Archives of Environmental Protection
Vol. 46 no. 1 pp. 93–99

PL ISSN 2083-4772
DOI 10.24425/aep.2020.132530

Explicit and implicit description of the factors impact on the NO_2 concentration in the traffic corridor

Joanna Amelia Kamińska*, Tomasz Turek

Wrocław University of Environmental and Life Sciences

*Corresponding author's e-mail: joanna.kaminska@upwr.edu.pl

Keywords: nitrogen dioxide; traffic flow; meteorological conditions; random forest; linear regression.

Abstract: High concentrations of nitrogen dioxide in the air, particularly in heavily urbanized areas, have an adverse effect on many aspects of residents' health. A method is proposed for modelling daily average, minimal and maximal atmospheric NO_2 concentrations in a conurbation, using two types of modelling: multiple linear regression (LR) an advanced data mining technique – Random Forest (RF). It was shown that Random Forest technique can be successfully applied to predict daily NO_2 concentration based on data from 2015–2017 years and gives better fit than linear models. The best results were obtained for predicting daily average NO_2 values with R^2=0.69 and RMSE=7.47 μg/m^3. The cost of receiving an explicit, interpretable function is a much worse fit (R^2 from 0.32 to 0.57). Verification of models on independent material from the first half of 2018 showed the correctness of the models with the mean average percentage error equal to 16.5% for RF and 28% for LR modelling daily average concentration. The most important factors were wind conditions and traffic flow. In prediction of maximal daily concentration, air temperature and air humidity take on greater importance. Prevailing westerly and south-westerly winds in Wrocław effectively implement the idea of ventilating the city within the studied intersection. Summarizing: when modeling natural phenomena, a compromise should be sought between the accuracy of the model and its interpretability.

Introduction

The main source of emissions of nitrogen oxides are exhaust gases, which come mainly from high-temperature ignition in vehicle engines. The purpose of the work is to assess the possibility of determining the impact of meteorological factors on the concentration of NO_2 in the air in the communication canyon and to compare the explicit and implicit method in modelling of extreme values and daily average. This paper presents two deterministic models to determine the pollutant concentration value named point-based model (without reference to previous values of the pollution). Among point-based models, there exist several principal methods of mathematical modelling. Multidimensional regression models are still popular (Ping and Harrison 1997, Aldrin and Haff 2005, Zhang et al., 2015). The main advantage of linear models is interpretable explicit function that can be used to determine the quantitative impact of each predictor on the value of the explaining variable. Models based on machine learning are more computationally advanced and successfully used in pollution concentrations modelling. Among them one can mention: artificial neural networks (Elangasinghe et al. 2014, Nejadkoorki and Baroutian 2012), single random trees (Singh et al. 2013), more complex structures – random forest (RF) (Zhu et al. 2019, Kamińska 2018a, Laña et al. 2016) and boosted regression trees e.g. (Kamińska 2018b). The aim of this study is to investigate the possibility of forecasting daily NO_2 concentration using the RF method and to compare the results with multivariate linear regression (LR) based on the same dataset. A non-time-sensitive approach is necessary to make it possible to apply the method to analysis of various scenarios of changes intended to reduce emissions and improve air quality. The development of environmental indicator-based assessment methods can be effectively implemented in location intelligence system (Szewrański et al. 2018) which constitute decision support systems for local decision makers (Kazak et al. 2018). Some analysis concerning air quality modelling has been also presented (Czechowski at al. 2013, Holnicki et al. 2017).

Methods

RF belongs to the machine learning methods. It is built from a set number of simple decision trees. Each component tree is created for another, randomly selected subset of data (sampling with replacement). In the present analysis, each training set includes another subset of 50% cases. The answer of the random forest is taken by aggregating and averaging the individual predictions of each component tree. This construction method improves modelling performance relative to other machine learning algorithms and linear regression models (Archer and Kimes 2001). The Classification and Regression Trees method

used to create random trees allows to determine the validity of a variable. The most important variable is assigned an importance of 100 (Breiman 2001).

