

# Assessment of Machine Learning Algorithms in Short-term Forecasting of PM<sub>10</sub> and PM<sub>2.5</sub> Concentrations in Selected Polish Agglomerations

Bartosz Czernecki<sup>1\*</sup>, Michał Marosz<sup>2,3</sup>, Joanna Jędruszkiewicz<sup>4</sup>

<sup>1</sup> Department of Meteorology and Climatology, Adam Mickiewicz University in Poznań, Poznań, Poland

<sup>2</sup> Institute of Meteorology and Water Management—National Research Institute, Warsaw, Poland

<sup>3</sup> University of Gdańsk, Faculty of Oceanography and Geography, Gdańsk, Poland,

<sup>4</sup> Institute of Geography, Pedagogical University of Cracow, Kraków, Poland

## ABSTRACT

Air pollution continues to have a significant impact on Europeans living in urban areas, and episodes of elevated PM<sub>x</sub> are responsible for a large number of premature deaths (mostly due to heart disease and stroke) each year. According to the annual EEA reports, Poland is one of the most polluted countries in Europe, experiencing high PM<sub>x</sub> concentrations during winter that mostly result from large emissions and unfavourable weather conditions in combination with environmental features. Thus, in addition to implementing municipal mitigation strategies, alerting residents to pollution episodes through accurate PM<sub>x</sub> forecasting is necessary.

This research aimed to assess the feasibility of short-term PM<sub>x</sub> forecasting via machine learning (ML) and the subsequent identification of the primary meteorological covariates. The data comprised 10 years of hourly winter PM<sub>10</sub> and PM<sub>2.5</sub> concentrations measured at 11 urban air quality monitoring stations, including background, traffic, and industrial sites, in four large Polish agglomerations, viz., Poznań, Kraków, Łódź, and Gdańsk, which cover areas with high population density and diverse environments that extend from the Baltic Sea coast (Tricity) through the lowlands (Poznań and Łódź) to the highlands (Kraków).

We tested four ML models: AIC-based stepwise regression, two tree-based algorithms (random forests and XGBoost), and neural networks. Employing analysis and cross-validation, we found that XGBoost performed the best, followed by random forests and neural networks, and stepwise regression performed the worst. This ranking was apparent in the threshold exceedance values of the binary forecasts obtained via regression. Overall, our results confirm the high applicability of ML to short-term air quality prediction with the perfect prog approach.

**Keywords:** PM<sub>10</sub>, PM<sub>2.5</sub>, Air quality, Machine learning, Short-term forecasting

OPEN ACCESS

Received: October 5, 2020

Revised: February 17, 2021

Accepted: March 20, 2021

\* Corresponding Author:

nwp@amu.edu.pl

Publisher:

Taiwan Association for Aerosol  
Research

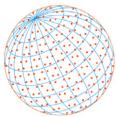
ISSN: 1680-8584 print

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.

## 1 INTRODUCTION

Particulate matter (PM) is a term widely used for a mixture of organic and inorganic substances suspended in the air. The most commonly measured are those with a diameter of less than 10 µm (PM<sub>10</sub>) and less than 2.5 µm (PM<sub>2.5</sub>), with the latter being more harmful to human health. PM can penetrate the human respiratory system, but the small particles can even pass through the respiratory barrier and enter into the circulatory system, spreading throughout the body (Pinkerton *et al.*, 2000; Feng *et al.*, 2016). PM<sub>10</sub> and PM<sub>2.5</sub> have adverse health effects that include acute and chronic respiratory illnesses such as pneumonia, chronic bronchitis, and asthma. It may also aggravate coughing or breathing difficulty. These small particles are also responsible for cardiovascular diseases such as coronary heart disease, congestive heart failure, stroke, or cardiac



arrhythmias. All cardiopulmonary diseases can lead to premature death (Polichetti *et al.*, 2009; Guaita *et al.*, 2011; Anderson *et al.*, 2012; Kim *et al.*, 2015; Lu *et al.*, 2015; Zheng *et al.*, 2015). The European Environment Agency (EEA) (2018) estimated that in Europe, the number of premature deaths that could have been a result of PM<sub>2.5</sub> exposure reached 420,000, and the World Health Organization (WHO) (2013) indicated that approximately 1–3% of cardiopulmonary and 2–5% of lung cancer deaths are attributed to PM. According to WHO (2014, 2018), the number of deaths related to ambient air pollution increased globally from 3 million in 2012 to 4.2 million in 2016. In the most polluted cities, life expectancy is reduced by 20 months. People living in urban areas with heart or lung diseases, as well as children, and elderly ( $\geq 75$  years of age) are the most vulnerable to air pollution (Samoli *et al.*, 2008).

Furthermore, a high PM concentration inevitably leads to economic losses due to health problems. In 2009 in China, such losses reached USD 106.5 billion (2.1% of China's GDP) (Hou *et al.*, 2012). The European Commission (EC) reported that the total health-related external costs in 2010 were within the range of EUR 330–940 billion (EC, 2013), while European GDP losses are predicted to increase by up to 2% by 2060 according to the Organisation for Economic Co-operation and Development (OECD, 2016).

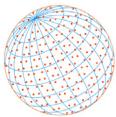
EEA (2018) reported that Poland (together with the Balkans and Turkey) has the worst air quality among European countries and one of the highest rates of premature deaths (44,500) related to PM<sub>2.5</sub> concentrations. Each year, daily and annual PM concentration limits are frequently exceeded (EEA, 2015). One of the most extreme incidents occurred in January 2017. The Polish Central Statistical Office recorded 35% (about 11,000) more fatalities than in January 2016. In Kraków, one of the most polluted cities in Poland, the total annual mortality rate (2005–2013) attributed to exposure to PM<sub>10</sub> ranged from 30 to 50 per 100,000 inhabitants (half of which is due to cardiovascular diseases), and from 40 to 58 per 100,000 inhabitants due to exposure to PM<sub>2.5</sub> (Samek, 2016).

According to Awe *et al.* (2019), the primary and still-growing source of PM<sub>2.5</sub> and PM<sub>10</sub> emissions in Poland is the residential sector (respectively 48% and 44% in 2015). The manufacturing industry is the second-largest emission source (more than 20% and 10%, respectively). Road transport and agriculture are also key contributors to air pollution. Another problem is the widespread plastic and chemicals combustions in illegal landfills, which in recent years have risen to enormous levels.

Poland's highest air pollution levels occur in winter. The problem is twofold. First of all, the primary sources of air pollution are coal combustion for heating, as well as regional and long-range transport (Juda-Rezler, 2006; Mira-Salama, 2008; Junninen *et al.*, 2009). The high emission level is intensified by burning coal of inferior quality and wastes in small residential units. Secondly, the poor air quality in winter is a result of the year's most unfavourable weather conditions. These comprise low temperature, stable atmosphere, calm winds, and limited convection, all of which reduce dispersion (Czernecki *et al.*, 2017; Jędruszkiewicz *et al.*, 2017).

Therefore, new mitigation schemes are being implemented by local authorities to improve or at least not to aggravate air quality in urban areas due to human activity. According to a decree of the Minister for the Environment on 8 October 2019, the daily thresholds of the smog alarm and information for PM<sub>10</sub> were respectively reduced to 150  $\mu\text{g m}^{-3}$  and 100  $\mu\text{g m}^{-3}$ —albeit still at least twice as much as those in most European countries. If the daily alarm threshold is exceeded, both local and regional authorities are obliged to take short-term actions. These may comprise free public transport, reducing car traffic, banning use of fireplaces, controlling industrial plants, or forbidding combustions in domestic furnaces. For example, in Poznań, it is not allowed to use a fireplace if the daily mean PM<sub>10</sub> concentration exceeds 50  $\mu\text{g m}^{-3}$ . Another one might be free public transport in days when the short-term forecasted PM<sub>10</sub> level exceeds 200  $\mu\text{g m}^{-3}$ .

These regulations for any random day are based on the daily means PM<sub>x</sub> concentration forecasts. Making the best-fitted forecast for that season is crucial for people, as it might help to alert them on time, especially in urban areas. Releasing the correct prediction on time might limit exposure to high PM concentrations, improve life expectancy, and allow authorities to introduce immediate relief actions such as private vehicle movement restrictions or free municipal transport. Therefore, the primary goal of this research is to create and compare an ensemble of short-term forecasts by means of a range of machine learning tools. These can be cheaper and more accurate than the presently used, computationally demanding numerical weather prediction models with chemical



modules. For that purpose, we compared the results from (1) linear regression with stepwise selection (stepAIC), (2) random forests (RFs), (3) XGBoost regression trees (XGBoost), and (4) neural networks (nnets). Regression models are limited by certain statistical assumptions, such as the independence of observations or distribution. More advanced machine learning methods overcome these limitations via data analysis with the use of self-learning statistical models (Hu *et al.*, 2017). Machine learning uses computational statistical methods to ascertain the best models that can learn from historical data and make predictions on such bases. Here, we used the daily means of PM<sub>2.5</sub> and PM<sub>10</sub> concentration levels, as well as atmospheric conditions in four agglomerations: Kraków, Łódź, Poznań, and Tricity. These cities represent different locations, topography, and climatic conditions. The 1-day forecasts were prepared for wintertime, as this season tends to stand out in terms of very high pollution levels.

In general, there are two main approaches for predicting PM concentrations. One is with the use of numerical models, and the other is the application of statistical methods. Awe *et al.* (2019) have identified high uncertainties about real-life emissions from residential units, which is a primary PM's emission carrier. Furthermore, it is difficult to estimate the exact amount of illegal waste combustions in residential units and landfills. It is also challenging to obtain access to PM emission data from individual industrial plants or to assess the pollution input from the old cars and diesel cars commonly deprived of filters. The lack of precise emission inventory data (Dore *et al.*, 2014) and initial air conditions is a significant disadvantage for the use of dynamical models that need reliable chemical and dispersion model outputs. Such a situation results in poor initial conditions for those dynamical models, making them flawed from the start and, as a result, rendering their output error-prone. Additionally, those dynamical models need vast computational resources that makes modelling time-consuming. For that reason, preparing forecasts or giving alerts on time is difficult if not impossible, especially in the case of nowcasting or short-term forecasts.

In this study, we decided to apply machine learning methods based on statistical assumptions. These models are becoming more and more reliable and less time consuming to calibrate. We use the perfect prog methods (Benestad *et al.*, 2008) not only to make PM concentration forecasts, but also to choose which variables are the most decisive for actual air quality. The perfect prog techniques do not take into account errors driven by weather forecasts. It needs to be emphasised that we investigated PM's concentration forecasts only for the nearest 24 hours (1 day ahead). This approach is entirely appropriate since most of the up-to-date numerical short-term weather predictions models provide very reliable data assimilation and yield negligible forecast error within such a temporal horizon (Kendzierski *et al.*, 2018).

