in Obrenovac and Veliki Crljeni [17]. In addition, several sources of O_3 were observed, and the dominant ones contributing to the concentrations higher than 40 μg m^{-3} were identified in N, E, and SE direction of the measurement site.

Photochemical formation of O_3 in the troposphere is facilitated by the presence of sunlight and elevated levels of precursor pollutants including NO_x and VOCs.

The correlation matrix shows that the two organic compounds, benzene and toluene, were well correlated (r=0.96) (*Fig. 2*), which might suggest their common origin [18]. However, this type of analysis does not provide an indication of which type of source and to what extent contributes to total benzene concentrations. No significant relationships between benzene and NO, NO_2, O_3, and wind parameters were registered although very low values of Pearson's correlation coefficient (r=0.29) between benzene and NO could be an indication of common pollutant emission sources, probably one of them being fossil fuel combustion.

When the benzene and other pollutant concentrations were plotted with a correlation statistic and slope binned by wind speed and direction (*Fig. 3*), the results were revealing more than the mean concentration polar plots and the correlation matrix together.

The strong relationship between benzene and toluene indicates the same dominant source of these compounds. The polar plot of the slope shows that the high toluene to benzene ratio (>2) was almost independent of wind speed and direction, indicating that the whole area was severely exposed to fresh vehicular emission sources [19]. Furthermore, during the N winds (NE and NW, more precisely), benzene and NO were highly correlated.

The zones of high correlation and the relationship between benzene and NO (slope > 6) in the N and NE indicate the impact of intensive anthropogenic activities in the central urban area of Belgrade and petrochemical sources near Pančevo [20].

The absence of a relationship between benzene and fossil fuel combustion gaseous pollutants including nitrogen oxides and O_3 indicated that benzene and toluene evaporations from small chemical industry complexes situated in the S direction from the study site can be considered the main pollution source in the area.

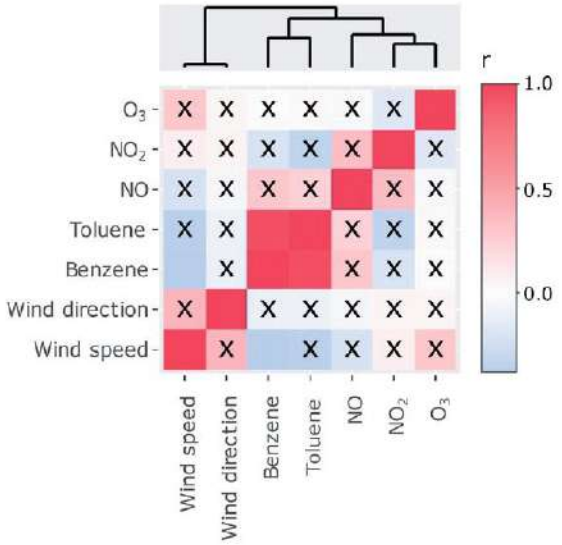


Fig. 2. Correlation matrix of pollutant concentrations and wind parameters.

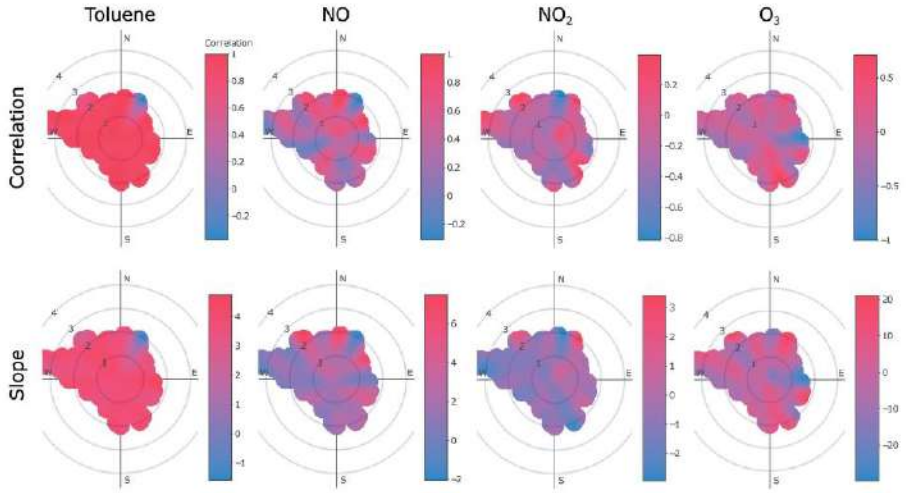


Fig. 3 Polar plot of the correlation (above) and slope (down) between benzene and toluene, NO, NO_2, NO_x, and O_3

4. CONCLUSION

In the field of atmospheric and environmental science, research on the relationship between chemical pollutants in the air and meteorological parameters is tremendously common, and there are diverse techniques of determination of the relationship and comparison. An analysis that also takes into account the correlation of the two pollutants can often be useful because it can lead to the identification of emission source characteristics.

Statistical analysis that includes wind characteristics – bivariate polar plot, applied to the concentrations of benzene, toluene, nitrogen oxides, and ozone, suggests that local sources of air pollution dominate in the study area. With pollutants mainly originating from local emissions, the elevated concentrations occur due to the lack of dispersion, in cases where wind speeds are low (< 1 m s^{-1}). Polar correlations and slopes indicate that there were several types of sources that affected benzene concentrations in the investigated area, and the most important include exhaust from motor vehicles and emissions from industrial processes.

In addition to the statistical receptor modeling, analysis related to spatio-temporal variations and the contribution of other species and sources is required for reliable identification of benzene sources in complex atmospheric environments. The accompanying study should focus on factors such as the pattern of the weekend/weekday activities, regional and long-range transport, planetary boundary layer dynamics, and others meteorological parameters that significantly affect the observed VOC concentrations.

ACKNOWLEDGMENT

The authors acknowledge funding provided by the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS.

REFERENCES

[1] X. Zhang, B. Gao, A. E. Creamer, C. Cao and Y. Li, "Adsorption of VOCs onto engineered carbon materials: A review", Journal of hazardous materials, 338, 2017, pp.102-123.

[2] R. Montero-Montoya, R. López-Vargas and O. Arellano-Aguilar, "Volatile organic compounds in air: sources, distribution, exposure and associated illnesses in children", Annals of global health, 84(2), 2018, pp. 225.

[3] A. Mohammadi, Y. Ghassoun, M.O. Löwner, M. Behmanesh, M. Faraji, S. Nemati, ... and M. Miri, "Spatial analysis and risk assessment of urban BTEX compounds in Urmia, Iran", Chemosphere, 246, 2020, pp. 125769.

[4] F. Hedayatzade and N. Hassanzadeh, "Occurrence, Probable Source, and Health Risk Assessment of Benzene, Toluene, Ethylbenzene, and Xylene Compounds in Ambient Urban Atmosphere in Ahvaz, Iran", Archives of Hygiene Sciences, 9(2), 2020, pp. 152-167.

[5] US EPA 2002. "Toxicological review of benzene (noncancer effects)", In Support of Summary Information on the Integrated Risk Information System (IRIS), U.S. Environmental Protection Agency, Washington, DC, October 2002

[6] IARC, 2014," Monographs on the Evaluation of Carcinogenic Risks to Humans", World Health Organization, International Agency for Cancer Research, Internal Report 14/002, Lyon.

[7] M. Dehghani, M. Fazlzadeh, A. Sorooshian, H. R. Tabatabaee, M. Miri, A. N. Baghani, ... and M. Rashidi, "Characteristics and health effects of BTEX in a hot spot for urban pollution", Ecotoxicology and environmental safety, 155, 2018, pp. 133-143.

[8] Z. Jiang, B. Grosselin, V. Daële, A. Mellouki and Y. Mu, "Seasonal and diurnal variations of BTEX compounds in the semi-urban environment of Orleans, France", Science of The Total Environment, 574, 2017, pp. 1659-1664.

[9] B. T. Thera, P. Dominutti, F. Öztürk, T. Salameh, S. Sauvage, C. Afif, ... and A. Borbon, "Composition and variability of gaseous organic pollution in the port megacity of Istanbul: source attribution, emission ratios, and inventory evaluation", Atmospheric Chemistry & Physics, 19(23), 2019.

[10] S. Stanišić, M. Perišić, G. Jovanović, T. Milićević, S. Herceg Romanić, A. Jovanović, A. Šoštarić, V. Udovičić and A. Stojić, "The PM2.5-bound polycyclic aromatic hydrocarbons: the identification of emission sources and XGBoost pollutant level prediction in indoor and outdoor environment", 2020. (submitted for publication)

[11] 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, 173(3), 1998, pp.191-241.

[12] R. Taipale, T.M. Ruuskanen, J. Rinne, M.K. Kajos, H. Hakola, T. Pohja and M. Kulmala, "Quantitative long-term measurements of VOC concentrations by PTR-MS? measurement, calibration, and volume mixing ratio calculation methods", 2008, hal. archives-ouvertes.fr.

[13] S. K. Grange, A. C. Lewis and D.C. Carslaw, "Source apportionment advances using polar plots of bivariate correlation and regression statistics", Atmospheric Environment, 145, 2016, pp. 128-134.

[14] W. Javed, M. Iakovides, E. G. Stephanou, J. M. Wolfson, P. Koutrakis and B. Guo, "Concentrations of aliphatic and polycyclic aromatic hydrocarbons in ambient PM2.5 and PM10 particulates in Doha, Qatar", Journal of the Air & Waste Management Association, 69(2), 2019, pp. 162-177.

[15] Team, R. C. (2012). R: A language and environment for statistical computing. 2012. Vienna, Austria: R Foundation for Statistical Computing, 10.

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

[17] M. Perišić, A. Stojić, S. S. Stojić, A. Šoštarić, Z. Mijić and S. Rajšić, "Estimation of required PM 10 emission source reduction on the basis of a 10-year period data", Air Quality, Atmosphere & Health, 8(4), 2015, pp. 379-389.

[18] A. Stojić, S. S. Stojić, Z. Mijić, A. Šoštarić, and S. Rajšić, "Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling", Atmospheric Environment, 106, 2015, pp. 71-79.

[19] B. T. Thera, P. Dominutti, F. Öztürk, T. Salameh, S. Sauvage, C. Afif, ... and A. Borbon, "Composition and variability of gaseous organic pollution in the port megacity of Istanbul: source attribution, emission ratios, and inventory evaluation", Atmospheric Chemistry & Physics, 19(23), 2019.

[20] A. Stojić, D. Maletić, S. S. 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, 521, 2015, pp. 19-26.

ARTIFICIAL INTELLIGENCE ATLAS SESSION

MULTIFRACTAL CHARACTERISTICS OF CRITERIA AIR POLLUTANT TIME SERIES IN URBAN AREAS

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

Abstract:
The complexity of spatial-temporal air pollutant concentration dynamics requires innovative modeling investigation approaches. The details of non-linear nature of pollutant behavior cannot be revealed by conventional approaches, but fractal and Hurst rescaled analyses allow the quantification of pollutant dynamics structure via self-similarity and scale invariance. In this study, we applied multiscale multifractal analysis (MMA) to investigate the complex time-series of criteria air pollutants (PM_{10}, $PM_{2.5}$, NO_X, SO_2, CO and O_3). The results showed that PM_{10} and $PM_{2.5}$ concentrations are more stable compared to gaseous oxides and exhibit less prominent multifractality. Out of gaseous contaminants, CO is confirmed to be less chemically reactive than NO, NO_2, NO_x, SO_2 and O_3 under the same atmospheric conditions in urban and semi-urban area. As concluded, the multifractal analysis presented herein can enhance our understanding of specific pollutant dynamic features and support relevant sectors to control the pollutant release and distribution.

Keywords:
multiscale multifractal analysis (MMA), suspended particle matter (PM), criteria air pollutants, air pollution, urban environment.

1. INTRODUCTION

Air pollution poses a major threat to health and climate, and it has been estimated to become the leading cause of mortality worldwide by 2050 [1]. Seven million premature deaths every year occur as a result of stroke, cardiovascular and respiratory diseases, all of which can be related to detrimental effects of high air pollutant concentration exposure [2]. Although the contaminant emissions have been reduced, about 30% of citizens in European urban areas are still exposed to air pollutant levels that exceed EU air quality guidelines defined with the aim of health protection [3]. Therefore, the research of air pollutant spatial-temporal behavior in urban areas appears to be one with highest potential to contribute to society.

Correspondence:
Gordana Jovanović

e-mail:
gjovanovic@singidunum.ac.rs

DOI: 10.15308/Sinteza-2020-29-34

In the atmosphere, air pollutant lifetimes are in a range from hours to years, depending on air mass transport processes, undergoing transformations via chemical reactions and/or particle deposition. Although usually assumed to appear by a chance as a linear or a single-compartment processes, variations of air pollutant levels over time are neither simple nor random in their behavior. The factors like mutual pollutant interactions, meteorological parameters, and the emission source type and intensity, affect the air quality significantly and make the air pollutant temporal dynamics complex and difficult to analyze and understand.

The most striking feature of the pollutant concentration time-series refers to their long-term memory, which reflects how the pollutant past behavior patterns affects its future levels [4]. Self-similarity and scale-invariance are considered to be the memory characteristics of decisive importance concerning the pollutant time-series and therefore, vital for the estimation of long-range correlations and precise forecasting [5]. These structural properties can be captured by fractal and Hurst rescaled analyses that define fractals as self-similar shapes of the "structure holding the key to the whole structure" [6, 7, 8]. The multifractal approach has been applied for estimation of constitutional features and forecasting of air pollutant time-series [9, 10, 11, 12, 13].