The classical multiple linear regression model (LR) was used as the basis method for the assessment of RF effectiveness. The main advantage of the modelling methods using the explicit function is interpretability of coefficients and, therefore, the relationships that take place and their strength. However, the use of linear regression requires fulfilling a number of assumptions, in particular: no redundant predictors, normal distribution of an explanatory variable and for the created model: homoscedantity and normal distribution of residues.

Data sources

The presented methods were used to predict current NO_2 concentration values based on traffic flow and meteorological conditions in selected intersection in Wrocław (Poland, Europe) in 2015–2017. Due to the significance of the traffic impact on the concentration of nitrogen dioxide in the air, the proximity of air quality measurement stations and traffic volume is very important. The only location in Wroclaw where the measurement of pollution concentration is carried out in the immediate vicinity of the traffic measurement is the Hallera and Powstańców Śląskich intersection (Fig. 1).

Pollution concentration data are collected by the Provincial Environment Protection Inspectorate and measured at hourly intervals. The air inlet to the system is located 3m above ground level. Daily extreme values and averages have been calculated.

Traffic data are provided by the Traffic and Public Transport Management Department of the Roads and City Maintenance Board in Wrocław. The data contain number of all vehicles (cars, buses, trucks etc.) passing through the measurement plane in a given traffic lane or lanes. The daily average, daily maximum and daily minimum values were determined on the basis of a hourly sums of vehicles passing the intersection. In the 2015–2017 research period, there were five days on which there was a failure of the counting system as well as seven days with no data. Therefore, the data from 1084 days (of 1096 total) were subjected to analysis. Meteorological data are provided by the Institute of Meteorology and Water Management (IMGW) at only one station in Wrocław, located on the outskirts of the city (9 km from the intersection in a straight line). The meteorological data set contains hourly air temperature, wind speed, wind direction and relative humidity. The existing differences between the values of meteorological factors registered at the airport and the actual at the analysed intersection certainly exist. They are not, however, an obstacle to using data from a distant point because both mathematical models used will take into account possible differences in equation coefficients (LR model) or learning process (RF). Wind direction data were initially obtained in continuous numerical form, but it was not appropriate to use the wind direction in degrees as an explanatory variable because values with a large difference may correspond to a very similar direction (for example, 1° and 360°). For this reason, wind direction in RF was instead expressed using eight categories with 45° separations (N, NE, E, etc.). In linear regression the wind direction was transformed from degrees to interval according to the equation:

$$num\_direction = sin\left(\frac{2\pi\alpha}{180} - \frac{\pi}{2} - \frac{2\pi}{9}\right) \tag{1}$$

where α is wind direction in degrees. Transformation corresponds to the geographical arrangement of intersecting streets. Translation (in function) or in another words rotation (in Fig. 2) by the angle of $90^o = \frac{\pi}{2}$ provides the wind blowing in the axis of the communication canyon with the value 1. The angle between street and W–E directions is $20^o = \frac{\pi}{9}$. Due to a twofold increase in the frequency of the sinus function, a translation of $40^o = \frac{2\pi}{9}$ was applied. Figure 2 presents graphically transformation of wind direction to *num_direction* variable.

