



# THE USE OF INNOVATIVE METHODOLOGY FOR THE CHARACTERIZATION OF BENZENE, TOLUENE, ETHYLBENZENE AND XYLENE SOURCES IN THE BELGRADE AREA

Svetlana Stanišić<sup>1\*</sup>,  
Mirjana Perišić<sup>1,2</sup>,  
Andreja Stojčić<sup>1,2</sup>

<sup>1</sup>Singidunum University,  
Belgrade, Serbia

<sup>2</sup>Environmental Laboratory, Institute of  
Physics Belgrade,  
National Institute of the Republic of Serbia,  
University of Belgrade,  
Belgrade, Serbia

## Abstract:

The growth of urban population, economic development, urbanization and transport have a strong impact on environmental pollution. The increase in air pollutant concentrations over the last few decades has been in focus of contemporary science and research mainly for its adverse effects on public health, environment and climate change. In this paper, we are using the innovative integrated methodology for spatio-temporal air pollution modeling, based on receptor-oriented air circulation modeling and artificial intelligence implemented through machine learning methods for detailed characterization of toxic, mutagenic and carcinogenic representatives of volatile organic species – benzene, toluene, ethylbenzene and xylene, in the Belgrade area. Also, the study evaluates the possibilities of spatio-temporal forecast based on the integrated methodology. The results suggest that temperature and wind speed represent the main parameters which govern the spatio-temporal distribution of benzene, while the impact of other factors shows significant variations depending on the characteristics of receptor location.

## Keywords:

Keywords – BTEX, artificial intelligence, machine learning, volatile organic compounds.

## 1. INTRODUCTION

Ambient air pollution accounts for an estimated 4.2 million deaths per year due to cardiovascular, malignant and chronic respiratory diseases [1]. Around 91% of the world's population lives in places where air pollution levels exceed World Health Organization limits [2]. Holgate (2017) emphasizes that 40,000 excess deaths in the UK annually can be associated with low air quality, and society would be much more aware of its significance if this mortality was the consequence of drinking polluted water [3].

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

## Correspondence:

Svetlana Stanišić

## e-mail:

sstanic@singidunum.ac.rs



group of organic species with boiling points  $<250$  °C. Their representatives are benzene, toluene, ethylbenzene and xylene, commonly referred to as BTEX. Over the last few decades in developed countries, reducing the levels of BTEX is still challenging [4], due to their enormous chemical diversity and abundance, their numerous emission sources, their complex atmospheric chemistry, insufficient funds for establishment and maintenance of monitoring networks, and the fact that abatement programs might have negative impacts on economic output.

The health effects of BTEX are diverse. For instance, the research has shown that long-term exposure to benzene increases the risk of developing malignant blood disorders, while long-term exposure to toluene causes renal tubular acidosis [5]. Furthermore, the studies have shown that after reduction of benzene, styrene, and tetrachloroethylene concentrations in industrial and urban areas, lifetime cancer risk decreased by one order of magnitude [6]. Populations in highly industrialized areas, socioeconomically deprived, as well as children, pregnant women and elderly people, appear to be more susceptible to pollution-related morbidity and mortality [7]. Apart from their impact on human health, BTEX and other VOCs are associated with climate change and increases in the oxidation capability of the atmosphere [8]. Not only volatile species directly and indirectly contribute to climate change, but their emissions and fates are expected to be influenced and increased by the forthcoming global warming.

Despite the fact that the big shifts in development and integration of different approaches in the area of environmental science have been made recently, spatio-temporal air pollution modeling remained a challenge. Two main approaches are typically employed to forecast air quality and to identify the factors that govern certain pollutant concentrations. The first approach relies on atmospheric diffusion models, while the second refers to statistical models that capture the essential relationships between the variables [9]. Thereby, multidimensionality and size of data sets, as well as the complexity of air pollutant processes and interactions, set too high requirements for conventional statistical methods. For this reason, methods of machine learning, a subfield of artificial intelligence that enables automatized big data analysis and development of learning algorithms, have been introduced into environmental science and research. In this paper, we used the innovative and integrated methodology for spatio-temporal air pollution modeling, based on artificial intelligence and imple-

mented through machine learning methods for detailed characterization of dominant and particular sources of BTEX in a wider region surrounding receptor site that was not covered by regular monitoring. The presented methodology has the potential to provide the basis for establishment of unique and sustainable system for air pollution source identification and enhanced air pollution data coverage that doesn't require additional investments in monitoring equipment. In long term, results of such an approach would provide a solid basis for establishing the sustainable system aimed at improved air pollution management and control.

