ENVIRONMENTAL DATA SCIENCE SESSION

METEOROLOGICAL FACTORS GOVERNING PARTICULATE MATTER DISTRIBUTION IN AN URBAN ENVIRONMENT

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

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

[2]Singidunum University,
Belgrade, Serbia

[3]Institute of Public Health Belgrade,
Belgrade, Serbia

Correspondence:
Svetlana Stanišić

e-mail:
sstanisic@singidunum.ac.rs

Abstract:

In this study, the impact of meteorological factors on PM10 concentrations in the Belgrade urban area was investigated by using eXtreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP) attribution methods. As shown, XGBoost provided reliable PM10 predictions with relative errors in the range from approx. 19% to 26% and correlation coefficients higher than 0.95. The change in emission source intensity, momentum flux intensity, lifted index, humidity, and temperature, as well as concentrations of benzene, NO, NO_x and SO_2 were the most important variables that described the PM concentration dynamics in Belgrade urban area.

Keywords:

Particulate Matter, Meteorological Factors, Machine Learning, Explainable Artificial Intelligence.

INTRODUCTION

Suspended particulate matter refers to a complex mixture of compounds in a solid and liquid state, of organic and inorganic origin. Depending on the size, they are characterized as small/fine or $PM_{2.5}$ (with a diameter of up to 2.5 μm) and large/coarse fraction or PM_{10} (with a diameter of 2.5 μm to 10 μm). In the short run, the consequences of exposure to high concentrations of PM are irregular heartbeat and bronchial asthma exacerbation. In the long run, the adverse health effects include reduced lung capacity, increased risk of malignant diseases, increased susceptibility to systemic inflammation, as well as diabetes and its complications, exacerbation of chronical conditions, higher susceptibility to infectious viral or bacterial diseases, and increased risk of atherosclerosis and its consequences, heart attack and stroke. In addition to the impact on human health, PM has effects on the environment and other living beings. For instance, it has been widely recognized that PM contributes to the formation of acid rain, which changes the acidity of freshwater systems, reduces soil fertility, damages plant species and agricultural crops, threatens biodiversity and endangers world cultural heritage.

89

In this study, we used regression analysis by means of machine learning eXtreme Gradient Boosting method (XGBoost) for estimating the relationships between PM_{10} concentrations and a number of environmental parameters in Belgrade, Serbia [1]. The influence of meteorological factors on PM_{10} concentrations in the Belgrade urban area was investigated and explained by using SHapley Additive exPlanations (SHAP) attribution method [2]. The provided methodology has already been approved in several case studies [3], [4], [5].

2. MATERIALS AND METHODS

The ground-based data, including benzene, inorganic gaseous pollutants (SO_2, NO, NO_2, NO_x), were provided by the Institute of Public Health Belgrade, Serbia. Meteorological data were provided by the Global Data Assimilation System (GDAS1).

The relationships between PM_{10} and other environmental parameters were obtained by XGBoost. XGBoost is an ensemble method of supervised machine learning based on a sequential tree growing algorithm. Each decision tree aims to complement all the others and correct for residuals in the predictions made by the previous trees by iteratively reweighing the training data to improve regression performance. XGBoost uses a gradient descent algorithm to minimize loss when adding new models. The method includes many optimizations and enhancements. The dataset was split into training (80%) and validation (20%) sets. Hyperparameter tuning was implemented using a brute-force grid search and 10-fold stratified cross-validation. The best performing hyperparameter values were used for the final model.

SHapley Additive exPlanations (SHAP) is a method based on Shapley values, calculated as a measure of feature importance using a game-theory approach that provide an impact of features on individual predictions. SHAP values represent the only possible locally accurate and globally consistent feature attribution method.

In this paper, XGBoost and SHAP method implementations within the Python software environment were used.

3. RESULTS AND DISCUSSION

XGBoost provided reliable PM10 predictions with relative errors in the range from approx. 19% to 26% and correlation coefficients higher than 0.95 (Figure 1).