In this study, we focus on the investigation of time-series of criteria air pollutants (PM_{10}, $PM_{2.5}$, NO_2, SO_2, CO and O_3) across Belgrade with the aim to obtain more comprehensive understanding of the investigated species behavior and fate. Unlike the previous findings related to the Balkans, that were limited in terms of the result interpretations and sampling time, the findings presented herein are based on a long-term data and advanced investigation methods that enable pollutant temporal dynamics to be explored more profoundly.

MATERIALS AND METHODS

The data used for the analyses were obtained from the regular monitoring network of Institute of Public Health Belgrade and multi-year measurement campaign conducted from 2011 to 2016 at 9 sites covering 17 municipalities located in Belgrade (Serbia) urban, semi-urban and rural areas. The dataset comprised the concentrations of inorganic gaseous pollutants (CO, SO_2, NO, NO_2, NO_x and tropospheric O_3) and particulate matter (PM10 and PM2.5) registered at the following locations: four of urban-traffic type (Institute of Public Health – IPB, New Belgrade – NBG, Zemun – ZEM and Slavija – SLA), two of urban-industry type (Obrenovac – OBR and Lazarevac – LAZ), one suburban location (Ovča – OVC), and two locations classified as rural-industry sites (Grabovac – GRA and Veliki Crljeni – VC).

We applied multiscale multifractal analysis (MMA) [13] to investigate fractal characteristics of complex PM_{10}, $PM_{2.5}$, NO_x, SO_2, CO and O_3 time series. MMA is a generalization of the standard MF-DFA, which adds the dependence on scale, providing a broader analysis of the fluctuation properties, as well as more general and stable results [14].

The aim of this study was to investigate the pollutant time-series, as described below. We delineated the strength of multifractality by the multifractal singularity, Q. The bigger the Q is, the more pollutant concentration time-series is influenced by Q interval fluctuations in negative or positive direction, while the absence of fluctuations results in Q equals to 0 and represents a mono-fractal behavior [10]. In addition, the span of multifractal singularity ($\Delta\alpha = \alpha_{max} - \alpha_{min}$) is an alternative way to study the strength of multifractality; the bigger $\Delta\alpha$ is, the stronger the multifractality degree is [5].

Besides, we considered Hurst exponent (H) which indicates the persistence level of the pollutant concentration time series denoted by the following values [5, 15, 16]:

- $H \geq 1.5$ represents brown noise uncorrelated processes with infinite memory;
- $H=1$ refers to pink noise, i.e. the most prominent, stable and adaptable fractal phenomena with long-term memory;
- $0.5<H<1.5$ describe random walk processes with long-range correlated and persistent structure;
- $H=0.5$ refers to white noise involving uncorrelated random variables with no memory;
- $H<0.5$ is associated with the anti-persistent increments which possess anti-correlated structure.

RESULTS AND DISCUSSION

In general, a sharp negative slope of H versus Q surfaces clearly shows the multifractal behavior of all examined pollutant concentrations, which is opposite to the straight line with zero gradient, being an indicator of monofractal systems [15]. The H values above 0.5 suggest nonrandom pollutant fluctuations over time and reflect non-stationary signals with long-range dependent structure, which follow a power law.

The very few cases of H<5, which indicate oscillating processes have occurred only at Q=5 for: O_3 being measured at urban-traffic and urban-industry sampling sites (NBG and LAZ), at scales between 50 and 130 hours; NO and NO_x registered at urban–ZEM location between 30 and 140 hours, and SO_2 being registered at urban and rural locations (OBR and GRA) impacted by the coal-fired power station emissions between 200 and 360 hours. Discrepancies in SO_2 dynamics in rural area could be attributed to intermittent pollution accidents related to the emissions from industrial sources and strong wind events. More comprehensive evaluation of O_3 and NO_X pathways in near-ground atmosphere, as well as the analysis of noted concentration deviations, would require more detailed investigation on volatile organic compounds (VOCs) and their interactions with NOX and O_3. Unlike many pollutants, ground-level O_3 is a product of photolytic decomposition of precursors rather than a component of primary emissions. Generation and removal of O_3 is strongly affected by the presence of NO_x, their type, lifetime and the way they interact with VOC, HO· and ROO· in a cycle of catalytic reactions. The interactions are impacted by location-specific factors, including pollutant sources and meteorological parameters.

As far as individual pollutants are concerned, H values between 0.70 and 1.5 showed the long-range persistent PM10 time-series at almost all sampling sites (*Fig. 1*). The most prominent variations of H values were apparent at lower and upper Q boundaries (-5, 5) exhibiting maximums at the small scales and in the scale range between 100 and 200 hours. Since PM_{10} represents a mixture of microscopic solid and liquid suspended matters involved in inherent condensation and nucleation, the registered cross-correlation behavior of meteorological factors and PM_{10} was not observed in the case of meteorological factors and the investigated gases [10, 17]. Hurst exponent exceeded 1.5 only at rural GRA and suburban OVC sampling sites, indicating pollutant concentration patterns which can be referred to a brown noise. These behavior pattern appeared at small temporal scales, up to 60 hours, and in the domain of negative Q values implying that the PM_{10} segments are more prone to small and large spatial-temporal variations. Air quality in rural areas being dominated by industrial emissions is more sensitive to fluctuations of meteorological factors compared to air quality in urban sites which is mainly burdened by alternating traffic emissions. After reached Brownian increments, the extremely steep slopes of H versus Q surfaces characterized the PM_{10} concentration behavior patterns at a time scale of up to 150 hours.

The results suggest a strong multifractal nature of PM_{10} which weakened over time and approached to the "healthy complex system" of pink noise.

Time-series of $PM_{2.5}$ were investigated at different sites, suburban OVC in residence area influenced by domestic heating, and VEL rural site in in the coal mining industrial area. Strong multifractal nature of $PM_{2.5}$ referring to persistence and long-range correlations, was apparent in the scale range between 30 and 150 hours, as well as in the range between the 30 and 90 hours at OVC and VEL, respectively (*Fig. 2*). While multifractal features of $PM_{2.5}$ trend decayed with time, they were maintained in the area of negative Q domains.

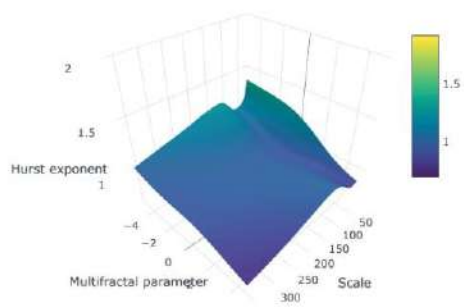


Fig. 1. MMA-derived Hurst surfaces for PM_{10} sampled at typical urban location in the Belgrade centre.

However, at a time scale from the 150 to 360 hours $PM_{2.5}$ concentration behavior patterns were different for Q>0, reflecting the probable mono-fractal characteristics. $PM_{2.5}$ and PM_{10} multifractal patterns resemble for Q interval between 0 and 5, which illustrates their monofractal nature, whereas differences in their behavior are reflected in the Q interval between -5 and 0, which supports previous findings on dissimilarities among differently sized PMs in urban areas [18]. A strong multifractal behavior of PM sized below 5 μm, and a weak and/or monofractal stochastic behavior of PM>5 μm was pronounced. In addition, the generalized distance coefficient (0.069), which was above threshold value of 0.065, indicated that the $PM_{2.5}$ and PM_{10} time-series are statistically different.

Multiscale multifractal analysis of SO_2 concentrations showed highly distinctive patterns depending on the sampling site (*Fig. 3*). The absolute similarities in the pollutant time-series between different sampling sites could not be expected due to numerous emission

sources, chemical reactions and meteorology, which are location-specific. Correspondingly to the findings for PM_{10}, the highest H values (from 1.64 at IPH to 2.37 at OBR) were observed at small time scales ranging from the 30 and 100 hours and in the area of Q negative domains (from -5 to -1).

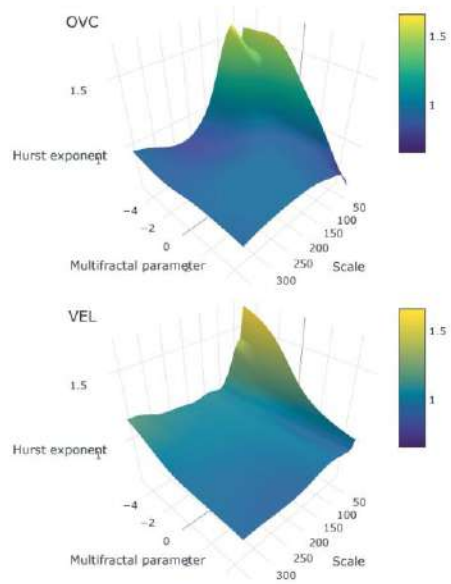


Fig. 2. MMA-derived Hurst surfaces for $PM_{2.5}$ sampled at two rural/semi-urban sites in Belgrade.

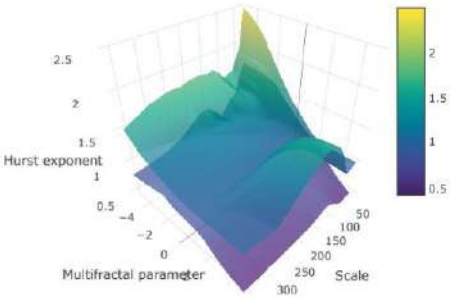


Fig. 3. MMA derived Hurst surfaces for SO_2 in Belgrade.

Such high H maximums induced steeper slopes of H versus Q surfaces implying unequivocal multifractality of SO_2 time-series. In addition, the results suggest that random fluctuations of SO_2 do not dissipate over time and possess brown-noise infinite memory. However, the most dominant characteristics of the SO_2 concentration behavior pattern are a long-range correlated structure and fast evolving fluctuations (H<1). Influence of PM_{10} presence and potential SO2 adsorption is a possible cause of SO_2 level variability, and therefore, multifractality. The studies aimed at chemical composition analysis identified material containing inorganic/black carbon and SO_2 as key constituents of PM_{10} and $PM_{2.5}$ [19, 20].

The time-series of nitrogen oxides (NO, NO_2 and NO_x) resemble each other at almost all sampling sites confirming the same origin and interrelations between these compounds (*Fig. 4*). Contrary to the PM_{10} and SO_2 segments, H value reached plateau-shaped maximums higher than 1.5 at a larger scale from the 150 to 360 hours in the negative domains of Q interval (small fluctuations between -5 and ≈ -2). However, the dominant H values (0.5–1.5) reflect persistent/stable pollutant concentration behavior patterns and this trend is likely to be maintained over time. Chemical reactions with O_3 and volatile precursors in which nitrogen oxides are involved provide a possible explanation for the strong multifractality of NO, NO_2 and NO_x. The exceptions from described phenomenon are behavior patterns of NO, NO_2 and NO_x levels registered at two urban sites – SLA and LAZ.

In contrast to the other gaseous pollutants, CO showed less pronounced multifractality (*Fig. 5*) in urban areas. The highest peaks were observed at the time scale of 30 hours, and H exceeded 1.5 only between the 30 and 40 hours at an urban site – IPH. However, the multifractality in CO level behavior patterns weakened in the area of large fluctuations corresponding to positive Q domains approaching mono-fractal behavior between 100 and 360 hours, which reflects the stability of CO compared to NO, NO_2, NO_x and SO_2.

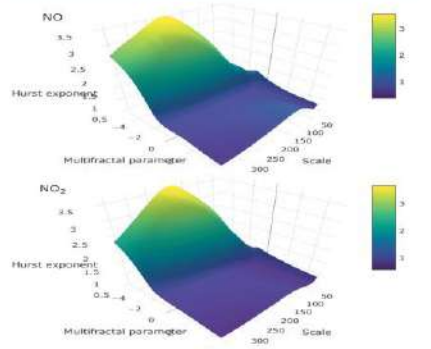


Fig. 4. MMA-derived Hurst surfaces for nitrogen oxides (IPH).

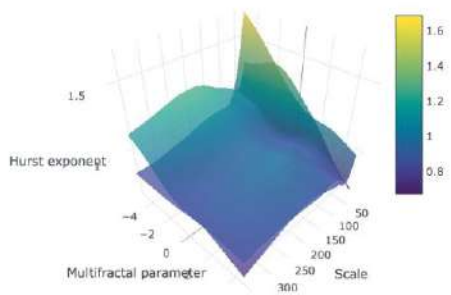


Fig. 5. MMA derived Hurst surfaces for CO in Belgrade.

CONCLUSION

The behavior of suspended particles (PM_{10} and $PM_{2.5}$) appeared to be more consistent in nature and exhibit less prominent multifractality than gaseous pollutants which is expected concerning the chemical reactivity of NO, NO_2, NO_x, and SO_2. Long-range correlated structure of PM_{10} time-series is likely to be driven by the variability of meteorological factors, while the corresponding patterns of SO_2 and nitrogen oxides could be a consequence of accidental pollution episodes and chemical reactions. Bearing upon the previously described complexity of the parameters affecting pollutant time-series, we emphasize that discussion above could not be used as a definite source of information regarding the pollutant behavior. However, what is of outmost importance is that given results are not biased by mean pollutant concentrations and may be regarded as representative for the studied area. For this reason, the study could serve as a standpoint to evaluate pollutant predictability over urban/rural areas based on the pollutant multifractal nature. Beside commonly used methods of statistical analysis and source apportionment, the multifractal analysis presented herein could reveal specific pollutant pathways and support relevant economy and governmental sectors to control the pollutant release and distribution.

