

Fires Impact on Air Quality: Extensive Analysis of Practical Indicators

Jan Stefan Bihałowicz *, Wioletta Rogula-Kozłowska , Adam Krasuski

Institute of Safety Engineering, The Main School of Fire Service, 01-629 Warsaw, Poland

ABSTRACT

The work aimed to build the best possible machine learning model predicting the concentration of selected air pollutants and evaluate model accuracy on the days with fires in a station vicinity. The underestimation of a pollutant concentration that coincides with fire would indicate its impact on air quality. Over 1353 thousand cases of fires in Poland and data from 410 air quality stations were analyzed (from 2012–2021). Models for prediction of NO₂, NO_x, PM₁₀, SO₂, and BTEX (benzene, toluene, ethylbenzene, m,p-xylene, and o-xylene) concentrations were built for the carefully selected station (rural background; Borówiec). The accuracy of models was checked as a function of distance from the fire and validated with the dispersion of plumes emitted during big landfill fires in 2018. The share of underpredicted concentrations of PM₁₀, benzene, toluene, and ethylbenzene on days when fire appeared in a range 30 km from the air quality station was significantly higher than the model performance. The concentrations of PM₁₀, SO₂, and BTEX, during plumes from landfill fires, were underestimated at least 30% of the duration of exposure. Hence, it was shown that it is very likely fires' contribution to air pollution can be evaluated using machine learning model misclassification.

Keywords: Fires, Air quality, Machine learning, Landfill fires, Aromatic hydrocarbons

1 INTRODUCTION

OPEN ACCESS

Received: April 12, 2022

Revised: June 23, 2022

Accepted: September 5, 2022

* Corresponding Author:

jbihalowicz@sgsp.edu.pl

Publisher:

Taiwan Association for Aerosol Research

ISSN: 1680-8584 print

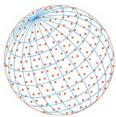
ISSN: 2071-1409 online

Copyright: The Author(s).

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

The consumption in economies is connected with waste production. In developing countries, as well as in some parts of developed countries, the waste is openly burned. The uncontrolled burning of waste possibly involves household waste, agricultural waste, or any type of waste stored at a burning landfill. Household waste burning covers backyard burning (campfire- or bonfire-like) (Bihałowicz *et al.*, 2021b; Gullett *et al.*, 2001; Hedman *et al.*, 2005; Mihai *et al.*, 2019) or barrel burning which is typical for North America (Lemieux, 1998; Lemieux *et al.*, 2004, 2000). Crop burning is either done before harvesting as a part of cultivation and harvesting (Arbex *et al.*, 2007; Gullett *et al.*, 2006; Rangel *et al.*, 2018) or after harvesting, as a quick way of disposal of residues (Chaudhary and Soni, 2020; Sarkar *et al.*, 2018; Yevich and Logan, 2003). There are also cases when waste is accidentally or intentionally burned in landfills (Aderemi and Otitoloju, 2012; Bihałowicz *et al.*, 2021a; Ettala *et al.*, 1996; Kumari *et al.*, 2019). During all these burnings greenhouse gases and air pollutants are emitted, including nitrogen oxides (NO, NO₂, NO_x) (Cheng *et al.*, 2009; EEA, 2016; Gadi *et al.*, 2011; Pansuk *et al.*, 2018; U.S. EPA, 1995), sulfur dioxide (SO₂) (Courtemanche and Levendis, 1998; EEA, 2016; Lönnemark and Blomqvist, 2006; Pansuk *et al.*, 2018), particulate matter (PM) (Bihałowicz *et al.*, 2021b; Hoffer *et al.*, 2020; Wiedinmyer *et al.*, 2014), PAHs (Downard *et al.*, 2015; Hoffer *et al.*, 2020; Ruokojärvi *et al.*, 1995), toxins (PCDD, Fs, dioxins) (Chen, 2004; Gullett *et al.*, 2006, 2001; Nadal *et al.*, 2016; Zhang *et al.*, 2019), VOCs (Blomqvist *et al.*, 2007; Lemieux *et al.*, 2004).

Waste fires are a significant problem in developing countries (Sharma *et al.*, 2022) as well as developed countries (Bihałowicz *et al.*, 2021a). In Poland, the problem of landfill fires was growing in the 2010's reaching 79 fires covering more than 301 m² or 1501 m³ in 2018 (KG PSP, 2020). The most generic pollutant emitted in every fire was PM₁₀, which was the most generic pollutant for



all types of waste (Bihałowicz *et al.*, 2021a), as presented in (Bihałowicz *et al.*, 2021c). The 79 fires a year means that landfill was burning more often than every 5 days. It is estimated that the landfill fires emitted an additional 5.3 ± 0.6 Gg of PM₁₀, i.e., 2.17% of national emission. The summary of the emission of air pollutants in Poland is yearly provided in the Informative Inventory Reports (KOBIZE, 2020). Although these reports divide emissions into over 120 sectors including even specialized industry contribution, the fires are not mentioned as one of the sources of air pollution.

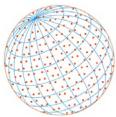
In this context, it is important to identify periods when the air quality is affected by fires. One of the ways of evaluating the contribution of sources of air pollution is the modeling of air quality, from the source-receptor approach (Weil *et al.*, 1992) through time-series modeling (Shahriar *et al.*, 2021) and machine learning (ML) models of concentration of PM_{2.5} (Bihałowicz, 2022; Shahriar *et al.*, 2021), the concentration of tomorrow PM₁₀ based on random forest (Jeong *et al.*, 2020), toluene, ethylbenzene, and xylene wet depositions for extreme gradient boosting (Stojić *et al.*, 2019) ending source apportionment evaluation and air quality management (Mach *et al.*, 2021; Thunis *et al.*, 2019). The model of concentration of PM₁₀ (Jeong *et al.*, 2020) basing on four meteorological variables (air temperature, relative humidity, wind speed, boundary layer height) had very good results except the coastal zone which implies that models should be developed locally. The model of PM_{2.5} (Shahriar *et al.*, 2021) used four meteorological variables: relative humidity, rainfall, temperature and wind speed, however, used also air quality data (SO₂, NO_x) to predict PM_{2.5} concentrations. The model of toluene, ethylbenzene, and xylene (Stojić *et al.*, 2019) was based on rain intensity, wind speed and direction, pressure, humidity, and temperature.

The quantitative contribution of fires to air pollution is difficult to evaluate, however, there is research in which its impact is quantified (Kollanus *et al.*, 2016; Lehtomäki *et al.*, 2018; Soares and Sofiev, 2014). It was shown that the problem of air pollution caused by landfill fires in Poland is significant on the sub-continental scale (Bihałowicz, 2021; Bihałowicz *et al.*, 2021c), while the impact of all fires on air quality was not evaluated. This study aims to identify whether there are periods when the impact of fires can be observed on air quality stations. The simple “looking for the peak” is not enough in the case of Poland, since the emission structure is complex (Bun *et al.*, 2019; Pyta *et al.*, 2020; Rogula-Kozłowska *et al.*, 2014, 2013a, 2019a). The concentrations in air are dependent mainly on two sources: transport and individual heating. The impact of the transport on air quality dominates in warm periods while in cold periods the impact of individual heating overwhelms other sources (Dzikuc *et al.*, 2017; Nidzgorska-Lencewicz and Czarnecka, 2015; Podstawczyńska and Chambers, 2019; Rogula-Kozłowska *et al.*, 2013b; Rozbicka *et al.*, 2020). Apart from sources of origin, the air quality depends also on the meteorological conditions which determine plume transport and dilution (Jeong and Park, 2013; Megaritis *et al.*, 2014; Oleniacz *et al.*, 2016). The investigation of the trends of concentrations should be done including the interference of other parameters. The point is to determine whether concentration at a given moment is typical for such set of conditions, even if it is not a “peak” concentration. In order to include meteorological variables, an ML approach based on Gradient Boosting Machine models (GBM) (Malohlava and Candel, 2020) in H2O Auto ML (H2O.ai, 2021; LeDell and Poirier, 2020) and Python (H2O.ai, 2020; Stetsenko, 2020) is applied. The concentration of modelled pollutant can be predicted using GBM. The impact of sources of pollution on concentration can be observed if the predictions of GBM do not match the measured concentration. The impact of the fire can be identified if the GBM prediction is lower than actual and there is a spatiotemporal coincidence with the fire occurrence.

2 MATERIALS AND METHODS

2.1 Selection of the Location

The current standard of fire inventory in Poland was fully introduced in early 2010's. Between the 1st April and 30th October 2021 database has 1,353,475 fire records, which will be analyzed in the work. The entry contains, among others, information about the size of fire, duration, and location. Prior to choice of modelling location, this database was analyzed using geographic information system (GIS) software QGIS (QGIS, 2021) together with the database of 410 air quality stations (GIOŚ, 2021). These stations are classified, according to (EC, 2008), into four types: urban, suburban, rural and rural background. The representativeness of each type of station are not

