

ИНСТИТУТ ЗА ФИЗИКУ

ПРИМЉЕНО: 08. 07. 2021			
Рад.јед.	б р о ј	Арх.шифра	Прилог
0801	501/1		

Научном већу Института за физику

Београд, 8.7.2021.

Предмет: Покретање поступка у звање истраживач сарадник

Молим научно веће Института за физику у Београду да покрене поступак за мој избор у звање **Истраживач сарадник**, имајући у виду да испуњавам све критеријуме прописане од стране Министарства просвете, науке и технолошког развоја Републике Србије за стицање наведеног звања.

У прилогу достављам:

1. мишљење руководиоца пројекта са предлогом комисије за избор у звање;
2. стручну биографију;
3. преглед научне активности;
4. списак и копије објављених научних радова и других публикација;
5. уверење о последњем овереном и уписаном семестру на докторским студијама;
6. фотокопију диплома са основних и мастер студија;
7. потврду о прихватању теме докторске дисертације.

С'поштовањем,



Душан Жигић
Истраживач приправник

8. 7. 2021.

Научном већу Института за физику

Предмет: Мишљење руководиоца лабораторије за избор Душана Жигића у звање истраживач сарадник

Поштовани,

Душан Жигић завршио је основне, а потом и мастер студије на Физичком факултету Универзитета у Београду. За мастер тезу одбраћену септембра 2018. године добио је и награду "Љубомир Ћирковић" 2019. године. Докторске студије на смеру Физика високих енергија и нуклеарна физика на Физичком факултету Универзитета у Београду уписао је 2018. године. Положио је све предвиђене испите. Тема његове докторске дисертације представљена је и прихваћена на Колегијуму докторских студија Физичког факултета Универзитета у Београду 30. јуна 2021. године, Ангажован је у групи др Магдалене Ђорђевић и бави се истраживањем кварк-глуонске плазме.

С обзиром да Душан Жигић испуњава све услове предвиђене Правилником о поступку и начину вредновања, и квантитативном исказивању научноистраживачких резултата истраживача, сагласна сам са покретањем поступка за избор у звање истраживач сарадник.

За чланове комисије за избор Душана Жигића у звање истраживач сарадник предлажем следећи састав:

1. др Магдалена Ђорђевић, научни саветник, Институт за физику;
2. др Игор Салом, виши научни сарадник, Институт за физику;
3. проф. др Воја Радовановић, редовни професор Физичког факултета;

Руководилац лабораторије за
физику високих енергија,



др Лидија Живковић
Научни саветник

Биографија Душана Жигића

Душан Жигић је рођен 19.10.1991. године у Сремској Митровици где је завршио основну школу и Гимназију. Основне студије уписао је 2012. на Физичком факултету, смер Примењена и компјутерска физика и дипломирао је 2017. године са просечном оценом 9,75. Мастер студије, уписане 2017. на Физичком факултету, завршава 2018. године са просечном оценом 9,67. Мастер тезу под називом *"Предвиђања пригушења и елиптичког шока високо-енерџијских честица при релативистичким сударима $Pb+Pb$ на LHC -у"* ради под менторством др Магдалене Ђорђевић у оквиру Лабораторије за физику високих енергија Института за физику у Београду. Септембра 2019. године Физички факултет Универзитета у Београду, доделио му је награду "проф. др Љубомир Ђирковић" за најбољи мастер рад одбрањен у претходној години.

2018. године на Физичком факултету Универзитета у Београду уписује докторске студије у области истраживања релативистичких судара тешких јона и кварк-глуонске плазме. Истраживање ради под менторством др Магдалене Ђорђевић и др Игора Салома. Био је ангажован на пројекту основних истраживања Министарства просвете, науке и технолошког развоја Републике Србије ОИ 171004 ("ATLAS експеримент и физика честица на LHC енергијама") у Лабораторији за физику високих енергија Института за физику у Београду. Ангажован је и на пројекту Horizon 2020 ERC-2016-CoG:72574 ("A novel Quark-Gluon Plasma tomography tool: from jet quenching to exploring the extreme medium properties"). Тему докторске дисертације под називом "Развој DREENA модела за томографију кварк-глуонске плазме" успешно брани пред Колегијумом докторских студија 30.06.2021.

Душан Жигић започео је рад на Институту за физику 01.02.2018. по основу ангажовања на горе наведеном ERC пројекту чији је руководилац др Магдалена Ђорђевић. 11.12.2018. године добија звање *Истраживач и управник* и од 01.01.2019. године запослен је на Институту за физику.

До сада је објавио 6 научних радова у области теоријске нуклеарне физике (четири категорије M21 и два категорије M22) од којих је наведен као први аутор на три рада. Такође, коаутор је на три рада у области квантитативне биологије од којих је један категорије M21a, један категорије M21 и један категорије M22.

Истраживање које се спроводи у оквиру пројеката на којима је ангажован представио је на осам престижних међународних научних конференција из области кварк-глуонске плазме. Неке од њих су: Quark Matter 2019, Strangness in Quark Matter 2019, Hard Probes 2018, Hard Probes 2020... Душан Жигић је такође учествовао и на четири школе.

Преглед научне активности Душана Жигића

Истраживање Душана Жигића је у области теоријског пручавања кварк-глуонске плазме. Циљ његовог истраживања је развијање нумеричких метода које ће користити високо-енергијске опсервабле за екстракцију особина овог новог стања материје. У свом досадашњем истраживању прво је развио DREENA-C модел који је био део његове мастер тезе за коју је био награђен наградом "проф. др Љубомир Тирковић" за најбољу мастер тезу, коју додељује Физички факултет.

Ови резултати су објављени у *J. Phys. G* **46**, no.8, 085101 (2019) раду на којем је Душан Жигић први аутор. У овом раду је презентован DREENA-C модел за рачунање пригушења високо-енергијских честица заснован на динамичком формализму губитака енергије у средини константне температуре и коначних димензија. Генерисана су заједничка R_{AA} и v_2 предвиђања за лаке и тешке честице, као и за различите централности код $Pb + Pb$ судара на *LHC*-у и упоређена су са доступним експерименталним подацима. Упркос чињеници да DREENA-C не узима у обзир еволуцију КПП (на коју је v_2 изразито осетљив), као и чињеници да су се други приступи сусрели са проблемима у покушајима објашњења података за v_2 (предвиђања испод података), DREENA-C доводи до квалитативно доброг слагања са подацима, мада квантитативно, предвиђања су видно изнад података. Објашњење ових резултата, приказано у раду, подржава применљивост модела и аналитичко разматрање указује да би увођење температурске еволуције средине у DREENA модел требало да доведе до бољег заједничког слагања са R_{AA} и v_2 подацима и омогући боље разумевање КПП. Обзиром да R_{AA} опсервабла није осетљива на еволуцију средине, DREENA-C модел је искоришћен да се предложи нова опсервабла за екстраковање зависности губитака енергије од дужине пређеног пута што је објављено у раду M. Djordjevic, D. Zigic, M. Djordjevic and J. Auvinen, *Phys. Rev. C Rapid Communications* **99**, no.6, 061902 (2019) .

Његово истраживање се наставило у смеру увођења температурских еволуција у DREENA модел. Кренуо је са увођењем Бјоркен експанзије у којој температура има аналитичку зависност само од времена (истраживање је објављено у D. Zigic, I. Salom, J. Auvinen, M. Djordjevic and M. Djordjevic, *Phys. Lett. B* **791**, 236-241 (2019)). DREENA-B модел је искориштен за испитивање различитих сценарија губитака енергије у стадијуму КПП пре термализације. Ови резултати су објављени у раду D. Zigic, B. Ilic, M. Djordjevic and M. Djordjevic, *Phys. Rev. C* **101**, no.6, 064909 (2020) . Тренутни фокус његовог истраживања је развијање DREENA-A модела који ће користити реалистичне (3+1)-димензионе температурске еволуције и на основу губитака енергије високо-енергијских честица ће се проучавати особине КПП.

Искуство из нумеричке анализе података је од недавно применио у квантитативној биологији, а на једном од тренутно најактуелнијих проблема, разумевању преносивости COVID-19 у популацији (три објављена рада).

Листа М20 публикација:

А. Теоријска нуклеарна физика:

1. M. Djordjevic, S. Stojku, D. Zigic, B. Ilic, J. Auvinen, I. Salom, M. Djordjevic and P. Huovinen, *From high p_{\perp} theory and data to inferring anisotropy of Quark-Gluon Plasma* Nucl. Phys. A **1005**, 121900 (2021) (M22, IF 1.695)
2. D. Zigic, B. Ilic, M. Djordjevic and M. Djordjevic, *Exploring the initial stages in heavy-ion collisions with high- p_{\perp} R_{AA} and v_2 theory and data*, Phys. Rev. C **101**, no. 6, 064909 (2020) (M21, IF 3.296)
3. D. Zigic, I. Salom, J. Auvinen, M. Djordjevic and M. Djordjevic, *DREENA-B framework: first predictions of R_{AA} and v_2 within dynamical energy loss formalism in evolving QCD medium*, Phys. Lett. B **791**, 236 (2019) (M21, IF 4.384)
4. D. Zigic, I. Salom, J. Auvinen, M. Djordjevic and M. Djordjevic, *DREENA-C framework: joint R_{AA} and v_2 predictions and implications to QGP tomography*, J. Phys. G **46**, no. 8, 085101 (2019) (M21, IF 3.456)
5. M. Djordjevic, D. Zigic, M. Djordjevic and J. Auvinen, *How to test path-length dependence in energy loss mechanisms: analysis leading to a new observable*, Phys. Rev. C **99**, no. 6, 061902(R) (2019) (M21, IF 3.304)
6. M. Djordjevic, D. Zigic, B. Blagojevic, J. Auvinen, I. Salom and M. Djordjevic, *Dynamical energy loss formalism: from describing suppression patterns to implications for future experiments*, Nucl. Phys. A **982**, 699-702 (2019) (M22, IF 1.992)

Б. Квантитативна биологија:

7. O. Milicevic, I. Salom, M. Tumbas, A. Rodic, S. Markovic, D. Zigic, M. Djordjevic, M. Djordjevic, *PM2.5 as a major predictor of COVID-19 basic reproduction number in the USA*, Environmental Research, in press (2021).
<https://doi.org/10.1016/j.envres.2021.111526> (M21a, IF 6.498)
8. I. Salom, A. Rodic, O. Milicevic, D. Zigic, M. Djordjevic, M. Djordjevic, *Effects of demographic and weather parameters on COVID-19 basic reproduction number*, Frontiers in Ecology and Evolution, 8, 617841 (2021). (M21, IF 4.171)
9. M. Djordjevic, A. Rodic, I. Salom, D. Zigic, O. Milicevic, B. Ilic, M. Djordjevic, *A systems biology approach to COVID-19 progression in population*, Advances in Protein Chemistry and Structural Biology, in press (2021).
<https://doi.org/10.1016/bs.apcsb.2021.03.003> (M22, IF 3.507)

Саопштења са међународног скупа штампана у изводу (категорија М34):

Напомена: испод су укључена само саопштења која је кандидат сам излагао, не и саопштења коаутора.

А. Предавања

1. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *Using DREENA framework to explore properties of QGP*, Online Strangeness in Quark Matter Conference 2021, May 2021
2. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *DREENA framework as a multipurpose tool for QGP tomography*, Hard Probes 2020: 10th International Conference on Hard & Electromagnetic Probes of High-Energy Nuclear Collisions, June 2020, Austin, TX, USA
3. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *QGP tomography with DREENA framework*, 3rd JETSCAPE Winter School and Workshop 2020, March 2020, Knoxville, Tennessee, USA
4. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *DREENA framework as a multipurpose tool for QGP tomography*, Zimanyi school 2019: 19th Zimanyi school - Winter workshop on heavy ion physics, Dec 2019, Budapest, Hungary
5. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *Recent results with full DREENA framework as a multipurpose tool for QGP tomography*, 2019 COST Action CA15213 THOR Annual Meeting, Sep 2019, Istanbul, Turkey

Б. Постери

6. Dusan Zigic, Igor Salom, Jussi Auvinen, Marko Djordjevic, Magdalena Djordjevic, *DREENA framework as a multipurpose tool for QGP tomography*, Quark Matter 2019 - the XXVIIIth International Conference on Ultra-relativistic Nucleus-Nucleus Collisions, Nov 2019, Wuhan, China
7. Dusan Zigic, Igor Salom, Jussi Auvinen, Magdalena Djordjevic, Marko Djordjevic, *DREENA framework: predictions and comparison with experimental data*, Strangeness in Quark Matter, Jun 2019, Bari, Italy
8. Dusan Zigic, Igor Salom, Jussi Auvinen, Marko Djordjevic, Magdalena Djordjevic, *Numerical predictions of DREENA-C and DREENA-B frameworks*, Hard Probes 2018: International Conference on Hard & Electromagnetic Probes of High-Energy Nuclear Collisions, Oct 2018, Aix-Les-Bains, France

PAPER • OPEN ACCESS

DREENA-C framework: joint R_{AA} and v_2 predictions and implications to QGP tomography

To cite this article: Dusan Zigic *et al* 2019 *J. Phys. G: Nucl. Part. Phys.* **46** 085101

View the [article online](#) for updates and enhancements.

Recent citations

- [Predicting parton energy loss in small collision systems](#)
Alexander Huss *et al*
- [Extracting the temperature dependence in high- \$p\$ particle energy loss](#)
Stefan Stojku *et al*
- [From high \$p\$ theory and data to inferring anisotropy of Quark-Gluon Plasma](#)
Magdalena Djordjevic *et al*

DREENA-C framework: joint R_{AA} and v_2 predictions and implications to QGP tomography

Dusan Zigic¹, Igor Salom¹, Jussi Auvinen¹,
Marko Djordjevic² and Magdalena Djordjevic^{1,3} 

¹Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia

²Faculty of Biology, University of Belgrade, Belgrade, Serbia

E-mail: magda@ipb.ac.rs

Received 26 November 2018, revised 7 May 2019

Accepted for publication 21 May 2019

Published 26 June 2019



CrossMark

Abstract

In this paper, we presented our recently developed Dynamical Radiative and Elastic ENergy loss Approach (DREENA-C) framework, which is a fully optimized computational suppression procedure based on our state-of-the-art dynamical energy loss formalism in constant temperature finite size QCD medium. With this framework, we have generated, for the first time, joint R_{AA} and v_2 predictions within our dynamical energy loss formalism. The predictions are generated for both light and heavy flavor probes, and different centrality regions in Pb + Pb collisions at the LHC, and compared with the available experimental data. While R_{AA} predictions agree with experimental data, v_2 predictions qualitatively agree with, but are quantitatively visibly above, the experimental data (in disagreement with other models, which underestimate v_2). Consistently with numerical predictions, through simple analytic analysis, we show that R_{AA} is insensitive to medium evolution (though highly sensitive to energy loss mechanisms), while v_2 is highly sensitive to the evolution. As a major consequence for precision quark-gluon plasma (QGP) tomography, this then leaves a possibility to calibrate energy loss models on R_{AA} data, while using v_2 to constrain QGP parameters that are in agreement with both high and low p_{\perp} data.

³ Author to whom correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: quark-gluon plasma, parton energy loss, heavy quarks, perturbative QCD, QGP tomography, high p_{\perp} suppression predictions

(Some figures may appear in colour only in the online journal)

1. Introduction

Quark-gluon plasma (QGP) is a new state of matter [1, 2] consisting of interacting quarks, antiquarks and gluons. Such a new state of matter is created in ultra-relativistic heavy ion collisions at Relativistic Heavy Ion Collider (RHIC) and Large Hadron Collider (LHC). Rare high momentum probes, which are created in such collisions and which traverse QGP, are excellent probes of this extreme form of matter [3–5]. Different observables (such as angular averaged nuclear modification factor R_{AA} and angular anisotropy v_2), together with probes with different masses, probe this medium in a different manner. Therefore, comparing comprehensive set of joint predictions for different probes and observables, with available experimental data from different experiments, collision systems and collision energies, allows investigating properties of QCD medium created in these collisions [6–12].

However, to implement this idea, it is necessary to have a model that realistically describes high- p_{\perp} parton interactions with the medium. With this goal in mind, we developed state-of-the-art dynamical energy loss formalism [13, 14], which includes different important effects (some of which are unique to this model). Namely, (i) the formalism takes into account finite size, finite temperature QCD medium consisting of dynamical (that is moving) partons, contrary to the widely used static scattering approximation and/or medium models with vacuum-like propagators. (ii) The calculations are based on the finite temperature field theory [15, 16], and generalized HTL approach, in which the infrared divergencies are naturally regulated, so the model does not have artificial cutoffs. (iii) Both radiative [13] and collisional [17] energy losses are calculated under the same theoretical framework, applicable to both light and heavy flavor. (iv) The formalism is generalized to the case of finite magnetic [18] mass and running coupling [19], and most recently, we also applied first steps towards removing widely used soft-gluon approximation from radiative energy loss calculations, enhancing the applicability region of this formalism [20]. This formalism was further integrated into numerical procedure [19], which includes initial p_{\perp} distribution of leading partons [21, 22], energy loss with path-length [23, 24] and multi-gluon [25] fluctuations, and fragmentation functions [26–28], to generate the final medium modified distribution of high p_{\perp} hadrons. While all the above effects have to be included based on theoretical grounds, it is plausible to ask whether all of these ingredients are necessary for accurately interpreting the experimental data, particularly since other available approaches [29–33] commonly neglect some—or many—of these effects. To address this important issue, in [34], we showed that, while abolishing widely used static approximation is the most important step for accurate suppression predictions, *including all other effects* is necessary for a fine agreement with high- p_{\perp} R_{AA} (and v_2 , not published) data.

To be able to generate predictions that can reasonably explain the experimental data, all ingredients stated above have to be preserved (with no additional simplifications used in the numerical procedure), as all of these ingredients were shown to be important for reliable theoretical predictions of jet suppression [34]. From computational perspective, it is also necessary to develop a framework that can efficiently generate wide set of theoretical predictions, to be compared with a broad range of available (or upcoming) experimental data. We here present DREENA-C (Dynamical Radiative and Elastic ENergy loss Approach) framework, which is the first step towards this goal. Due to the complexity of the underlying parton-medium

interaction model, this first step takes into account the medium evolution in its simplest form, through mean (constant) medium temperature (thus ‘C’ in DREENA-C framework). In addition to presenting the necessary baseline to be compared with future redevelopments of the dynamical energy loss to more fully account for the medium evolution, DREENA-C is also an optimal numerical framework for studying the medium evolution effects on certain observables. That is, as this framework takes into account state-of-the-art parton-medium interaction model, but only rudimental medium evolution, comparison of its predictions with experimental data allows assessing sensitivity of certain variables to QGP evolution.

DREENA-C framework corresponds to, in its essence, the numerical procedure presented in [19], with a major new development that the code is now optimized to use minimal computer resources and produce predictions within more than two orders of magnitude shorter time compared to [19]. Such step is necessary, as all further improvements of the framework, necessarily need significantly more computer time and resources. So, without this development, further improvements, e.g. towards nontrivially evolving QGP medium, would not be realistically possible. That is, DREENA-C framework, addresses the goal of efficiently generating predictions for diverse observables.

Exploiting the ability to generate predictions for a wide range of observables, we will here use DREENA-C framework to, for the first time, present joint R_{AA} and v_2 theoretical predictions within our dynamical energy loss formalism; these predictions will be generated for different experiments (ALICE, CMS and ATLAS), probes (light and heavy) and experimental conditions (wide range of collision centralities). Note that some of our results correspond to true predictions (some centrality intervals for B and D mesons), while for other cases, e.g. for charged hadrons, they correspond to postdictions, as the experimental data are already available. Motivation for generating these predictions is the following: (i) the theoretical models up to now were not able to jointly explain these data, which is known as v_2 puzzle [35, 36]. That is, the models lead to underprediction of v_2 , unless new phenomena (e.g. magnetic monopoles) are introduced [37]. (ii) Having this puzzle in mind, and the fact that other available models employ the complementary approach, i.e. combine simplified energy loss models with more sophisticated medium evolutions, this work will enable assessing to what extent state-of-the-art energy loss model, but with simplest QGP evolution, is able to jointly explain R_{AA} and v_2 data. To obtain additional understanding of this important issue, we will bellow complement DREENA-C predictions with analytical estimates. (iii) DREENA-C predictions will establish an important baseline for testing how future introduction of the medium evolution will improve the formalism. Moreover, such step-by-step introduction of different medium evolution effects in the model will also allow to investigate their importance in explaining the experimental data, which is highly relevant for QGP tomography.

2. Methods

The DREENA-C framework is a fully optimized numerical procedure, which contains all ingredients presented in detail in [19]. We below briefly outline the main steps in this procedure.

The quenched spectra of light and heavy flavor observables are calculated according to the generic pQCD convolution:

$$\frac{E_f d^3\sigma}{dp_f^3} = \frac{E_i d^3\sigma(Q)}{dp_i^3} \otimes P(E_i \rightarrow E_f) \otimes D(Q \rightarrow H_Q). \quad (1)$$

Subscripts i and f correspond, respectively, to ‘initial’ and ‘final’, and Q denotes initial light or heavy flavor jet. $E_i d^3\sigma(Q)/dp_i^3$ denotes the initial momentum spectrum, which are computed

according to [21, 22], $P(E_i \rightarrow E_f)$ is the energy loss probability, computed within the dynamical energy loss formalism [13, 14], with multi-gluon [25], path-length fluctuations [24] and running coupling [19]. $D(Q \rightarrow H_Q)$ is the fragmentation function of light and heavy flavor parton Q to hadron H_Q , where for light flavor, D and B mesons we use, DSS [26], BCFY [27] and KLP [28] fragmentation functions, respectively.

Regarding the numerical procedure, a major new development is that the code is now optimized, so that it is two orders of magnitude faster compared to the brute-force approach applied in [19]. Technically, the main optimization method we used was a combination of tabulation and interpolation of values of intermediary functions that appear at various steps of the energy loss calculation. This approach significantly reduces the number of necessary integrations. However, it must be preceded by careful analysis of the behavior of interpolated functions and the function sampling must be tailored to this behavior, so that effectively no loss of precision is introduced. Furthermore, in comparison to the computation of [19], different and better suited methods of numerical integration were used (mostly quasi Monte Carlo integration), producing a large speedup, higher integration precision and stability of the underlying results. Finally, the code was parallelized to take advantage of contemporary multi-core workstations. Furthermore, the optimization also allowed for further improvements of the physical model: (i) due to numerical constraints, in the previous multi-gluon fluctuation procedure, the number of radiated gluons was limited to 3. The procedure is now redeveloped to include the arbitrary number of radiated gluons; the detailed numerical analysis (both from the point of numerical precision and time efficiency) showed that the optimal limit of gluons to be included in the procedure is 4–5. (ii) Both radiative and collisional energy losses are now combined gradually along the traversed path of the parton, unlike in [19], where radiative and collisional losses were accounted separately.

As noted above, we model the medium by assuming a constant average temperature of QGP. We concentrate on the central rapidity region in 5.02 TeV Pb + Pb collisions at the LHC, though we note that these predictions will be applicable for 2.76 TeV Pb + Pb collisions as well, since the predictions for these two collision energies almost overlap [38]. To determine the temperature for each centrality region in 5.02 TeV Pb + Pb collisions, we use [39, 40] $T^3 \sim \frac{dN_g}{dy} \frac{1}{A_\perp \bar{L}} \rightarrow T = c \left(\frac{dN_{ch}}{dy} \frac{1}{A_\perp \bar{L}} \right)^{1/3}$, where $\frac{dN_g}{dy}$ is gluon rapidity density, A_\perp is the overlap area and \bar{L} is the average size of the medium for each centrality region. At mid rapidity, $\frac{dN_g}{dy}$ is directly proportional to experimentally measured charged particle multiplicity $\frac{dN_{ch}}{dy}$, which is measured for 5.02 TeV Pb + Pb collisions at the LHC across different centralities [41]. Furthermore, c is a constant, which can be fixed through ALICE measurement of effective temperature for 0%–20% centrality at 2.76 TeV Pb + Pb collisions LHC [42]. For each centrality region, path-length distributions (as well as overlap area A_\perp and average size of the medium \bar{L}) are calculated following the procedure described in [23], with an additional hard sphere restriction $r < R_A$ in the Woods–Saxon nuclear density distribution to regulate the path lengths in the peripheral collisions.

In numerical calculations, we use no fitting parameters in generating predictions for comparison with the data, i.e. all the parameters correspond to standard literature values. We consider a QGP with $\Lambda_{\text{QCD}} = 0.2$ GeV and $n_f = 3$. The temperature dependent Debye mass $\mu_E(T)$ is obtained from [43], while for the light quarks, we assume that their mass is dominated by the thermal mass $M \approx \mu_E/\sqrt{6}$, and the gluon mass is $m_g \approx \mu_E/\sqrt{2}$ [44]. The charm (bottom) mass is $M = 1.2$ GeV ($M = 4.75$ GeV). Finite magnetic mass effect is also included in our framework [18], as various non-perturbative calculations [45, 46] have shown that magnetic mass μ_M is different from zero in QCD matter created at the LHC and RHIC.

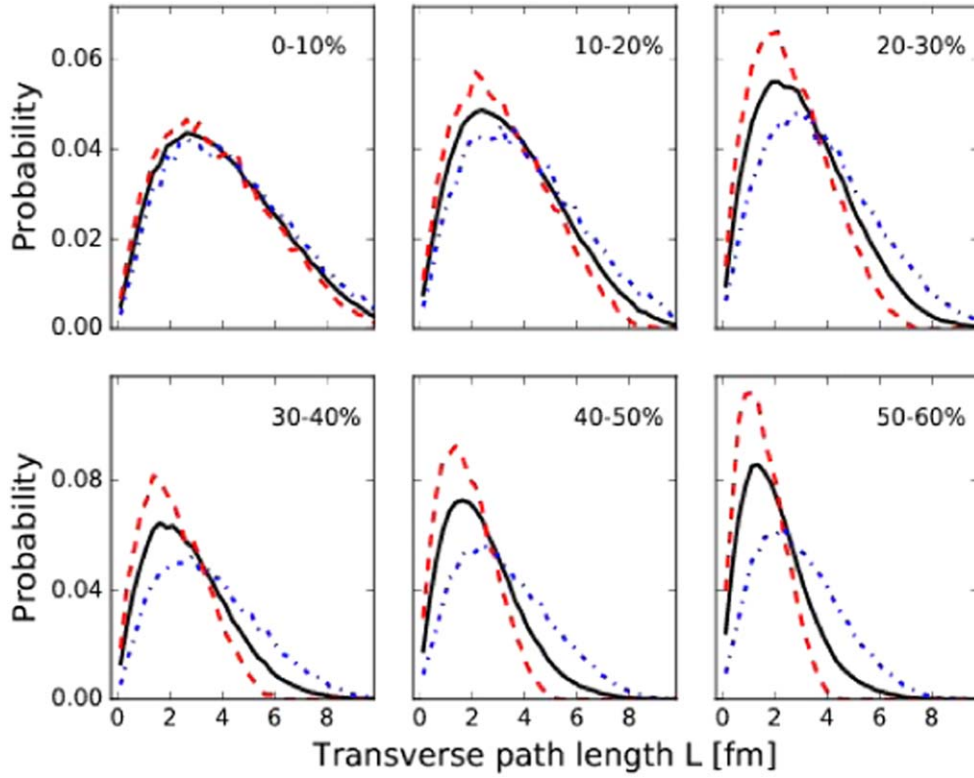


Figure 1. Path-length distributions. Probability distributions for hard parton path lengths in Pb + Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV for (0–10)%–(50–60)% centrality classes. Solid black curves: the total distributions with all hard partons included are represented; Dashed red curves: the distributions include only in-plane particles ($|\phi| < 15^\circ$ or $|\phi - 180^\circ| < 15^\circ$); dashed–dotted blue curves: the distributions include only out-of-plane partons ($|\phi - 90^\circ| < 15^\circ$).

Magnetic to electric mass ratio is extracted from these calculations to be $0.4 < \mu_M/\mu_E < 0.6$, so presented uncertainty in the predictions comes from this range of screening masses ratio. Note that other uncertainties (e.g. in quark masses or effective temperature), are not included in this study. However, we have checked that uncertainties in the quark masses lead to small (up to 4% for $p_\perp > 8$ GeV, and decreasing with increasing p_\perp) difference in the resulting predictions. Regarding effective temperature, as this temperature comes with large error bars, in [47] we presented a detailed study of how this uncertainty affects the R_{AA} calculations. We found that R_{AA} dependence on T is almost linear (and the same for all parton energies and all types of flavor) and does not significantly affect the suppression, concluding that uncertainty in the effective temperature would basically lead to a systematic (constant value) shift in the predictions, i.e. the results presented in this paper would not be affected by this uncertainty.

3. Results and discussion

In this section, we will present joint R_{AA} and v_2 predictions for high p_\perp charged hadrons, D and B mesons in Pb + Pb collisions at the LHC. In figure 1 we first show probability

distributions for hard parton path lengths in Pb + Pb collisions for different centralities, obtained by the procedure specified in the previous section. For most central collisions, we observe that in-plane and out-of-plane distributions almost overlap with the total (average) path-length distributions, as expected. As the centrality increases, in-plane and out-of-plane distributions start to significantly separate (in different directions) from average path-length distributions. Having in mind that [48]

$$v_2 \approx \frac{1}{2} \frac{R_{AA}^{\text{in}} - R_{AA}^{\text{out}}}{R_{AA}^{\text{in}} + R_{AA}^{\text{out}}}, \quad (2)$$

this leads to the expectation of v_2 being small in most central collisions and increasing with increasing centrality. Regarding the equation (2) above, note that this estimate presents a conventional way [48–51] to calculate high p_\perp v_2 , and it leads to exact result if the higher harmonics v_4 , v_6 , etc. are zero at high p_\perp , and the opening angle (where R_{AA}^{in} and R_{AA}^{out} are evaluated) goes to zero.

Based on path-length distributions from figure 1, we can now calculate average R_{AA} , as well as in-plane and out-of-plane R_{AA} s (R_{AA}^{in} and R_{AA}^{out}), and consequently v_2 for both light and heavy flavor probes and different centralities. We start by generating predictions for charged hadrons, where data for both R_{AA} and v_2 are available. Comparison of our joint predictions with experimental data is shown in figure 2, where left and right panels correspond, respectively, to R_{AA} and v_2 . We see good agreement with R_{AA} data, which is also robust, i.e. achieved across wide range of centralities and experiments. Regarding v_2 , we surprisingly see that our v_2 predictions are visibly above the data. This is in contrast with other energy loss models which consistently lead to underprediction of v_2 , where to resolve this, new phenomena (e.g. magnetic monopoles) were introduced [37]. Despite this quantitative disagreement, we see a reasonable qualitative agreement between the model and the data, i.e. the predictions are just shifted above the data; this will be further discussed below.

In figure 3, we provide predictions for D meson average R_{AA} (left panel) and v_2 (right panel) data, for four different centrality regions. The predictions are compared with the available 5.02 TeV Pb + Pb experimental data. For average R_{AA} , we observe good agreement with the data. Regarding v_2 , we observe similar behavior as for charged hadron: i.e. while we obtain a reasonable qualitative agreement with the measurements, quantitatively there is again an unexpected (having in mind predictions of other models) overestimation of the data. Figure 4 shows equivalent predictions as figure 3, only for B mesons. For R_{AA} , we compare our predictions with the available B^\pm [58], B_s^0 [59], non-prompt J/Ψ [60, 61] and non-prompt D_0 [62] data. Note that we can compare B meson predictions with these indirect b quark suppression data, as due to interplay of collisional and radiative energy loss, B meson suppression is almost independent on p_\perp for $p_\perp > 10$ GeV [47], so the fragmentation/decay functions will not play a large role for different types of b quark observables. Also, note that our predictions are provided for mid-rapidity region; for non-prompt D_0 (which are given for $|y| < 1$), we see good agreement between our predictions and the data. For B^\pm and non-prompt J/Ψ , our predictions show qualitatively good agreement, but overprediction of R_{AA} data. This is expected, having in mind that those data are given for $|y| \lesssim 2$, where both experiments show 30%–50% increase in R_{AA} with decreasing rapidity. Our predictions do not agree with B_s^0 , but these data come with very large error bars. For v_2 , we predict values significantly different from zero for all centrality regions, and see that our predictions agree with the available non-prompt J/Ψ data [61, 63], though we note that these predictions are given with very large error bars. This does not necessarily mean that heavy B meson flows, as flow is inherently connected with *low* p_\perp v_2 , and here we show predictions for high p_\perp . On

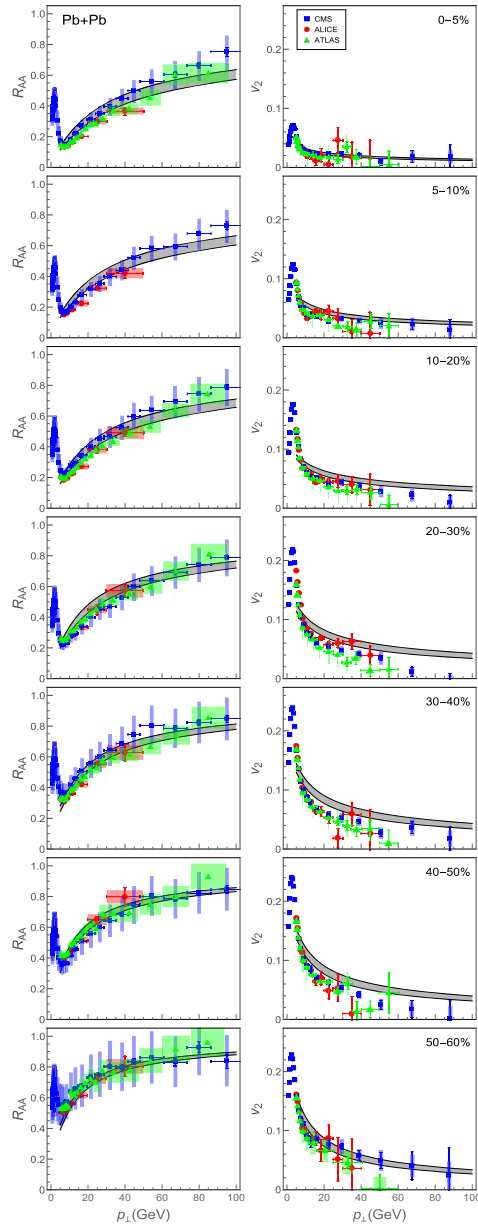


Figure 2. Joint R_{AA} and v_2 predictions for charged hadrons. *Left panels:* theoretical predictions for R_{AA} versus p_{\perp} are compared with ALICE [52] (red circles), CMS [53] (blue squares) and ATLAS [54] (green triangles) charged hadron experimental data for 5.02 TeV Pb + Pb collisions at the LHC. *Right panels:* theoretical predictions for v_2 versus p_{\perp} are compared with ALICE [55] (red circles), CMS [56] (blue squares) and ATLAS [57] (green triangles) charged hadron experimental data for 5.02 TeV Pb + Pb collisions at the LHC. The gray band boundaries correspond to $\mu_M/\mu_E = 0.4$ and $\mu_M/\mu_E = 0.6$. Rows 1–7 correspond to, respectively, 0%–5%, 5%–10%, 10%–20%, ..., 50%–60% centrality regions.

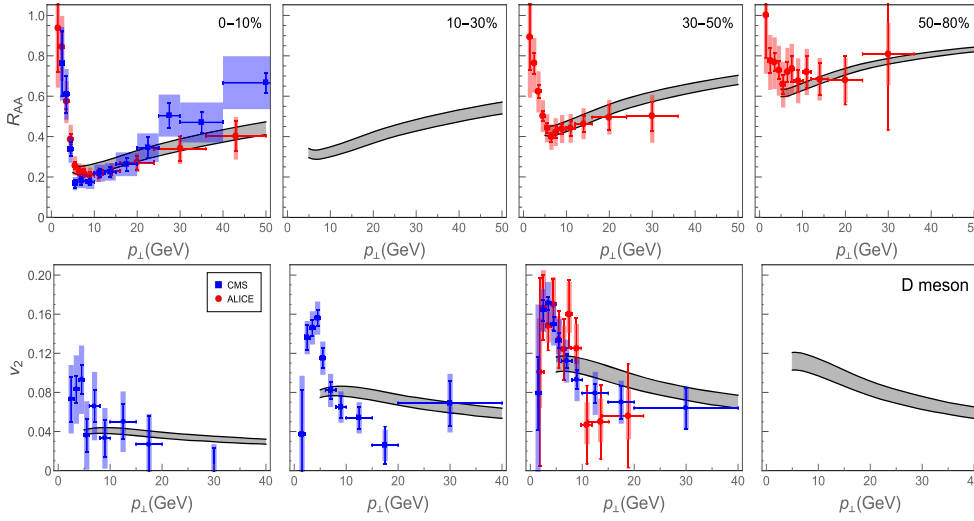


Figure 3. Joint R_{AA} and v_2 predictions for D mesons. *Upper panels:* theoretical predictions for R_{AA} versus p_{\perp} are compared with ALICE [64] (red circles) and CMS [65] (blue squares) D meson experimental data for 5.02 TeV Pb + Pb collisions at the LHC. *Lower panels:* theoretical predictions for v_2 versus p_{\perp} are compared with ALICE [51] (red circles) and CMS [66] (blue squares) D meson experimental data for 5.02 TeV Pb + Pb collisions at the LHC. The gray band boundaries correspond to $\mu_M/\mu_E = 0.4$ and $\mu_M/\mu_E = 0.6$. First to fourth column correspond to, respectively, 0%–10%, 10%–30%, 30%–50% and 50%–80% centrality regions.

the other hand, high p_{\perp} v_2 is connected with the difference in the energy loss (i.e. suppression) for particles going in different (e.g. in-plane and out-of-plane) directions; this difference then leads to our predictions of non-zero v_2 for *high* p_{\perp} B mesons.

Overall, we see that our predicted R_{AA} s agree well with all measured (light and heavy flavor) data, while our v_2 predictions are consistently above the experimental data. Since our model has sophisticated description of parton-medium interactions, but highly simplified medium evolution model (through average medium temperature), these robust numerical results imply the following: (i) R_{AA} is largely insensitive to the medium evolution, in contrast to its (previously shown [34]) large sensitivity to parton-medium interactions. (ii) v_2 is sensitive to the details of medium evolution. These two conclusions have important implications for QGP tomography, in particular (i) R_{AA} can be used to calibrate parton-medium interaction models, while (ii) v_2 can be used to constrain QGP medium evolution parameters also from the point of high p_{\perp} data (in addition to constraining them from low p_{\perp} predictions/measurements). One should note that insensitivity of R_{AA} and sensitivity of v_2 predictions to QGP evolution were also observed by using very different models and numerical frameworks [67, 68]. This then clearly suggests that such (in)sensitivity may be a general phenomenon, but to claim this, one should also gain an analytical understanding, which we provide below. Furthermore, the numerical results presented above also lead to the following questions, which are important from the point of future precision QGP tomography: (i) what is the reason behind the observed overestimation of v_2 within DREENA-C framework, and can expanding medium lead to a better agreement with the experimental data? (ii) Do we expect that B meson v_2 predictions will still be non-zero, once the expanding medium is introduced?

To intuitively approach the issues raised above, we start by noting that, within our dynamical energy loss formalism, $\Delta E/E \sim T^a$ and $\Delta E/E \sim L^b$, where $a, b \rightarrow 1$ ($\Delta E/E$ is

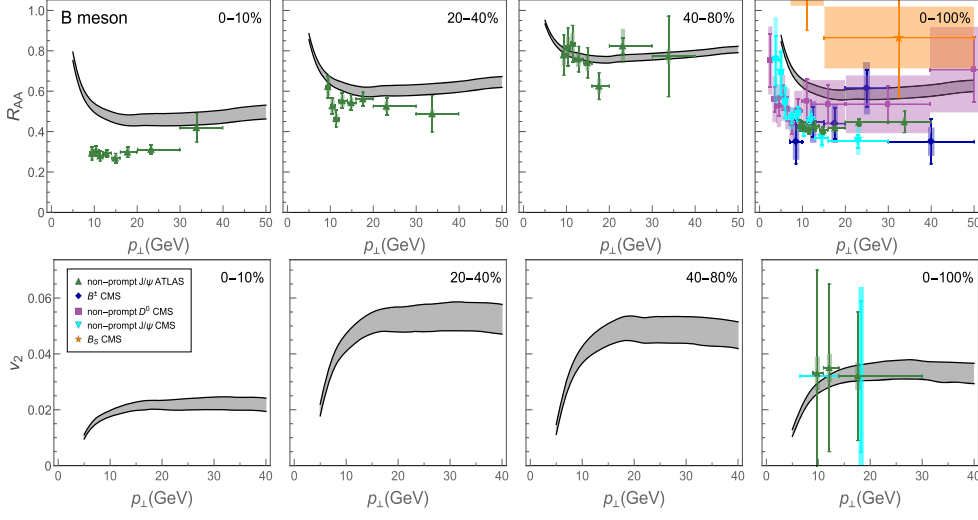


Figure 4. Joint R_{AA} and v_2 predictions for B mesons. *Upper panels:* theoretical predictions for B meson R_{AA} versus p_{\perp} are compared with ATLAS [60] (green triangles), CMS [61] (cyan triangles) non-prompt J/Ψ , and CMS non-prompt D^0 [62] (purple squares), B^{\pm} [58] (blue diamonds) and B_s^0 [59] (orange stars) experimental data for 5.02 TeV Pb + Pb collisions at the LHC. *Lower panels:* theoretical predictions for B meson v_2 versus p_{\perp} are compared with ATLAS [63] (green triangles) and CMS [61] (cyan triangles) non-prompt J/Ψ for 5.02 TeV Pb + Pb collisions at the LHC. The gray band boundaries correspond to $\mu_M/\mu_E = 0.4$ and $\mu_M/\mu_E = 0.6$. First to fourth column correspond to, respectively, 0%–10%, 20%–40%, 40%–80% and 0%–100% centrality regions.

fractional energy loss, T is the average temperature of the medium, while L is the average path-length traversed by the jet). To be more precise, note that both dependencies are close to linear, though a and b are still significantly different from 1 [38]. However, for the purpose of this estimate, let us assume that $a = b = 1$, leading to

$$\Delta E/E \approx \chi TL, \quad (3)$$

where χ is a proportionality factor.

Another commonly used estimate [25] is that

$$R_{AA} \approx \left(1 - \frac{1}{2} \frac{\Delta E}{E}\right)^{n-2}, \quad (4)$$

where n is the steepness of the initial momentum distribution function (i.e. approximate exponent of a power-law of initial momentum distribution p_{\perp}^{-n}), and $\Delta E/E$ is notably smaller than 1.

In the case when fractional energy loss $\Delta E/E \ll 1$, equation (4) becomes

$$R_{AA} \approx \left(1 - \frac{n-2}{2} \frac{\Delta E}{E}\right) \approx (1 - \xi TL), \quad (5)$$

where $\xi = (n-2) \chi/2$.

In the DREENA-C approach, T is constant, and the same in in-plane and out-of-plane directions, while $L_{in} = L - \Delta L$ and $L_{out} = L + \Delta L$, leading to

$$\begin{aligned}
R_{AA} &\approx \frac{1}{2}(R_{AA}^{\text{in}} + R_{AA}^{\text{out}}) \approx \frac{1}{2}(1 - \xi TL_{\text{in}} + 1 - \xi TL_{\text{out}}) \\
&= 1 - \xi T \frac{L_{\text{in}} + L_{\text{out}}}{2} = 1 - \xi TL,
\end{aligned} \tag{6}$$

and

$$\begin{aligned}
v_2 &\approx \frac{1}{2} \frac{R_{AA}^{\text{in}} - R_{AA}^{\text{out}}}{R_{AA}^{\text{in}} + R_{AA}^{\text{out}}} \approx \frac{1}{2} \frac{(1 - \xi TL_{\text{in}}) - (1 - \xi TL_{\text{out}})}{2(1 - \xi TL)} \\
&= \frac{1}{2} \frac{\xi T \Delta L}{1 - \xi TL} \approx \frac{\xi T \Delta L}{2}.
\end{aligned} \tag{7}$$

If the medium evolves, and by assuming 1 + 1D Bjorken time evolution [69] (as qualitatively sufficient for the early time dynamics [70]), the average temperature along in-plane will be larger than along out-of-plane direction [71], leading to $T_{\text{in}} = T + \Delta T$ and $T_{\text{out}} = T - \Delta T$ (where $\Delta L/L \cdot \Delta T/T \ll 1$). By repeating the above procedure in this case, it is straightforward to obtain

$$R_{AA} \approx 1 - \xi TL \tag{8}$$

and

$$\begin{aligned}
v_2 &\approx \frac{1}{2} \frac{(1 - \xi T_{\text{in}} L_{\text{in}}) - (1 - \xi T_{\text{out}} L_{\text{out}})}{2(1 - \xi TL)} = \frac{1}{2} \frac{\xi T \Delta L - \xi \Delta TL}{1 - \xi TL} \\
&\approx \frac{\xi T \Delta L - \xi \Delta TL}{2}.
\end{aligned} \tag{9}$$

We see that, while v_2 explicitly depends on ΔT and ΔL , R_{AA} does not. Therefore, it follows that, consistently with previous numerical results, R_{AA} can be only weakly sensitive to QGP evolution, while v_2 is quite sensitive to this evolution; note that this is to our knowledge, the first time that analytical argument to sensitivity of R_{AA} and v_2 to medium evolution is provided. Moreover, from equations (7) and (9), we see that introduction of temperature evolution is expected to lower v_2 compared to constant T case. Consequently, an accurate/complete energy loss models, when applied in the context of constant temperature medium should lead to higher v_2 than expected, while introduction of T evolution in such models would lower the v_2 compared to non-evolving case. Based on this, and the fact that previous theoretical approaches were not able to reach high enough v_2 without introducing new phenomena [37], we argue that accurate description of high- p_{\perp} parton-medium interactions is crucial for accurate description of high- p_{\perp} experimental data. With regards to this, the above results strongly suggest that the dynamical energy loss formalism has the right features needed to accurately describe jet-medium interactions in QGP, which is crucial for high precision QGP tomography.

Regarding the second question mentioned above, for B meson to have $v_2 \approx 0$, it is straightforward to see that one needs $\Delta T/T \approx \Delta L/L$. Having in mind that $\Delta L/L$ is quite large for larger centralities (see figure 1), $\Delta T/T$ would also have to be about the same magnitude. We do not expect this to happen, based on our preliminary estimates of the temperature changes in in-plane and out-of-plane in 1 + 1D Bjorken expansion scheme [69]. That is, our expectations is that B meson v_2 will be smaller than presented here, but still

significantly larger than zero, at least for large centrality regions. However, this still remains to be tested in the future with the introduction of full evolution model within our framework.

4. Conclusion

In this paper, we introduced the DREENA-C framework, which is a computational suppression procedure based on our dynamical energy loss formalism in finite size QCD medium with constant (mean) medium temperature. This approach, which combines a state-of-the-art energy loss model, but with including QGP evolution in its simplest form, is complementary to other available models that combine simplified energy loss models with more sophisticated medium evolutions. As such, DREENA-C can provide an important insight to what extent the accurate description of high- p_{\perp} parton-medium interactions *versus* accurate description of medium evolution is necessary for accurately explaining high p_{\perp} R_{AA} and v_2 measurements.

We here used the DREENA-C framework to, for the first time, generate joint R_{AA} and v_2 predictions for both light and heavy flavor probes and different centrality regions in Pb + Pb collisions at the LHC, and compare them with the available experimental data. We consistently, through both numerical and analytical calculations, obtained that R_{AA} is sensitive to the average properties of the medium, while v_2 is highly sensitive to the details of the medium evolution. Analytical calculations brought another advantage of DREENA-C, as they would likely not be possible in frameworks with more complex medium evolution models, but bring simple and intuitive predictions/explanations for our results, which is necessary for better qualitative and quantitative understanding of the obtained results.

Since different medium evolution profiles have both different average properties and different details of the evolution, in precision QGP tomography, both R_{AA} and v_2 have to be jointly used to extract the QGP properties. The DREENA-C framework presents an optimal starting point for QGP tomography, as R_{AA} predictions (obtained through DREENA-C) can be first used to calibrate the energy loss model itself; that is, DREENA-C is fast (which is important for efficient energy loss calibration), and it does not contain the details of the medium evolution, which could provide an unwanted background for such a purpose. Once this crucial step of accurate description and calibration of parton-medium interactions is achieved, different more-detailed profiles of medium evolution (generated through different bulk medium models and parameters, with and without event by event fluctuations) can be tested (through our future advancement of DREENA framework) to assess which of these profiles provide a simultaneous agreement with both high p_{\perp} R_{AA} and v_2 data, across wide range of diverse experimental data and without further adjustment of energy loss models. In this way, QGP parameters can be constrained from both low and high p_{\perp} measurements.

Furthermore, other approaches face difficulties in jointly explaining R_{AA} and v_2 data, where smaller v_2 , than experimentally observed, is obtained. In distinction to other approaches, we here obtained an overprediction of v_2 , where the analytical estimates moreover show that inclusion of more realistic medium evolution models would lead to better agreement with the data. This, together with the fact that v_2 prediction provided here already qualitatively (though not quantitatively) agree with the data, indicate an important (and highly non-trivial) conclusion that accurate description of high- p_{\perp} parton interactions with QGP is likely the most important ingredient for generating high- p_{\perp} predictions. These results therefore strongly suggest that our dynamical energy loss formalism provides a suitable basis for the QGP tomography (outlined above), which is our main future goal.

Acknowledgments

This work is supported by the European Research Council, grant ERC-2016-COG: 725741, and by the Ministry of Science and Technological Development of the Republic of Serbia, under project numbers ON171004, ON173052 and ON171031.

ORCID iDs

Magdalena Djordjevic  <https://orcid.org/0000-0001-9229-4648>

References

- [1] Collins J C and Perry M J 1975 *Phys. Rev. Lett.* **34** 1353
- [2] Baym G and Chin S A 1976 *Phys. Lett. B* **62** 241
- [3] Gyulassy M and McLerran L 2005 *Nucl. Phys. A* **750** 30
- [4] Shuryak E V 2005 *Nucl. Phys. A* **750** 64
- [5] Jacak B and Steinberg P 2010 *Phys. Today* **63** 39
- [6] Bjorken J D 1982 FERMILAB-PUB-82-059-THY 287 292
- [7] Djordjevic M, Gyulassy M and Wicks S 2005 *Phys. Rev. Lett.* **94** 112301
- [8] Dokshitzer Yu L and Kharzeev D 2001 *Phys. Lett. B* **519** 199
- [9] Burke K M *et al* (JET Collaboration) 2014 *Phys. Rev. C* **90** 014909
- [10] Aarts G *et al* 2017 *Eur. Phys. J. A* **53** 93
- [11] Akiba Y *et al* arXiv:1502.02730 [nucl-ex]
- [12] Brambilla N *et al* 2014 *Eur. Phys. J. C* **74** 2981
- [13] Djordjevic M 2009 *Phys. Rev. C* **80** 064909
- [14] Djordjevic M and Heinz U 2008 *Phys. Rev. Lett.* **101** 022302
- [15] Kapusta J I 1989 *Finite-Temperature Field Theory* (Cambridge: Cambridge University Press)
- [16] Bellac M Le 1996 *Thermal Field Theory* (Cambridge: Cambridge University Press)
- [17] Djordjevic M 2006 *Phys. Rev. C* **74** 064907
- [18] Djordjevic M and Djordjevic M 2012 *Phys. Lett. B* **709** 229
- [19] Djordjevic M and Djordjevic M 2014 *Phys. Lett. B* **734** 286
- [20] Blagojevic B, Djordjevic M and Djordjevic M 2019 *Phys. Rev. C* **99** 024901
- [21] Kang Z B, Vitev I and Xing H 2012 *Phys. Lett. B* **718** 482
- [22] Sharma R, Vitev I and Zhang B W 2009 *Phys. Rev. C* **80** 054902
- [23] Dainese A 2004 *Eur. Phys. J. C* **33** 495
- [24] Wicks S, Horowitz W, Djordjevic M and Gyulassy M 2007 *Nucl. Phys. A* **784** 426
- [25] Gyulassy M, Levai P and Vitev I 2002 *Phys. Lett. B* **538** 282
- [26] de Florian D, Sassot R and Stratmann M 2007 *Phys. Rev. D* **75** 114010
- [27] Cacciari M and Nason P 2003 *J. High Energy Phys.* **JHEP09(2003)006**
Braaten E, Cheung K-M, Fleming S and Yuan T C 1995 *Phys. Rev. D* **51** 4819
- [28] Kartvelishvili V G, Likhoded A K and Petrov V A 1978 *Phys. Lett. B* **78** 615
- [29] Baier R, Dokshitzer Yu L, Mueller A H, Peigne S and Schiff D 1997 *Nucl. Phys. B* **483** 291
Baier R, Dokshitzer Yu L, Mueller A H, Peigne S and Schiff D 1997 *Nucl. Phys. B* **484** 265
Zakharov B G 1996 *JETP Lett.* **63** 952
Zakharov B G 1997 *JETP Lett.* **65** 615
- [30] Armesto N, Salgado C A and Wiedemann U A 2004 *Phys. Rev. D* **69** 114003
- [31] Arnold P B, Moore G D and Yaffe L G 2001 *J. High Energy Phys.* **JHEP11(2001)057**
Arnold P B, Moore G D and Yaffe L G 2002 *J. High Energy Phys.* **JHEP06(2002)030**
Arnold P B, Moore G D and Yaffe L G 2003 *J. High Energy Phys.* **JHEP01(2003)030**
- [32] Gyulassy M, Levai P and Vitev I 2001 *Nucl. Phys. B* **594** 371
Djordjevic M and Gyulassy M 2004 *Nucl. Phys. A* **733** 265–98
- [33] Wang X N and Guo X F 2001 *Nucl. Phys. A* **696** 788
Majumder A and Leeuwen M Van 2011 *Prog. Part. Nucl. Phys.* **A 66** 41
- [34] Blagojevic B and Djordjevic M 2015 *J. Phys. G: Nucl. Part. Phys.* **42** 075105
- [35] Noronha-Hostler J, Betz B, Noronha J and Gyulassy M 2016 *Phys. Rev. Lett.* **116** 252301

- [36] Das S K, Scardina F, Plumari S and Greco V 2015 *Phys. Lett. B* **747** 260
- [37] Xu J, Liao J and Gyulassy M 2015 *Chin. Phys. Lett.* **32** 092501
Shi S, Liao J and Gyulassy M 2018 *Chin. Phys. C* **42** 104104
- [38] Djordjevic M and Djordjevic M 2015 *Phys. Rev. C* **92** 024918
- [39] Gyulassy M, Levai P and Vitev I 2001 *Nucl. Phys. B* **594** 371
- [40] Xu J, Buzzatti A and Gyulassy M 2014 *J. High Energy Phys.* **JHEP08(2014)063**
- [41] Adam J *et al* (ALICE Collaboration) 2016 *Phys. Rev. Lett.* **116** 222302
- [42] Adam J *et al* (ALICE Collaboration) 2016 *Phys. Lett. B* **754** 235
Wilde M and (for the ALICE Collaboration) 2013 *Nucl. Phys. A* **904-905** 573c
- [43] Peshier A 2006 arxiv:[hep-ph/0601119](https://arxiv.org/abs/hep-ph/0601119)
- [44] Djordjevic M and Gyulassy M 2003 *Phys. Rev. C* **68** 034914
- [45] Yu Maezawa *et al* (WHOT-QCD Collaboration) 2010 *Phys. Rev. D* **81** 091501
- [46] Nakamura A, Saito T and Sakai S 2004 *Phys. Rev. D* **69** 014506
- [47] Djordjevic M 2016 *Phys. Lett. B* **763** 439
- [48] Christiansen P, Tywoniuk K and Vislavicius V 2014 *Phys. Rev. C* **89** 034912
- [49] Xu J, Buzzatti A and Gyulassy M 2014 *J. High Energy Phys.* **JHEP08(2014)063**
- [50] Abelev B B *et al* (ALICE Collaboration) 2014 *Phys. Rev. C* **90** 034904
- [51] Acharya S *et al* (ALICE Collaboration) 2018 *Phys. Rev. Lett.* **120** 102301
- [52] Acharya S *et al* (ALICE Collaboration) 2018 *J. High Energy Phys.* **JHEP11(2018)013**
- [53] Khachatryan V *et al* (CMS Collaboration) 2017 *J. High Energy Phys.* **JHEP04(2017)039**
- [54] Aaboud M *et al* [ATLAS Collaboration] 2019 *Phys. Lett. B* **790** 108–28
- [55] Acharya S *et al* (ALICE Collaboration) 2018 *J. High Energy Phys.* **JHEP07(2018)103**
- [56] Sirunyan A M *et al* (CMS Collaboration) 2018 *Phys. Lett. B* **776** 195
- [57] Aaboud M *et al* (ATLAS Collaboration) 2018 *Eur. Phys. J. C* **78** 997
- [58] Sirunyan A M *et al* (CMS Collaboration) 2017 *Phys. Rev. Lett.* **119** 152301
- [59] Sirunyan A M *et al* (CMS Collaboration) arXiv:[1810.03022](https://arxiv.org/abs/1810.03022)
- [60] Aaboud M *et al* (ATLAS Collaboration) 2018 *Eur. Phys. J. C* **78** 762
- [61] Khachatryan V *et al* (CMS Collaboration) 2017 *Eur. Phys. J. C* **77** 252
- [62] Sirunyan A M *et al* (CMS Collaboration) arXiv:[1810.11102](https://arxiv.org/abs/1810.11102)
- [63] Aaboud M *et al* (ATLAS Collaboration) 2018 *Eur. Phys. J. C* **78** 784
- [64] Jaelani S and (ALICE Collaboration) 2018 *J. High Energy Phys.* **JHEP10(2018)174**
- [65] Sirunyan A M *et al* (CMS Collaboration) 2018 *Phys. Lett. B* **782** 474
- [66] Sirunyan A M *et al* (CMS Collaboration) 2018 *Phys. Rev. Lett.* **120** 202301
- [67] Molnar D and Sun D 2014 *Nucl. Phys. A* **932** 140
Molnar D and Sun D 2013 *Nucl. Phys. A* **910-911** 486
- [68] Renk T 2012 *Phys. Rev. C* **85** 044903
- [69] Bjorken J D 1983 *Phys. Rev. D* **27** 140
- [70] Kolb P F and Heinz U W *Quark Gluon Plasma* ed R C Hwa *et al* (Singapore: World Scientific)
pp 634–714 [nucl-th/0305084]
- [71] Djordjevic M, Stojku S, Djordjevic M and Huovinen P arXiv:[1903.06829](https://arxiv.org/abs/1903.06829) [hep-ph]

How to test path-length dependence in energy-loss mechanisms: Analysis leading to a new observable

Magdalena Djordjevic,^{1,*} Dusan Zigic,¹ Marko Djordjevic,² and Jussi Auvinen¹

¹*Institute of Physics Belgrade, University of Belgrade, Serbia*

²*Faculty of Biology, University of Belgrade, Serbia*



(Received 17 November 2018; published 19 June 2019)

When traversing the QCD medium, high- p_{\perp} partons lose energy, which is typically measured by suppression, and predicted by various energy-loss models. A crucial test of different energy-loss mechanisms is their functional dependence on the length of traversed medium (i.e., path-length dependence). The upcoming experimental measurements will, for the first time, generate data that may allow to clearly assess this dependence, in particular, by comparing results from Pb + Pb collisions with future measurements in smaller systems. However, to perform such a test, it is crucial to choose an optimal observable. To address this, we here use both analytical and numerical analyses to propose a novel—simple, yet accurate and robust—observable for assessing the path-length dependence of the energy loss. Our numerical results show that, by using this observable, different (underlying) energy-loss mechanisms may be directly differentiated from the experimental data, which is, in turn, crucial for understanding the properties of the created QCD medium.

DOI: [10.1103/PhysRevC.99.061902](https://doi.org/10.1103/PhysRevC.99.061902)

Introduction. Understanding properties of quark-gluon plasma (QGP) [1] created at the Large Hadron Collider (LHC) and Relativistic Heavy Ion Collider (RHIC) experiments is a major goal of ultrarelativistic heavy ion physics [2], which would allow understanding properties of QCD matter at its most basic level. Energy loss of high- p_{\perp} partons traversing this medium is an excellent probe of its properties [3], which provided a crucial contribution [2] to establishing that QGP is created in these experiments. Comparing predictions of different energy-loss models [4], and, consequently, different underlying energy-loss mechanisms with experimental data is, therefore, crucial for understanding properties of created QGP. However, an open question is how to provide the most direct comparison of energy-loss predictions with experimental data.