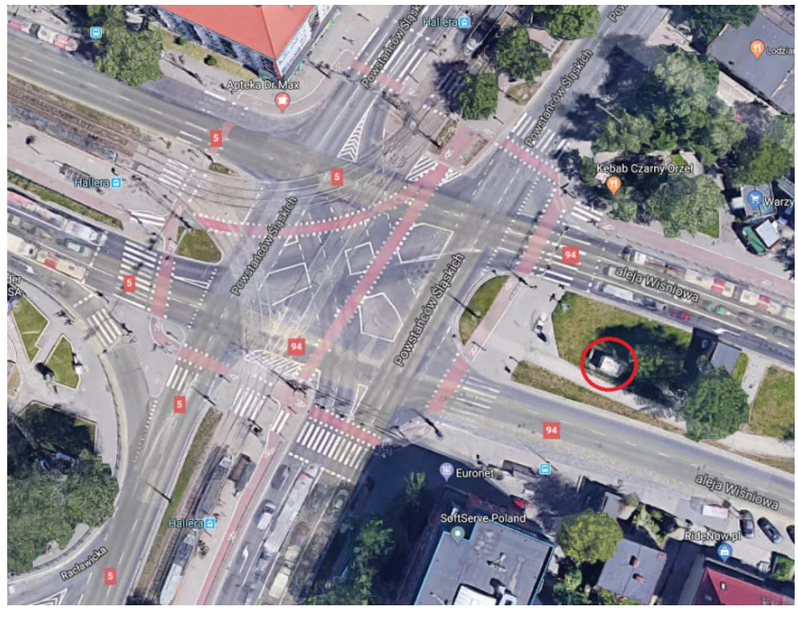

Fig. 1. Monitoring intersection in Wrocław: pollution station in red circle (source: www.google.maps)

Results

On the basis of the data presented above, three random forests were built to solve the regression problem. Each forest was built of 100 trees. As explanatory variables in each of the models considered were: wind condition (daily maximum, daily average wind speed [m/s], and daily average wind direction [–]), traffic flow (daily maximum, average and minimum traffic flow [veh/h]), daily average relative humidity [%] and daily average air temperature [°C]. Minimum daily wind speed due to the significant impact of only strong winds in the evacuation of pollutants was not included in the model. The phenomenon of the occurrence of pollutants in windless conditions is also described by the low value of daily average and daily maximum wind speed.

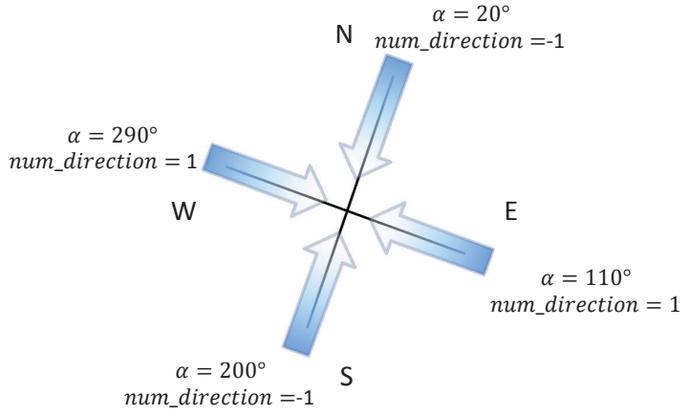

Fig. 2. Graphical presentation of wind direction transformation

Table 1. Descriptive statistics for all variables (N=1084).

Variable		Mean	Median	Min.	Max	St. Dev.
NO_2 concentration [$\mu g/m^3$]	Daily maximum	86.0	83.5	25.7	231.6	25.2
	Daily average	50.4	49.9	13.6	112.0	13.3
	Daily minimum	21.5	19.8	1.7	62.7	10.5
Traffic flow [*veh*]	Daily maximum (X_1)	4822	5094	2066	5712	647
	Daily average (X_2)	2768	3013	597	3361	498
	Daily minimum (X_3)	241	203	44	731	92
Wind speed [m/s] (daily average)	Daily maximum (X_4)	5.65	5	2	19	2.22
	Daily average (X_5)	3.13	2.79	0.67	9.21	1.38
	Daily minimum (X_6)	1.05	1.00	0.00	8.00	1.05
Air temperature [°C] (daily average) (X_7)		10.7	10.2	-12.9	29.4	7.7
Relative humidity [%] (daily average) (X_8)		74.8	75.1	40.0	98.5	11.6

Table 2. Goodness of fit coefficients

Coefficient	Equation	Minimum		Average		Maximum	
		RF	LR	RF	LR	RF	LR
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$	0.50	0.33	0.69	0.59	0.58	0.42
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}$	7.43	11.59	7.47	15.38	16.32	27.42
Mean Absolute Deviation Error	$MADE = \frac{1}{N}\sum_{i=1}^{N}\lvert\hat{y}_i - y_i\rvert$	5.84	9.00	5.65	11.81	12.00	20.41
Mean Absolute Percentage Error [%]	$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert\hat{y}_i - y_i\rvert}{\lvert y_i\rvert}$	27.1	51.6	11.4	25.7	13.9	25.48

where $\hat{y}_i$ is the *i*th theoretical value (from the model), y_i is the *i*th empirical (real) value, $\bar{y}$ is the mean empirical value, *N* is the sample size.