## 2. METHODOLOGY

Machine learning algorithms are based on the extraction of patterns and selection of specific attributes from a large number of data, while eliminating irrelevant information. By identification of prediction most important attributes, machine learning methods acquire knowledge and define substantial relationships that exist between input and output parameters by placing a special focus on the data aspect that is most useful for efficient forecasting. The fact that methods based on decision trees, such as Gradient Boosting and Random Forest, have been shown to provide inconsistent attribute contributions, has led to the development of SHAP (SHapley Additive exPlanation), a method that estimates the contribution of each instance of an attribute, which further enables interpretation of the model outputs [10, 11].

EXtreme Gradient Boosting (XGBoost) can be assigned as ensemble method of supervised machine learning which combines the results of more than one decision tree approaches. The main feature of the XGBoost method is focus on obtaining more precise prediction, compared to the one that could be provided by applying a single constitutive decision algorithm. XGBoost method is based on boosting technique that sequentially defines smaller series of decision trees for classifying input data into two or more attribute-defined classes. Each consecutive decision tree is trained through iterations by taking into account the registered errors of previous classification.



### 3. RESULTS AND DISCUSSION

The variations of particular meteorological parameter affect the changes in other related parameters, which makes it difficult to distinguish between their particular impact on air pollution phenomena. Thus, the impact of meteorological factors is not observed as isolated impact of a single parameter and its variations, but rather as an impact of a certain weather type. In addition to this, a number of other factors can contribute to final impact of meteorological conditions, including the distribution of pollutant emission sources, local topography, street geometry and distribution of all elements and surfaces that can be of significance for air flow regime, pollutant dispersion conditions, their transport pathways and thus, the spatio-temporal variability of their levels. For instance, Liao et al. (2017) have identified ten typical air circulation types within one of the most polluted areas of China and explored their synergetic contribution with topography to local air quality [12]. Ning et al. (2019) have shown that air pollution forecast can be significantly impacted by the complexity of terrain areas because topographic features can to a certain extent limit pollutant dispersion under different weather conditions [13]. In addition to this, the consistency of meteorological conditions significantly affects the extent of volatile pollutant dispersion. For instance, previous study which has been dealing with the accidental benzene release risk assessment in an urban area using an atmospheric dispersion model has shown that benzene spreads over a much larger area during the nighttime due to a stable boundary layer, while during the daytime the enhanced vertical mixing results in limited dispersion of pollutant over the study area [14].

The results of this study suggest that low temperatures and weak to moderate wind represent the main parameters which govern the spatio-temporal distribution of benzene, while the impact of other factors shows significant variations depending on the characteristics of receptor location (Fig. 1). In a similar manner, the presented figures demonstrating TEX distribution can be interpreted (Fig. 2). As can be seen, the figures also include the relative errors and correlations between the observed and estimated BTEX concentrations, which can contribute to better XGBoost method performance evaluation.

