

Article

Regression Modeling of Daily PM_{2.5} Concentrations with a Multilayer Perceptron

Szymon Hoffman ^{1,*} , Rafał Jasiński ¹  and Janusz Baran ²

¹ Faculty of Infrastructure and Environment, Czestochowa University of Technology, 69 Dabrowskiego St., 42-200 Czestochowa, Poland; rafal.jasinski@pcz.pl

² Faculty of Electrical Engineering, Czestochowa University of Technology, 17 Armii Krajowej, 42-200 Czestochowa, Poland; janusz.baran@pcz.pl

* Correspondence: szymon.hoffman@pcz.pl

Abstract: Various types of energetic fuel combustion processes emit dangerous pollutants into the air, including aerosol particles, marked as PM₁₀. Routine air quality monitoring includes determining the PM₁₀ concentration as one of the basic measurements. At some air monitoring stations, the PM₁₀ measurement is supplemented by the simultaneous determination of the concentration of PM_{2.5} as a finer fraction of suspended particles. Since the PM_{2.5} fraction has a significant share in the PM₁₀ fraction, the concentrations of both types of particles should be strongly correlated, and the concentrations of one of these fractions can be used to model the concentrations of the other fraction. The aim of the study was to assess the error of predicting PM_{2.5} concentration using PM₁₀ concentration as the main predictor. The analyzed daily concentrations were measured at 11 different monitoring stations in Poland and covered the period 2010–2021. MLP (multilayer perceptron) artificial neural networks were used to approximate the daily PM_{2.5} concentrations. PM₁₀ concentrations and time variables were tested as predictors in neural networks. Several different prediction errors were taken as measures of modeling quality. Depending on the monitoring station, in models with one PM₁₀ predictor, the RMSE error values were in the range of 2.31–6.86 µg/m³. After taking into account the second predictor D (date), the corresponding RMSE errors were lower and were in the range of 2.06–5.54 µg/m³. Our research aimed to find models that were as simple and universal as possible. In our models, the main predictor is the PM₁₀ concentration; therefore, the only condition to be met is monitoring the measurement of PM₁₀ concentrations. We showed that models trained at other air monitoring stations, so-called foreign models, can be successfully used to approximate PM_{2.5} concentrations at another station.

Keywords: air protection; air monitoring; environmental management; air quality modeling; particular matter; PM_{2.5} prediction; regression models; artificial neural networks; multilayer perceptrons



Citation: Hoffman, S.; Jasiński, R.; Baran, J. Regression Modeling of Daily PM_{2.5} Concentrations with a Multilayer Perceptron. *Energies* **2024**, *17*, 2202. <https://doi.org/10.3390/en17092202>

Academic Editor: George Halkos

Received: 14 March 2024

Revised: 25 April 2024

Accepted: 30 April 2024

Published: 3 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As knowledge about air pollution deepens, new information about the threats resulting from the presence of these pollutants reaches public awareness. Threats concern various aspects of social life. The most obvious and longest historically studied threat is the adverse impact of pollution on human life and health. Research confirms that air pollution also affects animals, plants and other living organisms [1–3]. It can cause large losses in animal husbandry and crop yields, therefore causing losses in agriculture [4–6]. Pollution also has a direct or indirect adverse impact on other sectors of the economy [7–11].

Basic air pollutants are emitted by natural processes. Anthropogenic emissions, mainly related to energy production, introduce additional amounts of pollutants into the air and cause pollutant concentrations to reach unnaturally high levels [12]. In the case of large anthropogenic emissions and in unfavorable weather, concentrations of toxic pollutants may be so high that they threaten not only health but also the lives of people, animals and

other organisms. The first historically documented strong smog episode was recorded on 5–9 December 1952 in London [13]. According to medical statistics, about 4000 people died and about 150,000 were hospitalized. According to the reports, people suffered from respiratory failure and hypoxia, i.e., oxygen deficiency. These were the first documented medical diagnoses of the reasons for death and hospitalization. Until today, it has been recognized that air pollution can cause bronchial and lung diseases [14–16], cancer [17,18] and also cardiovascular diseases [19–22]. Smog also affects the brain and can deepen mental illnesses and neurological ailments [23–26].

After this tragic event, the first legal regulations forcing the improvement of air quality (including the Clean Air Act) were passed in Great Britain [27]. There was also a need to monitor the air quality. Intensive research in the field of environmental chemistry was initiated to identify pollutants, mechanisms of their formation and threats to the broadly understood natural environment, including humans, animals and plants. The development of pollution detection techniques enabled the construction of automatic air monitoring stations [28,29]. Currently, networks of air monitoring stations operate in almost all developed and emerging countries of the world. This enables continuous monitoring of air quality and warning against smog episodes.