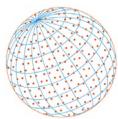


strictly defined and are few km², some tens of km², some hundreds of km², 1,000 to 10,000 km² respectively. Thus, for every station we created circular buffers, assuming radius (i.e., distance from station) 1.5 km, 5 km, 17 km, and 50 km respectively. In these buffers we analyzed total number and temporal distribution of fires for each station. Since the air pollution model using machine learning has to be developed at the air quality station, where the impact of some fires was investigated, we were examining the dispersion of the PM₁₀ from all big and very big waste fires in 2018 ([Bihałowicz et al., 2021c](#)) simulated using HYSPLIT ([Stein et al., 2015](#)) and. These fires are about 15% of all fires in Poland ([KG PSP, 2020](#)) but if we consider only the biggest fires in Poland, this share increases to more than 23%. Thus, the sample of landfill fires in Poland is representative of all fires and the dispersion from these fires can be used for selection air quality station. The map of the polygons with the increase of concentrations of PM₁₀ was included in GIS analysis. The aim was to find the stations which were exposed for a 1-h average increase in concentration by > 100 µg m⁻³ or > 10 µg m⁻³. We found out that concentrations only on 5 stations were in the range of impact of concentrations > 100 µg m⁻³, i.e., station PL0529A Sosnowiec located at Lubelska Street, PL0567A Katowice, Plebiscytowa/A4, PL0273A Skawina os. Ogrody, PL0184A Częstochowa Baczyńskiego Steet, PL0008A Katowice, Kossutha, and PL0552A Trzebinia, os. Związków Walki Młodych. The location of the stations with the map of dispersion of PM₁₀ from landfill fires is provided in Supplementary [Fig. S1](#). All of these stations are measuring air pollutants presented in Supplementary [Table S2](#). One could expect that PL0567A is the best station for evaluation of landfill fire impact on air quality since according to [Bihałowicz et al. \(2021c\)](#) two fires induced concentration increase by more than 100 µg m⁻³ and four fires by more than 10 µg m⁻³. Nonetheless, the station is located 10 m from the lanes of the European route E40, which at that point is a motorway, used for international transit and local transport in Silesia agglomeration, hence, it was not considered since the high impact of transport emission ([GIOŚ, 2021](#)). The remaining three stations are also urban stations ([GIOŚ, 2021](#)) according to the European Commission Directive on ambient air quality ([EC, 2008](#)), which are representative for a few km². Focusing on stations where the increase of concentration caused by landfill fires (sample representing around a quarter of all fires in 2018) was higher than 10 µg m⁻³ we found out that there were no stations affected five times or more, 9 stations were in the range of impact four times each, 20 stations three times, 61 stations two times, and 87 stations once. Among 9 stations impacted four times, all the stations were urban except the PL0573A Borówiec, Drapałka Street, which is a rural station, representative for some hundreds of km² ([EC, 2008; GIOŚ, 2021](#)). Apart from the concentrations of NO_x, NO₂, PM₁₀, SO₂ (further in work called basic air pollutants), there are measured at the station also additional pollutants, i.e., benzene, toluene, ethylbenzene, m,p-xylene, and o-xylene (BTEX). The basic pollutants are emitted in many situations like individual heating, transport, industrial production while emission of BTEX is not so common, and as possible sources of BTEX in air fires of waste are mentioned ([U.S. EPA, 1998, 1993a, 1993b](#)). Therefore, as the location of test the air quality station in Borówiec was chosen.

2.2 Data

The data measured at Borówiec station are available in the database of the Chief Inspectorate for Environmental Protection in Poland (GIOŚ) ([GIOŚ, 2021](#)). The BTEX (except ethylbenzene and m,p-xylene), are measured at Borówiec since 1st January 2016; for this reason in the work five years, from 1st Jan. 2016 to 31st Dec. 2020, are analyzed. The air pollutants concentrations were measured with a 1-h averaging period hence for the purpose of the study, daily average concentrations were calculated. The daily averages used in this work were calculated as an average of values provided by GIOŚ only if the daily time coverage was > 75%, otherwise, the value was left as NA (not available). The summary of parameters measured at Borówiec station with annual time coverages data is presented in [Table 1](#).

The meteorological data in Poland are gathered and analyzed by the Institute of Meteorology and Water Management – National Research Institute (IMGW-PIB) ([IMGW-PIB, 2021](#)). The system meteorological observations are based on 3 types of stations: synoptic, climate and precipitation. The highest number of parameters is measured on the synoptic station. The closest to Borówiec synoptic station is Poznań synoptic station, located 20 km northwest. In the work 37 daily parameters observed between 1st Jan. 2016 and 31st Dec. 2020 at Poznań station were used, whose descriptions and units are provided in Supplementary [Table S1](#).

**Table 1.** The air quality parameters measured at Borówiec station which were used in work with annual time coverage of data.

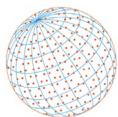
Target variable	Abbreviation	Unit	Time coverage (%)				
			2016	2017	2018	2019	2020
Nitrogen dioxide	NO ₂		99	98	89	99	100
Sum of nitrogen oxides	NO _x		100	98	97	100	100
Particulate matter $\leq 10 \mu\text{m}$	PM ₁₀		91	97	95	99	98
Sulfur dioxide	SO ₂		99	98	100	100	0
Benzene	C ₆ H ₆	$\mu\text{g m}^{-3}$	99	92	96	82	96
Toluene	C ₇ H ₈		0	92	91	82	96
Ethylbenzene	C ₈ H ₁₀		0	92	91	82	96
m,p-xylene	m,p-C ₈ H ₁₀		0	92	91	82	96
o-xylene	o-C ₈ H ₁₀		91	92	91	82	96

The data about the fires in Poland were taken from the National Headquarters of the State Fire Service of Poland which provides evidence of all incidents (SWD-EWID) in which State Fire Service units take part (KG PSP, 2019). A more detailed description of data structure and collected parameters is provided in (Bihałowicz *et al.*, 2021a). The database, as in case of air quality, was also limited to period from 1st Jan. 2016 to 31st Dec. 2020. Moreover, only data about fires bigger than 71 m² or 351 m³ were used in the work (according to SWD-EWID nomenclature, small fires were excluded). As a part of the validation of the models, the quantitative data about the spatial dispersion of PM₁₀ emitted in the biggest fires of waste reported in SWD-EWID in 2018 were used (Bihałowicz *et al.*, 2021c).

2.3 Machine Learning

The air pollution assessment can be treated as a classification problem. The concentrations, target variables in Table 1, are continuous, hence data points were divided into three classes. Since the concentrations are site-specific, we decided to divide days according to the local distribution of concentrations. We used boxplot-like division, i.e., the day was assigned to Class A if concentration on this day smaller than the first quartile of the whole range of data, $c_A < Q_1$. Similarly, Class B assignment was made if the concentration was between first and third quartile, $Q_1 \leq c_B < Q_3$, and Class C was made if concentration was higher than the third quartile $c_C \geq Q_3$. This procedure was repeated for each target variable separately (with the separate determination of quartiles for each pollutant). If the concentration of the pollutant was not measured on a given day, the concentration class is left as NA. The ML models are created for each target variable separately, i.e., the classification model of NO_x concentration has only 37 meteorological parameters as inputs, none of the air quality variables are input variables.

The choice of ML package was preceded with tests and evaluating of possibilities of modeling of air quality using Orange (Bihałowicz, 2022; Demšar *et al.*, 2013), scikit-learn (Feurer *et al.*, 2020; Pedregosa *et al.*, 2011), MLJAR (Płońska and Płoński, 2021). As the ML suite, an H2O software package (H2O.ai, 2021) was chosen. The H2O provides the automated machine learning models building (AutoML) which was used in this work. Due to the results of the test with all the above-mentioned software, the gradient boosting model for multinomial classification was chosen. The tree based gradient boosting was used in this work (H2O.ai, 2022). The principle of gradient boosting algorithm is to improve classification tree by creating new trees which classify better at the point where the initial was not performing well (Hastie *et al.*, 2009). The metrics assessing the quality of the models was the logarithm of loss function. In case of all models the boosting had learning parameter (shrinkage parameter) 0.1. To determine best performing model parameters, trees of depth from 3 to 17 were analyzed. The boosting algorithm creates new trees basing on the previously created trees, the number of trees was limited to 10000 but the new trees were not created in the boosting algorithm if the improvement of the next tree was smaller than reciprocal of square root of number of non-empty values. Models created for each target variable used multinomial distributions. The models in this work were validated using 5-fold cross-validation. Gradient boosting models are successful in modeling atmosphere-related phenomena (Cheng *et al.*, 2021; Liu *et al.*, 2020, 2021).



3 RESULTS AND DISCUSSION

3.1 General Description of the Results

The data about the meteorological conditions were used to build ML classification models of the concentration of air pollutants. The AutoML was looking for the best model according to the logarithmic loss value (logloss) (H2O.ai, 2021). The higher value of the loss is, and hence its logarithm, the worse is model. The procedure was conducted separately for all investigated pollutants. All models share the same variables—37 meteorological parameters. The values of parameters indicating the performance of best-logloss models are provided in **Table 2**. The model with the lowest mean-per-class error was PM₁₀, close to 0.25, the logloss and RMSE was smallest for the classification of C₆H₆. The model with poorest values of all three parameters was the classification of concentration of o-xylene.

In **Fig. 1** the heatmap of the relative importance of variables, normalized to the most important variable per each model is provided. The number of variables is limited so that the four most

Table 2. The mean per class error, logarithmic loss value (logloss), and root mean square error (RMSE) for the best performing models of classification of pollutant concentration.

	Mean per class error	logloss	RMSE
C ₆ H ₆	0.4428	0.6258	0.4430
C ₇ H ₈	0.5783	0.9064	0.5553
C ₈ H ₁₀	0.5990	0.9082	0.5638
m,p-C ₈ H ₁₀	0.5998	0.9388	0.5742
o-C ₈ H ₁₀	0.6460	0.9875	0.5989
NO ₂	0.2865	0.8041	0.5187
NO _x	0.2953	0.8272	0.5299
PM ₁₀	0.2552	0.7321	0.4893
SO ₂	0.5574	0.8680	0.5438

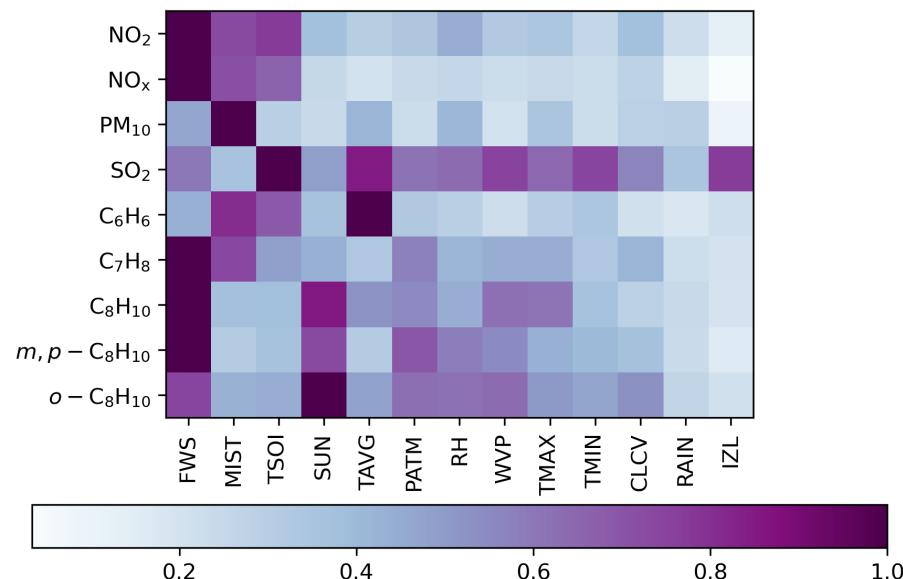
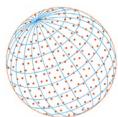


Fig. 1. The heatmap of variable importance in best performing GBM models of concentration classification. The values are normalized for each model (row) and variables are in columns. The names of variables are FWS – Average wind speed, MIST – mist duration, TSOI – minimal ground temperature, SUN – sun duration, TAVG – average temperature, PATM – atmospheric pressure, RH – relative humidity, WVP – water vapor pressure, TMAX – maximum temperature, TMIN – minimum temperature, CLCV – cloud cover, RAIN – precipitation, IZL – isotherm low, according to Supplementary **Table S1**.



important variables are shown for each model. The heatmap including all variables is provided as Supplementary Fig. S1. Since the concentration of pollutants is related to wind, and more general, plume dispersion, one could anticipate that the highest importance has FWS (average wind speed). In fact, for the NO₂, NO_x, C₇H₈, C₈H₁₀, and m,p-C₈H₁₀ the most important variable was FWS (Clarkson *et al.*, 1996; Grundström *et al.*, 2015; Kelessis *et al.*, 2006; Kourtidis *et al.*, 2002; Zhang *et al.*, 2018). The FWS was also important variable for the o-C₈H₁₀, however, not so important as SUN. In the case of PM₁₀ most important was MIST, and for C₆H₆ TAVG. All these modes have few most important variables while for SO₂ many variables are relatively important, the most important is the temperature at ground level TSOI. The models of the C₈H₁₀, m,p-C₈H₁₀, and o-C₈H₁₀ have similar most important variables while PM₁₀, SO₂, and C₆H₆ are different from both BTEX and NO models.