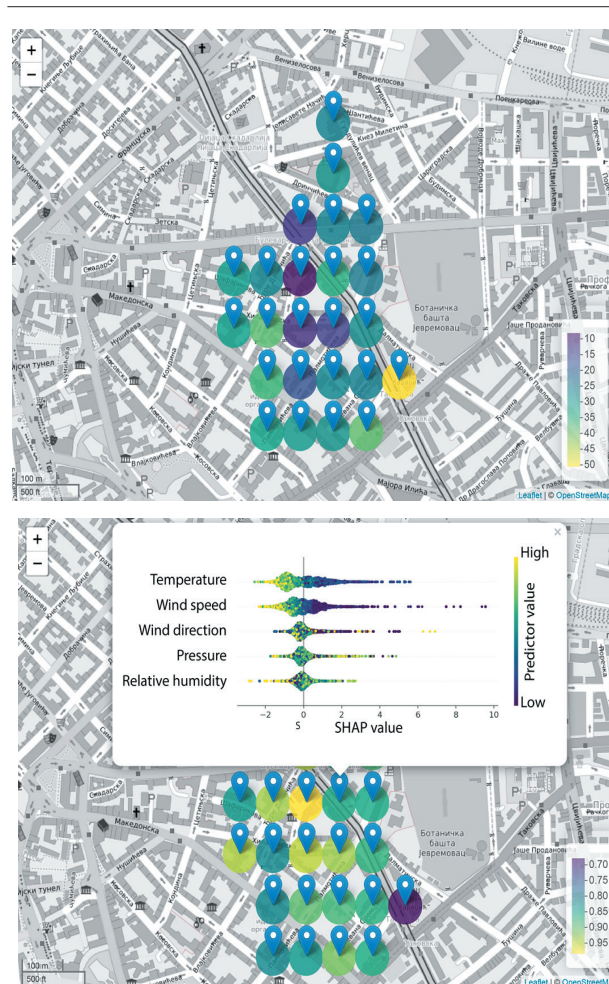


Fig. 1. Benzene forecast based on meteorological parameters – relative error [%] (above) and SHAP values and predicted/observed correlation coefficients (below).

Based on high correlation coefficients ( $>0,80$ ) that were obtained for most of the analyzed data, it can be concluded that XGBoost can be rated as successful and efficient method for air pollution forecasting in the urban area. It should be emphasized that the estimated method errors are significantly lower than uncertainty (50%) which has been requested for evaluation of average annual benzene concentrations obtained by conventional modeling.

The results have also shown high correlations ( $r>0,70$ ) between toluene and nitrate oxide concentrations, which can be considered as indicators of fossil fuel burning, which suggests the common origin of these pollutants in all locations being covered by the conducted analysis, except the old city area and Kalemegdan (Fig. 3), where the toluene concentrations most probably reflected the oxygenated air masses in some narrow canyon-type streets. Furthermore, high correlations ( $r>0,70$ ) between benzene and inorganic oxides ( $\text{NO}_x$ ,  $\text{CO}$ ,  $\text{SO}_2$ ) in western city region suggest the detrimental impact of remote air pollution sources, such as thermal plant Nikola Tesla A and B in Obrenovac (Fig. 4).