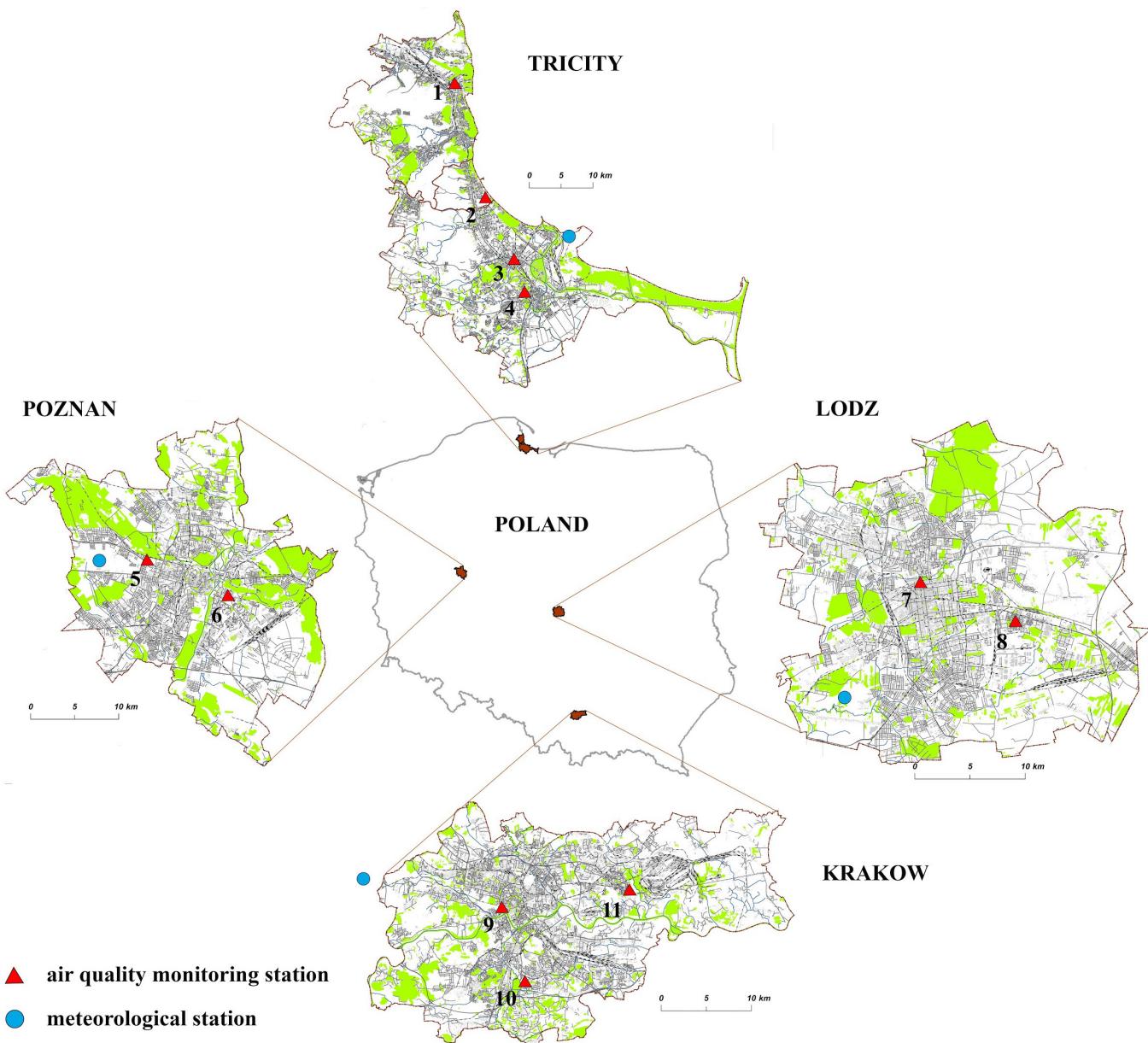
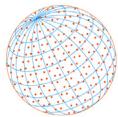
## 2 DATASET AND METHODS

### 2.1 Study Area

Four large urban areas (Kraków, Łódź, Poznań, and Tricity), each with more than 500,000 inhabitants and a population density above 2000 persons km<sup>-2</sup> (except Tricity, ca. 1800 persons km<sup>-2</sup>), were taken into consideration. These cities are located in various parts of Poland (Fig. 1) and are characterised by diverse topography, emission sources, and climate. Tricity, the polycentric urban agglomeration of Gdańsk, Gdynia, and Sopot, is located on the coast of Gdańsk Bay and hemmed in from the north by the Baltic Sea. Poznań is the fifth-largest city in Poland, located mainly on the plateau areas of the Wielkopolskie Lakeland at an elevation of 50–154 m ASL. Łódź is the third-largest city in central Poland on the edge of the Łódź Hills with an elevation ranging from 161.8 to 278.5 m ASL. Kraków is the second most populous city in Poland. It is located in the southern part of the country in the Vistula River valley (extending from west to east in the region of Lesser Poland). The city borders the hills of the Kraków-Częstochowa Upland to the north and Carpathian Foothills to the south, with an elevation ranging from about 200 to 370 m ASL.

### 2.2 Air Quality Data

Daily and hourly ground-level measurements of PM<sub>2.5</sub> and PM<sub>10</sub> for winter from Poznań, Łódź and Kraków were taken from the Voivodeship air pollution networks which, during the investigated period of years 2006–2016 belonged to the Polish Voivodeship Environment Protection Inspectorates. The hourly data for Tricity were obtained from the Agency of Regional Air Quality

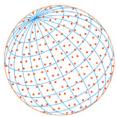


**Fig. 1.** Location of air quality and meteorological stations on the background of selected cities and country boundaries (Jędruszkiewicz *et al.*, 2017).

Monitoring in the Gdańsk Metropolitan Area (ARMAAG). The 11 urban stations (2–4 stations per investigated area) represent the official measurements for background, traffic, and industrial areas, as shown in Table 1. The locations of the air quality monitoring sites and meteorological stations are shown in Fig. 1.

### 2.3 Predictor Variables

Air quality for the winter season in Polish cities is usually shaped by complicated interactions of meteorological and human-related factors (Jędruszkiewicz *et al.*, 2017). With that in mind, both groups of predictors were taken into account to create short-term ML-based forecasting models. The meteorological data (planetary boundary layer, near-surface air temperature, dew point, wind speed and direction, wind gust, sea level pressure, sunshine duration, total and low-level cloudiness, and total precipitation) were acquired from the nearest meteorological stations maintained by the Institute of Meteorology and Water Management – National Research Institute that adheres to the measurement standards of the World Meteorological Organization.

**Table 1.** The locations of the air quality monitoring sites and meteorological stations.

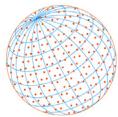
Station name	Coordinates	Altitude (m ASL)	Type	Parameters	Data timeline
Gdańsk-Wrzeszcz	54°22'49", 18°37'13"	7	Background	PM <sub>2.5</sub>	2010–2016
	54°21'12" 18°38'07"			PM <sub>10</sub>	2006–2016
Sopot	54°25'54" 18°34'47"	5	Background	PM <sub>10</sub>	2006–2016
	54°31'31" 18°32'11"			PM <sub>10</sub>	2006–2016
Poznań-Polanka	52°23'53.73" 16°57'34.60"	66	Background	PM <sub>2.5</sub>	Nov 2007–2016
	52°25'13.20" 16°52'38.20"			PM <sub>10</sub>	2006–2016
Łódź-Widzew	19°31'47.23" 51°45'28.98"	238	Background	PM <sub>2.5</sub>	2006–2016
	51°46'39.35" 19°27'09.72"			PM <sub>10</sub>	2006–Oct 2015
Kraków-Krasińskiego	50°03'27.91" 19°55'34.60"	208	Traffic	PM <sub>2.5</sub>	Dec 2009–2016
	50°00'37.88" 19°56'57.17"			PM <sub>10</sub>	2006–2016
Kraków-Kurdwanów	50°04'10.00" 20°03'12.58"	225	Background	PM <sub>2.5</sub>	Mar 2010–2016
	50°04'10.00" 20°03'12.58"			PM <sub>10</sub>	2006–2016
Kraków-Nowa Huta	50°04'10.00" 20°03'12.58"	203	Industrial	PM <sub>2.5</sub> PM <sub>10</sub>	2008–2016 2006–2016

Due to a lack of atmospheric vertical profile data from analysed locations nearby, we decided to use the nearest grid point of ERA-Interim reanalysis (Haylock *et al.*, 2008; Dee *et al.*, 2011) as a data representing the entire dynamics of near-surface air mixing. Among a group of all meteorological predictors, the daily mean, minimum, and maximum values were computed (where applicable). For wind parameters, its vector components (zonal and meridional) were calculated to account for the direction of air mass advection.

Although meteorological parameters are the most relevant in shaping air quality conditions, especially in the wintertime, they usually only emphasise only the effect of human-related emissions. Both phenomena are often correlated (e.g., a drop in air temperature increases emissions from heating systems). However, source inventory, especially for small heating installations, is usually far from being perfect, which causes extra uncertainty for all air quality modelling systems (Dore *et al.*, 2014; Kryza *et al.*, 2018). Moreover, some crucial parameters that potentially could have been used for a better estimation of human activity like traffic anomalies, industrial production, etc. are either not available on a daily scale or are not reliable enough. Therefore, this kind of information normally has to be usually parameterised to be correctly estimated by statistical models. In this study, we wanted to capture this type of factors by incorporating the information about the day of the week, and whether the current and previous day was a holiday or workday.

The last group of predictor variables were PM<sub>10</sub> and PM<sub>2.5</sub> daily means from the previous three days at a given location. Earlier studies have shown (Czernecki *et al.*, 2017) that the autocorrelation signal may diminish the importance of other factors if there is no evident change in other parameters. The 1-day autocorrelation for PM<sub>10</sub> in the winter season may reach (Pearson's correlation) coefficient value of 0.65 and remains statistically significant for even up to 3-day lagged daily mean PM<sub>10</sub> concentration.

In total, 32 predictor variables served to create machine learning models. These comprised 26 meteorological parameters, 3 parameters related to human activity, and 3 autocorrelation signals (Table 2). All investigated predictor variables were initially tested with the Boruta algorithm (Kursa and Rudnicki, 2010) which were proven to exert more impact on the predicted parameter than a random noise for all investigated monitoring stations.