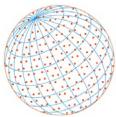
3.2 Basic Air Pollutants

The best performing GBM models of classification of concentration were best in the classification of PM₁₀, then NO₂ and NO_x, and significantly worse SO₂. The performance of these models can be expressed by the confusion matrix in Table 3. The concentrations are in order A < B < C, as it was described in Section 2.3. The error rate of the PM₁₀ model is only 11/1769 = 0.63% while for the remaining models it is of the order of a few percent with the highest rate for the NO_x model 141/1808 = 7.8%. The precision of the model in each class is the number of points correctly assigned over the total number of points assigned to this class. The recall of the model in each class is one minus error rate.

In all models of NO₂, NO_x, PM₁₀ similar set of variables is in the four most important for each model. For NO₂ they are FWS, TSOI, MIST, and RH, for NO_x FWS, MIST, TSOI and CLCV, for PM₁₀ MIST, FWS, TAVG and RH while for the model of SO₂ they are TSOI, TAVG, IZL and WVP. The partial dependence plots of models' classes on these variables are provided in Supplementary material. In the case of FWS (for NO₂ Supplementary Fig. S3(a), for NO_x Supplementary Fig. S4(a)) the highest mean response in the class of highest concentrations C is observed for FWS < 2 m s⁻¹, i.e., days without wind, the highest response in Class B (intermediate concentrations) is observed for FWS around 4 m s⁻¹ while the Class A (lowest concentrations) has highest mean responses for FWS > 6 m s⁻¹. The same pattern of dependence was observed for the PM₁₀ model Supplementary Fig. S4(b), however, the slopes of mean response plots are lower and interval endpoints are not sharp. The reported dependences of NO_x and PM₁₀ concentrations on wind speed also provide similar conclusions (Grundström *et al.*, 2015; Jones *et al.*, 2010). The second variable is TSOI (Supplementary Fig. S3(b) for NO₂, Supplementary Fig. S4(c) for NO_x, Supplementary Fig. S6(a) for SO₂). The dependence of SO₂ model has slightly different character, since intervals with

Table 3. Confusion matrix of models with a classification error rate, recall, and precision for each class, of NO₂ model, top left, of NO_x model, top right, of PM₁₀ model, bottom left, of SO₂ model, bottom right.

		NO ₂					NO _x						
		Predicted class			Error Rate	Recall	Predicted class			Error Rate	Recall		
		A	B	C			A	B	C				
Actual	A	401	38	3	41/442	0.91	Actual	A	387	62	2	64/451	0.86
	B	5	871	8	13/884	0.99		B	15	873	17	32/905	0.96
	C	1	21	421	22/443	0.95		C	2	43	407	45/452	0.90
	Total	407	930	432	76/1,769			Total	404	978	426	141/1,808	
	Precision	0.99	0.94	0.97				Precision	0.96	0.89	0.96		
PM ₁₀													
		Predicted class			Error Rate	Recall	Predicted class			Error Rate	Recall		
		A	B	C			A	B	C				
Actual	A	436	2	0	2/438	1.0	Actual	A	327	33	0	33/360	0.91
	B	1	873	1	2/875	1.0		B	4	723	1	5/728	0.99
	C	0	7	431	7/438	0.98		C	1	21	341	22/363	0.94
	Total	437	882	432	11/1,751			Total	332	777	342	60/1,451	
	Precision	1.0	0.99	1.0				Precision	0.98	0.93	1.0		
SO ₂													
		Predicted class			Error Rate	Recall	Predicted class			Error Rate	Recall		
		A	B	C			A	B	C				
Actual	A	436	2	0	2/438	1.0	Actual	A	327	33	0	33/360	0.91
	B	1	873	1	2/875	1.0		B	4	723	1	5/728	0.99
	C	0	7	431	7/438	0.98		C	1	21	341	22/363	0.94
	Total	437	882	432	11/1,751			Total	332	777	342	60/1,451	
	Precision	1.0	0.99	1.0				Precision	0.98	0.93	1.0		



highest responses for different classes are overlapping what can be caused by high relative importance of few first variables. The highest mean response for low concentrations (*A*) can be observed for highest temperatures, over 15°C (for SO₂ 5°C), for intermediate concentrations (*B*) for temperatures between -5°C and 15°C (SO₂: < 15°C) and for class of high concentrations (*C*) for TSOI < -5°C (SO₂ < 5°C). This can be justified by the increase of demand on individual heating during the cold days. The residential buildings are located on the ground hence the most crucial for the thermal comfort of people living there is not the average air temperature TAVG, but the lowest ground temperature TSOI. The TAVG is not representative for individual heating since the daily temperature can affect the average, while the TSOI is the daily lowest reading of temperature at 5 cm above ground. The data about the fuel consumption for individual heating could be evaluated basing on electricity household consumption and natural gas consumption. The electricity consumption is affected by the numerous of electric devices which are also powered from this source, while the natural gas household consumption is only affected by gas cookers whose consumption is much lower than consumption of boilers (Aras and Aras, 2004; Błaszcak *et al.*, 2016; Jedlikowski and Englart, 2018; Majewski *et al.*, 2018). There is a negative correlation between ambient temperature and natural gas consumption (Hribar *et al.*, 2019; Sabo *et al.*, 2011; Szoplik, 2015) and the similar dependence of consumption of temperature is observed for solid fuels as coal (Li *et al.*, 2016). Thus, the heating induced emissions to the atmosphere are also negatively correlated with temperature. The dependence on time of mist is not so clear for NO₂ and NO_x (Supplementary Fig. S3(c) and Fig. S4(b), respectively), the highest responses for Classes *A* and *B* are for MIST < 7.5 h, while the Class *C* has highest mean response for MIST > 7.5 h. It can be interpreted that for longer times of mist the Class *C* has higher responses than for shorter, while both Classes, *A* and *B*, are sensitive for shorter times of mist. In case of PM₁₀ the variability of mean response is much higher, Supplementary Fig. S5(a), as it is the most important variable for best performing model of basic air pollutants. As MIST → 24 h the mean response for Class *C* tends to 0.8, while mean response for Classes *B* and *A* tends to their maximum as MIST → 0 h, however, at MIST < 10 h for Class *B*, and MIST < 5 h for Class *A*. This dependence shows that meteorological phenomena classified at station as mist is in fact drop in the visibility (Li *et al.*, 2019; Tsai *et al.*, 2007) which is proven to be dependent on the concentration of PM₁₀ (Huang *et al.*, 2016; Majewski *et al.*, 2021, 2015). The partial dependence of models on RH, CLCV shows that there are very low changes in mean response curve for each class. The curves are rather constant, without slopes and extrema, therefore impact of related variables cannot be explained in the same way as it was in case of MIST.

To evaluate the impact of fires on the measurement on station Borówiec we compared the actual and predicted by model class of concentration at the station during the fire reported in SWD-EWID and plotted it against distance from the fire. As the resulting plot of the share of under- and overestimated concentrations on days with fires as a function of distance from the fire. The distance between fire and Borówiec was calculated using QGIS. The plot is shown in Fig. 2. The share of underestimated and overestimated reaches the model performance at the distance around 25 km for NO₂, 18 km for NO_x, 70 km for PM₁₀, and 40 km for SO₂. For the range where the station can be representative models for all pollutants except PM₁₀ under- and overestimates the concentration. It may suggest that the concentration of NO₂, NO_x, SO₂ at this site is affected also by other factors which influence the concentrations stronger than the emission from fires. Such sources can be emissions from transport and individual heating. Both, NO and NO₂, are produced during the combustion, and one of the primary sources of these oxides is the combustion in car engines (Klejnowski *et al.*, 2013; Rogula-Kozłowska *et al.*, 2019b). There were no cases of overestimation of PM₁₀ concentration on days with fires in the neighborhood of air quality station. It may suggest that the model of concentration class of PM₁₀, which is according to its training confusion matrix the best, properly found the influence of the fire—underpredicted concentration. Hence, the aim is to find are there any different air pollutants that can be treated.

3.3 BTEX

The GBM model of concentration of benzene at the air quality station in Borówiec is well-performing. The confusion matrix shown in Table 4 top left indicates that model error rate of the model is only 8/1422 = 0.56%. The precision and recall of the model are also provided in Table 4. The classification of concentration mostly depends on four variables TAVG, MIST, TSOI and FWS.