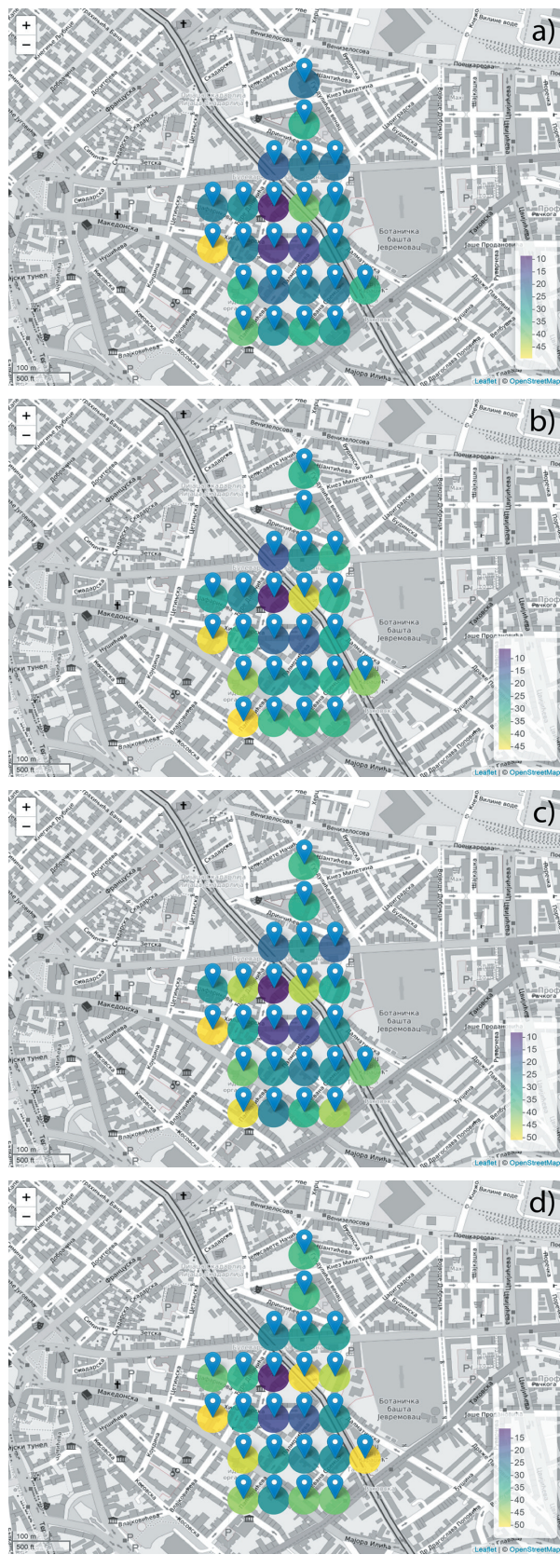


Fig. 2. Toluene (a), m,p-xylene (b), o-xylene (c), and ethylbenzene (d) relative error [%] forecasts based on meteorological parameters.

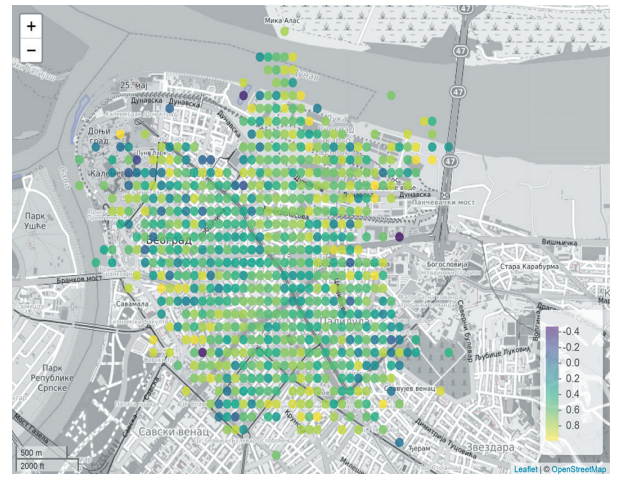


Fig. 3. Toluene and NO correlation coefficient.

Relatively low correlations between benzene and inorganic oxides in the northern and eastern city area suggest that benzene in this urban region can be related to evaporations and emissions coming from petrochemical industry, Oil refinery Pančevo and chemical industry Petrohemija.

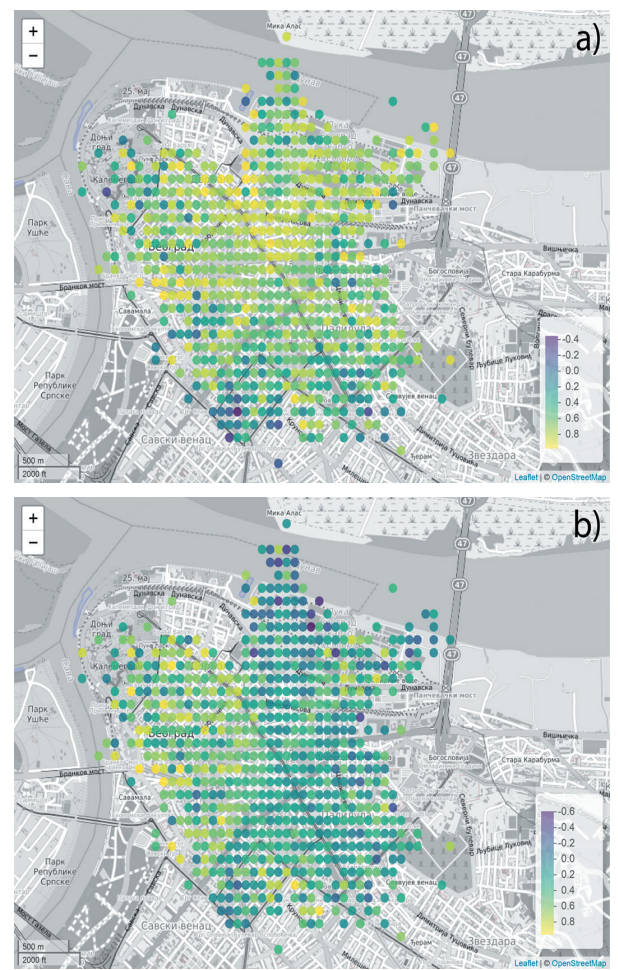


Fig. 4. Benzene correlation coefficients with CO (above) and NOx (below).