**Table 2.** List of predictor variables used in this study.

Parameter	Co-products	Abbreviation(s)	Units
Planetary boundary layer	Daily mean, min, max	PBL_avg, PBL_min, PBL_max	m
Air temperature at 2 m AGL	Daily mean, min, max	T2_avg, T2_min, T2_max	°C
Dew point temperature at 2 m AGL	Daily mean, min, max	DPT_avg, DPT_min, DPT_max	°C
Wind speed at 10 m AGL	Daily mean, min, max	WS_avg, WS_min, WS_max	m s <sup>-1</sup>
Zonal wind component at 10 m	Daily mean, min, max	U10_avg, U10_min, U10_max	m s <sup>-1</sup>
Meridional wind component at 10 m	Daily mean, min, max	V10_avg, V10_min, V10_max	m s <sup>-1</sup>
Sea-level pressure	Daily mean	SLP	hPa
Sunshine duration	Daily total	SUN	hours
Precipitation	Daily total	PREC	mm
Total and low cloud cover	Daily mean	TOT_CL, LOW_CL	octants
Wind gust	Daily max	GUST	m s <sup>-1</sup>
Day of week	–	DOW	factor
Working day/holiday	–	HOLIDAY	factor
PM <sub>10</sub> concentration 1–3 days earlier	Daily mean	PM10_lag1, PM10_lag2, PM10_lag3	µg m <sup>-3</sup>

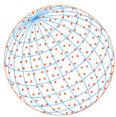
## 2.4 Machine Learning Models

A few commonly applied statistical machine learning methods, including linear regression with stepwise selection, random forests, Extreme Gradient Boosting regression trees, and multilayer perceptron neural networks were tested and evaluated against the measured PM<sub>10</sub> and PM<sub>2.5</sub> values, separately for each station. Repeated leave-group-out cross-validation (LGOCV; also known as *Monte Carlo cross-validation*) was used to prevent model overfitting and provide the best set of hyperparametric values. The randomly selected (without replacement) fraction of data to form the training set contained 66% of data, which was then used to create the model and test it on the rest of the data points for a chosen station (i.e., LGOCV approach was applied separately for every station). Subsequently, this process was repeated five times, generating new training and test partitions each time for the newly evaluated model. The final regression model for every ML method and station was the one with the lowest root-mean-square error (RMSE) among all performed iterations.

The selected variables were scaled and transformed according to the Yeo-Johnson transformation to fit into the methodological assumptions of the chosen models. In the case of a neural network-based model, the min-max scaling was performed (where applicable). In the case of the multiple regression, the selection of the best predictors that would subsequently result in the optimized model was conducted with the application of the stepwise Akaike Information Criterion (AIC) algorithm (Akaike, 1973). It is similar to ordinary stepwise multiple regression but with the difference that instead of the F-statistics, the value of the AIC determines the selection of the best combination of possible features. All the calculations were carried out using the R programming language (R Core Team, 2019) as well as packages for machine learning modelling included in the caret framework (Kuhn *et al.*, 2019).

Two of the selected ML models represent decision tree-based algorithms. The first is the random forest approach, which is a non-parametric machine learning algorithm that could be used to predict with high accuracy (Liu *et al.*, 2018). RF consists of a collection of classifiers with a tree structure. These classifiers are randomly and independently selected vectors with the same distribution that vote for the most popular class (Breiman, 2001). RF models have been successfully used to predict PM<sub>2.5</sub> regardless of the climatic zone (Hu *et al.*, 2017; Joharestani *et al.*, 2019). The chosen RF hyperparameters used in this study vary for each model as the calibration was done separately for each station. The general rule was to create up to 500 trees in every cross-validation procedure with splitting variables at each tree node according to the square root of possible combinations while maximising the obtained variance (Kuhn *et al.*, 2019).

Another decision tree algorithm represented here is the Extreme Gradient Boosting framework based on the gradient boosting technique. The main differences between XGBoost and RFs are how the trees are built (XGBoost—one tree at a time, RFs—trees are independent). In the case of XGBoost, there is also an adapted function loss to improve weak learners and combine the



obtained results in a stage-wise manner. Therefore, XGBoost tends to outperform RFs only if data between classes are relatively balanced, to the extent that there is not much noise in the data, and when tuned parameters are carefully set up. The last issue was particularly challenging in our study, as the default settings tend to overfit the final model easily. However, the finally obtained hyperparameters with grid search were similar to those presented in other studies dealing with particulate matter concentrations (Joharestani *et al.*, 2019) with slightly smaller `n_rounds` parameter ( $\sim 150$ ) and higher `eta` ( $= 0.3$ ).

Another possible approach in modelling of nonlinear systems is the utilisation of **artificial neural networks**. **ANNs** are commonly considered to be the most advanced AI tool which allow for the analysis of multidimensional systems. Here, we used a multilayer perceptron with two hidden layers yielding a total of four layers. Resilient back propagation (Riedmiller and Braun, 1993; Riedmiller, 1994) was selected as an error propagation method. We tested multiple network architectures in an attempt to find the optimum number of neurons in hidden layers. The first layer size equals the number of independent variables, whereas the last one is a neuron providing final information on the daily average PM concentration, given that the regression approach was applied. Hidden layers' structure was a subject of tuning, and in the case of the second layer, we tested sizes from 10 to 20. In the third one, we investigated the models' quality using 3–7 neurons. As in the case of other ML schemes, here also we used an LGOCV cross-validation scheme to assess the network architectures.

## 2.5 Verification Measures

### 2.5.1 Regression

The models' performance for the regression approach was characterised using **Pearson's correlation (r)**, **coefficient of determination ( $R^2$ )**, **root-mean-square error (RMSE)** and **mean absolute error (MAE)** following formulas presented by Wilks (2008).  $R^2$  provides information on the dependant variable's fraction of variance as explained by the model. RMSE reflects the average difference between predicted and observed values (average model's error), whereas MAE is the mean absolute error. Additionally, the standard deviation (SD) of the model's performance statistics was calculated to account for variance inflation problems that often occur while applying statistical modelling techniques (Maraun, 2013). The verification statistics for the Pearson correlation coefficients, RMSEs and SDs were additionally presented in the form of Taylor diagrams (Taylor, 2001), which allowed us to address different aspects of the models' accuracy on a single chart.

### 2.5.2 Binary events

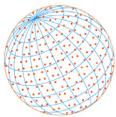
The numerical values of forecasted  $PM_{10}$  and  $PM_{2.5}$  concentrations obtained via the regression approach were classified into a binary form for situations according to daily threshold limits set by the Ambient Air Quality Directive (EU, 2008) ( $50 \mu\text{g m}^{-3}$  for  $PM_{10}$  and  $25 \mu\text{g m}^{-3}$  for  $PM_{2.5}$ ). The scheme of the contingency table adopted in this study follows Wilks's convention (2008) with observed and forecasted events split into four categories: (a) a number of correct hits, (b) false alarms, (c) misses, and (d) correct rejections.

Based on the created contingency tables, five verification measures were calculated to assess the robustness of the model (Table 3) for days that crossed the allowable  $PM_x$  concentration values. The most fundamental ones were chosen, including the **probability of detection (POD)**, **proportion correct (PC)**, **false alarm ratio (FAR)** and **success ratio (SR)**. Based on FAR and SR values,

**Table 3.** Formulas and ranges of selected verification measures used in this study, according to Jolliffe and Stephenson (2012).

Measure (abbreviation)	Formula/definition	Range
Probability of detection (POD)	$POD = a/(a + c)$	0 (no skill), 1 (perfect model)
Proportion correct (PC)	$PC = (a + d)/n$ , where $n = a + b + c + d$	0 (no skill), 1 (perfect model)
False alarm ratio (FAR)	$FAR = b/(a + b)$	[0, 1]
Success ratio (SR)	$SR = 1 - FAR$	[0, 1]
Heidke Skill Score (HSS)	$HSS = 2(ad - bc)/[(a + c)(c + d) + (a + b)(b + d)]$	$-\infty$ (no skill), 1 (perfect model)

*a*—number of correct hits, *b*—number of false alarms, *c*—misses, and *d*—correct rejections.



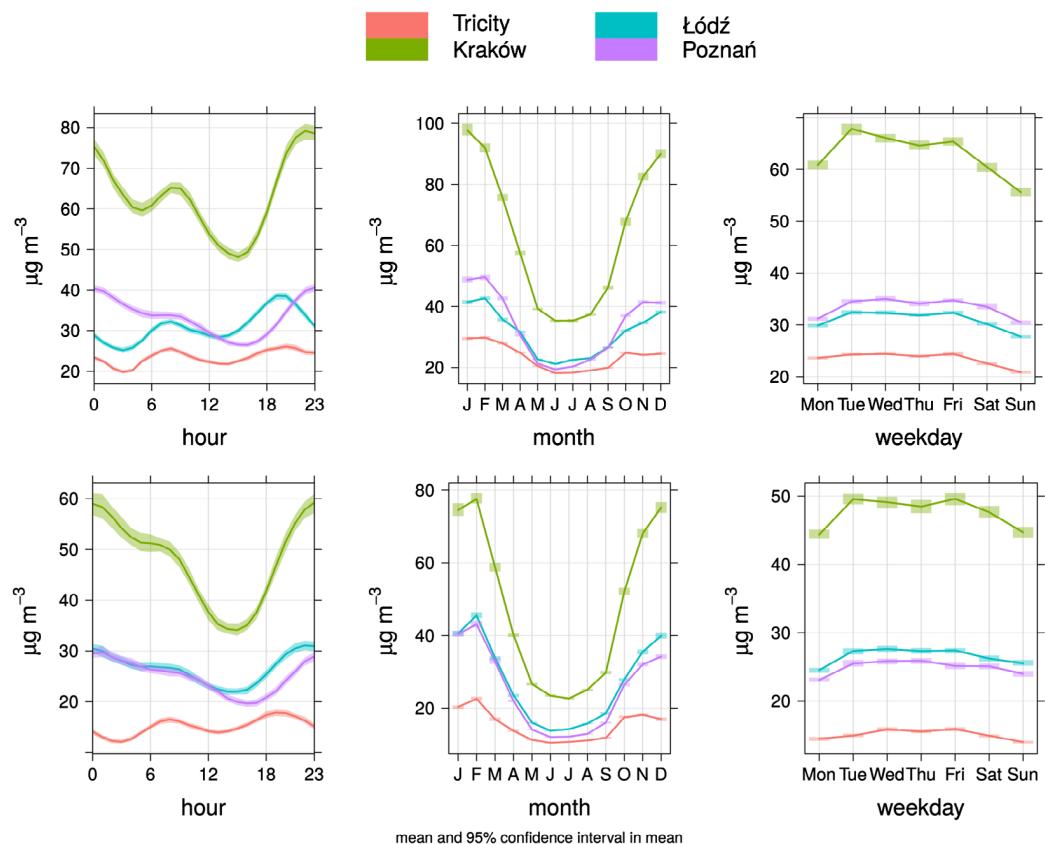
the performance diagrams (Roebber, 2009) were created to summarise model performances for those two complimentary indices.