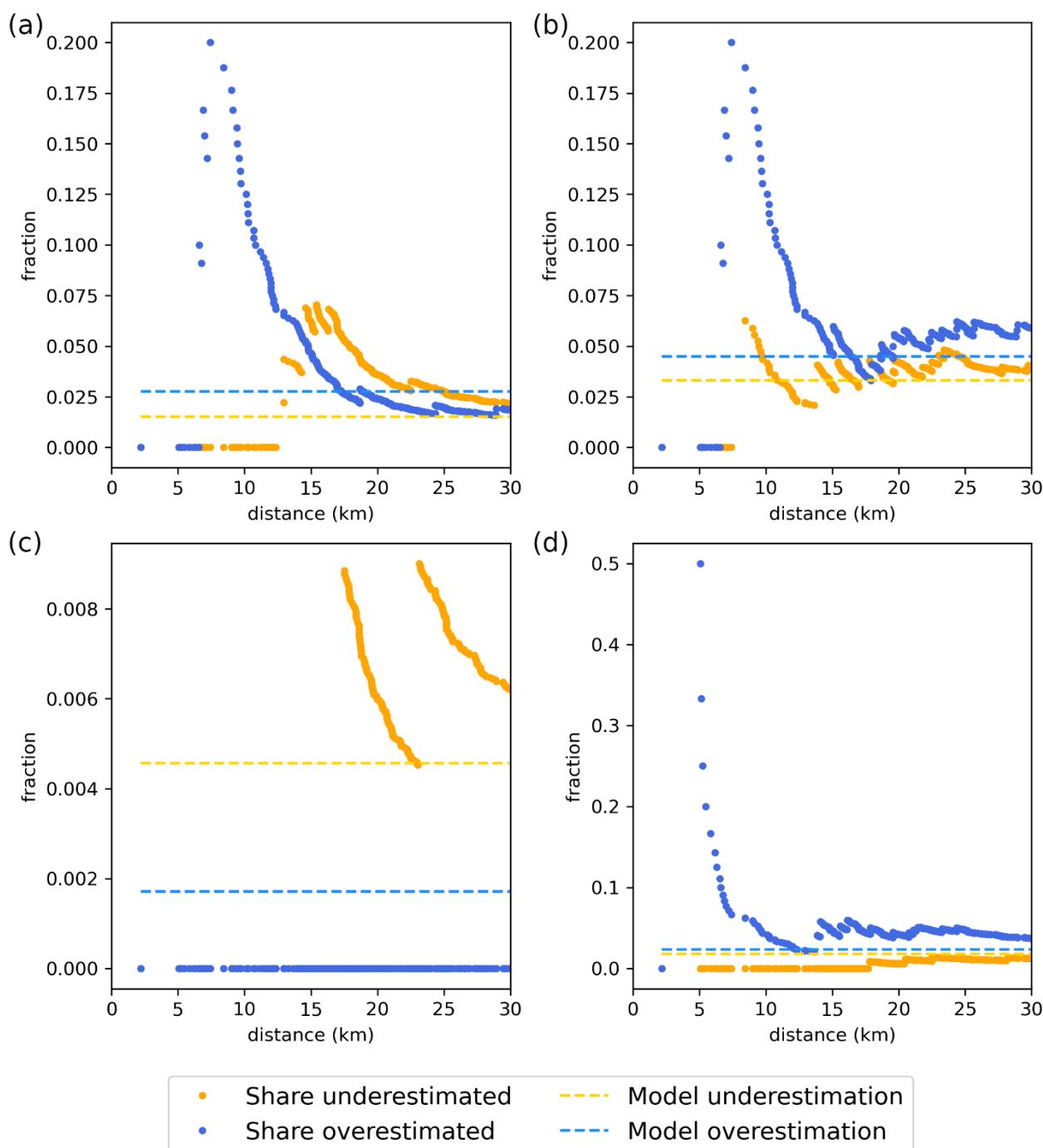
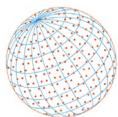


Fig. 2. The share of under- and overestimated concentrations at days with fire as a function of distance from the Borówiec air quality station, top left (a) NO₂, top right (b) NO_x, bottom left (c) PM₁₀, bottom right (d) SO₂.

The main difference of benzene model to other models is the fact that the most important variable is TAVG. The dependence on value of TAVG for particular classes is not as clear as dependence of TSOI, which exhibits the same pattern related to individual heating as for basic air pollutants. The mean response for class of highest concentrations C is highest for low TAVG while the pattern of dependence for class of intermediate concentrations B and low concentration A is varying at high TAVG values and the intervals where mean response is highest. In general, the partial dependence (Supplementary Fig. S7) on MIST, TSOI and FWS presents similar patterns for Classes A, B, C as for basic air pollutants, only the values and intervals where the mean response is highest are slightly shifted.

**Table 4.** Confusion matrix of the model of C₆H₆ (top), C₇H₈ (middle), C₈H₁₀ (bottom) with classification error rate, recall, and precision for each class.

Benzene C ₆ H ₆									Toluene C ₇ H ₈											
		Predicted class			Error Rate	Recall			Predicted class			Error Rate	Recall							
		A	B	C					A	B	C									
Actual	A	350	11	0	11/361	0.97	Actual	A	290	37	2	39/329	0.88	Actual	A	290	37	2	39/329	0.88
	B	19	659	11	30/689	0.96		B	1	656	1	2/658	1.0		B	1	656	1	2/658	1.0
	C	0	23	349	23/372	0.94		C	0	27	303	27/330	0.92		C	0	27	303	27/330	0.92
	Total	369	693	360	64/1,422			Total	291	720	306	68/1,317			Total	291	720	306	68/1,317	
	Precision	0.97	0.96	0.94				Precision	0.96	0.89	0.96				Precision	0.96	0.89	0.96		
Ethylbenzene C ₈ H ₁₀																				
		Predicted class			Error Rate	Recall			Predicted class			Error Rate	Recall							
		A	B	C					A	B	C									
Actual	A	308	18	3	21/329	0.94	Actual	A	308	18	3	21/329	0.94	Actual	A	308	18	3	21/329	0.94
	B	3	654	1	4/658	0.99		B	3	654	1	4/658	0.99		B	3	654	1	4/658	0.99
	C	2	24	304	26/330	0.92		C	2	24	304	26/330	0.92		C	2	24	304	26/330	0.92
	Total	313	696	308	51/1,317			Total	313	696	308	51/1,317			Total	313	696	308	51/1,317	
	Precision	0.98	0.94	0.99				Precision	0.98	0.94	0.99				Precision	0.98	0.94	0.99		

The point which makes model benzene different from models of NO_x, NO₂, and SO₂ is the dependence on fire occurrence. The share of under- and overpredicted concentrations as a distance from fire occurrence is shown in Fig. 3(a). There are no overpredicted concentrations by the model for days with fires, which appeared even 60 km from the air quality station. There are cases of underprediction of concentration, which are much often than model performance underestimate rate. It may be evidence that fires that are in the representative area of air quality stations have an impact on the concentration of benzene. The analysis of all fires in the range where station representativeness is decreasing shows that up to 160 km from the station is 22 times more probable that in the day with fire the concentration is underestimated than overestimated Supplementary Fig. S12. The confusion matrix suggests that this ratio should be more than 3 times lower. We can conclude that fires cause errors in modeling concentrations of benzene at Borówiec air quality station. It can be the reflection of possible sources of benzene in the atmosphere (Rogula-Kozłowska *et al.*, 2020; U.S. EPA, 1998) are fires of tire shreds, biomass burning, agricultural film fires, and forest wildfires.

The toluene model confusion matrix (Table 4, top right) shows that this model is less precise, especially in classification of Class B than model of benzene. Similarly, as in other models the toluene model's four most important parameters include FWS, MIST, TSOI and this time PATM (Supplementary Fig. S8). The patterns of FWS and MIST are similar to already discussed while pattern of TSOI is different than in all previously discussed cases since although the general pattern is roughly similar, the variability of the mean response is much lower and allow only for rough identification of intervals with the highest mean response for each class. The atmospheric pressure is a parameter related to other meteorological phenomena, like the passing of weather fronts, change of temperature, and relative humidity. The partial dependence on PATM shows complex relation shown in Supplementary Fig. S8.

One of the potential sources of emission toluene in the atmosphere is firefighting at solid waste disposal sites (Łukawski, 2019; Rogula-Kozłowska *et al.*, 2020; U.S. EPA, 1993a) hence we expected that also fires, in general, affect the concentration of toluene. The share of under- and overestimated concentrations of toluene are shown in Fig. 3(b). The model underestimates the concentration if the fire appears in range to approx. 16–18 km from the air quality station. Additionally, the overestimation does not take place if fires are up to 12 km. The performance of the model suggests that the number of overestimations should exceed the number of underestimation while in the range of almost 30 km it is the opposite. The analysis of fires up to 160 km from Borówiec shows that for days with fires up to 130 km the ratio of underestimated to overestimated is greater than one, especially up to 60 km when it is significantly greater than one (Supplementary Fig. S13).

The performance of the ethylbenzene model was slightly worse than benzene and toluene

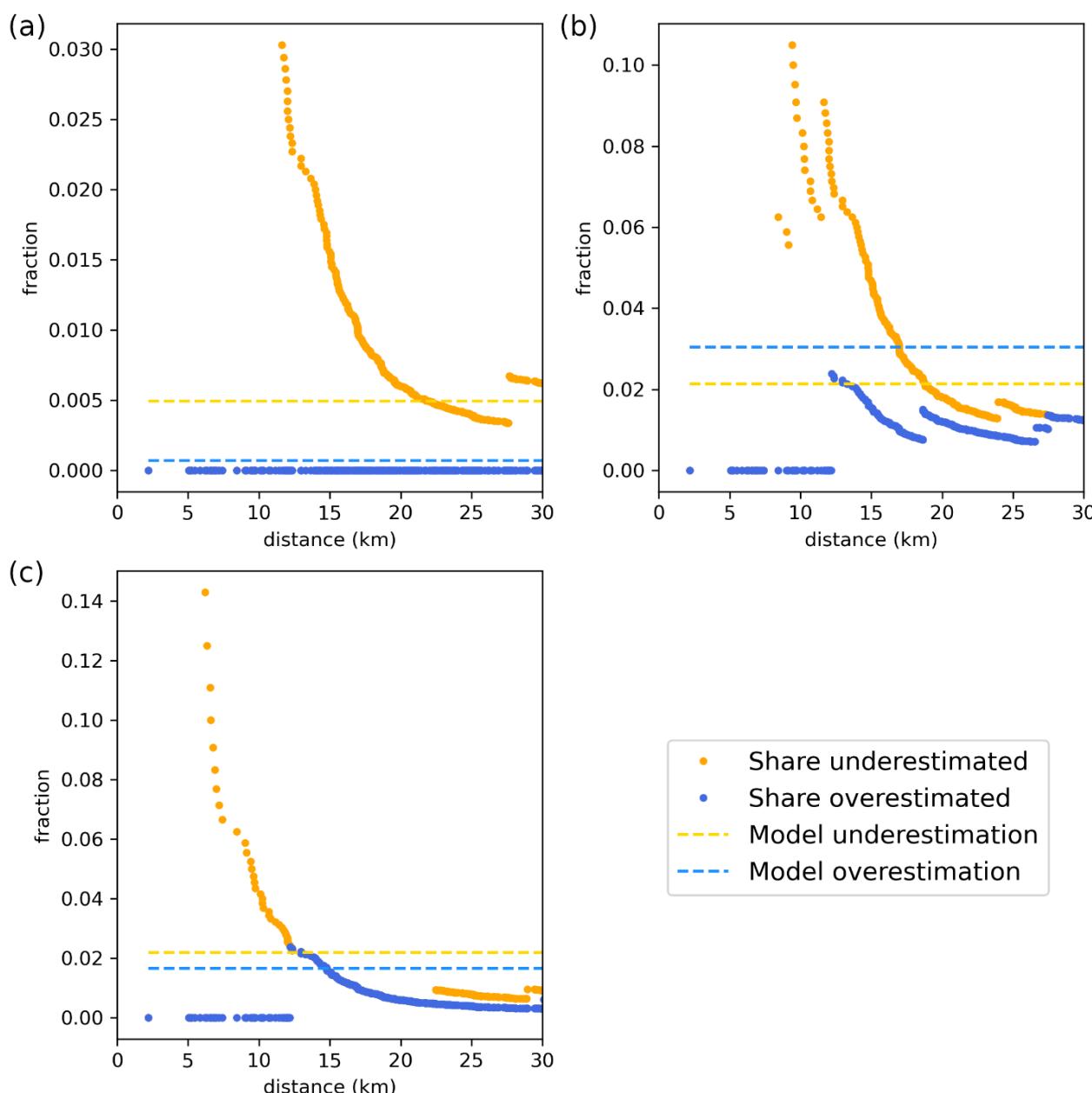
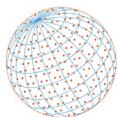


