ARTIFICIAL INTELLIGENCE IN REVEALING AIR POLLUTION-RELATED PROCESSES

Andreja M. Stojić[1,2], Mirjana D. Perišić[1,2], Gordana P. Jovanović[1,2], Svetlana M. Stanišić[2]

[1] Institute of Physics, Pregrevica 118, 11080 Belgrade, Serbia
e-mail: andreja.stojic@ipb.ac.rs
[2] Singidunum University, Danijelova 32, 11000 Belgrade, Serbia
e-mail: mirjana.perisic@ipb.ac.rs, gordana.vukovic@ipb.ac.rs, sstanisic@singidunum.ac.rs

Abstract:

In this study, we present one of the results of the Air Quality Plan for the Agglomeration of the City of Belgrade for the Period 2021-2031 project, supported by the City Administration of Belgrade. We used machine learning regression by means of the eXtreme Gradient Boosting (XGBoost) to interrelate PM_{10} concentrations and a number of environmental parameters in Belgrade, Serbia. The influence of meteorological factors affecting PM_{10} concentrations was investigated and explained using the SHapley Additive exPlanations (SHAP) attribution method. The ambient temperature is one of the most important parameters responsible for altering the PM_{10} concentrations in the range from -4 to 20 μg m^{-3} relative to the average level.

Keywords: air pollution, particulate matter, machine learning, explainable artificial intelligence

1. Introduction

The modern world is facing many environmental issues, with all environmental compartments being affected. According to recent estimates, the mortality rate due to exposure to high levels of air pollution accounts for 8.9 million deaths annually [1]. Containing several hundred types of chemical species, some of which are toxic, mutagenic, and carcinogenic, particulate matter (PM) pollution rises as a significant problem in urban areas. The environmental fate of PM is governed by a diversity of emission sources, meteorological factors, or topographic features, as well as their mutual interrelations.

The understanding of environmental pollution-related processes is yet to be enhanced based on data-driven research. The enormous potential for the enhancement lies in the effective interrogation of environmental data using artificial intelligence, advanced statistical analysis, and numerical modelling, as well as modelling hybridization. This paper elaborates on the concept required for an in-depth understanding of environmental pollution from the perspective of contextual data analysis and the ATLAS Project (*Artificial Intelligence Theoretical Foundations for Advanced Spatio-Temporal Modelling of Data and Processes*) supported by the Science Fund of the Republic of Serbia [2]. ATLAS aims to shift the methodology and current approaches to modelling spatio-temporal data and processes related to the global fate of air pollutants. The shift will enhance the understanding of the global environmental fate of air pollutants and lead to more thoughtful environmental protection practices, policies, and strategies.

2. Methodology

In this study, we used the air pollution data (PM_{10}, benzene, SO_2, NO, NO_2, and NOx) obtained from air quality stations within the Institute of Public Health Belgrade network and meteorological parameters obtained from ARLs Global Data Assimilation System (GDAS1). The eXtreme Gradient Boosting (XGBoost) regression machine learning method was used to investigate the relation between PM and other air pollutants, and meteorological parameter dynamics in the urban area of Belgrade, Serbia. We used the SHapley Additive exPlanations (SHAP) explainable artificial intelligence method to investigate and interpret the governing factors in shaping PM_{10} levels. The method uses Shapley values, calculated as a measure of feature importance based on a game-theory approach, that provide an impact of features on individual predictions [3]. These values are considered as fairly distributed payouts among the cooperating players (features) depending on their contribution to the joint payout (prediction). The main advantage of the approach is that SHAP represents the only possible locally accurate and globally consistent feature attribution method. We used Python XGBoost and SHAP implementations, and the TreeExplainer which reduces the complexity of exact Shapley value computation from exponential to low-order polynomial time by leveraging the internal structure of tree-based models [4]. The relative SHAP values were calculated according to Stojić et al. 2022 [5]. The stabilities of the obtained SHAP values were evaluated the bootstrap method.

3. Results

The main focus of the Air Quality Plan for the Agglomeration of the City of Belgrade for the period 2021-2031 project, supported by the City Administration of Belgrade [6], was time-resolved, contextual, in-depth, and synergetic modelling of particulate matter pollution, primarily based on machine learning, explainable artificial intelligence, and numerical modelling within the ATLAS software platform.