In some of the analysed locations' binary, events tend to be heavily unbalanced. For example, in Kraków there is a clear predominance of situations with bad air quality which makes it relatively easy for a forecaster to issue the correct binary forecast (assuming predominance of typical "bad air quality" conditions). The opposite situation with a relatively low share of bad air quality days as in Tricity also eases binary forecasting. Therefore, the Heidke Skill Score (HSS) was calculated in order to account for this kind of situation. The HSS returns fractional improvements over random chance and therefore makes it a more robust single measure for determining a model performance (Jolliffe and Stephenson, 2012). All formulas for binary events used in this study and its interpretability are presented in Table 3.

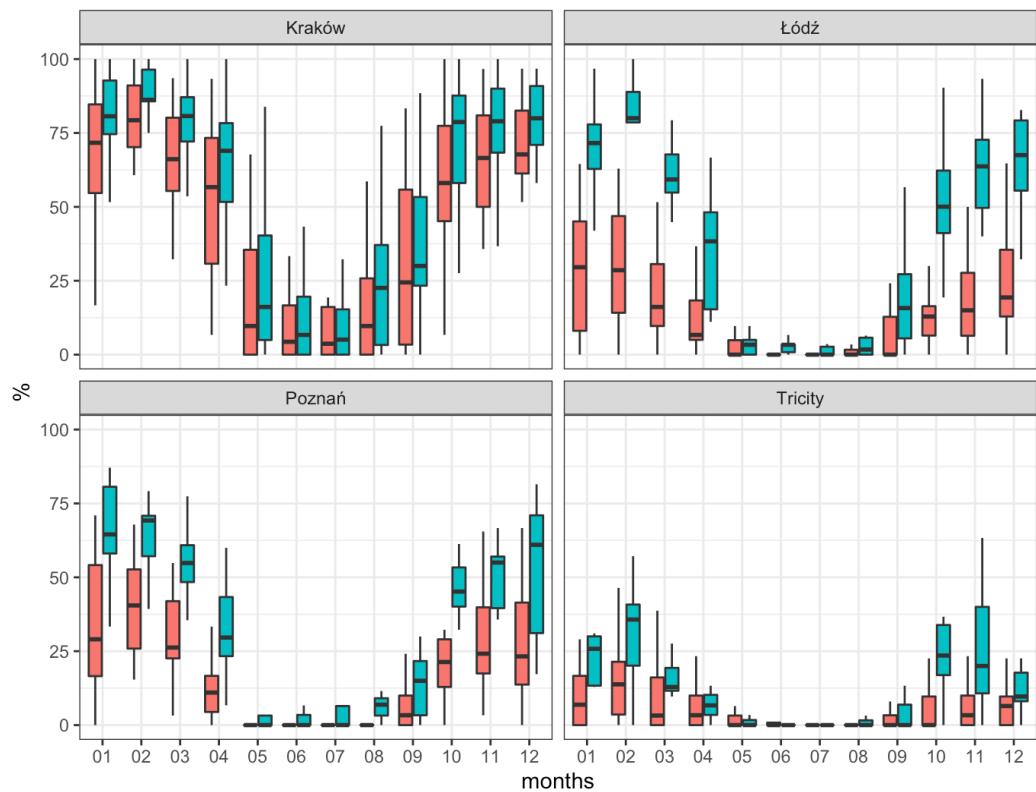
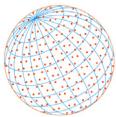
## 3 RESULTS

### 3.1 PM<sub>10</sub> and PM<sub>2.5</sub> Concentration Variability

There are clear seasonal, weekly, and daily cycles visible in the PM<sub>2.5</sub> and PM<sub>10</sub> concentrations among Polish agglomerations related to changes in atmospheric conditions and human-related emissions. These regularities are also confirmed in the analysed data (Fig. 2). The courses of PM<sub>2.5</sub> and PM<sub>10</sub> concentrations are very similar regardless of the analysed location, with PM<sub>10</sub> concentration generally higher and PM<sub>10</sub>-to-PM<sub>2.5</sub> ratio changing according to season. On average, the lowest concentration of PM<sub>10</sub> was found in the Tricity (25 µg m<sup>-3</sup>), then Łódź and Poznań (32–35 µg m<sup>-3</sup>), with the highest in Kraków (64–68 µg m<sup>-3</sup>). The ratio of PM<sub>10</sub> to PM<sub>2.5</sub> is the lowest in the summer months, where it contributes 45–65% of PM<sub>10</sub> concentrations, with these values rising to 70–90% in winter months. Also in wintertime, the most unfavourable air condition occurred with



**Fig. 2.** Annual variability of PM<sub>10</sub> (upper panel) and PM<sub>2.5</sub> (lower panel) concentrations in Gdańsk, Poznań, Łódź and Kraków, 2006–2016. Calculated as a mean of available data for the stations listed in Table 1.



**Fig. 3.** Frequency of daily PM<sub>10</sub> means above 50 µg m<sup>-3</sup> (red) and daily PM<sub>2.5</sub> above 25 µg m<sup>-3</sup> (blue) in Kraków, Łódź, Poznań and Tricity, 2006–2016.

the concentration of PM<sub>10</sub> from 25–30 µg m<sup>-3</sup> in Tricity, 42–50 µg m<sup>-3</sup> in Poznań, 38–42 µg m<sup>-3</sup> in Łódź and up to 98 µg m<sup>-3</sup> in Kraków. The most favourable air quality conditions are in summer when daily mean PM<sub>10</sub> concentrations range from 19 µg m<sup>-3</sup> in Tricity to 36 µg m<sup>-3</sup> in Kraków, while PM<sub>2.5</sub> concentration ranges from 12 µg m<sup>-3</sup> to 22 µg m<sup>-3</sup>, respectively.

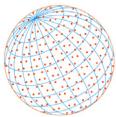
In the weekly cycle, the highest values were observed from Wednesday to Friday. In contrast, significantly better air quality conditions are typical for weekends on account for lower traffic and industrial activity (Juda-Rezler, 2006). At the weekend, the air quality is improved by about 10–20 µg m<sup>-3</sup> due to less traffic and people leaving the city for a weekend away. Also, the autocorrelation signal plays an essential role in a weekly cycle, which is visible in terms of lower particulate matter values on Mondays than any other working days.

In the sub-daily scale, two maxima (morning, late evening/night) and minima (around 3:00–5:00 AM, early afternoon) were distinguished. For the most part, this is related mainly to the diurnal cycle of air temperature and Planetary Boundary Layer (PBL), as much as daily cycle of human activity.

From the standpoint of human health, the most important is the time when threshold limits set by the Ambient Air Quality Directive (EU, 2008) were exceeded (50 µg m<sup>-3</sup> for PM<sub>10</sub> and 25 µg m<sup>-3</sup> for PM<sub>2.5</sub>). The monthly ratio of days when the limits set for PM<sub>10</sub> and PM<sub>2.5</sub> were exceeded on each station is presented in Fig. 3. Generally, it was found that the PM<sub>10</sub> limit exceedance occurred mainly from October to April–May (heating season). In the Tricity, it is about 10–30% of days, in Łódź and Poznań about 10–40%, and in Kraków 50–90%. However, these values are highly variable year to year due to the meteorological conditions and stations' locations. The situation is quite similar in the case of PM<sub>2.5</sub> concentration but with a significantly higher number of days with exceedance due to lower threshold value as presented in Fig. 3.

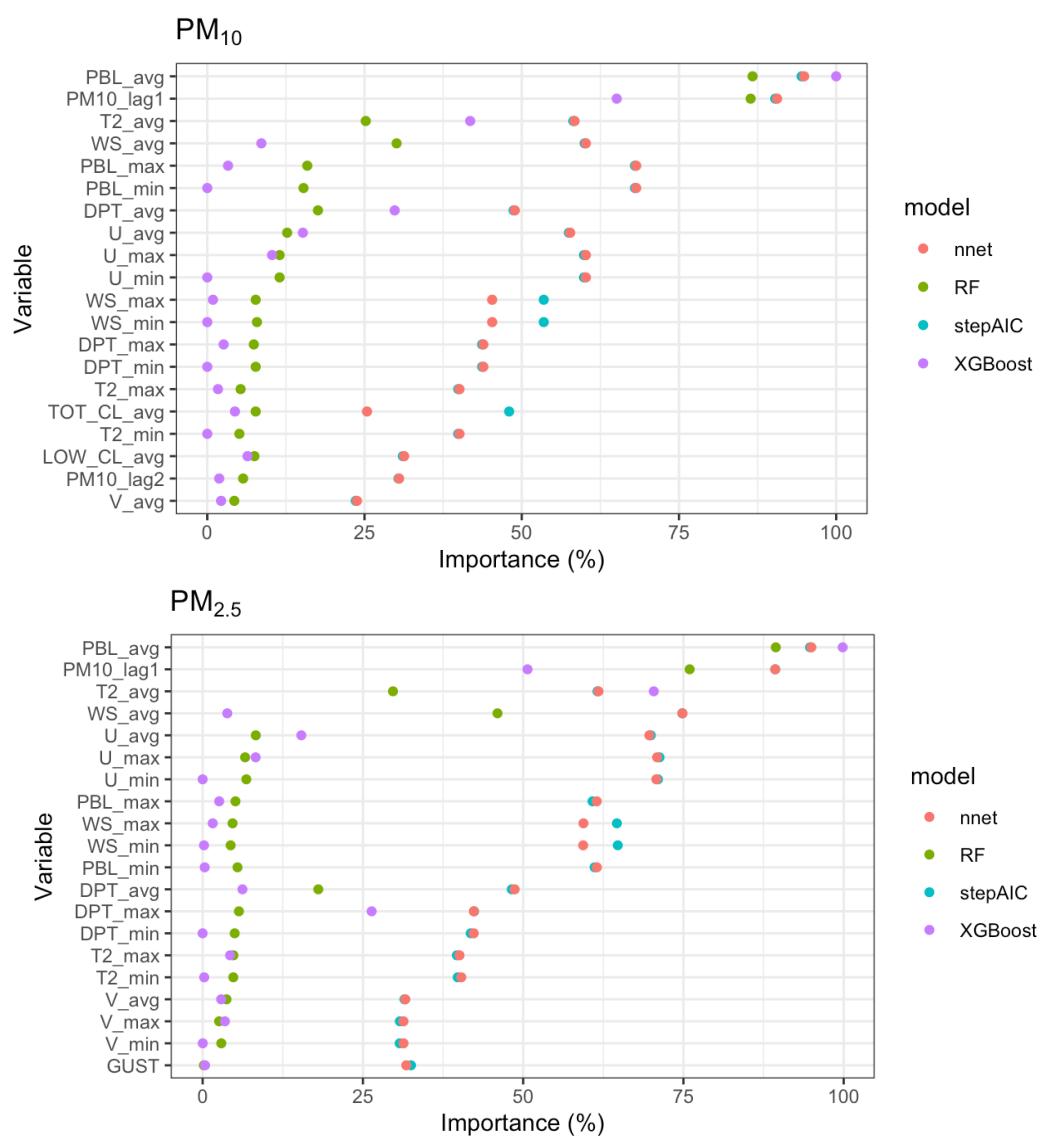
### 3.2 The Variable Importance of Predictors