Fig. 3. The share of under- and overestimated concentrations at days with fire as a function of distance from the Borówiec air quality station, top left (a) benzene, top (b) toluene, middle (c) ethylbenzene, bottom.

according to Table 1, however, the values presented in Table 4 bottom, the confusion matrix, precision, and recall show that performance is good. The model's confusion matrix shows that it tends to underestimate since the sum above the diagonal of the confusion matrix is lower than below the diagonal. The four most important variables for this model are FWS, SUN, WVP, and TMAX (Supplementary Fig. S9). This is the first discussed model where the dependence on the time of the sun plays important role in the classification of concentration. The ethylbenzene is photodegrading in the air (Banton, 2014) which explains dependence on SUN. The SUN also is one of the most important variables for models of m,p-xylene and o-xylene. Xylene is also subject of photolysis into less toxic compounds (Kandyala *et al.*, 2010). The model of ethylbenzene is underpredicting the concentration for fires up to 12 km from Borówiec air quality station Fig. 3(b). Among sources of ethylbenzene in atmospheric air are incomplete combustion and burning of waste (Banton, 2014). Hence, it is justified that changes in air quality induced by fire are the errors in the prediction of ethylbenzene concentration class.



The models of xylenes concentration were least performing among models of BTEX. Moreover, there were no cases of underestimation of concentration for days with fires in the range of representativeness of Borówiec station. The possible explanation is that set of meteorological variables is not an appropriate set for predicting the concentration of xylene and other factors, which are not included in the models are governing the concentration of xylene.

3.4 Concentration during Landfill Fires - Validation

We validated the predictions of the models of PM_{10} , C_6H_6 , C_7H_8 , and C_8H_{10} based on the calculated dispersion of PM_{10} from landfill fires in Poland in 2018 (Bihałowicz *et al.*, 2021c). The Borówiec air quality station was in the range of impact of plume increasing 1-h average concentration of PM_{10} by $> 10 \mu\text{g m}^{-3}$ four times. It was on the 24 to 28 April, 25 to 31 May, 27 to 30 May, and 16 to 18 December 2018. The regions, where 1-h average concentration increased by more than $10 \mu\text{g m}^{-3}$ and the location of the Borówiec station are shown in Fig. 4. In the validation we wanted to include PM_{10} , benzene, toluene, and ethylbenzene, nevertheless, during the landfill fire in Zgierz on 25–31 May, the Central Contamination Analysis Center of Polish Armed Forces found in benzene, toluene, styrene, methylstyrene, sulfur dioxide, ethylbenzene, and o-xylene (Cygańczuk *et al.*, 2020). For this reason, we included in the analysis also the validation of SO_2 and xylene concentration. The rates of the number of days with underpredicted concentrations in days when plume from the given fire was at the Borówiec are provided in Fig. 4. The fire on 24–28 April was a fire of general waste on landfill. The plume with increase of 1-h average concentration of PM_{10} from this fire was 5 days at the Borówiec station (Bihałowicz *et al.*, 2021a, 2021c). The concentrations

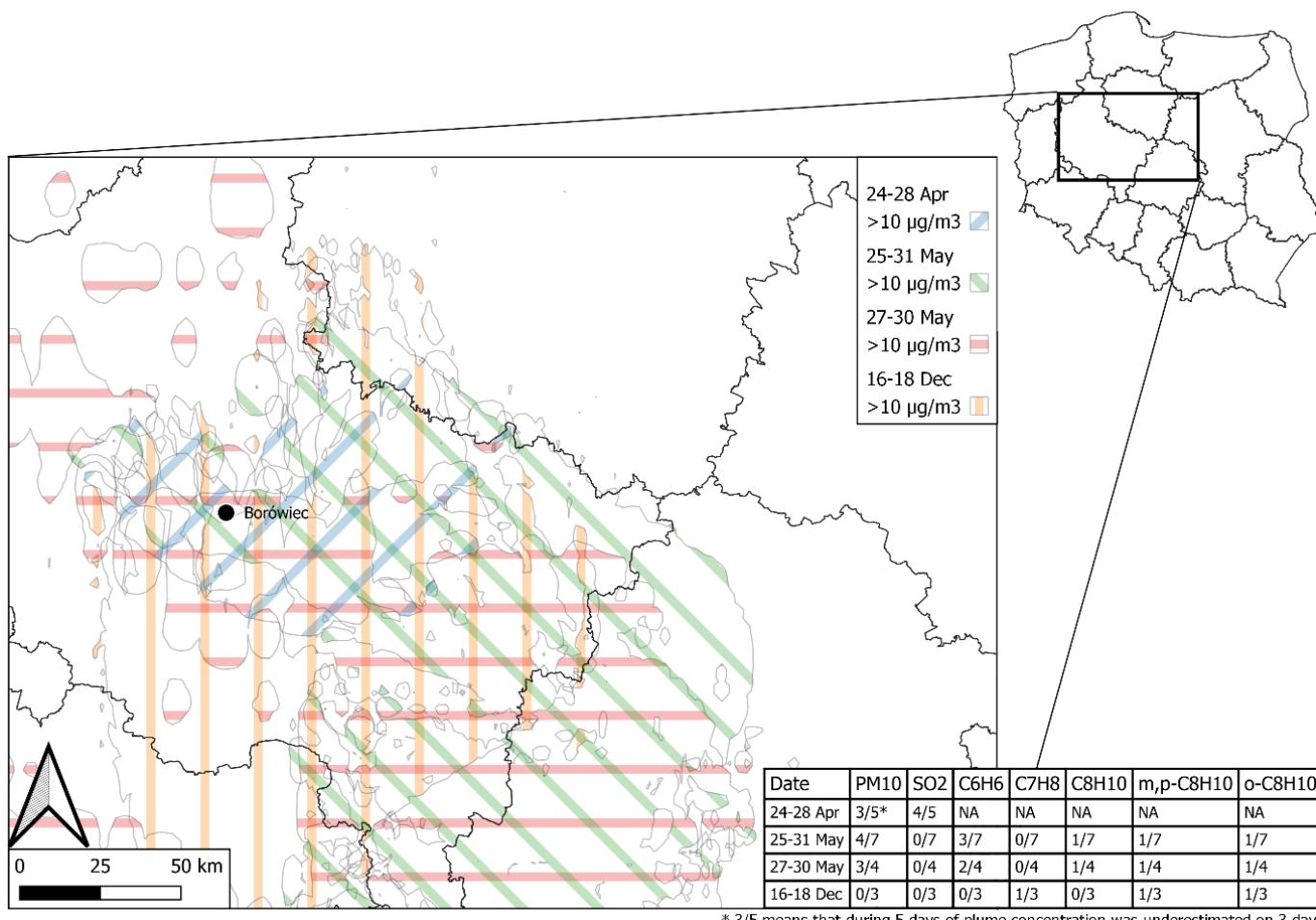
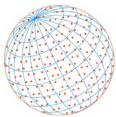


Fig. 4. The location of Borówiec air quality station and plumes of four landfill fires which caused increase in 1-h average concentration of PM_{10} by more than $10 \mu\text{g m}^{-3}$ with the underprediction rate on days when plumes from landfill fires were at the Borówiec air quality station.



class of PM₁₀ and SO₂ were underestimated. The fire on 25–31 May induced underestimation of PM₁₀, C₆H₆, C₈H₁₀, m,p- and o-C₈H₁₀, however, the concentration class of PM₁₀ was once during this fire overestimated. The third fire was on 27–30 May, the subset of previous days, thus the same pollutants' misclassification was observed. The last landfill fire which impacted Borówiec was in December when the concentrations of pollutants are high due to individual heating demand (Błaszczał *et al.*, 2016; Rogula-Kozłowska *et al.*, 2014, 2013a, 2013b). It can be a reason that only concentrations of C₇H₈ and m,p-, and o-C₈H₁₀ were only once underestimated, moreover, the concentration of m,p-C₈H₁₀ was once overestimated. It is important that the concentrations of xylenes are underpredicted during every landfill fire. It is plausible that landfill fires are characterized by the underestimation of both m,p-xylene and o-xylene concentration. All the rates presented in Fig. 4 significantly exceed the underestimation rates which would be caused by the performance of the models. The landfill fires are the reason for underestimation of concentration class on these days.