To compare the quality of the model fit, several goodness of fit measures were considered: RMSE, MADE, MAPE and R^2 (tab. 2). The values of goodness of fit coefficients explicitly favor RF-mean model. R^2 equal 0.69 for RF average NO_2 concentration model indicates a good fit of the model, compared with the results of other researchers: 0.52 based on 5220 monitors in 58 countries (Larkin et al. 2017), R^2 up to 0.54 (Sayed et al. 2016), R^2 up to 0.58 (Kamińska 2018a). MAPE and R^2 indicate the best accuracy in reflecting reality for this model. RSME values are comparable to those obtained by Zhu et al. (2019) for *monthly* average NO_2 concentration (RSME=11.0 μg/m^3). It can therefore be concluded that the model presented in this paper effectively predicts daily mean values of NO_2 concentration. The daily mean values of NO_2 concentration are the lowest and the daily minimum is the most diverse (coefficients of variation of 26% and 49% respectively). Due to the diversity of this natural phenomenon, prediction of the daily minimum value is the most difficult among all considered cases, and the matching of the proposed model is the smallest for this variable. All models underestimate high NO_2 values, while they overestimate low values. This phenomenon has also been observed in models presented by other researchers e.g. (Singh et al. 2013). This is a consequence of the use of the method of averaging the results obtained from all decision trees considered in the model. This technique prevents overshooting and, consequently, over-adjustment to the data used in the learning process at the expense of the loss of detail for extreme values.

For determining NO_2 concentration in any form (minimum, average, maximum) the most important predictor is the average wind speed (Fig.3.). Batista and de Lieto Vollaro (2017) and Laña et al. (2016) received similar results. Wind conditions and traffic flow have almost an equal feature importance for daily mean and daily maximum NO_2 concentration, as was also demonstrated by Ping Shi and Harrison (1997). For daily minimum NO_2 concentration, wind conditions are more important than traffic flow. In the maximum daily NO_2 concentration modelling, a greater impact is noted for relative humidity and air temperature and maximal traffic flow. However, when NO_2 concentrations and air temperature are high, but relative humidity and wind speed low, the chemical reactions that transform primary pollutants into secondary ones become more intense (Altenstedt 1998). The air temperature and air humidity then take on the greater importance. The high maximum concentration of NO_2 is influenced by the maximum daily traffic directly related to the main source of nitrogen oxides from exhaust gases in more significant way than in the case of the average and minimum. The pollution ambient concentration becomes stronger then. The observed relationship between the distribution of the importance of variables and the modelled concentration range (low, medium, high) is consistent with the conclusions of Kamińska (2019), where low and high *hourly* values of NO_2 concentration were modelled independently.

According to NO_2 concentration histograms, Q-Q plots and the values of , K-S and W Shapiro-Wilk normality tests it was decided that the dependent variable in linear regression will be logarithmic transformation of each of the variables (to ensure normality of distribution). Based on correlation coefficient from the full set of variables analyzed in the RF model as redundant: maximum daily traffic flow (r=0.95 with average traffic flow) and maximum daily wind speed (r=0.86 with average wind speed) were excluded. Then the outliers detection for independent variables values based on 3σ interval analysis has been made. The significant exceedance was recorded for minimum traffic flow. The 3σ interval for this variable is <0;520> and the values considered to be significantly outliers and therefore removed from the data set are equal: 721 and 703 vehicles passing the intersection. Thus, six explanatory