In our study, we opted to use the meteorological parameters with the greatest impact on high PM<sub>x</sub> concentrations according to previous studies (Jędruszkiewicz *et al.*, 2017). In the cold half of

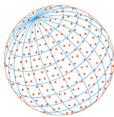


the year, the low temperature (below +5°C), thin PBL (i.e., below 300–500 m), stable atmosphere, anticyclonic weather, and low wind speed (below 2 m s<sup>-1</sup>) played the most important role in the accumulation of pollutants. In terms of Pearson's correlation coefficient, the highest absolute numbers against PM<sub>10</sub> concentrations were obtained for the daily means of PBL (from  $r = -0.76$  in Poznań to  $r = -0.65$  in Gdańsk), air temperature (from  $r = -0.56$  in Poznań to  $r = -0.43$  in Kraków), and wind speed (from  $r = -0.62$  in Łódź to  $r = -0.42$  in Gdańsk). Other meteorological parameters were below the absolute Pearson correlation value of 0.50. The presence of high-pressure systems in winter months along with relatively low dynamics characteristic for such a sea-level pressure pattern leads to situations when high PM<sub>x</sub> concentrations tend to persist for a few days in a row. This also allows us to include the autocorrelation signal of previous PM<sub>x</sub> concentration and treat it like a starting point for a model prediction. The autocorrelation signal reaches up to 0.60–0.70 of the Pearson's correlation coefficient which confirms that this is often a crucial factor that impacts air quality in European cities (Draghicescu and Ignaccolo, 2005; Holst *et al.*, 2008).

The main factors which favoured high PM<sub>x</sub> concentrations and had been previously confirmed in other studies on Polish cities were also the most significant ones for the created ML models. In almost all cases, the most important determinants were: average height of the PBL and 1-day-lagged PM<sub>x</sub> concentration, followed by the mean air temperature and wind speed (Fig. 4). There



**Fig. 4.** Mean variable importance (scaled) of predictors used to create ML models for daily PM<sub>10</sub> (upper panel) and PM<sub>2.5</sub> (lower panel) concentration in winter months.



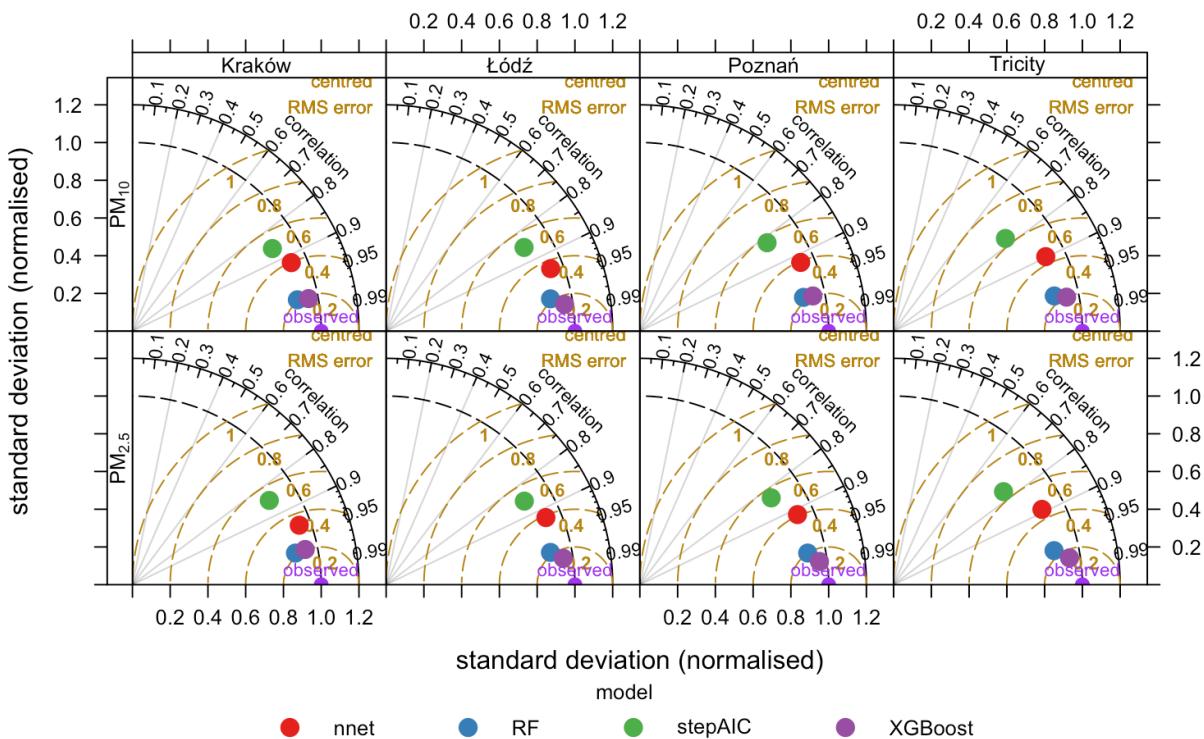
is also a clearly visible tendency for most models to benefit mainly from PBL and autocorrelation predictors, with XGBoost being an exception as it uses only about 50% of importance for the 1-day-lagged autocorrelation signal compared to changes in PBL height. Another thing worth mentioning is that decision tree methods tend to focus only on 4–5 main predictors which gives more than 25% of the importance signal and tend to ignore most of the other factors, with variable importance for nnet and stepAIC methods changing smoothly among all predictors. This is probably related to the high cross-correlation between variables and the diminishing significance of other parameters in decision tree models that even if capable of adding value to the output, they are eliminated due to redundancy or partial signal overlap with other variables. This can be seen in the example of wind gusts (which are potentially included in changes to other wind components), or precipitation (causing wet scavenging) which is strongly correlated with simultaneous changes, for example, in dew point and air temperature. It is worth noting that the human-related factors such as (non-)working day or day of week were considered unimportant in most cases, even if it was classified higher than, for instance, the importance of precipitation's occurrence.

### 3.3 Machine Learning-driven Short-term Forecasts

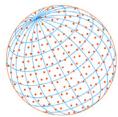
As mentioned above, the models' quality assessment involved a twofold approach. The first utilised the pure regression approach, where we attempted to forecast  $\text{PM}_{10}$  and  $\text{PM}_{2.5}$  concentrations. The latter involved a categorical forecast where administratively regulated thresholds were applied to identify the  $\text{PM}_x$  concentrations exceedances, which yielded the confusion matrices for each city and model, and in turn allowed for the calculation of quality metrics.

### 3.4 Regression Forecast

The results clearly show that there is a significant and consistent improvement depending on the type of algorithm used (Fig. 5). The sequence of models (from worst to best) is stepAIC, nnets, RFs, and finally XGBoost. What may be striking is a relatively poor performance of neural networks compared to decision tree methods (i.e., RFs and XGBoost). **It would seem that in this case, the size of the learning sample might be a problem as neural networks usually work best when provided with more data.** The shift in quality between RF and XGBoost algorithms is small,



**Fig. 5.** Taylor diagrams of daily  $\text{PM}_{10}$  (upper panel) and  $\text{PM}_{2.5}$  (lower panel) in winter months for created ML models in Kraków, Łódź, Poznań and Tricity, 2006–2016.



and in the case of PM<sub>10</sub> it is mainly indicated by a better variance reproduction and only slight improvement in RMSE (Table 4). In the case of PM<sub>2.5</sub> forecasts in all agglomerations (except Kraków) the difference between RFs and XGBoost is apparent for both RMSE and variance reproduction. The correlation coefficients between the observed and forecasted series of PM<sub>x</sub> are generally very high. The lowest values (Tables 4 and 5) occur for the stepAIC algorithm and vary between 0.77 and 0.86 (PM<sub>10</sub>) and between 0.76 and 0.86 (PM<sub>2.5</sub>). One might suggest that taking the overall abilities of this algorithm into consideration, those values are satisfactory. Much higher values of correlation coefficients occur for the other three applied algorithms. In nearly all cases, values exceed 0.9, which indicates a high ability to represent the variability of the modelled variable. In both cases (PM<sub>10</sub> and PM<sub>2.5</sub>), the XGBoost algorithm provided the highest correlation values (ca. 0.98). A high concordance in variability for both RF and XGBoost algorithms is further confirmed by the RMSE values (Tables 4 and 5).