The most basic signature for distinguishing different energy-loss models is how the predicted energy loss depends on the length of the traversed QCD medium (so-called path-length dependence). This path-length dependence directly relates to different underlying energy-loss mechanisms, such as perturbative QCD collisional (with typically linear [5,6]), radiative (with typically quadratic [7–11]), or alternatively conformal anti-De Sitter holography models (with third-power [12] energy-loss path-length dependence). Moreover, even in such cases, the division is not so clear as there are numerous other effects that can significantly alter these path-length dependencies [13–15]: inclusion of the mass of the leading particle, finite-size, and finite temperature effects in QGP, interference effects, etc. Therefore, accurately assessing the path-length dependence is also crucial for understanding

mechanisms that underly the observed energy loss, which is, in turn, necessary for investigating the properties of QCD matter created at RHIC and LHC, i.e., for precision QGP tomography.

However, despite its essential importance and longstanding interest in this subject, it is still not possible to directly infer the energy-loss path-length dependence from experimental measurements and, consequently, provide a possibility to discriminate between different energy-loss models. To our knowledge, the most comprehensive study in this subject [16,17], attempted to extract the energy-loss path-length dependence from a thorough simultaneous study of R_{AA} and v_2 predictions and data (at Au + Au collisions at RHIC and Pb + Pb collisions at the LHC) but was not able to constrain this dependence based on the existing observables and data. With this in mind, the goal of this Rapid Communication is to propose a novel approach for extracting the energy-loss path-length dependence.

It is intuitively clear that the most direct probe of the path-length dependence would involve comparing experimental data (and the related theoretical predictions) for two-collision systems of different sizes. Moreover, it would be optimal if the size would be the only property distinguishing these two systems, i.e., that other properties and parameters needed for generating relevant predictions would be the same between the two systems. Equally important, it is necessary to propose an appropriate observable from which the path-length dependence can be reliably extracted. Consequently, the aim of the analysis presented in this Rapid Communication is to infer an optimal system and an optimal observable for assessing the energy-loss path-length dependence. We will also test how reliable and robust is the inferred observable to different types of energy loss, probes, centralities, and collision systems.

* magda@ipb.ac.rs

Appropriate observable. In this section, we first start by asking what is an appropriate observable to assess the energy-loss path-length dependence? To start addressing this question, we note that such an observable should be sensitive to jet-medium interactions (so that energy-loss path-length dependence can be reliably extracted). On the other hand, it should not be sensitive to the medium evolution as the details of the medium evolution would, for such a purpose, present an unwanted background. Keeping this in mind, it is evident that such an observable should be a function of R_{AA} since R_{AA} has exactly these desired properties—i.e., it is highly sensitive to the energy-loss mechanisms in QGP [18,19], whereas being insensitive to the medium evolution (i.e., it can be characterized by mean QGP temperature) [19]. The medium evolution insensitivity is also consistent with our recent result [20] of almost identical R_{AA} for constant medium temperature and (1 + 1)-dimensional Bjorken expansion; however, this still remains to be further verified by using more realistic medium evolution calculations, including event-by-event fluctuations [17,21].

Appropriate systems. Measurements for 5.02-TeV Pb + Pb collisions are available, whereas precision measurements for 5.44-TeV smaller systems (Xe + Xe, Kr + Kr, Ar + Ar, and O + O) will become available in the future with the planned beam size scan (BSS) at the LHC. As these systems have similar collision energies but different sizes (atomic mass numbers are $A = 208, 129, 78, 40, 16$ for Pb, Xe, Kr, Ar, and O), comparison of Pb + Pb with smaller systems appears to be a good candidate for the path-length dependence study. Note that the BSS at the LHC is complementary to the current beam energy scan (BES) at RHIC, as in the BES, the systems of the same size but different collision energies are tested, whereas in the BSS, the systems of the same energy but different sizes will be explored, thus providing a crucial insight in how properties of the created matter depend on the size of the colliding ions.

Computational framework. In this Rapid Communication, the R_{AA} predictions will be generated by our full-fledged numerical procedure, recently developed in Ref. [22]. The procedure is based on our state-of-the-art dynamical energy-loss formalism [5,15], which contains different important effects (some of which are unique to this model): (i) *Finite-size finite temperature* QGP, consisting of *dynamical* (that is moving) constituents. This abolishes the widely used approximations of static scattering centers, vacuumlike propagators, and/or infinite-size QGP (e.g., Refs. [7,8,10,11]). (ii) Our calculations are based on the finite temperature generalized hard-thermal-loop approach [23] in which the infrared divergencies are naturally regulated [15]. (iii) Both collisional [5] and radiative [15] energy losses are computed under *the same* theoretical framework, which is applicable to *both light and heavy flavor*. (iv) The model is generalized to the case of finite magnetic mass [24] and running coupling [25]; recently, we also applied first steps towards removing widely used soft-gluon approximation [26]. Moreover, in Ref. [18], we showed that all these ingredients are necessary for accurately explaining the high- p_{\perp} parton-medium interactions in QGP.

To generate the final medium modified distribution of high- p_{\perp} hadrons, the formalism was integrated into fully optimized numerical framework DREENA [22], which integrates the

initial- p_{\perp} distribution of leading partons [27], energy loss with multigluon [28], and path-length [29] fluctuations and fragmentation functions [30]. To generate R_{AA} predictions for Pb + Pb collisions, we use the set of parameters specified in Ref. [22], which correspond to standard literature values (not repeated here for brevity).

The dynamical energy-loss formalism was previously used to obtain a comprehensive set of R_{AA} predictions at the RHIC and LHC [22]; it shows wide agreement with the existing data [25], explaining puzzling data and generating nonintuitive predictions for future experiments [31,32] (some of which were already confirmed by subsequent data [33,34]). This then strongly indicates that our formalism can realistically describe high- p_{\perp} parton-medium interactions and that it provides a suitable framework for the goal that we want to achieve in this Rapid Communication.

Smaller systems. For R_{AA} predictions in smaller systems and their comparison with Pb + Pb collisions, one should note that R_{AA} depends on: (i) initial distribution of high- p_{\perp} partons, (ii) average temperature of the created QGP, and (iii) path-length distributions. Regarding initial distributions, we previously showed [31] that, when the collision energy is changed almost two times (from 2.76 to 5.02 TeV), the influence of the change in p_{\perp} distributions leads to only a small change (less than 10%) in the resulting suppression. Consequently, for the increase of less than 10% in the collision energy (from 5.02 to 5.44 TeV), the same high- p_{\perp} distributions can be assumed. The average temperature (\bar{T}) for each centrality region in 5.02-TeV Pb + Pb collisions is estimated according to Ref. [22]. Note that \bar{T} is directly proportional to the charged multiplicity, whereas inversely proportional to the overlap area and average size of the medium, i.e., $\bar{T} = \left(\frac{dN_{ch}/d\eta}{A_{\perp}L}\right)^{1/3}$ [22,35]. To estimate \bar{T} in smaller systems, we note that, for each centrality region, all the above quantities change in the two-collision systems: $A_{\perp} \sim A^{2/3}$; $\bar{L} \sim A^{1/3}$ [36,37]; $dN_{ch}/d\eta \sim N_{part}$, where $N_{part} \sim A$ since, for the same collision energy, $\frac{dN_{ch}/d\eta}{N_{part}}$ should remain constant with decreasing the systems' size [38,39]. This therefore leads to $\bar{T} \sim \left(\frac{A}{A^{2/3}A^{1/3}}\right)^{1/3} \sim \text{const}$, i.e., we expect that, for a fixed centrality region, \bar{T} will remain unchanged when moving from large Pb + Pb to smaller systems. Finally, the path-length distributions for smaller systems, at different centralities, can be calculated in the same manner as previously for Pb + Pb [22]. It is straightforward to see that the two distributions are similar up to a rescaling factor corresponding to $A^{1/3}$. Consequently, we see that comparison of Pb + Pb with smaller systems is, in fact, close to ideal when it comes to probing the path-length dependencies.

R_{AA} ratio. The next question is what is the exact variable (i.e., its functional dependence on R_{AA}) that should be compared for the two systems in order to extract the path-length dependence. Since R_{AA} increases when the system size decreases, it may seem that the ratio of R_{AA} for the two systems is a natural choice [40]. To test this proposal, in Fig. 1, we show momentum dependence of the R_{AA} ratio for the Xe + Xe and Pb + Pb systems (note that, for easier reading, we will first concentrate on Xe + Xe and Pb + Pb and we will

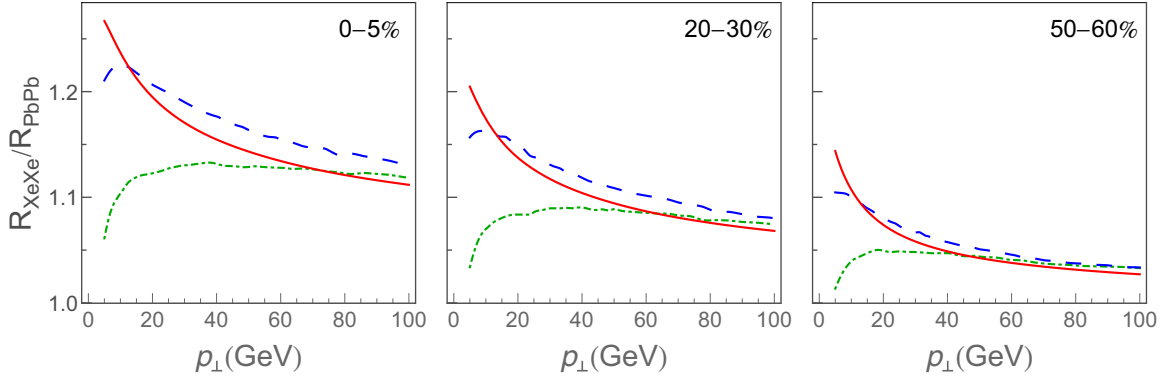


FIG. 1. Ratio of R_{XeXe} and R_{PbPb} is shown as a function of p_{\perp} for charged hadrons, D and B mesons (full, dashed, and dot-dashed curves, respectively). Centrality regions are denoted in the upper right corners of each panel.

discuss smaller systems subsequently). We see that it would be very hard to extract the path-length dependence from such a ratio, e.g., for high p_{\perp} , this ratio approaches 1, naively suggesting that the underlying model has no (or only weak) path-length dependence. However, the dynamical energy-loss model has, in fact, a strong (between linear and quadratic) path-length dependence. The same problem would emerge if experimental data would be plotted in that way, i.e., one may naively conclude that high- p_{\perp} suppression does not depend on the system size. Moreover, we see that this quantity is not robust with respect to the changes in collision centrality, which would further complicate extracting the path-length dependence from simple R_{AA} ratio.

The problem above can be intuitively understood by using scaling arguments. Fractional energy-loss $\Delta E/E$ can be estimated as [22]

$$\Delta E/E \approx \chi \bar{T}^a \bar{L}^b, \quad (1)$$

where a, b are proportionality factors, \bar{T} is the average temperature of the medium, \bar{L} is the average path-length traversed by the jet, and χ is a proportionality factor (which depends on initial jet p_{\perp}). $b \rightarrow 1$ corresponds to the linear, whereas $b \rightarrow 2$ corresponds to the quadratic [Landau-Pomeranchuk-Migdal- (LPM-) like] dependence of the energy loss.

If $\Delta E/E$ is small (i.e., for higher p_{\perp} of the initial jet, and for higher centralities), we can make the following estimate [22]:

$$R_{AA} \approx 1 - \xi \bar{T}^a \bar{L}^b, \quad (2)$$

where $\xi = (n-2)\chi/2$ and n is the steepness of the initial momentum distribution function.

The ratio of R_{XeXe} and R_{PbPb} then becomes

$$\frac{R_{XeXe}}{R_{PbPb}} \approx 1 + \xi \bar{T}^a \bar{L}^b \left[1 - \left(\frac{A_{Xe}}{A_{Pb}} \right)^{b/3} \right]. \quad (3)$$

This quantity is rather complicated, depending explicitly on the initial jet energy (through ξ), average medium temperature, and average size of the medium. Also, it explicitly depends on centrality (through \bar{T} and \bar{L}_{Pb} , which decrease with increasing centrality), consistently with what is seen in Fig. 1. Furthermore, as centrality and initial energy of the

jet increase, ξ , \bar{T} , and \bar{L}_{Pb} become smaller, explaining why the ratio in Fig. 1 goes to 1 for high p_{\perp} and high centrality, which results in the problem of concealing the path-length dependence. Consequently, the ratio of R_{AA} s for different collision systems is not a suitable observable for extracting path-length dependence.

Suitable observable. It is clear that such an observable should expose coefficient b in a simplest possible manner. To initially gauge the appropriate functional dependence, we again resort to the scaling arguments given above for which we have shown to provide a reasonable description of the full-fledged numerical model results in Fig. 1. We proceed by subtracting R_{AA} s [obtained from Eq. (2)] from 1, which, in the case of Xe and Pb, reduces to

$$R_L^{XePb} \equiv \frac{1 - R_{XeXe}}{1 - R_{PbPb}} \approx \frac{\xi \bar{T}^a \bar{L}_{Xe}^b}{\xi \bar{T}^a \bar{L}_{Pb}^b} \approx \left(\frac{A_{Xe}}{A_{Pb}} \right)^{b/3}. \quad (4)$$

This new quantity R_L^{XePb} has a very simple form, which depends only on the medium size (through A_{Xe}/A_{Pb}) and on the path-length dependence, i.e., coefficient b , which is now directly exposed. Note again that this simple dependence is expected to hold for *higher centralities and higher initial p_{\perp}* where Eqs. (2) and (4) are applicable. Consequently, as one plots R_L^{XePb} at higher centrality regions, one may expect that this value will approach a limit that directly reflects the path-length dependence, i.e., relation given by Eq. (4).

To numerically test our proposal and assess the applicability of the analytically derived scaling in Eq. (4), we further concentrate only on higher centrality regions and calculate $(1 - R_{XeXe})/(1 - R_{PbPb})$ using our full-fledged numerical procedure [22]. This ratio is shown in Fig. 2; full, dashed, and dot-dashed curves show our full results for charged hadrons, D and B mesons, respectively; the dashed lines correspond to the $b = 1$ and 2 limits from Eq. (4). From Fig. 2, one can see that R_L^{XePb} is almost independent of centrality, which is exactly what one needs for such observable. At high- $p_{\perp} \rightarrow 100$ GeV, we clearly see that R_L^{XePb} for all types of particles reaches a limiting value as expected. Moreover, this limiting value ($R_L^{XePb} \approx 0.8$) directly reflects the underlying path-length dependence, which is, in our case (the dynamical energy-loss formalism with radiative and collisional energy

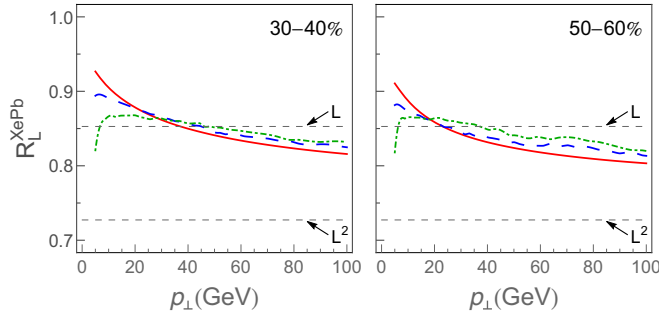


FIG. 2. Predictions for R_L^{XePb} as a function of p_\perp are shown for charged hadrons (full curves), D mesons (dashed curves), and B mesons (dot-dashed curves). Upper (lower) dashed gray lines correspond to the case in which energy-loss path-length dependence is linear (quadratic). Centrality regions are denoted in the upper right corners of each panel.

losses in a finite-size QCD medium), between linear and quadratic (i.e., $b \approx 1.4$), regardless of the particle flavor; note that this extracted path-length dependence is different from a common assumption of heavy flavor having linear, whereas light flavor having quadratic (LPM-like) dependence. It is, therefore, clear that, making such plots from experimental data and extracting the corresponding path-length dependence (exponent b), can be used to differentiate between different energy-loss models in a simple and direct manner. Also, note that, in distinction to Fig. 2, where the gray dashed lines are simple and intuitive (allowing straightforward inference of path-length dependence), defining such lines in Fig. 1 would not be possible.

Testing robustness and reliability. To address the robustness of the R_L^{AB} observable, i.e., if the observable is applicable to systems of diverse sizes, we further test R_L^{AB} on other smaller systems (Kr + Kr, Ar + Ar, and O + O). With this goal in mind, in Fig. 3, we concentrated on charged hadrons and generated full-fledged predictions for R_L^{AB} for Xe – Pb, Kr – Pb, Ar – Pb, and O – Pb as a function of p_\perp . From this figure, we first observe that, for all four systems, this observable is almost independent of centrality as expected from the arguments presented above. Second, we also observe that, independent of the collision system, this observable

shows the same behavior, so it is very robust with respect to extracting path-length dependence. We, moreover, observe that going to smaller systems makes extracting the path-length dependence even more straightforward since the separation between L and L^2 lines becomes larger when going to smaller systems, i.e., it increases for a factor of 2 when going from Xe – Pb to Ar – Pb and O – Pb. This then motivates using this observable across systems of different sizes and provides another argument for the utility of high- p_\perp measurements at the BSS at the LHC.

Finally, to address the reliability of this R_L^{AB} observable, in Fig. 3, we also show R_L^{AB} , calculated by using full numerical procedure stated above but if only collisional [5] (upper curves) or radiative [15] (lower curves) energy losses are taken into account—we here again concentrate on higher centrality regions where Eqs. (2) and (4) are applicable. Within the dynamical energy-loss model, collisional energy loss is close to—although somewhat less than—linear ($b \approx 0.9$) due to finite-size effects [5]. From Fig. 3, we see that this path-length dependence scenario is directly recovered where the approach to the appropriate dashed line (indicating $\lesssim L$ dependence) is almost ideal. For the radiative energy loss due to the LPM effect, path-length dependence approaches L^2 for higher p_\perp [15], and we see that, for such a scenario, R_L^{AB} also unambiguously recovers this tendency, although the spread of curves for different centralities is somewhat larger compared to the collisional energy-loss case. This, therefore, leads to the conclusion that, in addition to being simple and robust, R_L^{AB} is also an accurate observable for extracting path-length dependence.

Summary and outlook. Experimental measurements for smaller collision systems at the future BSS at the LHC will provide previously unprecedented opportunities to distinguish between different energy-loss mechanisms and, consequently, to better understand properties of created QGP. We here proposed a new—simple, robust, and reliable—observable for assessing the path-length dependence of the energy loss, which is a main signature of high- p_\perp parton-medium interactions. Based on our results, this observable can be used to straightforwardly extract the path-length dependence from experimental data, which can, consequently, be directly compared with such dependencies from various theoretical

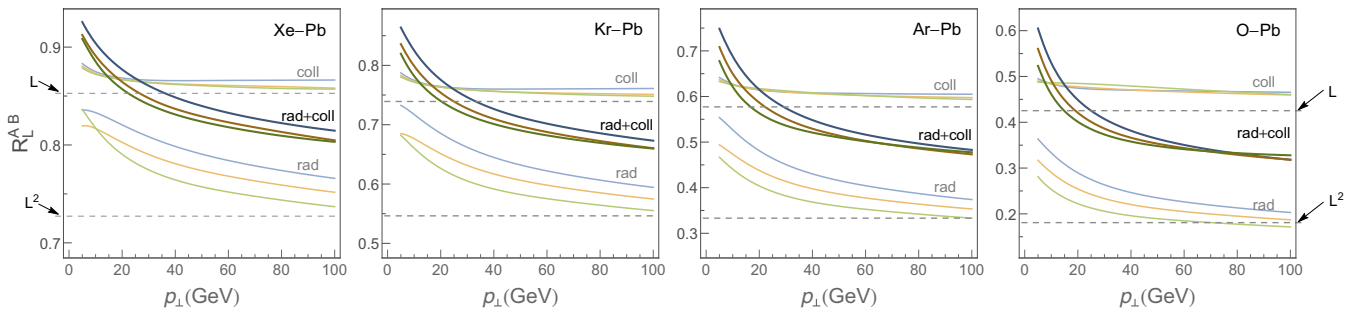


FIG. 3. Predictions for R_L^{AB} as a function of p_\perp are shown for charged hadrons where the darker sets of curves are obtained by using full dynamical energy loss, whereas upper and lower lighter sets or curves, respectively, correspond to the cases where only collisional or only radiative energy loss is considered. The first to fourth panels correspond to R_L^{XePb} , R_L^{KrPb} , R_L^{ArPb} , and R_L^{OPb} , respectively. In each panel, three centrality regions 30–40%, 40–50%, and 50–60% are marked by blue, orange, and green, respectively.

models as a major test of our understanding of energy-loss mechanisms.

Furthermore, our Rapid Communication also suggests that $(1 - R_{AA})$ might be a more suitable observable for the exploration of QGP than commonly used R_{AA} as we have here shown that it more directly reflects the underlying energy loss of the jet traversing the QGP. Furthermore, $(1 - R_{AA})$ observable appears to be highly correlated to v_2 (as noted in our recent study [41]). Since high- p_{\perp} observables are shown [41,42] to be sensitive to global QGP properties, we expect that including the full-medium evolution models (together with

event-by-event fluctuations) into the high- p_{\perp} predictions and providing a detailed joint study of high- p_{\perp} $(1 - R_{AA})$ and v_2 (and possibly higher harmonics) for different collision systems will prove to be an excellent tool for high-precision QGP tomography, which is a future major goal of relativistic heavy ion physics.

Acknowledgments. This work was supported by the European Research Council Grant No. ERC-2016-COG: 725741 and by the Ministry of Science and Technological Development of the Republic of Serbia under Projects No. ON171004 and No. ON173052.

-
- [1] J. C. Collins and M. J. Perry, *Phys. Rev. Lett.* **34**, 1353 (1975).
- [2] M. Gyulassy and L. McLerran, *Nucl. Phys. A* **750**, 30 (2005); E. V. Shuryak, *ibid.* **750**, 64 (2005); B. Jacak and P. Steinberg, *Phys. Today* **63**(5), 39 (2010).
- [3] J. D. Bjorken, FERMLAB-PUB-82-059-THY (1982).
- [4] JET Collaboration, K. M. Burke *et al.*, *Phys. Rev. C* **90**, 014909 (2014); Y. Xu *et al.*, *ibid.* **99**, 014902 (2019).
- [5] M. Djordjevic, *Phys. Rev. C* **74**, 064907 (2006).
- [6] M. H. Thoma and M. Gyulassy, *Nucl. Phys. B* **351**, 491 (1991); E. Braaten and M. H. Thoma, *Phys. Rev. D* **44**, R2625 (1991).
- [7] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, *Nucl. Phys. B* **483**, 291 (1997); B. G. Zakharov, *JETP Lett.* **63**, 952 (1996).
- [8] N. Armesto, C. A. Salgado, and U. A. Wiedemann, *Phys. Rev. D* **69**, 114003 (2004).
- [9] P. B. Arnold, G. D. Moore, and L. G. Yaffe, *J. High Energy Phys.* **11** (2001) 057; **01** (2003) 030.
- [10] M. Gyulassy, P. Levai, and I. Vitev, *Nucl. Phys. B* **594**, 371 (2001).
- [11] X. N. Wang and X. F. Guo, *Nucl. Phys. A* **696**, 788 (2001); A. Majumder and M. Van Leeuwen, *Prog. Part. Nucl. Phys. A* **66**, 41 (2011).
- [12] C. Marquet and T. Renk, *Phys. Lett. B* **685**, 270 (2010); F. Dominguez, C. Marquet, A. H. Mueller, B. Wu, and B. W. Xiao, *Nucl. Phys. A* **811**, 197 (2008).
- [13] C. A. G. Prado, J. Noronha-Hostler, R. Katz, A. A. P. Suaide, J. Noronha, and M. G. Munhoz, *Phys. Rev. C* **96**, 064903 (2017).
- [14] M. Nahrgang, J. Aichelin, P. B. Gossiaux, and K. Werner, *Phys. Rev. C* **93**, 044909 (2016).
- [15] M. Djordjevic, *Phys. Rev. C* **80**, 064909 (2009); M. Djordjevic and U. Heinz, *Phys. Rev. Lett.* **101**, 022302 (2008).
- [16] B. Betz and M. Gyulassy, *J. High Energy Phys.* **08** (2014) 090; **10** (2014) 043(E).
- [17] J. Noronha-Hostler, B. Betz, M. Gyulassy, M. Luzum, J. Noronha, I. Portillo, and C. Ratti, *Phys. Rev. C* **95**, 044901 (2017).
- [18] B. Blagojevic and M. Djordjevic, *J. Phys. G: Nucl. Part. Phys.* **42**, 075105 (2015).
- [19] T. Renk, *Phys. Rev. C* **85**, 044903 (2012); D. Molnar and D. Sun, *Nucl. Phys. A* **932**, 140 (2014); **910**, 486 (2013).
- [20] D. Zigic, I. Salom, M. Djordjevic, and M. Djordjevic, *Phys. Lett. B* **791**, 236 (2019).
- [21] T. Renk, H. Holopainen, J. Auvinen, and K. J. Eskola, *Phys. Rev. C* **85**, 044915 (2012).
- [22] D. Zigic, I. Salom, J. Auvinen, M. Djordjevic, and M. Djordjevic, [arXiv:1805.03494](https://arxiv.org/abs/1805.03494) [*J. Phys. G: Nucl. Part. Phys.* (to be published)].
- [23] J. I. Kapusta, *Finite-Temperature Field Theory* (Cambridge University Press, Cambridge, UK, 1989).
- [24] M. Djordjevic and M. Djordjevic, *Phys. Lett. B* **709**, 229 (2012).
- [25] M. Djordjevic and M. Djordjevic, *Phys. Lett. B* **734**, 286 (2014).
- [26] B. Blagojevic, M. Djordjevic, and M. Djordjevic, *Phys. Rev. C* **99**, 024901 (2019).
- [27] Z. B. Kang, I. Vitev, and H. Xing, *Phys. Lett. B* **718**, 482 (2012); R. Sharma, I. Vitev, and B. W. Zhang, *Phys. Rev. C* **80**, 054902 (2009).
- [28] M. Gyulassy, P. Levai, and I. Vitev, *Phys. Lett. B* **538**, 282 (2002).
- [29] A. Dainese, *Eur. Phys. J. C* **33**, 495 (2004).
- [30] D. de Florian, R. Sassot, and M. Stratmann, *Phys. Rev. D* **75**, 114010 (2007); M. Cacciari and P. Nason, *J. High Energy Phys.* **09** (2003) 006.
- [31] M. Djordjevic and M. Djordjevic, *Phys. Rev. C* **92**, 024918 (2015).
- [32] M. Djordjevic, *Phys. Lett. B* **763**, 439 (2016).
- [33] S. Acharya *et al.* (ALICE Collaboration), *J. High Energy Phys.* **11** (2018) 013.
- [34] V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **04** (2017) 039.
- [35] J. Xu, A. Buzzatti, and M. Gyulassy, *J. High Energy Phys.* **08** (2014) 063.
- [36] G. Giacalone, J. Noronha-Hostler, M. Luzum, and J. Y. Ollitrault, *Phys. Rev. C* **97**, 034904 (2018).
- [37] C. Loizides, J. Kamin, and D. d'Enterria, *Phys. Rev. C* **97**, 054910 (2018).
- [38] K. J. Eskola, H. Niemi, R. Paatelainen, and K. Tuominen, *Phys. Rev. C* **97**, 034911 (2018).
- [39] S. Acharya *et al.* (ALICE Collaboration), *Phys. Lett. B* **790**, 35 (2019).
- [40] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **10** (2018) 138.
- [41] M. Djordjevic, S. Stojku, M. Djordjevic, and P. Huovinen, [arXiv:1903.06829](https://arxiv.org/abs/1903.06829).
- [42] C. Andres, N. Armesto, H. Niemi, R. Paatelainen, and C. A. Salgado, [arXiv:1902.03231](https://arxiv.org/abs/1902.03231).



DREENA-B framework: First predictions of R_{AA} and v_2 within dynamical energy loss formalism in evolving QCD medium



Dusan Zigic^a, Igor Salom^a, Jussi Auvinen^a, Marko Djordjevic^b, Magdalena Djordjevic^{a,*}

^a Institute of Physics Belgrade, University of Belgrade, Serbia

^b Faculty of Biology, University of Belgrade, Serbia

ARTICLE INFO

Article history:

Received 16 July 2018

Received in revised form 10 December 2018

Accepted 15 February 2019

Available online 25 February 2019

Editor: W. Haxton

Keywords:

Relativistic heavy ion collisions

Quark-gluon plasma

LHC

Heavy flavor suppression

High p_T hadrons

ABSTRACT

Dynamical energy loss formalism allows generating state-of-the-art suppression predictions in finite size QCD medium, employing a sophisticated model of high- p_{\perp} parton interactions with QGP. We here report a major step of introducing medium evolution in the formalism through 1 + 1D Bjorken (“B”) expansion, while preserving all complex features of the original dynamical energy loss framework. We use this framework to provide joint R_{AA} and v_2 predictions, for the first time within the dynamical energy loss formalism in evolving QCD medium. The predictions are generated for a wide range of high p_{\perp} observables, i.e. for all types of probes (both light and heavy) and for all centrality regions in both $Pb + Pb$ and $Xe + Xe$ collisions at the LHC. Where experimental data are available, DREENA-B framework leads to a good joint agreement with v_2 and R_{AA} data. Such agreement is encouraging, i.e. may lead us closer to resolving v_2 puzzle (difficulty of previous models to jointly explain R_{AA} and v_2 data), though this still remains to be thoroughly tested by including state-of-the-art medium evolution within DREENA framework. While introducing medium evolution significantly changes v_2 predictions, R_{AA} predictions remain robust and moreover in a good agreement with the experimental data; R_{AA} observable is therefore suitable for calibrating parton-medium interaction model, independently from the medium evolution. Finally, for heavy flavor, we observe a strikingly similar signature of the dead-cone effect on both R_{AA} and v_2 - we also provide a simple analytical understanding behind this result. Overall, the results presented here indicate that DREENA framework is a reliable tool for QGP tomography.

© 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Funded by SCOAP³.

1. Introduction

It is by now established that quark-gluon plasma (QGP), being a new state of matter [1,2] consisting of interacting quarks, antiquarks and gluons, is created in ultra-relativistic heavy ion collisions at the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC). Energy loss of rare high p_{\perp} particles, which are created in such collisions and which transverse QGP, is considered to be an excellent probe of this form of matter [3–6]. Such energy loss is reflected through different observables, most importantly angular averaged (R_{AA}) [7–14] and angular differential (v_2) [15–22] nuclear modification factor, which can be measured and predicted for both light and heavy flavor probes. Therefore, comparing a comprehensive set of predictions, created under the same model and parameter set, with the corresponding experi-

mental data, allows for systematical investigation of QCD medium properties, i.e. QGP tomography.

We previously showed that the dynamical energy loss formalism [23–25] provides an excellent tool for such tomography. In particular, we demonstrated that the formalism shows a very good agreement [27–30] with a wide range of R_{AA} data, coming from different experiments, collision energies, probes and centralities. Recently, we also used this formalism to generate first v_2 predictions, within DREENA-C framework [26], where DREENA stands for Dynamical Radiative and Elastic ENergy loss Approach, and “C” denotes constant temperature QCD medium. These predictions were compared jointly with R_{AA} and v_2 data, showing a very good agreement with R_{AA} data, while visibly overestimating v_2 data. This overestimation also clearly differentiates the dynamical energy loss from other models, which systematically underestimated the v_2 data, leading to the so called v_2 puzzle [31–33]. On the other hand, it is also clear that v_2 predictions have to be further improved - in particular v_2 was shown to be sensitive to medium evolution, while in DREENA-C medium evolution

* Corresponding author.

E-mail address: magda@ipb.ac.rs (M. Djordjevic).

was introduced in the simplest form, through constant medium temperature. This problem then motivated us to introduce medium evolution in DREENA framework.

While several existing energy loss approaches already contain a sophisticated medium evolution, they employ simplified parton energy loss models. On the other hand, our dynamical energy loss formalism corresponds to the opposite “limit”, where constant (mean) medium temperature was assumed, combined with a sophisticated model of parton-medium interactions, which includes: i) QCD medium composed of dynamical (i.e. moving) scattering centers, which is contrary to the widely used static scattering centers approximation, ii) finite size QCD medium, iii) finite temperature QCD medium, modeled by generalized HTL approach [34, 35], naturally regularizing all infrared and ultraviolet divergencies [23–25,36]. iv) collisional [25] and radiative [23] energy losses, calculated within the same theoretical framework, v) finite parton mass, making the formalism applicable to both light and heavy flavor, vi) finite magnetic [37] mass and running coupling [27].

Note that we have previously showed that all the ingredients stated above are important for accurately describing experimental data [38]. Consequently, introducing medium evolution in the dynamical energy loss, is a major step in the model development, as all components in the model have to be preserved, and no additional simplifications should be used in the numerical procedure. In addition to developing the energy loss expressions with changing temperature, we also wanted to develop a framework that can efficiently generate a set of predictions for all types of probes and all centrality regions. That is, we think that for a model to be realistically compared with experimental data, the comparison should be done for a comprehensive set of light and heavy flavor experimental data, through the same numerical framework and the same parameter set. To implement this principle, we also had to develop a numerical framework that can efficiently (i.e. in a reasonably short time frame) generate such predictions, which will be presented in this paper.

We will start the task of introducing the medium evolution in the dynamical energy loss formalism with DREENA-B framework presented here, where “B” stands for Bjorken. In this framework, QCD medium is modeled by the ideal hydrodynamical $1 + 1D$ Bjorken expansion [39], which has a simple analytical form of temperature (T) dependence. This simple T dependence will be used as an intermediate between constant (mean) temperature DREENA-C framework and the full evolution QGP tomography tool. While, on one hand, inclusion of Bjorken expansion in DREENA framework is a major task (having in mind complexity of our model, see above), it on the other hand significantly simplifies the numerical procedure compared to full medium evolutions. This will then allow step-by-step development of full QGP tomography framework, and assessing improvements in the predictions when, within the same theoretical framework, one is transitioning towards more complex QGP evolution models within the dynamical energy loss framework.

2. Computational framework

To calculate the quenched spectra of hadrons, we use the generic pQCD convolution, while the assumptions are provided in [27]:

$$\frac{E_f d^3\sigma}{dp_f^3} = \frac{E_i d^3\sigma(Q)}{dp_i^3} \otimes P(E_i \rightarrow E_f) \otimes D(Q \rightarrow H_Q) \otimes f(H_Q \rightarrow e, J/\psi), \quad (1)$$

where “i” and “f”, respectively, correspond to “initial” and “final”, Q denotes quarks and gluons (partons). $E_i d^3\sigma(Q)/dp_i^3$ denotes

the initial parton spectrum, computed at next to leading order [40] for light and heavy partons. $D(Q \rightarrow H_Q)$ is the fragmentation function of parton Q to hadron H_Q ; for charged hadrons, D and B mesons we use DSS [41], BCFY [42] and KLP [43] fragmentation functions, respectively. $P(E_i \rightarrow E_f)$ is the energy loss probability, generalized to include both radiative and collisional energy loss in a realistic finite size dynamical QCD medium in which the temperature is changing, as well as running coupling, path-length and multi-gluon fluctuations. In below expressions, running coupling is introduced according to [27], where the temperature T now changes with proper time τ ; the temperature dependence along the jet path is taken according to the ideal hydrodynamical $1 + 1D$ Bjorken expansion [39]. Partons travel different paths in the QCD medium, which is taken into account through path length fluctuations [44]. Multi-gluon fluctuations take into account that the energy loss is a distribution, and are included according to [27,45] (for radiative energy loss) and [44,46] (for collisional energy loss).

The dynamical energy loss formalism was originally developed for constant temperature QCD medium, as described in detail in [23–25]. We have now derived collisional and radiative energy loss expressions for the medium in which the temperature is changing along the jet path; detailed calculations will be presented elsewhere, while the main results are summarized below.

For the collisional energy loss, we obtain the following analytical expression:

$$\begin{aligned} \frac{dE_{col}}{d\tau} &= \frac{2C_R}{\pi v^2} \alpha_S(E T) \alpha_S(\mu_E^2(T)) \\ &\int_0^\infty n_{eq}(|\vec{k}|, T) d|\vec{k}| \left(\int_0^{|\vec{k}|/(1+v)} d|\vec{q}| \int_{-v|\vec{q}|}^{v|\vec{q}|} \omega d\omega \right. \\ &\quad \left. + \int_{|\vec{k}|/(1+v)}^{|\vec{q}|_{max}} d|\vec{q}| \int_{|\vec{q}|-2|\vec{k}|}^{v|\vec{q}|} \omega d\omega \right) \\ &\left(|\Delta_L(q, T)|^2 \frac{(2|\vec{k}| + \omega)^2 - |\vec{q}|^2}{2} \right. \\ &\quad \left. + |\Delta_T(q, T)|^2 \frac{(|\vec{q}|^2 - \omega^2)((2|\vec{k}| + \omega)^2 + |\vec{q}|^2)}{4|\vec{q}|^4} (v^2|\vec{q}|^2 - \omega^2) \right). \end{aligned} \quad (2)$$

Here E is initial jet energy, τ is the proper time, T is the temperature of the medium, α_S is running coupling [27] and $C_R = \frac{4}{3}$. k is the 4-momentum of the incoming medium parton, v is the velocity of the incoming jet and $q = (\omega, \vec{q})$ is the 4-momentum of the gluon. $n_{eq}(|\vec{k}|, T) = \frac{N}{e^{|\vec{k}|/T} - 1} + \frac{N_f}{e^{|\vec{k}|/T} + 1}$ is the equilibrium momentum distribution [47] at temperature T including quarks and gluons (N and N_f are the number of colors and flavors, respectively). $\Delta_L(T)$ and $\Delta_T(T)$ are effective longitudinal and transverse gluon propagators [48]:

$$\Delta_L^{-1}(T) = \vec{q}^2 + \mu_E(T)^2 \left(1 + \frac{\omega}{2|\vec{q}|} \ln \left| \frac{\omega - |\vec{q}|}{\omega + |\vec{q}|} \right| \right), \quad (3)$$

$$\begin{aligned} \Delta_T^{-1}(T) &= \omega^2 - \vec{q}^2 - \frac{\mu_E(T)^2}{2} \\ &\quad - \frac{(\omega^2 - \vec{q}^2)\mu_E(T)^2}{2\vec{q}^2} \left(1 + \frac{\omega}{2|\vec{q}|} \ln \left| \frac{\omega - |\vec{q}|}{\omega + |\vec{q}|} \right| \right), \end{aligned} \quad (4)$$

while the electric screening (the Debye mass) $\mu_E(T)$ can be obtained by self-consistently solving the expression [49] (n_f is num-

ber of the effective degrees of freedom, Λ_{QCD} is perturbative QCD scale):

$$\frac{\mu_E(T)^2}{\Lambda_{QCD}^2} \ln \left(\frac{\mu_E(T)^2}{\Lambda_{QCD}^2} \right) = \frac{1 + n_f/6}{11 - 2/3 n_f} \left(\frac{4\pi T}{\Lambda_{QCD}} \right)^2. \quad (5)$$

The gluon radiation spectrum takes the following form:

$$\begin{aligned} \frac{dN_{\text{rad}}}{dx d\tau} = & \int \frac{d^2k}{\pi} \frac{d^2q}{\pi} \frac{2C_R C_2(G) T}{x} \frac{\alpha_s(ET) \alpha_s(\frac{k^2 + \chi(T)}{x})}{\pi} \\ & \times \frac{\mu_E(T)^2 - \mu_M(T)^2}{(\mathbf{q}^2 + \mu_M(T)^2)(\mathbf{q}^2 + \mu_E(T)^2)} \\ & \times \left(1 - \cos \frac{(\mathbf{k} + \mathbf{q})^2 + \chi(T)}{xE^+} \tau \right) \frac{(\mathbf{k} + \mathbf{q})}{(\mathbf{k} + \mathbf{q})^2 + \chi(T)} \\ & \times \left(\frac{(\mathbf{k} + \mathbf{q})}{(\mathbf{k} + \mathbf{q})^2 + \chi(T)} - \frac{\mathbf{k}}{\mathbf{k}^2 + \chi(T)} \right), \end{aligned} \quad (6)$$

where $C_2(G) = 3$ and $\mu_M(T)$ is magnetic screening. \mathbf{k} and \mathbf{q} are transverse momenta of radiated and exchanged (virtual) gluon, respectively. $\chi(T) \equiv M^2 x^2 + m_E(T)^2/2$, where x is the longitudinal momentum fraction of the jet carried away by the emitted gluon, M is the mass of the quark or gluon jet and $m_g(T) = \mu_E(T)/\sqrt{2}$ is effective gluon mass in finite temperature QCD medium [36]. We also recently abolished the soft-gluon approximation [50], for which we however showed that it does not significantly affect the model results; consequently, this improvement is not included in DREENA-B, but can be straightforwardly implemented in the future DREENA developments, if needed.

Note that, as a result of introducing medium evolution, the dynamical energy loss formalism now explicitly contains changing temperature in the energy loss expression. This is contrary to most of the other models, in which temperature evolution is introduced indirectly, through transport coefficient \hat{q} or gluon rapidity density $\frac{dN_g}{dy}$ (see [51] and references therein). This feature makes the dynamical energy loss a natural framework to incorporate diverse temperature profiles as a starting point for QGP tomography. As the first (major) step, we will below numerically implement this possibility through Bjorken 1 + 1D expansion [39].

Regarding the numerical procedure, computation efficiency of the algorithm implemented in DREENA-C framework [26] was already two orders of magnitude higher with respect to the basic (unoptimized) brute-force approach applied in [27]. However, straightforward adaptation of the DREENA-C code to the case of the Bjorken type evolving medium was not sufficient. This was dominantly due to additional integration over proper time τ , which increased the calculation time for more than two orders of magnitude. The computation of e.g. radiative energy losses alone, for a single probe, took around 10 hours on the available computer resources (a high performance workstation). Taking into account that it requires $\sim 10^3$ such runs to produce the results presented in this paper, it is evident that a substantial computational speedup was necessary.

The main algorithmic tool that we used to optimize the calculation was a combination of sampling and tabulating various intermediate computation values and their subsequent interpolation. We used nonuniform adaptive grids of the sampling points, denser in the parts of the parameter volume where the sampled function changed rapidly. Similarly, the parameters used for the numerical integration (the number of quasi-Monte Carlo sampling points and the required accuracy) were also suitably varied throughout the parameter space. Finally, while the computation in DREENA-C was performed in a software for symbolic computation, the new algorithm was redeveloped in C programming language.

The combined effect of all these improvements was a computational speedup of almost three orders of magnitude, which was a necessary prerequisite for both current practical applicability and future developments of DREENA framework.

Regarding the parameters, we implement Bjorken 1 + 1D expansion [39], with commonly used $\tau_0 = 0.6$ fm [52,53], and initial temperatures for different centralities calculated according to $T_0 \sim (dN_{ch}/dy/A_\perp)^{1/3}$ [54], where dN_{ch}/dy is charged multiplicity and A_\perp is overlap area (based on the Glauber model nuclear overlap function) for specific collision system and centrality. We use this equation, starting from $T_0 = 500$ MeV in 5.02 TeV $Pb + Pb$ most central collisions at the LHC, which is estimated based on average medium temperature of 348 MeV in these collisions, and QCD transition temperature of $T_c \approx 150$ MeV [55]. Note that the average medium temperature of 348 MeV in most central 5.02 TeV $Pb + Pb$ collisions comes from [28] the effective temperature (T_{eff}) of 304 MeV for 0-40% centrality 2.76 TeV $Pb + Pb$ collisions at the LHC [56] experiments (as extracted by ALICE). Once T_0 s for most central $Pb + Pb$ collisions are fixed, T_0 for both different centralities and different collision systems ($Xe + Xe$ and $Pb + Pb$) are obtained from the expression above.

Other parameters used in the calculation remain the same as in DREENA-C [26]. In particular, the path-length distributions for both $Xe + Xe$ and $Pb + Pb$ are calculated following the procedure described in [57], with an additional hard sphere restriction $r < R_A$ in the Woods-Saxon nuclear density distribution to regulate the path lengths in the peripheral collisions. Note that the path-length distributions for $Pb + Pb$ are explicitly provided in [26]; we have also checked that, for each centrality, our obtained eccentricities remain within the standard deviation of the corresponding Glauber Monte Carlo results [58]. For $Xe + Xe$, it is straightforward to show that $Xe + Xe$ and $Pb + Pb$ distributions are the same up to rescaling factor ($A^{1/3}$, where A is atomic mass number), as we discussed in [59]. Furthermore, the path-length distributions correspond to geometric quantity, and are therefore the same for all types of partons (light and heavy). For QGP, we take $\Lambda_{QCD} = 0.2$ GeV and $n_f = 3$. As noted above, temperature dependent Debye mass $\mu_E(T)$ is obtained from [49]. For light quarks and gluons, we, respectively, assume that their effective masses are $M \approx \mu_E(T)/\sqrt{6}$ and $m_g \approx \mu_E(T)/\sqrt{2}$ [36]. The charm and bottom masses are $M = 1.2$ GeV and $M = 4.75$ GeV, respectively. Magnetic to electric mass ratio is extracted from non-perturbative calculations [60,61], leading to $0.4 < \mu_M/\mu_E < 0.6$ - this range of screening masses leads to presented uncertainty in the predictions. We note that no fitting parameters are used in the calculations, that is, all the parameters correspond to standard literature values.

3. Results and discussion

In this section, we will present joint R_{AA} and v_2 predictions for light (charged hadrons) and heavy (D and B mesons) flavor in $Pb + Pb$ and $Xe + Xe$ collisions at the LHC, obtained by DREENA-B framework. Based on the path-length distributions from Figure 1 in [26], we will, in Figs. 1 to 2, show R_{AA} and v_2 predictions for light and heavy flavor, in 5.02 TeV $Pb + Pb$ and 5.44 TeV $Xe + Xe$ collisions, at different centralities. We start by presenting charged hadrons predictions, where R_{AA} data are available for both $Pb + Pb$ and $Xe + Xe$, while v_2 data exist for $Pb + Pb$ collisions. Comparison of our joint predictions with experimental data is shown in Fig. 1, where 1st and 2nd columns correspond, respectively, to R_{AA} and v_2 predictions at $Pb + Pb$, while 3rd and 4th columns present equivalent predictions/data for $Xe + Xe$ collisions at the LHC. From this figure, we see that DREENA-B is able to well explain joint R_{AA} and v_2 data. For 5.44 TeV $Xe + Xe$ collisions at the LHC, we observe good agreement of our predictions with prelim-

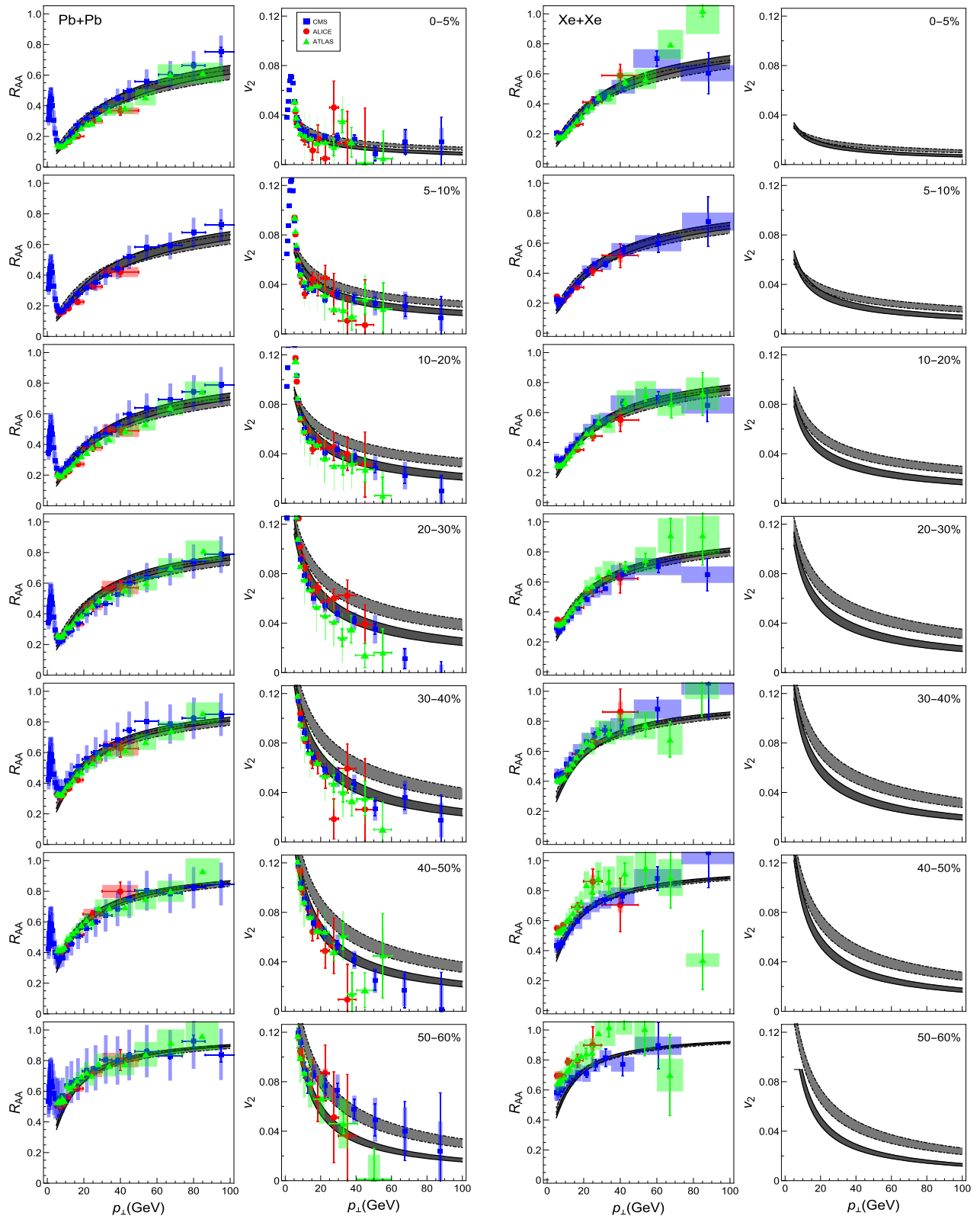


Fig. 1. First column: R_{AA} vs. p_{\perp} predictions are compared with 5.02 TeV $Pb+Pb$ ALICE [7], ATLAS [8] and CMS [9] h^{\pm} experimental data. Second column: Equivalent comparison for v_2 vs. p_{\perp} (data [15–17]). Third column: R_{AA} vs. p_{\perp} predictions are compared with 5.44 TeV $Xe+Xe$ ALICE [62], ATLAS [63] and CMS [64] preliminary data. Fourth column: Equivalent predictions for v_2 vs. p_{\perp} . ALICE, ATLAS and CMS data are respectively represented by red circles, green triangles and blue squares, while centrality regions are indicated in the relevant subfigures. Full and dashed curves correspond to, respectively, DREENA-B and DREENA-C frameworks. The gray band boundaries correspond to $\mu_M/\mu_E = 0.4$ and $\mu_M/\mu_E = 0.6$.

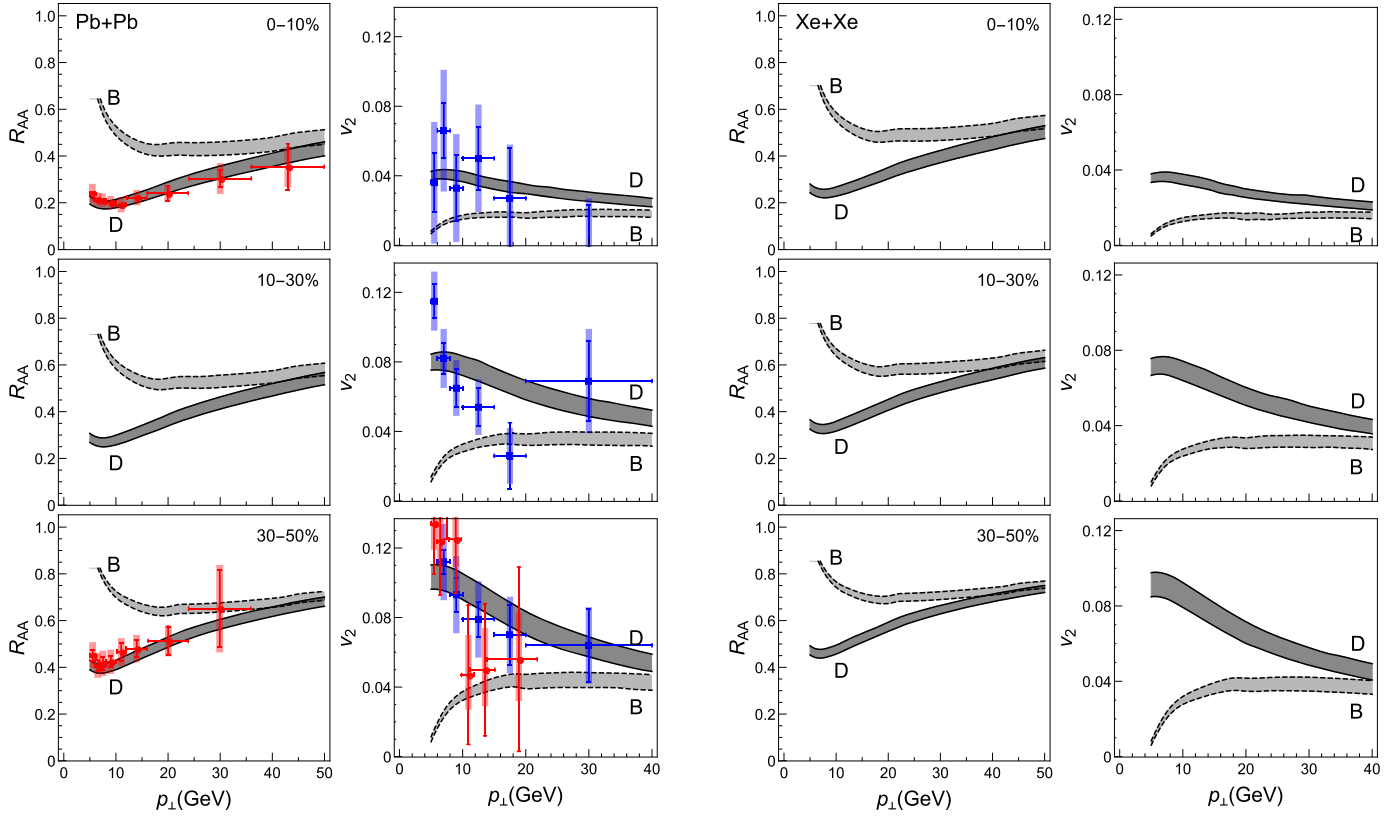


Fig. 2. First column: Theoretical predictions for D and B meson R_{AA} vs. p_{\perp} are compared with the available 5.02 TeV $Pb + Pb$ ALICE [10] (red circles) D meson experimental data. Second column: v_2 vs. p_{\perp} predictions are compared with 5.02 TeV $Pb + Pb$ ALICE [19] (red circles) and CMS [20] (blue squares) D meson experimental data. Third and fourth column: Heavy flavor R_{AA} and v_2 vs. p_{\perp} predictions are, respectively, provided for 5.44 TeV $Xe + Xe$ collisions at the LHC. First to third row, respectively, correspond to 0–10%, 10–30% and 30–50% centrality regions. The gray band boundaries correspond to $\mu_M/\mu_E = 0.4$ and $\mu_M/\mu_E = 0.6$.

inary R_{AA} data from ALICE, ATLAS and CMS data (where we note that these predictions were generated, and posted on arXiv, before the data became available), except for high centrality regions, where our predictions do not agree with ALICE (and also partially with ATLAS) data; however, note that in these regions ALICE, ATLAS and CMS data also do not agree with each other.

Furthermore, comparison of predictions obtained with DREENA-B and DREENA-C frameworks in Fig. 1, allows to directly assess the importance of inclusion of medium evolution on different observables, as the main difference between these two frameworks is that DREENA-B contains Bjorken evolution, while DREENA-C accounts for evolution in the simplest form (through constant mean temperature). We see that inclusion of Bjorken evolution has a negligible effect on R_{AA} , while having a significant effect on v_2 . That is, it keeps R_{AA} almost unchanged, while significantly decreasing v_2 . Consequently, small effect on R_{AA} , supports the fact that R_{AA} is weakly sensitive to medium evolution, making R_{AA} an excellent probe of jet-medium interactions in QGP; i.e. in QGP tomography, R_{AA} can be used to calibrate parton medium interaction models. On the other hand, medium evolution clearly influences v_2 predictions, in line with previous conclusions [65,66]; this sensitivity makes v_2 an ideal probe to constrain QGP medium parameters also from the point of high p_{\perp} measurements (in addition to constraining them from low p_{\perp} predictions and data).

In Fig. 2, we provide joint predictions for D and B meson R_{AA} (left panel) and v_2 (right panel) predictions for both 5.02 TeV $Pb + Pb$ and 5.44 TeV $Xe + Xe$ collisions at the LHC. Predictions are compared with the available experimental data. For D mesons, we again observe good joint agreement with the available R_{AA} and v_2 data. For B mesons (where the experimental data are yet to become available), we predict a notable suppression (see also [27,

67]), which is consistent with non-prompt J/Ψ R_{AA} measurements [68] (indirect probe of b quark suppression). Additionally, we predict non-zero v_2 for higher centrality regions. This does not necessarily mean that heavy B meson flows, since we here show predictions for high p_{\perp} , and flow is inherently connected with low p_{\perp} v_2 . On the other hand, high p_{\perp} v_2 is connected with the difference in the B meson suppression for different (in-plane and out-of-plane) directions, leading to our predictions of non zero v_2 for high p_{\perp} B mesons. Additionally, by comparing D and B meson v_2 s in Fig. 2, we observe that their difference is large and that it qualitatively exhibits the same dependence on p_{\perp} as R_{AA} . This v_2 comparison therefore presents additional important prediction of the heavy flavor dead-cone effect in QGP, where a strikingly similar signature of this effect is observed for R_{AA} and v_2 .

The predicted similarity between R_{AA} and v_2 dead-cone effects can be analytically understood by using simple scaling arguments. Fractional energy loss can be estimated as [26]

$$\Delta E/E \sim \eta T^a L^b, \quad (7)$$

where a, b are proportionality factors, T and L are, respectively, the average temperature of the medium and the average path-length traversed by the jet. η is a proportionality factor that depends on initial jet mass M and transverse momentum p_{\perp} .

Under the assumption of small fractional energy loss, we can make the following estimate [26]:

$$R_{AA} \approx 1 - \xi(M, p_{\perp}) T^a L^b, \\ v_2 \approx \xi(M, p_{\perp}) \frac{(T^a L^{b-1} \Delta L - T^{a-1} L^b \Delta T)}{2}, \quad (8)$$

where ΔL and ΔT are, respectively, changes in average path-lengths and average temperatures along out-of-plane and in-plane directions. $\xi = (n - 2)\eta/2$, where n is the steepness of the initial momentum distribution function.

The difference between R_{AA} and v_2 for D and B mesons then becomes:

$$R_{AA}^B - R_{AA}^D \approx (\xi(M_c, p_\perp) - \xi(M_b, p_\perp)) T^a L^b, \\ v_2^D - v_2^B \approx (\xi(M_c, p_\perp) - \xi(M_b, p_\perp)) \\ \times \frac{(T^a L^{b-1} \Delta L - T^{a-1} L^b \Delta T)}{2}, \quad (9)$$

where M_c and M_b are charm and bottom quark masses respectively. From Eq. (9), we see the same mass dependent prefactor for both R_{AA} and v_2 comparison, intuitively explaining our predicted dead-cone effect similarity for high- p_\perp R_{AA} and v_2 .

4. Summary

Overall, we see that comprehensive joint R_{AA} and v_2 predictions, obtained with our DREENA-B framework, lead to a good agreement with all available light and heavy flavor data. This is, to our knowledge, the first study to provide such comprehensive predictions for high p_\perp observables. In the context of v_2 puzzle, this study presents a significant development, as the other models were not able to achieve this agreement without introducing new phenomena [69]. However, for more definite conclusions, the inclusion of more complex QGP evolution within DREENA framework is needed, which is our main ongoing - but highly non-trivial - task, due to the complexity of underlying energy loss formalism.

As an outlook, for $Xe + Xe$, we also showed an extensive set of predictions for both R_{AA} and v_2 , for different flavors and centralities, to be compared with the upcoming experimental data. Reasonable agreement with these data would present a strong argument that the dynamical energy loss formalism can provide a reliable tool for precision QGP tomography. Moreover, such comparison between predictions and experimental data can also confirm interesting new patterns in suppression data, such as our prediction of strikingly similar signature of the dead-cone effect between R_{AA} and v_2 data.

