

Explainable Modelling of Particulate Air Pollution in Warsaw

Małgorzata Łazęcka¹, Katarzyna Woźnica¹

¹Warsaw University of Technology, Faculty of Mathematics and Information Sciences
75 Koszykowa, 00-662 Warsaw, Poland
m.lazeczka@mini.pw.edu.pl; k.woznica@mini.pw.edu.pl

Extended Abstract

Despite controlling emissions, air quality has remained global issue for authorities and non-governmental organisations. Cities are one of the most affected area by particulate matter concentrations (PM10 and PM2.5, the number stands for particles with diameters less or equal than 10 μm and 2.5 μm respectively). Both PM10 and PM2.5 seriously affect human health and they increased concentrations cause e.g. severe respiratory diseases. A variety of machine learning methods such as regression models, neural networks, time series analysis have been employed for forecasting air quality and revealing the most influential predictors [1, 2, 3].

In our paper, on the example of Warsaw, we statistically model the relationship between PM10 and PM2.5 and considered as significant factors such as vehicle traffic and weather conditions (e.g. wind speed, temperature, pressure and humidity). The input data used in this study is available online with no restrictions. The main objective is to provide a machine learning model predicting level of particular air pollution and to research which factors are the most correlated with releasing pollutants into the atmosphere in Warsaw. We put special attention to verifying if the vehicle traffic has significant impact on particulate matter concentrations in the urban area.

First approach consists in modelling hourly concentrations of PM10 and PM2.5 with generalised linear models built on subgroups chosen with respect to the weather condition. Based on collected data, estimated density functions of PM10 and PM2.5 are investigated and the assumption that these concentrations of air pollution have distributions from exponential family is made. Applied cluster analysis regarding meteorological variables captures and simplifies the problem of seasonal nature of weather during calendar year. Parametric models and cluster-wise regression have been used previously in the analysis of PM concentrations (cf. for example [4, 5]). The benefits of clustering and parametric settings are based on interpretability and simplicity of the models on subgroups. This approach also might give better results in terms of performance of the method.

Due to COVID-19 and unexpected restriction in relocating new data from time when vehicle traffic is scarce and in big cities there are no traffic jams and traffic flow is smooth has been collected. Because of the limited fuel emission, this data has been analysed and checked if there is significant decline in reported PM10 and PM2.5 concentration. Since this decrease in the newest data the impact of vehicle traffic may be treated as regardless. What is more, this enables collation of an inference from statistical modelling from average vehicle traffic with observed limited fuel emission during COVID-19 restriction time and validate consistency of conclusions from parametric models and cluster-wise generalised linear models.

References

- [1] G. Corani, “Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning“, *Ecological Modelling*, vol. 185 no. 2, pp. 513-529, 2005.
- [2] A. Elangasinghe, N. Singhal, K. Dirks, J. Salmond, S. Samarasinghe, “Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modelling and k-means clustering“, *Atmospheric Environment*, vol. 94, pp. 106-116, 2014.
- [3] S. Abdullah, M. Ismail, A-M. Najah, A. Abdullah, “Forecasting Particulate Matter Concentration Using Linear and Non-Linear Approaches for Air Quality Decision Support“, *Atmosphere*, vol. 20, pp. 667, 2019.
- [4] D. Zhu, C. Cai, T. Yang, X. Zhou, “A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization“, *Big Data and Cognitive Computing*, vol. 2, pp. 5, 2018.
- [5] J.-M. Poggi, B. Portier, “PM10 forecasting using clusterwise regression“, *Atmospheric Environment*, vol. 45, pp. 7005-7014, 2011.