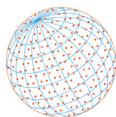
As was the case for the correlation coefficients, the poorest values occur for the stepAIC algorithm where the average deviation from the forecast usually exceeds 10 µg m<sup>-3</sup>, with values as high as 32.8 µg m<sup>-3</sup> for PM<sub>10</sub> in Kraków. The quality of the forecast via the neural network

**Table 4.** PM<sub>10</sub> forecast models' quality metrics in selected agglomerations in Poland.

City	Model	r	RMSE	MAE	POD	FAR	PC	HSS
Tricity	nn	0.898	10.2	7.0	0.628	0.259	0.936	0.644
Tricity	ranger	0.977	5.5	3.3	0.806	0.112	0.968	0.827
Tricity	stepAIC	0.770	14.8	9.7	0.503	0.412	0.908	0.491
Tricity	XGBoost	0.981	4.6	3.4	0.801	0.065	0.972	0.847
Kraków	nn	0.918	25.5	18.6	0.960	0.094	0.899	0.734
Kraków	ranger	0.982	13.5	9.2	0.994	0.062	0.948	0.863
Kraków	stepAIC	0.861	32.8	23.6	0.949	0.111	0.877	0.675
Kraków	XGBoost	0.983	11.9	8.7	0.983	0.050	0.950	0.872
Kraków	nn	0.935	9.3	6.7	0.850	0.161	0.904	0.775
Łódź	ranger	0.981	5.6	3.9	0.925	0.087	0.950	0.883
Łódź	stepAIC	0.855	13.5	9.9	0.843	0.260	0.862	0.687
Łódź	XGBoost	0.989	4.0	2.9	0.925	0.071	0.956	0.895
Poznań	nn	0.919	13.8	9.9	0.844	0.211	0.871	0.716
Poznań	ranger	0.979	7.9	4.7	0.940	0.093	0.947	0.883
Poznań	stepAIC	0.821	20.1	13.8	0.869	0.282	0.840	0.660
Poznań	XGBoost	0.980	7.2	5.1	0.917	0.104	0.936	0.857

**Table 5.** PM<sub>2.5</sub> forecast models' quality metrics in selected agglomerations in Poland.

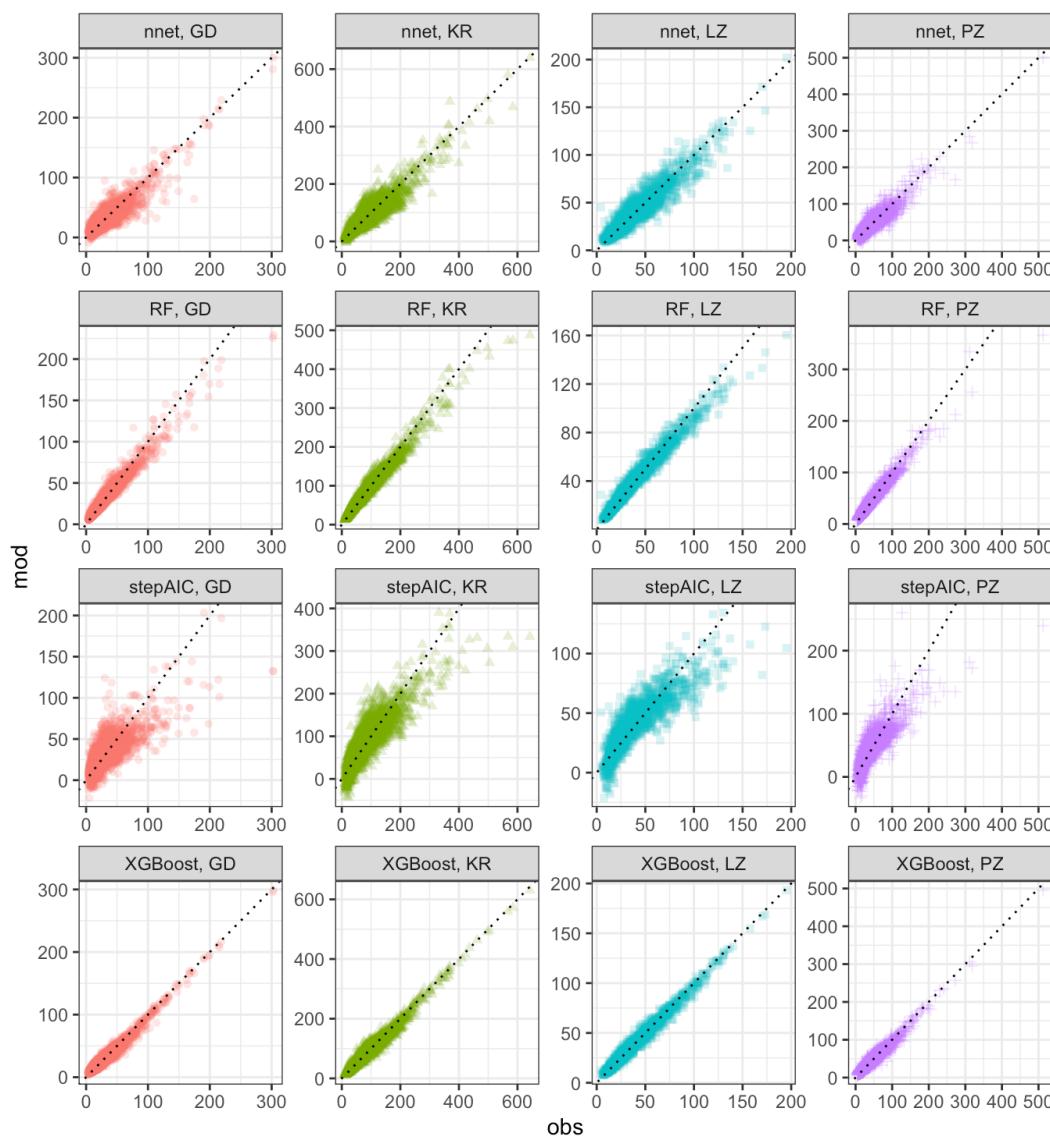
City	Model	r	RMSE	MAE	POD	FAR	PC	HSS
Tricity	nn	0.892	6.3	4.4	0.615	0.200	0.977	0.684
Tricity	ranger	0.978	3.3	2.1	0.692	0.053	0.985	0.793
Tricity	stepAIC	0.763	9.0	6.4	0.154	0.200	0.963	0.248
Tricity	XGBoost	0.989	2.2	1.6	0.846	0.000	0.993	0.913
Kraków	nn	0.942	15.0	10.7	0.909	0.120	0.882	0.762
Kraków	ranger	0.982	9.6	6.7	0.971	0.098	0.927	0.851
Kraków	stepAIC	0.851	23.4	17.5	0.929	0.196	0.837	0.666
Kraków	XGBoost	0.980	9.1	6.4	0.961	0.088	0.928	0.853
Łódź	nn	0.922	9.9	7.4	0.782	0.194	0.882	0.711
Łódź	ranger	0.981	5.5	3.8	0.903	0.069	0.952	0.884
Łódź	stepAIC	0.856	13.2	10.0	0.812	0.272	0.857	0.665
Łódź	XGBoost	0.989	3.9	3.0	0.909	0.068	0.954	0.888
Poznań	nn	0.914	11.8	8.7	0.785	0.215	0.886	0.708
Poznań	ranger	0.983	5.8	4.0	0.939	0.065	0.967	0.914
Poznań	stepAIC	0.834	16.0	12.1	0.780	0.264	0.868	0.666
Poznań	XGBoost	0.992	3.9	2.9	0.939	0.038	0.974	0.933



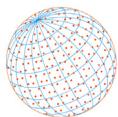
algorithm is better but compared to RFs and XGBoost it is also not satisfactory. RMSE values for neural network algorithms range from  $9.3 \mu\text{g m}^{-3}$  to  $25.5 \mu\text{g m}^{-3}$  ( $\text{PM}_{10}$ ) and from  $6.3 \mu\text{g m}^{-3}$  to  $15.5 \mu\text{g m}^{-3}$  ( $\text{PM}_{2.5}$ ).

RF and XGBoost models performed fairly well with comparable RMSE values ranging from  $4.0 \mu\text{g m}^{-3}$  (XGBoost, Łódź) to  $13.5 \mu\text{g m}^{-3}$  (RFs, Kraków) for  $\text{PM}_{10}$ . In the case of  $\text{PM}_{2.5}$ , the differences in RMSE between those two methods are negligible and do not exceed  $2 \mu\text{g m}^{-3}$ . RMSE itself varies from  $2.2 \mu\text{g m}^{-3}$  (XGBoost, Gdańsk) to  $9.6 \mu\text{g m}^{-3}$  (RFs, Kraków). When looking at the differences between agglomerations, Kraków stands out negatively with RMSE values being usually more than twice as high as those in other cities.

Scatter plots of observed and forecasted  $\text{PM}_x$  concentrations (Fig. 6) provide additional insight into the abilities of the algorithms within the context of a whole range of predicted values. The stepAIC method is the least capable forecaster of high  $\text{PM}_{10}$  concentrations. This is visible, especially for the higher concentrations where underestimation of  $\text{PM}_{10}$  concentrations might exceed locally even up to  $200 \mu\text{g m}^{-3}$  (Kraków). For other cities, the scale of underestimation of the highest concentration levels with stepAIC is lower and usually does not exceed  $100 \mu\text{g m}^{-3}$ . For lower concentrations, this algorithm shows a much higher over- and underestimation compared to the other methods. This is to be expected as the linear methods usually perform poorly when modelling highly non-linear processes. Other ML algorithms utilised in this research



**Fig. 6.** Observed vs. forecasted  $\text{PM}_{10}$  concentrations in selected agglomerations in Poland via the usage of ML algorithms.



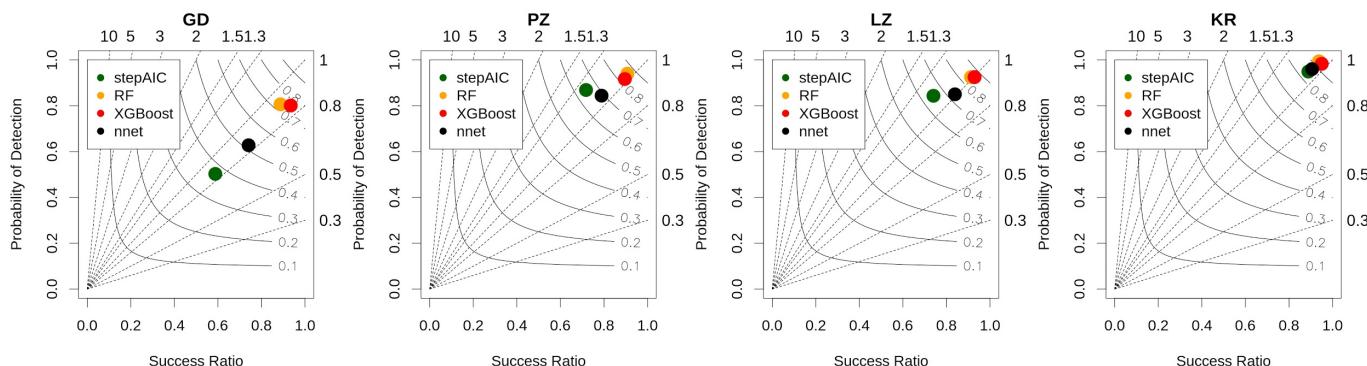
are more consistent with the linear shape of observed vs. forecasted scatter plots. The RF algorithm forecasts tend to underestimate higher PM<sub>10</sub> concentrations although this usually does not exceed 50 µg m<sup>-3</sup>. Neural networks, despite showing considerably higher spread along the range of PM<sub>10</sub> concentrations, seem to be more consistent for the higher levels of PM<sub>10</sub> concentration than RF. XGBoost algorithm shows the highest consistency. The spread is minimal across the whole range of observed PM<sub>10</sub> concentrations, while the minute underestimation is visible only in the case of Poznań.