Acknowledgements

We thank Bojana Blagojevic and Pasi Huovinen for useful discussions. We thank ALICE, ATLAS and CMS Collaborations for providing the shown data. This work is supported by the European Research Council, grant ERC-2016-COG: 725741, and by the Ministry of Science and Technological Development of the Republic of Serbia, under project numbers ON171004, ON173052 and ON171031.

References

- [1] J.C. Collins, M.J. Perry, *Phys. Rev. Lett.* **34** (1975) 1353.
- [2] G. Baym, S.A. Chin, *Phys. Lett. B* **62** (1976) 241.
- [3] M. Gyulassy, L. McLerran, *Nucl. Phys. A* **750** (2005) 30.
- [4] E.V. Shuryak, *Nucl. Phys. A* **750** (2005) 64; *E.V. Shuryak, Rev. Mod. Phys.* **89** (2017) 035001.
- [5] B. Jacak, P. Steinberg, *Phys. Today* **63** (2010) 39.
- [6] B. Muller, J. Schukraft, B. Wyslouch, *Annu. Rev. Nucl. Part. Sci.* **62** (2012) 361.
- [7] S. Acharya, et al., ALICE Collaboration, *J. High Energy Phys.* **1811** (2018) 013.
- [8] ATLAS Collaboration, ATLAS-CONF-2017-012.
- [9] V. Khachatryan, et al., CMS Collaboration, *J. High Energy Phys.* **1704** (2017) 039.
- [10] S. Jaelani, ALICE Collaboration, *Int. J. Mod. Phys. Conf. Ser.* **46** (2018) 1860018.

- [11] B. Abelev, et al., ALICE Collaboration, *Phys. Lett. B* **720** (2013) 52.
- [12] B. Abelev, et al., ALICE Collaboration, *J. High Energy Phys.* **1209** (2012) 112.
- [13] A. Adare, et al., PHENIX Collaboration, *Phys. Rev. Lett.* **101** (2008) 232301; A. Adare, et al., *Phys. Rev. C* **87** (3) (2013) 034911.
- [14] B.I. Abelev, et al., STAR Collaboration, *Phys. Lett. B* **655** (2007) 104.
- [15] S. Acharya, et al., ALICE Collaboration, *J. High Energy Phys.* **1807** (2018) 103.
- [16] M. Aaboud, et al., ATLAS Collaboration, *Eur. Phys. J. C* **78** (12) (2018) 997.
- [17] A.M. Sirunyan, et al., CMS Collaboration, *Phys. Lett. B* **776** (2018) 195.
- [18] B.B. Abelev, et al., ALICE Collaboration, *Phys. Rev. C* **90** (3) (2014) 034904.
- [19] S. Acharya, et al., ALICE Collaboration, *Phys. Rev. Lett.* **120** (10) (2018) 102301.
- [20] A.M. Sirunyan, et al., CMS Collaboration, *Phys. Rev. Lett.* **120** (20) (2018) 202301.
- [21] G. Aad, et al., ATLAS Collaboration, *Phys. Lett. B* **707** (2012) 330.
- [22] S. Chatrchyan, et al., CMS Collaboration, *Phys. Rev. Lett.* **109** (2012) 022301.
- [23] M. Djordjevic, *Phys. Rev. C* **80** (2009) 064909.
- [24] M. Djordjevic, U. Heinz, *Phys. Rev. Lett.* **101** (2008) 022302.
- [25] M. Djordjevic, *Phys. Rev. C* **74** (2006) 064907.
- [26] D. Zigic, I. Salom, J. Auvinen, M. Djordjevic, M. Djordjevic, arXiv:1805.03494 [nucl-th].
- [27] M. Djordjevic, M. Djordjevic, *Phys. Lett. B* **734** (2014) 286.
- [28] M. Djordjevic, M. Djordjevic, B. Blagojevic, *Phys. Lett. B* **737** (2014) 298.
- [29] M. Djordjevic, M. Djordjevic, *Phys. Rev. C* **92** (2015) 024918.
- [30] M. Djordjevic, *Phys. Rev. Lett.* **734** (2014) 286; M. Djordjevic, *Phys. Lett. B* **763** (2016) 439.
- [31] J. Noronha-Hostler, B. Betz, J. Noronha, M. Gyulassy, *Phys. Rev. Lett.* **116** (25) (2016) 252301.
- [32] B. Betz, M. Gyulassy, *J. High Energy Phys.* **1408** (2014) 090; B. Betz, M. Gyulassy, *J. High Energy Phys.* **1410** (2014) 043, Erratum.
- [33] D. Molnar, D. Sun, arXiv:1305.1046 [nucl-th].
- [34] J.I. Kapusta, *Finite-Temperature Field Theory*, Cambridge University Press, 1989.
- [35] M. Le Bellac, *Thermal Field Theory*, Cambridge University Press, 1996.
- [36] M. Djordjevic, M. Gyulassy, *Phys. Rev. C* **68** (2003) 034914.
- [37] M. Djordjevic, *Phys. Lett. B* **709** (2012) 229.
- [38] B. Blagojevic, M. Djordjevic, *J. Phys. G* **42** (2015) 075105.
- [39] J.D. Bjorken, *Phys. Rev. D* **27** (1983) 140.
- [40] Z.B. Kang, I. Vitev, H. Xing, *Phys. Lett. B* **718** (2012) 482; R. Sharma, I. Vitev, B.W. Zhang, *Phys. Rev. C* **80** (2009) 054902.
- [41] D. de Florian, R. Sassot, M. Stratmann, *Phys. Rev. D* **75** (2007) 114010.
- [42] M. Cacciari, P. Nason, *J. High Energy Phys.* **0309** (2003) 006; E. Braaten, K.-M. Cheung, S. Fleming, T.C. Yuan, *Phys. Rev. D* **51** (1995) 4819.
- [43] V.G. Kartvelishvili, A.K. Likhoded, V.A. Petrov, *Phys. Lett. B* **78** (1978) 615.
- [44] S. Wicks, W. Horowitz, M. Djordjevic, M. Gyulassy, *Nucl. Phys. A* **784** (2007) 426.
- [45] M. Gyulassy, P. Levai, I. Vitev, *Phys. Lett. B* **538** (2002) 282.
- [46] G.D. Moore, D. Teaney, *Phys. Rev. C* **71** (2005) 064904.
- [47] E. Braaten, M.H. Thoma, *Phys. Rev. D* **44** (1991) 1298.
- [48] A.V. Selikhov, M. Gyulassy, *Phys. Lett. B* **316** (1993) 373; A.V. Selikhov, M. Gyulassy, *Phys. Rev. C* **49** (1994) 1726.
- [49] A. Peshier, arXiv:hep-ph/0601119, 2006.
- [50] B. Blagojevic, M. Djordjevic, M. Djordjevic, *Phys. Rev. C* **99** (2) (2019) 024901.
- [51] K.M. Burke, et al., JET Collaboration, *Phys. Rev. C* **90** (1) (2014) 014909.
- [52] P.F. Kolb, U.W. Heinz, *Hydrodynamic description of ultrarelativistic heavy ion collisions*, in: R.C. Hwa, X.-N. Wang (Eds.), *Quark-Gluon Plasma 3*, World Scientific, Singapore, 2004, p. 634, arXiv:nucl-th/0305084.
- [53] J.E. Bernhard, J.S. Moreland, S.A. Bass, *Nucl. Phys. A* **967** (2017) 293.
- [54] M. Djordjevic, M. Gyulassy, R. Vogt, S. Wicks, *Phys. Lett. B* **632** (2006) 81.
- [55] A. Bazavov, et al., HotQCD Collaboration, *Phys. Rev. D* **90** (2014) 094503.
- [56] M. Wilde, ALICE Collaboration, *Nucl. Phys. A* **904–905** (2013) 573c.
- [57] A. Dainese, *Eur. Phys. J. C* **33** (2004) 495.
- [58] C. Loizides, J. Kamin, D. d'Enterria, *Phys. Rev. C* **97** (5) (2018) 054910.
- [59] M. Djordjevic, D. Zigic, M. Djordjevic, J. Auvinen, arXiv:1805.04030 [nucl-th].
- [60] Yu. Maizawa, et al., WHOT-QCD Collaboration, *Phys. Rev. D* **81** (2010) 091501.
- [61] A. Nakamura, T. Saito, S. Sakai, *Phys. Rev. D* **69** (2004) 014506.
- [62] S. Acharya, et al., ALICE Collaboration, *Phys. Lett. B* **788** (2019) 166.
- [63] ATLAS Collaboration, ATLAS-CONF-2018-007.
- [64] CMS Collaboration, CMS-PAS-HIN-18-004.
- [65] D. Molnar, D. Sun, *Nucl. Phys. A* **932** (2014) 140; D. Molnar, D. Sun, *Nucl. Phys. A* **910–911** (2013) 486.
- [66] T. Renk, *Phys. Rev. C* **85** (2012) 044903.
- [67] M. Djordjevic, B. Blagojevic, L. Zivkovic, *Phys. Rev. C* **94** (4) (2016) 044908.
- [68] J. Mihe, CMS Collaboration, *Nucl. Phys. A* **904–905** (2013) 657c.
- [69] J. Xu, J. Liao, M. Gyulassy, *Chin. Phys. Lett.* **32** (2015) 092501; S. Shi, J. Liao, M. Gyulassy, *Chin. Phys. C* **42** (10) (2018) 104104.



XXVIIth International Conference on Ultrarelativistic Nucleus-Nucleus Collisions
(Quark Matter 2018)

Dynamical energy loss formalism: from describing suppression patterns to implications for future experiments

Magdalena Djordjevic^a, Dusan Zigic^a, Bojana Blagojevic^a, Jussi Auvinen^a, Igor Salom^a and Marko Djordjevic^b

^a Institute of Physics Belgrade, University of Belgrade, Serbia

^b Faculty of Biology, University of Belgrade, Serbia

Abstract

We overview our recently developed DREENA-C and DREENA-B frameworks, where DREENA (Dynamical Radiative and Elastic ENergy loss Approach) is a computational implementation of the dynamical energy loss formalism; C stands for constant temperature and B for the medium evolution modeled by Bjorken expansion. At constant temperature our predictions overestimate v_2 , in contrast to other models, but consistent with simple analytical estimates. With Bjorken expansion, we obtain good agreement with both R_{AA} and v_2 measurements. We find that introducing medium evolution has a larger effect on v_2 predictions, but for precision predictions it has to be taken into account in R_{AA} predictions as well. We also propose a new observable, which we call *path length sensitive suppression ratio*, for which we argue that the path length dependence can be assessed in a straightforward manner. We also argue that $Pb + Pb$ vs. $Xe + Xe$ collisions make a good system to assess the path length dependence. As an outlook, we expect that introduction of more complex medium evolution (beyond Bjorken expansion) in the dynamical energy loss formalism can provide a basis for a state of the art QGP tomography tool – e.g. to jointly constrain the medium properties from the point of both high- p_{\perp} and low- p_{\perp} data.

Keywords: relativistic heavy ion collisions, quark-gluon plasma, energy loss, hard probes, heavy flavor

1. Introduction

Energy loss of high- p_{\perp} particles traversing QCD medium is considered to be an excellent probe of QGP properties [1, 2, 3]. The theoretical predictions can be generated and compared with a wide range of experimental data, coming from different experiments, collision systems, collision energies, centralities, observables. This comprehensive comparison of theoretical predictions and high p_{\perp} data, can then be used together with low p_{\perp} theory and data to study the properties of created QCD medium [4, 5, 6, 7], that is, for precision QGP tomography. However, to implement this idea, it is crucial to have a reliable high p_{\perp} parton energy loss model. With this goal in mind, during the past several years, we developed the dynamical energy loss formalism [8]. Contrary to the widely used approximation of static scattering centers, this model

takes into account that QGP consists of dynamical (moving) partons, and that the created medium has finite size. The calculations are based on the finite temperature field theory, and generalized HTL approach. The formalism takes into account both radiative and collisional energy losses, is applicable to both light and heavy flavor, and has been recently generalized to the case of finite magnetic mass and running coupling [9]. Most recently, we also relaxed the soft-gluon approximation within the model [15]. Finally, the formalism is integrated in an up-to-date numerical procedure [9], which contains parton production [10], fragmentation functions [11], path-length [12, 13] and multi-gluon fluctuations [14].

The model up-to-now explained a wide range of R_{AA} data [9, 16, 17, 18], with the same numerical procedure, the same parameter set, and with no fitting parameters, including explaining puzzling data and generating predictions for future experiments. This then strongly suggests that the model provides a realistic description of high p_{\perp} parton-medium interactions. However, the model did not take into account the medium evolution, so we used it to provide predictions only for those observables that are considered to be weakly sensitive to QGP evolution.

Therefore, our goal, which will be addressed in this proceedings, is to develop a framework which will allow systematic comparison of experimental data and theoretical predictions, obtained by the same formalism and the same parameter set. In particular, we want to develop a framework, which can systematically generate predictions for different observables (both R_{AA} and v_2), different collision systems ($Pb + Pb$ and $Xe + Xe$), different probes (light and heavy), different collision energies and different centralities [19, 20, 21]. Within this, our major goal is to introduce medium evolution in the dynamical energy loss formalism [20], where we start with 1+1D Bjorken expansion [22], and where our developments in this direction, will also be outlined in this proceedings. Finally, we also want to address an important question of how to differentiate between different energy loss models; in particular, what is the appropriate observable, and what are appropriate systems, to assess energy loss path-length dependence [21]. Note that only the main results are presented here; for a more detailed version, see [19, 20, 21], and references therein.

2. Results and discussion

As a first step towards the goals specified above, we developed DREENA-C framework [19], which is a fully optimized computational suppression procedure based on our dynamical energy loss formalism in constant temperature finite size QCD medium. Within this framework we, for the first time, generated joint R_{AA} and v_2 predictions based on our dynamical energy loss formalism. We generated predictions for both light and heavy flavor probes, and different centrality regions in $Pb + Pb$ collisions at the LHC (see [19] for more details). We obtained that, despite the fact that DREENA-C does not contain medium evolution (to which v_2 is largely sensitive), it leads to qualitatively good agreement with this data, though quantitatively, the predictions are visibly above the experimental data.

The theoretical models up-to-now, faced difficulties in jointly explaining R_{AA} and v_2 data, i.e. lead to underprediction of v_2 , unless new phenomena are introduced, which is known as v_2 puzzle [23]. Having this in mind, the overestimation of v_2 , obtained by DREENA-C, seems surprising. However, by using a simple scaling arguments, where fractional energy loss is proportional to T^a and L^b , and where, within our model a, b are close to 1, we straightforwardly obtain that in constant T medium, $R_{AA} \approx 1 - \xi TL$ and $v_2 \approx \frac{\xi T \Delta L}{2}$, while in evolving medium R_{AA} retains the same expressions and $v_2 \approx \frac{\xi T \Delta L - \xi \Delta T L}{2}$ (see [19] for more details, ξ is a proportionality factor that depends on initial jet p_{\perp}). So, it is our expectation that, within our model, the medium evolution will not significantly affect R_{AA} , while it will notably lower the v_2 predictions.

To check the reliability of these simple estimates, we developed DREENA-B framework [20], which is our most recent development within dynamical energy loss formalism. Here B stands for 1+1D Bjorken expansion [22], i.e. the medium evolution is introduced in dynamical energy loss formalism in a simple analytic way. We provided first joint R_{AA} and v_2 predictions with dynamical energy loss formalism in expanding QCD medium, which are presented in Fig. 1 (for charged hadrons), and we observe very good agreement with both R_{AA} and v_2 data. We equivalently obtained the same good agreement for D mesons, and predicted non-zero v_2 for high p_{\perp} B mesons.

In Fig. 2, we further present predictions for $Xe + Xe$ data [21], where we note that these predictions were generated before the data became available. In this figure (see also Fig. 1), we compare DREENA-C and

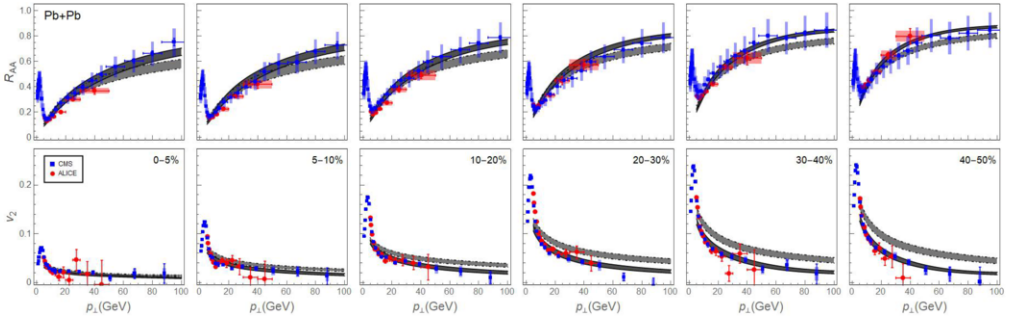


Fig. 1. **Joint R_{AA} and v_2 predictions for charged hadrons in 5.02 TeV $Pb + Pb$ collisions.** *Upper panels:* Predictions for R_{AA} vs. p_{\perp} are compared with ALICE [24] (red circles) and CMS [25] (blue squares) charged hadron experimental data in 5.02 TeV $Pb + Pb$ collisions. *Lower panels:* Predictions for v_2 vs. p_{\perp} are compared with ALICE [26] (red circles) and CMS [27] (blue squares) experimental data in 5.02 TeV $Pb + Pb$ collisions. Full and dashed curves correspond, respectively, to the predictions obtained with DREENA-B and DREENA-C frameworks. In each panel, the upper (lower) boundary of each gray band corresponds to $\mu_M/\mu_E = 0.6$ ($\mu_M/\mu_E = 0.4$). Columns 1-6 correspond, respectively, to 0 – 5%, 5 – 10%, 10 – 20%,..., 40 – 50% centrality regions. The figure is adapted from [19, 20] and the parameter set is specified there.

DREENA-B frameworks, to assess the importance of including medium evolution on R_{AA} and v_2 observables. We see that introduction of expanding medium affects both R_{AA} and v_2 data. That is, it systematically somewhat increase R_{AA} , while significantly decreasing v_2 ; this observation is in agreement with our estimate provided above. Consequently, we see that this effect has large influence on v_2 predictions, confirming previous arguments that v_2 observable is quite sensitive to medium evolution. On the other hand, this effect is rather small on R_{AA} , consistent with the notion that R_{AA} is not very sensitive to medium evolution [28, 29]. However, our observation from Figs. 1 and 2 is that medium evolution effect on R_{AA} , though not large, should still not be neglected in precise R_{AA} calculations, especially for high p_{\perp} and higher centralities.

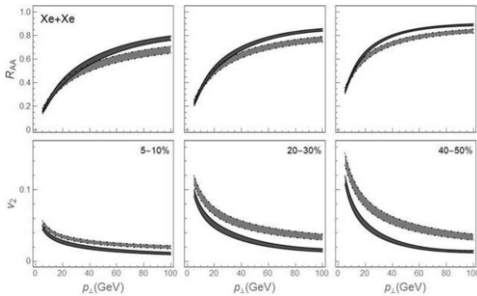


Fig. 2. **Joint R_{AA} and v_2 predictions for charged hadrons in 5.44 TeV $Xe + Xe$ collisions.** Predictions for R_{AA} vs. p_{\perp} and v_2 vs. p_{\perp} are shown on upper and lower panels, respectively. Columns 1-3, respectively, correspond to 5 – 10%, 20 – 30% and 40 – 50% centrality regions. Full and dashed curves correspond, respectively, to the predictions obtained with DREENA-B and DREENA-C frameworks. The figure is adapted from [20] and the parameter set is specified there.

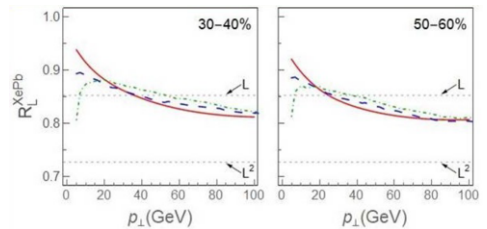


Fig. 3. **Path-length sensitive suppression ratio (R_L^{XePb}) for light and heavy probes.** Predictions for R_L^{XePb} vs. p_{\perp} is shown for charged hadrons (full), D mesons (dashed) and B mesons (dot-dashed). First and second column, respectively, correspond to 30 – 40% and 50 – 60% centrality regions. $\mu_M/\mu_E = 0.4$. The figure is adapted from [21] and the parameter set is specified there.

Finally, as the last topic of this proceedings, we address a question on how to differentiate between different energy loss models. With regard to this, note that path length dependence provides an excellent signature differentiating between different energy loss models, and consequently also between the underlying energy loss mechanisms. For example, some energy loss models have linear, some have quadratic, and our dynamical energy loss has the path-length dependence between linear and quadratic, which is due to both collisional and radiative energy loss mechanisms included in the model. To address this question, we

first have to answer what is an appropriate system for such a study. We argue that comparison of suppressions in $Pb + Pb$ and $Xe + Xe$ is an excellent way to study the path length dependence: From the suppression calculation perspective, almost all properties of these two systems are the same. That is, we show [21] that these two systems have very similar initial momentum distributions, average temperature for each centrality region and path length distributions (up to rescaling factor $A^{1/3}$). That is, the main property differentiating the two systems is its size, i.e. rescaling factor $A^{1/3}$, which therefore makes comparison of suppressions in $Pb + Pb$ and $Xe + Xe$ collisions an excellent way to study the path length dependence.

The second question is what is appropriate observable? With regards to that, the ratio of the two R_{AA} seems a natural choice, as has been proposed before. However, in this way the path length dependence cannot be naturally extracted, as shown in [21]. For example, this ratio approaches one for high p_{\perp} and high centralities, suggesting no path length dependence, while the dynamical energy loss has strong path length dependence. Also, the ratio has strong centrality dependence. That is, from this ratio, no useful information can be deduced. The reason for this is that this ratio includes a complicated relationship (see [21] for more details) which depends on the initial jet energy and centrality; so extracting the path-length dependence from this observable would not be possible.

However, based on the derivation presented in [21], we propose to use the ratio of $1-R_{AA}$ instead. From this estimate, we see that this ratio $R_L^{XePb} \equiv \frac{1-R_{XeXe}}{1-R_{PbPb}} \approx \left(\frac{A_{Xe}}{A_{Pb}}\right)^{b/3}$ has a simple dependence on only the size of the medium ($A^{1/3}$ ratio) and the path length dependence (exponent b). In Fig. 3 we plot this ratio, where we see that the path length dependence can be extracted from this ratio in a simple way, and moreover there is only a weak centrality dependence. Therefore, $1-R_{AA}$ ratio seems as a natural observable, which we propose to call *path-length sensitive suppression ratio*.

Acknowledgements: This work is supported by the European Research Council, grant ERC-2016-COG: 725741, and by the Ministry of Science and Technological Development of the Republic of Serbia, under project numbers ON171004, ON173052 and ON171031.

References

- [1] M. Gyulassy and L. McLerran, Nucl. Phys. A **750**, 30 (2005).
- [2] E. V. Shuryak, Nucl. Phys. A **750**, 64 (2005).
- [3] B. Jacak and P. Steinberg, Phys. Today **63**, 39 (2010).
- [4] K. M. Burke et al. [JET Collaboration], Phys. Rev. C **90**, no. 1, 014909 (2014).
- [5] G. Aarts et al., Eur. Phys. J. A **53**, no. 5, 93 (2017).
- [6] Y. Akiba et al., arXiv:1502.02730 [nucl-ex].
- [7] N. Brambilla et al., Eur. Phys. J. C **74**, no. 10, 2981 (2014).
- [8] M. Djordjevic, Phys. Rev. C **80**, 064909 (2009); M. Djordjevic and U. Heinz, Phys. Rev. Lett. **101**, 022302 (2008).
- [9] M. Djordjevic and M. Djordjevic, Phys. Lett. B **734**, 286 (2014).
- [10] Z. B. Kang, I. Vitev and H. Xing, Phys. Lett. B **718**, 482 (2012); R. Sharma, I. Vitev and B.W. Zhang, Phys. Rev. C **80**, 054902 (2009).
- [11] D. de Florian, R. Sassot and M. Stratmann, Phys. Rev. D **75**, 114010 (2007).
- [12] A. Dainese, Eur. Phys. J. C **33**, 495 (2004).
- [13] S. Wicks, W. Horowitz, M. Djordjevic and M. Gyulassy, Nucl. Phys. A **784**, 426 (2007).
- [14] M. Gyulassy, P. Levai and I. Vitev, Phys. Lett. B **538**, 282 (2002).
- [15] B. Blagojevic, M. Djordjevic and M. Djordjevic, arXiv:1804.07593 [nucl-th].
- [16] M. Djordjevic, B. Blagojevic and L. Zivkovic, Phys. Rev. C **94**, no. 4, 044908 (2016).
- [17] M. Djordjevic and M. Djordjevic, Phys. Rev. C **92**, 024918 (2015).
- [18] M. Djordjevic, Phys. Rev. Lett. **734**, 286 (2014); Phys. Lett. B **763**, 439 (2016).
- [19] D. Zigic, I. Salom, J. Auvinen, M. Djordjevic and M. Djordjevic, arXiv:1805.03494 [nucl-th].
- [20] D. Zigic, I. Salom, M. Djordjevic and M. Djordjevic, arXiv:1805.04786 [nucl-th].
- [21] M. Djordjevic, D. Zigic, M. Djordjevic and J. Auvinen, arXiv:1805.04030 [nucl-th].
- [22] J. D. Bjorken, Phys. Rev. D **27**, 140 (1983).
- [23] J. Noronha-Hostler, B. Betz, J. Noronha and M. Gyulassy, Phys. Rev. Lett. **116**, no. 25, 252301 (2016).
- [24] S. Acharya et al. [ALICE Collaboration], arXiv:1802.09145 [nucl-ex].
- [25] V. Khachatryan et al. [CMS Collaboration], JHEP **1704**, 039 (2017).
- [26] S. Acharya et al. [ALICE Collaboration], arXiv:1804.02944 [nucl-ex].
- [27] A. M. Sirunyan et al. [CMS Collaboration], Phys. Lett. B **776**, 195 (2018).
- [28] T. Renk, Phys. Rev. C **85** 044903 (2012).
- [29] D. Molnar and D. Sun, Nucl. Phys. A **932** 140 (2014); Nucl. Phys. A **910-911** 486 (2013).

Exploring the initial stages in heavy-ion collisions with high- p_{\perp} theory and data

Dusan Zigic,¹ Bojana Ilic ¹, Marko Djordjevic ² and Magdalena Djordjevic ^{1,*}

¹*Institute of Physics Belgrade, University of Belgrade, 11080 Belgrade, Serbia*

²*Faculty of Biology, University of Belgrade, 11000 Belgrade, Serbia*



(Received 16 October 2019; accepted 27 April 2020; published 25 June 2020)

Traditionally, the low- p_{\perp} sector is used to infer the features of initial stages before quark-gluon plasma thermalization. On the other hand, a recently acquired wealth of high- p_{\perp} experimental data paves the way to utilize the high- p_{\perp} particles' energy loss in exploring the initial stages. We here study how four different commonly considered initial-stage scenarios—which have the same temperature profile after thermalization, but differ in the “temperature” profile before thermalization—affect predictions of high- p_{\perp} R_{AA} and v_2 observables. Contrary to common expectations, we obtain that high- p_{\perp} v_2 is insensitive to the initial stages of medium evolution, being unable to discriminate between different conditions. On the other hand, R_{AA} is sensitive to these conditions; however, within the current error bars, the sensitivity is not sufficient to distinguish between different initial stages. Moreover, we also reconsider the validity of the widely used procedure of fitting the energy loss parameters, individually for different initial-stage cases, to reproduce the experimentally observed R_{AA} . We here find that previously reported sensitivity of v_2 to different initial states is mainly a consequence of the R_{AA} fitting procedure, which may lead to incorrect conclusions. On the other hand, if a global property, in particular the same average temperature, is imposed to tested temperature profiles, high sensitivity of high- p_{\perp} v_2 is again obtained. We show, however, that this sensitivity would be a consequence of differences not in initial stages but rather in final stages. Consequently, the simultaneous study of high- p_{\perp} R_{AA} and v_2 , with consistent energy loss parametrization and stringently controlled temperature profiles, is necessary to assess sensitivity of different variables to differences in initial and final stages.

DOI: [10.1103/PhysRevC.101.064909](https://doi.org/10.1103/PhysRevC.101.064909)

I. INTRODUCTION

It is by now firmly confirmed that a new state of matter—the quark-gluon plasma (QGP) [1,2], in which quarks, anti-quarks and gluons are deconfined—is formed at the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC). Rare high transverse momentum (high- p_{\perp}) particles, which are created immediately upon collision, are sensitive to all stages of QGP evolution, and are considered to be excellent probes [3–6] of this extreme form of matter. As these probes traverse the QGP, they lose energy, which is commonly assessed through high- p_{\perp} angular averaged (R_{AA}) [7–14] and high- p_{\perp} angular differential (v_2) [15–19] nuclear modification factors.

Commonly, high- p_{\perp} particles are used to study the nature of jet-medium interactions, while low- p_{\perp} particles are used to infer the bulk QGP properties. Accordingly, the scarce knowledge of the features of initial stages before QGP thermalization ($\tau < \tau_0$) was mostly inferred by utilizing data from the low- p_{\perp} sector [20–22] ($p_{\perp} \lesssim 5$ GeV). However, since

high- p_{\perp} partons effectively probe QGP properties, which in turn depend on initial stages, the idea of utilizing high- p_{\perp} theory and data in exploring the initial stages emerged. This idea acquired an additional boost, since a wealth of precision high- p_{\perp} R_{AA} [7–12] and v_2 [15–19] data have recently become available. Thus, the main goal of this paper is to assess to what extent and through what observables the initial stages of QGP evolution can be restrained by exploiting the energy loss of high- p_{\perp} particles in the evolving medium.

While clarifying these issues is clearly intriguing, the results of current theoretical studies on this subject are either inconclusive or questionable [23–25] as, e.g., the energy loss parameters are fitted to reproduce the experimentally observed high- p_{\perp} R_{AA} data individually for different analyzed initial stages. The energy loss parametrization should, however, clearly be a property of high- p_{\perp} parton interactions with the medium, rather than of individual temperature profiles. Consequently, to more rigorously study this issue, one needs a high control on both the energy loss and the analyzed temperature (T) profiles. To achieve this, we here use our state-of-the-art dynamical energy loss formalism, embedded in Bjorken one-dimensional (1D) medium evolution [26] (DREENA-B framework). While Bjorken 1D medium evolution is not a full bulk QGP evolution model, for this particular study it has a major advantage, as it allows one to analytically introduce different evolutions before thermalization, with the same evolution after thermalization, which therefore allows one to clearly isolate only the effects of different initial stages (which would not be possible with full hydrodynamics models).

*Corresponding author: magda@ipb.ac.rs

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

Consequently, we will here consider the effects on high- p_{\perp} R_{AA} and v_2 predictions of four common initial-stage cases [23], which have the same T profiles after thermalization, but differ in T profiles before the thermalization.

Furthermore, we recently demonstrated that the DREENA-B framework [27] is able to accurately reproduce both high- p_{\perp} R_{AA} and v_2 data for diverse colliding systems and energies (Pb + Pb at 2.76 and 5.02 TeV and Xe + Xe at 5.44 TeV), for both light and heavy flavors (h^{\pm} , B , D) and all available centralities, without introducing new phenomena [28,29]. This is distinct from many other formalisms, which employ more advanced medium evolution models, but contain simplified energy loss models, which have a tendency to underestimate v_2 relative to the experimental data; this is widely known as the v_2 puzzle [30,31]. Moreover, we recently obtained that going from 1D Bjorken to full 3+1D hydrodynamics evolution [32] does not significantly change the agreement between our predictions and experimental data, strongly suggesting that, for high- p_{\perp} data, accurate energy loss description is more important than the medium evolution. Consequently, for this study, using 1D Bjorken evolution has a major advantage of tight control over the temperature profiles used to mimic different initial states, while, at the same time, providing a reasonably realistic description of the data within our model.

The paper is organized as follows. In Sec. II, theoretical and computational frameworks are outlined. In Sec. III, we first assess the sensitivity of R_{AA} and v_2 to the aforementioned initial stages. We then adopt the approach of fitting initial temperature (T_0) to reproduce the same R_{AA} in all cases, and then assess the effect of thus obtained “modified” temperature profiles on R_{AA} and v_2 . We finally reexamine the validity of the widely used procedure [23–25] of fitting the energy loss parameters for different initial-stage cases to reproduce the same R_{AA} . For all these studies, we analytically pinpoint the origin of the obtained results. Our conclusions are presented in Sec. IV.

II. THEORETICAL AND COMPUTATIONAL FRAMEWORKS

To obtain the medium modified distribution of high- p_{\perp} light and heavy flavor particles, the generic pQCD convolution

formula [33,34] is utilized:

$$\frac{E_f d^3\sigma}{dp_f^3} = \frac{E_i d^3\sigma(Q)}{dp_i^3} \otimes P(E_i \rightarrow E_f) \otimes D(Q \rightarrow H_Q), \quad (1)$$

where indexes f and i refer to the final hadron (H_Q) and initial parton (Q), respectively. $\frac{E_i d^3\sigma(Q)}{dp_i^3}$ denotes the parton initial momentum distribution, calculated according to [35]. $P(E_i \rightarrow E_f)$ represents the energy loss probability based on our dynamical energy loss formalism (see below). $D(Q \rightarrow H_Q)$ stands for the fragmentation function of a parton into a hadron (H_Q), where for the light hadrons and D and B mesons we apply fragmentation functions of de Florian, Sassot, and Stratmann (DSS) [36], Braaten, Cheung, Fleming, and Yuan (BCFY) [37], and Kartvelishvili, Likhoded, and Petrov (KLP) [38], respectively.

The dynamical energy loss formalism [39–41] includes several unique features in modeling jet-medium interactions, whereby we previously showed [42] that all the ingredients are important for accurately describing experimental data: (1) The finite size QCD medium, that consists of dynamical (moving) as opposed to static scattering centers, which allows longitudinal momentum exchange with the medium constituents. (2) The calculations are done within the finite temperature generalized hard-thermal-loop approach [43], so that infrared divergences are naturally regulated in a highly nontrivial manner, contrary to many models which apply tree-level (vacuumlike) propagators [44–47]. (3) Both radiative [40,41] and collisional [39] contributions are calculated within the same theoretical framework. (4) The generalizations to a finite magnetic mass [48], running coupling [33], and beyond the soft-gluon approximation [49] are performed. In this paper, for the magnetic to electric mass ratio we assume the value $\mu_M/\mu_E = 0.5$, since various nonperturbative [50,51] approaches reported it to be in the range 0.4–0.6. (5) The energy loss probability comprises also multigluon [52] and path-length [34] fluctuations. The path-length fluctuations are calculated according to the procedure presented in [53], and are provided in Ref. [54].

As outlined in Ref. [27], the analytical expression for a single gluon radiation spectrum, in an evolving medium, reads

$$\begin{aligned} \frac{dN_{\text{rad}}}{dx d\tau} = & \frac{C_2(G)C_R}{\pi} \frac{1}{x} \int \frac{d^2\mathbf{q}}{\pi} \frac{d^2\mathbf{k}}{\pi} \frac{\mu_E^2(T) - \mu_M^2(T)}{[\mathbf{q}^2 + \mu_E^2(T)][\mathbf{q}^2 + \mu_M^2(T)]} T\alpha_s(ET)\alpha_s\left(\frac{\mathbf{k}^2 + \chi(T)}{x}\right) \\ & \times \left[1 - \cos\left(\frac{(\mathbf{k} + \mathbf{q})^2 + \chi(T)}{xE^+} \tau\right) \right] \frac{2(\mathbf{k} + \mathbf{q})}{(\mathbf{k} + \mathbf{q})^2 + \chi(T)} \left[\frac{\mathbf{k} + \mathbf{q}}{(\mathbf{k} + \mathbf{q})^2 + \chi(T)} - \frac{\mathbf{k}}{\mathbf{k}^2 + \chi(T)} \right], \end{aligned} \quad (2)$$

where \mathbf{k} and \mathbf{q} denote transverse momenta of radiated and exchanged gluons, respectively, $C_2(G) = 3$, $C_R = 4/3$ ($C_R = 3$) for the quark (gluon) jet, while $\mu_E(T)$ and $\mu_M(T)$ are electric (Debye) and magnetic screening masses, respectively. Temperature dependent Debye mass [55] is obtained by self-consistently solving Eq. (5) from Ref. [27]. α_s is the (temperature dependent) running coupling [56], E is the initial jet energy, while $\chi(T) = M^2 x^2 + m_g^2(T)$, where x is the longitudinal momentum fraction of the jet carried away by the emitted gluon, M is the mass of the quark ($M_{u,d,s} \approx \mu_E(T)/\sqrt{6}$, i.e., the thermal mass, whereas $M_c = 1.2$ GeV and $M_b = 4.75$ GeV) or gluon jet, and $m_g(T) = \mu_E(T)/\sqrt{2}$ [57] is the effective gluon mass in finite temperature QCD medium. Note that for all parameters we use standard literature values, i.e., we do not include additional fitting parameters when comparing our predictions with experimental data.

The analytical expression for the collisional energy loss per unit length in the evolving medium is given by [27]

$$\frac{dE_{\text{coll}}}{d\tau} = \frac{2C_R}{\pi v^2} \alpha_s(ET) \alpha_s(\mu_E^2(T)) \int_0^\infty n_{\text{eq}}(|\vec{k}|, T) d|\vec{k}| \left[\int_0^{|\vec{k}|/(1+v)} d|\vec{q}| \int_{-v|\vec{q}|}^{v|\vec{q}|} \omega d\omega + \int_{|\vec{k}|/(1+v)}^{|\vec{q}|_{\text{max}}} d|\vec{q}| \int_{|\vec{q}|-2|\vec{k}|}^{v|\vec{q}|} \omega d\omega \right] \times \left[|\Delta_L(q, T)|^2 \frac{(2|\vec{k}| + \omega)^2 - |\vec{q}|^2}{2} + |\Delta_T(q, T)|^2 \frac{(|\vec{q}|^2 - \omega^2)[(2|\vec{k}| + \omega)^2 + |\vec{q}|^2]}{4|\vec{q}|^4} (v^2|\vec{q}|^2 - \omega^2) \right], \quad (3)$$

where $n_{\text{eq}}(|\vec{k}|, T) = \frac{N}{e^{|\vec{k}|/T-1}} + \frac{N_f}{e^{|\vec{k}|/T+1}}$ is the equilibrium momentum distribution [58] comprising gluons, quarks, and antiquarks ($N = 3$ and $N_f = 3$ are the numbers of colors and flavors, respectively). k is the four-momentum of the incoming medium parton, v is velocity of the initial jet, and $q = (\omega, \vec{q})$ is the four-momentum of the exchanged gluon. $|\vec{q}|_{\text{max}}$ is provided in Ref. [39], while $\Delta_T(T)$ and $\Delta_L(T)$ are effective transverse and longitudinal gluon propagators given by Eqs. (3) and (4) in Ref. [27].

One of the assets of our energy loss formalism is the fact that energy loss explicitly depends on T , which makes it naturally suited for examining the QGP properties via implementation of various temperature profiles. In this paper, the temperature dependence on proper time (τ) is taken according to the ideal hydrodynamical 1D Bjorken expansion [26] $T(\tau) \sim \sqrt[3]{(\tau_0/\tau)}$, with thermalization time $\tau_0 = 0.6$ fm [59,60]. The initial QGP temperature T_0 for the chosen centrality bin is not a free parameter, i.e., it is constrained starting from the ALICE effective temperature [61] and following the numerical procedure outlined in Ref. [62]. In this paper, we will concentrate on mid-central 30–40% centrality region at 5.02 TeV Pb + Pb at the LHC, which corresponds to $T_0 = 391$ MeV [27]. However, we performed an extensive study on all centrality regions (as in [27]), and checked that the results and conclusions obtained here are the same irrespectively of the centrality region (results not shown for brevity). The QGP transition temperature is considered to be $T_C \approx 160$ MeV [63].

The DREENA-B framework is applied to generate predictions for two main high- p_\perp observables: R_{AA} and v_2 . R_{AA} is defined as the ratio of the quenched $A + A$ spectrum to the $p + p$ spectrum, scaled by the number of binary collisions N_{bin} :

$$R_{AA}(p_T) = \frac{dN_{AA}/dp_T}{N_{\text{bin}} dN_{pp}/dp_T}; \quad (4)$$

while for intuitive understanding of the underlying effects we also use [54]

$$R_{AA} \approx \frac{R_{AA}^{\text{in}} + R_{AA}^{\text{out}}}{2}, \quad (5)$$

where R_{AA}^{in} and R_{AA}^{out} denote in-plane and out-of-plane nuclear modification factors, respectively. The expression for the high- p_\perp elliptic flow is derived in [23] (see also [54,64,65]), under the assumption of negligible higher harmonics at high $p_\perp \gtrsim 10$ GeV, leading to

$$v_2 \approx \frac{1}{2} \frac{R_{AA}^{\text{in}} - R_{AA}^{\text{out}}}{R_{AA}^{\text{in}} + R_{AA}^{\text{out}}}. \quad (6)$$

The advantage of using Eq. (6) for high- p_\perp predictions is that it is computationally significantly less demanding than the commonly used v_2 expression (see, e.g., Eq. (1) from [15]). However, to explicitly verify its applicability, we checked that, for average temperature profiles, Eq. (6) will lead to the same result (up to less than 1% difference) as the commonly used azimuthally dependent expression.

We also note that the approach to experimentally infer v_2 (see, e.g., Eq. (16) in [15]) is different from the above-mentioned theoretical approaches. However, that approach could lead to different v_2 predictions only if event-by-event fluctuations are considered (which we do not do in this study). We also note that the importance of event-by-event fluctuations in adequately addressing high- p_\perp v_2 is currently an open question; i.e., in [30], it was proposed that event-by-event fluctuations may increase the high- p_\perp v_2 , while this was not supported by two subsequent independent studies [29,66].

III. RESULTS AND DISCUSSION

In the first part of this section we address how different initial stages (before the thermalization time τ_0) affect our predictions of high- p_\perp R_{AA} and v_2 . To this end, we consider the following four common cases of initial stages [23], which assume the same 1D Bjorken hydro temperature (T) profile [26] upon thermalization (for $\tau \geq \tau_0$), but have different T profiles before the thermalization (for $\tau < \tau_0$):

- $T = 0$, the so-called *free-streaming case*, which corresponds to neglecting interactions (i.e., energy loss) before the QGP thermalization.
- The *linear case*, corresponding to linearly increasing T with time from transition temperature ($T_C = 160$ MeV at $\tau_C = 0.25$ fm) to the initial temperature T_0 .
- The *constant case*, $T = T_0$.
- The *divergent case*, corresponding to 1D Bjorken expansion from $\tau = 0$.

These initial stages are depicted in Fig. 1, and it is clear that (a)–(d) case ordering corresponds to gradually increasing prethermal interactions. Note that we use this classification (a)–(d) consistently throughout the paper to denote initial stages (for $\tau < \tau_0$), as well as for the entire evolution. Also, note that in this part of the study we will include experimental data for comparison with our predictions. However, to allow better visualization of our obtained numerical results, in the other two parts of the study we will omit the comparison with the data, as the error bars are large and the data remain the same. Intuitively, one would expect that introducing these prethermal interactions would increase the energy loss

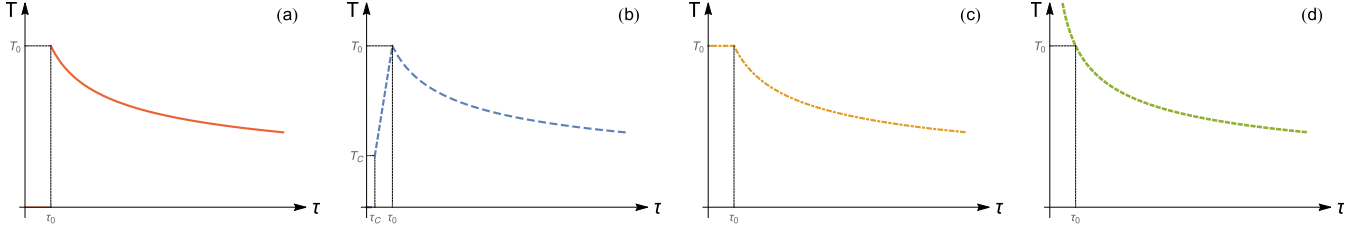


FIG. 1. Four temperature evolution profiles, which differ at the initial stages. At $\tau \geq \tau_0$, all profiles assume the same temperature dependence on the proper time (1D Bjorken [26]). At the initial stage, i.e., for $0 < \tau < \tau_0$, the temperature is considered to be (a) equal to zero; (b) increasing linearly from T_C to T_0 between τ_C and τ_0 , otherwise zero; (c) constant and equal to T_0 ; and (d) a continuous function of τ matching the dependence for $\tau \geq \tau_0$. Note that, in each panel, T_0 has the same value at τ_0 .

compared to the commonly considered free-streaming case, and consequently lead to smaller R_{AA} . In Fig. 2 we indeed observe that R_{AA} is sensitive to the initial stages for all types of particles. That is, as expected, we see that the suppression progressively increases from case (a) to case (d). However, these differences are not very large, and the current error bars at the LHC do not allow distinguishing between these scenarios, as can be seen in Fig. 2 (left).

In contrast to R_{AA} , the effect of initial stages on v_2 is intuitively less clear, as this observable nontrivially depends on the energy loss or R_{AA} 's (see Eq. (6)). From Fig. 3, we surprisingly infer that v_2 is insensitive to the presumed initial stage for all types of particles (in distinction to the results obtained in [24]), so that v_2 is unable to distinguish between different initial-stage scenarios.

To quantitatively understand this unexpected observation, in Fig. 4 we show transverse momentum dependence of R_{AA}^{in} , R_{AA}^{out} , and R_{AA} in $i = b, c, d$ cases relative to the baseline case (a) for charged hadrons. The conclusions for heavy particles are the same and therefore omitted. We distinguish three sets of curves, which correspond to the ratio of R_{AA} 's in linear (b), constant (c), and divergent (d) cases relative to the free-streaming (a) case. Note that the free-streaming case is used

as a baseline, as it corresponds to the most commonly used scenario, both in low- and high- p_\perp calculations.

Each set of curves in Fig. 4 contains three lines, representing proportionality functions $\gamma(p_\perp)$, which are defined as follows:

$$\gamma_{ia}^{\text{in}} = \frac{R_{AA,i}^{\text{in}}}{R_{AA,a}^{\text{in}}}, \quad \gamma_{ia}^{\text{out}} = \frac{R_{AA,i}^{\text{out}}}{R_{AA,a}^{\text{out}}}, \quad \gamma_{ia} = \frac{R_{AA,i}}{R_{AA,a}}, \quad (7)$$

where $i = b, c, d$ denotes the corresponding cases from Fig. 1. From Fig. 4 we see that for the same i (i.e., within the same set of curves (b), (c), or (d)) the proportionality functions $\gamma_{ia}(p_\perp)$ are practically identical for the relations involving in-plane, out-of-plane, and angular averaged R_{AA} 's:

$$\gamma_{ia}^{\text{in}} \approx \gamma_{ia}^{\text{out}} \approx \gamma_{ia}. \quad (8)$$

Note also that $\gamma_{ia} < 1$, while γ_{ia} s from distinct sets significantly differ from one another (i.e., for $i \neq j \rightarrow \gamma_{ia}(p_\perp) \neq \gamma_{ja}(p_\perp)$).

Consequently, by implementing Eq. (7) in Eq. (6) and acknowledging Eq. (8), we obtain

$$v_{2,i} \approx \frac{1}{2} \frac{\gamma_{ia}(R_{AA,a}^{\text{in}} - R_{AA,a}^{\text{out}})}{\gamma_{ia}(R_{AA,a}^{\text{in}} + R_{AA,a}^{\text{out}})} = v_{2,a}, \quad (9)$$

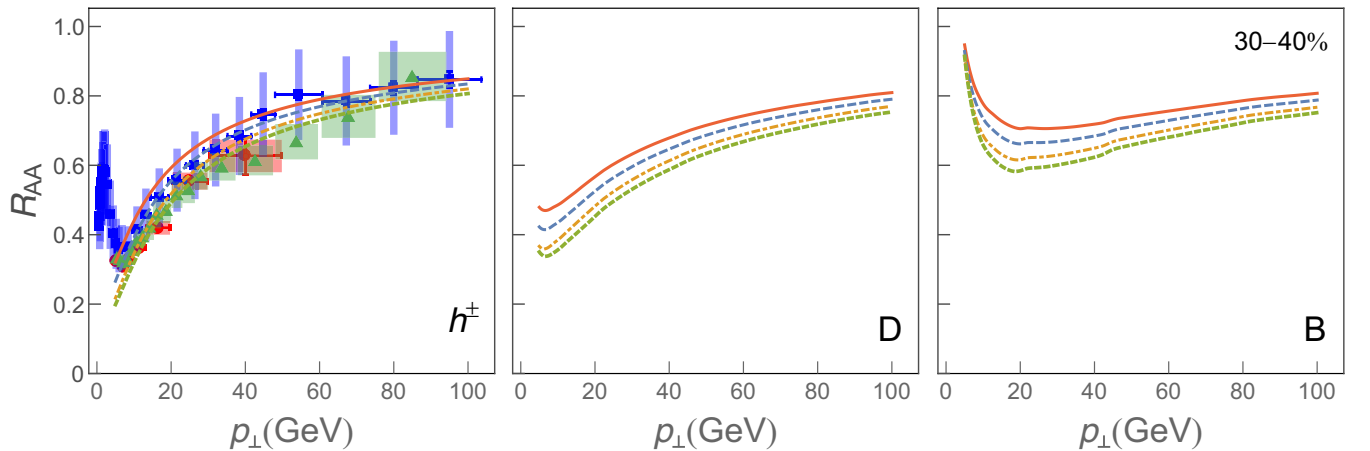


FIG. 2. R_{AA} dependence on p_\perp for four different initial stages depicted in Fig. 1 is shown for charged particles (left panel), D mesons (central panel) and B mesons (right panel). For charged hadrons, the predictions are compared with 5.02 TeV Pb + Pb ALICE [7] (red circles), ATLAS [8] (green triangles), and CMS [9] (blue squares) h^\pm R_{AA} experimental data. In each panel, temperature profiles from Fig. 1 are represented by a full red curve (case (a)), by a dashed blue curve (case (b)), by a dot-dashed orange curve (case (c)), and by dotted green curve (case (d)). The results correspond to the centrality bin 30–40%, and $\mu_M/\mu_E = 0.5$.

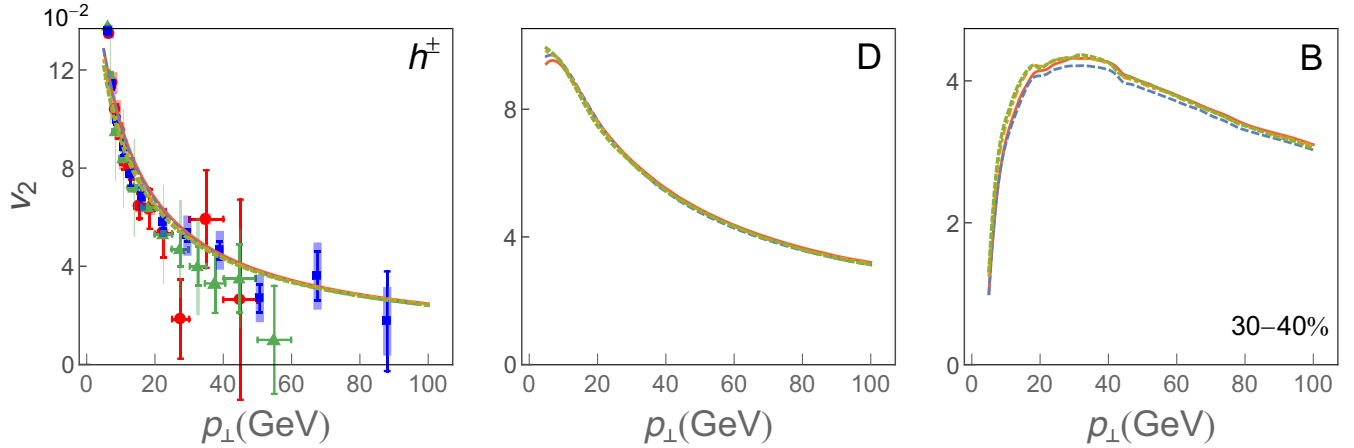


FIG. 3. v_2 dependence on p_\perp for four different initial stages depicted in Fig. 1. Left, central, and right panels correspond to charged hadrons, D mesons and B mesons, respectively. For charged hadrons, the predictions are compared with 30–40% centrality 5.02 TeV Pb + Pb ALICE [15] (red circles), ATLAS [16] (green triangles), and CMS [17] (blue squares) h^\pm v_2 experimental data. The labeling and remaining parameters are the same as in Fig. 2.

for any choice of $i = b, c, d$, as observed in Fig. 3. Therefore, we here showed that initial stages alone do not affect v_2 , i.e., they affect only R_{AA} . R_{AA} susceptibility to the initial stages is in qualitative agreement with papers [27,67,68], where R_{AA} is shown to be only sensitive to the averaged properties of the evolving medium, i.e., average temperature (\bar{T}). Since R_{AA} is proportional to \bar{T} , and since for all four initial-stage cases (a)–(d) the \bar{T} value is different ($\bar{T}_a < \bar{T}_b < \bar{T}_c < \bar{T}_d$), it is evident that R_{AA} will be different in these cases.

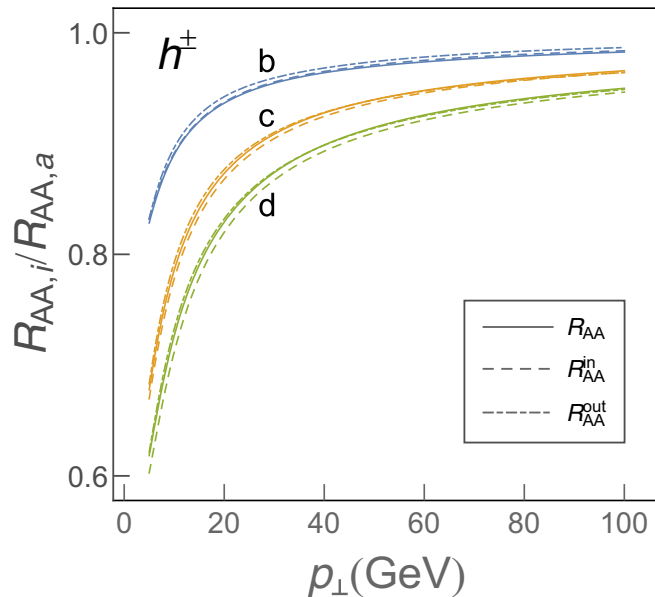


FIG. 4. Transverse momentum dependence of in-plane (dashed), out-of-plane (dot-dashed), and angular averaged (full curves) R_{AA} relative to the free-streaming (a) case for charged hadrons. Blue (upper), orange (middle), and green (lower) sets of curves correspond, respectively, to (b), (c) and (d) cases. The remaining parameters are the same as in Fig. 2.

The fact that R_{AA} depends on the average temperature of the medium, motivates us to further explore the case in which we modify the above temperature profiles to reproduce the same average temperature. This is equivalent to reevaluating the initial temperatures for different cases from Fig. 1, and, based on the reasoning above, it is evident that new initial temperatures should satisfy the following ordering: $T_{0,d'} < T_{0,c'} < T_{0,b'} < T_{0,a'}$. This leads to T profiles, which do not differ only at early times ($\tau < \tau_0$), but represent *different evolutions altogether*. These new evolutions, that are illustrated in Fig. 5 (which is a counterpart of Fig. 1 for the second

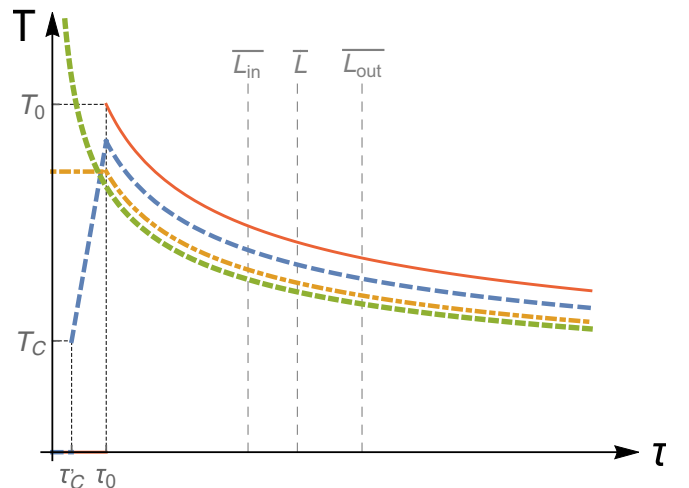


FIG. 5. Temperature dependence on the proper time in the setup with the same average temperatures. The labeling is the same as in Fig. 1, apart from the fact that initial temperatures (T_0 's) now differ in these four cases. As in Fig. 1, $T_C = 160$ MeV, $\tau_0 = 0.6$ fm, and $\tau_C' = 0.27$ fm. Vertical gray dashed lines correspond to average in-medium path length (\bar{L}), and to the path lengths along in-plane (\bar{L}_{in}) and out-of-plane (\bar{L}_{out}) directions, as labeled in the figure.

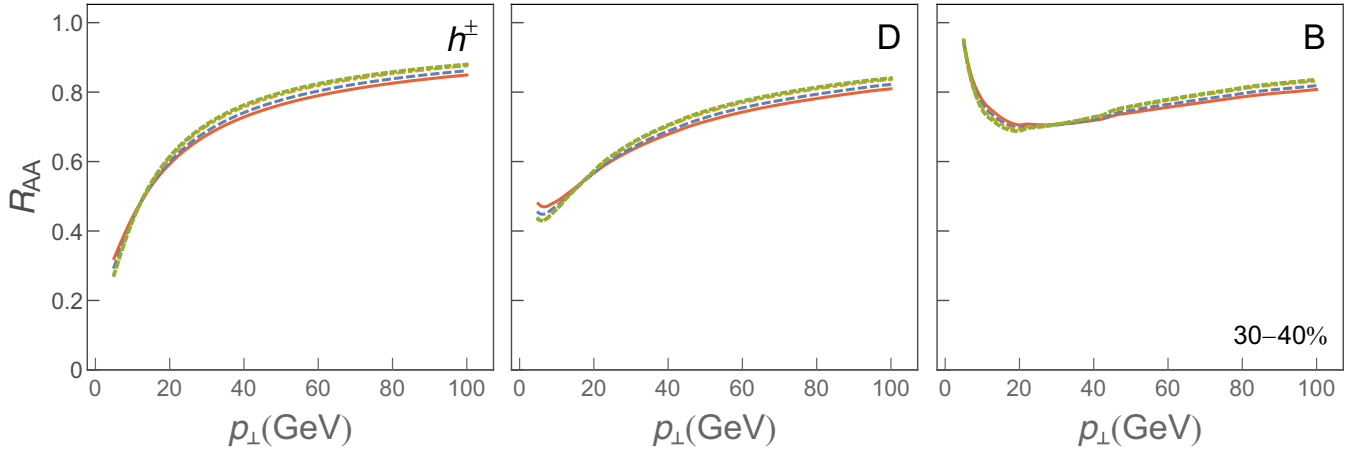


FIG. 6. R_{AA} dependence on p_{\perp} for four different medium evolutions depicted in Fig. 5. Left, central, and right panels correspond to charged hadrons, D mesons, and B mesons, respectively. In each panel, the T profile corresponding to the case (a') from Fig. 5 is represented by a full red curve, case (b') by a dashed blue curve, case (c') by a dot-dashed orange curve, and case (d') by a dotted green curve. The results correspond to the centrality bin 30–40%, and $\mu_M/\mu_E = 0.5$.

part of this section), are denoted as (a')–(d') and referred to as “modified” T profiles ((a) \equiv (a')).

In this second T -profiles setup, we first verify from Fig. 6 that R_{AA} s in all four cases practically overlap, as expected. We next address how these modified evolution cases (a')–(d') affect v_2 . From Fig. 7 we see that v_2 is now very sensitive to the transition from the free-streaming case to other modified T profiles. More accurately, for all types of particles, the lowest v_2 is observed in the modified divergent case, while the highest v_2 is observed in the free-streaming case.