ACKNOWLEDGMENT

The authors acknowledge funding provided by the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI – ATLAS.

REFERENCES

[1] Organization for Economic Co-Operation and Development (OECD) Environmental Outlook to 2050: The Consequences of Inaction. Organization for Economic Co-Operation and Development (OECD); Paris, France: 2012.

[2] WHO, https://www.who.int/health-topics/air-pollution#tab=tab_1

[3] European Environment Agency (EEA) Air Quality in Europe—2017 Report EEA Report No 13/2017. European Environment Agency (EEA); Copenhagen, Denmark: 2017

[4] A. Chelani, "Long-memory property in air pollutant concentrations." Atmos. Res. 2016, pp. 1-4.

[5] T. Stadnitski, "Measuring fractality." Front. Physiol. 2012, pp. 1-13.

[6] H.E. Hurst, "Long-term storage capacity of reservoirs." Transactions of the American Society of Civil Engineers 1951, pp 770-799.

[7] B.B. Mandelbrot, J.R. Wallis, "Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence." Water Resour. Res. 1969, pp. 967-988.

[8] BB. Mandelbrot, "The Fractal Geometry of Nature." New York: W.H. Freeman, 1977.

[9] Q. Dong, Y. Wang, P. Li, "Multifractal behavior of an air pollutant time series and the relevance to the predictability." Environ. Pollut. 2017, pp. 444-457.

[10] H.d, He, "Multifractal analysis of interactive patterns between meteorological factors and pollutants in urban and rural areas." Atmos. Environ. 2017, pp. 47-54.

[11] T. Plocoste, R. Calif, S. Jacoby-Koaly, "Temporal multiscaling characteristics of particulate matter PM10 and ground-level ozone O3 concentrations in Caribbean region." Atmos. Environ. 2017, pp. 22-35.

[12] A. Stojić, S. Stanišić Stojić, I. Reljin, M. Čabarkapa, A. Šoštarić, M. Perišić, Z. Mijić, "Comprehensive analysis of PM10 in Belgrade urban area on the basis of long-term measurements." Environ. Sci. Pollut. Res., 2016, pp. 10722-10732.

[13] A. Stojić, S. Stanišić Stojić, M. Perišić, Z. Mijić, "Multiscale multifractal analysis of nonlinearity in particulate matter time series." Paper presented at the sixth international WeBIOPATR workshop and conference, Serbia, Belgrade, September 6-8, 2017.

[14] J. Gieraltowski, J.J. Żebrowski, R. Baranowski, "Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia." Phys. Rev. 2012, p. 021915.

[15] E.A.F. Ihlen, "Introduction to multifractal detrended fluctuation analysis in Matlab." Front. Physiol. 2012, pp. 1-18.

[16] E. Molino-Minero-Re, F. García-Nocetti, H. Benítez-Pérez, "Application of a Time-Scale Local Hurst Exponent analysis to time series." Digit. Signal Process. 2015, pp. 92-99.

[17] C. Zhang, Z.W. Ni, L.P. Ni, "Multifractal detrended cross-correlation analysis between PM2.5 and meteorological factors." Physica A 2015, pp. 114-123.

[18] Y. Xue, W. Pan, W.Z. Lu, H.D. He, "Multifractal nature of particulate matters (PMs) in Hong Kong urban air." Sci. Total Environ. 2015, pp. 744-751.

[19] N.P. Ivleva, U. McKeon, R. Niessner, U. Pösch, "Raman Microspectroscopic Analysis of Size-Resolved Atmospheric Aerosol Particle Samples Collected with an ELPI: Soot, Humic-Like Substances, and Inorganic Compounds." Aerosol Sci. Tech. 2007, pp. 655-671.

[20] U. Pöschl, "Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects." Angewandte Chemie 2005, pp. 7520-7540.

ARTIFICIAL INTELLIGENCE ATLAS SESSION

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.

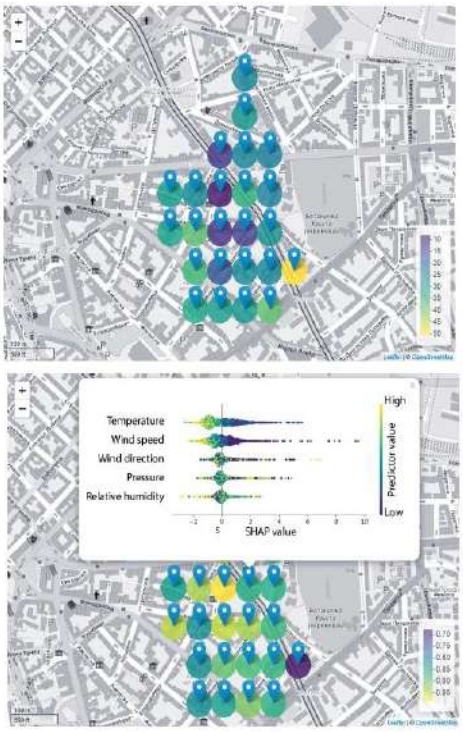


Fig. 1. Benzene forecast based on meteorological parameters – relative error [%] (above) and SHAP values and predicted/observed correlation coefficients (below).

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.

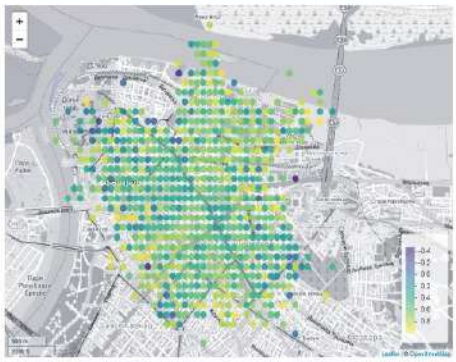


Fig. 3. Toluene and NO correlation coefficient.

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.

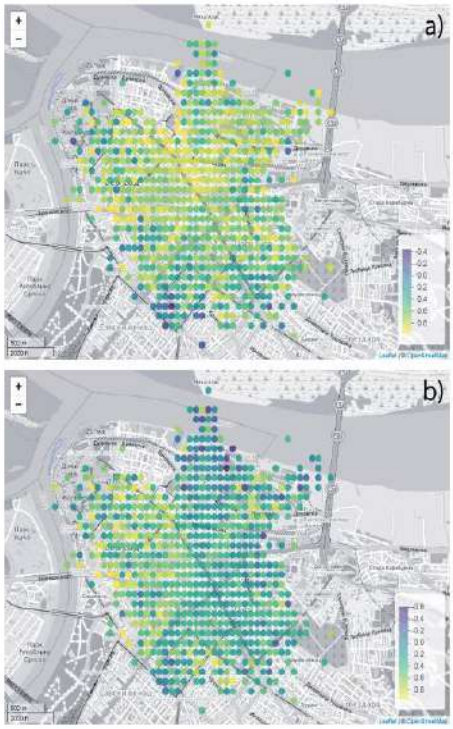


Fig. 4. Benzene correlation coefficients with CO (above) and NOx (below).

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.

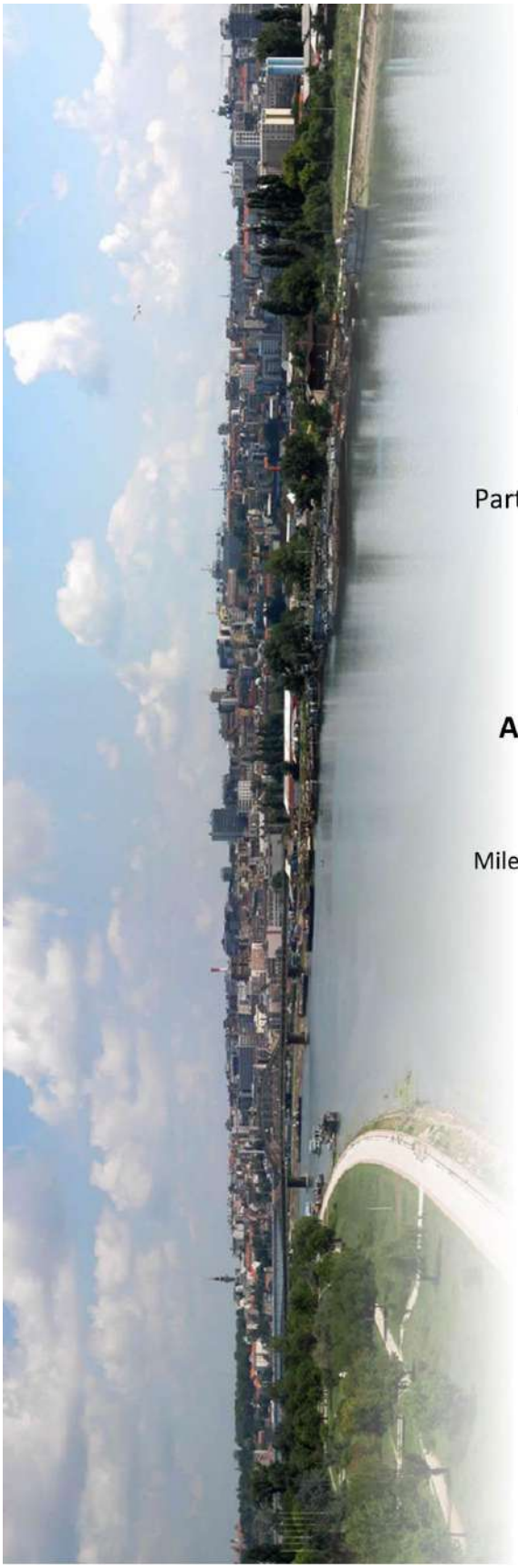
[10] M. V. García, J. L. Aznarte, "Shapley additive explanations for NO_2 forecasting," *Ecol. Inform,* vol. 56, pp. 101039, 2020.

[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 2019

The Seventh International WEBIOPATR
Workshop & Conference
Particulate Matter: Research and Management

Abstracts of Keynote Invited Lectures and Contributed Papers

Milena Jovašević-Stojanović and Alena Bartoňová, Eds

Public Health Institute of Belgrade

Belgrade 2019

ABSTRACTS OF KEYNOTE INVITED LECTURES AND CONTRIBUTED PAPERS

The Seventh International WeBIOPATR Workshop & Conference

Particulate Matter: Research and Management

WeBIOPATR 2019

1st to 3rd October, 2019

Belgrade, Serbia

Editors

Milena Jovašević-Stojanović

Alena Bartoňová

Publisher

Public Health Institute of Belgrade

Prof. Dr Dušanka Matijević, Director

Boulevar Despota Stefana 54a

Serbia, 11000 Belgrade

Printed by

Printing office of the Public Health Institute of Belgrade

Number of copies

150

ISBN 978-86-83069-56-9

www.zdravlje.org.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
Radim J. Šrám, Czech Republic
Renata Kovačević, Serbia
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
Dejan Lekić, Serbia
Dragan Alavantić, Serbia
Ivan Lazović, Serbia
Jasmina Jović-Stošić, Serbia
Maja Jovanović (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
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 Seventh WeBIOPATR Workshop and Conference,

Particulate Matter: Research and Management, WEBIOPATR 2019

is supported by:

Ministry of Education, Science and Technological Development of Republic of Serbia

PREFACE

The International Workshop and Conference, Particulate Matter: Research and Management – WeBIOPATR is a biennial event held in Serbia since 2007. The conference addresses air quality in general and particulate matter specifically. Atmospheric particulate matter arises both from primary emissions and from secondary formation in the atmosphere. It is one of the least well understood local and regional air pollutants, has complex implications for climate change, and is perhaps the pollutant with the highest health relevance. It also poses many challenges to monitoring.

By WeBIOPATR, we aim to link the research communities with relevance to particulate matter with the practitioners of air quality management on all administrative levels, in order to facilitate professional dialogue and uptake of newest research into practice. The workshops usually draw an audience of about 70, and attract media attention in Serbia. It enjoys support of the responsible authorities: Ministry of Education, Science and Technological Development, Ministry of Health, Ministry of Environment, and the Serbian Environmental Agency whose sponsorship is indispensable and gratefully acknowledged. We enjoy also support of international bodies such as the WHO.

The 1st WeBIOPATR Workshop was held in Beograd, 20.-22. May 2007, associated with a project funded by the Research Council of Norway. The 2nd workshop was held in Mecavnik, Serbia, 28.8.-1.9. 2009. WeBIOPATR2011 was held in Beograd 14.-17. 11. 2011 and for the first time, included a dedicated student workshop. WeBIOPATR2013 was held in Beograd 2.-4. 10. 2013. It covered the traditional PM research and management issues, discussions on how to encourage citizens to contribute to environmental governance, and how to develop participatory sensing methods. WeBIOPATR2015 was held in Beograd 14.-16.10. 2015. Own sessions were devoted to sensor technologies for air quality monitoring, utilizing information and input from the EU FP7 funded project CITI-SENSE (http://co.citi- sense.eu) and the EU COST action EuNetAir (www.eunetair.it). WeBIOPATR2017, the 6th conference, was held in Beograd 6.-8.9. 2017, with a wider than before Western Balkan participation.

WeBIOPATR2019 will be held 1.-3 -10-2019 in the Mechanical Faculty, University of Belgrade. It has attracted a record 58 contributions, and is bringing together scientists from 12 countries, documenting that the issues of atmospheric pollution, with their wide implications for climate change, human health and ecosystem services, are no less important today.

We are grateful to our unrelenting national and international partners for their support for this event.