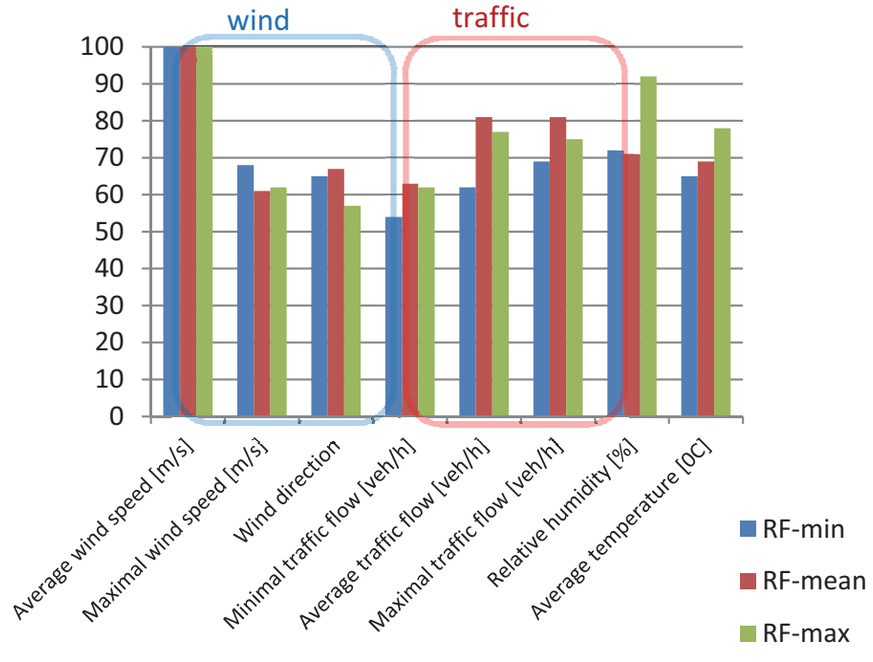

Fig. 3. Feature importance of each variable for NO_2 concentration

variables in 1082 samples took part in the regression analysis. The analyzed linear model describes the following equation:

$$\log(NO_2) = b_0 + b_1X_1 + b_2X_2 + b_3X_5 + b_4X_7 + b_5X_8 + b_6X_9 \quad (2)$$

where X_i *for* $i = 1, ..., 8$ are as in table 1 and is wind direction (1).

The linear coefficients are listed in Tab. 3. Residuals have a normal distribution for all three models therefore, the models can be assesed as correct.

Coefficients indicating a positive or negative relationship between predictors and dependent variables are in all models consistent with the exception of variable traffic flow for which there is also a lack of statistical significance of the coefficient. Generally, the higher wind speed the lower nitrogen dioxide concentration in the air. Moreover, there was a stronger effect of the average wind speed on the value of the minimum daily concentration of NO_2. The negative coefficients with the *num_direction* variable confirm the efficiency of the city's ventilation. The closer the wind direction is to the direction of the communication canyon axis, the lower is pollution concentration. Traffic flow, as the main nitrogen oxides source, is positive correlated with this pollution concentration. Increase of relative humidity (cloudy) is associated with decrease of insolation and deceleration of chemical reactions in the atmosphere. Therefore, an increase in humidity is conducive to reducing the concentration of NO_2. It should be noted that both models were characterized by the best efficiency for average daily values. With the extension of the averaging period, the effectiveness of linear models increases but the usefulness in the assessment of the impact of dynamic changes in the predictors' values decreases.

Verification

The models described above were verified on independent material containing values of both explained variables and predictors for the first half of 2018. A good correlation between the predicted values and the empirical values for the RF-mean with MAPE equal to 16.6% and 17.5% for RF and LR model respectively was obtained. For the RF-max model MAPE is equal 19.4% and 20.1% (RF an LR respectively).

The biggest errors were made when predicting minimum daily NO_2 concentration – MAPE=35.9% for RF and 36.4% for LR model. Both the above-described MAPE, Poisson's correlation coefficient (0.52, 0.67, 0.54; 0.46, 0.62, 0.49; min,

Table 3. Linear regression coefficients

Independent variable	Linear function coefficients		
	LR – log(min)	LR – average	LR – log(max)
Average wind speed	-0.0857	-0.0438	-0.0411
Wind direction (num)	-0.0572	-0.0039	-0.0223
Min traffic flow	0.00047	-0.00003*	-0.00009
Average traffic flow	0.000088	0.000119	0.000067
Relative humidity	-0.0039	-0.0028	-0.0038
Average air temperature	-0.0040	-0.0014	-0.0037*
Constant	1.5419	1.7340	2.1800

* coefficient statistically NOT significant for $\alpha=0.05$.