The observation from Fig. 7 leads to the following two questions: (i) Why is v_2 altered by these modified T profiles (a')–(d')? (ii) Are these discrepancies a consequence of different initial stages? To answer these questions, we first note that, within this setup, the differences between v_2 (observed in Fig. 7) are proportional to $R_{AA}^{\text{in}} - R_{AA}^{\text{out}}$, as the denominator in Eq. (6) (as a starting premise) remains practically unchanged (see Fig. 6). The transverse momentum dependence of $R_{AA}^{\text{in}} -$

R_{AA}^{out} is further shown in Fig. 8 for charged hadrons (as results for D and B mesons will lead to the same conclusion). We see a clear hierarchy, i.e., the largest $R_{AA}^{\text{in}} - R_{AA}^{\text{out}}$ for free streaming, descending towards the divergent case. To quantitatively understand this observation, we note that for R_{AA}^{in} the high- p_{\perp} probes traverse, on average, the medium up to \bar{L}_{in} , while for R_{AA}^{out} the medium is traversed up to \bar{L}_{out} . Consequently, if we refer to Fig. 5, $R_{AA}^{\text{in}} - R_{AA}^{\text{out}}$ comes from the T -profile difference in the time region between \bar{L}_{in} and \bar{L}_{out} , i.e., upon thermalization. Since in this region $\bar{T}_{d'} < \bar{T}_{c'} < \bar{T}_{b'} < \bar{T}_{a'}$ holds, $R_{AA}^{\text{in}} - R_{AA}^{\text{out}}$ is the largest for the free-streaming case and the smallest for the divergent case, as observed in Fig. 8, and in agreement with v_2 ordering in Fig. 7. This therefore provides clarification of why $R_{AA}^{\text{in}} - R_{AA}^{\text{out}}$, and consequently v_2 , is affected by these four different QGP evolution profiles, and that this difference originates primarily from the interactions of high- p_{\perp} partons with thermalized QGP and not the initial stages. This agrees with the first part of this section

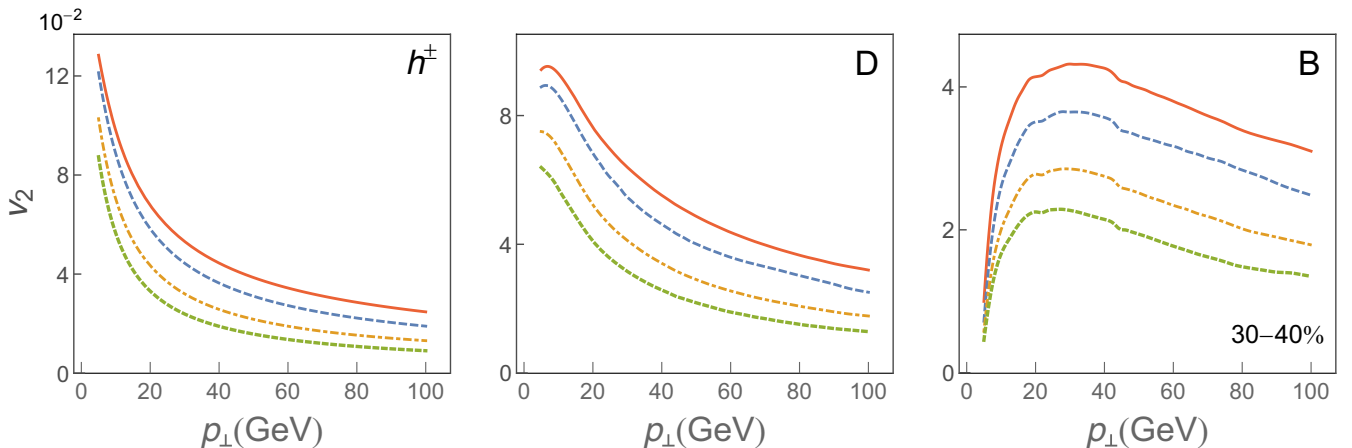


FIG. 7. v_2 dependence on p_{\perp} for four different medium evolutions depicted in Fig. 5. Left, central, and right panels correspond to charged hadrons, D mesons, and B mesons, respectively. The labeling and remaining parameters are the same as in Fig. 6.

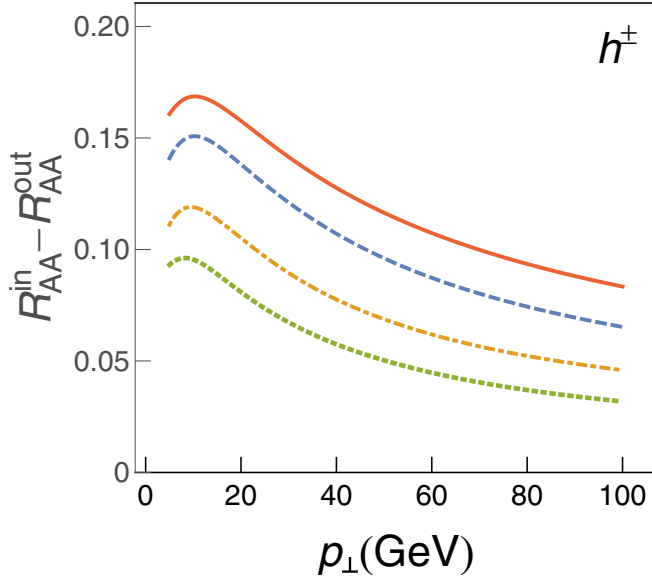


FIG. 8. $R_{AA}^{\text{in}} - R_{AA}^{\text{out}}$ dependence on p_{\perp} for charged hadrons. The labeling and remaining parameters are the same as in Fig. 6.

(Figs. 2 and 3), where we showed and explained insensitivity of v_2 to different initial stages. It is worth emphasizing that, contrary to the first part of this section, in the second part we tested the effects on R_{AA} and v_2 not from distinctive initial stages but instead from four entirely different evolutions of the QCD medium (related by the same global property, i.e., average temperature).

In the final, third part of this section we adopt a commonly used approach, in which the energy loss is fitted through change of multiplicative fitting factor in the energy loss, to reproduce the desired high- p_{\perp} R_{AA} , e.g., the one that best fits the experimental data (see, e.g., [24,30,65,69–71]). To this end, we use the same four T profiles from the first part of

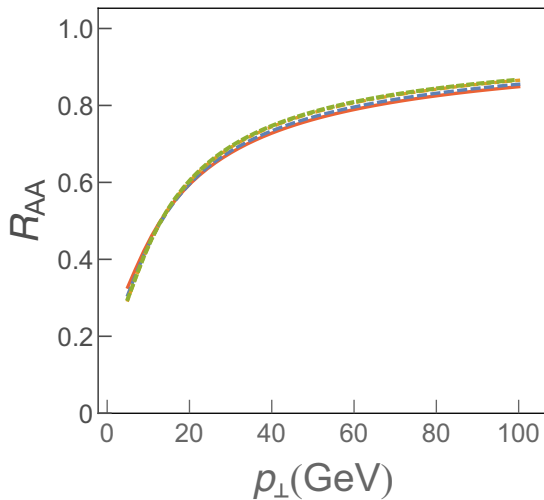


TABLE I. Fitting factors values.

T profile case	C_i^{fit}
Free-streaming case (a)	1
Linear case (b)	0.87
Constant case (c)	0.74
Divergent case (d)	0.67

this section (Fig. 1), while in our full-fledged calculations (see Sec. II) we introduce an additional multiplicative fitting factor (free parameter) C_i^{fit} , $i = b, c, d$. C_i^{fit} is then estimated for each initial-stage case as a best fit to the free-streaming R_{AA} (see Table I). Thus-obtained R_{AA} s are shown in the left panel of Fig. 9 only for the representative case of h^{\pm} , as the same conclusions stand for both light and heavy flavor hadrons. From the left panel of this figure we observe practically overlapping R_{AA} s in all (a)–(d) cases, as anticipated, which is obtained by decreasing C_i^{fit} consistently from the free-streaming to the divergent case (each $C_i^{\text{fit}} \leq 1$) in order to *compensate* for the higher energy losses in the corresponding cases compared to the case (a).

The effect of different T profiles from Fig. 1 after introduction of multiplicative fitting factor C_i^{fit} in the full-fledged numerical procedure on v_2 is depicted in the right panel of Fig. 9, where we see that elliptic flow in (a)–(d) cases notably differs; i.e., it is the highest in the free-streaming case while it is the lowest in the divergent case. Based on this observation, one could naively infer that initial stages, i.e., the $\tau < \tau_0$ region (the only region in which T profiles differ), have a significant effect on v_2 , as recently observed by an alternative approach [24].

However, this kind of reasoning is inconsistent with our analysis outlined in the first two parts of this section, as well as with intuitive expectation that introduction of the

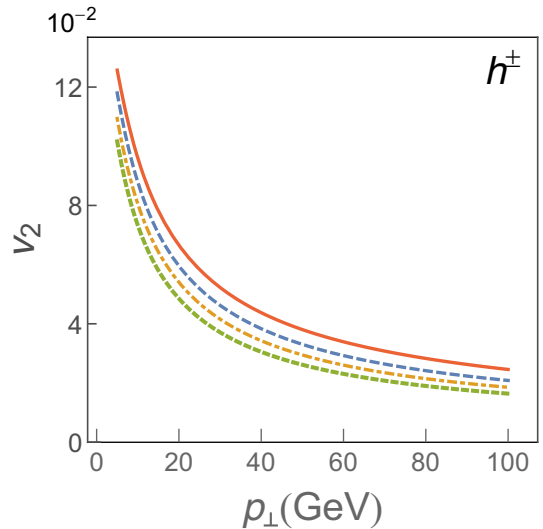


FIG. 9. R_{AA} (left panel) and v_2 (right panel) dependence on p_{\perp} for charged hadrons, when an additional energy loss multiplicative factor is introduced to reproduce the free-streaming R_{AA} , in four different initial-stage cases depicted in Fig. 1. The labeling and remaining parameters are the same as in Fig. 2.

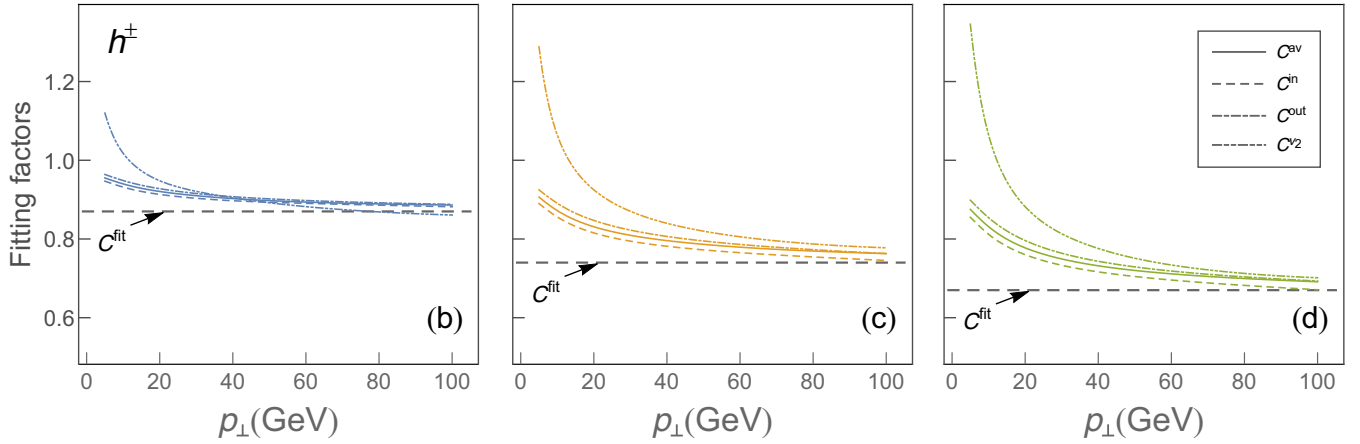


FIG. 10. Comparison of four fitting factors defined by Eq. (16) with the C_i^{fit} value, obtained from full-fledged numerical procedure, in *linear* (b) (left), *constant* (c) (central) and *divergent* (d) (right panel) cases. C factors represented by full, long-dashed, dot-dashed, and dot-dot-dashed curves correspond to h^{\pm} angular averaged, in-plane, and out-of-plane R_{AA} and v_2 cases, respectively. The horizontal gray dashed line represents the energy loss fitted value C_i^{fit} . The results correspond to the centrality bin 30–40%, and $\mu_M/\mu_E = 0.5$.

energy loss at the initial stage affects R_{AA} . To quantitatively understand this result, we introduce asymptotic scaling behavior [27,54,72]. That is, for higher p_{\perp} of the initial jet, and for higher centralities (where fractional energy loss is expected to be small), we can make the following estimates:

$$\Delta E/E \approx \chi \bar{T}^m \bar{L}^n, \quad R_{AA} \approx 1 - \frac{l-2}{2} \frac{\Delta E}{E} = 1 - \xi \bar{T}^m \bar{L}^n, \quad (10)$$

where m, n are proportionality factors, \bar{T} is the average temperature of the QGP, \bar{L} denotes the average path length traversed by the jet, χ is a proportionality factor (that depends on p_{\perp} and flavor of the jet). $\xi = \frac{l-2}{2} \chi$, where l is the steepness of a power law fit to the transverse momentum distribution.

If $\Delta E/E$ is fitted by additional multiplicative factor C , the new R_{AA}^{fit} becomes

$$R_{AA,i}^{fit} \approx 1 - C_i \xi \bar{T}^m \bar{L}^n \approx 1 - C_i (1 - R_{AA,i}), \quad (11)$$

where $i = b, c, d$ and C_i ($C_i < 1, \forall i$) denotes the fitting factor, and the last part of Eq. (11) is obtained by using Eq. (10), leading to

$$C_i \approx \frac{1 - R_{AA,i}^{fit}}{1 - R_{AA,i}}. \quad (12)$$

We note that Eq. (12) is applicable to the average, in-plane and out-of-plane R_{AA} s, since the same fitting factor is consistently applied in all three cases. By imposing the condition (which quantifies the equivalence of fitted R_{AA} in (b)–(d) cases to the free-streaming case)

$$R_{AA,i}^{fit} = R_{AA,a}, \quad (13)$$

and by applying Eqs. (5)–(8) and (13), together with Eqs. (10) and (11) and their in-plane and out-of-plane analogs, we obtain

$$\begin{aligned} v_{2,i}^{fit} &\approx \frac{1}{2} \frac{C_i (R_{AA,i}^{in} - R_{AA,i}^{out})}{2R_{AA,a}} = \frac{1}{2} \frac{C_i \gamma_{ia} (R_{AA,a}^{in} - R_{AA,a}^{out})}{R_{AA,a}^{in} + R_{AA,a}^{out}} \\ &= C_i \gamma_{ia} v_{2,a}, \end{aligned} \quad (14)$$

which can also be written as

$$C_i \approx \frac{v_{2,i}^{fit}}{\gamma_{ia} v_{2,a}}. \quad (15)$$

From Eq. (14), we see that decrease of v_2^{fit} in (b)–(d) cases compared to (a) is a result of a fitting factor $C_i(p_{\perp})$ (which is smaller than 1), as well as the proportionality function $\gamma_i(p_{\perp})$ (also smaller than 1). However, note that Eq. (14) describes asymptotic behavior at very high p_{\perp} , where, as shown earlier, γ_s approach 1. Consequently, the diminishing of elliptic flow compared to the case (a) is predominantly due to a decrease of the *artificially imposed fitting factor* C . Therefore, we obtain that, contrary to [24], *initial stages are not* mainly responsible for the obtained differences (the right panel of Fig. 9) in the v_2^{fit} curves for different T profiles. Moreover, this argument, as well as the obtained inconsistency of the results in this and the first two parts of the paper, implies that application of multiple fitting procedure for each different initial stage may result in incorrect energy loss estimates and in misinterpreting the underlying physics.

To assess if this qualitative conclusion indeed holds, i.e., that v_2 susceptibility observed in Fig. 9 (as well as in [24]) is indeed mainly a consequence of a fitting factor in the energy loss, in Fig. 10 we check the consistency of Eqs. (12) and (15) with the full-fledged numerical calculations. That is, a nontrivial consequence of Eqs. (12) and (15) is that C_i factors for the average, in-plane, and out-of-plane R_{AA} 's (Eq. (12)) and v_2 (Eq. (15)), should be the same in high- p_{\perp} limit, and moreover overlap with C_i^{fit} in this limit. To this end, we define the C factors (originating from Eqs. (12) and (15))

$$\begin{aligned} C_i^{in} &= \frac{1 - R_{AA,i}^{in,fit}}{1 - R_{AA,i}^{in}}, & C_i^{out} &= \frac{1 - R_{AA,i}^{out,fit}}{1 - R_{AA,i}^{out}}, \\ C_i^{av} &= \frac{1 - R_{AA,i}^{fit}}{1 - R_{AA,i}}, & C_i^{v_2} &= \frac{1}{\gamma_{ia}} \frac{v_{2,i}^{fit}}{v_{2,a}} \end{aligned} \quad (16)$$

and compare them with C_i^{fit} , for each separate initial-stages case, $i = b, c, d$. Note that, while terms themselves on the right-hand side of each expression in Eq. (16) are obtained from Eqs. (12) and (15) in the high- p_\perp limit (and consequently are expected to overlap in this limit, if our analytical estimate is valid), we calculate C_i^{fit} , and the terms on the right-hand side of each expression in Eq. (16), through full-fledged numerical procedure. We indeed observe that, for each i and at high- p_\perp , C_i^{in} , C_i^{out} , C_i^{av} , and $C_i^{v_2}$ factors are practically overlapping, and approach the value C_i^{fit} . Consequently, this highly nontrivial observation confirms that our qualitative conclusion is valid, and that v_2 susceptibility in this case is indeed mainly a consequence of an additionally introduced fitting factor.

IV. CONCLUSIONS

Traditionally, the features of initial stages before QGP thermalization are explored through comparison of bulk medium simulations and low- p_\perp data. On the other hand, the recent abundance of high- p_\perp experimental data motivates exploiting the high- p_\perp energy loss in studying the initial stages. We here utilized state-of-the-art dynamical energy loss embedded in analytical 1D Bjorken medium expansion (DREENA-B framework), which allowed us to tightly control the analyzed temperature profiles. In particular, we considered four temperature profiles, which are identical after thermalization but are different before thermalization, corresponding to four commonly considered initial-stage cases. This allowed us to study the effects of different initial-stage cases on high- p_\perp

R_{AA} and v_2 predictions, under highly controlled conditions, by combining full-fledged numerical results and analytical estimates used to interpret the experimental results.

We found that high- p_\perp R_{AA} is sensitive to the prethermalized stages of the medium evolution; however, within the current error bars, the sensitivity is not sufficient to distinguish between different scenarios. On the other hand, the high- p_\perp v_2 is unexpectedly insensitive to the initial stages. We furthermore found that previously reported sensitivity [24] of high- p_\perp v_2 to initial stages is mainly a consequence of the fitting procedure in which the parameters in the energy loss are adjusted to reproduce experimentally observed R_{AA} individually for different initial-stage cases. On the other hand, if the same global property, in particular the same average temperature, is imposed to tested temperature profiles, high sensitivity of high- p_\perp v_2 is again obtained. This sensitivity is, however, a consequence of differences in final rather than initial stages. Overall, our results underscore that the simultaneous study of high- p_\perp R_{AA} and v_2 , with consistent that is, fixed energy loss parameters across the entire study and controlled temperature profiles (reflecting only the differences in the initial stages), is crucial to impose accurate constraints on the initial stages.

ACKNOWLEDGMENTS

We thank Pasi Huovinen and Jussi Auvinen for useful discussions. This work is supported by the European Research Council, Grant No. ERC-2016-COG:725741, and by the Ministry of Science and Technological Development of the Republic of Serbia, under Projects No. ON171004 and No. ON173052.

-
- [1] J. C. Collins and M. J. Perry, *Phys. Rev. Lett.* **34**, 1353 (1975).
 - [2] G. Baym and S. A. Chin, *Phys. Lett. B* **62**, 241 (1976).
 - [3] M. Gyulassy and L. McLerran, *Nucl. Phys. A* **750**, 30 (2005).
 - [4] E. V. Shuryak, *Nucl. Phys. A* **750**, 64 (2005); *Rev. Mod. Phys.* **89**, 035001 (2017).
 - [5] B. Jacak and P. Steinberg, *Phys. Today* **63**(5), 39 (2010).
 - [6] B. Muller, J. Schukraft, and B. Wyslouch, *Annu. Rev. Nucl. Part. Sci.* **62**, 361 (2012).
 - [7] S. Acharya *et al.* (ALICE Collaboration), *J. High Energy Phys.* **11** (2018) 013.
 - [8] ATLAS Collaboration, Report No. ATLAS-CONF-2017-012 (unpublished).
 - [9] V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **04** (2017) 039.
 - [10] S. Jaelani (ALICE Collaboration), *Int. J. Mod. Phys. Conf. Ser.* **46**, 1860018 (2018).
 - [11] J. Wang (CMS Collaboration), *Nucl. Part. Phys. Proc.* **289–290**, 249 (2017).
 - [12] T. W. Wang (CMS Collaboration), *Nucl. Part. Phys. Proc.* **289–290**, 229 (2017).
 - [13] A. Adare *et al.* (PHENIX Collaboration), *Phys. Rev. Lett.* **101**, 232301 (2008); *Phys. Rev. C* **87**, 034911 (2013).
 - [14] B. I. Abelev *et al.* (STAR Collaboration), *Phys. Lett. B* **655**, 104 (2007).
 - [15] S. Acharya *et al.* (ALICE Collaboration), *J. High Energy Phys.* **07** (2018) 103.
 - [16] M. Aaboud *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **78**, 997 (2018).
 - [17] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **776**, 195 (2018).
 - [18] S. Acharya *et al.* (ALICE Collaboration), *Phys. Rev. Lett.* **120**, 102301 (2018).
 - [19] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **120**, 202301 (2018).
 - [20] F. Gelis and B. Schenke, *Annu. Rev. Nucl. Part. Sci.* **66**, 73 (2016).
 - [21] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **11** (2013) 183.
 - [22] H. Niemi, G. S. Denicol, H. Holopainen, and P. Huovinen, *Phys. Rev. C* **87**, 054901 (2013).
 - [23] J. Xu, A. Buzzatti, and M. Gyulassy, *J. High Energy Phys.* **08** (2014) 063.
 - [24] C. Andres, N. Armesto, H. Niemi, R. Paatelainen, and C. A. Salgado, *Phys. Lett. B* **803**, 135318 (2020).
 - [25] R. Katz, C. A. G. Prado, J. Noronha-Hostler, J. Noronha, and A. A. P. Suaide, [arXiv:1906.10768](https://arxiv.org/abs/1906.10768).
 - [26] J. D. Bjorken, *Phys. Rev. D* **27**, 140 (1983).
 - [27] D. Zigic, I. Salom, M. Djordjevic, and M. Djordjevic, *Phys. Lett. B* **791**, 236 (2019).

- [28] J. Xu, J. Liao, and M. Gyulassy, *Chin. Phys. Lett.* **32**, 092501 (2015).
- [29] S. Shi, J. Liao, and M. Gyulassy, *Chin. Phys. C* **42**, 104104 (2018); **43**, 044101 (2019).
- [30] J. Noronha-Hostler, B. Betz, J. Noronha, and M. Gyulassy, *Phys. Rev. Lett.* **116**, 252301 (2016).
- [31] S. K. Das, F. Scardina, S. Plumari, and V. Greco, *Phys. Lett. B* **747**, 260 (2015).
- [32] M. Djordjevic *et al.* (unpublished).
- [33] M. Djordjevic and M. Djordjevic, *Phys. Lett. B* **734**, 286 (2014).
- [34] S. Wicks, W. Horowitz, M. Djordjevic, and M. Gyulassy, *Nucl. Phys. A* **784**, 426 (2007).
- [35] Z. B. Kang, I. Vitev, and H. Xing, *Phys. Lett. B* **718**, 482 (2012); R. Sharma, I. Vitev, and B. W. Zhang, *Phys. Rev. C* **80**, 054902 (2009).
- [36] D. de Florian, R. Sassot, and M. Stratmann, *Phys. Rev. D* **75**, 114010 (2007).
- [37] M. Cacciari and P. Nason, *J. High Energy Phys.* **09** (2003) 006; E. Braaten, K.-M. Cheung, S. Fleming, and T. C. Yuan, *Phys. Rev. D* **51**, 4819 (1995).
- [38] V. G. Kartvelishvili, A. K. Likhoded, and V. A. Petrov, *Phys. Lett. B* **78**, 615 (1978).
- [39] M. Djordjevic, *Phys. Rev. C* **74**, 064907 (2006).
- [40] M. Djordjevic, *Phys. Rev. C* **80**, 064909 (2009).
- [41] M. Djordjevic and U. Heinz, *Phys. Rev. Lett.* **101**, 022302 (2008).
- [42] B. Blagojevic and M. Djordjevic, *J. Phys. G* **42**, 075105 (2015).
- [43] J. I. Kapusta, *Finite-Temperature Field Theory* (Cambridge University Press, Cambridge, 1989).
- [44] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, *Nucl. Phys. B* **483**, 291 (1997); R. Baier, Yu. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, **484**, 265 (1997).
- [45] N. Armesto, C. A. Salgado, and U. A. Wiedemann, *Phys. Rev. D* **69**, 114003 (2004).
- [46] M. Gyulassy, P. Levai, and I. Vitev, *Nucl. Phys. B* **594**, 371 (2001).
- [47] X. N. Wang and X. F. Guo, *Nucl. Phys. A* **696**, 788 (2001).
- [48] M. Djordjevic, *Phys. Lett. B* **709**, 229 (2012).
- [49] B. Blagojevic, M. Djordjevic, and M. Djordjevic, *Phys. Rev. C* **99**, 024901 (2019).
- [50] Y. Maezawa, S. Aoki, S. Ejiri, T. Hatsuda, N. Ishii, K. Kanaya, N. Ukita, and T. Umeda (WHOT-QCD Collaboration), *Phys. Rev. D* **81**, 091501(R) (2010).
- [51] A. Nakamura, T. Saito, and S. Sakai, *Phys. Rev. D* **69**, 014506 (2004).
- [52] M. Gyulassy, P. Levai, and I. Vitev, *Phys. Lett. B* **538**, 282 (2002).
- [53] A. Dainese, *Eur. Phys. J. C* **33**, 495 (2004).
- [54] D. Zigic, I. Salom, J. Auvinen, M. Djordjevic, and M. Djordjevic, *J. Phys. G* **46**, 085101 (2019).
- [55] A. Peshier, [arXiv:hep-ph/0601119](https://arxiv.org/abs/hep-ph/0601119).
- [56] R. Field, *Applications of Perturbative QCD* (Perseus Books, Cambridge, MA, 1995).
- [57] M. Djordjevic and M. Gyulassy, *Phys. Rev. C* **68**, 034914 (2003).
- [58] E. Braaten and M. H. Thoma, *Phys. Rev. D* **44**, 1298 (1991).
- [59] P. F. Kolb and U. W. Heinz, in *Quark-Gluon Plasma 3*, edited by R. C. Hwa and X.-N. Wang (World Scientific, Singapore, 2004), p. 634.
- [60] J. E. Bernhard, J. S. Moreland, and S. A. Bass, *Nucl. Phys. A* **967**, 293 (2017).
- [61] M. Wilde (ALICE Collaboration), *Nucl. Phys. A* **904–905**, 573c (2013).
- [62] M. Djordjevic, M. Gyulassy, R. Vogt, and S. Wicks, *Phys. Lett. B* **632**, 81 (2006).
- [63] A. Bazavov *et al.* (HotQCD Collaboration), *Phys. Rev. D* **90**, 094503 (2014).
- [64] P. Christiansen, K. Tywoniuk, and V. Vislavicius, *Phys. Rev. C* **89**, 034912 (2014).
- [65] B. Betz and M. Gyulassy, *J. High Energy Phys.* **08** (2014) 090 [Erratum: **10** (2014) 043].
- [66] S. Cao, L. G. Pang, T. Luo, Y. He, G. Y. Qin, and X. N. Wang, *Nucl. Part. Phys. Proc.* **289–290**, 217 (2017).
- [67] T. Renk, *Phys. Rev. C* **85**, 044903 (2012).
- [68] D. Molnar and D. Sun, *Nucl. Phys. A* **932**, 140 (2014); **910–911**, 486 (2013).
- [69] J. Noronha-Hostler, B. Betz, M. Gyulassy, M. Luzum, J. Noronha, I. Portillo, and C. Ratti, *Phys. Rev. C* **95**, 044901 (2017).
- [70] C. A. G. Prado, J. Noronha-Hostler, R. Katz, A. A. P. Suaide, J. Noronha, and M. G. Munhoz, *Phys. Rev. C* **96**, 064903 (2017).
- [71] S. Cao *et al.* (JETSCAPE Collaboration), *Phys. Rev. C* **96**, 024909 (2017).
- [72] M. Djordjevic, D. Zigic, M. Djordjevic, and J. Auvinen, *Phys. Rev. C* **99**, 061902(R) (2019).



XXVIIIth International Conference on Ultrarelativistic Nucleus-Nucleus Collisions
(Quark Matter 2019)

From high p_{\perp} theory and data to inferring anisotropy of Quark-Gluon Plasma

Magdalena Djordjevic^a, Stefan Stojku^a, Dusan Zigic^a, Bojana Ilic^a, Jussi
Auvinen^a, Igor Salom^a, Marko Djordjevic^b and Pasi Huovinen^a

^a *Institute of Physics Belgrade, University of Belgrade, Serbia*

^b *Faculty of Biology, University of Belgrade, Serbia*

Abstract

High p_{\perp} theory and data are commonly used to study high p_{\perp} parton interactions with QGP, while low p_{\perp} data and corresponding models are employed to infer QGP bulk properties. On the other hand, with a proper description of high p_{\perp} parton-medium interactions, high p_{\perp} probes become also powerful tomography tools, since they are sensitive to global QGP features, such as different temperature profiles or initial conditions. This tomographic role of high p_{\perp} probes can be utilized to assess the spatial anisotropy of the QCD matter. With our dynamical energy loss formalism, we show that a (modified) ratio of R_{AA} and v_2 presents a reliable and robust observable for straightforward extraction of initial state anisotropy. We analytically estimated the proportionality between the $v_2/(1-R_{AA})$ and anisotropy coefficient ϵ_{2L} , and found surprisingly good agreement with full-fledged numerical calculations. Within the current error bars, the extraction of the anisotropy from the existing LHC data using this approach is still inaccessible. However, with the expected accuracy improvement in the upcoming LHC runs, the anisotropy of the QGP formed in heavy ion collisions can be straightforwardly derived from the data. Such a data-based anisotropy parameter would present an important test to models describing the initial stages of heavy-ion collision and formation of QGP, and demonstrate the usefulness of high p_{\perp} theory and data in obtaining QGP properties.

Keywords: Quark-gluon plasma, High p_{\perp} probes, Initial anisotropy

1. Introduction

Understanding the properties of the new form of matter named Quark-Gluon Plasma (QGP) is the major goal of relativistic heavy ion physics [1, 2]. However, to explore the properties of QGP, one needs good probes. With regards to that, it is commonly assumed that high p_{\perp} theory and data are good probes for exploring the high p_{\perp} parton interactions with QGP, while low p_{\perp} theory and data are considered as good probes for bulk QGP properties. Contrary to this common assumption, the goal of this contribution is to demonstrate that high p_{\perp} particles can also be useful independent probes of *bulk* QGP properties.

To put it simply, the main idea is that when high p_{\perp} particles transverse QGP, they lose energy, where this energy loss is sensitive to bulk QGP properties, such as its temperature profiles or initial conditions.

Consequently, with a realistic and sophisticated high p_{\perp} parton energy loss model, high p_{\perp} probes can indeed become powerful tomographic tools. So, in this contribution, we will present how we can use these probes to infer some of the bulk QGP properties, i.e., for precision QGP tomography. Note that only the main results are presented here; for a more detailed version, see [3], and references therein.

2. DREENA framework

To achieve the goal of utilizing high p_{\perp} theory and data for inferring the bulk QGP properties, as previously implied, a reliable high p_{\perp} parton energy loss model is necessary. With this goal in mind, we developed a dynamical energy loss formalism [4, 5], which takes into account some more realistic and unique features, such as: i) The calculations are performed within finite temperature field theory and generalized Hard-Thermal-Loop [6] approach, in which the infrared divergences are naturally regulated, excluding the need for artificial cutoffs. ii) The formalism assumes QCD medium of finite size and finite temperature, consisting of dynamical partons (i.e., energy exchange with medium constituents is included), in distinction to commonly considered static scatterers approximation and/or models with vacuum-like propagators. iii) Both radiative [4] and collisional [5] energy losses are calculated within the same theoretical framework, and are equally applicable to light and heavy flavors. iv) The formalism is generalized to include a finite chromomagnetic mass [7], running coupling, and to relax the widely used soft-gluon approximation [8]. Finally, the formalism is integrated in a numerical framework DREENA (Dynamical Radiative and Elastic ENergy loss Approach) [9, 10], to provide predictions for high p_{\perp} observables.

Within this framework, we generated a wide set of high p_{\perp} predictions using 1D Bjorken expansion [11] (i.e., DREENA-B framework [10]). Thus we obtained a good joint agreement with a wide range of high p_{\perp} R_{AA} and v_2 data, by applying the same numerical procedure, the same parameter set, and no fitting parameters in model testing. That is, there is no v_2 puzzle [12] within our model, which then strongly suggests that the model provides a realistic description of high p_{\perp} parton-medium interactions. Moreover, our preliminary findings suggest that, within our formalism, moving from 1D Bjorken to full 3D hydrodynamical expansion does not significantly affect the agreement of our predictions with high p_{\perp} R_{AA} and v_2 data [13]. Consequently, in order to adequately address the high p_{\perp} measurements, a proper description of high p_{\perp} parton interactions with the medium appears to be much more important than an advanced medium evolution description. Furthermore, we have also analyzed the sensitivity of high p_{\perp} R_{AA} and v_2 to different initial stages, giving an additional insight in the usefulness of both high p_{\perp} observables in the precision QGP tomography [14].

3. Inferring QGP anisotropy through high p_{\perp} theory and data

As one example of QGP tomography, in this contribution, we will address how to infer the QGP anisotropy from high p_{\perp} R_{AA} and v_2 data. The initial state anisotropy is one of the main properties of QGP and a major limiting factor for precision QGP tomography. However, despite its essential importance, it is still not possible to directly infer the initial anisotropy from experimental measurements. Several theoretical studies [15, 16, 17, 18] have provided different methods for calculating the initial anisotropy, leading to notably different predictions, with a notable effect in the resulting predictions for both low and high p_{\perp} data. Therefore, approaches for inferring anisotropy from the data are necessary. Optimally, these approaches should be complementary to existing predictions, i.e., based on a method that is fundamentally different from models of early stages of QCD matter.

To this end, we here propose a novel approach to extract the initial state anisotropy. Our method is based on inference from high p_{\perp} data, by using already available R_{AA} and v_2 measurements, which will moreover be measured with much higher precision in the future. Such an approach is substantially different from the existing approaches, as it is based on the inference from experimental data (rather than on calculations of early stages of QCD matter) exploiting the information from interactions of rare high p_{\perp} partons with the QCD medium. This also presents an improvement/optimization in utilizing high p_{\perp} data as, to date, these data were mostly constrained on studying the parton-medium interactions, rather than assessing bulk QGP parameters, such as spatial asymmetry.

In the literature, the initial state anisotropy is quantified in terms of eccentricity parameter e_2

$$e_2 = \frac{\langle y^2 - x^2 \rangle}{\langle y^2 + x^2 \rangle} = \frac{\int dx dy (y^2 - x^2) \rho(x, y)}{\int dx dy (y^2 + x^2) \rho(x, y)}, \quad (1)$$

where $\rho(x, y)$ denotes the initial density distribution of the formed QGP. Regarding high p_\perp observables, we note that v_2 is sensitive to both the anisotropy of the system and its size, while R_{AA} is sensitive only to the size of the system. Therefore, it is plausible that the adequate observable for extracting eccentricity from high p_\perp data depends on both v_2 and R_{AA} , and the question is how.

To address this question, we will use the dynamical energy loss formalism, and DREENA-B framework outlined above. For high p_\perp , the fractional energy loss scales as [3] $\Delta E/E \sim \chi \langle T \rangle^a \langle L \rangle^b$, where $\langle T \rangle$ stands for the average temperature along the path of high p_\perp parton, $\langle L \rangle$ is the average path-length traversed by the parton, χ is a proportionality factor that depends on the initial parton transverse momentum, and a and b are exponents which govern the temperature and path-length dependence of the energy loss. Within our model, $a \approx 1.2$ and $b \approx 1.4$, which is contrary to simpler models, and consistent with a wide range of experimental data [19, 20]. From this simple scaling argument, we can straightforwardly obtain the following expressions for R_{AA} and v_2 (for more details we refer the reader to [3]):

$$R_{AA} \approx 1 - \xi(\chi) \langle T \rangle^a \langle L \rangle^b, \quad v_2 \approx \frac{1}{2} \frac{R_{AA}^{in} - R_{AA}^{out}}{R_{AA}^{in} + R_{AA}^{out}} \approx \xi(\chi) \langle T \rangle^a \langle L \rangle^b \left(\frac{b}{2} \frac{\Delta L}{\langle L \rangle} - \frac{a}{2} \frac{\Delta T}{\langle T \rangle} \right), \quad (2)$$

where we see that $\xi(\chi) \langle T \rangle^a \langle L \rangle^b$ corresponds to $1 - R_{AA}$. Therefore, if we divide v_2 by $(1 - R_{AA})$, we see that this ratio is given by the following simple expression:

$$\frac{v_2}{1 - R_{AA}} \approx \left(\frac{b}{2} \frac{\Delta L}{\langle L \rangle} - \frac{a}{2} \frac{\Delta T}{\langle T \rangle} \right). \quad (3)$$

Note that, while this ratio exposes the dependence on the asymmetry of the system (through spatial $(\Delta L/\langle L \rangle)$ and temperature $(\Delta T/\langle T \rangle)$ parts), the dependence only on spatial anisotropy is still not isolated. However, by plotting together spatial and temperature anisotropy, we obtain a linear dependence [3], with a proportionality factor given by $c \approx 4.3$. Therefore, $v_2/(1 - R_{AA})$ reduces to the following expression:

$$\frac{v_2}{1 - R_{AA}} \approx \frac{1}{2} \left(b - \frac{a}{c} \right) \frac{\langle L_{out} \rangle - \langle L_{in} \rangle}{\langle L_{out} \rangle + \langle L_{in} \rangle} \approx 0.57\zeta, \quad \text{where } \zeta = \frac{\langle L_{out} \rangle - \langle L_{in} \rangle}{\langle L_{out} \rangle + \langle L_{in} \rangle} \text{ and } \frac{1}{2} \left(b - \frac{a}{c} \right) \approx 0.57. \quad (4)$$

Consequently, the asymptotic scaling behavior of observables v_2 and R_{AA} , at high p_\perp , reveals that their (moderated) ratio is determined only by the geometry of the initial QGP droplet. Therefore, the anisotropy parameter ζ could, in principle, be directly obtained from the high p_\perp experimental data.

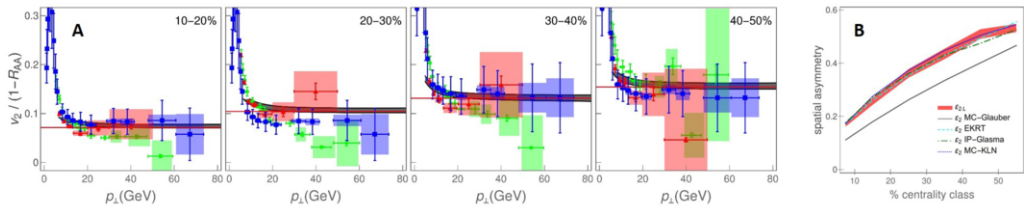


Fig. 1. **A)** Comparison of theoretical predictions for charged hadron $v_2/(1 - R_{AA})$ as a function of p_\perp with 5.02 TeV $Pb + Pb$ CMS [21, 22] (blue squares), ALICE [23, 24] (red triangles) and ATLAS [25, 26] (green circles) data. Each panel corresponds to different centrality range, as indicated in the upper right corners, while red lines denote the limit 0.57ζ from Eq. (4). **B)** Comparison of e_{2L} (red band) extracted from our full-fledged calculations, with e_2 obtained from MC-Glauber [15] (gray full curve), EKRT [16] (cyan dashed curve), IP-Glasma [17] (green dot-dashed curve) and MC-KLN [18] (blue dotted curve) models. MC-Glauber and EKRT curves correspond to 5.02 TeV, whereas IP-Glasma and MC-KLN curves correspond to 2.76 TeV $Pb + Pb$ collisions at the LHC.

To test the adequacy of the analytical estimate given by Eqs. (2)-(4), Fig. 1A is displayed, which comprises our $v_2/(1 - R_{AA})$ predictions (gray bands), stemming from our full-fledged recently developed

DREENA-B framework (outlined in the previous section), the ALICE, CMS and ATLAS data, and analytically derived asymptote 0.57ζ (red lines). Importantly, for each centrality range and for $p_{\perp} \gtrsim 20$ GeV, $v_2/(1 - R_{AA})$ is independent on p_{\perp} , and approaches the asymptote, i.e., is determined by the geometry of the system - depicted by the solid red line, up to 5% accuracy. Moreover, the experimental data for all three experiments also display the independence on the p_{\perp} and agree with our predictions, although the error bars are rather large. Therefore, we conclude that our scaling estimates are valid and that $v_2/(1 - R_{AA})$ indeed carries the information about the anisotropy of the fireball, which can be simply (from the straight line fit to data at high p_{\perp} limit) and robustly (in the same way for each centrality) inferred from the experimental data.

However, note that the anisotropy parameter ζ is not the widely-considered anisotropy parameter ϵ_2 (given by Eq. (1)). To facilitate comparison with ϵ_2 values in the literature, we define $\epsilon_{2L} = \frac{(L_{out})^2 - (L_{in})^2}{(L_{out})^2 + (L_{in})^2} = \frac{2\zeta}{1+\zeta^2}$, and in Fig. 1B compare it with the results from different initial-state models [15, 16, 17, 18]. First, we should note that as a starting point, our initial ϵ_2 , through which we generate our path-length distributions, agrees with EKRT and IP-Glasma. However, what is highly non-trivial is that, as an outcome of this procedure, in which $v_2/(1 - R_{AA})$ is calculated (based on the full-fledged DREENA-B framework), we obtain ϵ_{2L} which practically coincides with our initial ϵ_2 and also with some of the conventional initial-state models. As an overall conclusion, the straightforward extraction of ϵ_{2L} and its agreement with values of the prevailing initial-state models' eccentricity (and our initial ϵ_2) is highly non-trivial and supports $v_2/(1 - R_{AA})$ as a reliable and robust observable for anisotropy. Additionally, the width of our ϵ_{2L} band is smaller than the difference in the ϵ_2 values obtained by using different models (e.g., MC-Glauber vs. MC-KLN). Therefore, our approach provides genuine resolving power to distinguish between different initial-state models, although it may not be possible to separate the finer details of more sophisticated models. This resolving power, moreover, comes from an entirely different perspective, i.e., from high p_{\perp} theory and data, supporting the usefulness of utilizing high p_{\perp} theory and data for inferring the bulk QGP properties.

Acknowledgements: This work is supported by the European Research Council, grant ERC-2016-COG: 725741, and by the Ministry of Science and Technological Development of the Republic of Serbia, under project numbers ON171004, ON173052 and ON171031.

References

- [1] M. Gyulassy and L. McLerran, Nucl. Phys. A **750**, 30 (2005); E. V. Shuryak, Nucl. Phys. A **750**, 64 (2005).
- [2] B. Jacak and P. Steinberg, Phys. Today **63**, 39 (2010); C. V. Johnson and P. Steinberg, Phys. Today **63**, 29 (2010).
- [3] M. Djordjevic, S. Stojku, M. Djordjevic and P. Huovinen, Phys. Rev. C **100**, no. 3, 031901 (2019).
- [4] M. Djordjevic, Phys. Rev. C **80**, 064909 (2009); M. Djordjevic and U. Heinz, Phys. Rev. Lett. **101**, 022302 (2008).
- [5] M. Djordjevic, Phys. Rev. C **74**, 064907 (2006).
- [6] J. I. Kapusta, *Finite-Temperature Field Theory* (Cambridge University Press, 1989).
- [7] M. Djordjevic and M. Djordjevic, Phys. Lett. B **709**, 229 (2012).
- [8] B. Blagojevic, M. Djordjevic and M. Djordjevic, Phys. Rev. C **99**, 024901 (2019).
- [9] D. Zigic, I. Salom, J. Auvinen, M. Djordjevic and M. Djordjevic, J. Phys. G **46**, 085101 (2019).
- [10] D. Zigic, I. Salom, M. Djordjevic and M. Djordjevic, Phys. Lett. B **791**, 236 (2019).
- [11] J. D. Bjorken, Phys. Rev. D **27**, 140 (1983).
- [12] J. Noronha-Hostler, B. Betz, J. Noronha and M. Gyulassy, Phys. Rev. Lett. **116**, no. 25, 252301 (2016); B. Betz and M. Gyulassy, JHEP **1408**, 090 (2014); S. K. Das, F. Scardina, S. Plumari and V. Greco, Phys. Lett. B **747**, 260 (2015).
- [13] D. Zigic, *et al.*, to be submitted.
- [14] D. Zigic, B. Ilic, M. Djordjevic and M. Djordjevic, arXiv:1908.11866 [hep-ph].
- [15] C. Loizides, J. Kamin and D. d'Enterria, Phys. Rev. C **97**, 054910 (2018).
- [16] K. J. Eskola, H. Niemi, R. Paatelainen and K. Tuominen, Phys. Rev. C **97**, 034911 (2018).
- [17] J. E. Bernhard, J. S. Moreland, S. A. Bass, J. Liu and U. Heinz, Phys. Rev. C **94**, 024907 (2016).
- [18] T. Hirano, P. Huovinen, K. Murase and Y. Nara, Prog. Part. Nucl. Phys. **70**, 108 (2013).
- [19] M. Djordjevic and M. Djordjevic, Phys. Rev. C **92**, 024918 (2015).
- [20] M. Djordjevic, D. Zigic, M. Djordjevic and J. Auvinen, Phys. Rev. C **99**(R), 061902 (2019).
- [21] V. Khachatryan *et al.* [CMS Collaboration], JHEP **1704**, 039 (2017).
- [22] A. M. Sirunyan *et al.* [CMS Collaboration], Phys. Lett. B **776**, 195 (2018).
- [23] S. Acharya *et al.* [ALICE Collaboration], JHEP **1811**, 013 (2018).
- [24] S. Acharya *et al.* [ALICE Collaboration], JHEP **1807**, 103 (2018).
- [25] [ATLAS Collaboration], ATLAS-CONF-2017-012.
- [26] M. Aaboud *et al.* [ATLAS Collaboration], Eur. Phys. J. C **78**, no. 12, 997 (2018).

A systems biology approach to COVID-19 progression in population

Magdalena Djordjevic^a, Andjela Rodic^b, Igor Salom^a, Dusan Zigic^a,
Ognjen Milicevic^c, Bojana Ilic^a, and Marko Djordjevic^{b,*}

^aInstitute of Physics Belgrade, University of Belgrade, Belgrade, Serbia

^bComputational Systems Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia

^cDepartment for Medical Statistics and Informatics, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

*Corresponding author: e-mail address: dmarko@bio.bg.ac.rs

Contents

1. Introduction	2
2. An overview of compartmental models of epidemic progression	5
3. Systems biology approach to compartmental modeling of the COVID-19 epidemic	7
3.1 Virus transmission in the early stages of epidemics	10
4. Parameter analysis and inference	11
5. Analysis of COVID-19 transmission in China	13
6. Conclusions	21
Acknowledgment	22
References	22

Abstract

A number of models in mathematical epidemiology have been developed to account for control measures such as vaccination or quarantine. However, COVID-19 has brought unprecedented social distancing measures, with a challenge on how to include these in a manner that can explain the data but avoid overfitting in parameter inference. We here develop a simple time-dependent model, where social distancing effects are introduced analogous to coarse-grained models of gene expression control in systems biology. We apply our approach to understand drastic differences in COVID-19 infection and fatality counts, observed between Hubei (Wuhan) and other Mainland China provinces. We find that these unintuitive data may be explained through an interplay of differences in transmissibility, effective protection, and detection efficiencies between Hubei and other provinces. More generally, our results demonstrate that regional differences may drastically shape infection outbursts. The obtained results demonstrate the applicability of our developed method to extract key infection parameters directly from publically available data so that it can be globally applied to outbreaks of COVID-19

in a number of countries. Overall, we show that applications of uncommon strategies, such as methods and approaches from molecular systems biology research to mathematical epidemiology, may significantly advance our understanding of COVID-19 and other infectious diseases.



1. Introduction

As the novel COVID-19 disease caused by the SARS-CoV-2 virus took the world by a storm, the new pandemic quickly gained priority in scientific research in a wide range of biological and medical science disciplines. Despite that their prior expertise was in unrelated research fields, many researchers have successfully adapted their approaches and methods to examine various aspects of this viral infection and, thus, contributed to finding the necessary solutions. The systems biology community is not an exception (Alon, Mino, & Yashiv, 2020; Bar-On, Flamholz, Phillips, & Milo, 2020; Djordjevic, Djordjevic, Ilic, Stojku, & Salom, 2021; Eilersen & Sneppen, 2020; Karin et al., 2020; Saad-Roy et al., 2021; Vilar & Saiz, 2020; Wong et al., 2020): those involved in modeling the dynamics of biological systems at the molecular and cellular level can directly apply the similar methodology in epidemiological studying of the virus spread—and this exactly is the central point of the present paper. In particular, dynamic models of biochemical reaction networks, in which the reaction kinetics follow the law of mass action, are analogous to compartmental epidemiological models which, instead of concentrations of chemical species, track the prevalence of individuals in defined population classes over time (Voit, Martens, & Omholt, 2015). Moreover, gene expression dynamics is usually a result of the interplay between the changing rate of cell growth, on which the global physiological rates of molecule synthesis and degradation depend, and complex transcription regulation (Djordjevic, Rodic, & Graovac, 2019). Therefore, modeling dynamics of gene circuits implies combining kinetic models, often relying on the law of mass action, with appropriate non-linear functions describing the regulation part. In the case of the COVID-19 epidemic, one can note that the virus transmission in a population, driven by the biological capacity of the particular virus in the given environment, is coupled with strong, time-dependent regulation, represented by the epidemic mitigation measures imposed by governments. These similarities between the modeled systems may facilitate the application of

the systems biology techniques to the epidemiology field of research. In this paper, we will show how such an approach can be used to assess the basic parameters of the COVID-19 epidemic progression in a given population. In particular, we will use the analogy outlined above to study the COVID-19 spread in Mainland China and test the hypothesis about possible reasons for the uneven disease spread in China provinces.

Our interest in Mainland China infection progression comes from Fig. 1. The progression seems highly intriguing, as Hubei (with only 4% of China population) shows an order of magnitude larger number of detected infection cases (Fig. 1A) and two orders of magnitude higher fatalities (Fig. 1B) compared to the *total* sum in all other Mainland China provinces. The epidemic was unfolding well before the Wuhan closure (with the reported symptom onset of the first patient on December 1, 2019) and within the period of huge population movement, which started 2 weeks before January 25 (the Chinese Lunar New Year) (Chen, Yang, Yang, Wang, & Bärnighausen, 2020). As a rough baseline, a modeling study of the infection spread from Wuhan (Wu, Leung, & Leung, 2020) estimated more than 10^5 new cases per day in Chongqing alone—instead, the actual (reported) peak number for *all* Mainland China provinces outside Hubei was just 831.

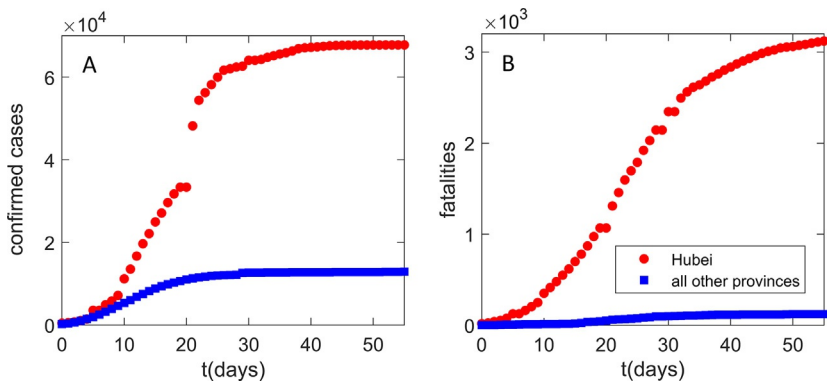


Fig. 1 Infection and fatality counts for Hubei vs all other provinces. The number of (A) detected infections, (B) fatality cases. Zero on the horizontal axis corresponds to the time from which the data (Hu et al., 2020) are taken (January 23), which also coincides with the Wuhan closure. Red circles correspond to the observed Hubei counts. Blue squares correspond to the sum of the number of counts for all other provinces. The figure illustrates a puzzling difference in the number of counts between Hubei alone and the sum of all other Mainland China provinces.

Consequently, it is a notable challenge for computational modeling to understand drastic differences in COVID-19 infection and fatality counts observed between Hubei (Wuhan) and other Mainland China provinces. These drastic differences may be a consequence of an interplay between the virus transmissibility (influenced by environmental and demographic factors) and the effectiveness of the protection measures. Both can significantly change between different provinces (more generally different countries/regions), and the model has to infer this from available data (commonly the number of confirmed cases, publicly available for a large number of countries/regions).

The study presented here will therefore demonstrate the usefulness of the systems biology approach to the analysis of non-trivial COVID-19 data from China. In particular, the developed method will allow us to analyze the puzzling differences in dynamics trajectories in Mainland China provinces, and it will also turn out to be more generally applicable for understanding regional differences in outburst dynamics. The surprising differences in COVID-19 progression in different provinces may put strong constraints on the underlying infection progression parameters and allow us to understand:

- i. What interplay between the inherent disease transmissibility and the effects of social distancing is responsible for the large difference in the count numbers between Hubei and the rest of Mainland China? Addressing this question in a proper way would make easier to comprehend how regional differences may shape the infection outbursts, which is important both locally (for explaining this puzzle), and more generally in the context of global COVID-19 pandemics progression.
- ii. What is the Infected Fatality Rate (IFR, the number of fatalities per total number of *infected* cases) in China? Case Fatality Rate (CFR, the number of fatalities per *confirmed/detected* cases) can be obtained directly from the data but is highly sensitive to the testing coverage. IFR is a more fundamental mortality parameter, as it does not depend on the testing coverage, but is however much harder to determine, due to the unknown number of infected cases.

Addressing these questions allows understanding both the different response policies, and the inherent risks posed by the pandemics and will enable future cross-country comparisons. The developed methodology (i) demonstrates the usefulness of applying transdisciplinary expertise to efficiently analyze problems of nationwide importance, (ii) allows to readily analyze future

outbreaks of COVID-19 and other infectious diseases, as it depends only on inference from straightforward and publically available data.



2. An overview of compartmental models of epidemic progression

In epidemiology, for practical and ethical reasons, it is fairly impossible to conduct scientific experiments in controlled conditions in order to investigate the spread of the disease in the human population (Brauer, 2008). Therefore, epidemiologists usually resort to collecting data from clinical reports on the observed situation in the field and, then, using mathematical models to interpret these data, i.e., to infer the principles underlying the process of disease spreading. These principles may point to potentially successful control strategies, as well as to the probable future status of the disease in the population. Epidemiological data can often be incomplete or inaccurate due to poorly controlled or non-standardized collection methods, which significantly complicates modeling. However, even a qualitative agreement of the model with the data can provide useful information of great practical importance. Hence, model predictions are widely used for making various estimates and answering important questions about the seriousness of the epidemic consequences. For example, how many people will be infected, require treatment, or die, or how many patients should the public health facilities expect at any given time? Also, how long will the epidemic last? To what extent could quarantine and self-isolation of the infected contribute to mitigating the effects of the epidemic? Model predictions guide the development of strategies to control the epidemic spread, including vaccination programs.

When the goal is to discover the general principles of epidemic progression, simple mathematical models, which can be solved and analyzed with a “pencil on paper,” are a logical choice as they give insight into the properties of the examined process despite failing to reproduce it in detail. In 1927, Kermack and McKendrick formulated a simple model that predicted behavior similar to that observed in numerous epidemics (Kermack & McKendrick, 1927). It was a type of compartmental model describing the infection spread in a population by analogy with a system of vessels connected by pipes through which a fluid flows. Namely, the population is divided into compartments, and assumptions are made about the nature and the rate of the flow between them. The structure of the compartmental

model—which sections and how many of them it will contain and how they will be connected—depends on the characteristics of transmission of a given infectious disease and whether the past disease provides immunity to re-infections or not. The model set by these two scientists is known as SIR (from **S**usceptible–**I**nfected–**R**ecovered). It divides the population into three classes which correspond to compartments (Fig. 2): Susceptible (S) class includes healthy individuals susceptible to infection, which have never been exposed to the virus; Those who are infected and can infect others belong to the Infected (I) class; Recovered (R) class encompasses those who are excluded from the population, either by quarantining the infected, or by acquiring immunity through recovery from disease or immunization, or by the death of the infected (Brauer, 2008).

Mathematically, this model is represented by a system of ordinary differential equations. The time derivative of the number of individuals in a compartment, i.e., the rate of their change, is given by the difference between the rates at which the compartment is filled and emptied. Analogous to the processes in which chemical species (e.g., proteins) are degraded or converted into others within a biochemical reaction network (Ingalls, 2013), the rate of transition of individuals from one compartment to another follows the law of mass action. For example, a person moves from compartment S to compartment I at the rate which is proportional to the product of the S and I , as the encounter with an infected person enables virus transmission to the susceptible one (Voit et al., 2015).

By formulating such (or similar) models, one assumes that the epidemic is a deterministic process. Namely, the state of the population at all times is completely determined by its previous state and the rules described by the model. This is a reasonable approximation in cases where the numbers of individuals in the compartments are large, i.e., in a commonly considered



Fig. 2 Schematic representation of the SIR model. Rectangles denote model compartments containing susceptible (S), infected (I), and recovered (R) individuals in the population of size N . Permitted directions of flow between compartments are denoted by arrows, with the rates of flow indicated above them. The rates are expressed according to the law of mass action, where κ_1 and κ_2 are the rate constants. The dashed curve corresponds to bimolecular reaction, where newly infected are generated through interactions (contacts) between susceptible and already infected individuals.

deterministic range (>10). Such approximation (i.e., deterministic modeling) is well suited for the spread of COVID-19, which is up to now known for a large number of individuals in all compartments.



3. Systems biology approach to compartmental modeling of the COVID-19 epidemic

The above-introduced SIR model is likely the simplest compartmental model in mathematical epidemiology and many subsequent models are derivatives of this basic form. Among others, these extensions have also been developed toward including control measures such as vaccination or quarantine (Diekmann, Heesterbeek, & Britton, 2012; Keeling & Rohani, 2011; Martcheva, 2015). However, COVID-19 brought a challenge to account for previously unprecedented social distancing measures, taken by most countries. When included, these effects have been, up to now, accounted for by the direct changes in the transmissibility term (Chowell, Sattenspiel, Bansal, & Viboud, 2016; Tian et al., 2020), which, however, corresponds to introducing a phenomenological dependence in otherwise mechanistic models. That is, to be included consistently in the model, social distancing should move individuals from one compartment to the other, just as vaccination and quarantine are usually implemented. On the other hand, it is necessary to construct a minimal mechanistic model in terms of the ability to explain the data with the smallest number of parameters, so that relevant infection progression properties can be inferred without overfitting. With this goal in mind, we used our systems biology background to develop a minimal model that accounts for all the main qualitative features of the SARS-CoV-2 infection spread under epidemic mitigation measures. As outlined above, we opt for a deterministic model due to the robust and computationally less demanding parameter inference (Wilkinson, 2018).

To describe the COVID-19 epidemic, we developed SPEIRD model depicted schematically in Fig. 3. It assumes that healthy persons susceptible to infection (S), can be infected, but in the case of this (and many other) viruses they do not immediately become contagious to other people, but first spend some time in the compartment E (Exposed to the virus) and then develop symptoms and pass to the compartment I . Infected persons can either recover at home, moving to the compartment R , or they can be diagnosed with SARS-CoV-2 virus infection (Active detected cases). A (Active) cases can, further, either become healed (H) or die from the disease (F). To consistently implement the social distancing within this model structure, we

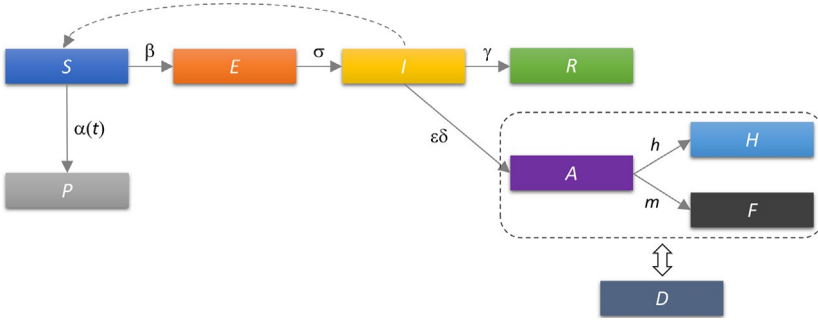


Fig. 3 Schematic representation of the SPEIRD model. Compartments and the transition rates are as indicated in the text, where transitions between different compartments are marked by arrows. The time-dependent transition rate from susceptible to protected category $\alpha(t)$ is indicated by the solid arrow. The infected can transition to the recovered category either without being diagnosed (transition to R), or being diagnosed and then transitioning to confirmed healed or fatality cases. The dashed rectangle indicates that A , H , and F categories in the starting model are substituted for the cumulative case counts (D), which removes h and m from the analysis, where D is fitted to the observed data.

included a compartment P (**P**rotected) in the model, which contains susceptible persons who are protected from exposure to the virus as a result of the epidemic mitigation measures, such as self-imposed isolation, social distancing, and advised changes in individual behavior.