Welcome to Beograd, and have a stimulating and productive time!

Milena Jovašević-Stojanović and Alena Bartoňová

TABLE OF CONTENTS

4.2. PARSING ENVIRONMENTAL FACTORS WHICH SHAPE PARTICULATE MATTER POLLUTION USING EXPLAINABLE ARTIFICIAL INTELLIGENCE

A. Stojić (1), M. Perišić (1), G. Jovanović (1), S. Stanišić (2), N. Stanić (2), T. Milićević (1)
(1) Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Serbia
(2) Singidunum University, Serbia
andreja.stojic@ipb.ac.rs

The unpredicted rate and the diversity of modern world development lead to unprecedented changes in the environment which require deep understanding of their nature and measures that might be undertaken to prevent further environment deteoritation. To tackle the root causes which shape air pollution, understanding the fundamental mechanisms of nature must rely on highly sophisticated machine learning algorithms and the interpretation frameworks aimed at delivering explainable predictive analytis (Stojić et al. 2018). In this paper, we utilize the statistical analysis of SHapley Additive exPlanation (SHAP) values to reveal the environmental conditions which shape PM_{10} pollution in an urban area (Belgrade, Serbia).

To examine the evolution of PM levels in the context of the urban environment, eXtreme Gradient Boosting regression analysis (XGBoost) was performed to obtain dependency between PM_{10} and criteria air pollutants (NO_x, NO, NO_2, SO_2, CO), volatile aromatics (benzene, toluene, ethylbenzene and xylene), meteorological factors (visibility, ceil height, wind speed and direction, relative humidity, dew point, atmospheric pressure, temperature and 25 1-degree-Global Data Assimilation System surface parameters), as well as temporal and seasonal variations (trend, day length, daylight, weekday, weekend, sunrise angle, month and season). All the measured concentrations and parameters were obtained from the automatic monitoring network of the Institute of Public Health Belgrade, Serbia. XGBoost is a supervised ensemble learning method which implements iterative combining of ensembles of weak prediction models into a single strong learner (Stojić et al. 2019). The dataset was divided into stratified training (80%) and validation (20%) sets. Hyperparameter tuning was implemented using an advanced grid search and stratified cross-validation replicated ten times. Moreover, to test the stability of the obtained model, 100 times replicated bootstrap procedure was performed.

Subsequently, SHAP framework was applied on the obtained regression function to deliver model explanations. The framework is based on unification of additive attribution algorithms, individualized for each prediction, offering uniquely consistent and locally accurate attribution values. It overcomes the drawback of other methods inconsistency, suppressing the possibility of underestimating the importance of a feature with a certain attribution value. Finally, fuzzy clustering of SHAP attributions was performed to obtain clusters of environmental factors (forces) which govern PM evolution in complex urban environment.

Six clusters of forces were identified, all dominated by CO, but with the ambivalent impact on PM levels. Namely, the clusters which represent the ambient in which the highest PM_{10} concentrations occur, are related to the highest concentrations of CO and benzene. On the contrary, essentially different interrelations between these compounds can be attributed to a lower concentration range of PM, suggesting different emission sources regime and different atmospheric chemistry. Also, visibility appears to be extracted as the most important variable, which clearly depicts fundamentally different atmospheric conditions regarding PM occurrence in different environmental clusters.

Acknowledgments

This study was performed as a part of projects no. III43007 and no. III41011, which were funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, within the framework of integrated and interdisciplinary research for the period 2011–2019.

REFERENCES

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, Nova Science Publishers, NY, USA.

10.3. RECEPTOR ORIENTED MODELING OF URBAN PARTICULATE AIR POLLUTION: SOURCE CHARACTERIZATION AND SPATIAL DISTRIBUTION

M. Perišić (1), A. Stojić (1), S. Stanišić (2), G. Jovanović (1)
(1) Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Serbia
(2) Singidunum University, Serbia
mirjana.perisic@ipb.ac.rs

Urban atmosphere is a complex system, in which air pollutants' levels are not only driven by the features of emission sources and variations of meteorological conditions, but also by the pollutant interactions and area-specific factors which have an impact on atmospheric chemistry. Nowadays, the application of advanced analytical methods is required to gain reliable information for better understanding of underlying factors which shape the air pollution phenomena in the urban environment. In this paper, we used the methodology based on receptor oriented modeling (Stojić and Stojić 2017) to investigate the spatial distribution of pollutants, their concentrations and potential emission sources in the urban core of the City of Belgrade, Serbia.

The database used in this study included the concentrations of suspended particulates matter (PM_{10}), inorganic and organic gaseous pollutants (NO_X, SO_2, CO, benzene and toluene), measured during the period of two years at the monitoring site Institute of Public Health Belgrade (Serbia). The chosen station, within the Air Quality Monitoring Network of Belgrade, is located in a densely populated part of the city, near intensive traffic activities, exposed to emission of local fireboxes and central district heating, as well as under the influence of various industrial emissions. Based on the pollutant concentrations and meteorological parameters (wind speed and wind direction), measured at the sampling site, the developed receptor-oriented model provides a detailed information on pollutants' concentrations and their mutual correlations in a wide area, not covered by the regulatory monitoring network.

The obtained interactive maps contain the results of the correlation analysis, as well as the relations between the concentrations of benzene and toluene, and measured air pollutants (PM_{10}, NOx, CO, and SO_2). The highest correlations of benzene and PM_{10} ($r = 0.8$) were observed along the large traffic routes, in Brankova Street and Bulevar Kralja Aleksandra Street, as well as in the northwestern part of the city center, suggesting that the intensive traffic represents the common source of gaseous and particulate pollution (Stojić et al. 2018). Furthermore, high correlations ($r > 0.7$) between benzene and combustion gases (NOx, CO, SO_2) in the western region reflect the influence of distant sources associated with the thermal power plants Nikola Tesla in Obrenovac, while relatively low correlations between benzene and combustion gases in the northern and eastern part of the city indicate that benzene in this area possibly originates from industrial-petrochemical emissions near Pančevo (Stojić et al. 2015). Also, high correlations ($r > 0.7$) between toluene and NOx, which are indicators of fossil fuel combustion from traffic and heating, suggest the shared origin of these compounds in all the parts of the city included in the analysis, except for the old city core and Kalemegdan (north-west), where the toluene is probably present due to the enhanced retention of aged air in streets of urban-canyon type.

Acknowledgments

This study was performed as a part of projects no. III43007 and no. III41011, which were funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, within the framework of integrated and interdisciplinary research for the period 2011–2019. Also, this paper presents a part of the results of the project realized with the support of the Green Fund of the The Ministry of Environmental Protection of the Republic of Serbia.

REFERENCES

Stojić, A., Stojić, S. S., Šoštarić, A., Ilić, L., Mijić, Z. and 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.

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

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, Nova Science Publishers, NY, USA.

11.6. EXPLAINABLE RELATIONS OF PARTICULATE MATTER AND ENVIRONMENTAL FACTORS IN AN URBAN AREA

G. Jovanović (1), A. Stojić (1), M. Perišić (1), S. Stanišić (2), N. Stanić (2) and T. Milićević (1)
(1) Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Serbia
(2) Singidunum University, Serbia
gordana.vukovic@ipb.ac.rs

In this study we focused on examining the dependencies between particulate matter (PM), and other air pollutants and atmospheric conditions in an urban environment. Briefly, eXtreme Gradient Boosting regression (XGBoost) was performed to obtain the relations between concentrations of PM_{10} and volatile organic compounds, inorganic gaseous pollutants, measured and modeled meteorological parameters, as well as parameters representing temporal and seasonal variations (Stojić et al. 2018). The relations were further analyzed by the use of Shapley Additive exPlanation (SHAP) summary and dependence distributions. The details about the methods applied are given elsewhere (Stojić et al. 2019).

The results indicate that, although CO individually achieves the highest impact on PM_{10} levels, meteorological conditions play the major role in shaping its environmental fate in an urban environment. Among other polluting species, the relation with benzene can be considered to be substantial, while the impact of other compounds found in the urban atmosphere, such as inorganic gaseous pollutants or other aromatics, can be considered to be significantly lower. The influence of CO is different, depending on the CO concentration range. The concentrations below 1 mg m^{-3} are associated with lower PM; in the CO concentration range between 1 and 1.5 mg m^{-3} the influence on PM levels can be considered negligible, while an increase in CO above 1.5 mg m^{-3} is accompanied by an increase of PM_{10}. This impact is largely determined by seasonality, indicating a strong influence of emission source, particularly the combustion of fossil fuels for heating purposes. However, it can be noticed that, even during the colder part of a year, low CO concentrations, being always followed by low concentrations of NO_x, SO_2 and volatile aromatics, stay related with lower PM levels. High-level CO concentration range is associated with complex interactions with other environmental factors, which need to be further addressed.

Low ceil height, even when being registered along with low visibility, does not have to be unambiguously associated with increased levels of PM. However, it can be seen that low cloudiness generally leads to a decrease in PM concentrations, which cannot be attributed to an increase in humidity or wet deposition, because with the highest relative humidity (Rh) values, the contribution of low cloudiness to an increase in PM can be extremely high. On the other hand, several of the most extreme PM events, associated with the highest impact on PM concentration, occurred with good visibility and max ceil height. Low visibility conditions, on their own, lead to an increase in PM levels. The lowest impact was observed for the highest concentrations of CO and benzene, and the lowest concentrations of other aromatics during the colder part of a year, and thus can be attributed to the the activation of the combustion emission sources emitting a lower share of PM. The ambient conditions which correspond to lower Rh (lower than 60%) and higher visibility contribute to the decrease in PM concentrations. Only with humidity above 80% and reduced visibility, an increase in PM concentrations up to about 10 μg m^{-3} was evident. The highest positive impact of relative humidity cannot be associated with high low and medium cloudiness, while the situation is indeterminate as for high cloudiness. At the end of the analyzed period, there was a noticeable change in the impact of the temporal trend on PM levels. The results clearly identify the suppression of the combustion sources which, in addition to CO emission, contain PM. This event was followed by the appearance of a pronounced combustion source which, besides CO, emits a large amount of particulatte pollution.

ACKNOWLEDGMENTS

This study was performed as a part of projects no. III43007 and no. III41011, which were funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, within the framework of integrated and interdisciplinary research for the period 2011–2019.

REFERENCES

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, Nova Science Publishers, NY, USA.

49th International October Conference on Mining and Metallurgy

PROCEEDINGS

Editors:
Nada Štrbac
Ivana Marković
Ljubiša Balanović

Bor Lake, Serbia
October 18-21, 2017

IOC 2017
International October Conference

PROCEEDINGS,

49th INTERNATIONAL OCTOBER CONFERNCE

on Mining and Metallurgy

Editors:

Prof. dr Nada Štrbac

Doc. dr Ivana Marković

Doc. dr Ljubiša Balanović

University of Belgrade, Technical Faculty in Bor

Technical Editor:

M. Sc. Uroš Stamenković

University of Belgrade, Technical Faculty in Bor

Publisher: University of Belgrade, Technical Faculty in Bor

For the publisher: Dean Prof. dr Nada Štrbac

Circulation: 200 copies

Printed by "Happy trend DOO", Zaječar, 2017

ISBN 978-86-6305-066-2

CIP - Каталогизација у публикацији - Народна библиотека Србије, Београд

622(082)
669(082)

INTERNATIONAL October Conference on Mining and Metallurgy (49 ; 2017 ; Bor Lake)
Proceedings / 49th International October Conference on Mining and Metallurgy - IOC 2017, Bor Lake, Serbia, October 18-21, 2017;
[organized by] University of Belgrade, Technical Faculty Bor and Mining and Metallurgy Institute Bor; editors Nada Štrbac, Ivana Marković, Ljubiša Balanović. - Bor : University of Belgrade, Technical Faculty, 2017 (Zaječar : Happy trend). - XXIII, 664 str. : ilustr. ; 25 cm

Tiraž 200. - Bibliografija uz svaki rad. - Registar.