The best performing model with the lowest relative error and the highest correlation coefficient was obtained for the monitoring station of rural/industrial type located in Ovča.
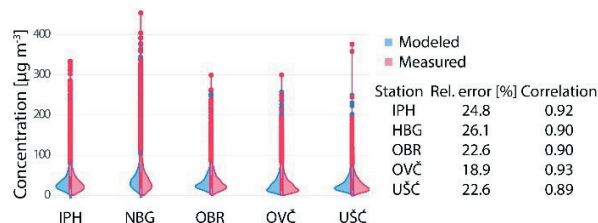

Figure 1 - XGBoost model evaluation.

On the other hand, the highest difference between modeled and measured values were detected at urban-type monitoring stations Novi Beograd and the Institute of Public Health Belgrade, both of which are exposed to traffic emissions. The modeling results were not satisfying for rural/industrial monitoring station located in Veliki Crljeni (relative error>30%, correlation coefficient<0.8), which implies that the PM10 level dynamic was mostly governed by variables other than available pollutant concentrations and meteorological parameters.

PM_{10} concentrations in Belgrade were predominantly determined by a variable that is defined as a trend of changing the intensity of emission sources (Figure 2).
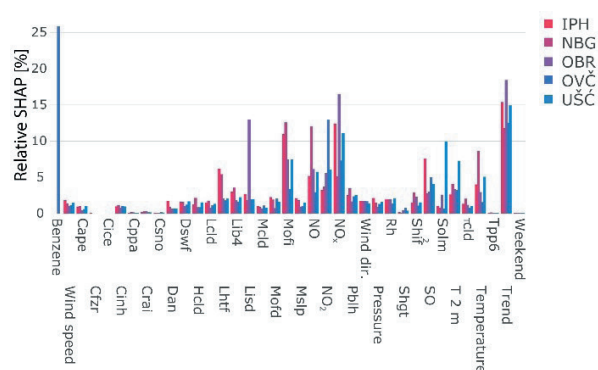

Figure 2 - Feature importance.

This variable appears to be the most important at three monitoring locations (Institute of Public Health Belgrade, Obrenovac, and Ušće), while at the stations New Belgrade and Ovča it was among the first three most significant.

Meteorological parameters including momentum flux intensity – Mofi, standard lifted index – Lisd, volumetric soil moisture content – Solm, and temperature, as well as pollutants such as benzene, NO, NO_x, and SO_2, were among the five most important variables that described the dynamics of suspended particulate matter in the territory of Belgrade area.

3.1. VOLUMETRIC SOIL MOISTURE CONTENT

In urban areas, suspended particulate matter, benzene, nitrogen oxides and SO_2 originate from common anthropogenic sources that include emissions from traffic and industrial activities, as well as the combustion of fossil fuels in thermal power plants, heating plants, and households. After the emission, the pollutants are subject to a variety of physical, chemical, and photochemical reactions. Suspended particulate matter, benzene, nitrogen oxides, and SO_2 participate in the formation of secondary atmospheric aerosols. A number of processes take place on the surface of suspended particulate matter, including gas-particles conversion, adsorption, desorption, absorption and gas dissolution, condensation of volatile compounds, as well as nucleation and coagulation. Under conditions of increased humidity in the presence of soot and inorganic oxides as catalysts (for example MgO_2 or Fe_2O_3), SO_2 will be adsorbed on the surface of suspended particulate matter to form a secondary sulfate aerosol. On the other hand, nitrogen oxides are less soluble in water compared to SO_2, so they will be less adsorbed on the surface of the particles (Figure 3). Nevertheless, in the conditions of high temperatures and intense solar activity, nitrogen oxides and volatile organic compounds such as benzene will rather participate in photochemical reactions with hydroxy, peroxy, and organic radicals in the air in which tropospheric ozone is formed.
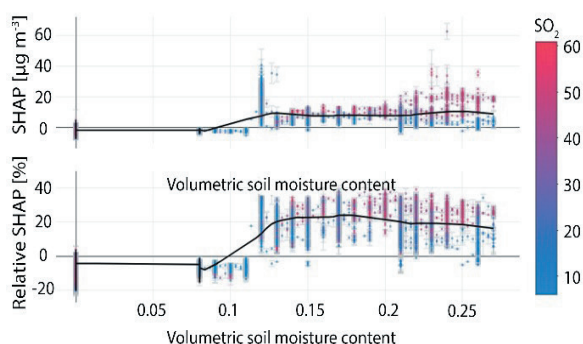

Figure 3 - PM_{10} SHAP dependency on volumetric soil moisture content and SO_2.