## 4. CONCLUSION

As can be concluded, we have demonstrated the use of efficient methods for spatio-temporal BTEX concentration modeling in the Belgrade area, based on receptor-oriented air circulation modeling and artificial intelligence implemented through machine learning and explainable artificial intelligence methods. The estimated method errors were shown to be lower than the requested uncertainty for conventional modeling. According to the results, temperature and wind speed represented the main parameters which governed the spatio-temporal distribution of benzene. In addition to this, the correlations between different air pollutant concentrations were considered for determination of their origin in all locations covered by the conducted analysis.

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] S. Rajagopalan, A-K. Sadeer, D. R. Brook, "Air pollution and cardiovascular disease: JACC state-of-the-art review," *J. Am. Coll. Cardiol.* vol. 72, pp. 2054-2070, 2018.
- [2] C. Y. Wright, D. A. Millar, "A global statement for air pollution and health," *Clean Air J*, vol. 29, pp. 1-2, 2019.
- [3] S. T. Holgate, "Every breath we take: the lifelong impact of air pollution—a call for action," *Clin. Med*, vol. 17, pp. 8, 2017.
- [4] M. J. Milazzo, J. M. Gohlke, D. L. Gallagher, A. A. Scott, B. F. Zaitchik, L. C. Marr, "Potential for city parks to reduce exposure to BTEX in air," *Environ. Sci. Process. Impacts*, vol. 21, pp. 40-50, 2019.
- [5] E. J. Werder, L. S. Engel, A. Blair, R. K. Kwok, J. A. McGrath, D. P. Sandler, "Blood BTEX levels and neurologic symptoms in Gulf states residents," *Environ. res*, 175, pp. 100-107, 2019.
- [6] J. E. C. Lerner, T. Kohajda, M. E. Aguilar, L. A. Massolo, E. Y. Sánchez, A. A. Porta, P. Opitz, G. Wichmann, O. Herbarth, A. Mueller, "Improvement of health risk factors after reduction of VOC concentrations in industrial and urban areas," *Environ. Sci. Poll. Res.*, vol. 21, pp. 9676-9688, 2014.
- [7] S. Bose, G. B. Diette, "Health disparities related to environmental air quality," *In Health disparities in respiratory medicine*, pp. 41-58, Humana Press, Cham, 2016.
- [8] P. Campbell, Y. Zhang, F. Yan, Z. Lu, D. Streets, "Impacts of transportation sector emissions on future US air quality in a changing climate. Part II: Air quality projections and the interplay between emissions and climate change," *Environ. Poll.*, vol. 238, pp. 918-930, 2018.
- [9] H. B. Ly, L. M. Le, L. V. Phi, V. H. Phan, V. Q. Tran, B. T. Pham, T. T. Le, S. Derrible, "Development of an AI model to measure traffic air pollution from multisensor and weather data," *Sensors*, vol. 19, pp. 4941, 2019.
- [10] M. V. García, J. L. Aznarte, "Shapley additive explanations for NO<sub>2</sub> forecasting," *Ecol. Inform.*, vol. 56, pp. 101039, 2020.
- [11] A. Stojić, N. Stanić, G. Vuković, S. Stanišić, M. Perišić, A. Šoštarić, L. Lazić, "Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition" *Sci Total Environ*, vol. 653, pp.140-147, 2019.
- [12] Z. Liao, M. Gao, J. Sun, S. Fan, "The impact of synoptic circulation on air quality and pollution-related human health in the Yangtze River Delta region," *Sci. Tot. Environ*, vol. 607, pp. 838-846, 2017.
- [13] G. Ning, S. H. L. Yim, S. Wang, B. Duan, C. Nie, X. Yang, J. Wang, K. Shang, K, "Synergistic effects of synoptic weather patterns and topography on air quality: a case of the Sichuan Basin of China," *Climate Dynamics*, vol. 53, pp. 6729-6744, 2019.
- [14] S. C. Truong, M. I. Lee, G. Kim, D. Kim, J. H. Park, S. D. Choi, G. H. Cho, "Accidental benzene release risk assessment in an urban area using an atmospheric dispersion model," *Atmos. Environ*, vol. 144, pp. 146-159, 2016.