4 CONCLUSIONS

The impact of fires on air quality can be evaluated with machine learning models. The GBM classification models based on 37 meteorological variables works well for NO₂, NO_x, PM₁₀, and BTEX. The introduction of more variables than in previous studies of ML modeling of air quality and their appearance in four most important variables shows that use of synoptic station data is important for air quality modeling. The models of air quality, although the precision of models was 0.89 for worst model, and in general was around 0.95, reaching even 1.0, have some misclassification. The analysis of the distance and time of underestimation with fire incident data (spatiotemporal coincidence) showed that fires in the range of representativeness of air quality stations cause underestimations. For model of PM₁₀ and for model of C₆H₆ there were no overestimations of concentration class for fires in range up to 30 km from air quality station, while for C₇H₈ and C₈H₁₀, models were underestimating for fires up to 12 km from station. The landfill fires, which constituted a quarter of very big fires in Poland in 2018, emitted plumes which could be measured four times at Borówiec air quality stations. The concentrations of PM₁₀ and SO₂ were underestimated for more than 60% of time when plume from first validated landfill fire could be measured. In case of next three analyzed landfill fires PM₁₀ and C₆H₆ were underestimated for non-winter landfill fires, and remaining BTEX were underestimated at least one during these fires. Since it was shown that commonly measured air pollutants are also object of significant local impact, the BTEX, especially benzene, seems to be a marker of impact of fire on air quality. This work provides an example which can be adapted to different rural background stations. The approach has potential to be applied in an identification of fires impact on air quality stations worldwide, after tuning to the local meteorological conditions. The adaptation needs the integration with database of fires and validation set for which dispersion was evaluated.

ADDITIONAL INFORMATION

Funding

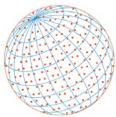
This work was supported by the National Science Centre, Poland, 2020/37/N/ST10/02997 *The impact of landfill fires on the atmospheric air quality - methodology and estimation of emission.*

Supplementary Material

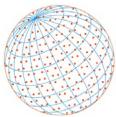
Supplementary material for this article can be found in the online version at <https://doi.org/10.4209/aaqr.220172>

REFERENCES

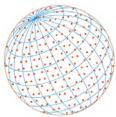
- Aderemi, A.O., Otitoloju, A.A. (2012). An assessment of landfill fires and their potential health effects-a case study of a municipal solid waste landfill in Lagos, Nigeria. Int. J. Environ. Prot. 2, 22–26.



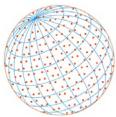
- Aras, H., Aras, N. (2004). Forecasting residential natural gas demand. *Energy Sources* 26, 463–472. <https://doi.org/10.1080/00908310490429740>
- Arbex, M.A., Martins, L.C., Oliveira, R.C. de, Pereira, L.A.A., Arbex, F.F., Cançado, J.E.D., Saldiva, P.H.N., Braga, A.L.F. (2007). Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in Brazil. *J. Epidemiol. Community Health* 61, 395–400. <https://doi.org/10.1136/jech.2005.044743>
- Banton, M. (2014). Ethylbenzene, in: Wexler, P. (Ed.), *Encyclopedia of Toxicology* (Third Edition), Academic Press, Oxford, pp. 516–518. <https://doi.org/10.1016/B978-0-12-386454-3.00311-0>
- Błaszcak, B., Rogula-Kozłowska, W., Mathews, B., Juda-Rezler, K., Klejnowski, K., Rogula-Kopiec, P. (2016). Chemical compositions of PM_{2.5} at two non-urban sites from the polluted region in Europe. *Aerosol Air Qual. Res.* 16, 2333–2348. <https://doi.org/10.4209/aaqr.2015.09.0538>
- Bihałowicz, J.S. (2021). The cross-boundary impact of the landfill fires in Poland on air quality. *Sci. Business. Soc.* 6, 11–14.
- Bihałowicz, J.S., Rogula-Kozłowska, W., Krasuski, A. (2021a). Contribution of landfill fires to air pollution – An assessment methodology. *Waste Manage.* 125, 182–191. <https://doi.org/10.1016/j.wasman.2021.02.046>
- Bihałowicz, J.S., Rogula-Kozłowska, W., Krasuski, A., Majder-Łopatka, M., Walczak, A., Fliszkiewicz, M., Rogula-Kopiec, P., Mach, T. (2021b). Characteristics of particles emitted from waste fires—A construction materials case study. *Materials* 15, 152. <https://doi.org/10.3390/ma15010152>
- Bihałowicz, J.S., Rogula-Kozłowska, W., Krasuski, A., Salamonowicz, Z. (2021c). The critical factors of landfill fire impact on air quality. *Environ. Res. Lett.* 16, 104026. <https://doi.org/10.1088/1748-9326/ac27cd>
- Bihałowicz, J.S. (2022). Application of machine learning in air pollutants modeling: A case study of PM_{2.5} in Szczecin (Poland). *Zesz. Nauk. SGSP* in press.
- Blomqvist, P., Persson, B., Simonson, M. (2007). Fire emissions of organics into the atmosphere. *Fire Technol.* 43, 213–231. <https://doi.org/10.1007/s10694-007-0011-y>
- Bun, R., Nahorski, Z., Horabik-Pyzel, J., Danylo, O., See, L., Charkovska, N., Topylko, P., Halushchak, M., Lesiv, M., Valakh, M., Kinakh, V. (2019). Development of a high-resolution spatial inventory of greenhouse gas emissions for Poland from stationary and mobile sources. *Mitigation Adapt. Strategies Global Change* 24, 853–880. <https://doi.org/10.1007/s11027-018-9791-2>
- Chaudhary, I.J., Soni, S. (2020). Crop Residue Burning and Its Effects on the Environment and Microbial Communities, in: *Microbes in Agriculture and Environmental Development*, CRC Press, pp. 87–106.
- Chen, C.M. (2004). The emission inventory of PCDD/PCDF in Taiwan. *Chemosphere* 54, 1413–1420. <https://doi.org/10.1016/j.chemosphere.2003.10.039>
- Cheng, B., Ma, Y., Feng, F., Zhang, Y., Shen, J., Wang, H., Guo, Y., Cheng, Y. (2021). Influence of weather and air pollution on concentration change of PM_{2.5} using a generalized additive model and gradient boosting machine. *Atmos. Environ.* 255, 118437. <https://doi.org/10.1016/j.atmosenv.2021.118437>
- Cheng, M.T., Horng, C.L., Su, Y.R., Lin, L.K., Lin, Y.C., Chou, C.C.K. (2009). Particulate matter characteristics during agricultural waste burning in Taichung City, Taiwan. *J. Hazard. Mater.* 165, 187–192. <https://doi.org/10.1016/j.jhazmat.2008.09.101>
- Clarkson, T.S., Martin, R.J., Rudolph, J., Graham, B.W.L. (1996). Benzene and toluene in New Zealand air. *Atmos. Environ.* 30, 569–577. [https://doi.org/10.1016/1352-2310\(95\)00312-6](https://doi.org/10.1016/1352-2310(95)00312-6)
- Courtemanche, B., Levendis, Y. (1998). A laboratory study on the NO, NO₂, SO₂, CO and CO₂ emissions from the combustion of pulverized coal, municipal waste plastics and tires. *Fuel* 77, 183–196. [https://doi.org/10.1016/S0016-2361\(97\)00191-9](https://doi.org/10.1016/S0016-2361(97)00191-9)
- Cygańczuk, K., Zielecka, M., Rabajczyk, A. (2020). Zasady bezpiecznej gospodarki odpadami jako czynnik bezpieczeństwa środowiskowego. *Przem. Chem.* 1, 36–41. <https://doi.org/10.15199/62.2020.7.5>
- Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevat, T., Milutinović, M., Možina, M., Polajnar, M., Toplak, M., Starić, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., Zupan, B. (2013). Orange: Data mining toolbox in python. *J. Mach. Learn. Res.* 14, 2349–2353.
- Downard, J., Singh, A., Bullard, R., Jayaratne, T., Rathnayake, C.M., Simmons, D.L., Wels, B.R., Spak, S.N., Peters, T., Beardsley, D., Stanier, C.O., Stone, E.A. (2015). Uncontrolled combustion



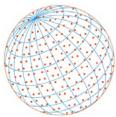
- of shredded tires in a landfill – Part 1: Characterization of gaseous and particulate emissions. *Atmos. Environ.* 104, 195–204. <https://doi.org/10.1016/j.atmosenv.2014.12.059>
- Dzikuć, M., Adamczyk, J., Piwowar, A. (2017). Problems associated with the emissions limitations from road transport in the Lubuskie Province (Poland). *Atmos. Environ.* 160, 1–8. <https://doi.org/10.1016/j.atmosenv.2017.04.011>
- EC (2008). Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal of the European Union.
- Ettala, M., Rahkonen, P., Rossi, E., Mangs, J., Keski-Rahkonen, O. (1996). Landfill fires in Finland. *Waste Manage. Res.* 14, 377–384. <https://doi.org/10.1177/0734242X9601400405>
- European Environment Age (EEA) (2016). EMEP/EEA air pollutant emission inventory guidebook 2016. EEA Report, Publications Office of the Europe, Luxembourg. <https://doi.org/10.2800/247535>
- Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M., Hutter, F. (2020). Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning. *arXiv* <https://doi.org/10.48550/arXiv.2007.04074>
- Gadi, R., Kulshrestha, U.C., Sarkar, A.K., Garg, S.C., Parashar, D.C. (2011). Emissions of SO₂ and NO_x from biofuels in India. *Tellus B* 55, 787–795. <https://doi.org/10.3402/tellusb.v55i3.16370>
- GIOŚ (2021). Portal Jakość Powietrza GIOŚ. <http://powietrze.gios.gov.pl/pjp/home> (accessed 2 December 2021).
- Grundström, M., Hak, C., Chen, D., Hallquist, M., Pleijel, H. (2015). Variation and co-variation of PM₁₀, particle number concentration, NO_x and NO₂ in the urban air – Relationships with wind speed, vertical temperature gradient and weather type. *Atmos. Environ.* 120, 317–327. <https://doi.org/10.1016/j.atmosenv.2015.08.057>
- Gullett, B.K., Lemieux, P.M., Lutes, C.C., Winterrowd, C.K., Winters, D.L. (2001). Emissions of PCDD/F from uncontrolled, domestic waste burning. *Chemosphere* 43, 721–725. [https://doi.org/10.1016/S0045-6535\(00\)00425-2](https://doi.org/10.1016/S0045-6535(00)00425-2)
- Gullett, B.K., Touati, A., Huwe, J., Hakk, H. (2006). PCDD and PCDF emissions from simulated sugarcane field burning. *Environ. Sci. Technol.* 40, 6228–6234. <https://doi.org/10.1021/es060806k>
- H2O.ai (2020). Python Interface for H2O. H2O. <https://github.com/h2oai/h2o-3>
- H2O.ai (2021). Scalable Machine Learning Platform. Version 3.32.1.5. H2O. <https://h2o.ai/>
- H2O.ai (2022). Gradient Boosting Machine (GBM). H2O documentation site. <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/gbm.html> (accessed 20 June 2022).
- Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Statistical Learning. Springer New York, New York. <https://doi.org/10.1007/b94608>
- Hedman, B., Näslund, M., Nilsson, C., Marklund, S. (2005). Emissions of polychlorinated dibenzodioxins and dibenzofurans and polychlorinated biphenyls from uncontrolled burning of garden and domestic waste (backyard burning). *Environ. Sci. Technol.* 39, 8790–8796. <https://doi.org/10.1021/es051117w>
- Hoffer, A., Jancsek-Turóczi, B., Tóth, Á., Kiss, G., Naghiu, A., Levei, E.A., Marmureanu, L., Machon, A., Gelencsér, A. (2020). Emission factors for PM₁₀ and polycyclic aromatic hydrocarbons (PAHs) from illegal burning of different types of municipal waste in households. *Atmos. Chem. Phys.* 20, 16135–16144. <https://doi.org/10.5194/acp-20-16135-2020>
- Hribar, R., Potočnik, P., Šilc, J., Papa, G. (2019). A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy* 167, 511–522. <https://doi.org/10.1016/j.energy.2018.10.175>
- Huang, L., Chen, M., Hu, J. (2016). Twelve-year trends of PM₁₀ and visibility in the Hefei metropolitan area of China. *Adv. Meteorol.* 2016, 1–9. <https://doi.org/10.1155/2016/4810796>
- IMGW-PIB (2021). Dane publiczne IMGW-PIB. <https://dane.imgw.pl/> (accessed 12 August 2021).
- Jedlikowski, A., Englart, S. (2018). The comparison of calculation methods of gas demand in the installation with combi boilers. *E3S Web Conf.* 44, 00058. <https://doi.org/10.1051/e3sconf/20184400058>
- Jeong, J.I., Park, R.J. (2013). Effects of the meteorological variability on regional air quality in East Asia. *Atmos. Environ.* 69, 46–55. <https://doi.org/10.1016/j.atmosenv.2012.11.061>
- Jeong, Y., Youn, Y., Cho, S., Kim, S., Huh, M., Lee, Y. (2020). Prediction of daily PM₁₀ concentration for air Korea stations using artificial intelligence with LDAPS weather data, MODIS AOD, and Chinese air quality data. *Korean J. Remote Sens.* 36, 573–586. <https://doi.org/10.7780/kjrs.2020.36.4.7>