The most important variables which explain PM level dynamics in the urban area of Belgrade include meteorological variables, such as momentum flux intensity, standard lifted index, volumetric soil moisture content, and temperature, as well as the concentrations of benzene, NO, NO_x, and SO_2. Figure 1 represents the example of the relation between the temperature and PM_{10} concentrations in the context of SO_2 concentrations.

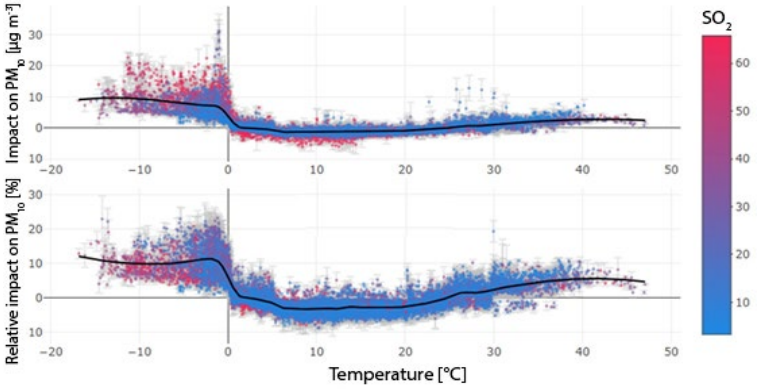

Fig. 1. The influence of temperature on PM_{10} concentrations in the context of SO_2 concentrations in Belgrade urban area

At temperatures below zero, an increase in PM_{10} concentrations by an average of 10 µg m^3 was observed. This may be the consequence of the impact of intensive combustion of fossil fuels for heating. In the ambient of high SO_2 concentrations, this increase may be as high as 20 µg m^{-3} which may be associated with the usage of fuels with high sulphur content. In the range from a few degrees above zero to about 25°C, the effect of temperature on

suspended particles is the least significant. During warmer weather, at temperatures above 25°C, the impact of temperature is related to an increase of PM concentrations to about 4 μg m^{-3} and may be related to the processes of particulates resuspension.

4. Conclusions

The understanding of environmental pollution-related processes must be enhanced by data-driven research. The concept supporting ATLAS can help to harmonize environmental research via facilitating access to environmental data, data analysis, exploration, and exploitation of the results. It can increase efficiency, creativity, and productivity of research and, at the same time, scale-up data analysis, support transdisciplinary, and lead to more thoughtful environmental protection practices, policies, and strategies. The inclusion of advanced technologies, such as artificial intelligence, in solving problems both locally and globally, clearly indicates benefits, as shown in the Air Quality Plan for the Agglomeration of the City of Belgrade project.

Acknowledgement: The authors acknowledge funding provided by the Institute of Physics Belgrade, through the grant by the Ministry of Education, Science and Technological Development of the Republic of Serbia, the Science Fund of the Republic of Serbia #GRANT No. 6524105, AI-ATLAS, as well as the City of Belgrade, Department of Environmental protection of the city administration, Serbia, Air quality plan for the City of Belgrade.

References

[1] Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C. A. et al, 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proceedings of the National Academy of Sciences, 115(38), 9592-9597.
[2] Science Fund of the Republic of Serbia - http://fondzanauku.gov.rs/atlas/?lang=en
[3] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 2522-5839.
[4] Stojić, A., Stanić, N., Vuković, G., Stanišić, S., Perišić, M., Šoštarić, A., Lazić, L., 2019. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci. Tot. Environ. 653, 140-147.
[5] Stojić, A., Jovanović, G., Stanišić, S., Romanić, S.H., Šoštarić, A., Udovičić, V., Perišić, M. and Milićević, T., 2022. The PM2. 5-bound polycyclic aromatic hydrocarbon behavior in indoor and outdoor environments, part II: explainable prediction of benzo [a] pyrene levels. Chemosphere, 289, p.133154.
[6] ATLAS Project - https://ai.ipb.ac.rs/
[7] AQP, 2021. Air quality plan for the City of Belgrade (In Serbian) – https://www.beograd.rs/images/data/7a0a3b18c076a6bfb21688ca7d314015_4302538578.pdf