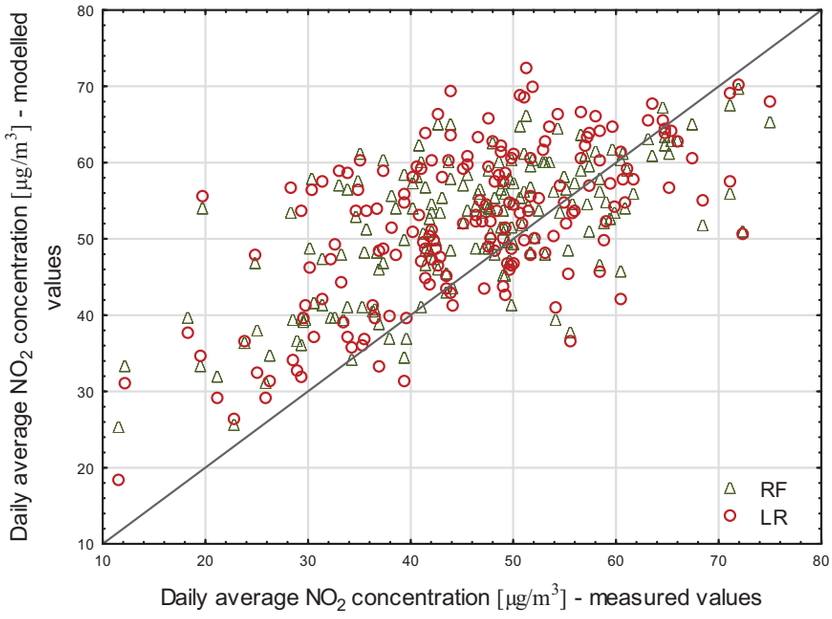

Fig. 4. Scatter plots for verification results – RF-average model with *y=x* line

average, max for RF and LR respectively) and other goodness of fit measures, indicate a significantly smaller difference in the accuracy of prediction of NO_2 concentration values for a test set. This reveals the unitarily of the linear approach, which allows to recognize the general structure of the phenomenon. Machine learning, although still giving a better fit for the test set, slightly loses the predominance. Point models do not make it possible to predict extreme values. Extremes can only be effectively predicted taking into account the time series as well as the values of environmental factors. The point-based model presented here allows for prediction of average daily values of NO_2 concentrations on the basis of average and extreme daily values for meteorological factors as well as the intensity of vehicle traffic with sufficient accuracy. The RF-mean model may therefore be used effectively to analyze and predict pollutant concentrations under *a priori* complex ambient conditions. Multiple linear model, although interpretable, is biased with a big error, bigger than RF. The next stage of the research will be the use of the model to assess the impact of the reduction of the number of vehicles travelling about the city (a restricted traffic zone) on NO_2 concentration.

Conclusions

The multiple linear model enables the interpretation of function coefficients in the explicit form at the expense of a significant deterioration in the quality of the match compared to the black--box model – Random Forest. The Random Forest method can be effectively used for prediction of minimum, average and maximum daily NO_2 concentration based on meteorological conditions and traffic flow information without using historical values. The described method can therefore be used to predict concentration values for various scenarios considered to reduce air pollution. The best modeling effects measured by the goodness of fit measures for both models were obtained for the daily average NO_2 concentration prediction and 0.59 for RF and LR respectively. The weakest fit was obtained (in both models) for the most diverse variable – daily minimal NO_2 concentration. Neither Linear Regression nor Random Forest method is suitable for predicting minimum daily values. The most important single predictor for NO_2 concentration is daily average wind speed. The applied method of transforming the wind direction from degrees to the numerical range taking into account the geographic location of the artery made it possible to assess the flow of wind direction to pollution evacuation. The negative coefficients of the linear model at the variable *num_direction* indicate effective ventilation of the city in the area of the examined intersection. Considering in general the environmental conditions, for daily NO_2 maximum and average concentration wind conditions and traffic flow are of equal importance. For minimal daily NO_2 concentration, wind conditions are more important. For maximal NO_2 concentration, relative humidity and air temperature play an important role.