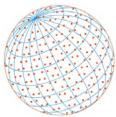
- Jones, A.M., Harrison, R.M., Baker, J. (2010). The wind speed dependence of the concentrations of airborne particulate matter and NO_x. *Atmos. Environ.* 44, 1682–1690. <https://doi.org/10.1016/j.atmosenv.2010.01.007>
- Kandyala, R., Raghavendra, S.P., Rajasekharan, S. (2010). Xylene: An overview of its health hazards and preventive measures. *J. Oral Maxillofac. Pathol.* 14, 1. <https://doi.org/10.4103/0973-029X.64299>
- Kelessis, A.G., Petrakakis, M.J., Zoumakis, N.M. (2006). Determination of benzene, toluene, ethylbenzene, and xylenes in urban air of Thessaloniki, Greece. *Environ. Toxicol.* 21, 440–443. <https://doi.org/10.1002/tox.20197>
- Komenda Główna Państwowej Straży Pożarnej (KG PSP) (2019). Zasady Ewidencjonowania Zdarzeń w Systemie Wspomagania Decyzji Państwowej Straży Pożarnej. Komenda Główna Państwowej Straży Pożarnej. <https://www.gov.pl/attachment/4a4a6c0e-28f4-4826-a60a-9e75c43e9d63>
- Komenda Główna Państwowej Straży Pożarnej (KG PSP) (2020). Interwencje PSP: lata 2010–2019 zestawienia. https://www.kgpsp.gov.pl/panstwowa_straz_pozarna/interwencje_psp (accessed 18 January 2020).
- Klejnowski, K., Krasa, A., Rogula-Kozłowska, W., Błaszcak, B. (2013). Number size distribution of ambient particles in a typical urban site: The first polish assessment based on long-term (9 Months) measurements. *Sci. World J.* 2013, 1–13. <https://doi.org/10.1155/2013/539568>
- Krajowy Ośrodek Bilansowania i Zarządzania Emisjami (KOBiZE) (2020). Poland's Informative Inventory Report 2020. Submission under the UN ECE Convention on Long-range Transboundary Air Pollution and the Directive (EU) 2016/2284 Air pollutant emissions in Poland 1990–2018. National Centre for Emission Management (KOBiZE) at the Institute of Environmental Protection – National Research Institute, Poland.
- Kollanus, V., Tiittanen, P., Niemi, J. V., Lanki, T. (2016). Effects of long-range transported air pollution from vegetation fires on daily mortality and hospital admissions in the Helsinki metropolitan area, Finland. *Environ. Res.* 151, 351–358. <https://doi.org/10.1016/j.envres.2016.08.003>
- Kourtidis, K.A., Ziomas, I., Zerefos, C., Kosmidis, E., Symeonidis, P., Christophilopoulos, E., Karathanassis, S., Mploutsos, A. (2002). Benzene, toluene, ozone, NO₂ and SO₂ measurements in an urban street canyon in Thessaloniki, Greece. *Atmos. Environ.* 36, 5355–5364. [https://doi.org/10.1016/S1352-2310\(02\)00580-0](https://doi.org/10.1016/S1352-2310(02)00580-0)
- Kumari, K., Kumar, S., Rajagopal, V., Khare, A., Kumar, R. (2019). Emission from open burning of municipal solid waste in India. *Environ. Technol.* 40, 2201–2214. <https://doi.org/10.1080/09593330.2017.1351489>
- LeDell, E., Poirier, S. (2020). H2O AutoML: Scalable Automatic Machine Learning. 7th ICML Workshop on Automated Machine Learning. https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_61.pdf
- Lehtomäki, H., Korhonen, A., Asikainen, A., Karvosenoja, N., Kupiainen, K., Paunu, V.V., Savolahti, M., Sofiev, M., Palamarchuk, Y., Karppinen, A., Kukkonen, J., Hänninen, O. (2018). Health impacts of ambient air pollution in Finland. *Int. J. Environ. Res. Public Health* 15, 736. <https://doi.org/10.3390/ijerph15040736>
- Lemieux, P. (1998). Evaluation of Emissions from the Open Burning of Household Waste in Barrels. EPA-600/R-97-134a, U.S. Environmental Protection Agency, USA.
- Lemieux, P.M., Lutes, C.C., Abbott, J.A., Aldous, K.M. (2000). Emissions of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans from the open burning of household waste in barrels. *Environ. Sci. Technol.* 34, 377–384. <https://doi.org/10.1021/es990465t>
- Lemieux, P.M., Lutes, C.C., Santoiani, D.A. (2004). Emissions of organic air toxics from open burning: A comprehensive review. *Prog. Energy Combust. Sci.* 30, 1–32. <https://doi.org/10.1016/j.pecs.2003.08.001>
- Li, M., Cao, J., Guo, J., Niu, J., Xiong, M. (2016). Response of energy consumption for building heating to climatic change and variability in Tianjin City, China. *Meteorol. Appl.* 23, 123–131. <https://doi.org/10.1002/met.1537>
- Li, X., Huang, L., Li, J., Shi, Z., Wang, Y., Zhang, H., Ying, Q., Yu, X., Liao, H., Hu, J. (2019). Source contributions to poor atmospheric visibility in China. *Resour. Conserv. Recycl.* 143, 167–177. <https://doi.org/10.1016/j.resconrec.2018.12.029>
- Liu, D.R., Lee, S.J., Huang, Y., Chiu, C.J. (2020). Air pollution forecasting based on attention-based



- LSTM neural network and ensemble learning. *Expert Systems* 37, e12511. <https://doi.org/10.1111/exsy.12511>
- Liu, M., Chen, H., Wei, D., Wu, Y., Li, C. (2021). Nonlinear relationship between urban form and street-level PM_{2.5} and CO based on mobile measurements and gradient boosting decision tree models. *Build. Environ.* 205, 108265. <https://doi.org/10.1016/j.buildenv.2021.108265>
- Lönnérmark, A., Blomqvist, P. (2006). Emissions from an automobile fire. *Chemosphere* 62, 1043–1056. <https://doi.org/10.1016/j.chemosphere.2005.05.002>
- Łukawski, A.S. (2019). Zawartość lotnych związków organicznych w powietrzu na terenie jednostki ratowniczo-gaśniczej. Ocena potencjalnych źródeł ich pochodzenia oraz negatywnego wpływu na zdrowie strażaka. *Zesz. Nauk. SGSP/Szkoła Główna Służby Pożarniczej*.
- Mach, T., Rogula-Kozłowska, W., Bralewska, K., Majewski, G., Rogula-Kopiec, P., Rybak, J. (2021). Impact of municipal, road traffic, and natural sources on PM₁₀: The hourly variability at a rural site in Poland. *Energies* 14, 2654. <https://doi.org/10.3390/en14092654>
- Majewski, G., Rogula-Kozłowska, W., Czechowski, P., Badyda, A., Brandyk, A. (2015). The impact of selected parameters on visibility: First results from a long-term campaign in Warsaw, Poland. *Atmosphere* 6, 1154–1174. <https://doi.org/10.3390/atmos6081154>
- Majewski, G., Rogula-Kozłowska, W., Rozbicka, K., Rogula-Kopiec, P., Mathews, B., Brandyk, A. (2018). Concentration, chemical composition and origin of PM₁: Results from the first long-term measurement campaign in Warsaw (Poland). *Aerosol Air Qual. Res.* 18, 636–654. <https://doi.org/10.4209/aaqr.2017.06.0221>
- Majewski, G., Szeląg, B., Mach, T., Rogula-Kozłowska, W., Anioł, E., Bihałowicz, J., Dmochowska, A., Bihałowicz, J.S. (2021). Predicting the number of days with visibility in a specific range in Warsaw (Poland) based on meteorological and air quality data. *Front. Environ. Sci.* 9, 623094. <https://doi.org/10.3389/fenvs.2021.623094>
- Malohlava, M., Candel, A. (2020). Gradient Boosting Machine with H2O. <https://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/GBMBooklet.pdf> (accessed 15 December 2021)
- Megaritis, A.G., Fountoukis, C., Charalampidis, P.E., Denier van der Gon, H.A.C., Pilinis, C., Pandis, S.N. (2014). Linking climate and air quality over Europe: Effects of meteorology on PM_{2.5} concentrations. *Atmos. Chem. Phys.* 14, 10283–10298. <https://doi.org/10.5194/acp-14-10283-2014>
- Mihai, F.C., Banica, A., Grozavu, A. (2019). Backyard burning of household waste in rural areas: environmental impact with focus on air pollution, in: International Multidisciplinary Scientific GeoConference on Ecology, Economics, Education and Legislation-SGEM 2019, MISC, pp. 55–62.
- Nadal, M., Rovira, J., Díaz-Ferrero, J., Schuhmacher, M., Domingo, J.L. (2016). Human exposure to environmental pollutants after a tire landfill fire in Spain: Health risks. *Environ. Int.* 97, 37–44. <https://doi.org/https://doi.org/10.1016/j.envint.2016.10.016>
- Nidzgorska-Lencewicz, J., Czarnecka, M. (2015). Winter weather conditions vs. air quality in Tricity, Poland. *Theor. Appl. Climatol.* 119, 611–627. <https://doi.org/10.1007/s00704-014-1129-8>
- Oleniacz, R., Bogacki, M., Szulecka, A., Rzeszutek, M., Mazur, M. (2016). Assessing the impact of wind speed and mixing-layer height on air quality in Krakow (Poland) in the years 2014–2015. *J. Civ. Eng. Environ. Archit.* 63, 315–342. <https://doi.org/10.7862/rb.2016.168>
- Pansuk, J., Junpen, A., Garivait, S. (2018). Assessment of air pollution from household solid waste open burning in Thailand. *Sustainability* 10, 2553. <https://doi.org/10.3390/su10072553>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Płońska, A., Płoński, P. (2021). State-of-the-art Automated Machine Learning Framework for Tabular Data. Version 0.10.3. MLJAR. <https://github.com/mljar/mljar-supervised>
- Podstawnińska, A., Chambers, S.D. (2019). Improved method for characterising temporal variability in urban air quality part I: Traffic emissions in central Poland. *Atmos. Environ.* 219, 117038. <https://doi.org/10.1016/j.atmosenv.2019.117038>
- Pyta, H., Widziewicz-Rzońca, K., Ślaby, K. (2020). Inhalation exposure to gaseous and particulate bound mercury present in the ambient air over the polluted area of Southern Poland. *Int. J. Environ. Res. Public Health* 17, 4999. <https://doi.org/10.3390/ijerph17144999>