ISBN 978-86-6305-066-2

a) Рударство - Зборници b) Металургија - Зборници

COBISS.SR-ID 246349324

Bor Lake, Serbia, October 18-21, 2017

Conference is financially supported by
The Ministry of Education, Science and
Technological Development of the
Republic of Serbia

SCIENTIFIC COMMITTEE

Prof. dr Nada Štrbac (Serbia) - president
Prof. dr Radoje Pantović (Serbia) - vice-president
Prof. dr Grozdanka Bogdanović (Serbia)- vice-president
Prof. dr Dragoslav Gusković (Serbia) - vice-president
Prof. dr Aleksandar Dimitrov (Macedonia)
Dr Ana Kostov (Serbia)
Dr Andrei Rotaru (Romania)
Prof. dr Andelka Mihajlov (Serbia)
Prof. dr Batrić Pešić (USA)
Prof. dr Boštjan Markoli (Slovenia)
Prof. dr Boyan Boyanov (Bulgaria)
Prof. dr Branka Jordović (Serbia)
Prof. dr Carl Heinz Spitzer (Germany)
Prof. dr Costas Matis (Greece)
Prof. dr Dejan Tanikić (Serbia)
Prof. dr Desimir Marković (Serbia)
Prof. dr Dimitris Panias (Greece)
Prof. dr Dimitriu Sorin (Romania)
Prof. dr Dragan Manasijević (Serbia)
Prof. dr Duško Minić (Serbia)
Prof. dr Endre Romhanji (Serbia)
Prof. dr Fathi Habashi (Canada)
Prof. dr Guven Onal (Turkey)
Prof. dr György Kaptay (Hungary)
Prof. dr Heikki Jalkanen (Finland)
Prof. dr Iwao Katayama (Japan)
Prof. dr Jakob Lamut (Slovenia)
Prof. dr Jelena Penavin Škundrić (B&H)
Prof. dr Jožef Medved (Slovenia)
Prof. dr Karlo Raić (Serbia)
Prof. dr Kemal Delijić (Montenegro)
Prof. dr Krzystof Fitzner (Poland)
Prof. dr Luis Filipe Malheiros (Portugal)
Dr Magnus Ericsson (Sweden)
Prof. dr Milan Antonijević (Serbia)
Prof. dr Milan Trumić (Serbia)
Prof. dr Mile Dimitrijević (Serbia)
Prof. dr Mirjana Rajčić Vujasinović (Serbia)
Prof. dr Mirko Gojić (Croatia)
Dr Mile Bugarin (Serbia)
Dr Milenko Ljubojev (Serbia)
Dr Mirjam Jan-Blažić (Slovenia)
Dr Miroslav Sokić (Serbia)
Prof. dr Mirsada Oruč (B&H)
Dr Nadežda Talijan (Serbia)
Prof. dr Nenad Radović (Serbia)
Prof. dr Nenad Vušović (Serbia)
Prof. dr Nobuyuki Masuda (Japan)
Prof. dr Onuralp Yucel (Turkey)
Prof. dr Petr M. Solozhenkin (Russia)
Prof. dr Rodoljub Stanojlović (Serbia)
Prof. dr Sanda Krausz (Romania)
Prof. dr Seshadri Seetharaman (Sweden)
Dr Slavomir Hredzak (Slovakia)
Prof. dr Snežana Šerbula (Serbia)
Prof. dr Stoyan Groudev (Bulgaria)
Prof. dr Sulejman Muhamedagić (B&H)
Prof. dr Svetlana Ivanov (Serbia)
Dr Srećko Stopić (Germany)
Prof. dr Tamara Holjevac Grgurić (Croatia)
Prof. dr Tatjana Volkov-Husović (Serbia)
Prof. dr Tomaš Havlik (Slovakia)
Prof. dr Velizar Stanković (Serbia)
Prof. dr Velimir Radmilović (USA)
Prof. dr Vitomir Milić (Serbia)
Dr Vladan Ćosović (Serbia)
Prof. dr Vladimir Krstić (Canada)
Prof. dr Vladislav Kecojević (USA)
Prof. dr Vlastimir Trujić (Serbia)
Prof. dr Yong Du (China)
Prof. dr Zoran Marković (Serbia)
Prof. dr Žarko Radović (Montenegro)
Prof. dr Željko Kamberović (Serbia)
Prof. dr Živan Živković (Serbia)
Dr Walter Valery (Australia)
Dr Zvonko Gulišija (Serbia)

ORGANIZING COMMITTEE

Doc. dr Ivana Marković - president
Doc. dr Ljubiša Balanović - vice-president
Doc. dr Saša Stojadinović - vice-president
Prof. dr Svetlana Ivanov
Prof. dr Dragan Manasijević
Prof. dr Snežana Uroševič
Dr Ana Kostov (IRM Bor)
Doc. dr Vesna Grekulovič
Doc. dr Aleksandra Mitovski
Doc. dr Dejan Petrović
Doc. dr Milan Gorgievski
Doc. dr Ana Simonović
Doc.dr Tanja Kalinović
Doc.dr Marija Petrović Mihajlović
M.Sc. Uroš Stamenković
M.Sc. Oliver Marković
Slavica Stevanović, prof. engl.
Sandra Vasković, prof. engl.
Predrag Stolić, dipl. ing.
Dr Ana Radojević
M.Sc. Jelena Milosavljević

PREFACE

On behalf of the Organizing Committee, it is a great honor and pleasure to wish all the participants a warm welcome to the 49th International October Conference on Mining and Metallurgy (IOC 2017) held at Bor Lake, Serbia, 18 – 21 October 2017.

The IOC 2017 has been organized by the University of Belgrade, Technical Faculty in Bor, in cooperation with Mining and Metallurgy Institute Bor. It is devoted to presenting recent research results and advances in the fields of geology, mining, metallurgy, materials science, technology, environmental protection, and related engineering topics. The primary goal of IOC is to bring together academics, researchers, and industry engineers to exchange their experiences, expertise and ideas, and also to consider possibilities for collaborative research.

This year's conference is dedicated to the memory of Professor Dragana Zivkovic who was one of our most loyal and active Committee members. The 4th International Student Conference on Technical Sciences (ISC 2017) will take place within the frame of IOC 2017. ISC provides a unique opportunity for the students from both the country and the region to promote scientific research and discuss future directions of research with the experts and specialists.

These proceedings include 153 papers from authors coming from universities, research institutes and industries in 30 countries: Austria, Bosnia and Herzegovina, Bulgaria, China, Croatia, Czech Republic, France, Germany, Hungary, India, Iran, Italy, Japan, Jordan, Kazakhstan, Libya, Macedonia, México, Montenegro, Norway, Poland, Romania, Russia, Slovakia, Slovenia, South Africa, Spain, Turkey, USA and Serbia.

Financial assistance provided by the Ministry of Education, Science and Technological Development of the Republic of Serbia is gratefully acknowledged. The support of the sponsors and their willingness and ability to cooperate has been of great importance for the success of IOC 2017. The Organizing Committee would like to extend their appreciation and gratitude to all the sponsors and friends of the Conference for their donations and support.

We would like to thank all the authors who have contributed to these proceedings, and also to the members of the scientific and organizing committees, reviewers, speakers, chairpersons and all the Conference participants for their support to IOC 2017. Sincere thanks to all the people who have contributed to the successful organization of IOC 2017.

We look forward to welcoming you to the 50th International October Conference on Mining and Metallurgy (IOC 2018), which will be held in October 2018.

On behalf of the 49th IOC Organizing Committee,

Assistant Professor Ivana Marković, PhD

TABLE OF CONTENTS

Plenary Lectures

Section Lectures

Conference Papers

AIR MASS TRANSPORT OVER BALKANS REGION IDENTIFIED BY ATMOSPHERIC MODELING AND AEROSOL LIDAR TECHNIQUE

Zoran Mijić, Mirjana Perišić, Luka Ilić, Andreja Stojić, Maja Kuzmanoski

Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia

Abstract

This study combines atmospheric modeling with lidar measurements in order to assess the origin of aerosols traveling over Balkan region, having an impact on regional radiative budget and air quality. Particulate matter potential source regions and transport pathways were investigated using hybrid receptor modeling and mass concentrations measured in Belgrade, Serbia. In addition, the case study evidencing transport of Saharan dust particles simulated by the DREAM model was presented. The capability of the lidar technique to derive range-resolved vertical profiles of aerosol optical parameters was used to analyze the aerosol layers altitude and temporal evolution.

***Keywords**: atmospheric modeling, transport, PM*

1. INTRODUCTION

Suspended particulate matter (PM) in the atmosphere, commonly known as aerosol, plays an important role in the climate system. Besides significant effect on climate change, air quality and human health, aerosols affect long-range transport and deposition of toxics and nutrients. The complexity of aerosol processes in the atmosphere leads to large uncertainties in understanding of their role in many of the major environmental issues [1]. The direct (scattering and absorbing incoming solar radiation) and indirect aerosol effects (as they act as a cloud condensation nuclei) make the two largest contributions to the total uncertainty of radiative forcing. Regarding the impact of aerosols on air quality, the same processes that govern the global distribution, control the aerosol properties on regional and local scales. While *in situ* measurements are most adequate for air quality monitoring at the ground level, the assessment of impact of remote sources and transformation processes requires aerosol vertical distribution observations. Key parameters to be observed for this purpose are the presence, altitude and extent of elevated aerosol layers, the height of the planetary boundary layer (PBL), aerosol type, and mass concentration. Since long-range transport occurs at elevated layers, surface-based measurements of aerosol properties, such as chemical composition and size distribution are not sufficient. For global coverage including all relevant parameters, an integrated approach including ground-level and airborne in-situ measurements, ground-based remote sensing, and space-borne observations in combination with advanced modeling is necessary. Large observational networks such as the European Aerosol Research Lidar Network (EARLINET) [2], provide the long-term measurement series needed to build an aerosol vertical profile climatology at the continental scale. The capability of the lidar system (Light Detection And Ranging) to derive range-resolved aerosol vertical profiles with high spatial and temporal resolution is used to identify layers altitude and temporal evolution of intrusions. Using altitudes as inputs in air mass back-trajectories tracing method identification of aerosol sources at large distances from the measurement point, if their contribution is important, can be conducted. In this study hybrid receptor models for identification of potential source regions of PM affecting air quality in Belgrade are presented together with a case study evidencing transport of Saharan dust particles.

2. METHODOLOGY

Lidar technique is an active remote sensing method based on laser emission of the short-duration light pulses to the atmosphere and the analysis of the return signal. The intensity of the light backscattered by atmospheric molecules and particles is measured versus time – through the telescope receiver, collimating optics, a bandpass filter for daylight suppression – by an appropriate detector. For vertical profiling and remote sensing of atmospheric aerosol layers, Raman lidar at the Institute of Physics Belgrade (44.860 N, 20.390 E) has been used. It is bi-axial system with combined elastic and Raman detection designed to perform continuous measurements of aerosols in the PBL layer and the lower free troposphere. It is based on the third harmonic frequency of a compact, pulsed Nd:YAG laser, emitting pulses of 65 mJ output energy at 355 nm with a 20 Hz repetition rate. The optical receiver is a Cassegrain reflecting telescope with a primary mirror of 250 mm diameter and a focal length of 1250 mm. Photomultiplier tubes are used to detect elastic backscatter lidar signal at 355 nm and Raman signal at 387 nm. The detectors are operated both in the analog and photon-counting mode with lidar profiles averaging time of the order of 1 min and the spatial raw resolution of the detected signals of 7.5 m. Lidar measurements can be used in synergy with numerical models in order to validate and compare information about aerosols. In this paper DREAM (Dust Regional Atmospheric Model) model, designed to simulate and/or forecast the atmospheric cycle of mineral dust aerosol [3], is used to analyze dust transport. To estimate potential PM remote emission sources and their impact at the receptor site, concentration weighted trajectory (CWT) hybrid receptor model [4] was applied to the data set comprised of hourly PM_{10} mass concentrations obtained from Belgrade suburban location "Ovča" during the period 2012-14, and 72-h air masses back-trajectories, calculated according to Perišić et al. [5]. Furthermore, to obtain the vertical profile of PM, concentration weighted boundary layer (CWBL) hybrid receptor model [6], which uses a two-dimensional grid and a planetary boundary layer height, or any altitude in general, as a frame of reference, was used. Although the model can be applied for analyzing the pollutant concentration vertical distribution along the transport pathway, in this paper we present its usage for the receptor site solely.

3. RESULTS AND DISCUSSION

According to the CWT analysis, the most prominent PM_{10} sources were located in neighboring countries and in the areas NW, E and S of Belgrade. Significant impact of Central and Eastern European sources was registered during the autumn season (Figure 1–left panel).

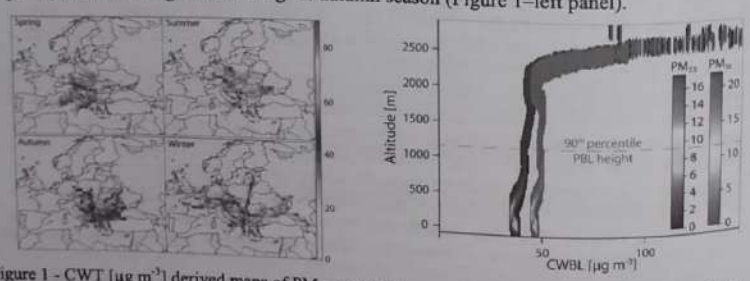


Figure 1 - CWT [μg m^{-3}] derived maps of PM_{10} potential sources in Europe – seasonal variations (left), and CWBL derived PM altitude distribution above the receptor site (right) – color scales indicate the number of events

Very similar, almost constant PM altitude distribution over the receptor site was observed for both coarse and fine particles (Figure 1–right panel), and the most common PBL heights (within 90th percentile). However, given the number of events (colored scale), it can be seen that concentrations exhibit decreasing trend to the height of about 400 m because the species emitted or generated near the ground level are mostly trapped and concentrated within the PBL, whereas free atmosphere concentrations remain low. Large CWBL values at higher altitudes correspond to rare PBL fluctuations which are not statistically significant, so the model results cannot be taken into consideration.

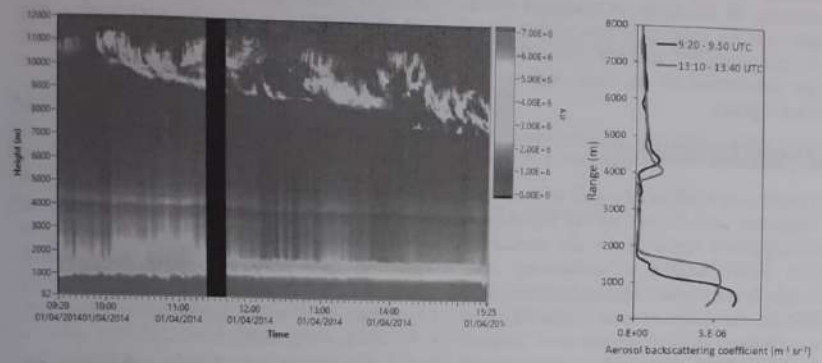


Figure 2 – Lidar range corrected signal (left) and backscatter coefficient at 355 nm (right) in Belgrade

Another aspect of aerosol climatology over Balkans region is related to the intrusions of Saharan dust which usually occurs during spring and summer periods. Such a case study evidencing transport of Saharan dust on 1st April 2014 is presented. From the RCS lidar time series (Figure 2), but also from the calculated backscatter coefficients profiles, the direct presence of an aerosol layer around 4-5 km altitude over Belgrade can been seen. This event was also successfully forecasted by DREAM model (Figure 3-left panel).