### 3.5 Binary Forecast

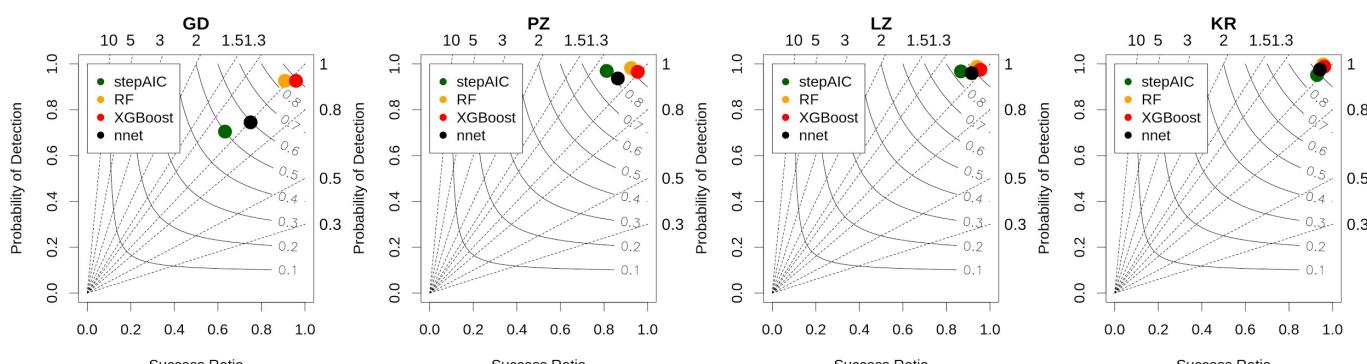
The second approach with a categorical forecast of the threshold exceedance for PM<sub>x</sub> concentrations yields similar outcomes when it comes to the hierarchical quality of ML algorithms applied in PM short-term forecasting. In all cases, the worst performance is for the stepAIC algorithm, then neural networks, and finally RFs and XGBoost with quality metrics at relatively similar levels (Figs. 7 and 8). What seems striking is the apparent variability between agglomerations. In the case of northern Poland and Tricity, there is a rather significant shift in forecast model metrics. For example, Success Ratio (defined as 1-FAR) changes from 0.59 (nnet) to 0.95 (XGBoost). It is also apparent for POD increase—from 0.50 (stepAIC) to 0.81 (RFs). The least improvement in case of categorical PM<sub>10</sub> forecasts occurs in Kraków where all used ML algorithms perform quite well with POD over 0.9 and Success Ratio near or over 0.9. In the case of PM<sub>2.5</sub>, a similar pattern emerges. However, in Poznań and Łódź the improvement appears for the Success Ratio metric (from ca. 0.75 to 0.97) with the POD at relatively similar levels. In Kraków, the differences in the quality metrics are minimal, and all the models perform very well in terms of POD and SR.

## 4 DISCUSSION AND CONCLUSIONS

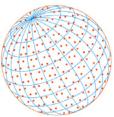
Unsurprisingly, air quality forecasts for Polish agglomerations have garnered increasing public attention—to the point of becoming a common element in national and local weather forecasting



**Fig. 7.** Performance diagrams for forecasts of PM<sub>10</sub> exceedance at selected stations in Poland, 2006–2016.



**Fig. 8.** Performance diagrams for forecasts of PM<sub>10</sub> exceedance at selected stations in Poland, 2006–2016.



services during the winter recently—as they warn residents of PM<sub>10</sub> and PM<sub>2.5</sub> exceedances and the consequent restrictions, e.g., bans on burning in fireplaces (when other heating systems are available), limits on outdoor time for children, and reduced services on free public transport. In most cases, these decision-making systems rely on numerical weather prediction models coupled with a chemistry module and often suffer from imperfect emission inventory input on a local scale (Dore *et al.*, 2014). Hence, this study applied the perfect prog approach to machine learning models for short-term air quality forecasting (Gutierrez *et al.*, 2019).

Of the ML models we tested, the decision tree algorithms—RFs and XGBoost—unquestionably exhibited the best performance in simulating PM regardless of the analysed location or particle diameter. According to the administrative threshold, the probability of detecting days with poor air quality varied from 0.80 in Tricity to 0.98 in Kraków (Tables 4 and 5), whereas the false alarm ratio varied from 0.05 to 0.11 for these algorithms. The more informative Heidke Skill Score, which factors the frequency of situations with bad or good air quality into its calculation, was relatively stable, with values of 0.83–0.88.

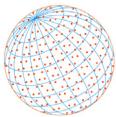
Depending on the location of the sampling station, we obtained Pearson correlation coefficients of 0.97–0.99 and RMSE values of 4.0–13.5 and 2.2–9.6 for the PM<sub>10</sub> and PM<sub>2.5</sub>, respectively, with the regression models. However, the corresponding statistics for the nnet models were significantly lower ( $r = 0.89\text{--}0.94$ ), which we attributed to the relatively small sample size—an assumption that was confirmed during the cross-validation tests, as models that have received larger amounts of data during training frequently display overfitting. Multiple linear regression using stepwise AIC achieved the worst results in all of the analysed cases, with  $r$  ranging from 0.76 (for PM<sub>2.5</sub> in Tricity) to 0.87 (for PM<sub>10</sub> in Kraków) and the RMSE ranging from 8.0 (for PM<sub>2.5</sub> in Tricity) to 32.8 (for PM<sub>10</sub> in Kraków) (Tables 4 and 5). As illustrated in the Taylor diagrams (Fig. 5), all of the models demonstrated high variance flattening, and the reconstructed standard deviation values were always lower than the actual ones—a persistent issue with statistical modelling (Maraun, 2013).

Nevertheless, in our opinion, the normalised standard deviation of the forecasted time series, despite being approximately 10–15% lower than the observed values, substantiates the high applicability of the XGBoost, RF, and nnet ensemble. Specifically, based on two rationales, we find that our verification of the regression models surpasses that performed in other studies on PM<sub>x</sub> forecasting with ML (e.g., Nidzgorska-Lencewicz, 2018; Joharestani *et al.*, 2019).

First, the majority of the previous research input instantaneous or hourly values from their datasets, whereas we targeted daily averages. Not unexpectedly, predicting weather conditions or anthropogenic emissions related to routine activity is significantly easier on a daily than a sub-daily scale. Additionally, PM<sub>x</sub> concentrations typically exhibit seasonal consistency, with rather gradual day-to-day shifts, and sharp, sporadic changes tend to be closely correlated with meteorological events.

Second, we applied perfect prog, which is based on a quality-controlled observational and reanalysed dataset. The models primarily benefitted from the PBL height, 1-day-lagged air quality dataset, and, to a lesser degree, temperature and wind data (Fig. 4), which may prove to be a challenge in real-time application, especially with regard to the PBL, since the typical RMSE for this parameter in a 1-day forecast is approximately 250 m (Lavers *et al.*, 2019). Furthermore, the RMSE values for the hourly air temperature and wind speed are approximately 2.3°C and 1.5 m s<sup>-1</sup>, respectively (Kendzierski *et al.*, 2018), although they are significantly better for daily aggregates (e.g., < 1°C for the daily mean of the near-surface air temperature). The variable importance of the precipitation is very low; therefore, wet scavenging processes, even though they are often considered the most challenging variable in weather forecasting models, produce negligible effects on the ML models' predictions.

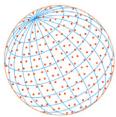
The perfect prog approach we used assumes a constant relationship between the predictors and the predictands, as diagnosed from the training dataset (Benestad *et al.*, 2008). Therefore, one must keep in mind that flawed meteorological data, changes to the PM<sub>x</sub> measurement devices or the stations' locations, or any other factors that decrease the homogeneity of the input data—including long-term strategies that reduce anthropogenic emissions, such as thermal modernisation or clean transportation programs—may negatively affect the predictive accuracy. Although these issues may need to be considered in real-world air quality forecasting, they were not detected during our period of analysis, nor did they impact the models in a statistically significant way.



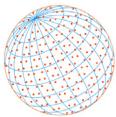
The tested machine learning models have proven to be a useful tool for forecasting short-term changes in air quality in the selected Polish agglomerations and thus can be utilized in air quality management as a component of any city's warning system. The current state of ML modelling offers cost- and computationally effective solutions that can be coupled with numerical weather prediction systems for a real-time hybrid approach without requiring a detailed emission inventory. In the future, we plan to extend the time scale of the predictions to at least three days.

## REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, I.F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest.
- Anderson, J.O., Thundiyil, J.G., Stolbach, A. (2012). Clearing the air: A review of the effects of particulate matter air pollution on human health. *J. Med. Toxicol.* 8, 166–175. <https://doi.org/10.1007/s13181-011-0203-1>
- Awe, Y.A., Murisic, M., Koziel, A., Wolsczak, G., Kochan, F.P., Green, J.M., Golub, E.S., Oguz, K.O., Kim, H.J., Kanakia, R., Brody, M.S., Amman, M., Cofala, J., Kiese wetter, G., Klimont, Z., Wagner, F., Borken, J., Juda-Rezler, K., Maciejewska, K., Reizer, M. (2019). Air Quality Management in Poland (English). World Bank Group, Washington, D.C. <http://documents.worldbank.org/curated/en/574171554178748054/Air-Quality-Management-in-Poland> (accessed 1 August 2020).
- Benestad, R.E., Hanssen-Bauer, I., Chen D., (2008). Empirical-statistical downscaling. *World Scientific*, p. 212. <https://doi.org/10.1142/6908>
- Breiman, L. (2001). Machine Learning. *Machine Learning*. 45, 5–32.
- Czernecki, B., Pórolnyczak, M., Kolendowicz, L., Marosz, M., Kendzierski, S., Pilgaj, N. (2017). Influence of the atmospheric conditions on PM<sub>10</sub> concentrations in Poznań. *J. Atmos. Chem.* 74, 115–139. <https://doi.org/10.1007/s10874-016-9345-5>
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., ..., Vitart, F. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 53–597. <https://doi.org/10.1002/qj.828>
- Dore, A.J., Hallsworth, S., McDonald, A.G., Werner, M., Kryza, M., Abbot, J., Nemitz, E., Dore, C.J., Malcolm, H., Vieno, M., Reis, S., Fowler, D. (2014). Quantifying missing annual emission sources of heavy metals in the United Kingdom with an atmospheric transport model. *Sci. Total Environ.* 479–480, 171–180. <https://doi.org/10.1016/j.scitotenv.2014.02.001>
- Draghicescu, D., Ignaccolo, R. (2005). Spatio-temporal maps of risk for PM<sub>10</sub> in Piemonte. In: Contributed papers for the Italian Statistical Society Meeting "Statistics and Environment", 275–278.
- European Union (EU) (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, OJ L152, 11.6.2008. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF> (accessed 15 August 2020).
- European Environment Agency (EEA) (2015). Air quality in Europe – 2015 report EEA report no 5/2015. Copenhagen, European Environment Agency, Denmark, p. 57.
- European Environment Agency (EEA) (2018). Air quality in Europe – 2018 report. EEA Report No 12/2018. Copenhagen, European Environment Agency, Denmark, p.83.
- Feng, S., Gao, D., Liao, F., Zhou, F., Wang, X., (2016). The health effects of ambient PM<sub>2.5</sub> and potential mechanisms. *Ecotoxicol. Environ. Saf.* 128, 67–74. <https://doi.org/10.1016/j.ecoenv.2016.01.030>
- Guaita, R., Pichiule, M., Mate, T., Linares, C., Diaz, J. (2011) Short-term impact of particulate matter (PM<sub>2.5</sub>) on respiratory mortality in Madrid. *Int. J. Environ. Health. Res.* 21, 260–274. <https://doi.org/10.1080/09603123.2010.544033>
- Gutiérrez, J.M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, R., Wibig, T., Wilcke, R., Kotlarski, S., San-Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., ..., Pagé C. (2018). An intercomparison of



- a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *Int. J. Climatol.* 39, 3750–3785. <https://doi.org/10.1002/joc.5462>
- Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M. (2008). A European daily high-resolution gridded dataset of surface temperature and precipitation. *J. Geophys. Res.* 113, D20119. <https://doi.org/10.1029/2008JD010201>
- Holst, J., Mayer, H., Holst, T. (2008). Effect of meteorological exchange conditions on PM<sub>10</sub> concentration. *Meteorol. Z.* 17, 273–282. <https://doi.org/10.1127/0941-2948/2008/0283>
- Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L., Strickland, M., Liu, Y. (2017). Estimating PM<sub>2.5</sub> concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* 51, 6936–6944. <https://doi.org/10.1021/acs.est.7b01210>
- Jędruszkiewicz, J., Czernecki, B., Marosz, M. (2017). The variability of PM<sub>10</sub> and PM<sub>2.5</sub> concentrations in selected Polish agglomerations: The role of meteorological conditions, 2006–2016. *Int. J. Environ. Health Res.* 27, 441–462. <https://doi.org/10.1080/09603123.2017.1379055>
- Joharestani, Z.M., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S. (2019). PM<sub>2.5</sub> Prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere* 10, 373. <https://doi.org/10.3390/atmos10070373>
- Jolliffe, I.T., Stephenson, D.B. (2012). Forecast Verification: A Practitioner's Guide in Atmospheric Science. John Wiley and Sons. <https://doi.org/10.1002/9781119960003>
- Juda-Rezler, K. (2006). Assessing the winter time air pollution in Cracow Poland in relation with possible influences on human health and cultural heritage. *Pol. J. Environ. Stud.* 15, 123–128.
- Junninen, H., Mönster, J., Rey, M., Cancelinha, J., Douglas, K., Duane, M., Forcina, V., Müller, A., Lagler, F., Marelli, L., Borowiak, A., Niedzialek, J., Paradiz, B., Mira-Salama, D., Jimenez, J., Hansen, U., Astorga, C., Stanczyk, K., Viana, M., ..., Larsen, B.R. (2009). Quantifying the Impact of Residential Heating on the Urban Air Quality in a Typical European Coal Combustion Region. *Environ. Sci. Technol.* 43, 7964–7970. <https://doi.org/10.1021/es8032082>
- Kendzierski, S., Czernecki, B., Kolendowicz, L., Jaczewski, A. (2018). Air temperature forecasts' accuracy of selected short-term and long-term numerical weather prediction models over Poland. *Geofizika* 35, 67–85. <https://doi.org/10.15233/gfz.2018.35.5>
- Kim, K.H., Kabir, E., Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environ. Int.* 74, 136–143. <https://doi.org/10.1016/j.envint.2014.10.005>
- Kryza, M., Werner, M., Guzikowski, J. (2020). Assimilation of Meteorological Data in Online Integrated Atmospheric Transport Model—Example of Air Quality Forecasts for Poland. In: Mensink C., Gong W., Hakami A. (Eds.), Air Pollution Modeling and its Application XXVI. ITM 2018. Springer Proceedings in Complexity. Springer, Cham.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, Ch., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T. (2019). Caret: Classification and Regression Training. R package version 6.0-84. <https://CRAN.R-project.org/package=caret> (accessed 1 August 2020).
- Kursa, M.B., Rudnicki, W.R. (2010). Feature Selection with the Boruta Package. *J. Stat. Softw.* 36, 1–13. <https://doi.org/10.18637/jss.v036.i11>
- Lavers, D.A., Beljaars, A., Richardson, D.S., Rodwell, M.J., Pappenberger, F. (2019). A forecast evaluation of planetary boundary layer height over the ocean. *J. Geophys. Res.* 124, 4975–4984. <https://doi.org/10.1029/2019JD030454>
- Liu, W., Guo, G., Chen, F., Chen, Y. (2019). Meteorological pattern analysis assisted daily PM<sub>2.5</sub> grades prediction using SVM optimized by PSO algorithm. *Atmos. Pollut. Res.* 10, 1482–1491. <https://doi.org/10.1016/j.apr.2019.04.005>
- Lu, F., Xu, D., Cheng, Y., Dong, S., Guo, C., Jiang, X., Zheng, X. (2015). Systematic review and meta-analysis of the adverse health effects of ambient PM<sub>2.5</sub> and PM<sub>10</sub> pollution in the Chinese population. *Environ. Res.* 136, 196–204. <https://doi.org/10.1016/j.envres.2014.06.029>
- Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Clim.* 26, 2137–2143. <https://doi.org/10.1175/JCLI-D-12-00821.1>
- Mira-Salama, D., Grüning, C., Jensen, N.R., Cavalli, P., Putaud, J.P., Larsen, B.R., Raes, F., Coe, H. (2008). Source attribution of urban smog episodes caused by coal combustion. *Atmos. Res.* 88, 294–304. <https://doi.org/10.1016/j.atmosres.2007.11.025>



- Organisation for Economic Co-operation and Development (OECD). (2016). The Economic Consequences of Outdoor Air Pollution, OECD Publishing, Paris. <https://doi.org/10.1787/9789264257474-en> (accessed 31 January 2019).
- Pinkerton, K.E., Green, F.H.Y., Saiki, C., Vallyathan, V., Plopper, C., Gopal, V., Hung, D., Bahne, E.B., Lin, S.S., Ménache, M.G., Schenker, M.B. (2000). Distribution of particulate matter and tissue remodeling in the human lung. *Environ. Health Perspect.* 108, 1063–1069. <https://doi.org/10.1289/ehp.001081063>
- Polichetti, G., Cocco, S., Spinali, A., Trimarco, V., Nunziata, A. (2009). Effects of particulate matter ( $PM_{10}$ ,  $PM_{2.5}$  and  $PM_1$ ) on the cardiovascular system. *Toxicology* 261, 1–8. <https://doi.org/10.1016/j.tox.2009.04.035>
- Riedmiller M., Braun H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. Proceedings of the IEEE International Conference on Neural Networks (ICNN), pp. 586–591. San Francisco. <https://doi.org/10.1109/ICNN.1993.298623>
- Riedmiller, M. (1994). Rprop - Description and Implementation Details. Technical Report. University of Karlsruhe.
- Roebber, P.J. (2009). Visualizing multiple measures of forecast quality. *Weather Forecasting* 24, 601–608. <https://doi.org/10.1175/2008WAF2222159.1>
- Samek, L. (2016). Overall human mortality and morbidity due to exposure to air pollution. *Int. J. Occup. Med. Environ. Health* 29, 417–426. <https://doi.org/10.13075/ijomeh.1896.00560>
- Samoli, E., Peng, R., Ramsay, T., Pipikou, M., Touloumi, G., Dominici, F., Burnett R., Cohen, A., Krewski, D., Samet, J., Katsouyanni, K. (2008). Acute effects of ambient particulate matter on mortality in Europe and North America: Results from the APHENA study. *Environ. Health. Perspect.* 116, 1480–1486. <https://doi.org/10.1289/ehp.11345>
- Taylor, K. (2001). Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* 106, 7183–7192. <https://doi.org/10.1029/2000JD900719>
- Wilks, D. (2011). Statistical Methods in the Atmospheric Sciences. Academic Press.
- World Health Organization (WHO) (2013). Health effects of particulate matter. Policy implications for countries in eastern Europe, Caucasus and central Asia. WHO 2013 Copenhagen O, Denmark, p. 20. [http://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf?ua=1](http://www.euro.who.int/__data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf?ua=1) (accessed 1 August 2020).
- World Health Organization (WHO) (2014). Burden of disease from ambient air pollution for 2012 — summary of results, World Health Organization. [http://www.who.int/phe/health\\_topics/outdoorair/databases/AAP\\_BoD\\_results\\_March2014.pdf](http://www.who.int/phe/health_topics/outdoorair/databases/AAP_BoD_results_March2014.pdf) (accessed 1 August 2020).
- World Health Organization (WHO) (2018). World health statistics 2018: monitoring health for the SDGs, sustainable development goals. <https://apps.who.int/iris/bitstream/handle/10665/272596/9789241565585-eng.pdf?ua=1> (accessed 1 August 2020).
- Zheng, S., Pozzer, A., Cao, C.X., Lelieveld, J. (2015). Long-term (2001–2012) concentrations of fine particulate matter ( $PM_{2.5}$ ) and the impact on human health in Beijing, China. *Atmos. Chem. Phys.* 15, 5715–5725. <https://doi.org/10.5194/acp-15-5715-2015>