The following differential equations describe how different categories change with time:

$$dS/dt = -\beta \cdot I \cdot S/N - \alpha(t) \cdot S \quad (1)$$

$$dE/dt = \beta \cdot I \cdot S/N - \sigma \cdot E \quad (2)$$

$$dI/dt = \sigma \cdot E - \gamma \cdot I - \varepsilon \cdot \delta \cdot I \quad (3)$$

$$dA/dt = \varepsilon \cdot \delta \cdot I - h \cdot A - m \cdot A \quad (4)$$

$$dH/dt = h \cdot A \quad (5)$$

$$dF/dt = m \cdot A \quad (6)$$

where β is the infection rate in a fully susceptible population; $\alpha(t)$, the time-dependent protection rate, i.e., the rate at which the population moves from susceptible to the protected category, quantifying the impact of the social protection measures; σ , the inverse of the exposed period; γ , the inverse of the infectious period; δ , the inverse of the period of the infection diagnosis; ε , the detection efficiency; h , the healing rate of diagnosed cases; m , the mortality rate.

The probability that an infected person will meet a susceptible person is proportional to S/N , where N is the total number of individuals in the population. The rate at which individuals move from S to E is obtained when the product of I and S/N is multiplied by the infection rate, β , which quantifies the efficiency of transmission of a particular virus in the population with certain demographic characteristics and meteorological conditions, and it does not depend on epidemic suppression measures. Thus, β is a characteristic of the virus, the population, and the external conditions in which the virus is transmitted. Since the compartment S is being emptied, the corresponding rate in the first equation is specified with the minus sign.

S also decays by moving the individuals to P with a protection rate that may vary with time. While mitigation measures are commonly accounted for by models with time-independent terms (Martcheva, 2015), we note that the social distancing term should depend on time, as this measure is introduced at a certain point in epidemics and may also evolve gradually. We denote the time point (more specifically, the date) of the onset of the social distancing measures in the examined population with t_0 . The protection rate $\alpha(t)$ is then taken as 0 before t_0 and a constant value α afterwards.

One may notice a direct parallel between the model outlined above, and e.g., modeling gene expression regulation in systems biology with a step function that approximates the activity of a promoter to which repressor proteins are highly cooperatively bound: the promoter is initially silenced and upon receiving a signal which leads to the abrupt removal of repression, promoter activity rises sharply to its maximum value. We notice that the step function is a satisfactory approximation of the dynamics of social distancing, i.e., it may not be necessary to further increase the number of parameters by applying the Hill function (which describes a more gradual activation), since governments quickly introduced these measures, together with their effective implementation. Note however that in (Djordjevic et al., 2021) we introduced a more complex model with Hill function, and provided analytical results for key properties of this model.

Compartment E is filled by infecting the susceptibles and emptied by moving the individuals to I , with the rate σ representing the inverse value of the latent period during which the person is not contagious. While compartment I is filled with individuals from E , it is depleted through two channels. Individuals move to R with the rate γ , which is the inverse of the period of contagiousness, and to A with the rate δ , which is the inverse of the time required for diagnosis, multiplied by ϵ , reflecting that only a fraction (likely small due to many asymptomatic infections) of the total infected are

detected. Note that case detection reduces the number of individuals in I that can infect susceptibles: the model assumes that the detected cases are quarantined and thus isolated from the general population. The numbers in compartments A , H , and F change following the same logic described for the other compartments.

We can further simplify the analysis by looking at the total number of detected cases (D), which is the sum of A , H , and F . By adding the Eqs. (4)–(6), we obtain:

$$dD/dt = \varepsilon \cdot \delta \cdot I, \quad (7)$$

and thus lose two parameters, h and m . The total number of detected cases in time is a measurable quantity from which we can determine the dynamics of other model compartments since this is the data that is available for various different regions and countries. Thereby, we assume that before t_0 social distancing does not take effect, and the measures introduced at t_0 will take effect on $D \sim 10$ days later, as this is about the time that elapses between infection and detection/diagnosis (Feng et al., 2020). Consequently, for the first $t_0 + 10$ days, the D curve reflects disease transmission without epidemic suppression measures.

3.1 Virus transmission in the early stages of epidemics

We will now focus on the dynamics of the infection spread at the very beginning of the epidemic, i.e., on the period before the introduction and practice of any control measures (Salom et al., 2021). Regarding the model, we assume that there is no social distancing (no transition from S to P), there is no quarantine, and almost the entire population consists of people susceptible to infection, so $S/N = 1$. This gives us an even simpler mathematical model which appears to be very useful because it allows analytical derivation of the expressions we need. Our system of Eqs. (1)–(3) and (7) is reduced to two linear differential equations that we can write in matrix form

$$\frac{d}{dt} \begin{pmatrix} E \\ I \end{pmatrix} = \begin{pmatrix} -\sigma & \beta \\ \sigma & -\gamma \end{pmatrix} \begin{pmatrix} E \\ I \end{pmatrix} = A \begin{pmatrix} E \\ I \end{pmatrix}, \quad (8)$$

determine the eigenvalues and eigenvectors of the matrix and, subsequently, the solutions of the system, $E(t)$ and $I(t)$. Specifically, the cumulative number of infected in time, $I(t)$, is obtained according to the following equation:

$$I(t) = C_1 \exp(\lambda_+ t) + C_2 \exp(\lambda_- t), \quad (9)$$

where λ_s are eigenvalues of the matrix. Since one of the eigenvalues, here denoted by λ_- , is negative, the corresponding term of the Eq. (9) will decrease over time, and $I(t)$ will be effectively described by the first term, already after few days from the epidemic outbreak (Salom et al., 2021). We can further derive this equation for the dependence of the logarithm of the number of detected cases in time:

$$\log(D(t)) = \log(\varepsilon \cdot \delta \cdot I(0)/\lambda_+) + \lambda_+ \cdot t \quad (10)$$

This is the straight line equation whose slope is given by the value of λ_+ (the dominant, positive eigenvalue of the matrix in Eq. (8)).

Once we know λ_+ , we can calculate the value of the so-called basic reproduction number, $R_{0,free}$, by fixing mean values of the latency period and the infectivity period ($\gamma = 0.4 \text{ days}^{-1}$, $\sigma = 0.2 \text{ days}^{-1}$), which are known from the literature and characterize the fundamental infection processes (Kucharski et al., 2020; Li et al., 2020):

$$R_{0,free} = \frac{\beta}{\gamma} = 1 + \frac{\lambda_+(\gamma + \sigma) + \lambda_+^2}{\gamma\sigma} \quad (11)$$

$R_{0,free}$ is an important epidemiological parameter that characterizes the inherent biological transmission of the virus in a completely unprotected population. In particular, it is the mean number of secondarily infected by one infected person introduced in a completely susceptible population. It depends on the biology of the specific virus, as well as the demographic characteristics of the population and the environmental conditions, while it does not depend on the applied infection control measures (Brauer, 2008). In Salom et al. (2021) we utilized a bioinformatics analysis, akin to those often used to understand complex data in systems biology, to pinpoint demographic and meteorological factors that affect $R_{0,free}$ (i.e., inherent virus transmissibility in population). This furthermore underlines that a rich array of techniques developed and/or widely used within systems biology can be successfully employed within infectious disease modeling.



4. Parameter analysis and inference

$R_{0,free}$, α , t_0 , two initial conditions (I_0 and E_0), and the detection efficiency ε , are unknown and may differ between the provinces. Is it possible to determine these unknown parameters from different properties of the D curve? Early in the infection, almost the entire population is susceptible ($S \approx N$), so Eqs. (2) and (3) become linear, and decoupled from the rest of the

system, as discussed in the previous section. This sets the ratio of I_0 to E_0 , through the eigenvector components with the dominant (positive) eigenvalue of the Jacobian for this subsystem. This eigenvalue, corresponding to the initial slope of the $\log(D)$ curve, sets the value of λ_+ and subsequently, of $R_{0,\text{free}}$ (see Eq. 11). From Eq. (7) one can see that the product of I_0 and $\varepsilon \cdot \delta$ is set by dD/dt at the initial time ($t=0$). Later dynamics of the D curve is determined solely by the combination $t_\alpha = t_0 + 1/\alpha$ (which we denote as protection time), setting the time at which $\sim 1/2$ of the population moves to the protected category. We also numerically checked this, and confirmed that t_0 can be lowered at the expense of increasing $1/\alpha$, without affecting the fit quality. We allowed for t_0 to vary in reasonable proximity of January 23, as the social distancing was generally introduced close to Wuhan closure (e.g., on that date, all major events in Beijing were canceled) (Chen et al., 2020; Du et al., 2020), but we cannot be sure when the measures effectively took place. Our inferred t_0 values are within a week from Wuhan closure, appearing as reasonable. The remaining independent parameter (I_0) is then left to be determined from D curve properties at the late infection stage, such as its saturation time. The number of characteristic dynamics features is thus at least equal to the number of fit parameters, leading to constrained numerical analysis, so that overfitting is not expected. For few provinces, we however observed that I_0 can be decreased compared to the best fit value, without noticeably affecting the fit quality. For these provinces, we chose the lowest I_0 value that still leads to a comparably good fit. This allows obtaining the most conservative (i.e., as high as possible while still consistent with data) IFR estimate, as the reported fatality counts for provinces other than Hubei is surprisingly low.

Parameter inference and uncertainties are estimated separately for each province. However, within a given province, demographic, special, or population activity (network effects) heterogeneities (Britton, Ball, & Trapman, 2020; Diekmann et al., 2012), or seasonality effects (Wong et al., 2020), are not taken into account. These are potentially important, particularly for projections (longer-term predictions of infection dynamics under different scenarios), and can be readily included in our model. Such extensions would however complicate parameter inference, due to an increase in parameter number, as this may either lead to overfitting or require special/additional data that may be available only for a limited number of countries/regions (which would limit the generality of our proposed method). A more complex model structure may also obscure a straightforward relationship between the model parameters and distinct dynamical features of the

confirmed case count curve analyzed above. While the inclusion of additional effects is left for future work, we here employ the model structure and parameter inference introduced above on widely available case count data, as proof of the principle for the generality of our proposed approach. Moreover, a major advantage of our approach is that it allows consistent analysis for all provinces with the same model, numerical procedure, and parameter set, allowing an objective comparison of the obtained results.

Our model was numerically solved by the Runge-Kutta method (Dormand & Prince, 1980) for each parameter combination. Parameter values were inferred by exhaustive search over a wide parameter range, to avoid reaching a local minimum of the objective function (R^2). To infer the unknown parameters, we fit (by minimizing R^2) the model to the observed total number of detected D for each province. As an alternative to exhaustive search, some of many optimization techniques used in epidemics modeling, such as the Markov chain Monte Carlo (MCMC) approach, can be used instead (Keeling & Rohani, 2011; Wong et al., 2020)—exhaustive search is however straightforward, guarantees that the global minimum is reached, and is in this case not computationally demanding. Errors were estimated through Monte-Carlo simulations (Press, Flannery, Teukolsky, & Vetterling, 1986), individually for each province with the assumption that count numbers follow the Poisson distribution. Monte-Carlo simulations were found as the most reliable estimate of the fit parameter uncertainties for a non-linear fit (Cunningham, 1993). This also serves as an independent check for overfitting, as in that case, data point perturbations would lead to large parameter uncertainties. We find no indication of this in the results reported below, as the inferred uncertainties (consistently indicated with all results) are reasonably small. In particular, the differences in the inferred parameter values, which are relevant for the reported results/conclusions, are statistically highly significant. P values for extracted parameter differences between provinces are estimated by the t -test.



5. Analysis of COVID-19 transmission in China

We used our SPEIRD model with the parameter inference described above, to analyze all Mainland China provinces, except Tibet, where only one COVID-19 case was reported. Parameters were estimated separately for each of the 30 provinces by the same model and parameter set, which enables an impartial comparison of the results presented below. To allow

for a straightforward comparison of the infection progression between different provinces, the starting date (i.e., $t=0$) in our analysis is the same for all the provinces and corresponds to January 23 (when the data for all the provinces became publically available and continuously tracked (Hu et al., 2020)).

In Fig. 4A and B, we show that our model can robustly explain the observed D , in the cases of large outburst (Hubei on Fig. 4A), as well as for all other provinces, where D is in the range from intermediate (e.g., Guangdong) to low (e.g., Inner Mongolia). Provinces in Fig. 2B were selected to cover the entire range of observed D (from lower to higher counts), while comparably good fits were obtained for other provinces, which were all included in the further analysis. Our method is also robust to data perturbations (which might be frequent), e.g., in the case of Hubei (Wuhan), a large number of counts was added on February 12, based on clinical diagnosis (CT scan) (Feng et al., 2020), which is apparent as a discontinuity in observed D in Fig. 4A. The model however interpolates this discontinuity, finding a reasonable description of the overall data.

We backpropagated the dynamics inferred for Hubei, to estimate that January 5 (± 4 days) was the onset of the infection's exponential growth in the population (not to be confused with the appearance of first infections, which likely happened in December (Feng et al., 2020)). This agrees well with (Feng et al., 2020) (cf. Fig. 3A), which tracked cases according to their symptom onset (shifted for ~ 12 days with respect to detection/diagnosis, cf. Fig. 3B), and coincides with WHO reports on social media that there is a cluster of pneumonia cases—with no deaths—in Wuhan (WHO, 2020). Since our analysis does not directly use any information before January 23, this agreement provides confidence in our I_0 estimate. Note that we infer I_0 separately for each province of interest, through which we also take into account different times of the infection onset in different provinces (so that earlier onset time would generally lead to a larger number of infected on January 23).

Key parameters inferred from our analysis are summarized in Fig. 4C–F, with individual results and errors for all the provinces shown in Table 1. Fig. 4C shows the distribution of $R_{0,\text{free}}$. Note that $R_{0,\text{free}}$ might depend on demographic (population density, etc.) and climate factors (temperature, humidity...), which are not controllable, but are unrelated to the applied social distancing measures (see above). It is known that the R_0 value can strongly depend on the model, e.g., the number of introduced compartments (Keeling & Rohani, 2011); accordingly, a wide range of R_0 values

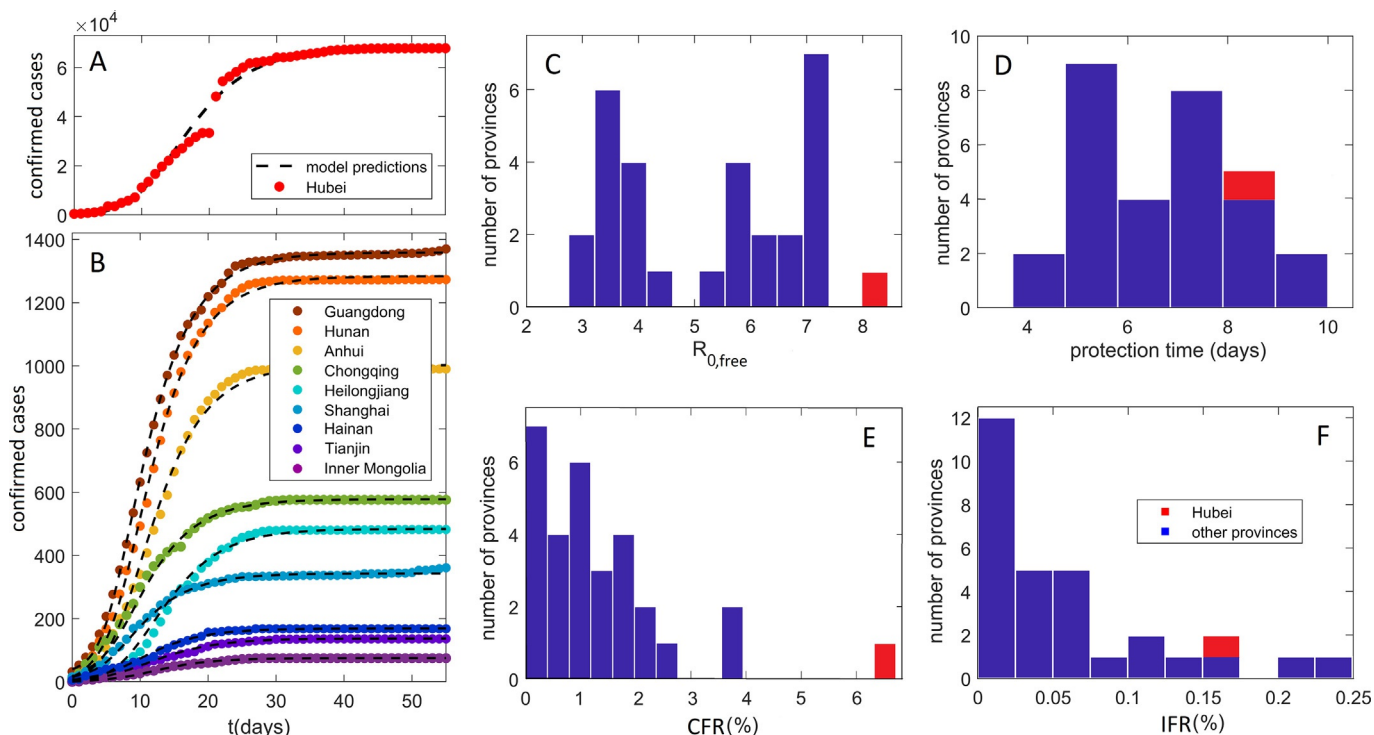


Fig. 4 Model predictions: comparison with data and key parameter estimates. Predictions (compared to data) of detected infection counts for (A) Hubei, (B) other Mainland China provinces. Zeros on the horizontal axis correspond to January 23, which is the initial time in our numerical analysis for all the provinces. The observed counts are shown by dots and our model predictions by dashed curves. Names of the provinces are indicated in the legend, with provinces selected to cover the full range of the observed total detected counts. The distribution with respect to provinces of (C) the basic reproduction number in the absence of social distancing, $R_{0,free}$, (D) the protection time t_p , (E) Case Fatality Rate, calculated directly from the reported data. (F) Infected Fatality Rate. The values for Hubei are indicated by the red bars.

Table 1 Inferred COVID-19 infection progression parameters for Mainland China provinces

Province	t_α (days)	R_0	E_0	I_0	IFR (%)	CFR (%)	Detected (%)
Anhui	6.6 ± 0.5	5.5 ± 0.8	920 ± 30	220 ± 20	0.04 ± 0.02	0.6 ± 0.3	6 ± 3
Beijing	7.9 ± 0.5	3.5 ± 0.4	610 ± 20	180 ± 10	0.12 ± 0.05	1.7 ± 0.7	7 ± 3
Chongqing	7.0 ± 0.2	3.5 ± 0.2	1900 ± 40	560 ± 20	0.04 ± 0.03	1.0 ± 0.5	4 ± 2
Fujian	3.7 ± 0.4	7 ± 2	1660 ± 40	360 ± 20	0.007 ± 0.003	0.3 ± 0.4	2 ± 1
Gansu	5 ± 1	6 ± 3	630 ± 20	150 ± 10	0.03 ± 0.04	1 ± 1	2 ± 3
Guangdong	5.0 ± 0.1	7 ± 1	1360 ± 40	290 ± 20	0.04 ± 0.01	0.6 ± 0.2	7 ± 2
Guangxi	7 ± 1	3.8 ± 0.8	1000 ± 30	290 ± 20	0.02 ± 0.02	0.8 ± 0.6	3 ± 3
Guizhou	8.1 ± 0.6	7 ± 1	53 ± 7	11 ± 3	0.06 ± 0.03	1 ± 1	4 ± 2
Hainan	7.6 ± 0.8	3.3 ± 0.7	300 ± 20	90 ± 10	0.21 ± 0.09	4 ± 2	6 ± 3
Hebei	6.0 ± 0.6	7 ± 2	240 ± 20	52 ± 7	0.11 ± 0.03	1.8 ± 0.8	6 ± 2
Heilongjiang	7 ± 1	6 ± 2	260 ± 20	59 ± 7	0.15 ± 0.07	2.9 ± 0.9	5 ± 3
Henan	7.0 ± 0.3	4.5 ± 0.5	1780 ± 40	460 ± 20	0.09 ± 0.04	1.7 ± 0.4	5 ± 2
Hubei	8.3 ± 0.2	8.2 ± 0.4	$31,900 \pm 400$	6600 ± 200	0.15 ± 0.09	6.5 ± 0.1	2 ± 2
Hunan	5.1 ± 0.1	6.8 ± 0.8	1430 ± 40	310 ± 20	0.02 ± 0.01	0.4 ± 0.2	5 ± 2
I. Mongolia	10.0 ± 0.8	2.8 ± 0.4	940 ± 30	300 ± 20	0.01 ± 0.03	1 ± 1	1 ± 3
Jiangsu	5.5 ± 0.5	7 ± 2	500 ± 20	110 ± 10	0 ± 0	0 ± 0	6 ± 2

Jiangxi	7.0 ± 0.2	5.6 ± 0.9	890 ± 30	210 ± 10	0.005 ± 0.002	0.1 ± 0.1	5 ± 2
Jilin	10.0 ± 0.7	4.0 ± 0.8	270 ± 20	76 ± 9	0.02 ± 0.02	1 ± 1	1 ± 2
Liaoning	7 ± 1	2.9 ± 0.7	1240 ± 40	390 ± 20	0.02 ± 0.04	2 ± 2	1 ± 2
Ningxia	5.3 ± 0.9	7 ± 3	72 ± 9	15 ± 4	0 ± 0	0 ± 0	6 ± 23
Qinghai	6.1 ± 0.6	4.0 ± 0.5	2260 ± 50	640 ± 30	0 ± 0	0 ± 0	0 ± 2
Shaanxi	5.2 ± 0.5	6 ± 1	380 ± 20	90 ± 10	0.07 ± 0.03	1.3 ± 0.8	6 ± 2
Shandong	9 ± 1	3.5 ± 0.5	900 ± 30	260 ± 20	0.06 ± 0.01	1.0 ± 0.4	6 ± 1
Shanghai	5.0 ± 0.4	6 ± 1	1570 ± 40	370 ± 20	0.02 ± 0.02	0.8 ± 0.5	2 ± 3
Shanxi	5.2 ± 0.5	6 ± 2	1600 ± 40	370 ± 20	0 ± 0	0 ± 0	1 ± 2
Sichuan	7.7 ± 0.8	3.7 ± 0.5	990 ± 30	280 ± 20	0.03 ± 0.02	0.6 ± 0.3	5 ± 3
Tianjin	7 ± 2	4 ± 2	170 ± 10	46 ± 7	0.14 ± 0.06	2 ± 1	7 ± 3
Xinjiang	7.3 ± 0.9	6 ± 1	42 ± 7	10 ± 3	0.25 ± 0.09	3 ± 2	8 ± 2
Yunnan	4.0 ± 0.2	7 ± 2	360 ± 20	76 ± 9	0.06 ± 0.03	1.2 ± 0.9	5 ± 2
Zhejiang	5.0 ± 0.1	7.2 ± 0.8	1340 ± 40	290 ± 20	0.005 ± 0.002	0.1 ± 0.1	7 ± 3

t_{gr} , protection time; $R_{0,free}$, basic reproduction number; E_0 , initial exposed; I_0 , initial infected; IFR , Infected Fatality Rate; CFR , Case Fatality Rate; *detected* %, fraction of the infected population that has been detected. Error of the quantities correspond to one standard deviation.

were reported for China in the literature (Sanche et al., 2020; Wu, Leung, Bushman, et al., 2020). Consequently, a clear advantage of our study is that parameters for all China provinces were determined from the same model and data set, which allows direct comparisons. Our obtained average $R_{0,\text{free}}$ for provinces outside of Hubei is 5.3 ± 0.3 , in a reasonable agreement with a recent estimate (≈ 5.7) (Sanche et al., 2020). Furthermore, we observe that $R_{0,\text{free}}$ for Hubei is a far outlier with a value of 8.2 ± 0.4 , which is notably larger than for other provinces with $p \sim 10^{-11}$. This then strongly suggests that demographic and climate factors that determine $R_{0,\text{free}}$, played a decisive role in a large outburst in Hubei vs other provinces, which we further address below.

The distribution of protection time t_α for the provinces is shown in Fig. 4D, with the value for Hubei indicated in red. The mean for the other provinces is 6.6 ± 0.2 days. That is, we observe that the suppression measures were efficiently implemented, with $\sim 1/2$ of the population moving to the protected category within a week from Wuhan closure. The protection time for Hubei of 8.3 ± 0.2 days was longer, which is statistically significant at the $p \sim 10^{-11}$ level. The estimated less efficient protection in the case of Hubei may also be an important contributing factor in the surprising difference in Hubei vs other provinces, which we further investigate below.

CFR distribution, based on the fatality numbers reported for Hubei and other provinces is shown in Fig. 4E. These numbers are not based on the model predictions, i.e., can be straightforwardly obtained by dividing the total number of fatalities by the total number of detected cases. CFR for other provinces with a mean of $1.2 \pm 0.4\%$ is significantly smaller compared to CFR for Hubei, which was 4.6% before the correction on April 17, and 6.5% after the correction (with 1290 fatalities added to Wuhan). This large difference in CFR between Hubei and other provinces further accentuates the differences noted in Fig. 1.

IFR is harder to determine than CFR, as a majority of COVID-19 infections correspond to asymptomatic or mild cases that are by large not diagnosed (Day, 2020). We consequently calculate IFR as the total number of fatalities divided by the total number of infections (cumulative incidence) for the entire outburst, where cumulative incidence is estimated from our model. As the infections precede fatalities, both the total number of fatalities and the cumulative incidence in our estimate correspond to the entire outburst, so that all the infections had a sufficient time to recover or lead to fatalities—this is directly feasible for the provinces in China, where all detected case counts reached saturation. Note that IFR calculated in this

way corresponds to an averaged quantity so that it does not capture possible time-dependent change over the outburst interval (in fact, for Wuhan it is known that the fatality rate was larger at the very beginning of the outburst). Nevertheless, the estimated IFR's present a reasonable measure of COVID-19 mortality across China provinces.

IFR distribution, which provides a much less biased measure of the infection mortality, is shown in Fig. 4F. In distinction to CFR, estimated IFR shows a much smaller difference between Hubei ($0.15 \pm 0.09\%$) and other provinces ($0.056 \pm 0.007\%$). Therefore, while Hubei is a clear outlier with respect to CFR, we observe similar IFR values for all Mainland China provinces, where few provinces have even higher IFR than Hubei. The ratio of IFR to CFR equals the fraction of all infected that got detected (*detection coverage*). We estimate that the mean detection coverage for all provinces except Hubei is higher than detection coverage for Hubei ($4.5 \pm 0.9\%$ vs $2 \pm 2\%$). This difference is responsible for a decrease by a factor of two from CFR to IFR for Hubei, compared to the other provinces, and consequently for more uniform mortality estimates at the IFR level. Xinjiang has the highest IFR of $0.25 \pm 0.09\%$ so that Hubei is not an outlier anymore. Estimated IFR's of up to 0.3% in China provinces are in general agreement with the estimates reported elsewhere (see e.g., (Bar-On et al., 2020; Djordjevic et al., 2021; Mizumoto, Kagaya, & Chowell, 2020)).

In Fig. 5A, two key infection progression parameters are plotted against each other: protection time t_α vs basic reproduction number $R_{0,\text{free}}$. Unexpectedly, there is a high negative correlation, with Pearson correlation coefficient $R = -0.70$, which is statistically highly significant $p \sim 10^{-5}$, where these two are a priori unrelated (see above). Actually, stronger social distancing measures—which by definition are not included in $R_{0,\text{free}}$ —would lead to a decrease in *effective* transmissibility. This would then lead to a tendency of transmissibility to positively correlate with t_α , oppositely from the strong negative correlation observed in Fig. 5A. Therefore, higher basic reproduction number is genuinely related to a shorter protection time (larger effect of the suppression measures). Intuitively, this could be understood as a negative feedback loop, commonly observed in systems biology (Alon, 2019; Phillips, Kondev, Theriot, & Garcia, 2012), where larger $R_{0,\text{free}}$ leads to steeper initial growth in the infected numbers, which may elicit stronger measures and better observing of these measures by the population faced with a more serious outbreak. Interestingly, similar negative feedback was also obtained in the context of epidemics research other than COVID-19 (Wang, Andrews, Wu, Wang, & Bauch, 2015).

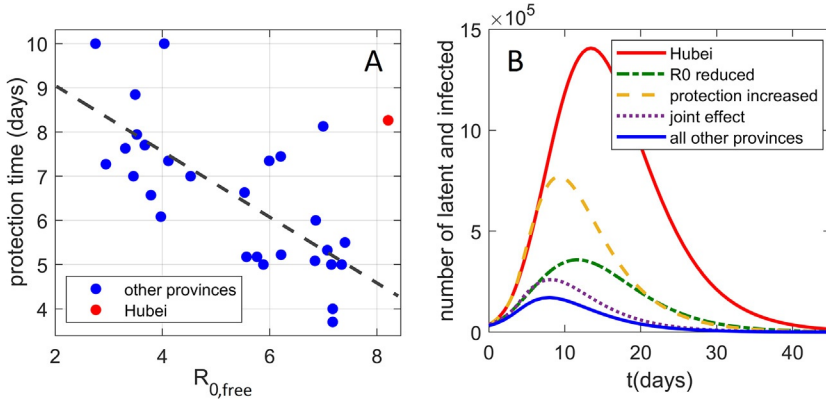


Fig. 5 The interplay of transmissibility and effective social distancing. (A) The correlation plot of t_α vs $R_{0,free}$ for all provinces, where the point corresponding to Hubei is marked in red. (B) The effect (on the Hubei dynamics of infected and latent cases) of reducing $R_{0,free}$ and t_α to the mean values of other Mainland China provinces. Both the unperturbed Hubei dynamics and the sum of infected and latent cases for all other provinces are included as references.

The two main properties of the Hubei outburst are therefore higher $R_{0,free}$ and t_α compared to other provinces. In Fig. 5B, we investigate how these two properties separately affect the Wuhan outburst for latent and infected cases, where unperturbed Hubei dynamics is shown by the red full curve. We first reduce only $R_{0,free}$ from the Hubei value, to the mean value for all other provinces (the dash-dotted green curve). We see that this reduction substantially lowers the peak of the curve, though it still remains wide. Next, instead of decreasing $R_{0,free}$, we decrease the protection time t_α to the mean value for all other provinces (dashed orange curve). While reducing t_α also significantly lowers the peak of the curve, its main effect is in narrowing the curve, i.e., reducing the outburst time. Finally, when $R_{0,free}$ and t_α are jointly reduced, we obtain the (dotted purple) curve that is both significantly lower and narrower than the original Hubei progression. This curve comes quite close to the curve that presents the sum of all other provinces (full blue curve)—the dotted curve remains somewhat above this sum, mainly because the initial number of latent and infected cases is somewhat higher for Hubei compared to the sum of all other provinces. This synergy between the transmissibility and the control measures will be further discussed below.



6. Conclusions

In this study, we applied a systems biology approach to develop a novel method of COVID-19 transmission dynamics. The model includes (time-dependent) social distancing measures in a simple manner, consistent with the compartmental mechanistic nature of the underlying process. The model has a major advantage that it is independent of the specific transmission process considered, and requires only commonly available count data as an input. The model allows extracting key infection parameters from the data that are readily available and publicly accessible (both for China and other countries), so that, in a nutshell, our approach is of wide applicability. To our best knowledge, such parameters (necessary to assess any future COVID-19 risks), were not extracted by other computational approaches.

The developed method is subsequently applied to the problem that appears highly non-trivial, i.e., to understand the puzzle created by the drastic differences in the infection and fatality counts between Hubei and the rest of Mainland China. The goal was to determine if it is possible to consistently explain such drastic differences by the same model, and what are the resulting numerical estimates and conclusions. We found that Hubei was a suitable ground for infection transmission, being an outlier with respect to two key infection progression parameters: having significantly larger $R_{0,free}$, and a longer time needed to move a sizable fraction of the population from susceptible to a protected category. While stricter measures were formally introduced in Hubei, the initial phase of the outburst put a large strain on the system, arguably leading to less effective measures compared to other provinces.

The fact that the initial epidemic in Hubei was not followed by similar outbursts in the rest of Mainland China may be understood as a serendipitous interplay of the two factors noted above. While both smaller $R_{0,free}$ and lower half-protection time (more efficient measures) significantly suppress the infection curve, their effect is also qualitatively different. While lowering $R_{0,free}$ more significantly suppresses the peak, decreasing the half-protection time significantly reduces the outburst duration. Consequently, the synergy of these two effects appears to lead to drastically suppressed infection dynamics in other Mainland China provinces compared to Hubei. The number of detected (diagnosed) cases in the entire Mainland China is, therefore, though unintuitive, well consistent with the model, and is explainable by a seemingly reasonable combination of circumstances. Our obtained

negative feedback between transmissibility and effects of social distancing may be understood in terms of larger transmissibility triggering more stringent social distancing measures, where a similar conclusion was also obtained through entirely different means (a combination of real-time human mobility data and regression analysis) (Kraemer et al., 2020).

In summary, we showed that unintuitive dissimilarity in the infection progression for Hubei vs other Mainland China provinces is consistent with our model, and can be attributed to the interplay of transmissibility and effective protection, demonstrating that regional differences may drastically shape the infection outbursts. This also shows that comparisons in terms of the confirmed cases, or fatality counts (even when normalized for population size), between COVID-19 and other infectious diseases, or between different regions for COVID-19, are not feasible, and that parameter inference from quantitative models (individually for different affected regions) is necessary. Consequently, this paper illustrates that utilization of uncommon strategies, such as systems biology application to mathematical epidemiology, may significantly advance our understanding of COVID-19 and other infectious diseases.

Acknowledgment

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

- Alon, U. (2019). *An introduction to systems biology: Design principles of biological circuits*. CRC press.
- Alon, U., Mino, R., & Yashiv, E. (2020). 10–4: How to reopen the economy by exploiting the coronavirus's weak spot. *The New York Times*.
- Bar-On, Y. M., Flamholz, A., Phillips, R., & Milo, R. (2020). Science forum: SARS-CoV-2 (COVID-19) by the numbers. *eLife*, 9, e57309.
- Brauer, F. (2008). Compartmental models in epidemiology. In *Mathematical epidemiology* (pp. 19–79). Springer.
- Britton, T., Ball, F., & Trapman, P. (2020). A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, 369(6505), 846–849.
- Chen, S., Yang, J., Yang, W., Wang, C., & Bärnighausen, T. (2020). COVID-19 control in China during mass population movements at New Year. *The Lancet*, 395(10226), 764–766.
- Chowell, G., Sattenspiel, L., Bansal, S., & Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18, 66–97.
- Cunningham, R. W. (1993). Comparison of three methods for determining fit parameter uncertainties for the Marquardt Compromise. *Computers in Physics*, 7(5), 570–576.
- Day, M. (2020). Covid-19: Identifying and isolating asymptomatic people helped eliminate virus in Italian village. *British Medical Journal*, 368(m1165)(Online).

- Diekmann, O., Heesterbeek, H., & Britton, T. (2012). Mathematical tools for understanding infectious disease dynamics. In (Vol. 7). Princeton University Press.
- Djordjevic, M., Djordjevic, M., Ilic, B., Stojku, S., & Salom, I. (2021). Understanding infection progression under strong control measures through universal COVID-19 growth strategies. *Global Challenges*, 5, 2000101.
- Djordjevic, M., Rodic, A., & Graovac, S. (2019). From biophysics to 'omics and systems biology. *European Biophysics Journal*, 48(5), 413–424.
- Dormand, J. R., & Prince, P. J. (1980). A family of embedded Runge–Kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1), 19–26.
- Du, Z., Xu, X., Wang, L., Fox, S. J., Cowling, B. J., Galvani, A. P., et al. (2020). Effects of proactive social distancing on COVID-19 outbreaks in 58 cities, China. *Emerging Infectious Diseases*, 26(9), 2267.
- Eilersen, A., & Sneppen, K. (2020). Cost–benefit of limited isolation and testing in COVID-19 mitigation. *Scientific Reports*, 10(1), 1–7.
- Feng, Z., et al. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly*, 2(8), 113–122.
- Hu, T., Weihe, W. G., Zhu, X., Shao, Y., Liu, L., Du, J., et al. (2020). Building an open resources repository for COVID-19 research. *Data and Information Management*, 4(3), 130–147.
- Ingalls, B. P. (2013). *Mathematical modeling in systems biology: An introduction*. MIT Press.
- Karin, O., Bar-On, Y. M., Milo, T., Katzir, I., Mayo, A., Korem, Y., et al. (2020). Cyclic exit strategies to suppress COVID-19 and allow economic activity. *medRxiv* 2020.04.04.20053579.
- Keeling, M. J., & Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*, 115(772), 700–721.
- Kraemer, M. U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368(6490), 493–497.
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., et al. (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *The Lancet Infectious Diseases*, 20(5), 553–558.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, 382(13), 1199–1207.
- Martcheva, M. (2015). An introduction to mathematical epidemiology. In (Vol. 61) Springer.
- Mizumoto, K., Kagaya, K., & Chowell, G. (2020). Early epidemiological assessment of the transmission potential and virulence of coronavirus disease 2019 (COVID-19) in Wuhan City, China, January–February, 2020. *BMC Medicine*, 18(1), 1–9.
- Phillips, R., Kondev, J., Theriot, J., & Garcia, H. (2012). *Physical biology of the cell*. Garland Science.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical recipes: The art of scientific computing*. Cambridge: Cambridge University Press.
- Saad-Roy, C. M., Grenfell, B. T., Levin, S. A., Pellis, L., Stage, H. B., van den Driessche, P., et al. (2021). Superinfection and the evolution of an initial asymptomatic stage. *Royal Society Open Science*, 8(1), 202212.
- Salom, I., Rodic, A., Milicevic, O., Zigic, D., Djordjevic, M., & Djordjevic, M. (2021). Effects of demographic and weather parameters on COVID-19 basic reproduction number. *Frontiers in Ecology and Evolution*, 8, 617841.
- Sanche, S., Lin, Y. T., Xu, C., Romero-Severson, E., Hengartner, N., & Ke, R. (2020). High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerging Infectious Diseases*, 26(7), 1470–1477.

- Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M. U., et al. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, *368*(6491), 638–642.
- Vilar, J. M., & Saiz, L. (2020). The evolving worldwide dynamic state of the COVID-19 outbreak. *medRxiv*.
- Voit, E., Martens, H., & Omholt, S. (2015). 150 years of the mass action law. *PLoS Computational Biology*, *11*(1), e1004012.
- Wang, Z., Andrews, M. A., Wu, Z.-X., Wang, L., & Bauch, C. T. (2015). Coupled disease–behavior dynamics on complex networks: A review. *Physics of Life Reviews*, *15*, 1–29.
- WHO. (2020). *WHO timeline—COVID-19*. <https://www.who.int/news-room/detail/27-04-2020-who-timeline—covid-19>: World Health Organization.
- Wilkinson, D. J. (2018). *Stochastic modelling for systems biology* (3rd ed.). London: Chapman and Hall/CRC.
- Wong, G. N., Weiner, Z. J., Tkachenko, A. V., Elbanna, A., Maslov, S., & Goldenfeld, N. (2020). Modeling COVID-19 dynamics in Illinois under nonpharmaceutical interventions. *Physical Review X*, *10*(4), 041033.
- Wu, J. T., Leung, K., Bushman, M., Kishore, N., Niehus, R., de Salazar, P. M., et al. (2020). Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine*, *26*(4), 506–510.
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet*, *395*(10225), 689–697.



Effects of Demographic and Weather Parameters on COVID-19 Basic Reproduction Number

Igor Salom¹, Andjela Rodic², Ognjen Milicevic³, Dusan Zigic¹, Magdalena Djordjevic¹ and Marko Djordjevic^{2*}

¹ Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia,

² Quantitative Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia, ³ Department for Medical Statistics and Informatics, School of Medicine, University of Belgrade, Belgrade, Serbia

It is hard to overstate the importance of a timely prediction of the COVID-19 pandemic progression. Yet, this is not possible without a comprehensive understanding of environmental factors that may affect the infection transmissibility. Studies addressing parameters that may influence COVID-19 progression relied on either the total numbers of detected cases and similar proxies (which are highly sensitive to the testing capacity, levels of introduced social distancing measures, etc.), and/or a small number of analyzed factors, including analysis of regions that display a narrow range of these parameters. We here apply a novel approach, exploiting widespread growth regimes in COVID-19 detected case counts. By applying nonlinear dynamics methods to the exponential regime, we extract basic reproductive number R_0 (i.e., the measure of COVID-19 inherent biological transmissibility), applying to the completely naïve population in the absence of social distancing, for 118 different countries. We then use bioinformatics methods to systematically collect data on a large number of potentially interesting demographics and weather parameters for these countries (where data was available), and seek their correlations with the rate of COVID-19 spread. While some of the already reported or assumed tendencies (e.g., negative correlation of transmissibility with temperature and humidity, significant correlation with UV, generally positive correlation with pollution levels) are also confirmed by our analysis, we report a number of both novel results and those that help settle existing disputes: the absence of dependence on wind speed and air pressure, negative correlation with precipitation; significant positive correlation with society development level (human development index) irrespective of testing policies, and percent of the urban population, but absence of correlation with population density *per se*. We find a strong positive correlation of transmissibility on alcohol consumption, and the absence of correlation on refugee numbers, contrary to some widespread beliefs. Significant tendencies with health-related factors are reported, including a detailed analysis of the blood type group showing consistent tendencies on Rh factor, and a strong positive correlation of transmissibility with cholesterol levels. Detailed comparisons of obtained results with previous findings, and limitations of our approach, are also provided.

Keywords: COVID-19 transmissibility, environmental factors, basic reproduction number, COVID-19 demographic dependence, COVID-19 weather dependence

OPEN ACCESS

Edited by:

Matteo Convertino,
Hokkaido University, Japan

Reviewed by:

Essam Rashed,
Nagoya Institute of Technology, Japan
Subhas Khajanchi,
Presidency University, India
Kankan Sarkar,
Malda College, India

*Correspondence:

Marko Djordjevic
dmarko@bio.bg.ac.rs

Specialty section:

This article was submitted to
Models in Ecology and Evolution,
a section of the journal
Frontiers in Ecology and Evolution

Received: 15 October 2020

Accepted: 28 December 2020

Published: 25 January 2021

Citation:

Salom I, Rodic A, Milicevic O, Zigic D,
Djordjevic M and Djordjevic M (2021)
Effects of Demographic and Weather
Parameters on COVID-19 Basic
Reproduction Number.
Front. Ecol. Evol. 8:617841.
doi: 10.3389/fevo.2020.617841

INTRODUCTION

The ancient wisdom teaches us that “knowing your adversary” is essential in every battle—and this equally applies to the current global struggle against the COVID-19 pandemic. Understanding the parameters that influence the course of the pandemic is of paramount importance in the ongoing worldwide attempts to minimize the devastating effects of the virus which, to the present moment, has already taken a toll of more than a million lives (Dong et al., 2020), and resulted in double-digit recession among some of the major world economies (World Bank, 2020a). Of all such factors, the ecological ones (both abiotic such as meteorological factors and biotic such as demographic and health-related population properties) likely play a prominent role in determining the dynamics of disease progression (Qu et al., 2020).

However, making good estimates of the effects that general demographic, health-related, and weather conditions, have on the spread of COVID-19 infection is beset by many difficulties. First of all, these dependencies are subtle and easily overshadowed by larger-scale effects. Furthermore, as the effective rate of disease spread is an interplay of numerous biological, medical, social, and physical factors, a particular challenge is to differentiate the dominating effects of local COVID-19-related policies, which are both highly heterogeneous and time-varying, often in an inconsistent manner. And this is precisely where, in our view, much of the previous research on this subject falls short.

There are not many directly observable variables that can be used to trace the progression of the epidemics on a global scale (i.e., for a large number of diverse countries). The most obvious one—the number of detected cases—is heavily influenced both by the excessiveness of the testing (which, in turn, depends on non-uniform medical guidelines, variable availability of testing kits, etc.) and by the introduced infection suppression measures (where the latter are not only non-homogeneous but are also erratically observed (Cohen and Kupferschmidt, 2020). Nevertheless, the majority of the research aimed to establish connections of the weather and/or demographic parameters with the spread of COVID-19 seeks correlations exactly with the raw number of detected cases (Adhikari and Yin, 2020; Correa-Araneda et al., 2020; Fareed et al., 2020; Gupta et al., 2020; Iqbal et al., 2020; Li et al., 2020; Pourghasemi et al., 2020; Rashed et al., 2020; Singh and Agarwal, 2020). For the aforementioned reasons, the conclusions reached in this way are questionable. Other variables that can be directly measured, such as the number of hospitalized patients or the number of COVID-19 induced deaths (Pranata et al., 2020; Tosepu et al., 2020; Ward, 2020), again depend on many additional parameters that are difficult to take into account: level of medical care and current hospital capacity, advancements, and changing practices in treating COVID-19 patients, the prevalence of risk groups, and even on the diverging definitions of when hospitalization or death should be attributed to the COVID-19 infection. As such, these variables are certainly not suitable as proxies of the SARS-CoV-2 transmissibility *per se*.

On the other hand, as we here empirically find [and as theoretically expected (Anderson and May, 1992; Keeling and

Rohani, 2011)] the initial stage of the COVID-19 epidemic (in a given country or area) is marked by a period of a nearly perfect exponential growth for a wide range of countries, which typically lasts for about 2 weeks (based on our analysis of the available data). One can observe widespread dynamical growth patterns for many countries, with a sharp transition between exponential, superlinear (growth faster than linear), and sublinear (growth slower than linear) regimes (see **Figure 1**)—the last two representing a subexponential growth. We here concentrate on the initial exponential growth of the detected-case data (marked in red in **Figure 1**), characterizing the period *before* the control measures took effect, and with a negligible fraction of the population resistant to infection. Note that dates which correspond to the exponential growth regime (included in **Supplementary Table 1**) are different for each country, corresponding to the different start of COVID-19 epidemic in those countries.

We use the exponential growth regime to deduce the basic reproduction number R_0 (Martcheva, 2015), following a simple and robust mathematical (dynamical) model presented here. R_0 is a straightforward and important epidemiological parameter characterizing the inherent biological transmissibility of the virus, in a completely naïve population, and the absence of social distancing measures (Bar-On et al., 2020; Eubank et al., 2020). To emphasize the absence of social distancing in the definition (and inference) of R_0 used here, the term $R_{0,free}$ is also used, — for simplicity, we further denote $R_0 \equiv R_{0,free}$. R_0 is largely independent of the implemented COVID-19 policies and thus truly reflects the characteristics of the disease itself, as it starts to spread unhampered through the given (social and meteorological) settings. Namely, the exponential period ends precisely when the effect of control measures kick in, which happens with a delay of ~ 10 days after their introduction

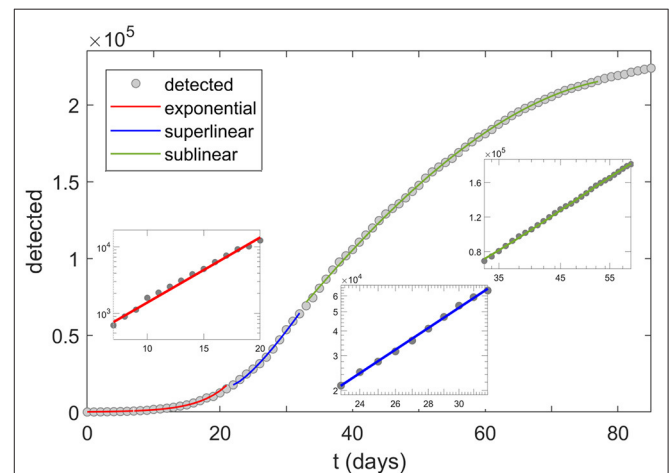


FIGURE 1 | COVID-19 growth regimes. Transitions of the growth patterns (here shown for Italy) from exponential (red), to superlinear (blue) and sublinear (green) regime. The three insets correspond to the log-linear scale (exponential), log-log scale (superlinear), and linear-log scale (sublinear). Dots correspond to detected infections, starting from 20.02.2020. In this study, $R_{0,free}$ is extracted from the slope of the first (exponential, i.e., log-linear) inset, corresponding to dates 29.02–13.03 in the case of Italy.

(The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020), corresponding to the disease latent period, and to the time between the symptom onset and the disease confirmation. Not only that very few governments had enacted any social measures before the occurrence of a substantial number of cases (Cohen and Kupferschmidt, 2020), but also the length of the incubation period makes it likely that the infection had been already circulating for some time through the community even before the first detected case (and that the effects of the measures are inescapably delayed in general). Also, the transition from the exponential to the subsequent subexponential phase of the epidemics is readily visible in the COVID data (see **Figure 1**). Furthermore, R_0 is invariant to the particular testing guidelines, as long as these do not significantly vary over the (here relatively short) studied period. Note that in **Figure 1** cumulative number of positive cases (also known as cumulative infection incidence) is shown, which has to monotonically increase—though with a decreasing rate, once the infection starts to slow down, i.e., once the subexponential growth (sublinear and superlinear regimes) is reached.

In the analysis presented here, we consider 42 different weather, demographic, and health-related population factors, whose analyzed ranges correspond to their variations exhibited in 118 world countries (not all of the parameters were available for all of the countries, as discussed in Section “Demographic and Weather Data Acquisition”). While some authors prefer more coherent data samples to avoid confusing effects of too many different factors (Adhikari and Yin, 2020; Correa-Araneda et al., 2020; Fareed et al., 2020; Rashed et al., 2020; Singh and Agarwal, 2020; Tosepu et al., 2020), this consideration is outweighed by the fact that large ranges of the analyzed parameters serve to amplify the effects we are seeking to recognize and to more reliably determine the underlying correlations. For example, while the value of the Human Development Index (HDI, a composite index of life expectancy, education, and per capita income indicators) varies from 0.36 to 0.96 over the set of analyzed countries, this range would drop by an order of magnitude (Global Data Lab, 2020) if the states of the US were chosen as the scope of the study (other demographic parameters exhibit similar behavior). The input parameters must take values in some substantial ranges to have measurable effects on R_0 (i.e., small variations may lead to effects that are easily lost in statistical fluctuations).

The number of considered parameters is also significant, especially when compared to other similar studies (Adhikari and Yin, 2020; Copat et al., 2020; Fareed et al., 2020; Iqbal et al., 2020; Rashed et al., 2020; Rychter et al., 2020; Singh and Agarwal, 2020; Thangriyal et al., 2020; Tosepu et al., 2020). In a model where a large number of factors are analyzed under the same framework, consistency of the obtained results, in terms of agreement with other studies, common-sense expectations, and their self-consistency, becomes an important check of applied methodology and analysis. Furthermore, a comprehensive and robust analysis is expected to generate new findings and lead to novel hypotheses on how environmental factors influence COVID-19 spread. Overall, we expect that the understanding achieved here will contribute to the ability to understand the behavior of the pandemics in the future and, by the same token,

to timely and properly take measures in an attempt to ameliorate the disease effects.

MODEL AND PARAMETER EXTRACTION

Modified SEIR Model and Relevant Approximations

There are various theoretical models and tools used to investigate and predict the progress of an epidemic (Keeling and Rohani, 2011; Martcheva, 2015). We here opted for the SEIR compartmental model, up to now used to predict or explain different features of COVID-19 infection dynamics (Maier and Brockmann, 2020; Maslov and Goldenfeld, 2020; Perkins and España, 2020; Tian et al., 2020; Weitz et al., 2020). The model is sufficiently simple to be applied to a wide range of countries while capturing all the features of COVID-19 progression relevant for extracting the R_0 values. The model assumes dividing the entire population into four (mutually exclusive) compartments with labels: (S)usceptible, (E)xposed, (I)nfected, and (R)ecovered.

The dynamics of the model (which considers gradual transitions of the population from one compartment to the other) directly reflects the disease progression. Initially, a healthy individual has no developed SARS-CoV-2 virus immunity and is considered as “susceptible.” Through contact with another infected individual, this person may become “exposed”—denoting that the transmission of the virus has occurred, but the newly infected person at this point has neither symptoms nor can yet transmit the disease. An exposed person becomes “infected”—in the sense of becoming contagious—on average after the so-called “latent” period which is, in the case of COVID-19, approximately 3 days. After a certain period of the disease, this person ceases to be contagious and is then considered as “recovered” (from the mathematical perspective of the model, “recovered” are all individuals who are no longer contagious, which therefore also includes deceased persons). In the present model, the recovered individuals are taken to be no longer susceptible to new infections (irrespective of whether the COVID-19 immunity is permanent or not, it is certainly sufficiently long in the context of our analysis).

Accordingly, almost the entire population initially belongs to the susceptible class. Subsequently, parts of the population become exposed, then infected, and finally recovered. SARS-CoV-2 epidemic is characterized by a large proportion of asymptomatic cases (or cases with very mild symptoms) (Day, 2020), which leads to a large number of cases that remain undiagnosed. For this reason, only a portion of the infected will be identified (diagnosed) in the population, and we classify them as “detected.” This number is important since it is the only direct observable in our model, i.e., the only number that can be directly related to the actual COVID-19 data.

This dynamic is schematically represented in **Figure 2**, and is governed by the following set of differential equations:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \quad (1)$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E \quad (2)$$

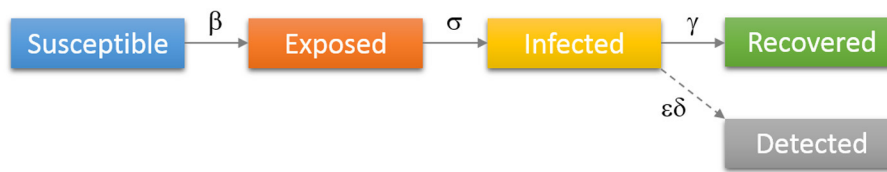


FIGURE 2 | Diagrammatic representation of the SEIR model with the added class of “Detected” patients. Individuals move (denoted by solid arrows) from Susceptible to Exposed to Infected to Recovered, with the rates indicated above arrows in the figure. Some of the infected are detected (diagnosed/confirmed), indicated by the dashed arrow.

$$\frac{dI}{dt} = \sigma E - \gamma I \quad (3)$$

$$\frac{dR}{dt} = \gamma I \quad (4)$$

$$\frac{dD}{dt} = \epsilon \delta I \quad (5)$$

In the above equations, S , E , I , and R denote numbers of individuals belonging to, respectively, susceptible, exposed, infected, and recovered compartments, D is the cumulative number of detected cases, while N is the total population. Parameter β denotes the transmission rate, which is proportional to the probability of disease transmission in contact between a susceptible and an infectious subject. Incubation rate σ determines the rate at which exposed individuals become infected and corresponds to the inverse of the average incubation period. Recovery rate γ determines the transition rate between infected and recovered parts of the population, (i.e., $1/\gamma$ is the average period during which an individual is infectious). Finally, ϵ and δ are detection efficiency and the detection rate. All these rate parameters are considered constant during the analyzed (brief) period. Also, note that the constants in our model do not correspond to transition probabilities *per se*, but rather to transition rates (with units 1/time), so that e.g. γ and $\epsilon\delta$ do not add to one. While rates in the model can be rescaled and normalized to directly correspond to transition probabilities, our formulation (with rates rather than probabilities) is rather common (see e.g., Keeling and Rohani, 2011), and also has a direct intuitive interpretation, where the transition rates correspond to the inverse of the period that individuals spend in a given compartment (see e.g., the explanation for γ above).

In the first stage of the epidemic, when essentially the entire population is susceptible (i.e., $S/N \approx 1$) and no distancing measures are enforced, the average number of secondary infections, caused directly by primary-infected individuals, corresponds to the basic reproduction number R_0 . The infectious disease can spread through the population only when $R_0 > 1$ (Khajanchi et al., 2020a), and in these cases, the initial growth of the infected cases is exponential. Though R_0 is a characteristic of the pathogen, it also depends on environmental abiotic (e.g., local weather conditions), as well as biotic factors (e.g., prevalence of health conditions, and population mobility tightly related to the social development level).

Note that, as we seek to extract the basic reproduction number R_0 from the model for a wide range of countries,

the social distancing effects are not included in the model presented above. That is, the introduced model serves only to explain the exponential growth phase—note that this growth regime characterizes part of the infection progression where the social distancing interventions still did not take effect, and where the fraction of resistant (non-susceptible) population is still negligible. It is only this phase which is relevant for extracting R_0 that is used in the subsequent analysis. R_0 should not be confused with the effective reproduction number R_e , which takes into account also the effects of social distancing interventions and the decrease in the number of susceptibles due to acquired infection resistance. R_e is not considered in this work, as we are concerned with the factors that affect the inherent biological transmissibility of the virus, independently from the applied measures. That is, by considering R_0 rather than R_e , we disentangle the influence of meteorological and demographic factors on transmissibility (the goal of this study), from the effects of social distancing interventions (not analyzed here). The model can, however, be straightforwardly extended to include social distancing measures, as we did in (Djordjevic et al., 2020)—social distancing measures were also included through other frameworks (Khajanchi and Sarkar, 2020; Maier and Brockmann, 2020; Maslov and Goldenfeld, 2020; Perkins and España, 2020; Samui et al., 2020; Sarkar et al., 2020; Tian et al., 2020; Weitz et al., 2020). Such extensions are needed to explain the subexponential growth that emerges due to intervention measures (i.e., superlinear and sublinear growth regimes that are illustrated in **Figure 1** for Italy but are common for other countries as well).

COVID-19 Growth Regimes

If we observe the number of total COVID-19 cases (e.g., in a given country) as a function of time, there is a regular pattern that we observe: the growth of the detected COVID cases is initially exponential but slows down after some time—when we say it enters the subexponential regime. The subexponential regime can be further divided into the superlinear (growing asymptotically faster than a linear function) and sublinear regime (the growth is asymptotically slower than a linear function). This typical behavior is illustrated, in the case of Italy, in **Figure 1** above. The transition to the subexponential regime occurs relatively soon, much before a significant portion of the population gains immunity, and is a consequence of the introduction of the infection suppression measures.

Inference of the Basic Reproduction Number R_0

In the initial exponential regime, a linear approximation to the model can be applied. Namely, in this stage, almost the entire population is susceptible to the virus, i.e., $S/N \approx 1$, which simplifies the Equation (2) to:

$$\frac{dE}{dt} = -\sigma E + \beta I. \quad (6)$$

By combining expressions (3) and (6) one obtains:

$$\frac{d}{dt} \begin{pmatrix} E \\ I \end{pmatrix} = \begin{pmatrix} -\sigma & \beta \\ \sigma & -\gamma \end{pmatrix} \begin{pmatrix} E \\ I \end{pmatrix} = A \begin{pmatrix} E \\ I \end{pmatrix}, \quad (7)$$

where we have introduced a two-by-two matrix:

$$A = \begin{pmatrix} -\sigma & \beta \\ \sigma & -\gamma \end{pmatrix} \quad (8)$$

The solution for the number of infected individuals can now be written:

$$I(t) = C_1 \cdot e^{\lambda_+ t} + C_2 \cdot e^{\lambda_- t}, \quad (9)$$

where λ_+ and λ_- denote eigenvalues of the matrix A, i.e., the solutions of the equation:

$$\det(A - \lambda I) = 0. \quad (10)$$

The eigenvalues must satisfy:

$$\begin{vmatrix} -\sigma - \lambda & \beta \\ \sigma & -\gamma - \lambda \end{vmatrix} = 0,$$

leading to:

$$(\lambda + \sigma) \cdot (\lambda + \gamma) - \beta \cdot \sigma = 0. \quad (11)$$

The solutions of (11) are:

$$\lambda_{\pm} = \frac{-(\gamma + \sigma) \pm \sqrt{(\gamma - \sigma)^2 + 4\beta\sigma}}{2}. \quad (12)$$

Since $\lambda_- < 0$, the second term in (9) can be neglected for sufficiently large t . More precisely, numerical analysis shows that this approximation is valid already after the second day, while, for the extraction of R_0 value we will anyhow ignore all data before the fifth day (for the analyzed countries, numbers of cases before the fifth day were generally too low, hence this early data is dominated by stochastic effects/fluctuations). Hence, $I(t)$ is proportional to $\exp(\lambda_+ t)$, i.e.:

$$I(t) = I(0) \cdot e^{\lambda_+ t}. \quad (13)$$

By using β from (12) and $R_0 = \frac{\beta}{\gamma}$ (Keeling and Rohani, 2011; Martcheva, 2015), we obtain:

$$R_0 = 1 + \frac{\lambda_+ \cdot (\gamma + \sigma) + \lambda_+^2}{\gamma \cdot \sigma}. \quad (14)$$

From (13) and (5) we compute:

$$D(t) = \varepsilon \cdot \delta \cdot I(0) \cdot \frac{(e^{\lambda_+ t} - 1)}{\lambda_+}. \quad (15)$$

By taking the logarithm, the above expression leads to:

$$\log(D(t)) = \log(\varepsilon \delta I(0) / \lambda_+) + \lambda_+ \cdot t, \quad (16)$$

from which λ_+ can be obtained as the slope of the $\log(D(t))$ function. From Equation (14), we thus obtain the R_0 value as a function of the slope of $\log(D(t))$, where the latter can be efficiently inferred from the plot of the number of detected COVID-19 cases for a large set of countries.