References

Aldrin, M. & Haff, I.H. (2005). Generalized additive modelling of air pollution, traffic volume and meteorology. *Atmospheric Environment*, 39, 11, pp. 2145−2155, DOI:10.1016/j.atmosenv.2004.12.020.

Altenstedt, A. (1997). Modelling of the high to low NO_x transition using the IVL model – a contribution to the EUROTRACK sub--project LOOP. IVL Swedish Environmental Research Institute, Rapport B-1301, pp. 35

Archer, K.J. & Kimes, R.V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4), pp. 2249−2260, DOI:10.1016/j.csda.2007.08.015.

Battista, G. & de Lieto Vollaro, R. (2017). Correlation between air pollution and weather data in urban areas: Assessment of the city of Rome (Italy) as spatially and temporally independent regarding pollutants. *Atmospheric Environment* 165, pp. 240−247, DOI:10.1016/j.atmosenv.2017.06.050.

Breiman, L. (2001). Random Forests, *Machine Learning*, 45, 1, pp. 5−32, DOI:10.1023/A:1010933404324.

Czechowski, P., Badyda, A. & Majewski, G. (2013). Data mining system for air quality monitoring networks. *Archives of Environmental Protection*, vol. 39, 4, pp. 123−144.

Elangasinghe, M.A., Singhal, N., Dirks, K.N. & Salmond, J.A. (2014). Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmospheric Pollution Research*, *5*, 4, pp. 696−708, DOI: 10.5094/APR.2014.079.

Holnicki, P., Kałuszko, A., Nahorski, Z., Stankiewicz, K. & Trapp, W. (2017). Air quality modeling for Warsaw agglomeration, *Archives of Environmental Protection*, 43, 1, pp. 48−64. DOI: 10.1515/aep-2017-0005

Kamińska, J.A. (2019). A random forest partition model for predicting NO2 concentrations from traffic flow and meteorological conditions. *Science of the Total Environment*, 651, pp. 475−483, DOI: 10.1016/j.scitotenv.2018.09.196.

Kamińska, J.A. (2018a). The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: a case study in Wrocław, *Journal of Environment Management*, 217C, pp. 164−174, DOI: 10.1016/j.jenvman.2018.03.094.

Kamińska, J.A. (2018b). Residuals in the modelling of pollution concentration depending on meteorological conditions and traffic flow, employing decision trees. XLVIII Seminar of Applied Mathematics, *ITM Web Conf.* 23, 00016, DOI: 10.1051/itmconf/20182300016.

Kazak, J., Chalfen, M., Kamińska, J., Szewrański, S. & Świąder, M. (2018). Geo-Dynamic Decision Support System for Urban Traffic Management. In: Ivan I., Horák J., Inspektor T. (Eds.), Dynamics in GIscience. GIS OSTRAVA 2017. *Lecture Notes in Geoinformation and Cartography.* Springer, Cham, pp. 195−207, DOI: 10.1007/978-3-319-61297-3_14.

Laña, I., Del Ser, J., Pedró, A., Vélez, M. & Casanova-Mateo, C. (2016). The role of local urban traffic and meteorological conditions in air pollution: A data-based study in Madrid, Spain. *Atmospheric Environment*, 145, pp. 424−438, DOI: 10.1016/j.atmosenv.2016.09.052.

Larkin, A., Geddes, J.A., Martin, R.V., Xiao, Q., Liu, Y., Marshall, J.D., Brauer, M. & Hystad, P. (2017). Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environmental Science & Technology*, 51, pp. 6957−6964, DOI: 10.1021/acs.est.7b01148.

Nejadkoorki, F. & Baroutian, S. (2012). Forecasting extreme PM10 concentrations using artificial neural networks. *International Journal of Environmental Research, 6,* pp. 277−284, DOI: 10.22059/IJER.2011.493.