Typical air monitoring stations are equipped with analyzers enabling continuous measurement of the concentrations of basic gaseous air pollutants: O_3 , SO_2 , NO_x and CO. Sometimes this measurement package is expanded to include other pollutants, e.g., from the group of volatile organic compounds (VOCs). Studies of atmospheric toxicity revealed that particularly dangerous pollutants can be sorbed on the surfaces of aerosols, which are also ubiquitous in atmospheric air. These pollutants include polycyclic aromatic hydrocarbons (PAHs), heavy metals and many others. Therefore, an important challenge for atmospheric monitoring is to measure the concentration of suspended particles and the pollutants they contain. Initially, attention was focused on the dust fraction with particle sizes up to $10\ \mu m$, designated as PM_{10} . The measurement of this aerosol fraction is performed at most automatic air monitoring stations, even those that are poorly equipped.

New research is providing more and more information about the relationship between dust particle size and human exposure risk. It turned out that the worst health effects are observed when inhaling particles smaller than $2.5\ \mu m$ (the dust fraction called $PM_{2.5}$). Due to their small size, $PM_{2.5}$ particles can penetrate much deeper into the respiratory system than PM_{10} particles. Pollutants collected in particles can reach the bronchi and even the alveoli of the lungs, enter the bloodstream and then spread throughout the body [30–32]. Currently, it is the concentration of the $PM_{2.5}$ fraction that is used to calculate mortality rates, such as the number of premature deaths. It is estimated that in Poland, this number is approximately 40,000 people, which constitutes approximately 0.1% of the population [33]. Particulate matter is considered the main cause of premature death in many other countries around the world [34,35].

The development of measurement techniques enables automatic measurement of $PM_{2.5}$ concentration. However, this measurement is complicated and expensive. It is performed only at selected air monitoring stations. It should be mentioned that the reference measurement method in the case of PM_{10} or $PM_{2.5}$ is the gravimetric method, which involves measuring the mass of dust collected on a special filter [36]. This method has its limitations and is usually used over longer averaging periods than gas concentration measurements. Throughout the European Union, including Poland, a 24 h (daily) measurement period has been adopted as the standard for this type of measurement [37]. To sum up, the results of $PM_{2.5}$ concentration measurements are key to assessing health effects, but appropriate monitoring is performed in only a few measurement stations. There is a need to increase the density of the $PM_{2.5}$ measurement network. In the long term, this need will probably be met by measurements of $PM_{2.5}$ concentrations at each air monitoring station. Until this happens, $PM_{2.5}$ concentrations can be estimated using modeling techniques.

Modeling air pollution concentrations has a long history. These techniques can be used to fill in missing data in air monitoring systems. They are also used to predict concentrations

of selected air pollutants. In the past, classical regression or autoregressive methods were used [38–40]. Since the 1990s, artificial intelligence techniques have been increasingly used to model air pollution concentrations [41–50]. The most popular were models that provide prediction without any data from outside of the monitoring system. They are sometimes called autonomous models [48,49]. Very complex neural models, including deep learning methods, are increasingly used to predict air pollution concentrations [51–53]. However, these networks require a lot of data, including external data, and therefore their practical applicability is limited.

Neural networks, even networks with a relatively simple structure, turn out to be useful in regression models if predictors are available that are highly correlated with the variable being modeled. If the goal is to model $PM_{2.5}$ concentrations at a selected air monitoring station, PM_{10} concentrations measured at the same location are often also available. In Polish conditions, both types of concentrations are always strongly correlated because $PM_{2.5}$ is a finer fraction within PM_{10} , and its percentage usually significantly exceeds 50% in the mass of PM_{10} aerosol [54]. Strong statistical dependencies between PM_{10} and $PM_{2.5}$ fractions were also found in reports from other countries [55,56]. PM_{10} concentration can therefore be considered a universal primary predictor in regression models approximating $PM_{2.5}$ concentration. This concept was tested in neural network models trained to approximate hourly $PM_{2.5}$ concentrations [54], and promising results were obtained. It turned out that using simple regression MLP models could yield a small $PM_{2.5}$ prediction error. It is enough to use time variables and PM_{10} concentration as explanatory variables. Introducing more predictors only slightly improves the accuracy of MLP regression models. In the studies described below, similar modeling was performed to predict 24 h $PM_{2.5}$ concentrations. The analysis was performed on data from various air monitoring stations in Poland. We took into consideration the stations where the concentrations of both PM_{10} and $PM_{2.5}$ were measured over many years. MLP regression neural networks were used for modeling.