3.2. STANDARD LIFTED INDEX

The lifted index indicates the degree of atmospheric stability. The temperature in the atmosphere decreases with an increase in altitude, and the air that rises from the surface of the ground cools. However, when a temperature inversion occurs, air that rises to higher altitudes is warmer than the one near the ground level, which can lead to atmospheric instability. At all measuring points included in the analysis, there was a significant influence of maximum positive values of this parameter on PM_{10} concentrations (on average about 8 µg m^{-3}), which indicates that the dynamics and transformations of PM_{10} depended on atmospheric stability, Figure 4.
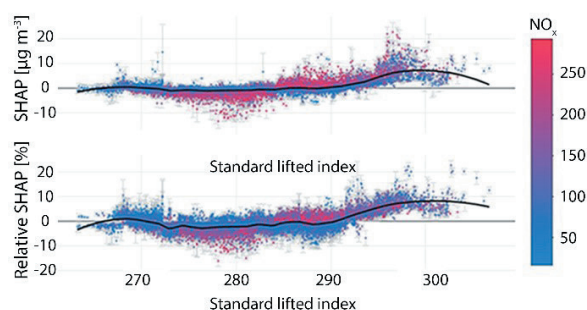

Figure 4 - PM_{10} SHAP dependency on standard lifted index and NO_x.

3.3. MOMENTUM FLUX INTENSITY

For the forecast and assessment of meteorological conditions, the momentum flux intensity is usually observed together with the wind speed. This parameter provides information important for understanding airflow in the vertical structure of the atmosphere. It can also be used to assess the stability of air mass flows in the planetary boundary layer and the occurrence of turbulent transmissions and vortices. Under stable meteorological conditions, the values of this parameter do not change significantly from the surface to the higher layers of the atmosphere and usually have lower values compared to the values measured in the case of turbulent movements. High SHAP values corresponding to increasing concentrations of suspended particulate matter up to several tens of µg m^{-3} at lower values of momentum flux intensity (<0.2) indicate a significant influence of vertical movements on the dynamics of PM_{10} and other pollutants

(NO) when they are present in high concentrations in the air (Figure 5). At all monitoring locations, a significant impact of this parameter is recorded at its lower values, which indicates that the stated relationships between pollutants are observed in conditions of stable meteorological conditions.
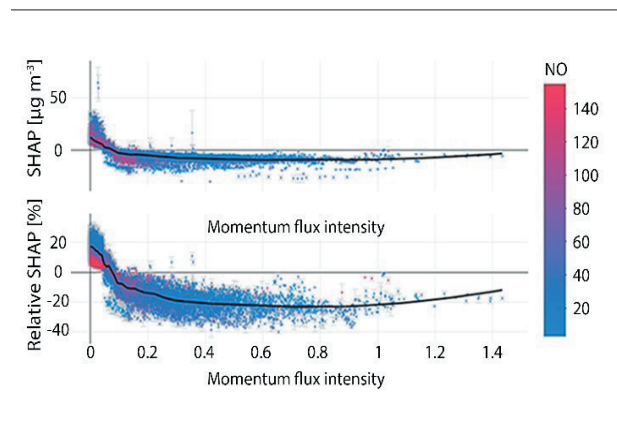

Figure 5 - PM$_{10}$ SHAP dependency on momentum flux intensity and NO.

3.4. TEMPERATURES

The intensive combustion of fossil fuels for heating at temperatures below zero contributes to an increase in PM$_{10}$ concentrations by an average of 10 μg m^{-3} (Figure 6). In the case of using fuels with high sulfur content, this increase can be as high as 20 μg m^{-3}. From only a few degrees above zero to about 25 °C, the effect of temperature on the suspended particulate matter is negligible, while during warmer weather, at temperatures above 25 °C, the resuspension of particles contributes to an increase in concentrations of about 4 μg m^{-3} on average.