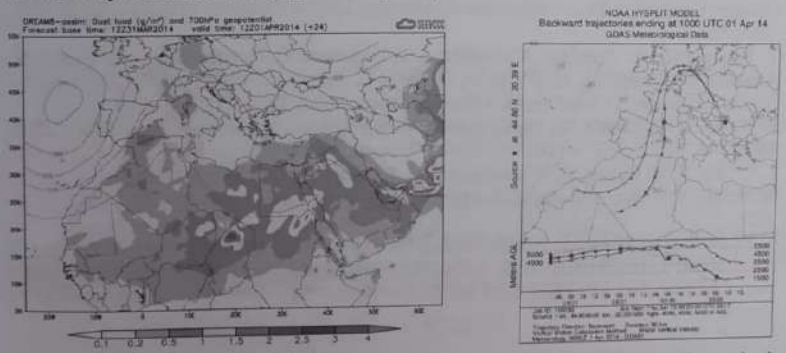


Figure 3 – Dust load over South Europe, estimated by the DREAM model (left) and air mass back-trajectories ending over Belgrade on 1st April, 2014 (righ)

Since the aerosols serve as a valuable tracer of air motion, using lidar observed altitudes of aerosol layer as inputs in the HYSPLIT [7] back-trajectory tracing method the source of aerosols was confirmed. As shown in Figure 3 (right panel) air masses reaching Belgrade traveled over

South Europe (Mediterranean Sea, Spain) and West Europe being influenced by continental pollution too.

4. CONCLUSION

The main advantage of lidar – real time observation of aerosol layering is that it can be used for air mass origin and path identification. Furthermore, in combination with statistical and numerical modeling, this technique can provide important information about aerosol type and distribution. In this paper we presented a case analysis of aerosol transport process affecting air quality over the Balkans region evidencing transport of Saharan dust particles over Serbia. Air mass back-trajectory analysis combined with hybrid receptor modeling were used to assess spatial distribution of the main regional sources for aerosols affecting air quality over the Balkans regions.

ACKNOWLEDGEMENTS

This paper was realized as a part of the projects III43007 and III41011 financed by the Ministry of Education and Science of the Republic of Serbia within the framework of integrated and interdisciplinary research for the period 2011-2017. The publication was supported by the project 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), Grant Agreement No. 690133, funded under European Union Horizon 2020 Programme - Topic: SC5-18b-2015, Integrating North African, Middle East and Balkan Earth Observation capacities in GEOSS. The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model and/or READY website (http://www.ready.noaa.gov) used in this publication.

REFERENCES

[1] T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley, IPCC: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[2] G. Pappalardo, A. Amodeo, A. Apituley, A. Comeron, V. Freudenthaler, H. Linné, A. Ansmann, J. Bösenberg, G. D'Amico, I. Mattis, L. Mona, U. Wandinger, V. Amiridis, L. Alados Arboledas, D. Nicolae, M. Wiegner, Atmos. Meas. Tech. 7 (2014) 2389.

[3] S. Nickovic, G. Kallos, A. Papadopoulos, O. Kakaliagou, J. Geophys. Res. 106 (2001) 1813.

[4] Y. K. Hsu, T. M. Holsen, P. K. Hopke, Atmos. Environ. 37(4) (2003) 545-562.

[5] M. Perišić, S. Rajšić, A. Šoštarić, Z. Mijić, A. Stojić, Air Qual Atmos Health 10 (2017) 93-103.

[6] A. Stojić, S. Stanišić Stojić., Atmos. Environ. 164 (2017) 216-223.

[7] A. F. Stein, R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, F. Ngan, Bull. Amer. Meteor. Soc. 96 (2015) 2059-2077.

ISBN 978-86-6305-066-2

WeBIOPATR2017

Particulate Matter: Research and Management

Proceedings from the
6th WeBIOPATR
Workshop & Conference
Belgrade, Serbia
6.-8.9.2017

Milena Jovašević-Stojanović
and Alena Bartoňová, eds.

Belgrade 2019

WeBIOPATR

The 6th WeBIOPATR Workshop and Conference,
Particulate Matter: Research and Management,
WEBIOPATR2017 is organized by:

Vinča Institute of Nuclear Sciences, Serbia
Public Health Institute of Belgrade, Serbia
NILU Norwegian Institute for Air Research, Norway

The 6th WeBIOPATR Workshop and Conference,
Particulate Matter: Research and Management,
WeBIOPATR2017 is supported by:

Ministry of Education, Science and Technological Development of

Republic of Serbia

PROCEEDINGS

The Sixth International WeBIOPATR Workshop & Conference
Particulate Matter: Research and Management
WeBIOPATR2017

6 - 8 September 2017
Belgrade, Serbia

Editors
Milena Jovašević-Stojanović
Alena Bartoňová

Publisher
Vinča Institute of Nuclear Sciences
Dr Zlatko Rakočević, Director
P.O. Box 522
11001 Belgrade, Serbia

Printed by
Vinča Institute of Nuclear Sciences

Number of copies
150

ISBN: 978-86-7306-152-8

Vinča Institute of Nuclear Sciences
www.vin.bg.ac.rs

SCIENTIFIC COMMITTEE

Dr Alena Bartoňová, Norway

Dr Bojan Radak, Serbia

Prof. Dr David Broday, Israel

Dr Med Elizabeta Paunović, Germany

Dr Maria Cruiz Min, Spain

Dr Milena Jovašević-Stojanović, Serbia

Prof. Dr Nenad Živković, Serbia

Prof. Dr Radim Šrám, Czech Republic

Dr Renata Kovačević, Serbia

Dr Slobodan Nićković, Serbia

Prof. Dr Simone Barreira Morais, Portugal

Zoran Mijić, Serbia

Prof. Dr Zoran Ristovski, Australia

Dr Zorana Jovanović-Andersen, Denmark

ORGANIZING COMMITTEE

Vinča Institute of Nuclear Sciences, Belgrade: Serbia
Dr Dragan Alavantić, Serbia
MS Ivan Lazović, Serbia
MS Maja Jovanović, Serbia (Secretary)
Dr Milena Jovašević-Stojanović (Co-chair)
Dr Miloš Davidović, Serbia (Secretary)
Dr Snežana Pašalić, Serbia

NILU - Norwegian Institute for Air Research, Kjeller
Dr Alena Bartoňová, Norway (Co-chair)

Public Health Institute of Belgrade, Belgrade
Dr Anka Cvetković, Serbia
MS Andrej Šoštarić, Serbia
Vesna Slapčević, Serbia

Ministry of Environmental Protection of RS
Ms Biljana Filipović, Serbia

Serbian Environmental Protection Agency
Mr Dejan Lekić, Serbia
Mr Tihomir Popović, Serbia

National Institute of Public Health "Dr Milan Jovanović-Batut", Belgrade
Dr Med Branislava Matić, Serbia

Military Medical Academy, Belgrade
Prof. Dr Jasmina Jović-Stošić, Serbia

Institute of Physics, Belgrade
Dr Mira Aničić Urošević, Serbia

Faculty of Occupational Protection, University of Niš
Prof. Dr Nenad Živković, Serbia

Medical Faculty, University of Niš
Prof. Dr Aleksandra Stanković, Serbia

Institute of Metallurgy and Mining, Bor
Dr Viša Tasić, Serbia

Mechanical Faculty, University of Belgrade
Prof. Dr Aleksandar Jovović, Serbia

CONFERENCE TOPICS

ATMOSPHERIC PARTICULATE MATTER - PHYSICAL AND CHEMICAL PROPERTIES	• *sources and formation of particulate matter* • *particulate matter composition and levels outdoors and indoors* • *environmental modeling* • *nanoparticles in the environment*
PARTICULATE MATTER AND HEALTH	• *exposure to particulate matter* • *health aspects of atmospheric particulate matter* • *full chain approach*
PARTICULATE MATTER AND REGULATORY ISSUES	• *issues related to monitoring of particulate matter* • *legislative aspects* • *abatement strategies*

PREFACE

The International Workshop and Conference, Particulate Matter: Research and Management – WeBIOPATR is a biennial event held in Serbia since 2007. The conference addresses air quality in general and particulate matter specifically. Atmospheric particulate matter arises both from primary emissions and from secondary formation in the atmosphere. It is one of the least well-understood local and regional air pollutants, has complex implications for climate change, and is perhaps the pollutant with the highest health relevance. It also poses many challenges to monitoring.

By WeBIOPATR, we aim to link the research communities with relevance to particulate matter with the practitioners of air quality management on all administrative levels, in order to facilitate professional dialogue and uptake of newest research into practice. The workshops usually draw an audience of about 70, and attract media attention in Serbia. It enjoys support of the responsible authorities, Ministry of Health, Ministry of Environment, and the Serbian Environmental Agency whose sponsorship is indispensable and gratefully acknowledged. We enjoy also support of international bodies such as the WHO.

The 1st WeBIOPATR Workshop was held in Beograd, 20.-22. May 2007, associated with a project funded by the Research Council of Norway. The 2nd workshop was held in Mecavnik, Serbia, 28.8.-1.9. 2009. WeBIOPATR2011 was held in Beograd 14.-17. 11. 2011 and for the first time, included a dedicated student workshop. WeBIOPATR2013 was held in Beograd 2.-4. 10. 2013. It covered the traditional PM research and management issues, discussions on how to encourage citizens to contribute to environmental governance, and how to develop participatory sensing methods. WeBIOPATR2015 was held in Beograd 14.-16.10. 2015. Own sessions were devoted to sensor technologies for air quality monitoring, utilizing information and input from the EU FP7 funded project CITI-SENSE (http://co.citi-sense.eu) and the EU COST action EuNetAir (www.eunetair.it).

We have now the pleasure to present to you the proceedings of the 6th conference held in Beograd 6.-8.9. 2017. We are excited to have contributions from old friends and new acquaintances, and we are especially pleased with a wider than before Western Balkan participation. The contributions were reviewed. The language editing was performed by Dr Simon Smith, PhD, to whom we would like to extend out sincere thanks. Technical manuscript preparation was graciously done by Dr Milos Davidovic, PhD, to whom we are very grateful.

We are hoping that you, the reader, will extend your support to WeBIOPATR also in the future. The issues of atmospheric pollution, with their wide implications for climate change, human health and ecosystem services, are no less important today than before. Addressing them requires a strong scientific community and commitment of all societal actors. Your contribution will make a difference.

Milena Jovašević-Stojanović and Alena Bartoňová

CONTENTS

5.4. RELATIVE IMPORTANCE OF GASEOUS POLLUTANTS AND AEROSOL CONSTITUENTS FOR IDENTIFICATION OF PM_{10} SOURCES OF VARIABILITY

M. Perišić (1), G. Vuković (1), Z. Mijić (1), A. Šoštarić (2) and A. Stojić (1)
(1) Institute of Physics, University of Belgrade, Pregrevica 118, Serbia
(2) Institute of Public Health of Belgrade, Boulevard of Despot Stefan 54a, Belgrade, Serbia
mirjana.perisic@ipb.ac.rs

ABSTRACT

This study combines advanced statistical methods including time series decomposition, source apportionment and supervised learning algorithms, to identify the main sources of particulate matter (PM_{10}) variability in an urban area within Belgrade. The analyses indicated that the season, (i.e., meteorological conditions) strongly influenced daily and annual PM_{10} variations particularly during the colder part of the year. A guided regularized random forest model estimated that As, Cd, BaP, CO, and benzene have the highest relative importance for the prediction of PM_{10}. Polar plot source apportionment revealed common sources of pollution at specific directions. Specifically, emissions of PM_{10}, CO and benzene could be attributed to heating and gasification processes, while processes in oil refineries and chemical industries produced PM_{10} and toluene.

INTRODUCTION

Due to adverse effects on human health and the increased risk of morbidity and mortality, particulate matter (PM) is one of the most studied atmospheric pollutants, and perhaps, the most pressing issue in worldwide air quality regulation (Fuzzi et al, 2015, Stanišić Stojić et al, 2016). Even though significant progress has been made through the integration of different scientific approaches, modelling of air pollution data remains a challenge due to the complexity and non-linear nature of atmospheric phenomena and processes (Pai et al, 2013). During the last decade, poor air quality in Belgrade, with many PM_{10} limit value exceedances (Directive 2008/50/EC), has been identified as an important environmental risk factor (Perišić et al, 2015, 2017). Identification of factors affecting PM_{10} concentration variability could provide better insight into the aerosol spatiotemporal distribution and source composition, revealing their dominant sources in an urban area (Stojić et al, 2016).

Apart from the commonly used methods for data analysis, this study adopts the advanced statistical classifier, guided regularized random forest (GRRF), widely applied in many fields for feature selection. Moreover, the study demonstrates the possibilities of source apportionment analysis, which combines correlation and regression statistics with the bivariate polar plot analysis, to offer considerably more insight into air pollution sources.