Ping, Shi J. & Harrison, R.M. (1997). Regression modelling of hourly NO_x and NO_2 concentration in urban air in London. *Atmospheric Environ*ment, 31, 24, pp. 4081−4094, DOI: 10.1016/S1352-2310(97)00282-3.

Sayegh. A., Tate, J.A. & Ropkins, K. (2016). Understanding how roadside concentrations of NOx are influenced by the

background levels, traffic density, and meteorological conditions using Boosted Regression Trees. *Atmospheric Environment* 127, pp. 163−175, https://doi.org/10.1016/j.atmosenv.2015.12.024.

Singh, K.P., Gupta, S. & Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environment*, 80, pp. 426−437, DOI: 10.1016/j.atmosenv.2013.08.023.

Siwek, K. & Osowski S. (2016). Data mining methods for prediction of air pollution. *International Journal of Applied Mathematics and Computer Science*, 26, 2, pp. 467−478, DOI: 10.1515/amcs-2016-0033.

Szewrański, Sz., Świąder, M., Kazak, J.K., Tokarczyk-Dorociak, K. & van Hoof, J. (2018). Socio-Environmental Vulnerability Mapping for Environmental and Flood Resilience Assessment: The Case of Ageing and Poverty in the City of Wrocław, Poland. *Society of Evironmental Toxicology and Chemistry*, 14(5), pp. 592−597, DOI: 10.1002/ieam.4077.

Zhang, Z., Zhang, X., Gong, D., Quan, W., Zhao, X., Ma, Z. & Kim, S-J. (2015). Evolution of Surface O_3 and $PM_{2.5}$ concentrations and their relationships with meteorological conditions over the last decade in Beijing, *Atmospheric Environment*, 108, pp. 67−75, DOI: 10.1016/j.atmosenv.2015.02.071.

Zhu, Y., Zhan, Y., Wang, B., Li, Z., Qui, Y. & Zhang, K. (2019). Spatiotemporally mapping of the relationship between NO_2 pollution and urbanization for a megacity in Southwest China during 2005–2016. *Chemosphere*, 220, pp. 155−162, DOI: 10.1016/j.chemosphere.2018.12.095.

Jawny i niejawny opis wpływu czynników na stężenie NO_2 w kanionie komunikacyjnym

Streszczenie: Celem pracy jest zbadanie możliwości prognozowania dziennego stężenia NO_2 za pomocą metody losowego lasu – RF i porównanie wyników z wielowymiarową regresją liniową (LR) w oparciu o ten sam zestaw danych. Ponadto zbadano wpływ zwiększenia interpretowalności modelu na jego dokładność.

W pracy przedstawiono dwie metody modelowania dziennych wartości minimalnych, średnich oraz maksymalnych stężeń NO_2 w aglomeracji miejskiej: wielowymiarowa regresja liniowa (LR) oraz losowy las (RF).

Wykazano, że metoda Lasu Losowego (Random Forest) może być skutecznie wykorzystywana do przewidywania dziennych wartości stężenia NO_2. Największą dokładność otrzymano dla przewidywania średnich wartości dziennych stężenia z R^2=0.69 oraz RMSE=7.47 μg/m^3. Kosztem otrzymania jawnej postaci funkcji w modeli liniowym (LR) jest znacząco niższa dokładność przewidywania wartości stężenia (R^2 od 0.32 do 0.57). Weryfikacja modeli na niezależnym materiale z pierwszej połowy 2018 roku potwierdziła poprawność modeli ze średnim błędem względnym dla średnich wartości dobowych stężeń równym 16.5% dla RF oraz 28% dla LR.

Największy wpływ na stężenia NO_2 w kanionie komunikacyjnym ma wiatr oraz natężenie ruchu. W modelowaniu maksymalnych wartości dobowych nabierają znaczenia temperatura powietrza oraz wilgotność względna powietrza. Przeważające zachodnie i północno-zachodnie wiatry we Wrocławiu skutecznie realizują koncepcję przewietrzania miasta w zakresie rozważanego skrzyżowania.