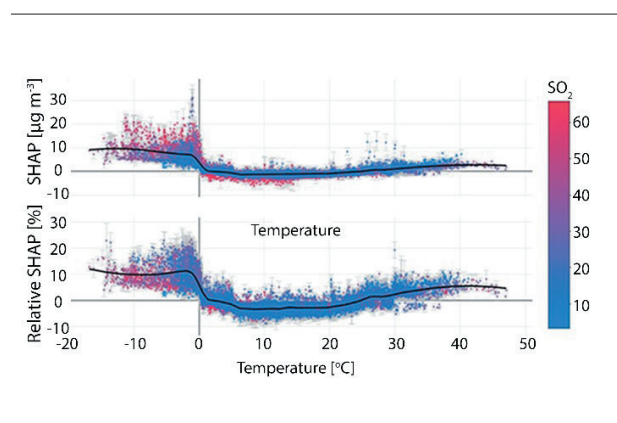Figure 6 - PM$_{10}$ SHAP dependency on temperature and SO$_2$.

3.5. ATMOSPHERIC PRESSURE

The effect of pressure on the concentrations of suspended particulate matter is relatively small and constant (Figure 7). Somewhat stronger impact on their level dynamics is recorded in the urban atmosphere being characterized by the presence of NO$_2$ higher concentrations. Low-pressure conditions can contribute to a reduction in PM concentrations of up to 3 μg m^{-3}.
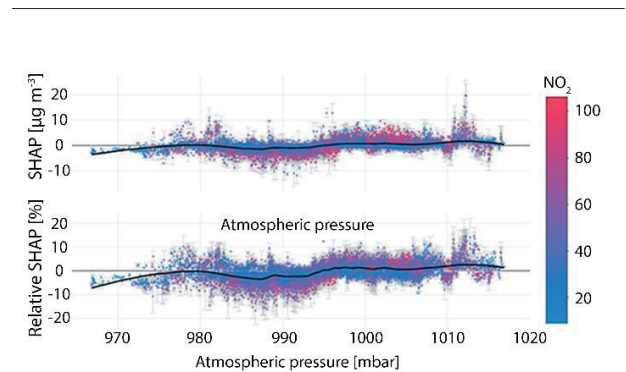

Figure 7 - PM$_{10}$ SHAP dependency on atmospheric pressure and NO$_2$.

4. CONCLUSION

Based on the analysis of the dependence of suspended particulate matter concentrations on environmental factors (concentrations of SO$_2$, NO, NO$_2$, NO$_x$, and benzene, modeled meteorological parameters - GDAS base, trend, daily and weekend variations), the change in emission source intensity is singled out as a variable that dominantly determines the dynamics of PM$_{10}$ concentration in Belgrade. This variable stands out as the most important one in three measuring points - Institute of Public Health Belgrade, Obrenovac, and Ušće. Also, meteorological parameters including momentum flux intensity, lifted index, humidity, and temperature, as well as concentrations of benzene, NO, NO$_x$, and SO$_2$ were among the five most important variables that described the PM concentration dynamics in Belgrade urban area.

5. ACKNOWLEDGEMENTS

AI-ATLAS, as well as the City of Belgrade, Department of Environmental protection of the city administration, Serbia, Air quality plan for the City of Belgrade.

REFERENCES

[1] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco California USA, 2016.

[2] S. Lundberg, G. Erion, H. Chen, A. DeGrave, J. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature machine intelligence*, vol. 2, no. 1, pp. 56-67, 2020.

[3] A. Stojić, D. Maletić, S. Stanišić Stojić, Z. Mijić and A. Šoštarić, "Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods," *Science of the Total Environment*, vol. 521, pp. 19-26, 2015.

[4] A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić and L. Lazić, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition," *Science of The Total Environment*, vol. 653, pp. 140-147, 2019.

[5] M. Perišić, D. Maletić, S. Stanišić Stojić, S. Rajšić and A. Stojić, "Forecasting hourly particulate matter concentrations based on the advanced multivariate methods," *Journal of Environmental Science and Technology*, vol. 14, no. 5, pp. 1047-1054, 2017.