METHODOLOGY

The analysed dataset, comprised of daily PM_{10} and its constituent concentrations (As, Cd, Cr, Mn, Ni, Pb, BaP, Cl^-, NO_3^-, NH_4^+, SO_4^{2-}, Na^+, K^+, Mg^{2+} and Ca^{2+}), and hourly PM_{10} and gaseous pollutant concentrations (CO, SO_2, NO, NO_2, NO_x, benzene, toluene, o- and m, p xylene) have been obtained from an Institute of Public Health regular monitoring station located within an urban area in Belgrade (Longitude 20.470, Latitude 44.817) from 2011 - 2016. The time series of PM_{10} concentrations was resolved into the additive components of the multi-year and seasonal trends, as well as the remainders using the Loess smoothing decomposition model (LSD) (Li et al, 2014). Daily, weekly and seasonal periodicity was analyzed by the use of Lomb-Scargle periodogram (*Lomb* package within the statistical software environment *R*) (Ruf, 1999; Team, 2014). Bivariate polar plot analysis was used for identification of the main PM_{10} emission sources (Carslaw and Ropkins, 2012), while the advanced bivariate polar plots, coupled with pair-wise statistics, were applied to distinguish specific sources and to gain information about pollutant relationships. The model includes a weighted Pearson correlation, linear regression slope and Gaussian kernel to locally weight the statistical calculations on a wind speed-direction surface together with variable-scaling (Grange et al, 2016). Feature selection was implemented using a GRRF ensemble learning method (Deng and Runger, 2013). GRRF can select compact feature subsets revealing higher order variable interactions, thus moderating the problem of dimensionality and avoiding the effort to analyze irrelevant or redundant features.

RESULTS AND DISCUSSION

Annual concentrations of PM_{10} and BaP exceeded prescribed limit values of 50 $\mu g\ m^{-3}$ and 1 $ng\ m^{-3}$, respectively (Directive 2000/69/EC, Directive 2008/50/EC) every year of the period examined. The most abundant aerosol constituents were Cr, Pb and Mn (Figure 1), while SO_4^{2-} and NO_3^- were the ions with the highest concentrations.

In an urban area, the dominance of sulfate and nitrate ions is related to fossil fuel burning and traffic exhaust emission of SO_2 and NO_x, which, in the presence of water, transform into these ions. In addition, NH_4^+ and Ca^{2+} cations are usually presented as neutralizing agents for SO_4^{2-} and NO_3^- in heterogeneous atmospheric chemical reactions.

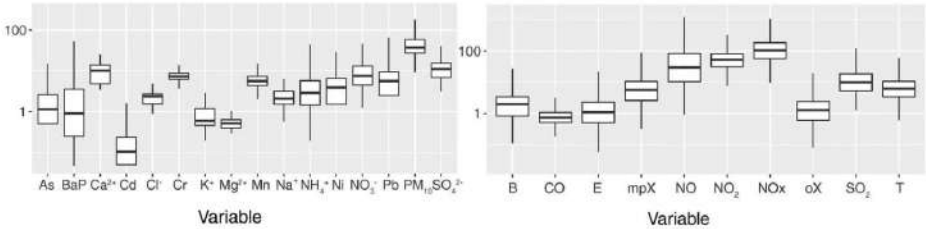


Figure 1. *PM_{10} concentration [$\mu g\ m^{-3}$], its chemical constituent (ions and BaP [$\mu g\ m^{-3}$], metals [$ng\ m^{-3}$]) (left) and gaseous pollutant [$\mu g\ m^{-3}$] (right) whisker plots*

Spectral analysis (Figure 2, left) reveals the highest normalized power values are attributed to the periods of 12 and 24 h, 7 days, and 1 and 3 months. This implies that meteorological conditions and anthropogenic emissions are strongly affected by aerosol daily and seasonal variations, and weekly periodicity, respectively (Bigi, 2016).

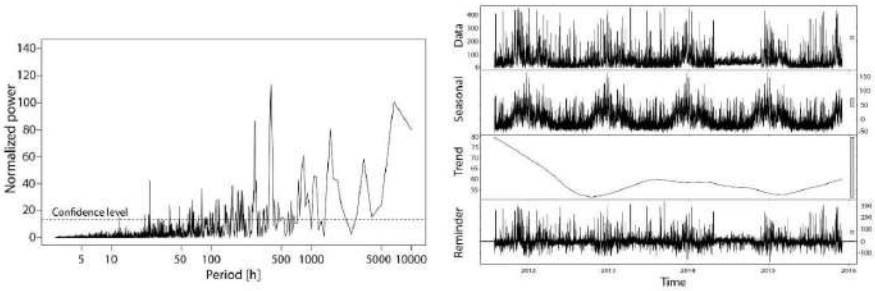


Figure 2. *PM_{10} Lomb-Scargle periodogram (left) and PM_{10} time series decomposition [$\mu g\ m^{-3}$] (right)*

Decomposed PM_{10} time series indicates a decreasing multi-year trend and significant impact of the seasonal component. Large variance of the remainder component possibly occurs as a result of short-term air pollution episodes (Figure 2, right). The conventional bivariate polar plot approach reveals the pronounced influence of both local and remote sources on PM_{10} variability (Figure 3).

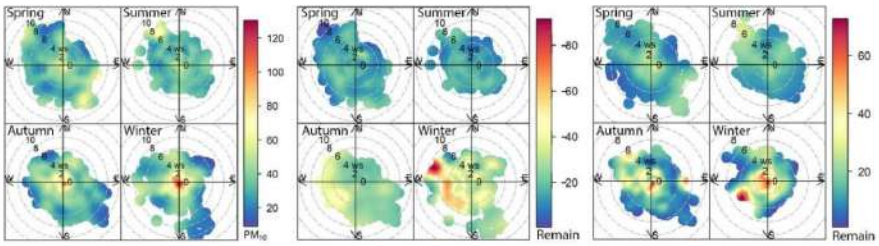


Figure 3. *Bivariate polar plot of PM_{10} concentrations (left) and its remainder components: negative (middle) and positive (right) [$\mu g\ m^{-3}$]*

Bivariate polar plot analysis of the remainder component, separately applied on positive and negative values, confirms that the episodes of the highest variations mainly occur during the colder part of the year. Positive variations related to SW and negative related to NW winds with speeds greater than 6 m s^{-1}.

The highest Pearson's correlation coefficients were obtained between concentrations of PM_{10} and its constituents (BaP (0.83), As (0.81), Cd (0.79) and Pb (0.66)), as well as for the gaseous pollutants: CO (0.56), benzene (0.46), NO (0.35), and NO_x (0.35). Similarly, the GRRF estimated the highest relative importance of As, Cd, BaP, CO and benzene for the prediction of PM_{10}, indicating that the environmental burden is mainly associated with fossil fuel combustion, particularly pronounced during the colder part of the year. An inconsistency between the correlation and GRRF analysis was observed for toluene. This compound had a higher importance for PM_{10} prediction than NO_x, but its correlation coefficient was among the lowest (0.25).

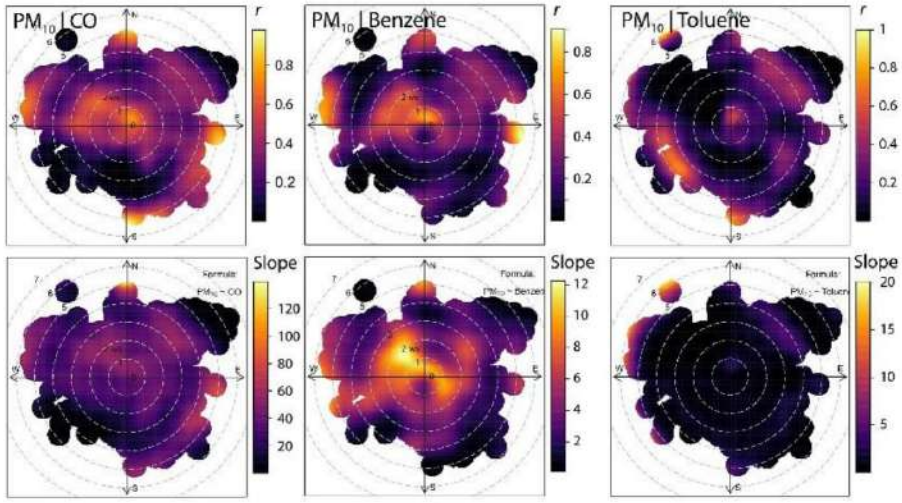


Figure 4. *Bivariate polar source apportionment*

Even though Pearson's coefficients did not indicate a significant correlation between PM_{10} and gaseous pollutants ($r < 0.6$), the bivariate polar source apportionment (Figure 4) showed that during episodes of north-westerly winds, concentrations of PM_{10} and benzene and CO were more correlated ($r \approx 0.7$) probably because of several common sources in the vicinity of the sampling site. Source composition obtained from slope diagrams reveals a 1:0.1 and 1:12 contribution of PM_{10}, CO and benzene, respectively. This could be associated with various biomass combustion processes (traffic activities, heating plants and individual heating units) (Yokelson et al, 2007). Besides the vicinity of the sampling site, particulate matter and toluene shared prominent sources located in the SW, S, NE and SE directions ($r > 0.8$, wind speed > 4 m s-1). Unlike southern and western sources, characterized by PM_{10} to toluene ratio of 1:1 which could be related to mineral oil and gas refineries, the source located on the north-east is characterized by the ratio of 1:6 indicating influences from the chemical industry, and chemical installations for production, on an industrial scale, of basic organic chemicals including aromatic hydrocarbons (European Commission, 2006).

CONCLUSIONS

Due to the pronounced nonlinearity and complexity of atmospheric processes in the troposphere of an urban environment, the application of multivariate and nonlinear methods is required to gain reliable information for a better understanding of the underlying factors which determine the air pollution phenomena. Methods such as feature selection based on advanced supervised learning algorithms, advanced source apportionment techniques and time series decomposition and detailed component analysis, are capable of providing this information, particularly for characterization of variable pollution sources. Summarizing this study, it has been shown that

locally emitted and transported pollution, as well as meteorological factors, have the highest impact on urban air quality.

ACKNOWLEDGMENTS

This paper was completed as part of the project titled "Studying climate change and its influence on the environment: impacts, adaptation and mitigation" (III43007) financed by the Ministry of Education and Science of the Republic of Serbia within the framework of integrated and interdisciplinary research for the period 2011-2017. The publication was supported by the project 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), Grant Agreement No. 690133, funded under European Union Horizon 2020 Programme - Topic: SC5-18b-2015, Integrating North African, Middle East and Balkan Earth Observation capacities in GEOSS.

REFERENCES

1. Bigi, A. and Ghermandi, G. 2016. Trends and variability of atmospheric $PM_{2.5}$ and $PM_{10-2.5}$ concentration in the Po Valley, Italy. Atmospheric Chemistry and Physics 16, 15777-15788.
2. Carslaw, D.C., Ropkins, K. 2012. Openair - an R package for air quality data analysis. Environmental Modelling and Software 27-28, 52-61.
3. Deng, H. and Runger, G. 2013. Gene selection with guided regularized random forest. Pattern Recognition 46, 3483-3489.
4. Directive 2000/69/EC of the European Parliament and of the council of 16 November 2000 relating to limit values for benzene and carbon monoxide in ambient air. Official Journal of the European Communities L313, 12-21 (13/12/2000).
5. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal of the European Union L152/3, 6-15 (11/06/2008).Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier Van Der Gon, H., Facchini, M.C., Fowler, D., Koren, I., Langford, B., Lohmann, U. and Nemitz, E. 2015. Particulate matter, air quality and climate: lessons learned and future needs. Atmospheric Chemistry and Physics 15, 8217-8299.
6. European Commission, 2006. Guidance Document for the implementation of the European Pollutant Release and Transfer Register (E-PRTR).
7. Grange S.K., Lewis A. and Carslaw D. 2016. Source apportionment advances using polar plots of bivariate correlation and regression statistics. Atmospheric Environment 145, 128-134.
8. 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.
9. Ruf, T. 1999. The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. Biological Rhythm Research 30, 178-201.
10. Pai, T.Y., Hanaki, K., Chiou, R.J. 2013. Forecasting hourly roadside particulate matter in Taipei County of Taiwan based on firstorder and one-variable grey model. CLEAN Soil Air Water 41, 737-742.
11. 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 and Health 10, 93-103.
12. Perišić, M., Stojić, A., Stojić, S. S., Šoštarić, A., Mijić, Z., and Rajšić, S. (2015). Estimation of required PM_{10} emission source reduction on the basis of a 10-year period data. *Air Quality, Atmosphere & Health*, *8*(4), 379-389.
13. Stanišić Stojić, S., Stanišić, N., Stojić, A., and Šoštarić, A. 2016a. 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, 17-27.
14. 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.
15. R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
16. Yokelson, R. J., Urbanski, S. P., Atlas, E. L., et. al, 2007. Emissions from forest fires near Mexico City. Atmospheric Chemistry and Physics 7, 5569-5584.