The SEIR model and the above derivation of R_0 assume that the population belonging to different compartments is uniformly mixed. Possible heterogeneities may tend to increase R_0 values (Keeling and Rohani, 2011). However, this would not influence the results obtained below, as our R_0 values are consistently inferred for all analyzed countries by using the same model, methodology, and parameter set. Moreover, our R_0 values are in agreement with the prevailing estimates in the literature (Najafimehr et al., 2020).

Demographic and Weather Data Acquisition

For the countries for which R_0 was determined through the procedure above, we also collect a broad spectrum of meteorological and demographic parameter values. Overall, 118 countries were selected for our analysis, based on the relevance of the COVID-19 epidemiological data. Namely, a country was considered as relevant for the analysis if the number of detected cases on June 15th was higher than a threshold value of 1,000. A few countries were then discarded from this initial set, where the case count growth was too irregular to extract any results, possibly due to inconsistent or irregular testing policies. As a source for detected cases, we used (World Bank, 2020b; Worldometer, 2020).

In the search for factors correlated with COVID-19 transmissibility, we have analyzed overall 42 parameters, 11 of which are related to weather conditions, 30 to demographics or health-related population characteristics, and one parameter quantifying a delay in the epidemic's onset (data provided in **Supplementary Tables 2–5**). Not all of these parameters were available for all of the considered countries. In particular, data on the prevalence of blood types (**Supplementary Table 4** in the Supplement) was possible to find for 83 of the 118 countries, while, primarily due to scarce data on pollutant concentrations during the epidemics, almost 30% of entries in **Supplementary Table 5** in Supplement had to be left blank for this category. Nevertheless, we opted to include these parameters in our report: despite the lower number of values, some of these parameters exhibited strong and highly statistically significant correlations with R_0 , warranting their inclusion.

Our main source of weather data was project POWER (Prediction of Worldwide Energy Resources) of the NASA agency (NASA Langley Research Center, 2020). A dedicated

Python script was written and used to acquire weather data via the provided API (Application Programming Interface). NASA project API allows a large set of weather parameters to be obtained for any given location (specified by latitude and longitude) and given date (these data are provided in the **Supplementary Table 7**). From this source, we gathered data on temperature (estimated at 2 m above ground), specific humidity (estimated at 2 m above ground), wind speed (estimated at 2 m above ground), and precipitation (defined as the total column of precipitable water). Data on air pressure (at ground level) and UV index (international standard measurement of the strength of sunburn-producing ultraviolet radiation) were collected via similar API from World Weather Online source (World Weather Online, 2020), using the same averaging methodology. Since we needed to assign a single value to each country (for each analyzed parameter), the following method was used for averaging meteorological data. In each country, a number of largest cities¹ were selected and weather data was taken for the corresponding locations. These data was then averaged, weighted by the population of each city, followed by averaging over the period used for R_0 estimation (more precisely, to account for the time between disease transmission and the case confirmation, we shifted this period 12 days into the past). The applied averaging method used here can be of limited adequacy in countries spreading over multiple climate zones, but is still expected to provide reasonable single-value estimates of the weather parameters, particularly since the averaging procedure was formulated to reflect the most likely COVID-19 hotspots in a given country.

Demographic data was collected from several sources. Percentage of the urban population, refugees, net migration, social and medical insurance coverage, infant mortality, and disease (CVD, cancer, diabetes, and CRD) risk was taken from the World Bank organization (World Bank, 2020b). The HDI was taken from the Our World in Data source (Our World in Data, 2020), while median age information was obtained from the CIA website. The source of most of the considered medical parameters: cholesterol, raised blood pressure, obesity, inactivity, BSG vaccination as well as data on alcohol consumption and smoking prevalence was World Health Organization (World Health Organization, 2020). Data for blood types were taken from the Wikidata web site. BUCAP parameter, representing population density in the built-up area, was taken from GHS Urban Center Database 2015 (European Commission Global Human Settlement, 2020). The onset parameter, determining the delay (in days) of the epidemic's start, was inferred from COVID-19 counts data. We used the most recent available data for all the parameters.

RESULTS

The $\log(D(t))$ function, for a subset of selected countries, is shown in **Figure 3**. The obvious linear dependence confirms that

¹This number was determined for each country by the following condition: the total population of the cities taken into consideration had to surpass 10 percent of the overall population of the country.

the progression of the epidemic in this stage is almost perfectly exponential. Note that our model exactly reproduces this early exponential growth (see Equation 13), happening under the assumption of a small fraction of the population being resistant, and the absence of the effect of social distancing interventions. From **Figure 3**, we see that this behavior, predicted by the model for the early stage of the epidemic, is also directly supported by the data, i.e., the exponential growth in the cumulative number of confirmed cases is indeed observed for a wide range of countries. For each country, the parameter λ_+ is directly obtained as the slope of the corresponding linear fit of the $\log(D(t))$, and the basic reproduction number R_0 is then calculated from Equation (14). Here, we used the following values for the incubation rate, $\sigma = 1/3 \text{ day}^{-1}$, and for the recovery rate $\gamma = 1/4 \text{ day}^{-1}$, per the commonly accepted values in the literature (Bar-On et al., 2020). Note that possible variations in these two values would not significantly affect any conclusions about R_0 correlations, due to the mathematical properties of the relation (14): it is a strictly monotonous function of λ_+ and the linear term $\lambda_+ \cdot (\gamma + \sigma) / \gamma \cdot \sigma$ dominantly determines the value of R_0 .

Supplementary tables contain the values for 42 variables, for all countries. Correlations of each of the variables with R_0 are given in **Supplementary Table 6**. Values for the Pearson correlation coefficient are further shown below, though consistent conclusions are also obtained by Kendall and Spearman correlation coefficients (which do not assume a linear relationship between variables). Correlation coefficients were calculated in the usual manner: as the correlation of the vector of parameter values with the vector of R_0 values, by taking into account all available data (for parameters that were available across all of the countries, both of the vectors were 118 dimensional; if values were missing for certain countries, these countries were simply ignored and lower-dimensional vectors were compared).

The first set of results that corresponds to, roughly speaking, general demographic data, is presented in **Figure 4**. The plot in panel A shows the distribution of R_0 vs. HDI values for all countries, where a higher HDI score indicates the more prosperous country concerning life expectancy, education, and per capita income (Sagar and Najam, 1998). This parameter was included in the study due to a reasonable expectation that a higher level of social development also implies a higher level of population interconnectedness and mixing (stronger business and social activity, more travelers, more frequent contacts, etc.), and hence that HDI could be related to the SARS-CoV-2 transmissibility. Indeed, we note a strong, statistically highly significant correlation between the HDI and the R_0 value, with $R = 0.37$, and $p = 4 \cdot 10^{-5}$, demonstrating that the initial expansion of COVID-19 was faster in more developed societies.

The social security and health insurance coverage (INS) “shows the percentage of population participating in programs that provide old age contributory pensions (including survivors and disability) and social security and health insurance benefits (including occupational injury benefits, paid sick leave, maternity, and other social insurance)” (World Bank, 2020b). Reflecting the percentage of the population covered by medical insurance and likely feeling more protected from the financial

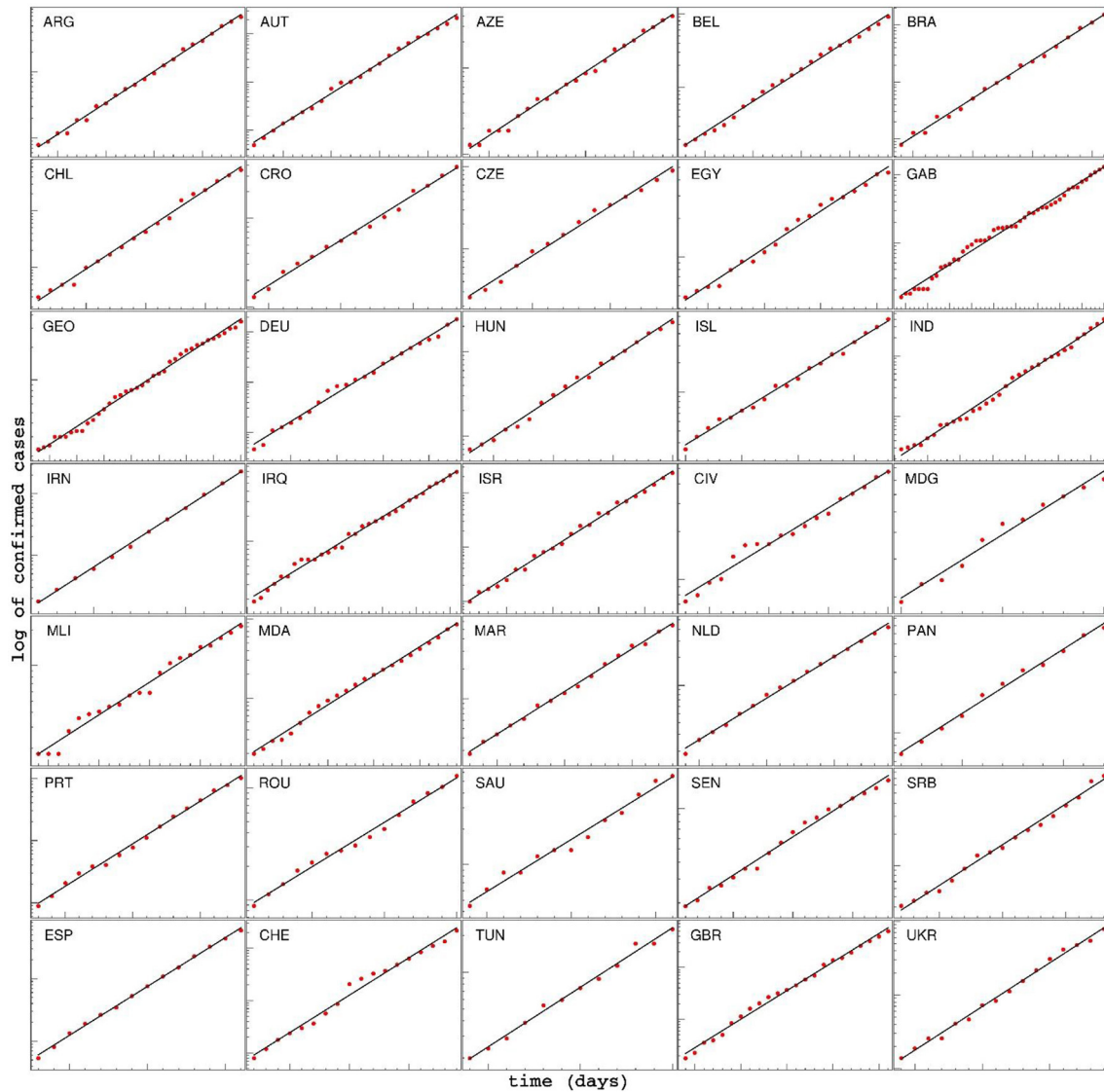
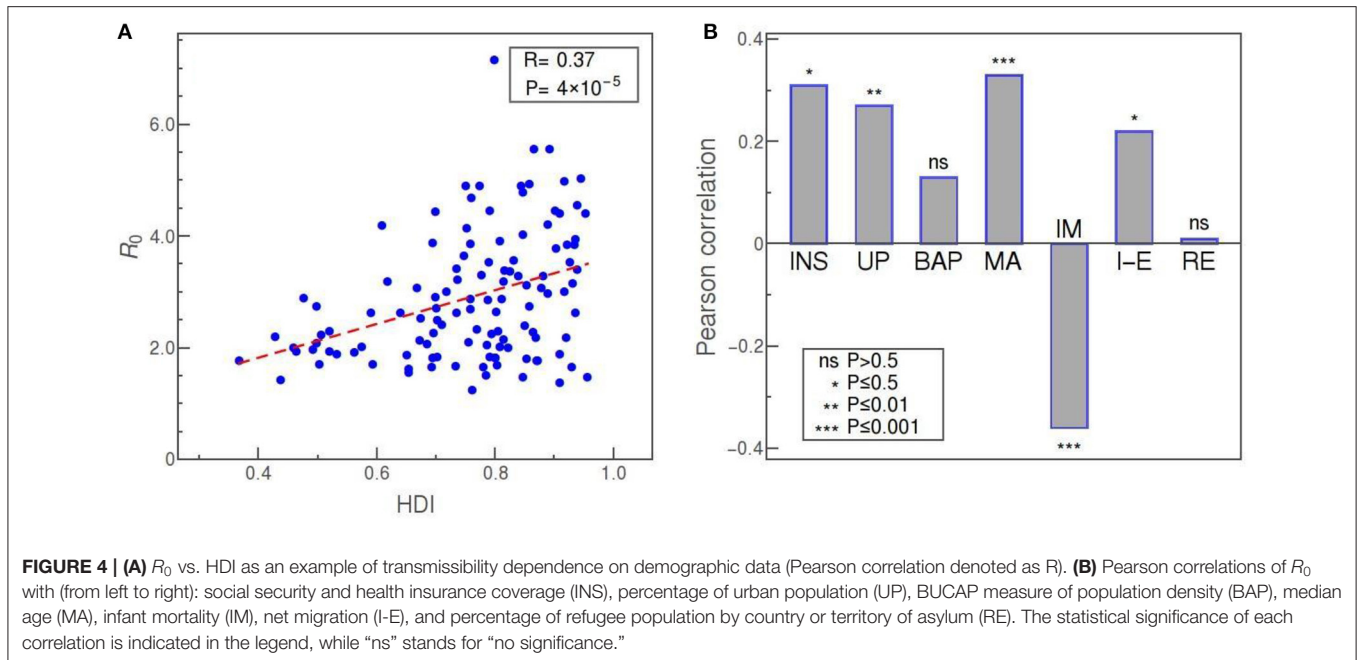


FIGURE 3 | Time dependence of the detected cases for various countries, during the initial period of the epidemic, shown on a log-linear scale. The linear fit of $\log(D)$ shows that the spread of COVID-19 in this phase is very well approximated by exponential growth. Note that the values on axes are chosen differently for each country, in order to emphasize the exponential character of the growth. For each country, the start and end dates of the exponential regime, together with the extracted slope λ_+ , are provided in the **Supplementary Table 1**. ARG, Argentine; AUT, Austria; AZE, Azerbaijan; BEL, Belgium; BRA, Brazil; CHL, Chile; CRO, Croatia; CZE, Czech Republic; EGY, Egypt; GAB, Gabon; GEO, Georgia; DEU, Germany; HUN, Hungary; ISL, Iceland; IND, India; IRN, Iran; IRQ, Iraq; ISR, Israel; CIV, Cote d'Ivoire; MDG, Madagascar; MLI, Mali; MDA, Moldova; MAR, Morocco; NLD, Netherlands; PAN, Panama; PRT, Portugal; ROU, Romania; SAU, Saudi Arabia; SEN, Senegal; SRB, Serbia; ESP, Spain; CHE, Switzerland; TUN, Tunis; GBR, Great Britain; UKR, Ukraine.

effects of the epidemics, this indicator shows a strong ($R = 0.4$) and highly significant ($p = 4 \cdot 10^{-4}$) positive correlation with R_0 . The percentage of urban population (UP) and BUCAP density (BAP) are both included as measures of how concentrated is the population of the country. While the UP value simply shows what percentage of the population lives in cities, the BUCAP parameter denotes the amount of built-up area per person. Of the two, the former shows a highly significant positive correlation with the COVID-19 basic reproduction number, whereas the latter shows no correlation. Median age (MA) should be of obvious potential relevance in COVID-19 studies since it is well

known that the disease more severely affects the older population (Jordan et al., 2020). Thus, we wanted to investigate also if there is any connection of age with the virus transmissibility. Our results are suggestive of such a connection, since we obtained a strong positive correlation of age with R_0 , with very high statistical confidence. Infant mortality (IM) is defined as the number of infants dying before reaching 1 year of age, per 1,000 live births. Lower IM rates can serve as another indicator of the prosperity of a society, and it turns out that this measure is also strongly correlated, but negatively, with $R = -0.36$ and $p = 8 \cdot 10^{-5}$ (showing again that more developed countries, i.e.,



those with lower IM rates, have experienced more rapid spread of the virus infection). Net migration (I-E) represents the 5-year estimates of the total number of immigrants less the annual number of emigrants, including both citizens and non-citizens. This number, related to the net influx of foreigners, turns out to be positively correlated, in a statistically significant way, with R_0 . However, according to our data, the percentage of refugees, defined as the percentage of the people in the country who are legally recognized as refugees and were granted asylum in that country, is not correlated with R_0 at all.

Another set of parameters corresponds to medically-related demographic parameters and is shown in the upper part of **Figure 5**. The plot in panel A represents the average blood cholesterol level (in mmol/L) in the population of various countries, plotted against the value of R_0 . The two parameters are strongly correlated, with $R = 0.4$, and $p = 6 \cdot 10^{-6}$. Another demographic parameter with clear medical relevance, that has a comparatively strong and significant positive correlation with R_0 , is the alcohol consumption per capita (ALC), as shown in panel B of **Figure 5**. Our data shows that R_0 is also positively correlated, with high statistical significance, with the prevalence of obesity and to a somewhat smaller extent with the percentage of smokers. Here, obesity is defined as having a body-mass index over 30. A medical parameter that is strongly, but negatively, correlated with R_0 , is a measure of prevalence and severity of COVID-19 relevant chronic diseases in the population (CD). This parameter is defined as “the percent of 30-year-old-people who would die before their 70th birthday from any of cardiovascular disease, cancer, diabetes, or chronic respiratory disease, assuming that s/he would experience current mortality rates at every age and s/he would not die from any other cause of death” (World Bank, 2020b). The percentage of people with raised blood pressure (RBP) is also negatively

correlated with R_0 , though this correlation is not as strong and as statistically significant as in the case of the CD parameter. Here, raised blood pressure is defined as systolic blood pressure over 140 or diastolic blood pressure over 90, in the population older than 18. The percentage of smokers exhibits statistically significant (though not large) positive correlation. Two medical-demographic parameters that show no correlation with R_0 in our data are the prevalence of insufficient physical activity among adults aged over 18 (IN) and BCG immunization coverage among 1-year-olds (BCG).

In **Figure 5C** we see that blood types are, in general, strongly correlated with R_0 . The highest positive correlation is exhibited by A^- and O^- types, with a Pearson correlation of 0.4 and 0.39, and a very high statistical significance of $p = 10^{-4}$ and $p = 2 \cdot 10^{-4}$, respectively. Taken as a whole, group A is still strongly and positively correlated with R_0 , albeit with a bit lower statistical significance (A^+ type correlation has p -value two orders of magnitude higher than A^-). This is not so for group O that, overall, does not seem to be correlated to R_0 (O^+ even shows a certain negative correlation but without statistical significance). Our data reveals a highly significant positive correlation also for AB^- subtype ($R = 0.31$, $p = 0.003$), while neither the AB^+ subtype nor overall AB group is significantly correlated with the basic reproduction number. Clear negative correlation is exhibited only by B blood group ($R = -0.31$, $p = 0.004$), mostly due to the negative correlation of its B^+ subtype ($R = -0.34$, $p = 0.001$), whereas B^- subtype is not significantly correlated with R_0 in our data. If we consider the rhesus factor alone, we again observe very strong correlations with R_0 and with very high statistical significance: Rh^- and Rh^+ correlate positively ($R = 0.4$) and negatively ($R = -0.4$), respectively, with very high statistical significance ($p = 2 \cdot 10^{-4}$). The tendency of Rh^- and Rh^+ to, respectively, increase and decrease the transmissibility,

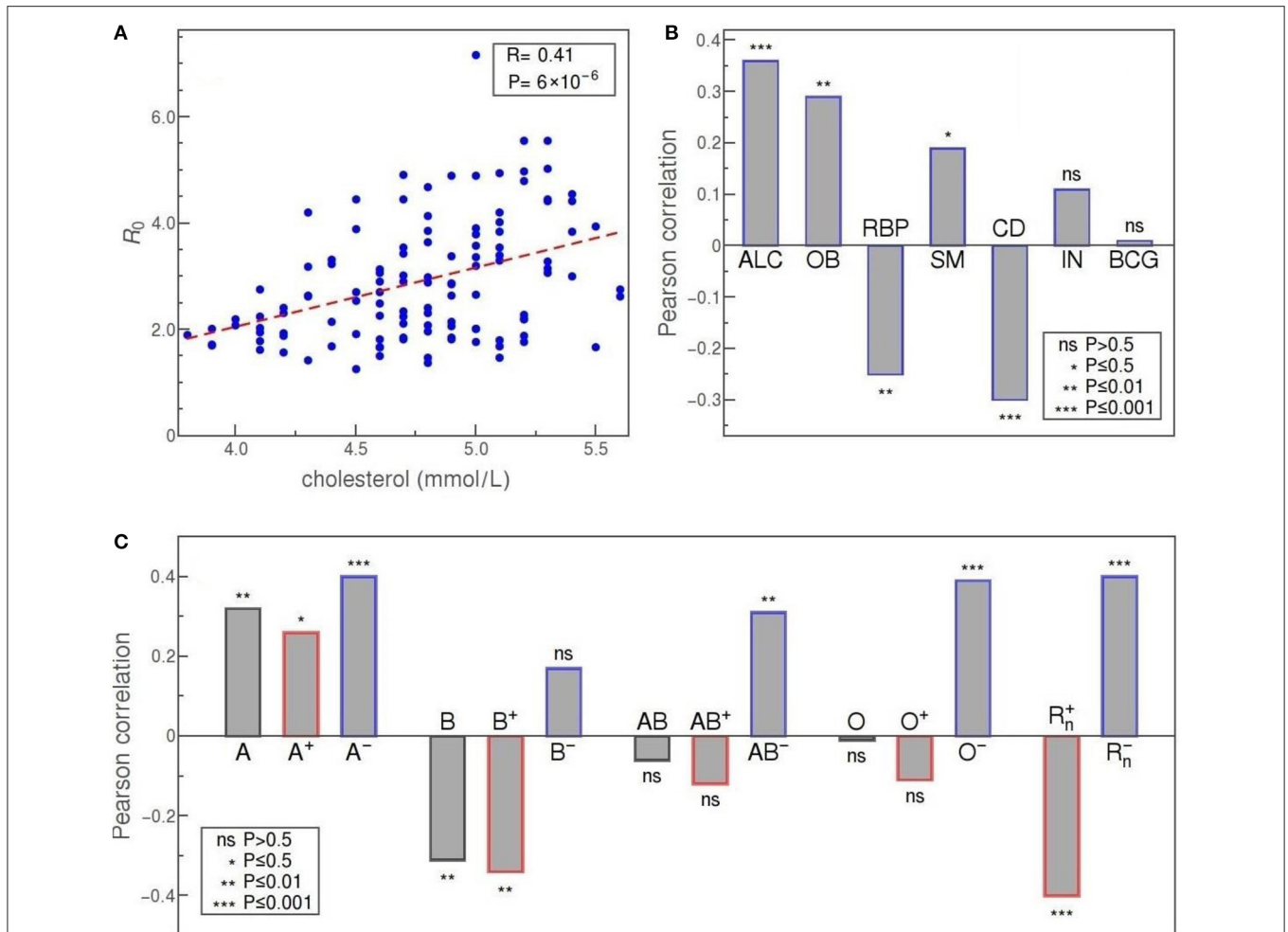


FIGURE 5 | (A) R_0 vs cholesterol level, as an example of a health-related parameter dependence. **(B)** Pearson correlation of R_0 with (from left to right): alcohol consumption per capita (ALC); the prevalence of obesity (OB); severity of COVID-19 relevant chronic diseases in the population (CD); a percentage of people with raised blood pressure (RBP); a percentage of smokers (SM); the prevalence of insufficient physical activity among adults (IN); BCG immunization coverage among 1-year-olds (BCG) **(C)** Correlation of blood types with R_0 in order: A, B, AB, and O (from left to right); overall value for that group, correlation only for Rh⁺ subtype of the group, and correlation for Rh⁻ subtype is shown. The two rightmost bars correspond to the overall correlation of Rh⁺ and the overall correlation of the Rh⁻ blood type with R_0 . The convention for representing the statistical significance of each correlation is the same as in **Figure 4**.

is therefore consistent with the results obtained for all four individual blood-groups.

In **Figure 6**, the onset represents the delay of the exponential phase and is defined, for each country, as the number of days from February 15 to the start of the exponential growth of detected cases. The motivation was to check for a possible correlation between the delay in the onset of the epidemic and the rate at which it spreads. Indeed, our data shows that such correlation exists and that it is strong and statistically significant: $R = -0.48$ and $p = 4 \cdot 10^{-8}$. In other words, the later the epidemic started, the lower (on average) is the basic reproduction number.

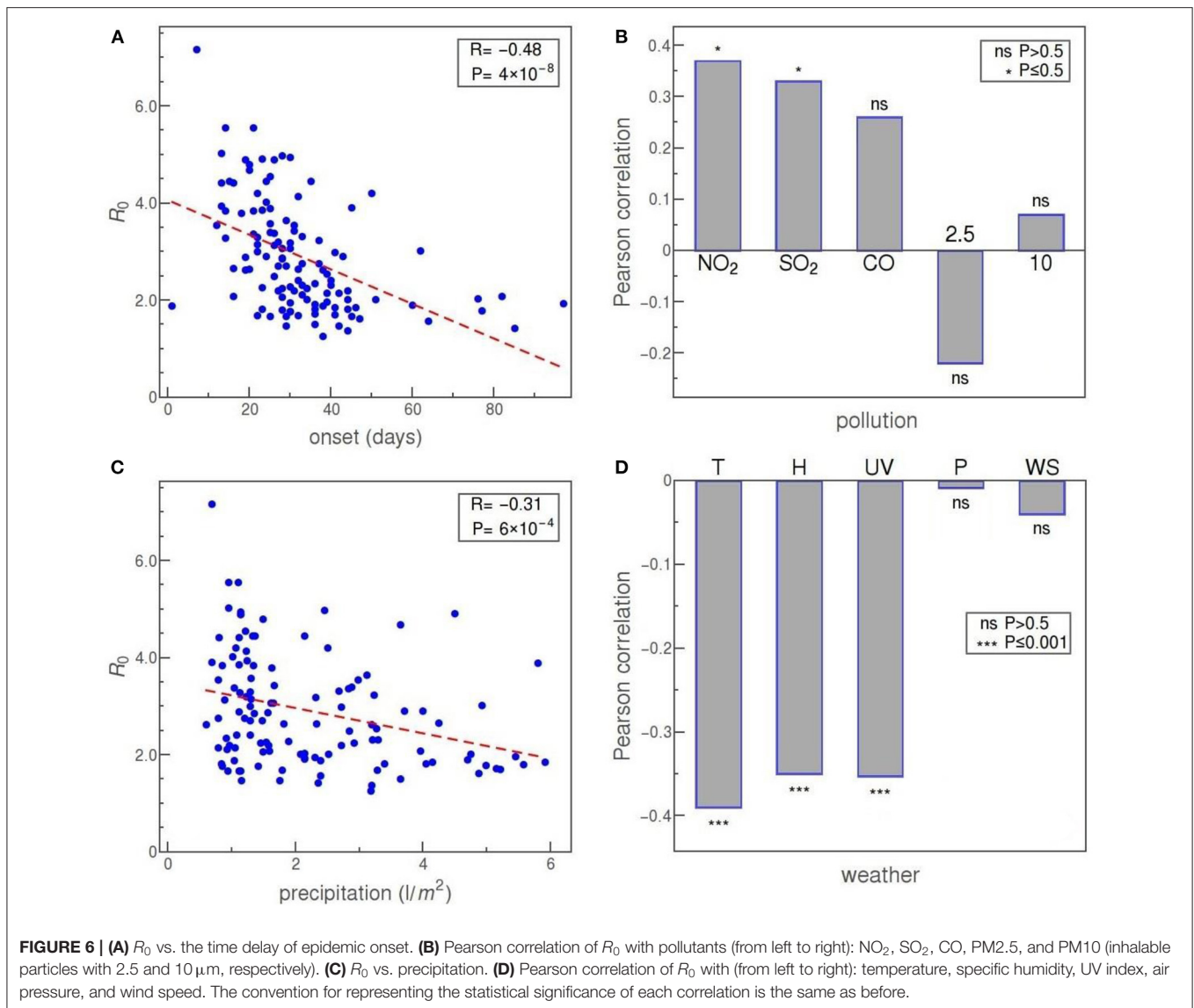
Panel B of **Figure 6** shows the correlation of R_0 with some of the commonly considered air pollutants. Our data reveal a statistically significant positive correlation of R_0 with NO_2 and SO_2 concentrations. Other pollutants—CO, PM_{2.5} (fine inhalable particles, with diameters that are generally 2.5 μm

and smaller), and PM₁₀ (inhalable particles, up to 10 nm in diameter)—show no statistically significant correlation with R_0 .

Next, we consider weather factors. Panels C and D of **Figure 6** show correlations of precipitation, temperature, specific humidity, UV index, air pressure, and wind speed with the reproduction number R_0 . Of these, precipitation, temperature, specific humidity, and UV index show a strong negative correlation, at a high level of statistical significance. Of the other two parameters, both air pressure and wind speed are not correlated at all with R_0 in our data.

DISCUSSION

The present paper aimed to establish relations between the COVID-19 transmissibility and a large number of



demographic and weather parameters. As a measure of COVID-19 transmissibility, we have chosen the basic reproduction number R_0 —a quantity that is essentially independent of the variations in both the testing policies and the introduced social measures (as discussed in the Introduction), in distinction to many studies on transmissibility that relied on the total number of detected case counts [see e.g., (Adhikari and Yin, 2020; Correa-Araneda et al., 2020; Fareed et al., 2020; Gupta et al., 2020; Iqbal et al., 2020; Li et al., 2020; Pourghasemi et al., 2020; Rashed et al., 2020; Singh and Agarwal, 2020)]. We have covered a substantial number of demographic and weather parameters, and included in our analysis all world countries that were significantly affected by the COVID-19 pandemic (and had a reasonable consistency in tracking the early phase of infection progression). While a number of manuscripts have been devoted to factors that may influence COVID-19 progression, only a few used an estimate of R_0 or some of its

proxies (Coccia, 2020; Contini and Costabile, 2020; Copiello and Grillenzoni, 2020)—these studies were however limited to China, and included a small set of meteorological variables, with conflicting results obtained for their influence on R_0 . Therefore, a combination of (i) using a reliable and robust measure of COVID-19 transmissibility, and (ii) considering a large number of factors that may influence this transmissibility within the same study/framework, distinguishes our study over prior work. We, however, must be cautious when it comes to further interpretation of the obtained data. As always, we must keep in mind that “correlation does not imply causation” and that further research is necessary to identify possible confounding factors and establish which of these parameters truly affect the COVID-19 transmissibility. Due to the sheer number of studied variables, an even larger number of parameters that might be relevant but are inaccessible to study (or even impossible to quantify), as well as due to possible intricate mutual relations of

the factors that may influence COVID-19 transmission, this is a highly non-trivial task. While we postpone any further analysis in this direction to future studies, we will, nevertheless, consider here the possible interpretations of the obtained correlations, assuming that they also probably indicate the existence of at least some causation. We below provide a detailed comparison between our results and previous findings. While a detailed discussion is presented, despite our best effort, we may have missed some of the relevant references due to an extremely rapidly developing field. Nevertheless, we point out a clear distinction of our work with previous studies, as outlined in this paragraph.

We will first consider the demographic variables presented in **Figure 4**. The obtained correlation of the HDI with the basic reproduction number is both strong and hardly surprising. The level of prosperity and overall development of a society is necessarily tied with the degree of population mobility and mixing, traffic intensity (in particular air traffic), business and social activity, higher local concentrations of people, and other factors that directly or indirectly increase the frequency and range of personal contacts (Gangemi et al., 2020), rendering the entire society more vulnerable to the spread of viruses. In this light, it is reasonably safe to assume that the obtained strong and highly statistically significant correlation of HDI with R_0 reflects a truly causal connection. However, some authors offer also a different explanation: that higher virus transmission in more developed countries is a consequence of more efficient detection of COVID-19 cases due to the better-organized health system (Gangemi et al., 2020)—but since our R_0 measure does not depend on detection efficiency, presented results can be taken as evidence against such hypothesis.

The interpretation is less clear for other demographic parameters, for example, the percentage of the population covered by medical and social insurance programs (INS). While there seem to be no previous studies discussing this parameter, one possibility is to attribute its strong positive correlation with R_0 to a hypothetical tendency of population to more easily indulge in the epidemiologically-risky behavior if they feel well-protected, both medically and financially, from the risks posed by the virus; conversely, that the population that cannot rely on professional medical care in the case of illness is likely to be more cautious not to contract the virus. The other is, of course, to see this correlation as an indirect consequence of the strong correlation of this parameter with HDI—which is also, almost certainly, the underlying explanation of the infant mortality (IM) correlation, where low mortality ratios point to a better medical system, which goes hand in hand with the overall prosperity and development of the country (Ruiz et al., 2015) (thus the negative correlation with R_0).

Similarly, the strong positive correlation of median age (MA) with R_0 might be a mere consequence of its clear relation with the overall level of development of the country (Gangemi et al., 2020), but it can be also considered in the light of the fact that clinical and epidemiological studies have unanimously shown that the elderly are at higher risk of developing a more severe clinical picture, and our result may indicate that the virus also spreads more efficiently in the elderly population. Possible

explanations may include: drugs frequently prescribed to this population that increase levels of ACE2 receptors (Shahid et al., 2020), a general weakening of the immune system with age leading to a greater susceptibility to viral infections (Pawelec and Larbi, 2008), and a large number of elderly people grouped in nursing homes, where the virus can expand very quickly (Kimball et al., 2020).

The correlation of population density with R_0 , or the lack of thereof, is more challenging to explain. Naively, one could expect that COVID-19 spreads much more rapidly in areas with a large concentration of people, but, if exists, this effect is not that easily numerically captured. As the standard population density did not show any correlation with the reproduction number R_0 (not shown), we explored some more subtle variants. Namely, the simplest reason why the data shows no correlation of R_0 with population density would be that the density, calculated in the usual way, is too averaged out: the most densely-populated country on our list, Monaco, has roughly 10,000 times more people per square kilometer than the least densely-populated Australia. However, Melbourne downtown has a similar population density as Monaco and far more people, so one would expect no a priori reason that its infection progression would be slower (and the R_0 rate for Australia as a whole will be dominantly determined by the fastest exponential expansion occurring anywhere on its territory). For this reason, we included the BUCAP parameter into the analysis, which takes into account only population density in built-up areas. Surprisingly, even this parameter did not exhibit any statistically significant correlation. Actually, several studies may serve as examples showing that the correlation of population density with the rate of COVID-19 expansion can be expected only under certain conditions since the frequency of contacts between people is to a large extent modulated by additional geographical, economic, and sociological factors (Berg et al., 2020; Carozzi, 2020; Pourghasemi et al., 2020; Rashed et al., 2020). Our observed absence of a correlation could be therefore expected and possibly indicates that such a correlation should be sought at the level of smaller populated areas—for example, individual cities (Yu et al., 2020). This conclusion is somewhat supported by the obtained highly significant and strong positive correlation of R_0 with the percentage of the population living in cities (UP) and which probably reflects the higher number of encounters between people in a more densely populated, urban environment (Li et al., 2020). It is also possible that virus spread might have a highly non-linear dependence on the population density—namely, that an outbreak in a susceptible population requires a certain threshold value of its density, while below that value population density ceases to be a significant factor influencing virus (Scheffer, 2009; Carozzi, 2020; Coro, 2020).

Another demographic parameter that exhibits a significant correlation with R_0 in our data is the net migration (I-E), denoting the number of immigrants less the number of emigrants. Unlike this number, which shows a positive correlation, the number of refugees (RE) seems not to be correlated at all. By definition, migrants deliberately choose to move to improve their prospects, while refugees have to move to save their lives or preserve their freedom. Migrants (e.g., in

economic or academic migration), arguably tend to stay in closer contact with the country of their origin and have more financial means for that, which likely contributes to more frequent border crossings and more intensive passenger traffic (Fan et al., 2020), thereby promoting the infection spread. On the other hand, refugees are mostly stationed in refugee camps, there is less possibility of spreading the virus outside through contacts with residents, but there is a high possibility of escalation of the epidemic within camps with a high concentration of people (Hargreaves et al., 2020). We did not find any other attempt in the literature to examine this issue. In any case, our results demonstrate that refugees are certainly not a primary cause of concern in the pandemics, contrary to fears expressed in some media.

Of the medical factors, the strongest correlation of R_0 is established with elevated cholesterol levels, as shown in **Figure 5**. Cholesterol may be associated with a viral infection and further disease development through a complex network of direct and indirect effects. *In vitro* studies of the role of cholesterol in virus penetration into the host body, done on several coronaviruses, indicate that its presence in the lipid rafts of the cell membrane is essential for the interaction of the virus with the ACE2 receptor, and also for the latter endocytosis of the virus (Radenkovic et al., 2020). Obesity prevalence (OB) also exhibits a highly significant, though somewhat weaker correlation with R_0 , which might be a consequence of the common connection between obesity and cholesterol: in principle, obesity might be a relevant factor in the COVID-19 epidemic exactly due to the effects of cholesterol on SARS-CoV-2 susceptibility. Of course, other effects might be at play, e.g., the fact that the adipose tissue of obese people excessively produces pro-inflammatory cytokines (Sattar et al., 2020). In the case of obesity, a simple explanation via relation to HDI is not available, since obesity does not show a simple correlation with the society development (Haidar and Cosman, 2011). Overall, while the correlation of obesity with a more severe prognosis in COVID-19 is well established in the literature, its relation to COVID-19 transmissibility is only mentioned in Li et al. (2020) and hitherto unexplained.

Often related to obesity is also raised blood pressure (RBP), and we have discovered that this factor is also correlated, at high statistical significance, with R_0 . While this seems to be the first study correlating high blood pressure with the SARS-CoV-2 transmission rate, it is known that, based on clinical studies, RBP appears to be a risk factor for hospitalization and death due to COVID-19 (Ran et al., 2020a; Schiffrin et al., 2020). In this light, it might be surprising that the correlation between RBP and R_0 turns out to be negative. On the other hand, this result supports the existing hypothesis about the beneficial effect of ACE inhibitors and ARBs (Ran et al., 2020a; Schiffrin et al., 2020) (standardly used in the treatment of hypertension). Similarly unintuitive correlation we report in the case of chronic diseases that are known to be relevant for the COVID-19 outcome. Namely, our data show, at very high statistical significance, a strong negative correlation of R_0 with the risk of death from a batch of chronic diseases (cardiovascular disease, cancer, diabetes, and chronic respiratory disease), agreeing in this regard with some recent research

(Chiang et al., 2020; Li et al., 2020). These diseases are identified as relevant comorbidities in the context of COVID-19, leading to a huge increase in the severity of the infection and poorer prognosis (An et al., 2020; Zheng et al., 2020) and, therefore, the discovered negative correlation comes as a surprise—particularly when contrasted to the positive correlation of obesity (where both are recognized risk factors in COVID-19 illness). One possible explanation is that the correlation may be due to potentially lower mobility of people with chronic diseases compared to the general mobility of the population. Additionally, it is possible that these people, being aware to belong to a high-risk group, behaved more cautiously even before the official introduction of social distancing measures.

According to our analysis, the prevalence of certain health-hazard habits is also significantly correlated to COVID-19 transmissibility. Chronic excessive alcohol consumption has, in general, a detrimental effect on immunity to viral and bacterial infections, which, judging by the strong positive correlation we obtained, most likely applies also to SARS-CoV-2 virus infection. This correlation contradicts the belief that alcohol can be used as a protective nostrum against COVID-19, which has spread in some countries and even led to cases of alcohol poisoning (Chick, 2020).

Regarding the impact of smoking on SARS-CoV-2 virus infection—the results are controversial (Chatkin and Godoy, 2020). The positive correlation of smoking with COVID-19 transmissibility that we obtained seems to support the reasoning that, since the SARS-CoV-2 virus enters cells by binding to angiotensin-converting enzyme 2 (ACE2) receptors and that the number of these receptors is significantly higher in the lungs of smokers, the smokers will be more affected and easily infected (Brake et al., 2020; Hoffmann et al., 2020). Accordingly, our result contradicts the hypothesis that a weakened immune response of smokers to virus infection may prove beneficial in the context of inflammation caused by intense cytokine release (Garufi et al., 2020).

Another result that addresses the association of unhealthy lifestyle with greater susceptibility to SARS-CoV-2 infection is the slight positive correlation we obtained for the prevalence of insufficient physical activity (IN) in adults, which is however not statistically significant. In this sense, in the case of COVID-19, we could not fully confirm the findings from (Jurak et al., 2020), who found that physical activity significantly reduces the risk of viral infections.

Despite the recent media interest (Gallagher, 2020), our findings neither could confirm that BCG immunization has any beneficial effect in the case of COVID-19, at least as far as reducing the risk of contracting and transmitting the disease is concerned. While it is known that the BCG vaccine provides some protection against various infectious agents, unfortunately, there is no clear evidence for such an effect against SARS-CoV-2 (O'Neill and Netea, 2020). Our analysis suggests that BCG immunization simply does not correlate with SARS-CoV-2 virus transmission.

SARS-CoV-2 target cells are typically capable of synthesizing ABH antigens and certain arguments exist, both theoretical and experimental, for a potential relation of blood groups

with COVID-19 progression and transmission (Guillon et al., 2008; Dai, 2020; Gérard et al., 2020). While the results of epidemiological studies on COVID-19 patients mostly support the proposed effect of blood groups on the development of COVID-19 disease, the relationship between virus transmission and blood group prevalence and Rh phenotype has been significantly less studied. Our analysis showed strong positive correlations of virus transmission with the presence of A blood group and Rh⁻ phenotype, as well as strong negative correlations for B blood group and Rh⁺ phenotype, while for AB and O blood group no significant correlations were obtained (**Figure 5C**). This result coincides significantly with the correlations obtained in a study conducted for 86 countries (Ansari-Lari and Saadat, 2020). However, another study focused on hospitalized patients in Turkey reported that the Rh⁺ phenotype represents a predisposition to infection (Arac et al., 2020), contradicting our findings. Similar results regarding the Rh factor were obtained in a study (Latz et al., 2020) on hospitalized patients in the US (this study further reported no correlation of blood types with the severity of the disease). One way to reconcile these results with ours would be to speculate that the virus is more efficiently transmitted in a population with a higher proportion of Rh⁻ phenotype because these people show a milder clinical picture compared to Rh⁺, so their movement is not equally limited, which is why they have more ability to pass on the infection.

Our data (**Figure 6A**) shows a strong negative correlation with the date of the epidemic onset. Curiously, it seems that the later the epidemic started in a given country, it is more likely that the disease expansion will be slower. Instead of interpreting this result as an indication that the virus has mutated and changed its properties over such a short period, we offer the following simpler explanation: pandemic reached first those countries that are most interconnected with the rest of the world (at the same time, those are the countries characterized by great mobility of people overall), so it is expected that also the progression of the local epidemics in these countries is more rapid. Another contributing factor could be the effect of media, which had more time to raise awareness about the risks of COVID-19 in the countries that were hit later (Khajanchi et al., 2020b).

Another segment of our interest were air pollutants, shown in **Figure 6B**. Air pollution can have a detrimental effect on the human immune system and lead to the development (or to worsening) of respiratory diseases, including those caused by respiratory viral infections (Becker and Soukup, 1999; Copat et al., 2020). Several papers have already investigated air pollution in the context of COVID-19 and reported a positive correlation between the death rate due to COVID-19 and the concentration of PM_{2.5} in the environment (Wu et al., 2020; Yao et al., 2020c). Positive correlations were also found for the spread of the SARS-CoV-2 virus, but mainly by considering daily numbers of newly discovered cases—a method that, as we have already argued, may strongly depend on testing policies, as well as on state measures to combat the epidemic (Copat et al., 2020). It has been suggested that virus RNA can be adsorbed to airborne particles facilitating thus its spread over greater distances (Coccia, 2020; Setti et al., 2020), but these arguments were contested by examination of air samples in Wuhan (Contini and Costabile, 2020; Liu et al., 2020).

The latter conclusions concur with the results of a study in which no correlation was obtained between the basic reproductive number of SARS-CoV-2 infection for 154 Chinese cities and the concentration of PM_{2.5} and PM₁₀ particles, while the correlation of these factors with the death rate (CFR) was shown (Ran et al., 2020b). The statistically insignificant and relatively weak correlations we obtained for PM_{2.5} and PM₁₀ pollutants also do not support the hypothesis of a potentially significant role of these particles in the transmission of this virus. In contrast, significant positive correlations were shown by our analysis for concentrations of NO₂, SO₂, and CO in the air (although the correlation for CO is not statistically significant), which is generally supported by the results of other studies. For example, a positive correlation of NO₂ levels with the basic reproductive number of infection was obtained from data for 63 Chinese cities (Yao et al., 2020a). Also, it has been shown that the number of detected cases of COVID-19 in China is strongly positively correlated with the level of CO, while in Italy and the USA such correlation exists with NO₂ (Pansini and Fornacca, 2020). The mentioned study failed to establish a clear correlation with the level of SO₂. Possible mechanisms of interaction were also proposed (Darai et al., 2020). Also, it is important to emphasize that the atmospheric concentration of NO₂ strongly depends on the levels of local exhaust emissions, so its correlation with virus transmission can be interpreted by the connection with the urban environment, characterized by more intensive traffic (Goldberg et al., 2020).

Finally, we have also obtained some interesting correlations of the meteorological parameters with R_0 , shown in **Figures 6C,D**. The statistically very highly significant negative correlation of the basic reproductive number of SARS-CoV-2 virus infection with both the mean temperature and humidity obtained in our research (**Figure 6D**) is consistent with the results of other relevant papers, e.g., (Mecenas et al., 2020). For example, a similar correlation was obtained in a study that analyzed COVID-19 outbreak in the cities of Chile—a country that covers several climate zones, but where it is still safe to assume that social patterns of behavior and introduced epidemic control measures do not drastically differ throughout the country (Correa-Araneda et al., 2020). Effectively the same conclusion—that fewer COVID-19 cases were reported in countries with higher temperatures and humidities—was reached in a study covering over 200 countries in the world (Iqbal et al., 2020). While an established correlation between virus transmission and a certain factor is not, in general, a telltale sign of a direct causal relationship between them, in the case of temperature and humidity such connection is firmly indicated also by results of experimental research (Lowen et al., 2007; Casanova et al., 2010; Chan et al., 2011; van Doremalen et al., 2020). Nevertheless, some studies yielded different conclusions, most likely due to the method of calculating R_0 or due to choosing a small/uninformative sample of populations in which the number of infected cases was monitored (Guo et al., 2020; Lin et al., 2020; Yao et al., 2020b). For example, a study focused on the suburbs of New York, Queens, obtained a positive correlation between virus transmission and temperature, which seems unexpected given the prevailing observations of other studies (Adhikari and Yin,

2020). This result is most likely a consequence of analyzing data for a small area (Queens only) where the temperature varies in a relatively narrow range of values, as well as correlating the number of detected cases, which may be sensitive to variations in the testing procedure.

Another environmental agent that can destroy or inactivate viruses is UV radiation from sunlight, and the properties of a particular virus determine how long it can remain infectious when exposed to radiation. For example, epidemics of influenza have a seasonal character precisely due to the susceptibility of influenza viruses to UV radiation (Sagripanti and Lytle, 2007). Our analysis found, at very high statistical significance, a strong negative correlation between the transmission of the SARS-CoV-2 virus and the intensity of UV radiation, which is consistent with the results of other studies obtained for the cities of Brazil and the provinces of Iran (Ahmadi et al., 2020; Mendonça et al., 2020). It is worth mentioning that lower temperatures, humidity, and sunlight levels usually occur in combination and directly affect not only the virus but also the human behavior, so the observed higher transmission of the virus in such conditions can alternatively be interpreted by indirect effects of other factors that act together in cold weather, such as more time spent indoors where the virus spreads more easily, or weakening of the immune system that increases susceptibility to infections (Abdullahi et al., 2020).

While the results related to COVID-19 correlations with temperature, humidity, and UV radiation are fairly frequent in the literature, this is less so for the results on the precipitation levels. Very few other studies have examined the association of precipitation with SARS-CoV-2 transmission, with either no correlation found (Pourghasemi et al., 2020), or looking at precipitation as a surrogate for humidity and generally receiving a negative correlation with infection rate (Araujo and Naimi, 2020; Coro, 2020). Our results, however, shown in **Figure 6D**, confirm natural expectations: just like humidity, the precipitation exhibits a strong negative correlation with R_0 , only slightly lower than in the case of T, H, and UV, at a very high level of statistical significance. Such results also concur with some general conclusions about the behavior of similar viruses (Agrawal et al., 2009; Pica and Bouvier, 2012).

Our analysis did not reveal any statistically significant correlation either between the wind speed or between air pressure and SARS-CoV-2 transmissibility. In the case of wind speed, this result agrees with the findings in some other papers (Gupta et al., 2020; Oliveiros et al., 2020). A positive correlation of wind speed with COVID-19 transmissibility was obtained in a study in Chilean cities, but, as the authors themselves note, the interpretation of the effect of this factor is complicated by its observed significant interaction with temperature (Correa-Araneda et al., 2020). The role of wind in transmitting the virus to neighboring buildings is predicted by the SARS virus spread model within the Amoy Gardens residential complex in Hong Kong, but such an effect may relate to local air currents and virus transmission over relatively short distances and does not imply a correlation of mean wind speeds in the area with virus transmission (McKinney et al., 2006; Pica and Bouvier, 2012). As for the air pressure, the potential connection is hardly at all investigated in the literature. An

exception is a study (Cambaza et al., 2020) reporting a positive correlation of air pressure with the number of COVID-19 cases in parts of Mozambique, but our results do not confirm such a conclusion.

CONCLUSION

While there is by now a significant amount of research on a crucial problem of how environmental factors affect COVID-19 spread, several features set this analysis apart from the existing research. First is the applied methodology: instead of basing analysis directly on the number of detected COVID-19 cases (or some of its simple derivatives), we employ an adapted SEIR model to extract the basic reproduction number R_0 from the initial stage of the epidemic. By taking into account only data in the exponential growth regime, i.e., before the social measures took effect (as explained in the “Methods” section), we ensured that the correlations we have later identified were not confounded with the effects of local COVID-19 policies. Even more importantly, our method is also invariant to variations in COVID-19 testing practices, which, as is well known, used to vary in quite an unpredictable manner between different countries. Another important factor is the large geographical scope of our research: we collected data from 118 countries worldwide, more precisely, from all the countries that were above a certain threshold for the number of confirmed COVID-19 cases (except for several countries with clearly irregular early growth data). The third factor was the number of analyzed parameters: we calculated correlations for the selected 42 different variables (of more than a hundred that we initially considered overall) and looked for viable interpretations of the obtained results.

These results should also help in resolving some of the existing disputes in the literature. For example, our findings indicate that correlation of HDI with R_0 is not a consequence of the COVID-19 testing bias, as was occasionally argued. Of the opposing opinions, our data seem to support assertions that blood types are indeed related to COVID-19 transmissibility, as well as arguments that the higher prevalence of smoking does increase the virus transmissibility (though weakly). On the other hand, in the dispute about the effects of the pollution, our correlations give an edge to claims that there is no correlation between PM2.5 and PM10 particles and transmissibility (whereas we agree with the prevailing conclusions about the positive correlation of other considered pollutants). In the case of the effects of the wind, based on the obtained results we tend to side with those denying any connection. In certain cases our findings contradict popular narratives: there are no clear indications that either number of refugees or physical inactivity intensifies the spread of COVID-19. Unfortunately, our data also suggest that BCG immunization may not help in subduing the epidemic. Additionally, the obtained correlations hint to possible new alleys of research, e.g., those that would help us understand the connection between cholesterol levels and SARS-CoV-2 transmissibility.

Overall, we believe that the presented results can be a useful contribution to the ongoing attempts to better understand the first pandemic of the twenty-first century—and the better we understand it, the sooner we may hope to overcome it.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MarD and MagD conceived the research. The work was supervised by IS, MagD, and MarD. Data acquisition by OM, IS, DZ, and AR. Data analysis by DZ, OM, and IS, with the help of MagD. Figures and tables made by DZ and AR with the help of MagD. A literature search by AR. Manuscript written by IS, AR, and MarD. All authors contributed to the article and approved the submitted version.