- QGIS (2021). QGIS Geographic Information System. Open Source Geospatial Foundation Project. ver 3.22 Białowieża. QGIS.
- Rangel, M.G.L., Henríquez, J.R., Costa, J.A.P., de Lira Junior, J.C. (2018). An assessment of dispersing pollutants from the pre-harvest burning of sugarcane in rural areas in the northeast of Brazil. *Atmos. Environ.* 178, 265–281. <https://doi.org/10.1016/j.atmosenv.2018.02.006>
- Rogula-Kozłowska, W., Kozielska, B., Klejnowski, K. (2013a). Concentration, origin and health hazard from fine particle-bound PAH at three characteristic sites in southern Poland. *Bull. Environ. Contam. Toxicol.* 91, 349–355. <https://doi.org/10.1007/s00128-013-1060-1>
- Rogula-Kozłowska, W., Kozielska, B., Klejnowski, K. (2013b). Hazardous compounds in urban PM in the central part of Upper Silesia (Poland) in winter. *Arch. Environ. Prot.* 39, 53–65. <https://doi.org/10.2478/aep-2013-0002>
- Rogula-Kozłowska, W., Klejnowski, K., Rogula-Kopiec, P., Ośródka, L., Krajny, E., Błaszcak, B., Mathews, B. (2014). Spatial and seasonal variability of the mass concentration and chemical composition of PM_{2.5} in Poland. *Air Qual. Atmos. Health* 7, 41–58. <https://doi.org/10.1007/s11869-013-0222-y>
- Rogula-Kozłowska, W., Majewski, G., Rogula-Kopiec, P., Mathews, B. (2019a). Mass concentration and chemical composition of submicron particulate matter (PM₁) in the Polish urban areas. *IOP Conf. Ser. Earth Environ. Sci.* 214, 012092. <https://doi.org/10.1088/1755-1315/214/1/012092>
- Rogula-Kozłowska, Wioletta, Majewski, G., Widziewicz, K., Rogula-Kopiec, P., Tytla, M., Mathews, B., Ciuka-Witrylak, M. (2019b). Seasonal variations of PM₁-bound water concentration in urban areas in Poland. *Atmos. Pollut. Res.* 10, 267–273. <https://doi.org/10.1016/j.apr.2018.08.004>
- Rogula-Kozłowska, W., Bralewska, K., Jureczko, I. (2020). BTEX concentrations and exposure assessment in a fire station. *Atmosphere* 11, 470. <https://doi.org/10.3390/atmos11050470>
- Rozbicka, K., Majewski, G., Rogula-Kozłowska, W., Rozbicki, T. (2020). Tropospheric ozone assessment in urban environment – Warsaw case study of selected heat waves. *J. Atmos. Sol.-Terr. Phys.* 209, 105418. <https://doi.org/10.1016/j.jastp.2020.105418>
- Ruokojärvi, P., Ruuskanen, J., Ettala, M., Rahkonen, P., Tarhanen, J. (1995). Formation of polycyclic aromatic hydrocarbons and polychlorinated organic compounds in municipal waste landfill fires. *Chemosphere* 31, 3899–3908. [https://doi.org/10.1016/0045-6535\(95\)00264-9](https://doi.org/10.1016/0045-6535(95)00264-9)
- Sabo, K., Scitovski, R., Vazler, I., Zekić-Sušac, M. (2011). Mathematical models of natural gas consumption. *Energy Convers. Manage.* 52, 1721–1727. <https://doi.org/10.1016/j.enconman.2010.10.037>
- Sarkar, S., Singh, R.P., Chauhan, A. (2018). Crop residue burning in Northern India: Increasing threat to greater India. *J. Geophys. Res.* 123, 6920–6934. <https://doi.org/10.1029/2018JD028428>
- Shahriar, S.A., Kayes, I., Hasan, K., Hasan, M., Islam, R., Awang, N.R., Hamzah, Z., Rak, A.E., Salam, M.A. (2021). Potential of ARIMA-ANN, ARIMA-SVM, DT and catboost for atmospheric PM_{2.5} forecasting in Bangladesh. *Atmosphere* 12, 100. <https://doi.org/10.3390/atmos12010100>
- Sharma, G., Annadate, S., Sinha, B. (2022). Will open waste burning become India's largest air pollution source? *Environ. Pollut.* 292, 118310. <https://doi.org/10.1016/j.envpol.2021.118310>
- Soares, J., Sofiev, M. (2014). A Global Wildfire Emission and Atmospheric Composition: Refinement of the Integrated System for Wild-Land Fires IS4FIRES, in: Steyn, D., Mathur, R. (Eds.), *Air Pollution Modeling and its Application XXIII*, Springer International Publishing, Cham, pp. 253–258. https://doi.org/10.1007/978-3-319-04379-1_41
- Stein, A.F., Draxler, R.R., Rolph, G.D., Stunder, B.J.B., Cohen, M.D., Ngan, F. (2015). NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bull. Am. Meteorol. Soc.* 96, 2059–2077. <https://doi.org/10.1175/bams-d-14-00110.1>
- Stetsenko, P. (2020). Machine Learning with Python and H2O. H2O.ai, Inc., USA. <https://h2o.ai/content/dam/h2o/en/marketing/documents/2018/01/Python-BOOKLET.pdf>
- 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. Total Environ.* 653, 140–147. <https://doi.org/10.1016/j.scitotenv.2018.10.368>
- Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy* 85, 208–220. <https://doi.org/10.1016/j.energy.2015.03.084>
- Thunis, P., Clappier, A., Tarrason, L., Cuvelier, C., Monteiro, A., Pisoni, E., Wesseling, J., Belis, C.A., Pirovano, G., Janssen, S., Guerreiro, C., Peduzzi, E. (2019). Source apportionment to support



- air quality planning: Strengths and weaknesses of existing approaches. *Environ. Int.* 130, 104825. <https://doi.org/10.1016/j.envint.2019.05.019>
- Tsai, Y.I., Kuo, S.C., Lee, W.J., Chen, C.L., Chen, P.T. (2007). Long-term visibility trends in one highly urbanized, one highly industrialized, and two Rural areas of Taiwan. *Sci. Total Environ.* 382, 324–341. <https://doi.org/10.1016/j.scitotenv.2007.04.048>
- U.S. Environmental Protection Agency (U.S. EPA) (1993a). Locating and estimating air emissions from sources of toluene. EPA-454/R-93-047, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.
- U.S. Environmental Protection Agency (U.S. EPA) (1993b). Locating and Estimating Air Emissions from Sources of Xylene. EPA-454/R-93-048, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.
- U.S. Environmental Protection Agency (U.S. EPA) (1995). AP-42: Compilation of Air Pollutant Emission Factors. <https://www.epa.gov/air-emissions-factors-and-quantification/ap-42-compilation-air-emissions-factors>
- U.S. Environmental Protection Agency (U.S. EPA) (1998). Locating and Estimating Air Emissions from Sources of Benzene. EPA-454/R-98-011, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.
- Weil, J.C., Sykes, R.I., Venkatram, A. (1992). Evaluating air-quality models: Review and outlook. *J. Appl. Meteorol.* 31, 1121–1145. [https://doi.org/10.1175/1520-0450\(1992\)031<1121:EAQMRA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1992)031<1121:EAQMRA>2.0.CO;2)
- Wiedinmyer, C., Yokelson, R.J., Gullett, B.K. (2014). Global emissions of trace gases, particulate matter, and hazardous air pollutants from open burning of domestic waste. *Environ. Sci. Technol.* 48, 9523–9530. <https://doi.org/10.1021/es502250z>
- Yevich, R., Logan, J.A. (2003). An assessment of biofuel use and burning of agricultural waste in the developing world. *Global Biogeochemical Cycles* 17, 1095. <https://doi.org/10.1029/2002GB001952>
- Zhang, B., Jiao, L., Xu, G., Zhao, S., Tang, X., Zhou, Y., Gong, C. (2018). Influences of wind and precipitation on different-sized particulate matter concentrations (PM_{2.5}, PM₁₀, PM_{2.5–10}). *Meteorol. Atmos. Phys.* 130, 383–392. <https://doi.org/10.1007/s00703-017-0526-9>
- Zhang, G., Huang, X., Liao, W., Kang, S., Ren, M., Hai, J. (2019). Measurement of dioxin emissions from a small-scale waste incinerator in the absence of air pollution controls. *Int. J. Environ. Res. Public Health* 16, 1267. <https://doi.org/10.3390/ijerph16071267>