6.1. MULTISCALE MULTIFRACTAL ANALYSIS OF NONLINEARITY IN PARTICULATE MATTER TIME SERIES

A. Stojić (1), S. Stanišić Stojić (2), M. Perišić (1), Z. Mijić (1)

(1) Institute of Physics, University of Belgrade, Belgrade, Serbia, (2) Faculty of Physical Chemistry, University of Belgrade, Belgrade, Serbia

andreja.stojic@ipb.ac.rs

ABSTRACT

In this study the multiscale multifractal method was used with aim of capturing the fractal behaviour of the particulate matter time series obtained from an urban area in Belgrade, Serbia, as well as investigating their persistence properties and heterogeneity features. As shown, the $PM_{2.5}$ time series exhibited persistency, slightly affected by the concentrations occurring randomly only at the level of small fluctuations and small scales. Compared to $PM_{2.5}$, PM_{10} concentrations were shown to display more stochastic behaviour with more frequent random fluctuations being observed at small scales. The results herein presented contribute to the current understanding of the structural complexity of the temporal evolution of particulate matter and provide a theoretical background for enhanced air pollution modelling.

INTRODUCTION

Comprehensive analyses, conducted over the past few years, of air pollutant emission sources, their subsequent distribution and relationship to mortality caused by circulatory, respiratory and malignant diseases suggest that the exposure to particulate matter (PM) has detrimental effects on human health in the Belgrade area (Stanišić Stojić et al, 2016a, 2016b). Besides the fact that PM levels in Serbia are higher than in most European cities, with a significant number of air quality standard exceedances, our studies have shown that suspended particles also contain high concentrations of carcinogenic contaminants, such as arsenic and benzo(a)pyrene (Stojić et al, 2015a, 2015b, 2016, Perišić et al, 2015, 2017).

Diverse methods have been implemented to provide relevant information for efficient air quality management, including deterministic models, statistical analysis, neural networks, fuzzy models, geographic information system, remote sensing and trend analysis (Yu et al, 2011). Multifractality is one of the inherent properties that can be recognized in physical, chemical, biological, social and other systems, that are described as very complex at different spatial and temporal scale levels (Glushkov et al, 2014). The atmosphere is a complex system that exhibits nonlinear behavior involving both deterministic and stochastic components (Lorenz and Haman, 1996). In previous studies, the multifractal approach has been applied to analyse average ozone concentrations (Kocak et al, 2000), nonlinearity in NO_2 and CO time series (Kumar et al, 2008) and the daily air pollution index (Sivakumar et al, 2007). The aim was to provide information essential to better understand the behaviour of pollution and to forecast the temporal evolution of the species (Dong et al, 2017). The multifractal method was used herein to reveal PM fluctuation properties, *i.e.* to investigate to what extent, and on which time scale, changes in $PM_{2.5}$ and PM_{10} concentration levels can be considered random or persistent.

METHODOLOGY

In this study, multiscale multifractal analysis (MMA) was used to investigate the presence of fractal behaviour in the complex time series of $PM_{2.5}$ and PM_{10} concentrations. Data was obtained during a period of almost three years (2012-2014) of regular pollutant monitoring in Belgrade (suburban site Ovča, Longitude 20.528, Latitude 44.884, Serbia) provided by the Institute of Public Health Belgrade. MMA is a generalization of the standard multifractal detrended fluctuation analysis (MF-DFA), which adds the dependence on scale, providing a broader analysis of the fluctuation properties, as well as, more general and stable results (Gierałtowski et al, 2012).

RESULTS AND DISCUSSION

Measured PM concentrations are presented in Figure 1. According to the results, multiscale multifractal derived Hurst surfaces confirmed the non-linear behavior of PM time series (Figure 2).

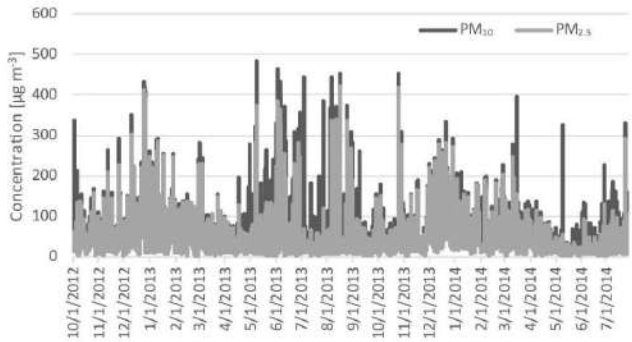


***Figure 1**. Measured $PM_{2.5}$ and PM_{10} concentrations.*

For most of the scale and multifractal parameter values, the local Hurst exponent remains in the interval between 1 and 1.5 indicating persistency of the $PM_{2.5}$ time series, while slightly affected by the concentrations occurring randomly. Such random concentration values occur only at the level of small fluctuations for scales below 44, which corresponds to a period of about 2 days. At this scale, there emerges a clear crossover resulting from the different correlation properties. Given that the sampling site was not directly exposed to intense PM bursts, the occurrence of concentrations in narrow bands (Hurst exponent equals 2) was not recorded. The PM_{10} Hurst surface reveals similar features, except that in the area of small variance and scales below 90, its growth to a maximum of approximately 1.9 is steeper, almost reaching black noise area values of local Hurst exponent. Compared to $PM_{2.5}$, the PM_{10} Hurst structure around its maximum corresponds to visibly more pronounced peaks in the time series (Figure 1). However, unlike $PM_{2.5}$, the PM_{10} Hurst surface shows no crossover.

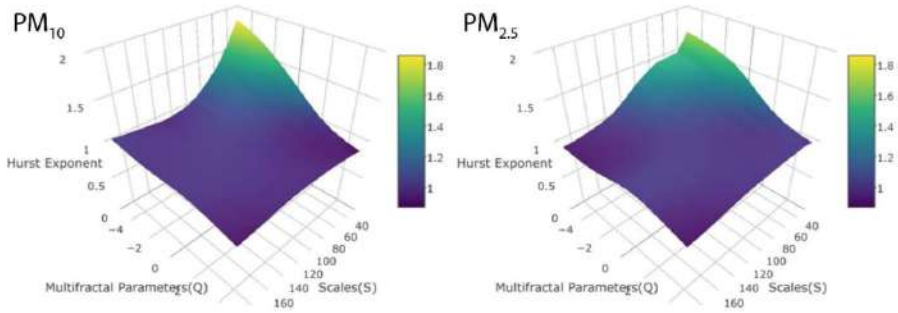


***Figure 2**. MMA derived particulate matter Hurst surfaces.*

In addition, the generalized distance coefficient (0.069) between Hurst surfaces of PM fractions is higher than the threshold value (0.065) and implies that the $PM_{2.5}$ and PM_{10} time series must be considered statistically different. The difference is particularly pronounced in the area of small fluctuations and medium scales (Figure 3).

Furthermore, it is shown that the source of multifractality, examined by PM time series randomization, originates from both nonlinear correlations and a fat-tailed probability distribution (Figure 4).

The findings of Lalwani (2016) and Liu et al. (2015) confirmed the existence of multifractality in the PM time series and found that daily pollutant concentrations exhibited high persistence in a period of approximately one

year. As argued, the persistence in the air pollutant concentrations over longer period of time may be governed by the impact of background levels, seasonal trend or intrinsic evolution of the system.

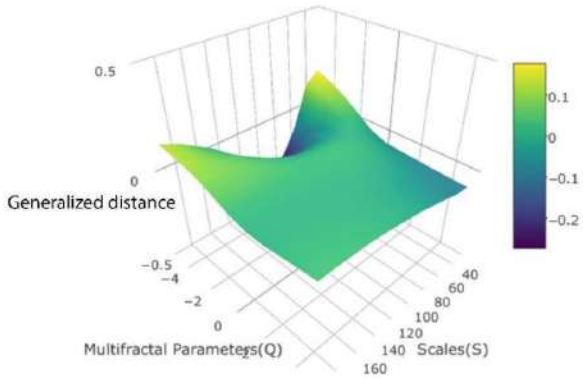


***Figure 3**. Generalized $PM_{10}/PM_{2.5}$ Hurst surface distance.*

The difference in the behavior of the $PM_{2.5}$ and PM_{10} time series was proven by Xue et al. (2015), who employed a multifractal analysis to explore temporal fluctuations and self-similarities within the PM time series and to understand their behaviour associated with diffusion, spreading and coagulation processes. Using the multifractal detrended fluctuation analysis method, the researchers registered the pronounced multifractality and long-term persistence of the $PM_{2.5}$ time-series, whereas the PM_{10} time series were shown to have stochastic behaviour.

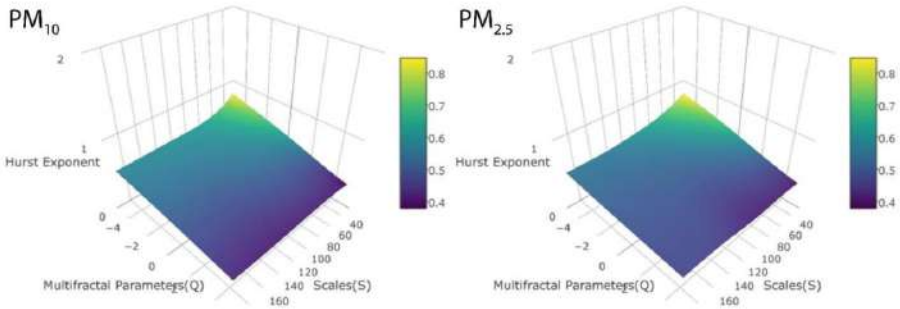


***Figure 4**. MMA derived Hurst surfaces for randomized PM time series.*

CONCLUSIONS

In this study, the multifractal approach was used to analyse the temporal dynamics of $PM_{2.5}$ and PM_{10} concentrations on the basis of a regular monitoring of data over a three-year period. As shown, the particulate matter time series possess a long-term memory of distant past events and require a large number of exponents, the so-called fractal dimensions, to be described. The presented analysis provides essential information for better understanding of the PM behaviour and the underlying factors, as well as for more accurate and reliable pollutant forecasting and efficient mitigation policy.

ACKNOWLEDGEMENTS

This study was performed as part of the projects Grant No III43007 and No III41011, which were supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia within the framework of integrated and interdisciplinary research for the period 2011-2017. The publication was supported by the project 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), Grant Agreement No. 690133, funded under European Union Horizon 2020 Programme - Topic: SC5-18b-2015, Integrating North African, Middle East and Balkan Earth Observation capacities in GEOSS.

REFERENCES

1. Dong, Q., Wang, Y. and Li, P. 2017. Multifractal behaviour of an air pollutant time series and the relevance to the predictability, Environmental Pollution 222, 444-457.
2. Gieraltowski, J., Żebrowski, J. J. and Baranowski, R. 2012. Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia, Physical Review E 85(2), 021915.
3. Glushkov, A. V., Svinarenko, A. A., Buyadzhi, V. V., Zaichko, P. A. and Ternovsky, V. B. 2014. Chaosgeometric attractor and quantum neural networks approach to simulation chaotic evolutionary dynamics during perception process, Advances in Neural Networks, Fuzzy Systems and Artificial Intelligence, Series: Recent Advances in Computer Engineering, Ed. J. Balicki. Gdansk, WSEAS Pub.
4. Koçak, K., Şaylan, L. and Şen, O. 2000. Nonlinear time series prediction of O_3 concentration in Istanbul, Atmospheric Environment 34(8), 1267-1271.
5. Kumar, U., Prakash, A. and Jain, V. K. 2008. Characterization of chaos in air pollutants: A Volterra-Wiener-Korenberg series and numerical titration approach, Atmospheric Environment 42(7), 1537-1551.
6. Lalwani, A. 2016. Long-Range Correlations in Air Quality Time Series: Effect of Differencing and Shuffling, Aerosol and Air Quality Research 16(9), 2302-2313.
7. Liu, Z., Wang, L. and Zhu, H. 2015. A time-scaling property of air pollution indices: a case study of Shanghai, China, Atmospheric Pollution Research 6(5), 886-892.
8. Lorenz, E. N. and Haman, K. 1996. The essence of chaos, Pure and Applied Geophysics 147(3), 598-599.
9. Sivakumar, B., Wallender, W. W., Horwath, W. R. and Mitchell, J. P. 2007. Nonlinear deterministic analysis of air pollution dynamics in a rural and agricultural setting, Advances in Complex Systems 10(04), 581-597.
10. Stojić, A., Stojić, S. S., Šoštarić, A., Ilić L., Mijić Z. and Rajšić S. 2015a. Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling, Environmental Science and Pollution Research 1-16.
11. Stojić, A., Stojić, S. S., Mijić, Z., Šoštarić, A. and Rajšić S. 2015b. Spatio-temporal distribution of VOC emissions in urban area based on receptor modelling, Atmospheric Environment 106, 71-79.
12. 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(11), 10722-10732.
13. Stojić, S. S., Stanišić, N., Stojić, A., and Šoštarić, A. 2016a. 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.
14. Stojić, S. S., Stanišić, N., and Stojić, A. 2016b. Temperature-related mortality estimates after accounting for the cumulative effects of air pollution in an urban area, Environmental Health 15(1), 73.
15. Perišić, M., Stojić, A., Stojić, S. S., Šoštarić, A., Mijić, Z. and Rajšić, S. 2015. Estimation of required PM_{10} emission source reduction on the basis of a 10-year period data, Air Quality, Atmosphere & Health 8(4), 379-389.
16. 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(1), 93-103.
17. Xue, Y., Pan, W., Lu, W. Z. and He, H. D. 2015. Multifractal nature of particulate matters (PMs) in Hong Kong urban air, Science of the Total Environment 532, 744-751.
18. Yu, B., Huang, C., Liu, Z., Wang, H. and Wang, L. 2011. A chaotic analysis on air pollution index change over past 10 years in Lanzhou, Northwest China, Stochastic Environmental Research and Risk Assessment 25(5), 643-653.