REFERENCES

- Abdullahi, I. N., Emeribe, A. U., Mustapha, J. O., Fasogbon, S. A., Ofor, I. B., Opeyemi, I. S., et al. (2020). Exploring the genetics, ecology of SARS-CoV-2 and climatic factors as possible control strategies against COVID-19. *Infez. Med.* 28, 166–173.
- Adhikari, A., and Yin, J. (2020). Short-term effects of ambient ozone, PM_{2.5}, and meteorological factors on COVID-19 confirmed cases and deaths in queens, New York. *Int. J. Environ. Res. Public Health* 17:4047. doi: 10.3390/ijerph17114047
- Agrawal, A. S., Sarkar, M., Chakrabarti, S., Rajendran, K., Kaur, H., Mishra, A. C., et al. (2009). Comparative evaluation of real-time PCR and conventional RT-PCR during a 2 year surveillance for influenza and respiratory syncytial virus among children with acute respiratory infections in Kolkata, India, reveals a distinct seasonality of infection. *J. Med. Microbiol.* 58, 1616–1622. doi: 10.1099/jmm.0.011304-0
- Ahmadi, M., Sharifi, A., Dorosti, S., Ghouschi, S. J., and Ghanbari, N. (2020). Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Sci. Total Environ.* 729:138705. doi: 10.1016/j.scitotenv.2020.138705
- An, C., Lim, H., Kim, D. W., Chang, J. H., Choi, Y. J., and Kim, S. W. (2020). Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Sci. Rep.* 10:18716. doi: 10.1038/s41598-020-75767-2
- Anderson, R. M., and May, R. M. (1992). *Infectious Diseases of Humans: Dynamics and Control*. New York, NY: Oxford University Press.
- Ansari-Lari, M., and Saadat, M. (2020). The morbidity and mortality of COVID-19 are associated with ABO and Rh blood groups. *Eur. J. Prev. Cardiol.* doi: 10.1177/2047487320939216. [Epub ahead of print].
- Arac, E., Solmaz, I., Akkoc, H., Donmezdil, S., Karahan, Z., Kaya, S., et al. (2020). Association between the Rh blood group and the Covid-19 susceptibility. *UHOD-Uluslar. Hematol.* 30, 81–86. doi: 10.4999/uhod.204247
- Araujo, M. B., and Naimi, B. (2020). Spread of SARS-CoV-2 coronavirus likely constrained by climate. medRxiv [Preprint]. Available online at <https://europepmc.org/article/ppr/ppr117563> (accessed December 16, 2020).
- Bar-On, Y. M., Flamholz, A., Phillips, R., and Milo, R. (2020). SARS-CoV-2 (COVID-19) by the numbers. *Elife* 9:e57309. doi: 10.7554/eLife.57309.sa2
- Becker, S., and Soukup, J. M. (1999). Exposure to urban air particulates alters the macrophage-mediated inflammatory response to respiratory viral infection. *J. Toxicol. Environ. Health Part A* 57, 445–457. doi: 10.1080/009841099157539
- Berg, M. K., Yu, Q., Salvador, C. E., Melani, I., and Kitayama, S. (2020). Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of COVID-19. *Sci. Adv.* 6:eabc1463. doi: 10.1126/sciadv.abc1463
- Brake, S. J., Barnsley, K., Lu, W., McAlinden, K. D., Eapen, M. S., and Sohal, S. S. (2020). Smoking upregulates angiotensin-converting enzyme-2 receptor: a potential adhesion site for novel coronavirus SARS-CoV-2 (Covid-19). *J. Clin. Med.* 9:841. doi: 10.3390/jcm9030841
- Cambaza, E. M., Viegas, G. C., and Cambaza, C. M. (2020). Potential impact of temperature and atmospheric pressure on the number of cases of COVID-19 in Mozambique, Southern Africa. *J. Public Health Epidemiol.* 12, 246–260. doi: 10.5897/JPHE2020.1258
- Carozzi, F. (2020). Urban density and COVID-19. *Institute for the Study of Labor (IZA)* [Online], 13440. Available online at <https://ssrn.com/abstract=3643204> (accessed December 16, 2020).
- Casanova, L. M., Jeon, S., Rutala, W. A., Weber, D. J., and Sobsey, M. D. (2010). Effects of air temperature and relative humidity on coronavirus survival on surfaces. *Appl. Environ. Microbiol.* 76, 2712–2717. doi: 10.1128/AEM.02291-09
- Chan, K.-H., Peiris, J. M., Lam, S., Poon, L., Yuen, K., and Seto, W. H. (2011). The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.* 2011:734690. doi: 10.1155/2011/734690
- Chatkin, J. M., and Godoy, I. (2020). Are smoking, environmental pollution, and weather conditions risk factors for COVID-19? *J. Bras. Pneumol.* 46:e20200183. doi: 10.36416/1806-3756/e20200183
- Chiang, W.-H., Liu, X., and Mohler, G. (2020). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *medRxiv [Preprint]*. Available online at: <https://europepmc.org/article/ppr/ppr172763> (accessed December 16, 2020).
- Chick, J. (2020). Alcohol and COVID-19. *Alcohol Alcohol.* 55, 341–342. doi: 10.1093/alcal/agaa039
- Coccia, M. (2020). The effects of atmospheric stability with low wind speed and of air pollution on the accelerated transmission dynamics of COVID-19. *Int. J. Environ. Stud.* doi: 10.1080/00207233.2020.1802937. [Epub ahead of print].
- Cohen, J., and Kupferschmidt, K. (2020). Countries test tactics in 'war' against COVID-19. *Science* 367, 1287–1288. doi: 10.1126/science.367.6484.1287
- Contini, D., and Costabile, F. (2020). Does air pollution influence COVID-19 outbreaks? *Atmosphere* 11:377. doi: 10.3390/atmos11040377
- Copat, C., Cristaldi, A., Fiore, M., Grasso, A., Zuccarello, P., Signorelli, S. S., et al. (2020). The role of air pollution (PM and NO₂) in COVID-19 spread and lethality: a systematic review. *Environ. Res.* 191:110129. doi: 10.1016/j.envres.2020.110129
- Copiello, S., and Grillenzoni, C. (2020). The spread of 2019-nCoV in China was primarily driven by population density. Comment on "Association between short-term exposure to air pollution and COVID-19 infection: evidence from China" by Zhu et al. *Sci. Total Environ.* 744:141028. doi: 10.1016/j.scitotenv.2020.141028
- Coro, G. (2020). A global-scale ecological niche model to predict SARS-CoV-2 coronavirus infection rate. *Ecol. Modell.* 431:109187. doi: 10.1016/j.ecolmodel.2020.109187
- Correa-Araneda, F., Ulloa-Yañez, A., Núñez, D., Boyero, L., Tonin, A., Cornejo, A., et al. (2020). Environmental determinants of COVID-19 transmission across a

FUNDING

This work was partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

ACKNOWLEDGMENTS

The authors thank Anica Brzakovic for reproducing a subset of the results presented in this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fevo.2020.617841/full#supplementary-material>

- wide climatic gradient in Chile. Research Square [Preprint]. Available online at: <https://europepmc.org/article/ppr/ppr166276> (accessed December 16, 2020).
- Dai, X. (2020). ABO blood group predisposes to COVID-19 severity and cardiovascular diseases. *Eur. J. Prev. Cardiol.* 27, 1436–1437. doi: 10.1177/2047487320922370
- Daraei, H., Toolabian, K., Kazempour, M., and Javanbakht, M. (2020). The role of the environment and its pollution in the prevalence of COVID-19. *J. Infect.* 81, e168–e169. doi: 10.1016/j.jinf.2020.06.019
- Day, M. (2020). Covid-19: four fifths of cases are asymptomatic, China figures indicate. *BMJ* 369:m1375. doi: 10.1136/bmj.m1375
- Djordjevic, M., Djordjevic, M., Salom, I., Rodic, A., Zigic, D., Milicevic, O., et al. (2020). COVID-19 puzzle in China: a serendipitous interplay between transmissibility and social distancing measures. arXiv [Preprint]. Available online at: <https://arxiv.org/abs/2005.09630> (accessed December 16, 2020).
- Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet. Infect. Dis.* 20, 533–534. doi: 10.1016/S1473-3099(20)30120-1
- Eubank, S., Eckstrand, I., Lewis, B., Venkatraman, S., Marathe, M., and Barrett, C. (2020). Commentary on Ferguson, et al., “Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand”. *Bull. Math. Biol.* 82:52. doi: 10.1007/s11538-020-00726-x
- European Commission Global Human Settlement (2020). *Urban Centre Database UCDB R2019A* [Online]. Available online at: https://ghsl.jrc.ec.europa.eu/ghs_stat_ucdb2015mt_r2019a.php (accessed September 26, 2020).
- Fan, C., Cai, T., Gai, Z., and Wu, Y. (2020). The relationship between the migrant population's migration network and the risk of COVID-19 transmission in China—empirical analysis and prediction in prefecture-level cities. *Int. J. Environ. Res. Public Health* 17:2630. doi: 10.3390/ijerph17082630
- Fareed, Z., Iqbal, N., Shahzad, F., Shah, S. G. M., Zulfiqar, B., Shahzad, K., et al. (2020). Co-variance nexus between COVID-19 mortality, humidity, and air quality index in Wuhan, China: new insights from partial and multiple wavelet coherence. *Air Qual. Atmos. Health* 8, 1–10. doi: 10.1007/s11869-020-00847-1
- Gallagher, J. (2020, October 11). BCG: can a vaccine from 1921 save lives from Covid-19? *BBC News*.
- Gangemi, S., Billeci, L., and Tonacci, A. (2020). Rich at risk: socio-economic drivers of COVID-19 pandemic spread. *Clin. Mol. Allergy* 18:12. doi: 10.1186/s12948-020-00127-4
- Garufi, G., Carbognin, L., Orlandi, A., Tortora, G., and Brija, E. (2020). Smoking habit and hospitalization for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-related pneumonia: the unsolved paradox behind the evidence. *Eur. J. Intern. Med.* 77, 121–122. doi: 10.1016/j.ejim.2020.04.042
- Gérard, C., Maggipinto, G., and Minon, J.-M. (2020). COVID-19 and ABO blood group: another viewpoint. *Brit. J. Haematol.* 190, e57–e94. doi: 10.1111/bjh.16884
- Global Data Lab (2020). *Subnational Human Development Index (4.0)* [Online]. Available online at: <http://globaldatalab.org/shdi/> (accessed September 27, 2020).
- Goldberg, D. L., Anenberg, S. C., Griffin, D., McLinden, C. A., Lu, Z., and Streets, D. G. (2020). Disentangling the impact of the COVID-19 lockdowns on urban NO₂ from natural variability. *Geophys. Res. Lett.* 47:e2020GL089269. doi: 10.1002/essoar.10503396.1
- Guillon, P., Clément, M., Sébille, V., Rivain, J.-G., Chou, C.-F., Ruvoën-Clouet, N., et al. (2008). Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies. *Glycobiology* 18, 1085–1093. doi: 10.1093/glycob/cwn093
- Guo, X.-J., Zhang, H., and Zeng, Y.-P. (2020). Transmissibility of COVID-19 in 11 major cities in China and its association with temperature and humidity in Beijing, Shanghai, Guangzhou, and Chengdu. *Infect. Dis. Poverty* 9:87. doi: 10.1186/s40249-020-00708-0
- Gupta, A., Banerjee, S., and Das, S. (2020). Significance of geographical factors to the COVID-19 outbreak in India. *Model. Earth Syst. Environ.* 6, 2645–2653. doi: 10.1007/s40808-020-00838-2
- Haidar, Y. M., and Cosman, B. C. (2011). Obesity epidemiology. *Clin. Colon Rectal Surg.* 24, 205–210. doi: 10.1055/s-0031-1295684
- Hargreaves, S., Kumar, B. N., McKee, M., Jones, L., and Veizis, A. (2020). Europe's migrant containment policies threaten the response to covid-19. *BMJ* 368:m1213. doi: 10.1136/bmj.m1213
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280. doi: 10.1016/j.cell.2020.02.052
- Iqbal, M. M., Abid, I., Hussain, S., Shahzad, N., Waqas, M. S., and Iqbal, M. J. (2020). The effects of regional climatic condition on the spread of COVID-19 at global scale. *Sci. Total Environ.* 739:140101. doi: 10.1016/j.scitotenv.2020.140101
- Jordan, R. E., Adab, P., and Cheng, K. K. (2020). Covid-19: risk factors for severe disease and death. *BMJ* 368:m1198. doi: 10.1136/bmj.m1198
- Jurak, G., Morrison, S. A., Leskošek, B., Kovač, M., Hadžić, V., Vodičar, J., et al. (2020). Physical activity recommendations during the coronavirus disease-2019 virus outbreak. *J. Sport Health Sci.* 9, 325–327. doi: 10.1016/j.jshs.2020.05.003
- Keeling, M. J., and Rohani, P. (2011). *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press. doi: 10.2307/j.ctvc4m4gk0
- Khajanchi, S., Bera, S., and Roy, T. K. (2020a). Mathematical analysis of the global dynamics of a HTLV-I infection model, considering the role of cytotoxic T-lymphocytes. *Math. Comput. Simulat.* 180, 354–378. doi: 10.1016/j.matcom.2020.09.009
- Khajanchi, S., and Sarkar, K. (2020). Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India. *Chaos* 30:071101. doi: 10.1063/5.0016240
- Khajanchi, S., Sarkar, K., Mondal, J., and Perc, M. (2020b). Dynamics of the COVID-19 pandemic in India. arXiv [Preprint]. Available at: <https://arxiv.org/abs/2005.06286> (accessed December 16, 2020). doi: 10.21203/rs.3.rs-27112/v1
- Kimball, A., Hatfield, K. M., Arons, M., James, A., Taylor, J., Spicer, K., et al. (2020). Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility—King County, Washington, March 2020. *MMWR Morb. Mortal. Wkly Rep.* 69, 377–381. doi: 10.15585/mmwr.mm6913e1
- Latz, C. A., DeCarlo, C., Boitano, L., Png, C. M., Patell, R., Conrad, M. F., et al. (2020). Blood type and outcomes in patients with COVID-19. *Ann. Hematol.* 99, 2113–2118. doi: 10.1007/s00277-020-04169-1
- Li, M., Zhang, Z., Cao, W., Liu, Y., Du, B., Chen, C., et al. (2020). Identifying novel factors associated with COVID-19 transmission and fatality using the machine learning approach. *Sci. Total Environ.* 142810. doi: 10.1016/j.scitotenv.2020.142810. [Epub ahead of print].
- Lin, S., Wei, D., Sun, Y., Chen, K., Yang, L., Liu, B., et al. (2020). Region-specific air pollutants and meteorological parameters influence COVID-19: a study from mainland China. *Ecotoxicol. Environ. Saf.* 204:111035. doi: 10.1016/j.ecoenv.2020.111035
- Liu, Y., Ning, Z., Chen, Y., Guo, M., Gali, N. K., Sun, L., et al. (2020). Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature* 582, 557–560. doi: 10.1038/s41586-020-2271-3
- Lowen, A. C., Mubareka, S., Steel, J., and Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* 3, 1470–1476. doi: 10.1371/journal.ppat.0030151
- Maier, B. F., and Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* 368, 742–746. doi: 10.1126/science.abb4557
- Martcheva, M. (2015). *An Introduction to Mathematical Epidemiology*. Boston, MA: Springer. doi: 10.1007/978-1-4899-7612-3
- Maslov, S., and Goldenfeld, N. (2020). Window of opportunity for mitigation to prevent overflow of ICU capacity in Chicago by COVID-19. medRxiv [Preprint]. Available online at: <https://europepmc.org/article/ppr/ppr118721> (accessed December 16, 2020).
- McKinney, K. R., Yu, Y. G., and Lewis, T. G. (2006). Environmental transmission of SARS at amoy gardens. *J. Environ. Health* 68, 26–30.
- Mecenas, P., Bastos, R., Vallinoto, A. C. R., and Normando, D. (2020). Effects of temperature and humidity on the spread of COVID-19: a systematic review. *PLoS ONE* 15:e0238339. doi: 10.1371/journal.pone.0238339
- Mendonça, F., Anjos, M., Collischonn, E., Murara, P., DE, F., Limberger, L., et al. (2020). Climate and Covid-19—upgrade and solar radiation influences based on Brazil cases. Research Square [Preprint]. Available online at: <https://europepmc.org/article/ppr/ppr174953> (accessed December 16, 2020).

- Najafimehr, H., Mohamed Ali, K., Safari, S., Yousefifard, M., and Hosseini, M. (2020). Estimation of basic reproduction number for COVID-19 and the reasons for its differences. *Int. J. Clin. Pract.* 74:e13518. doi: 10.1111/ijcp.13518
- NASA Langley Research Center (2020). *The Prediction of Worldwide Energy Resources (POWER) Project [Online]*. Available online at: <https://power.larc.nasa.gov/> (accessed October 6, 2020).
- Oliveiros, B., Caramelo, L., Ferreira, N. C., and Caramelo, F. (2020). Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. medRxiv [Preprint]. Available online at: <https://europepmc.org/article/ppr/ppr116498> (accessed December 16, 2020).
- O'Neill, L. A., and Netea, M. G. (2020). BCG-induced trained immunity: can it offer protection against COVID-19? *Nat. Rev. Immunol.* 20, 335–337. doi: 10.1038/s41577-020-0337-y
- Our World in Data (2020). *Research and Data to Make Progress Against the World's Largest Problems [Online]*. Available online at: <https://ourworldindata.org/> (accessed May, 2020).
- Pansini, R., and Fornacca, D. (2020). COVID-19 higher induced mortality in Chinese regions with lower air quality. medRxiv [Preprint]. Available online at <https://www.medrxiv.org/content/10.1101/2020.05.28.20115832v2> (accessed December 16, 2020).
- Pawelec, G., and Larbi, A. (2008). Immunity and ageing in man: annual review 2006/2007. *Exp. Gerontol.* 43, 34–38. doi: 10.1016/j.exger.2007.09.009
- Perkins, T. A., and España, G. (2020). Optimal control of the COVID-19 pandemic with non-pharmaceutical interventions. *Bull. Math. Biol.* 82:118. doi: 10.1007/s11538-020-00795-y
- Pica, N., and Bouvier, N. M. (2012). Environmental factors affecting the transmission of respiratory viruses. *Curr. Opin. Virol.* 2, 90–95. doi: 10.1016/j.coviro.2011.12.003
- Pourghasemi, H. R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S. R., Babaei, S., et al. (2020). Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *Int. J. Infect. Dis.* 98, 90–108. doi: 10.1016/j.ijid.2020.06.058
- Pranata, R., Lim, M. A., Huang, I., Raharjo, S. B., and Lukito, A. A. (2020). Hypertension is associated with increased mortality and severity of disease in COVID-19 pneumonia: a systematic review, meta-analysis and meta-regression. *J. Renin Angiotensin Aldosterone Syst.* 21:1470320320926899. doi: 10.1177/1470320320926899
- Qu, G., Li, X., Hu, L., and Jiang, G. (2020). An imperative need for research on the role of environmental factors in transmission of novel coronavirus (COVID-19). *Environ. Sci. Technol.* 54, 3730–3732. doi: 10.1021/acs.est.0c01102
- Radenkovic, D., Chawla, S., Pirro, M., Sahebkar, A., and Banach, M. (2020). Cholesterol in relation to COVID-19: should we care about it? *J. Clin. Med.* 9:1909. doi: 10.3390/jcm9061909
- Ran, J., Song, Y., Zhuang, Z., Han, L., Zhao, S., Cao, P., et al. (2020a). Blood pressure control and adverse outcomes of COVID-19 infection in patients with concomitant hypertension in Wuhan, China. *Hypertens. Res.* 43, 1267–1276. doi: 10.1038/s41440-020-00541-w
- Ran, J., Zhao, S., Han, L., Qiu, Y., Cao, P., Yang, Z., et al. (2020b). Effects of particulate matter exposure on the transmissibility and case fatality rate of COVID-19: a Nationwide Ecological Study in China. *J. Travel Med.* 27:taaa133. doi: 10.1093/jtm/taaa133
- Rashed, E. A., Kodera, S., Gomez-Tames, J., and Hirata, A. (2020). Influence of absolute humidity, temperature and population density on COVID-19 spread and decay durations: multi-prefecture study in Japan. *Int. J. Environ. Res. Public Health* 17:5354. doi: 10.3390/ijerph17155354
- Ruiz, J. I., Nuhu, K., McDaniel, J. T., Popoff, F., Izcovich, A., and Criniti, J. M. (2015). Inequality as a powerful predictor of infant and maternal mortality around the world. *PLoS ONE* 10:e0140796. doi: 10.1371/journal.pone.0140796
- Rychter, A. M., Zawada, A., Ratajczak, A. E., Dobrowolska, A., and Krela-Kazmierczak, I. (2020). Should patients with obesity be more afraid of COVID-19? *Obes. Rev.* 21:e13083. doi: 10.1111/obr.13083
- Sagar, A. D., and Najam, A. (1998). The human development index: a critical review. *Ecol. Econ.* 25, 249–264. doi: 10.1016/S0921-8009(97)00168-7
- Sagripanti, J.-L., and Lytle, C. D. (2007). Inactivation of influenza virus by solar radiation. *Photochem. Photobiol.* 83, 1278–1282. doi: 10.1111/j.1751-1097.2007.00177.x
- Samui, P., Mondal, J., and Khajanchi, S. (2020). A mathematical model for COVID-19 transmission dynamics with a case study of India. *Chaos Solitons Fractals* 140:110173. doi: 10.1016/j.chaos.2020.110173
- Sarkar, K., Khajanchi, S., and Nieto, J. J. (2020). Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fractals* 139:110049. doi: 10.1016/j.chaos.2020.110049
- Sattar, N., McInnes, I. B., and McMurray, J. J. (2020). Obesity a risk factor for severe COVID-19 infection: multiple potential mechanisms. *Circulation* 142, 4–6. doi: 10.1161/CIRCULATIONAHA.120.047659
- Scheffer, M. (2009). *Critical Transitions in Nature and Society*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400833276
- Schiffirin, E. L., Flack, J. M., Ito, S., Muntner, P., and Webb, R. C. (2020). Hypertension and COVID-19. *Am. J. Hypertens.* 33, 373–374. doi: 10.1093/ajh/hpaa057
- Setti, L., Passarini, F., De Gennaro, G., Barbieri, P., Perrone, M. G., Borelli, M., et al. (2020). SARS-Cov-2RNA found on particulate matter of Bergamo in Northern Italy: first evidence. *Environ. Res.* 188:109754. doi: 10.1016/j.envres.2020.109754
- Shahid, Z., Kalayanamitra, R., McClafferty, B., Kepko, D., Ramgobin, D., Patel, R., et al. (2020). COVID-19 and older adults: what we know. *J. Am. Geriatr. Soc.* 68, 926–929. doi: 10.1111/jgs.16472
- Singh, K., and Agarwal, A. (2020). Impact of weather indicators on the COVID-19 outbreak: a multi-state study in India. medRxiv [Preprint]. Available online at <https://europepmc.org/article/ppr/ppr176027> (accessed December 16, 2020).
- Thangriyal, S., Rastogi, A., Tomar, A., and Baweja, S. (2020). Impact of temperature and sunshine duration on daily new cases and death due to COVID-19. medRxiv [Preprint]. Available online at: <https://europepmc.org/article/ppr/ppr176097> (accessed December 16, 2020).
- The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly* [Online], 2. Available online at: <https://slma.lk/wp-content/uploads/2020/02/TheEpidemiologicalCharacteristicsofanOutbreakof2019NovelCoronavirusDiseases28COVID-1929E28094China2C20201.pdf>. doi: 10.46234/ccdcw2020.032
- Tian, H., Liu, Y., Li, Y., Wu, C. H., Chen, B., Kraemer, M. U. G., et al. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 368, 638–642. doi: 10.1126/science.abb6105
- Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, O. A. I., Lestari, H., Bahar, H., et al. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci. Total Environ.* 725:138436. doi: 10.1016/j.scitotenv.2020.138436
- van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., et al. (2020). Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* 382, 1564–1567. doi: 10.1056/NEJMc2004973
- Ward, D. (2020). *Actions Speak Louder Than Age: Explaining Wide Variations in COVID-19 Deaths* [Online]. Available online at: https://www.researchgate.net/publication/341599695_Actions_Speak_Louder_than_Age_Explaining_Wide_Variations_in_COVID-19_Deaths (accessed December 16, 2020).
- Weitz, J. S., Beckett, S. J., Coenen, A. R., Demory, D., Dominguez-Mirazo, M., Dushoff, J., et al. (2020). Modeling shield immunity to reduce COVID-19 epidemic spread. *Nat. Med.* 26, 849–854. doi: 10.1038/s41591-020-0895-3
- World Bank (2020a). *Global Economic Prospects, June 2020*. Washington, DC: World Bank.
- World Bank (2020b). *World Bank Open Data* [Online]. Available online at: <https://www.worldbank.org/> (accessed May, 2020).
- World Health Organization (2020). *The Global Health Observatory: Explore a World of Health Data* [Online]. Available online at: <https://www.who.int/data/gho> (accessed October 6, 2020).
- World Weather Online (2020). *Local Weather History API* [Online]. Available online at: <https://www.worldweatheronline.com/developer/api/historical-weather-api.aspx> (accessed December 9, 2020).

- Worldometer (2020). *COVID-19 Coronavirus Pandemic* [Online]. Available online at: <https://www.worldometers.info/coronavirus/> (accessed June 14, 2020).
- Wu, X., Nethery, R. C., Sabath, M. B., Braun, D., and Dominici, F. (2020). Air pollution and COVID-19 mortality in the United States: strengths and limitations of an ecological regression analysis. *Sci. Adv.* 6:eabd4049. doi: 10.1126/sciadv.abd4049
- Yao, Y., Pan, J., Liu, Z., Meng, X., Wang, W., Kan, H., et al. (2020a). Ambient nitrogen dioxide pollution and spread ability of COVID-19 in Chinese cities. *Ecotoxicol. Environ. Saf.* 208:111421. doi: 10.1016/j.ecoenv.2020.111421
- Yao, Y., Pan, J., Liu, Z., Meng, X., Wang, W., Kan, H., et al. (2020b). No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. *Eur. Respir. J.* 55:2000517. doi: 10.1183/13993003.00517-2020
- Yao, Y., Pan, J., Wang, W., Liu, Z., Kan, H., Qiu, Y., et al. (2020c). Association of particulate matter pollution and case fatality rate of COVID-19 in 49 Chinese cities. *Sci. Total Environ.* 741:140396. doi: 10.1016/j.scitotenv.2020.140396
- Yu, Q., Salvador, C., Melani, I., Berg, M., Neblett, E., and Kitayama, S. (2020). Racial residential segregation and economic disparity jointly exacerbate the COVID-19 fatality in large American cities. *PsyArXiv [Preprint]*. Available online at: <https://psyarxiv.com/xgbpy/> (accessed December 16, 2020).
- Zheng, Z., Peng, F., Xu, B., Zhao, J., Liu, H., Peng, J., et al. (2020). Risk factors of critical and mortal COVID-19 cases: a systematic literature review and meta-analysis. *J. Infect.* 81:e16–e25. doi: 10.1016/j.jinf.2020.04.021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Salom, Rodic, Milicevic, Zigic, Djordjevic and Djordjevic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

PM_{2.5} as a major predictor of COVID-19 basic reproduction number in the USA

Ognjen Milicevic¹, Igor Salom², Marko Tumbas³, Andjela Rodic³, Sofija Markovic³, Dusan Zigic², Magdalena Djordjevic², Marko Djordjevic^{3,*}

¹Department for Medical Statistics and Informatics, School of Medicine, University of Belgrade, Serbia

²Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Serbia

³Quantitative Biology Group, Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Serbia

* Correspondence:

Marko Djordjevic, e-mail: dmarko@bio.bg.ac.rs

Keywords: COVID-19 pollution dependence, outdoor air pollutants, basic reproduction number, principal component analysis, machine learning

Abstract

Many studies have proposed a relationship between COVID-19 transmissibility and ambient pollution levels. However, a major limitation in establishing such associations is to adequately account for complex disease dynamics, influenced by e.g. significant differences in control measures and testing policies. Another difficulty is appropriately controlling the effects of other potentially important factors, due to both their mutual correlations and a limited dataset. To overcome these difficulties, we will here use the basic reproduction number (R_0) that we estimate for USA states using non-linear dynamics methods. To account for a large number of predictors (many of which are mutually strongly correlated), combined with a limited dataset, we employ machine-learning methods. Specifically, to reduce dimensionality without complicating the variable interpretation, we employ Principal Component Analysis on subsets of mutually related (and correlated) predictors. Methods that allow feature (predictor) selection, and ranking their importance, are then used, including both linear regressions with regularization and feature selection (Lasso and Elastic Net) and non-parametric methods based on ensembles of weak-learners (Random Forest and Gradient Boost). Through these substantially different approaches, we robustly obtain that PM_{2.5} levels are the main predictor of R_0 in USA states, with corrections from factors such as other pollutants, prosperity measures, population density, chronic disease levels, and possibly racial composition. As a rough magnitude estimate, we obtain that a relative change in R_0 , with variations in pollution levels observed in the USA, is typically ~30%, going up to 70%, which further underscores the importance of pollution in COVID-19 transmissibility.

1 Introduction

Many studies have provided evidence or strong arguments for the importance of pollution (primarily $PM_{2.5}$ and to a lesser degree PM_{10} and NO_2) in COVID-19 transmissibility: *i*) Droplets with virus particles may bind to Particulate Matter (PM), which may promote the diffusion of virus droplets in the air (Chen et al., 2010; Comunian et al., 2020; Contini and Costabile, 2020). *ii*) Once the virus droplet bound to PM reaches a susceptible individual, it can penetrate deeper in alveolar and tracheobronchial regions (Qu et al., 2020). *iii*) Pollution has a general effect on weakening the immune system making the organism more susceptible to infection (Domingo and Rovira, 2020; Paital and Agrawal, 2020; Qu et al., 2020). *iv*) It promotes overexpression of ACE-2 receptors, which allows SARS-CoV-2 binding and entry into cells (Comunian et al., 2020; Paital and Agrawal, 2020; Sagawa et al., 2021).

While these arguments are compelling, and several studies pointed to correlations between pollutant levels and increased severity of COVID-19 progression (De Angelis et al., 2021; Lorenzo et al., 2021; Tello-Leal and Macías-Hernández, 2020; Yao et al., 2021; Zhu et al., 2020), there are also prominent methodological difficulties in establishing this link. Specifically, comparing case counts (Adhikari and Yin, 2020; Suhaimi et al., 2020) in different geographical regions may be influenced by significant differences in the epidemic onsets, applied control measures (social distancing and similar), and testing methodologies (most significantly the number of performed tests). Consequently, adequately controlling the infection dynamics, rather than relying on absolute case counts, is crucial. Secondly, confounding factors, which have to be used jointly with pollution in assessing transmissibility, such as social-demographic, medical, and meteorological variables can be mutually highly correlated (Salom et al., 2021). Such high correlations realistically present a problem for any statistical inference method, though modern machine-learning approaches can partially account for this difficulty (Gupta and Gharehgozli, 2020). Also, to obtain robust predictions that are not an artifact of the applied methodology and the underlying assumptions, it is crucial to perform analysis by several independent methods.

We will accordingly employ the following approach: *i*) We consider USA states as this dataset has sufficient variability in the relevant variables to extract reasonable conclusions while heterogeneities in sociodemographic and weather parameters are not too large to overshadow the dependence on pollution. *ii*) As a measure of transmissibility, we use the basic reproduction number (R_0), which is insensitive to specific testing policies and estimates SARS-CoV-2 transmissibility in the absence of social distancing and in a completely susceptible (non-resistant) population. To estimate R_0 for individual USA states, we will apply our previously developed methodology (Salom et al., 2021) based on observation of different dynamical regimes in COVID-19 infection counts during the disease outburst and disease dynamics model applied to one of these growth regimes (exponential). These R_0 estimates, instead of the disease counts (or other similar measures), will be used as a dependent (response) variable in further analysis. As independent (input) variables, we will assemble a large set of available sociodemographic, medical, and weather variables. Importantly, to assess the pollution levels in detail, we will assemble the data for ten different pollutants, with the levels determined in the time windows relevant for the analyzed exponential growth regimes. The weather parameters will be assembled in the same dynamically relevant manner. This will result in a large number of predictors, many of which are grouped in sets of similar and mutually often highly correlated variables. Additionally, the number of assembled variables will exceed the total sample size, so reducing the number of predictors to a smaller and less correlated set will become a priority. We will achieve this through data preprocessing (feature engineering), which includes variable transformations, removing all outliers, and grouping variables in mutually related and correlated subsets (e.g., measures of similarity, population age, prosperity, chronic disease). Principal Component Analysis (PCA) will be

applied on these subsets, resulting in dimensionality reduction (reducing the number of predictors) and smaller correlations within these reduced predictor sets. Finally, established machine learning approaches will be used with the goals to *i*) select important variables and rank their relative importance in explaining R_0 , *ii*) obtain an estimate of expected changes in R_0 based on observed variability in pollution levels. While this will be a rough estimate, due to the inability to assemble all relevant factors in determining R_0 , it will provide a quantitative assessment for the importance of pollution in SARS-CoV-2 transmissibility.

2 Methods

2.1 R_0 extraction

Basic reproduction number (R_0) is a measure of SARS-CoV-2 transmissibility in a fully susceptible population and in the absence of intervention measures (social distancing, quarantine). For extraction of R_0 , we used our previously published methodology, in particular analysis of widespread infection growth regimes (Magdalena Djordjevic et al., 2021) and extraction of R_0 from the exponential growth phase that we previously applied on a worldwide level (Salom et al., 2021). We below for completeness summarize this methodology.

To describe the SARS-CoV-2 transmission in a population, we constructed an adapted version of an SEIR compartmental model (Maier and Brockmann, 2020; Maslov and Goldenfeld, 2020; Perkins and España, 2020; Tian et al., 2020; Weitz et al., 2020), which takes into account all the relevant features of this process, while being simple enough to be used for R_0 estimation in a wide range of populations (Magdalena Djordjevic et al., 2021; Salom et al., 2021). In the early stages of epidemics and before social distancing measures are introduced, the flow between the model compartments leads to the changes of the compartment member abundances S (susceptible), E (exposed), I (infected), R (recovered), and D (cumulative detected cases) which are described by the following system of ordinary differential equations:

$$\frac{dS}{dt} = - \frac{\beta SI}{N} \quad (1.1.)$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E \quad (1.2.)$$

$$\frac{dI}{dt} = \sigma E - \gamma I \quad (1.3.)$$

$$\frac{dR}{dt} = \gamma I \quad (1.4.)$$

$$\frac{dD}{dt} = \varepsilon \delta I \quad (1.5.)$$

where N is the population size. Parameters represent: β - the rate of virus transmission from an infected to the encountered susceptible individual, σ - the inverse of the average incubation period (~ 3 days), γ - the inverse of the average period of infectiousness, ε - the detection efficiency (as not every infected individual becomes detected), and δ - the detection rate.

We here apply the model to the relatively brief, initial epidemics period when government mitigation measures were still absent, the virus spreads according to its biological potential, the characteristics of the particular population, and the environment. Therefore, the parameter values are considered constant in this period. The standard measure of the virus transmissibility in these, uncontrolled conditions is the basic reproduction number, R_0 , defined as the average number of secondary infections caused by a primary infected individual in a fully susceptible population ($S/N \approx 1$), and in the absence of social distancing measures (also sometimes denoted as $R_{0,free}$) (Maier and Brockmann, 2020). At the start of an epidemic, $R_0 > 1$ and the number of the infected individuals grows exponentially. At this time

interval, the model can be linearized by invoking $S/N \approx 1$, becoming represented by only two linear differential equations (1.2) and (1.3). Solving for the eigenvalues of this system,

$$\lambda_{\pm} = \frac{-(\gamma + \sigma) \pm \sqrt{(\gamma - \sigma)^2 + 4\beta\sigma}}{2}, \quad (3.7.)$$

provides the solution of the form $I(t) = C_1 \cdot e^{\lambda_+ t} + C_2 \cdot e^{\lambda_- t}$, which can be approximated by

$$I(t) = I(0) \cdot e^{\lambda_+ t} \quad (3.8.)$$

where the term containing the negative eigenvalue, λ_- can be neglected (see (Salom et al., 2021)). With $R_0 = \beta/\gamma$ (Keeling and Rohani, 2011; Martcheva, 2015), the equation for the basic reproduction number,

$$R_0 = 1 + \frac{\lambda_+ \cdot (\gamma + \sigma) + \lambda_+^2}{\gamma * \sigma}. \quad (3.9.)$$

can be obtained by expressing β from Eq. (3.7).

To estimate the R_0 values for 46 US states, we collect the detected case counts for each state from (Worldometer, 2020). The solution $D(t) = \varepsilon \cdot \delta \cdot I(0) \cdot (e^{\lambda_+ t} - 1)/\lambda_+$ of the Eq. (1.5) using the Eq. (3.8) models the dependence of the cumulative number of detected with time. Taking its logarithm,

$$\log(D(t)) = \log(\varepsilon \delta I(0)/\lambda_+) + \lambda_+ \cdot t, \quad (3.12.)$$

results in the equation of the straight line that can be fitted to the data on the semilogarithmic scale. Notably, the slope of that line is given by the positive eigenvalue of the system, λ_+ . Once that λ_+ is, thus, determined by fitting, the value of R_0 for a particular state can be calculated from Eq. (3.9).

2.1 Pollution data collection

Air quality information was obtained from the US environmental protection agency (EPA) Air Data service (US Environmental Protection Agency, 2020) . We used aggregated daily data for pollutant gases (O_3 , NO_2 , SO_2 , CO), particulates ($PM_{2.5}$ and PM_{10}) and other available species (VOCs, NO_x , and HAPs). Aggregation was done over all the cities with available information (smaller in comparison to the cities used for weather data). The population was obtained from the US Census Bureau (U.S. Census Bureau, 2020). All the variable values are averaged for each city over the identified time period, and the state average is calculated as the average of all included state cities weighted by population.

2.4 Weather data collection

Weather parameters were downloaded in bulk using a custom Python script from the NASA POWER project service (NASA Langley Research Center, 2020) . All the parameters were downloaded via the POWER API at the longitude and latitude coordinates matching the largest cities in each state that comprise above 10% of the state population. Variables include temperature at 2m and 10m, measures of humidity and precipitation (wet bulb temperature, relative humidity, total precipitation), and insolation indices. The maximum predicted UV index was downloaded from OpenUV (OpenUV, 2020). Geographical coordinates of the cities and populations of cities and states were adapted from Wikidata (Wikipedia, 2021a, 2021b).

1.3 Socio-demographic data collection

Demographic data was collected from several sources. The demographic composition of the U.S. population by gender, race, and percentage of the population under 18 and over 65 was taken from the Measure of America, a project of The Social Science Research Council website (Measure of America,

2018). Information about health insurance, GDP, life expectancy at birth, infant and child mortality was also taken from the Measure of America website. Medical parameters such as hypertension, cholesterol, cardiovascular disease, diabetes, cancer, obesity, inactivity, and chronic kidney and obstructive pulmonary disease were taken from the Americas Health Rankings website (America’s Health Ranking, 2021) hosting Centers for Disease Control and Prevention (CDC) data (CDC, 2019). Percentages of the population that are actively smoking and consuming alcohol are taken from the same source as well. The percentage of the foreign population was taken from the Census Reporter website (U.S. Census Bureau, 2019). The subnational HDI was taken from the Global Data Lab website (2020) (Smits and Permanyer, 2019). Population density, urban population percentage, and median age were taken from the U.S. Census Bureau website (U.S. Census Bureau, Population Division, 2019).

2.5 Data processing

Analysis of data distributions and QQ plots reveals non-normal distributions in a majority of variables. To reduce the skewness of the data we applied a number of transforms with different strengths (square root, cubic root, or log), adjusted in sign to maintain the data ranking (Spearman correlation). The table with all applied transformations is provided below. Also, note that the entire dataset used in this analysis (variable values for all 46 states) is provided in Supplement Table 1. In addition to the transformations applied, the table below also links the variables to the dataset, by relating a variable shortcut (used in Supplemental files) with its full name and units.

Data	Name(units)	Transformation f(x)
T2M, T2M _{MAX} , T2M _{MIN} , T10M, T10M _{MAX} , T10M _{MIN} , TS, T2MWET	Temperatures (°C)	None
RH2M	Relative humidity at 2 meters (%)	-log(max(x) - x)
QV2M	Specific humidity at 2 meters (g/kg)	log(x)
T2MDEW	Dew Point (°C)	None
PRECTOT	Precipitation (mm/day)	$x^{1/3}$
TQV	Total Column Precipitable Water (cm)	log(x)
CLRSKY_SFC_SW_DWN	Clear Sky Insolation Incident on a Horizontal Surface (MJ/m ² /day)	log(x)
ALLSKY_SFC_LW_DWN	Downward Thermal Infrared (Longwave) Radiative Flux (MJ/m ² /day)	-(max(x) - x) ^{1/3}
ALLSKY_SFC_SW_DWN	All Sky Insolation Incident on a Horizontal Surface (MJ/m ² /day)	log(x)
OpenUV _{max}	UV radiation	$x^{1/3}$
WS2M	Wind speed at 2 meters	None
WS10M	Wind speed at 10 meters	None
Population over 65 (%)	Population over 65 (%)	None
Life Expectancy	Life Expectancy at Birth (years)	-(max(x) - x) ^{1/2}
Median age	Median age (years)	-(max(x) - x) ^{1/2}
Youth population	Population under 18 (%)	log(x)
Population density	Population density (people/km ²)	log(x)
BUAPC	Built Up Area Per Capita (km ² /people)	log(x)

Pollution influence on COVID-19 R_0

Urban Population	Urban Population (%)	$-(\max(x) - x)^{1/2}$
HDI	Human development index (0-1) Average of education, health and standard of living. (Mean years of schooling of adults aged 25+, Expected years of schooling of children aged 6 + Life expectancy at birth + GNIpc) /3	$-(\max(x) - x)^{1/2}$
GDPpc	Gross domestic product per capita	$\log(x)$
Infant mortality rate	Infant Mortality Rate (per 1,000 live births)	$-\log(x)$
Child mortality	Child Mortality (age 1-4, per 1000 population)	$-\log(x)$
Alcohol consumption	Adults alcohol consumption binge drinking (%)	$\log(x)$
Foreign-born population	Foreign-born population (%)	$\log(x)$
Obesity	Obesity age 20 and older (%)	None
CVD deaths	Age 65+ Cardiovascular disease deaths per 100000 people	$\log(x)$
Hypertension	Adults with Hypertension (%)	$\log(x)$
High cholesterol	Population with high cholesterol (%)	None
Smoking	Population smoking (%)	None
Cardiovascular disease	Population with cardiovascular disease (%)	None
Diabetes	Population with diabetes (%)	$x^{1/3}$
Cancer	Population with cancer (%)	None
Chronic kidney disease	Population with chronic kidney disease (%)	$x^{1/2}$
Chronic obstructive pulmonary disease	Population with chronic obstructive pulmonary disease (%)	$\log(x)$
Multiple chronic conditions	Population with multiple chronic conditions (%)	None
Physical inactivity	Population physically inactive (%)	$x^{1/3}$
Male percent	Fraction of male in the population (%)	$\log(x)$
White percent	Fraction of white in the population (%)	$-\log(\max(x) - x)$
Black percent	Fraction of black in the population (%)	$x^{1/3}$
Native percent	Fraction of native in the population (%)	$\log(x)$
Asian percent	Fraction of Asian in the population (%)	$\log(x)$
Latino percent	Fraction of Latino in the population (%)	$\log(x)$
No health insurance children	No health insurance under 18 (%)	$x^{1/2}$
No health insurance adults	No health insurance 18-64 (%)	None
No health insurance all	No health insurance all population (%)	None

No insurance black	No health insurance black (%)	None
No insurance native	No health insurance native (%)	$x^{1/3}$
No insurance Asian	No health insurance Asian (%)	$x^{1/2}$
No insurance Latino	No health insurance Latino (%)	None
No insurance white	No health insurance white (%)	None
PM _{2.5}	PM _{2.5} concentration ($\mu\text{g}/\text{m}^3$)	None
PM ₁₀	PM ₁₀ concentration ($\mu\text{g}/\text{m}^3$)	$x^{1/2}$
CO	CO concentration (ppm, 10^{-6})	$x^{1/2}$
NO ₂	NO ₂ concentration (ppb, 10^{-9})	None
SO ₂	SO ₂ concentration (ppb)	$\log(x - \min(x))$
O ₃	O ₃ concentration (ppm)	None
VOC	Volatile organic compounds concentration (ppb Carbon)	$\log(x)$
Lead	Lead concentration ($\mu\text{g}/\text{m}^3$)	$\log(x)$
HAPs	Hazardous air pollutants concentration ($\mu\text{g}/\text{m}^3$)	$x^{1/2}$
NONOxNOy	Nitrous oxides concentration (ppb)	$x^{1/3}$
R ₀	Estimated basic reproduction number	$\log(x)$

Table 1. List of variables (with units) and the applied transformations. Variable shortcuts (first column) correspond to Supplement Table 1.

Individual data values that remained more than three median absolute deviations from the new median were substituted by the said median value.

2.6 Feature engineering and Principal Components Analysis

To reduce the number of variables, we divided them into groups by conceptual similarity and expected correlation and performed Principle Component Analysis (PCA) on each group. This also partially reduced data correlation (Joliffe, 2002). Variables are grouped according to two criteria: *i*) They represent similar quantities so that after PCA interpretation of the obtained PC remains unambiguous *ii*) That correlation between the variables in the same group are high, so that in this way after PCA the correlations in the predictor set are substantially reduced. Grouping of variables and their relation to PCA is provided in Table 2 below:

PC components	Variables
PC1 T	T2M, T2M _{MAX} , T2M _{MIN} , T10M, T10M _{MAX} , T10M _{MIN} , TS
PC1 humidity	QV2M, T2MDEW
PC1 percipitation	PRECTOT, TQV
PC1 radiation PC2 radiation	CLRSKY_SFC_SW_DWN, ALLSKY_SFC_SW_DWN, ALLSKY_SFC_LW_DWN
PC1 seasonality PC2 seasonality	PC1 T, PC1 humidity, PC1 percipitation, PC1 radiation, PC2 radiation, RH2M, UV
PC1 age	Population over 65, Youth population, Median age

PC2 age	
PC1 density PC2 density	1/BUCAP, Urban population, Population density
PC1 prosperity PC2 prosperity PC3 prosperity PC4 prosperity	Life expectancy, Infant mortality, GDP, HDI, Child mortality, Alcohol consumption, Foreign born population
PC1 disease PC2 disease PC3 disease PC4 disease	Obesity (% age 20 and older), Age 65+ CVD deaths, Adults with hypertension (%), Population with high cholesterol (%), Population smoking (%), Population with cardiovascular disease (%), Population with diabetes%, Population with cancer (%), Population chronic kidney disease (%), Population chronic obstructive pulmonary disease (%), Population multiple chronic conditions (%), Population physical inactivity (%),
PC1 ins. PC2 ins. PC3 ins.	No health insurance (% of _children_ under_18), No health insurance (% of _adults_ ages_18–64), No health insurance total population (%), No health insurance black (%), No health insurance native (%), No health insurance Asian (%), No health insurance Latino (%), No health insurance white (%),

Table 2. Grouping of variables and relation to PC.

Since different variables are expressed in different units and correspond to diverse scales, each variable in the dataset was standardized (the mean subtracted and divided by the standard deviation) before PCA. For each datasets, we retained as many PCs (starting from the most dominant one) as needed to (cumulatively) explain >85% of the data variance. It was inspected that PCs reasonably follow a normal distribution (as expected, based on the transformation of the original variables). Note that a number of variables did not enter any of the groups from Table 2, as they either have a distinct meaning from other variables (e.g. racial prevalence), or have a similar meaning but do not exhibit a high correlation with the related variables (e.g. relative humidity RH2M, which does not correlate well with the other two humidity measures, QV2M and T2MDEW). These variables enter further analysis independently, i.e. together with PCs obtained after PCA on grouped variables.

2.7 LASSO regression

To complement the feature selection already done through PCA, additional L1 regularization was done with Lasso (Tibshirani, 1996). All input variables were standardized. Hyperparameter λ controlling for the model complexity was optimized through grid search on an exponential scale from numerical zero (OLS regression) to the value yielding the intercept-only model. Mean Squared Error (MSE) on the cross-validation testing set (200 repeats, 80-20 split) was taken as the loss function, and we chose the λ_{1SE} as the simplest model still comparable to the optimal one (Krstajic et al., 2014). The final model was comprised of all the non-zero coefficients.

2.8 Elastic net regression

Elastic Net expands the Lasso regression with an L2 regularization and introduces a second hyperparameter α (Friedman et al., 2010; Hastie et al., 2008; Zou and Hastie, 2005). The same preprocessing was done for the input variables, after which the 2-dimensional grid-search with the same λ -scale as in Lasso, and the α linearly equidistant on the interval from 0 (Ridge regression) to 1 (Lasso regression) inclusive. Cross-validation was performed the same as for the Lasso regression, but each fold gave a distinct (α, λ) pair of hyperparameters. The final chosen value is the (α, λ) pair closest to the centroid of all the folds, and these hyperparameters were used to retrain the model on the whole dataset. Again, the final model was comprised of all the non-zero coefficients.

2.9 Random Forest and Gradient Boost

To avoid overfitting, variables were preselected so that they exhibit significant correlations with R_0 (with a liberal threshold of $P < 0.1$), by either Pearson, Kendall, or Spearman correlations. Cross-validation and hyperparameter selection for Gradient Boost and Random Forest (Breiman, 2001, 1996; Freund and Schapire, 1997; Friedman, 2001; Hastie et al., 2008) was done equivalently as for Lasso and Elastic net. For Gradient Boost, maximal number of splits, minimal leaf size, and learning rate were chosen through grid search, with the respective values: $\{1, 2, 3, 4, 5, 8, 16\}$; $\{1, 2, 3, 4, 5, 8, 16, 18\}$; $\{0.1, 0.25, 0.5, 0.75, 1\}$. For Random Forest, the grid values for maximal number of splits and minimal leaf size were, respectively: $\{6, 12, 18, 22, 24, 26, 30, 35\}$, $\{1, 2, \dots, 7\}$. The number of trained decision trees in the ensemble was also chosen to minimize Mean Square Error (MSE) on the testing set, for both methods. The obtained hyperparameters were used to retrain the models on the whole dataset, and predictor importance was estimated for both methods.

2.10 Model metrics

MSE for the testing data, averaged over all cross-validations, was used as a metric to compare the performance of different models. For easier interpretability, MSE values were scaled by those corresponding to the constant model (so that MSE of 1 corresponds to the constant model). To assess statistical significance with respect to the constant model, a t-test was applied to MSE values obtained through cross-validation.

3 Results

3.1 Extraction of R_0 and feature engineering

The $\log(D(t))$ in the exponential growth regime for a subset of selected USA states is shown in Fig. 1. The linear dependence confirms that the progression of the epidemic in this stage is almost perfectly exponential. Moreover, from Fig. 1, we see that this exponential growth in the cumulative number of confirmed cases is robustly observed for a wide range of USA states, while we previously also observed the same robust initial exponential growth for a wide range of world countries (Salom et al., 2021). This exponential growth happens in the early infection stage, when only a small fraction of the population is resistant, and before social distancing interventions take effect. Note that, even after introducing the measures, there is ~ 10 days delay in observing their effect in the confirmed case-counts curve, due to the incubation period and the time needed between the symptom onset and the infection detection/confirmation. We exploit this exponential regime to infer R_0 as described in Methods (see also (Magdalena Djordjevic et al., 2021) and (Salom et al., 2021)), which we further use as our independent (response) variable.

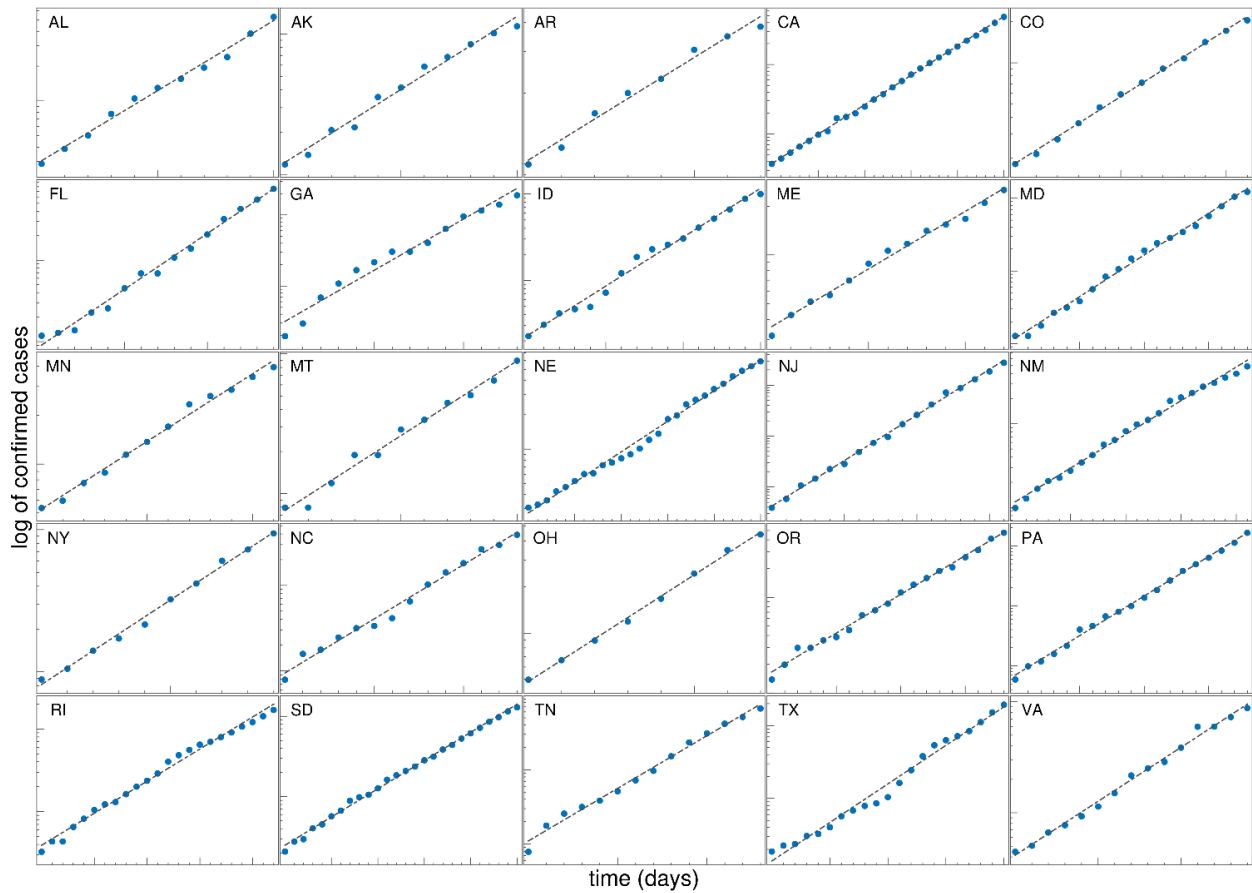


Figure 1. The time dependence of the detected cases for the different US states during the initial period of the epidemic is shown on a log-linear scale. The linear fit of $\log(D)$ shows that the spread of COVID-19 is well approximated by exponential growth in this phase. Values on axes are chosen differently for each state to emphasize the exponential growth phase. For each state, the start and end dates, the extracted slope λ_+ , of the exponential regime, are given in the Supplementary Table S1. AL – Alabama; AK – Arkansas; AR – Arizona; CA – California; CO – Colorado; FL – Florida; GA – Georgia; ID – Idaho; ME – Maine; MD – Maryland; MN – Montana; NE – Nebraska; NJ – New Jersey; NM – New Mexico; NY – New York; NC – North Carolina; OH – Ohio; OR – Oregon; PA – Pennsylvania; RI – Rhode Island; SD – South Dakota; TN – Tennessee; TX – Texas; VA – Virginia.

We further transformed the variables so that their distribution becomes as close to normal as possible and removed the outliers, as detailed in Methods. The main purpose of these transformations, and outliers' removal, is to account for more extreme variable values (such as heavy distribution tails), which may significantly affect some of the analysis methods that we further use (in particular, correlation analysis, Lasso and Elastic net regressions). On the other hand, methods based on the ensembles of decision trees (e.g., Random Forest and Gradient Boost) are fairly robust to outliers and non-normal variable distributions and will provide a consistency check of the obtained conclusions. Note that transformations of all variables are provided in Table 1. Further, the total number of variables (74) is larger than the sample size (46 states). While the regressions with feature selection (Lasso and Elastic net) can handle the number of variables that is significantly larger than the sample size (as long as the number of selected features is smaller than the sample size), this large number of variables (some highly correlated) is a major risk for overfitting, particularly for Random Forest and Gradient Boost methods.

To reduce both the number of variables and correlations between them, we employ PCA. The main disadvantage of this is possibly hard interpretation of the obtained PCs. We, thus, divided the initial set of variables into smaller subsets of similar meaning and high mutual correlation, where the division

of the variables into subgroups is provided in Table 2. Each group of variables is related to their corresponding PCs in that table. Note that the number of PCs for each variable group is chosen such that they explain more than 85% of the variability in the dataset (standard threshold). Finally, to each of the PCs, we assign an intuitive name (e.g., PC1 prosperity, PC1 age) according to the set of variables from which they are formed.

3.2 Feature extraction

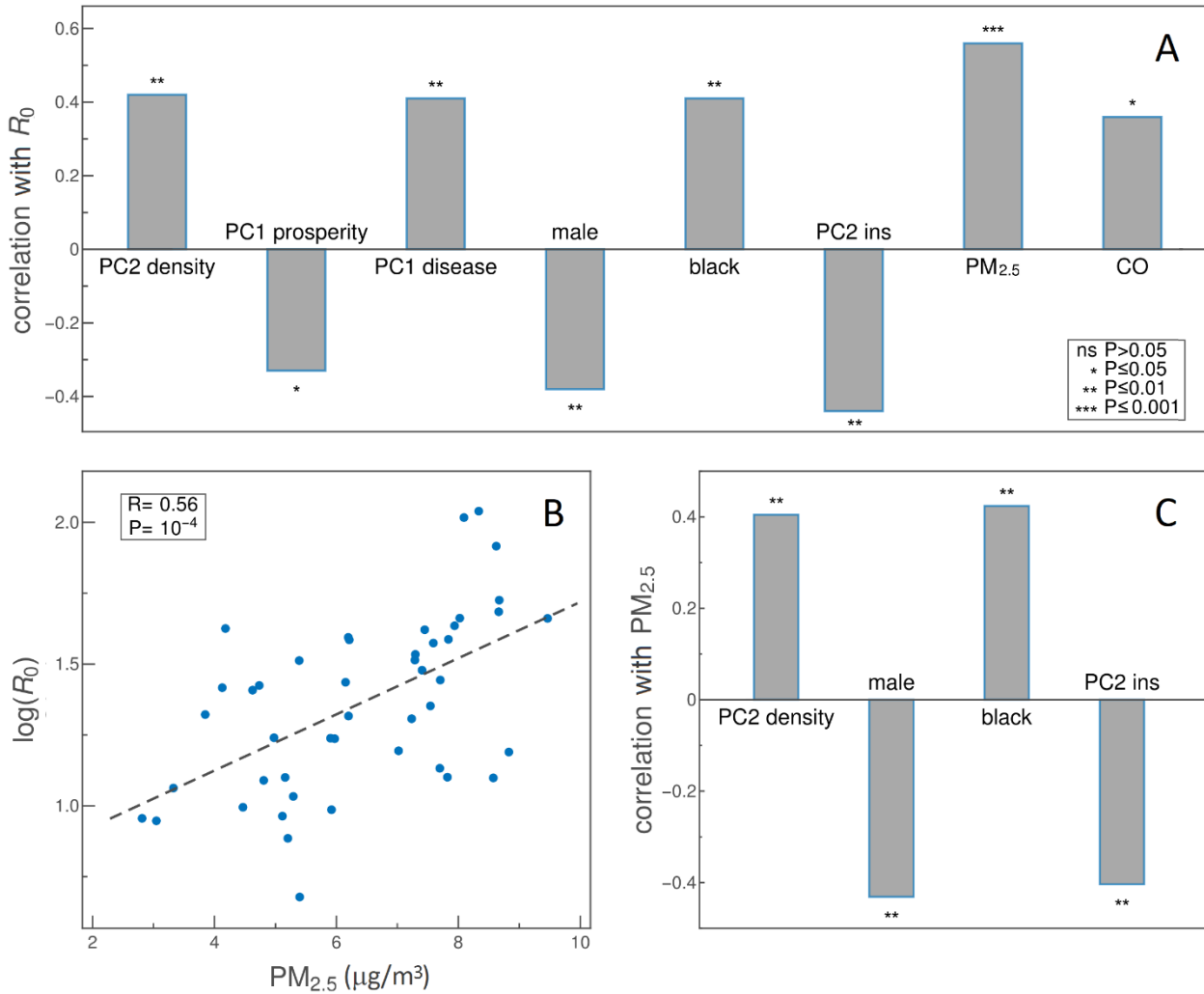


Figure 2. Pearson's correlations for relevant variables. A) Variables significantly correlated ($P < 0.05$) with the basic reproduction number R_0 are shown. The bars' height indicates the value of Pearson's correlation coefficient (R on the y axis). B) Scatter plot of R_0 vs. PM_{2.5}. The dashed line shows linear fit. C) Pearson's correlations of variables in A) with PM_{2.5}. Variable names are indicated on the horizontal axis. Stars in bar plots represent the level of statistical significance, as indicated in the figure legend.

We start from the basic assessment of the variable importance in explaining R_0 , which are pairwise correlations. Note that these do not control for the presence of other potentially important variables but are a straightforward initial assessment of the relation with R_0 . In Figure 2A, we show the Pearson correlation constant of the variables with R_0 , where predictors with statistically significant correlations ($P < 0.05$) are shown together with their correlation constants (represented by bars' heights) and statistical significance levels (indicated by stars). Somewhat surprisingly, we find that the highest correlation is with PM_{2.5}, with $R \sim 0.6$ and $P \sim 10^{-4}$. A large positive correlation between R_0 and PM_{2.5} levels can also be observed from the scatter plot in Figure 2B. Additionally, several other variables exhibit statistically significant correlations with R_0 , as indicated in Figure 2A. Note, however, that

some of these variables are also significantly correlated with $PM_{2.5}$. Moreover, their correlation with R_0 and $PM_{2.5}$ is in the same direction (Figure 2C). Consequently, their significant correlation with R_0 might, at least in part, be due to their correlation with $PM_{2.5}$.

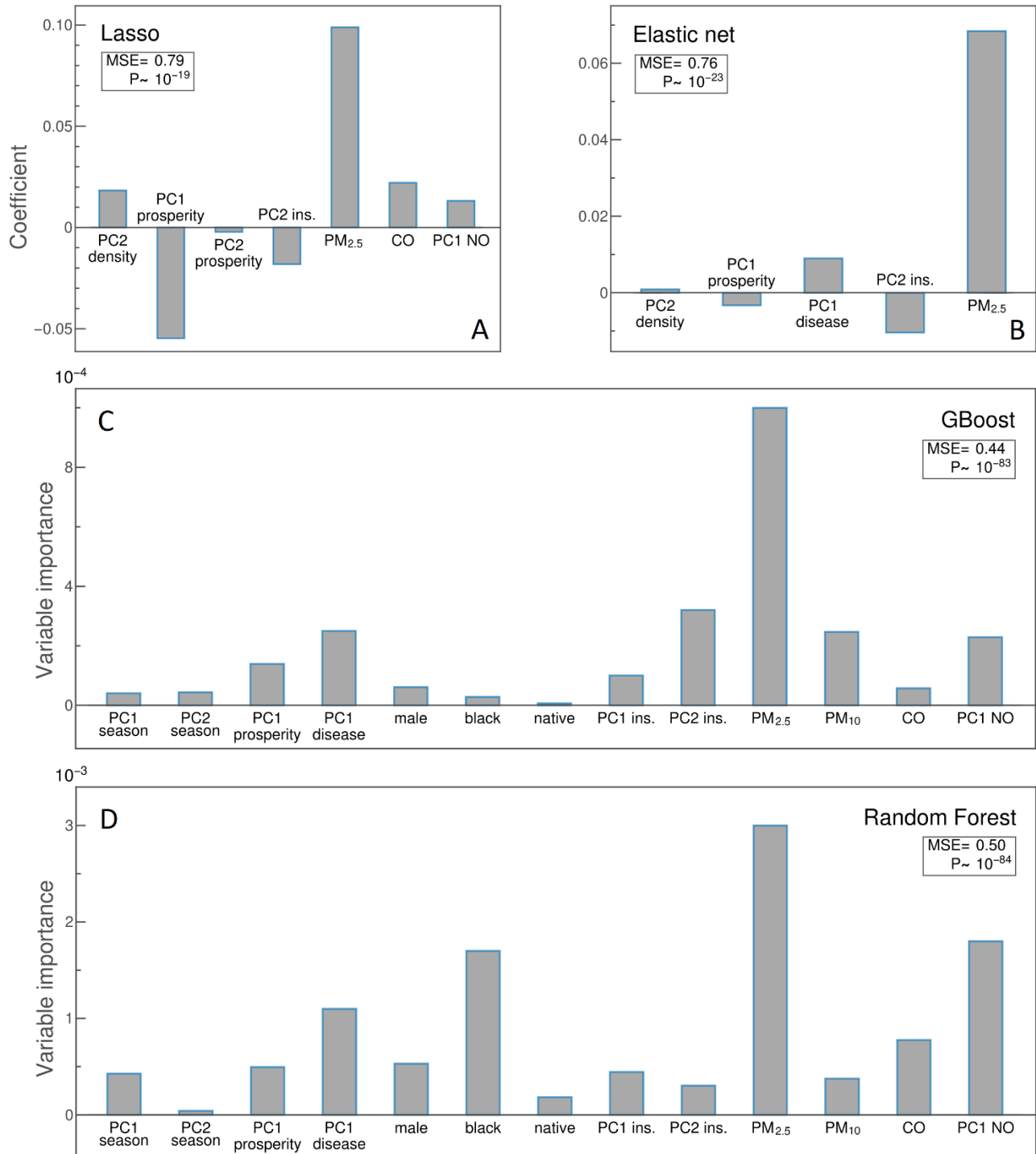


Figure 3: Values of regression coefficients in A) Lasso and B) Elastic Net regressions, respectively, where the bars' height corresponds to the coefficients' values for selected variables. Coefficients of all other variables are shrunk to zero (not shown) by the regressions. Variable importance in C) Gradient Boosting (GBoost) and D) Random Forest (RF) regressions, with the bars' height corresponding to estimated importance. Only variables with $P < 0.1$ (according to either Pearson, Kendall, or Spearman correlations with R_0) are included in GBoost and RF regressions. MSE values are scaled to the constant value model and averaged over 200 cross-validations. P -values correspond to the statistical significance of obtained MSE's compared to the baseline model. Variable names are indicated on the horizontal axis.

To partially address this, we proceed to an analysis that allows selecting the most important predictors from the set of correlated variables. Specifically, results of Lasso and Elastic net regressions are shown in Figures 3A and 3B. Both of these methods provide both regularization and the ability to select significant predictors through shrinking other coefficients to zero. Moreover, we standardize all the variables before using them in regressions, so that the absolute values of the regression coefficients provide relative importance of the selected variables. For each of the two methods, we perform repeated cross-validations, together with optimizations of hyperparameters, so that methods have maximal predictive power (i.e., minimal MSE) on the training set (see Methods for details). We obtain that the two methods are statistically highly significant compared to the constant model ($P \sim 10^{-19}$ and 10^{-23} , for Lasso and Elastic net, respectively). The predictive power of these methods is, however, only moderate, as can be seen for the obtained MSE values (MSEs are scaled, so that MSE of 1 corresponds to the constant model, which is not a large difference from 0.79 and 0.76, obtained by Lasso and Elastic net, respectively). Note, however, that the main purpose of these models is in feature selection, while predictability will be improved through models employed in the next subsection.

From both Lasso and Elastic net, we again obtain that $PM_{2.5}$ is the most important predictor, positively affecting COVID-19 transmissibility (so that higher $PM_{2.5}$ leads to higher transmissibility). A similar trend is obtained for CO and PC1 NO (formed from NO₂ and Nitrogen-oxides concentrations) – CO was also found to be significantly related with R_0 through pairwise correlations. Additionally, the population density (PC2 density) appears as an important predictor through both Lasso and Elastic net, though with smaller importance (regression coefficient), but consistently with pairwise correlations, and with a tendency to increase transmissibility. Also, through all three approaches employed so far (pairwise correlations, Lasso, and Elastic net), we obtain that the higher state prosperity (PC1 prosperity) negatively influences R_0 . Also, chronic diseases significantly influence (increase) R_0 as obtained by both pairwise correlations and Elastic net. Finally, PC2 ins., which is related to the fraction of the population (in particular Latinos) with medical insurance, also negatively correlates with R_0 (through all three methods). Interpretation of these dependencies is further addressed in the Discussion section.

3.3 Variable importance estimates

Our next goal is to assess variable importance and achieve better model predictability through methods that are considered state-of-the-art in machine learning for these types of problems. We will employ two methods based on ensembles of weak learners (decision trees), in particular Gradient Boost and Random Forest. They are substantially different from Lasso and Elastic net employed in the previous subsection, as they do not assume linear dependence of the response from input variables (so-called non-parametric models). Consequently, their employment provides an independent check for the importance of $PM_{2.5}$ in explaining R_0 . Moreover, we expect to obtain better predictability of these models, which can, in turn, be used for a quantitative estimate of pollution variation effects on R_0 .

The two methods are implemented similarly to Lasso and Elastic net, i.e., model hyperparameters are optimized to achieve maximal predictability through repeated cross-validations (see Method for details). As these models (i.e., decision trees in general) are prone to overfitting, we performed a simple variable selection. That is, only variables with $P < 0.1$ (according to either Pearson, Kendall, or Spearman correlations) are selected, resulting in 13 variables shown on the horizontal axes of Figures 3C and 3D, which were then used in further analysis. We obtained a much better predictive power for both Gradient Boost and Random Forest models (compared to regressions in the previous subsection) with MSE of 0.44 and 0.5, respectively, where these differences compared to the constant model (MSE=1) are statistically highly significant ($P \sim 10^{-83}$ and 10^{-84} , respectively).

Estimates of variable importance for both of these models are shown in Figures 3C and 3D. In both figures, the most prominent feature is $PM_{2.5}$, consistently with all other results obtained so far. Furthermore, PC1 disease and PC1 NO appear with moderate importance in both methods, where GBoost also emphasizes the importance of PC2 ins., which is all generally consistent with the analysis presented in the previous subsection. With respect to the pollution, the only difference is that PM_{10} appears as moderately important in GBoost, while not selected by other models. Also, CO is selected by Random Forest as moderately important (consistent with the previous analysis) but does not appear as such in GBoost. Finally, the racial factor (in particular, fraction of black population) is selected as important by Random Forest (and also appeared as significant through pairwise correlations) but does not appear as important in GBoost. A possible interpretation of these findings is addressed in the Discussion section.

3.4 Quantitative estimate of pollution influence on R_0

As we obtain a reasonable model accuracy through both GBoost and Random Forest, we will next estimate how pollution variations (observed through different USA states) affect R_0 . While we included a substantial number of variables (all that we could realistically assemble) in our analysis, these are of course not all the variables that can affect R_0 , so we here aim to provide rough estimates. Still, such an estimate is useful, as it provides the magnitude by which reasonably realistic changes in the pollution levels can affect R_0 . For example, the new SARS-CoV-2 strain that emerged in Great Britain (known as B.1.1.7), which is considered to become dominant in many other parts of the world, is estimated to lead to up to 1.9 increase in R_0 – this value can e.g. be compared with our estimated change due to pollution variations. To generate predictions for each of the analyzed states, we keep all other parameters fixed while changing the pollution values so that the changes correspond to the actual values observed in all 46 states. In this way, the relative change in R_0 , due to observed variations in pollution ($\Delta R_0/R_0$) was estimated, where ΔR_0 corresponds to the difference between maximal and minimal estimated R_0 values.

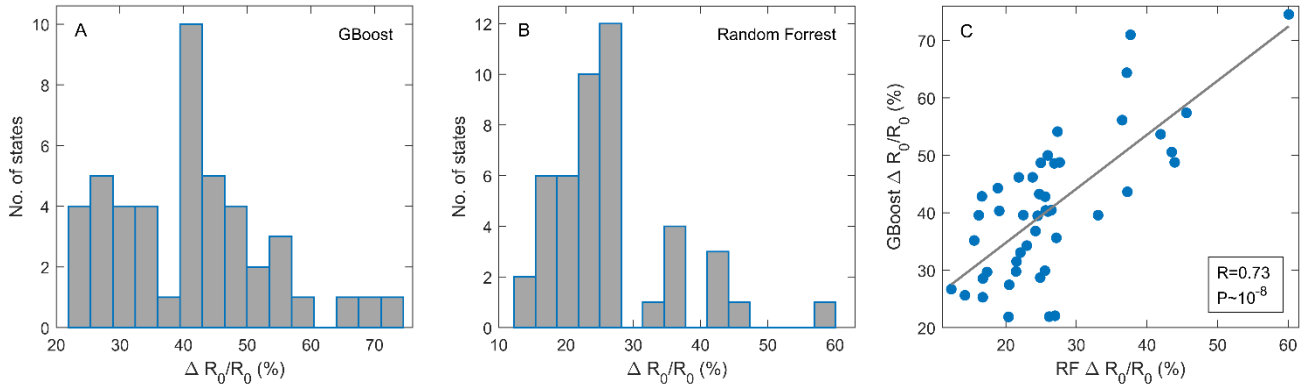


Figure 4: Relative change in R_0 due to pollution variations observed in USA states. For each state included in the analysis, R_0 was predicted for the range of pollution values observed throughout all other states. Relative variation in R_0 was estimated through both A) Gradient Boost (GBoost) and B) Random Forest (RF) regressions, with the models trained as in Fig. 3C) Scatter plot of $\Delta R_0/R_0$ predictions for GBoost and RF, with Pearson’s correlation coefficient and P -value indicated.

The obtained results for $\Delta R_0/R_0$ for all analyzed states are shown as histograms in Figure 4A (GBoost) and 4B (Random Forest). For GBoost, a somewhat larger $\Delta R_0/R_0$, corresponding to the median of $\sim 40\%$, is obtained, while for Random Forest, smaller values with a median of $\sim 25\%$ are estimated. This can e.g. be compared with $\Delta R_0/R_0$ of up to 90% for the B.1.1.7 strain (Davies et al., 2021), so that estimated changes due to pollution variation are smaller but still substantial. Finally, as the two

histograms are somewhat different, in Figure 4C we directly test the consistency of their $\Delta R_0/R_0$ predictions. It can be seen that they are well consistent, with reasonably high correlation ($R=0.73$ and $P\sim 10^{-8}$). Note that these two methods are independent and substantially different (though both based on ensembles of decision trees), so differences in their predictions are expected.

4 Discussion

Figures 2 and 3 reveal the main paper conclusion: $PM_{2.5}$ pollution is the main driver behind SARS-CoV-2 transmissibility in the US. This result is unambiguously obtained through both pairwise correlations of variables with R_0 and by the applied machine learning approaches. While the association of the $PM_{2.5}$ pollution with the rate of COVID19 spread is not a novel result (Gujral and Sinha, 2021; Gupta and Gharehgozli, 2020; Maleki et al., 2021; Stieb et al., 2020), this research is distinct from the existing studies: First, through its large robustness, where we reached the same conclusion by diverse approaches, as illustrated in Figure 3. Secondly, by explicitly taking into account the infection dynamics, i.e., the model-based estimate of R_0 as SARS-CoV-2 transmissibility measure, ensuring that the result does not depend on testing and/or the implemented social distancing measures. This study is also distinct due to the sheer number of 74 predictors initially considered, where PCA and different statistical learning approaches were used to analyze such a large feature set in a meaningful way that allows straightforward interpretation. Taken together, we believe that the results presented here establish the link of $PM_{2.5}$ pollution with COVID19 transmissibility much more firmly. Previous studies on the USA obtained non-consistent reports on pollution importance, underlying importance of more extensive modeling and statistical learning approaches that we employed here (Allen et al., 2021; Gupta and Gharehgozli, 2020; Luo et al., 2021).

A quantitative implication of the established connection is illustrated in Figure 4. We estimated that varying the pollutant levels (specifically, levels of $PM_{2.5}$, PM_{10} , CO and NO_2 , which enter Random Forest and Gradient Boost methods), where changes in $PM_{2.5}$ levels are by far the most important, they make a difference of $\sim 30\%$ in terms of the R_0 values. While this is smaller compared to reproduction number changes due to the appearance of new highly infective strains (that are up to $\sim 90\%$ higher) (Davies et al., 2021), it is still sizable, and clearly illustrates the importance of $PM_{2.5}$ in modulating the virus transmissibility. For example, in an exponential regime of infection progression (c.f. Eq. (3.8) in Methods) lasting for ~ 10 days (a typical period in which exponential growth is observed for USA states), and with typical parameter values, leads to two times larger number of infected, and (at least) equal proportion of lost human lives. An additional (and independent) effect of larger pollutant levels, is in potentially increased COVID-19 mortality due to higher pollutant levels, as suggested by several studies (Luo et al., 2021; Pozzer et al., 2020; X. Wu et al., 2020). Overall, this underscores the importance of reducing pollutant levels in the epidemiological context.

While we obtain that $PM_{2.5}$ pollution is the dominant predictor of virus transmissibility, our results also suggest the relevance of other factors. First, few other pollutants are also selected through our analysis, most notably NO_2 and its related nitrogen oxide derivatives (where its particularly high importance was assigned by the Random Forest method, see Fig. 3D), and to some extent CO and PM_{10} . These results are partially in line with findings that a number of pollutants, more precisely particulate matter (Comunian et al., 2020; Sagawa et al., 2021), but also NO_2 (Paital and Agrawal, 2020), cause overexpression of ACE-2 in respiratory cells, thus increasing the likelihood of infection. This is not necessarily the only relevant mechanism, as the prolonged exposure to pollutants can cause a general weakening of the immune system (Glencross et al., 2020; Qu et al., 2020). However, the relatively low importance of NO and CO pollutants that we obtained speaks in favor of the hypothesis that PM pollution, by binding to virus droplets, additionally mechanically facilitates SARS-CoV-2 spread through the air - both extending the range of virus diffusion and allowing its direct transport into deeper

pulmonary regions (Qu et al., 2020). On the other hand, the inferred large difference in the influence of $PM_{2.5}$ and PM_{10} particles may be understood through the difficulty of particulate matter larger than $5 \mu m$ to reach ACE2 receptors located in type II alveolar cells (Bontempi, 2020; Copat et al., 2020; Zhu et al., 2020).

Another factor (unsurprisingly) related to the susceptibility of an organism to infections, is the presence of different comorbidities and, in general, any diseases that could potentially compromise the immune system (Allel et al., 2020; Coccia, 2020; Liu et al., 2020). Indeed, all applied analysis methods except for Lasso, find the prevalence of chronic diseases in the population (i.e., its dominant principal component PC1-disease) to be an important R_0 predictor.

Additionally, the applied methods also identify a group of three mutually interrelated factors: the dominant PC reflecting the overall prosperity of the state (PC1 prosperity), the percentage of the black population, and the PC2 insurance component (this component effectively reflects the insurance coverage among the Hispanic population). Our recent study of the effects of various demographic and weather parameters on the spread of COVID19 based on the data from 118 world countries (Marko Djordjevic et al., 2021) also pointed to the essential role of the country's prosperity, but we note a stark disagreement in the sign of the correlation: whereas, worldwide, the more developed countries suffered from higher COVID19 expansion rates, data on US states show an opposite trend - wealthier and more developed areas of US on average seem to exhibit lower R_0 values (Gupta and Gharehgozli, 2020). However, this difference may be expected: on the global level, there are substantial variations in the development level between countries, where this level effectively becomes a proxy for the frequency of social contacts (reflecting business and cultural activity, population mixing due to work/education, international travel, etc.) (Gangemi et al., 2020). On the other hand, all states of America have highly developed societies and the dominant effect of these more subtle differences is likely different: within this prosperity range, the better off population has more means to prioritize and practice precautionary behavior (e.g. have professions that require less physical contacts, fewer comorbidities, healthier lifestyle, higher awareness of the infection risks, etc.). But a COVID19 pandemic has also emphasized a specific racial aspect of healthcare disparities. The correlation between the percentage of the black population and R_0 observed in our data (Figure 2A), as well as the results of the Random Forrest regression method (Figure 3D), agree with the already established conclusion that the black minority is by far overrepresented not only among COVID19 fatalities (Luo et al., 2021; Xiao Wu et al., 2020) but also among the total infected population (Chakraborty, 2021; Stieb et al., 2020). Another relevant factor is the health insurance coverage (PC2 insurance), which consistently through our analysis shows that COVID19 infection is spreading faster among people without medical insurance (Figure 2 and Figure 3). Both the percentage of the black population and the prevalence of insurance coverage are significantly correlated with pollution, in particular with $PM_{2.5}$, as can be seen in Figure 2C (curiously, our data do not show such correlation with the PC1 prosperity component). It has been argued that the influence of factors related to a more economically disadvantaged population (overrepresentation of minorities, absence of medical insurance,...) is inherently hard to disentangle from the effects of pollution. While this standpoint is also in part supported by our analysis, we also note that $PM_{2.5}$ consistently appeared with much larger importance through all analyses compared to these economically disadvantaged factors (Chakraborty, 2021; Stieb et al., 2020). Therefore, the most plausible interpretation is to associate $PM_{2.5}$ (rather than these other factors) with R_0 changes.

It is also interesting to consider which parameters did not show up as important in our results. The absence of seasonal principal components from the final sets of significant predictors may imply irrelevance of the weather parameters such as temperature, UV radiation, and humidity on the SarS-CoV2 transmission - in spite that it has become almost a common knowledge that high temperatures should suppress the virus transmission (Byun et al., 2021; Fu et al., 2021; Sarkodie and Owusu, 2020).

Thus, the results presented here side with the authors who disagree that weather factors bear a significant influence on the course of the COVID19 epidemic (Wang et al., 2021). One should however note that variations of meteorological factors are much larger on a global scale, where indeed we find out a larger significance of these factors (Salom et al., 2021). Another somewhat surprising conclusion is the moderate significance of the population density. While there is a significant correlation of PC2 density component with R_0 , it further appeared significant only in Lasso regression, and even there with not that high coefficient. This is however in line with several other studies, that also didn't assign a high significance to population density (Carozzi et al., 2020; Hamidi et al., 2020; Pourghasemi et al., 2020; Rashed et al., 2020).

5 Conclusion and outlook

Starting from 74 initial parameters and by using five different analysis approaches, we obtained the results that robustly select $PM_{2.5}$ pollution as the most important predictor of SARS-CoV-2 transmissibility in the USA. Using R_0 as a transmissibility measure and non-linear dynamics to extract its values for different USA states, these results are largely insensitive to the differences in the state policies. The obtained large quantitative estimate of the magnitude of the $PM_{2.5}$ effect on virus transmissibility may be intuitively unexpected and is not that far from estimated differences in transmissibility caused by virus mutations.

The main issue to be addressed in future studies is that of causality, i.e. disentangling the effects of pollution from those of socio-demographic factors with which it is correlated. This clearly cannot be achieved through studies with a larger resolution, such as the one employed here, even with applications of sophisticated statistical (machine) learning methods, and with taking into account the infection progression dynamics. Carefully crafted, and high resolution, longitudinal epidemiological studies may be a way forward in this regard. The results obtained here, and by other similar studies, may provide a basis for these high-resolution studies, particularly in terms of factors that should be considered, their expected relative importance, and the magnitude of the effects that may be expected.

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

MarD, IS and MagD conceived the research. The work was supervised by MarD, IS and AR. Data acquisition and supplementary material by OM, MT and DZ. Code writing and data analysis by OM, MarD, DZ, SM. Figures and tables made by OM, DZ, MT and MagD. Literature search by AR and SM. Result interpretation by MarD, MagD, IS, AR and SM. Manuscript written by MarD, IS, AR, MagD, OM and MT.

8 Data Availability Statement

Data is provided in provided in the Supplementary material.

9 Funding

This work was partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

10 References

- Adhikari, A., Yin, J., 2020. Short-Term Effects of Ambient Ozone, PM(2.5,) and Meteorological Factors on COVID-19 Confirmed Cases and Deaths in Queens, New York. *Int J Environ Res Public Health* 17, 4047. <https://doi.org/10.3390/ijerph17114047>
- Allel, K., Tapia-Muñoz, T., Morris, W., 2020. Country-level factors associated with the early spread of COVID-19 cases at 5, 10 and 15 days since the onset. *medRxiv* 15, 1589–1602. <https://doi.org/10.1080/17441692.2020.1814835>
- Allen, O., Brown, A., Wang, E., 2021. Socioeconomic Disparities in the Effects of Pollution on Spread of Covid-19: Evidence from US Counties. *medRxiv*. <https://doi.org/10.1101/2021.01.06.21249303>
- America's Health Ranking, 2021. America's Health Rankings analysis of CDC, Behavioral Risk Factor Surveillance System, United Health Foundation. America's Health Rankings. URL <https://www.americashealthrankings.org/> (accessed 3.28.21).
- Bontempi, E., 2020. First data analysis about possible COVID-19 virus airborne diffusion due to air particulate matter (PM): The case of Lombardy (Italy). *Environmental Research* 186, 109639. <https://doi.org/10.1016/j.envres.2020.109639>
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., 1996. Bagging predictors. *Machine Learning* 24, 123–140. <https://doi.org/10.1007/BF00058655>
- Byun, W.S., Heo, S.W., Jo, G., Kim, J.W., Kim, S., Lee, S., Park, H.E., Baek, J.-H., 2021. Is coronavirus disease (COVID-19) seasonal? A critical analysis of empirical and epidemiological studies at global and local scales. *Environmental Research* 196, 110972. <https://doi.org/10.1016/j.envres.2021.110972>
- Carozzi, F., Provenzano, S., Roth, S., 2020. Urban Density and COVID-19 (IZA Discussion Paper No. 13440). Institute of Labor Economics (IZA), Bonn.
- CDC, 2019. CDC - Behavioral Risk Factor Surveillance System. URL <https://www.cdc.gov/brfss/index.html> (accessed 3.28.21).
- Chakraborty, J., 2021. Convergence of COVID-19 and chronic air pollution risks: Racial/ethnic and socioeconomic inequities in the U.S. *Environmental Research* 193, 110586. <https://doi.org/10.1016/j.envres.2020.110586>
- Chen, P.-S., Tsai, F.T., Lin, C.K., Yang, C.-Y., Chan, C.-C., Young, C.-Y., Lee, C.-H., 2010. Ambient influenza and avian influenza virus during dust storm days and background days. *Environ Health Perspect* 118, 1211–1216. <https://doi.org/10.1289/ehp.0901782>
- Coccia, M., 2020. An index to quantify environmental risk of exposure to future epidemics of the COVID-19 and similar viral agents: Theory and practice. *Environ Res* 191, 110155–110155. <https://doi.org/10.1016/j.envres.2020.110155>
- Comunian, S., Dongo, D., Milani, C., Palestini, P., 2020. Air Pollution and Covid-19: The Role of Particulate Matter in the Spread and Increase of Covid-19's Morbidity and Mortality. *Int J Environ Res Public Health* 17, 4487. <https://doi.org/10.3390/ijerph17124487>
- Contini, D., Costabile, F., 2020. Does Air Pollution Influence COVID-19 Outbreaks? *Atmosphere* 11. <https://doi.org/10.3390/atmos11040377>
- Copat, C., Cristaldi, A., Fiore, M., Grasso, A., Zuccarello, P., Signorelli, S.S., Conti, G.O., Ferrante, M., 2020. The role of air pollution (PM and NO₂) in COVID-19 spread and lethality: A systematic review. *Environmental Research* 191, 110129. <https://doi.org/10.1016/j.envres.2020.110129>
- Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D., Wenseleers, T., Gimma, A., Waites, W., Wong,

- K.L.M., van Zandvoort, K., Silverman, J.D., Diaz-Ordaz, K., Keogh, R., Eggo, R.M., Funk, S., Jit, M., Atkins, K.E., Edmunds, W.J., 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372. <https://doi.org/10.1126/science.abg3055>
- De Angelis, E., Renzetti, S., Volta, M., Donato, F., Calza, S., Placidi, D., Lucchini, R.G., Rota, M., 2021. COVID-19 incidence and mortality in Lombardy, Italy: An ecological study on the role of air pollution, meteorological factors, demographic and socioeconomic variables. *Environmental Research* 195, 110777. <https://doi.org/10.1016/j.envres.2021.110777>
- Djordjevic, Magdalena, Djordjevic, Marko, Ilic, B., Stojku, S., Salom, I., 2021. Understanding Infection Progression under Strong Control Measures through Universal COVID-19 Growth Signatures. *Global Challenges* 2000101. <https://doi.org/10.1002/gch2.202000101>
- Djordjevic, Marko, Salom, I., Markovic, S., Rodic, A., Milicevic, O., Djordjevic, Magdalena, 2021. Inferring the main drivers of SARS-CoV-2 transmissibility, arXiv:2103.15123.
- Domingo, J.L., Rovira, J., 2020. Effects of air pollutants on the transmission and severity of respiratory viral infections. *Environmental Research* 187, 109650. <https://doi.org/10.1016/j.envres.2020.109650>
- Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences* 55, 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29, 1189–1232.
- Friedman, J.H., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software, Articles* 33, 1–22. <https://doi.org/10.18637/jss.v033.i01>
- Fu, S., Wang, B., Zhou, J., Xu, X., Liu, J., Ma, Y., Li, L., He, X., Li, S., Niu, J., Luo, B., Zhang, K., 2021. Meteorological factors, governmental responses and COVID-19: Evidence from four European countries. *Environmental Research* 194, 110596. <https://doi.org/10.1016/j.envres.2020.110596>
- Gangemi, S., Billeci, L., Tonacci, A., 2020. Rich at risk: socio-economic drivers of COVID-19 pandemic spread. *Clinical and Molecular Allergy* 18, 12. <https://doi.org/10.1186/s12948-020-00127-4>
- Glencross, D.A., Ho, T.-R., Camiña, N., Hawrylowicz, C.M., Pfeffer, P.E., 2020. Air pollution and its effects on the immune system. *Free Radical Biology and Medicine* 151, 56–68. <https://doi.org/10.1016/j.freeradbiomed.2020.01.179>
- Gujral, H., Sinha, A., 2021. Association between exposure to airborne pollutants and COVID-19 in Los Angeles, United States with ensemble-based dynamic emission model. *Environmental Research* 194, 110704. <https://doi.org/10.1016/j.envres.2020.110704>
- Gupta, A., Gharehgozli, A., 2020. Developing a Machine Learning Framework to Determine the Spread of COVID-19. <http://dx.doi.org/10.2139/ssrn.3635211>
- Hamidi, S., Sabouri, S., Ewing, R., 2020. Does Density Aggravate the COVID-19 Pandemic? *Journal of Urban Health* 86, 495–509. <https://doi.org/10.1080/01944363.2020.1777891>
- Hastie, T., Tibshirani, R., Friedman, J., 2008. *The Elements of Statistical Learning*, 2nd edition. ed. Springer, New York.
- Keeling, M.J., Rohani, P., 2011. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, NJ.
- Krstajic, D., Buturovic, L.J., Leahy, D.E., Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminform* 6, 10–10. <https://doi.org/10.1186/1758-2946-6-10>

- Liu, H., Chen, S., Liu, M., Nie, H., Lu, H., 2020. Comorbid Chronic Diseases are Strongly Correlated with Disease Severity among COVID-19 Patients: A Systematic Review and Meta-Analysis. *Aging Dis* 11, 668–678. <https://doi.org/10.14336/AD.2020.0502>
- Lorenzo, J.S.L., Tam, W.W.S., Seow, W.J., 2021. Association between air quality, meteorological factors and COVID-19 infection case numbers. *Environmental Research* 197, 111024. <https://doi.org/10.1016/j.envres.2021.111024>
- Luo, Y., Yan, J., McClure, S., 2021. Distribution of the environmental and socioeconomic risk factors on COVID-19 death rate across continental USA: a spatial nonlinear analysis. *Environ Sci Pollut Res Int* 28, 6587–6599. <https://doi.org/10.1007/s11356-020-10962-2>
- Maier, B.F., Brockmann, D., 2020. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* 368, 742–746. <https://doi.org/10.1126/science.abb4557>
- Maleki, M., Anvari, E., Hopke, P.K., Noorimotlagh, Z., Mirzaee, S.A., 2021. An updated systematic review on the association between atmospheric particulate matter pollution and prevalence of SARS-CoV-2. *Environmental Research* 195, 110898. <https://doi.org/10.1016/j.envres.2021.110898>
- Martcheva, M., 2015. *An Introduction to Mathematical Epidemiology*. Springer, Boston, MA.
- Maslov, S., Goldenfeld, N., 2020. Window of Opportunity for Mitigation to Prevent Overflow of ICU capacity in Chicago by COVID-19. <https://doi.org/10.1101/2020.03.20.20040048>
- Measure of America, 2018. *Mapping America: Demographic Indicators*. URL <http://measureofamerica.org/tools-old/> (accessed 3.28.21).
- NASA Langley Research Center, 2020. *The Prediction of Worldwide Energy Resources (POWER) Project*. URL <https://power.larc.nasa.gov/> (accessed 3.28.21).
- OpenUV, 2020. *Global UV Index API*. OpenUV. URL <https://www.openuv.io/> (accessed 3.28.21).
- Paital, B., Agrawal, P.K., 2020. Air pollution by NO₂ and PM_{2.5} explains COVID-19 infection severity by overexpression of angiotensin-converting enzyme 2 in respiratory cells: a review. *Environ Chem Lett* 1–18. <https://doi.org/10.1007/s10311-020-01091-w>
- Perkins, T.A., España, G., 2020. Optimal Control of the COVID-19 Pandemic with Non-pharmaceutical Interventions. *Bulletin of Mathematical Biology* 82, 118. <https://doi.org/10.1007/s11538-020-00795-y>
- Pourghasemi, H.R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S.R., Babaei, S., Khosravi, R., Etemadi, M., Ghanbarian, G., Farhadi, A., Safaeian, R., Heidari, Z., Tarazkar, M.H., Tiefenbacher, J.P., Azmi, A., Sadeghian, F., 2020. Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *International Journal of Infectious Diseases* 98. <https://doi.org/10.1016/j.ijid.2020.06.058>
- Pozzer, A., Dominici, F., Haines, A., Witt, C., Münzel, T., Lelieveld, J., 2020. Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovascular Research* 116, 2247–2253. <https://doi.org/10.1093/cvr/cvaa288>
- Qu, G., Li, X., Hu, L., Jiang, G., 2020. An Imperative Need for Research on the Role of Environmental Factors in Transmission of Novel Coronavirus (COVID-19). *Environ Sci Technol* 54, 3730–3732. <https://doi.org/10.1021/acs.est.0c01102>
- Rashed, E.A., Kodera, S., Gomez-Tames, J., Hirata, A., 2020. Influence of Absolute Humidity, Temperature and Population Density on COVID-19 Spread and Decay Durations: Multi-Prefecture Study in Japan. *International Journal of Environmental Research and Public Health* 17. <https://doi.org/10.3390/ijerph17155354>
- Sagawa, T., Tsujikawa, T., Honda, A., Miyasaka, N., Tanaka, M., Kida, T., Hasegawa, K., Okuda, T., Kawahito, Y., Takano, H., 2021. Exposure to particulate matter upregulates ACE2 and


- TMPRSS2 expression in the murine lung. *Environmental Research* 195, 110722. <https://doi.org/10.1016/j.envres.2021.110722>
- Salom, I., Rodic, A., Milicevic, O., Zigic, D., Djordjevic, Magdalena, Djordjevic, Marko, 2021. Effects of Demographic and Weather Parameters on COVID-19 Basic Reproduction Number. *Frontiers in Ecology and Evolution* 8, 524. <https://doi.org/10.3389/fevo.2020.617841>
- Sarkodie, S.A., Owusu, P.A., 2020. Impact of meteorological factors on COVID-19 pandemic: Evidence from top 20 countries with confirmed cases. *Environmental Research* 191, 110101. <https://doi.org/10.1016/j.envres.2020.110101>
- Smits, J., Permanyer, I., 2019. The Subnational Human Development Database. *Scientific Data* 6, 190038. <https://doi.org/10.1038/sdata.2019.38>
- Stieb, D.M., Evans, G.J., To, T.M., Brook, J.R., Burnett, R.T., 2020. An ecological analysis of long-term exposure to PM_{2.5} and incidence of COVID-19 in Canadian health regions. *Environmental Research* 191, 110052. <https://doi.org/10.1016/j.envres.2020.110052>
- Suhaimi, N.F., Jalaludin, J., Latif, M.T., 2020. Demystifying a Possible Relationship between COVID-19, Air Quality and Meteorological Factors: Evidence from Kuala Lumpur, Malaysia. *Aerosol and Air Quality Research* 20, 1520–1529. <https://doi.org/10.4209/aaqr.2020.05.0218>
- Tello-Leal, E., Macías-Hernández, B.A., 2020. Association of environmental and meteorological factors on the spread of COVID-19 in Victoria, Mexico, and air quality during the lockdown. *Environmental Research* 110442. <https://doi.org/10.1016/j.envres.2020.110442>
- Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M.U.G., Li, B., Cai, J., Xu, B., Yang, Q., Wang, B., Yang, P., Cui, Y., Song, Y., Zheng, P., Wang, Q., Bjornstad, O.N., Yang, R., Grenfell, B.T., Pybus, O.G., Dye, C., 2020. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 368, 638–642. <https://doi.org/10.1126/science.abb6105>
- Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- U.S. Census Bureau, 2020. U.S. Census. URL <https://www.census.gov/en.html> (accessed 3.28.21).
- U.S. Census Bureau, 2019. Nativity in the United States American Community Survey 1-year estimates. URL <https://censusreporter.org/> (accessed 3.28.21).
- U.S. Census Bureau, Population Division, 2019. Annual Estimates of the Resident Population by Single Year of Age and Sex for the United States, States, and Puerto Rico Commonwealth: April 1, 2010, to July 1, 2018. URL <https://www.census.gov/en.html> (accessed 3.28.21).
- US Environmental Protection Agency, 2020. Air Quality System Data. US EPA. URL <https://www.epa.gov/outdoor-air-quality-data> (accessed 3.28.21).
- Wang, Q., Zhao, Yu, Zhang, Yajuan, Qiu, J., Li, J., Yan, N., Li, N., Zhang, J., Tian, D., Sha, X., Jing, J., Yang, C., Wang, K., Xu, R., Zhang, Yuhong, Yang, H., Zhao, S., Zhao, Yi, 2021. Could the ambient higher temperature decrease the transmissibility of COVID-19 in China? *Environmental Research* 193, 110576. <https://doi.org/10.1016/j.envres.2020.110576>
- Weitz, J.S., Beckett, S.J., Coenen, A.R., Demory, D., Dominguez-Mirazo, M., Dushoff, J., Leung, C.-Y., Li, G., Măgălie, A., Park, S.W., Rodriguez-Gonzalez, R., Shivam, S., Zhao, C.Y., 2020. Modeling shield immunity to reduce COVID-19 epidemic spread. *Nature Medicine* 26, 849–854. <https://doi.org/10.1038/s41591-020-0895-3>
- Wikipedia, 2021a. List of states and territories of the United States by population. URL https://en.wikipedia.org/w/index.php?title=List_of_states_and_territories_of_the_United_States_by_population&oldid=1016990633 (accessed 3.28.21).
- Wikipedia, 2021b. List of United States cities by population. URL https://en.wikipedia.org/w/index.php?title=List_of_United_States_cities_by_population&oldid=1017904123 (accessed 3.28.21).

- Wu, X., Nethery, R.C., Sabath, M.B., Braun, D., Dominici, F., 2020. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Science Advances* 6. <https://doi.org/10.1126/sciadv.abd4049>
- Wu, Xiao, Nethery, R.C., Sabath, M.B., Braun, D., Dominici, F., 2020. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. *medRxiv* 2020.04.05.20054502. <https://doi.org/10.1101/2020.04.05.20054502>
- Yao, Y., Pan, J., Liu, Z., Meng, X., Wang, Weidong, Kan, H., Wang, Weibing, 2021. Ambient nitrogen dioxide pollution and spreadability of COVID-19 in Chinese cities. *Ecotoxicology and Environmental Safety* 208, 111421. <https://doi.org/10.1016/j.ecoenv.2020.111421>
- Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. *Science of The Total Environment* 727, 138704. <https://doi.org/10.1016/j.scitotenv.2020.138704>
- Zou, H., Hastie, T., 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67, 301–320.
- Worldometer, 2020. COVID-19 Coronavirus Pandemic [Online]. Available online at: <https://www.worldometers.info/coronavirus/> (accessed January 14, 2021).

PRESENTATIONS

A. Contributed talks at conferences and workshops

1. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *Recent results with full DREENA framework as a multipurpose tool for QGP tomography*, 2019 COST Action CA15213 THOR Annual Meeting, Sep 2019, Istanbul, Turkey



2019 COST Action CA15213 THOR Annual Meeting

2-6 September 2019
Istanbul Technical University
Europe/Istanbul timezone

- Overview
- Timetable
- Contribution List
- Registration
- Participant List

Recent results with full DREENA framework as a multipurpose tool for QGP tomography

Sep 2, 2019, 5:00 PM
40m
Istanbul Technical University

Speaker
Dusan Zigic (Institute of Physics ...)

Presentation materials

Istanbul_DusanZigic.pdf

[Help](#) | [Contact](#) | [Terms and conditions](#) | [URL Shortener](#)

2. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *DREENA framework as a multipurpose tool for QGP tomography*, Zimanyi school 2019: 19th Zimanyi school - Winter workshop on heavy ion physics, Dec 2019, Budapest, Hungary

Zimányi School 2019

Dec 2, 2019, 9:00 AM → Dec 6, 2019, 6:00 PM Europe/Budapest
 Bldg 3, Council room (Wigner Research Center for Physics)

Description

ZIMÁNYI SCHOOL 19

19. ZIMÁNYI SCHOOL

WINTER WORKSHOP ON
HEAVY ION PHYSICS

Dec. 2. - Dec. 6.,
Budapest, Hungary

Györfi András: Az úton (On the road)

József Zimányi (1931 - 2006)

10:55 AM → 12:40 PM **Jets (section chair: M. Strikman)** Bldg 3, Council room

10:55 AM	<p>Latest Results from RHIC and BNL</p> <p>Speaker: Michael Tannenbaum (Brookhaven National Laboratory (US))</p> <p></p>	🕒 30m
11:25 AM	<p>Dynamical jet energy loss</p> <p>Speaker: Magdalena Djordjevic (Institute of Physics Belgrade)</p> <p></p>	🕒 25m
11:50 AM	<p>Beyond the soft-gluon approximation in calculating hard probe radiative energy loss</p> <p>Speaker: Dr Bojana Ilic (Blagojevic) (Institute of Physics Belgrade)</p> <p></p>	🕒 20m
12:10 PM	<p>DREENA framework as a multipurpose tool for QGP tomography</p> <p>Speaker: Dusan Zigic (Institute of Physics Belgrade)</p> <p></p>	🕒 15m
12:25 PM	<p>Shape of the quark gluon plasma droplet reflected in the high pt data</p> <p>Speaker: Stefan Stojku (Institute of Physics University of Belgrade)</p> <p></p>	🕒 15m

3. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *QGP tomography with DREENA framework*, 3rd JETSCAPE Winter School and Workshop 2020, March 2020, Knoxville, Tennessee, USA



3rd JETSCAPE Winter School and Workshop 2020

16-20 March 2020
University of Tennessee Knoxville
US/Eastern timezone

THE SCHOOL HAS BEEN CANCELLED DUE TO CORONA VIRUS RELATED TRAVEL RESTRICTIONS. THE WORKSHOP WILL BE HELD VIRTUALLY.

- Overview
- Coronavirus disruptions
- Remote connection
- Conference photo
- Announcement
- Timetable
- Contribution List**
- Registration
- Participant List
- Travel Information
- Map
- Local Information
- School Material
- Directions and Speaker Instructions
- Organizing Committee

QGP tomography with DREENA framework



20 Mar 2020, 10:10

Workshop Session VIII

15m

University of Tennessee Knoxville

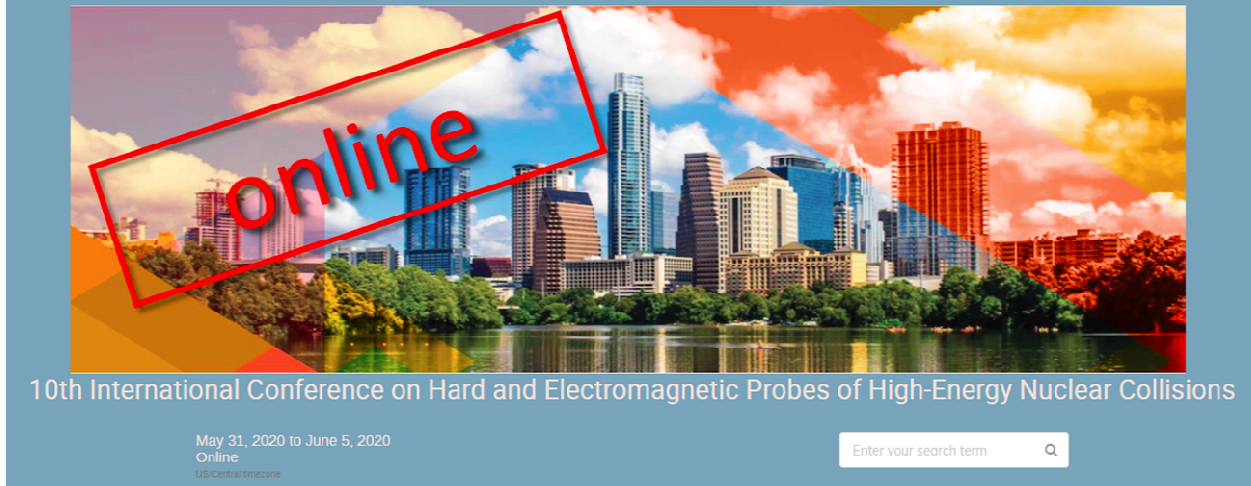
Speaker

Dusan Zigic (Institute of Physics ...)

Presentation Materials

JETSCAPEWS20_DZ.pdf

4. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *DREENA framework as a multipurpose tool for QGP tomography*, Hard Probes 2020: 10th International Conference on Hard & Electromagnetic Probes of High-Energy Nuclear Collisions, June 2020, Austin, TX, USA



- Overview
- Committees
- Scientific Program
- Timetable
- Zoom Links to Speaker Sessions
- Access To Poster Session
- Contribution List (all)
- Proceedings
- Instructions for Presenters & Chairs
- Registration
- Call for Abstracts
- Contribution List
- Book of Abstracts
- Participant List
- Important dates
- Code of Conduct
- Previous Conferences
- Zoom Background
- Contact
- ✉ hp2020@ph.utexas.edu

DREENA framework as a multipurpose tool for QGP tomography

📅 Jun 3, 2020, 11:30 AM
 ⌚ 20m
 📍 <https://cern.zoom.us/j/97442756310>

Oral Presentation Heavy Flavor and Q... Parallel

Speaker

👤 Dusan Zigic (Institute of Physics ...)

Description

DREENA framework is based on our dynamical energy loss formalism, which takes into account finite size, finite temperature QCD medium consisting of dynamical (moving) partons. Both radiative and collisional energy losses are calculated under the same theoretical framework in the dynamical energy loss formalism, which is applicable to both light and heavy flavor observables. We generalized the formalism to the case of finite magnetic mass, running coupling, and towards removing widely used soft-gluon approximation. Importantly, DREENA provides a natural framework where temperature profile from any medium evolution can be straightforwardly implemented. We exploit this by implementing different state-of-the-art medium evolutions (both event-by-event hydrodynamics and kinetic transport theory) within DREENA framework. DREENA does not use free parameters, i.e. its only input is the temperature profile that comes directly from various hydrodynamics and kinetic theory models. This opens possibility to use DREENA on both light and heavy flavor to test and differentiate between different available QGP evolution models, including both large and smaller systems, making DREENA a multipurpose QGP tomography tool. Our results on these tests will be presented, which enables us to gain a better understanding of the bulk QGP medium created at RHIC and LHC. As a highlight, contrary to the existing models, which for full hydro evolution models lead to v_2 puzzle, with DREENA we surprisingly obtain a very good joint agreement between R_{AA} and v_2 data. This well known puzzle therefore appears to be a consequence of a simplified energy loss commonly used by other models – once a proper description of parton medium interactions is used, v_2 puzzle is abolished. While a widely accepted paradigm is that proper medium evolution description dominates in explaining high p_T data, this result strongly suggests that proper description of parton-medium interactions is much more important.

Contribution type Contributed Talk
Track Heavy Flavor and Quarkonia

Primary author

👤 Dusan Zigic (Institute of Physics ...)

Co-authors

- 👤 Dr Igor Salom (Institute of Physics ...)
- 👤 Jussi Auvinen (Institute of Physics ...)
- 👤 Dr Marko Djordjevic (Faculty of Biology, U...)
- 👤 Magdalena Djordjevic (Institute of Physics ...)

📎 Presentation materials

- 📄 HP2020_DREENA_DZ.pdf
- 🔊 Recording

5. Dusan Zigic, Magdalena Djordjevic, Jussi Auvinen, Igor Salom, Marko Djordjevic and Pasi Huovinen, *Using DREENA framework to explore properties of QGP*, Online Strangeness in Quark Matter Conference 2021, May 2021



- Overview
- Scientific program
- ZOOM connection
- Call for Abstracts
- Timetable
- Contribution List**
- Registration
- Participant List
- Important dates
- Announcements
- Committees
- Code of conduct

Using DREENA framework to explore properties of QGP



📅 May 21, 2021, 10:30 AM

Theory talk

Open and New (IV)

🕒 20m

📍 Room C (Zoom)

Speaker

👤 [Dusan Zigic](#) (Institute of Physics ...)

Description

DREENA framework is based on our dynamical energy loss formalism, which takes into account finite size, finite temperature QCD medium consisting of dynamical (moving) partons. Both radiative and collisional energy losses are calculated under the same theoretical framework in the dynamical energy loss formalism, which is applicable to both light and heavy flavor observables. We generalized the formalism to the case of finite magnetic mass, running coupling, and towards removing widely used soft-gluon approximation. Importantly, DREENA provides a natural framework where temperature profile from any medium evolution can be straightforwardly implemented. We exploit this by implementing different state-of-the-art medium evolutions (both smooth and event-by-event hydrodynamics simulations) within DREENA framework. DREENA does not use free parameters, i.e. its only input is the temperature profile that comes directly from various QGP simulation models. This opens possibility to use DREENA on both light and heavy flavor to test and differentiate between different available QGP evolution models, including both large and smaller systems, making DREENA a multipurpose QGP tomography tool. Our results on these tests will be presented, which enables us to gain a better understanding of the bulk QGP medium created at RHIC and LHC. As a highlight, contrary to the existing models, which for full hydro evolution models lead to τ_2 puzzle, with DREENA we surprisingly obtain a very good joint agreement between R_{AA} and τ_2 data. This well known puzzle therefore appears to be a consequence of a simplified energy loss commonly used by other models – once a proper description of parton-medium interactions is used, τ_2 puzzle is abolished. While a widely accepted paradigm is that proper medium evolution description dominates in explaining high p_T data, this result strongly suggests that proper description of parton-medium interactions is much more important.

Primary author

👤 [Dusan Zigic](#) (Institute of Physics ...)

Co-authors

- 👤 [Jussi Auvinen](#) (Institute of Physics ...)
- 👤 [Dr Igor Salom](#) (Institute of physics ...)
- 👤 [Magdalena Djordjevic](#) (Institute of Physics ...)
- 👤 [Marko Djordjevic](#)
- 👤 [Pasi Huovinen](#) (University of Wroclaw ...)

📎 Presentation materials

📄 SQM_May2021_DZ.pdf

B. Contributed posters at conferences and workshops

6. Dusan Zigic, Igor Salom, Jussi Auvinen, Marko Djordjevic, Magdalena Djordjevic, *Numerical predictions of DREENA-C and DREENA-B frameworks*, Hard Probes 2018: International Conference on Hard & Electromagnetic Probes of High-Energy Nuclear Collisions, Oct 2018, Aix-Les-Bains, France

Hard Probes 2018: International Conference on Hard & Electromagnetic Probes of High-Energy Nuclear Collisions

September 30, 2018 to October 5, 2018
Aix-Les-Bains, Savoie, France
Europe/Zurich timezone

PROCEEDINGS OPEN UNTIL DECEMBER 15th 2018

- Overview
- Committees
- Timetable
- CALL FOR ABSTRACTS
- Registration
- Student Lectures Day
- Registration Payment
- Participant List
- PROCEEDINGS
- Photos
- Instructions for Speakers
- Instructions for Poster Presenters
- Venue and Travel Information
- Accommodations
- Info Student Residence Azuréa
- Excursions and Activities
- Social Events
- Previous Conferences
- Conference Poster
- Sponsors

Numerical predictions of DREENA-C and DREENA-B frameworks

Not scheduled 3b) Heavy-flavours and... Poster Session
 20m
 Aix-Les-Bains, Savoie, France

Speaker

Dusan Zigic (Institute of Physics ...)

Description

We here present two frameworks that allow generating a wide range of predictions from the dynamical energy loss formalism [1,2]. In distinction to majority of other methods, the dynamical energy loss formalism takes into account a realistic medium composed of dynamical scattering centers. The first framework (DREENA-C) [1] applies to the medium with constant temperature, while the second framework applies to evolving medium modeled by Bjorken 1+1D expansion (DREENA-B) [2]. We here present joint R_{AA} and v_2 predictions for light and heavy flavor, for different systems, centralities and collision energies. DREENA-C (constant temperature) predictions overestimate v_2 . For DREENA-B (Bjorken expansion) we obtain a good agreement with both R_{AA} and v_2 data. Introducing medium evolution has a larger effect on v_2 , but for precision predictions it also has to be taken into account for R_{AA} . These results argue the dynamical energy loss formalism can provide a basis for a state of the art QGP tomography tool.

[1] D. Zigic, I. Salom, J. Auvinen, M. Djordjevic and M. Djordjevic, arXiv:1805.03494 [nucl-th]
[2] D. Zigic, I. Salom, M. Djordjevic and M. Djordjevic, arXiv:1805.04786 [nucl-th]

Primary author

Dusan Zigic (Institute of Physics ...)


Co-authors


Dr Igor Salom (Institute of Physics ...)
 Jussi Auvinen (Institute of Physics ...)
 Marko Djordjevic
 Magdalena Djordjevic (Institute of Physics ...)


Support

7. Dusan Zigic, Igor Salom, Jussi Auvinen, Magdalena Djordjevic, Marko Djordjevic, *DREENA framework: predictions and comparison with experimental data*, Strangeness in Quark Matter, Jun 2019, Bari, Italy

Strangeness in Quark Matter 2019

 Jun 9, 2019, 4:00 PM → Jun 15, 2019, 1:00 PM Europe/Zurich

 Bari, Italy

 Domenico Elia (INFN Bari)

Description The 18th International Conference on Strangeness in Quark Matter (SQM 2019) will focus on new experimental and theoretical developments on the role of strange and heavy-flavour quarks in high energy heavy-ion collisions and in astrophysical phenomena.

Scientific topics include:

- Strangeness and heavy-quark production in nuclear collisions and hadronic interactions
- Hadron resonances in the strongly-coupled partonic and hadronic medium
- Bulk matter phenomena associated with strange and heavy quarks
- QCD phase structure
- Collectivity in small systems
- Strangeness in astrophysics
- Open questions and new developments


6:45 PM

DREENA framework: predictions and comparison with experimental data

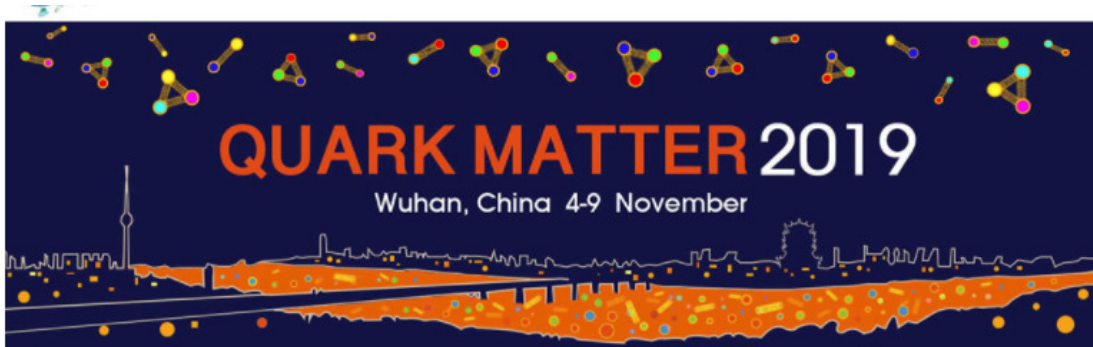
 2h

We will present our newly developed DREENA framework [1], which allows predicting energy loss of high p_{\perp} partons traversing quark gluon plasma (QGP). DREENA framework is based on state-of-the-art dynamical energy loss formalism, which takes into account finite size medium composed of dynamical (that is moving) constituents. In DREENA-B, this formalism is applied to 1+1D Bjorken medium expansion. Joint R_{AA} and v_2 predictions are generated for both light and heavy probes, and for all centrality regions in both $Pb + Pb$ and $Xe + Xe$ collisions at the LHC. DREENA-B framework leads to a good agreement with both R_{AA} and v_2 experimental data, which is a major step towards introducing complex medium evolution into DREENA framework. Examining influence of different initial conditions (stages) of QGP evolution on joint R_{AA} and v_2 predictions with DREENA framework, will also be discussed.

Speaker: Dusan **Zigic** (Institute of Physics Belgrade)

 SQM19_Poster_DZ....

8. Dusan Zigic, Igor Salom, Jussi Auvinen, Marko Djordjevic, Magdalena Djordjevic, *DREENA framework as a multipurpose tool for QGP tomography*, Quark Matter 2019 - the XXVIIIth International Conference on Ultra-relativistic Nucleus-Nucleus Collisions, Nov 2019, Wuhan, China



Quark Matter 2019 - the XXVIIIth International Conference on Ultra-relativistic Nucleus-Nucleus Collisions

3-9 November 2019
Wanda Reign Wuhan Hotel
Asia/Shanghai timezone

Enter your search term

- Overview
- Timetable
- QM2019 Web Home
- Scientific Program
- Participant List
- Committees
- Important Dates
- Conference Venue
- Travel
- Call for Abstracts
- Contribution List**
- Registration
- Local Information
- Visa Information
- Social Events
- Accommodation
- Satellite Meetings
- Conference Sponsors
- Proceedings
- Contact**
- qm2019@mail.ccnu.edu.cn
- qm2019-poster@mail.ccnu.edu.cn
- qm2019_visa@mail.ccnu.edu.cn

DREENA framework as a multipurpose tool for QGP tomography [Link](#) [Print](#) [Calendar](#)

Nov 4, 2019, 5:40 PM

20m

Wanda Han Show Theatre & Wanda Reign Wuhan Hotel

Poster Presentation

Heavy flavor and qu...

Poster Session

Speaker

Dusan Zigic (Institute of Physics ...)

Description

DREENA framework is based on our dynamical energy loss formalism, which takes into account finite size, finite temperature QCD medium consisting of dynamical (moving) partons. Both radiative and collisional energy losses are calculated under the same theoretical framework in the dynamical energy loss formalism, which is applicable to both light and heavy flavor observables. We generalized the formalism to the case of finite magnetic mass, running coupling, and towards removing widely used soft-gluon approximation. Importantly, DREENA provides a natural framework where temperature profile from any medium evolution can be straightforwardly implemented. We exploit this by implementing different state-of-the-art medium evolutions (both event-by-event hydrodynamics and kinetic transport theory) within DREENA framework. DREENA does not use free parameters, i.e. its only input is the temperature profile that comes directly from various hydrodynamic and kinetic theory models. This opens possibility to use DREENA on both light and heavy flavor to test and differentiate between different available QGP evolution models, including both large and smaller systems, making DREENA a multipurpose QGP tomography tool. Our results on these tests will be presented, which enables us to gain a better understanding of the bulk QGP medium created at RHIC and LHC. As a highlight, contrary to the existing models, which for full hydro evolution models lead to τ_2 puzzle, with DREENA we surprisingly obtain a very good joint agreement between R_{AA} and τ_2 data. This well known puzzle therefore appears to be a consequence of a simplified energy loss commonly used by other models once a proper description of parton-medium interactions is used. τ_2 puzzle is abolished. While a widely accepted paradigm is that proper medium evolution description dominates in explaining high p_T data, this result strongly suggests that proper description of parton-medium interactions is much more important.

Primary authors

Dusan Zigic (Institute of Physics ...)

Dr Igor Salom (Institute of Physics ...)

Jussi Auvinen (Institute of Physics ...)

Marko Djordjevic

Magdalena Djordjevic (Institute of Physics ...)



Република Србија
Универзитет у Београду
Физички факултет
Д.Бр.2018/8014
Датум: 06.07.2021. године

На основу члана 161 Закона о општем управном поступку и службене евиденције издаје се

УВЕРЕЊЕ

Жигић (Зоран) Душан, бр. индекса 2018/8014, рођен 19.10.1991. године, Сремска Митровица, Република Србија, уписан школске 2020/2021. године, у статусу: финансирање из буџета; тип студија: докторске академске студије; студијски програм: Физика.

Према Статуту факултета студије трају (број година): три.
Рок за завршетак студија: у двоструком трајању студија.

Ово се уверење може употребити за регулисање војне обавезе, издавање визе, права на дечији додатак, породичне пензије, инвалидског додатка, добијања здравствене књижице, легитимације за повлашћену возњу и стипендије.

Овлашћено лице факултета



M. J. J. J. J.



Република Србија
Универзитет у Београду
Физички факултет
Д.Бр.2018/8014
Датум: 06.07.2021. године

На основу члана 161 Закона о општем управном поступку и службене евиденције издаје се

УВЕРЕЊЕ

Жигић (Зоран) Душан, бр. индекса 2018/8014, рођен 19.10.1991. године, Сремска Митровица, Република Србија, уписан школске 2020/2021. године, у статусу: финансирање из буџета; тип студија: докторске академске студије; студијски програм: Физика.

Према Статуту факултета студије трају (број година): три.
Рок за завршетак студија: у двоструком трајању студија.

Ово се уверење може употребити за регулисање војне обавезе, издавање визе, права на дечији додатак, породичне пензије, инвалидског додатка, добијања здравствене књижице, легитимације за повлашћену возњу и стипендије.

Овлашћено лице факултета



M. J. J. J. J.



Република Србија
Универзитет у Београду
Физички факултет
Број индекса: 2018/8014
Датум: 30.06.2021.

На основу члана 29. Закона о општем управном поступку и службене евиденције издаје се

УВЕРЕЊЕ О ПОЛОЖЕНИМ ИСПИТИМА

Душан Жигић, име једног родитеља Зоран, рођен 19.10.1991. године, Сремска Митровица, Република Србија, уписан школске 2018/2019. године на докторске академске студије, школске 2020/2021. године уписан на статус финансирање из буџета, студијски програм Физика, током студија положио је испите из следећих предмета:

Р.бр.	Шифра	Назив предмета	Оцена	ЕСПБ	Фонд часова**	Датум
1.	ДС15ВО2	Монте Карло симулације у физици	10 (десет)	15	I:(8+0+0)	17.09.2019.
2.	ДС15ФРНД1	Рад на докторату 1. део	П.	30	I:(0+0+12) II:(0+0+12)	
3.	ДС15ПЕ8	Суперсиметрије	10 (десет)	15	I:(8+0+0)	20.09.2019.
4.	ДС15ТНФ	Теоријска нуклеарна физика	10 (десет)	15	III:(8+0+0)	25.09.2020.
5.	ДС15НФ7	Анализа података у физици високих енергија	9 (девет)	15	III:(8+0+0)	30.09.2020.
6.	ДС15ФРНД2	Рад на докторату 2. део	П.	30	III:(0+0+12) IV:(0+0+12)	
7.	ДС15ФРНД3	Рад на докторату 3. део	П.	15	V:(0+0+20)	
8.	ДС15ФРНД4	Рад на докторату 4. део	П.	15	VI:(0+0+20)	

* - еквивалентиран/признат испит.

** - Фонд часова је у формату (предавања+вежбе+остало).

Општи успех: 9,75 (девет и 75/100), по годинама студија (10,00, 9,50, /).



Овлашћено лице факултета

[Handwritten signature]



Универзитет у Београду
Физички факултет
Број индекса: 2012/3137
Број: 2432017
Датум: 26.09.2017.

На основу члана 161 Закона о општем управном поступку ("Службени лист СРЈ", бр. 33/97, 31/2001 и "Службени гласник РС", бр. 30/2010) и службене евиденције, Универзитет у Београду - Физички факултет, издаје

У В Е Р Е Њ Е

Душан Жићих

име једног родитеља Зоран, ЈМБГ 1910991890009, рођен 19.10.1991. године, Сремска Мићровица, оштинина Сремска Мићровица-град, Република Србија, уписан школске 2012/13. године, дана 20.09.2017. године завршио је основне академске студије на студијском програму Примењена и компјутерска физика, у трајању од четири године, обима 242 (двеста четирдесет два) ЕСПБ бодова, са просечном оценом 9,75 (девет и 75/100).

На основу наведеног издаје му се ово уверење о стеченом високом образовању и стручном називу **Дипломирани физичар**.

Декан



Јаблан Дојчиловић
Проф. др Јаблан Дојчиловић



Универзитет у Београду
Физички факултет
Број индекса: 2017/7014
Број: 2292018
Датум: 25.09.2018.

На основу члана 161 Закона о општем управном поступку ("Службени лист СРЈ", бр. 33/97, 31/2001 и "Службени гласник РС", бр. 30/2010) и службене евиденције, Универзитет у Београду - Физички факултет, издаје

У В Е Р Е Њ Е

Душан Жијић

име једног родитеља Зоран, ЈМБГ 1910991890009, рођен 19.10.1991. године, Сремска Мишковица, оштина Сремска Мишковица-град, Република Србија, уписан школске 2017/18. године, дана 19.09.2018. године завршио је мастер академске студије на студијском програму Теоријска и експериментална физика, у трајању од једне године, обима 60 (шездесет) ЕСПБ бодова, са просечном оценом 9,67 (девет и 67/100).

На основу наведеног издаје му се ово уверење о стеченом високом образовању и академском називу **мастер физичар**.



Декан

Prof. dr. Јаблан Дојчиловић
Проф. др Јаблан Дојчиловић



ДОКТОРСКЕ СТУДИЈЕ

ПРЕДЛОГ ТЕМЕ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ
КОЛЕГИЈУМУ ДОКТОРСКИХ СТУДИЈА

Школска година
2020/2021

Подаци о студенту

Име

ДУШАН

Презиме

ЖИГИЋ

Број индекса

8014/2018

Научна област дисертације

ФИЗИКА ВИСОКИХ ЕНЕРГИЈА
И НУКЛЕАРНА ФИЗИКА

Подаци о ментору докторске дисертације

Име

МАГДАЛЕНА

Презиме

БОРЂЕВИЋ

Научна област

ТЕОРИЈСКА НУКЛЕАРНА ФИЗИКА

Звање

НАУЧНИ САВЕТНИК

Институција

ИНСТИТУТ ЗА ФИЗИКУ

Име

ИГОР

Презиме

САХОМ

Научна област

ПРИМЕНА РАЧУНАРА У
МАТЕМАТИЧКОЈ ФИЗИЦИ

Звање

ВИШИ НАУЧНИ САРАДНИК

Институција

ИНСТИТУТ ЗА ФИЗИКУ

Предлог теме докторске дисертације

Наслов

РАЗВОЈ DREENA МОДЕЛА ЗА ТОМОГРАФИЈУ
КВАРК-ГЛУОНСКЕ ПЛАЗМЕ

Уз пријаву теме докторске дисертације Колегијуму докторских студија, потребно је приложити следећа документа:

1. Семинарски рад (дужине до 10 страница)
2. Кратку стручну биографију писану у трећем лицу јединине
3. Фотокопију индекса са докторских студија

Датум

8. 6. 2021.

Потпис ментора

Magdalena Borjević

Др Магдалена Ђорђевић

Потпис ментора

Igor Salom

Др Игор Салом

Потпис студента

Željko Ž.

Мишљење Колегијума докторских студија

Након образложења теме докторске дисертације Колегијум докторских студија је тему

прихватио

није прихватио

Датум

30. 06. 2021.

Продекан за науку Физичког факултета

Stojanović