The purpose of the analysis was not to create the most accurate predictive models of $PM_{2.5}$ concentrations. The main goal was to test the concept of possibly simple predictive models that would use data resources available at most air monitoring stations. It was decided that the key predictor would be the concentration of PM_{10} . Since the modeling concerns 24 h (daily) concentrations, the supporting predictor was a time variable providing information about the time of year. Therefore, the analysis was limited to testing only the two variables mentioned above as predictors. An additional goal was to test the universality of the obtained models. For this purpose, simulations were performed to check whether a neural network trained at one of the stations could be used to predict daily $PM_{2.5}$ concentrations at other air monitoring stations.

2. Materials and Methods

2.1. Air Monitoring Sites

We took to the examination data from 11 automatic air monitoring stations situated in Poland in the following sites: Jaslo, Katowice, Kosciierzyna, Krakow, Lodz, Lublin, Olsztyn, Osieczow, Puszcza Borecka, Zielona Gora and Zielonka. Measurement data cover a period from 2010 to 2021. Monitoring stations were selected to meet two conditions:

1. Their location was diverse enough to cover various regions of Poland;
2. Daily concentrations of both PM_{10} and $PM_{2.5}$ fractions were simultaneously measured at each station for at least several years.

The location of the station is shown in Figure 1, where the borders of the administrative division of Poland are marked. All the monitoring stations are operated by the Chief Inspectorate of Environmental Protection in Poland. Table 1 contains background information about the location of individual stations, including addresses, international codes, geographical coordinates, station types and types of monitored area.

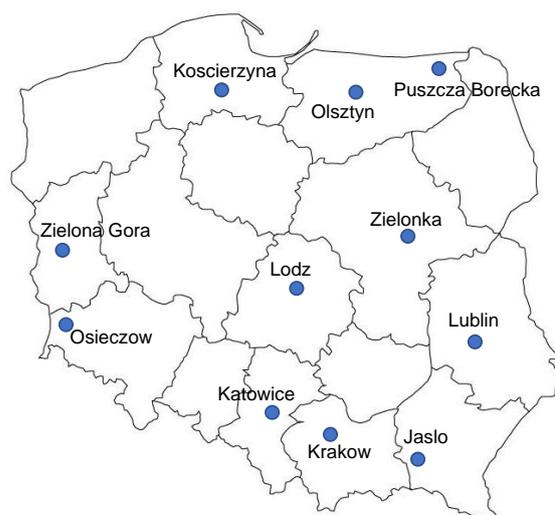


Figure 1. Map of Poland with locations of the air monitoring stations considered in the research.

Table 1. Background information about the considered air monitoring stations, from [57].

Air Monitoring Station	Address	International Code	Geographical Coordinates, WGS84	Type of Station	Area Type
Jaslo	Sikorskiego Str.	PL0518A	Φ 49.744886, λ 21.454617	background	urban
Katowice	6 Kossutha Str.	PL0008A	Φ 50.264611, λ 18.975028	background	urban
Koscierzyna	Targowa Str.	PL0558A	Φ 54.120694, λ 17.975861	background	urban
Krakow	Bujaka Str.	PL0501A	Φ 50.010575, λ 19.949189	background	urban
Lodz	1 Legionow Str.	PL0100A	Φ 51.776417, λ 19.452936	background	urban
Lublin	5 Sliwińskiego Str.	PL0085A	Φ 51.273078, λ 22.551675	background	urban
Olsztyn	16 Puskina Str.	PL0175A	Φ 53.789233, λ 20.486075	background	urban
Osieczow	(no street)	PL0505A	Φ 51.317630, λ 15.431719	background	rural
Puszczka Borecka	Diabla Gora	PL0005R	Φ 54.124819, λ 22.038056	background	rural
Zielona Gora	Krotka Str.	PL0213A	Φ 51.939783, λ 15.518861	background	urban
Zielonka	Bory Tucholskie	PL0077A	Φ 53.662136, λ 17.933986	background	rural

2.2. Air Monitoring Data

The 24 h PM_{10} and $PM_{2.5}$ concentration values recorded in 2010–2021 were used for the study. The data were provided by the Chief Inspectorate of Environmental Protection in Poland. The provided air monitoring data were validated and officially approved.

The concentration of PM_{10} and $PM_{2.5}$ was measured using the gravimetric method [35]. The PM concentrations were calculated on the basis of differences in filter masses before and after daily exposure, which involved passing a specific volume of dust-contaminated air through the filters. The filters were replaced automatically every 24 h. PM concentrations are given in $\mu\text{g}/\text{m}^3$.

The following symbols were used to describe the data used:

PM_{10} —daily averaged concentration of particles up to 10 μm in size.

$PM_{2.5}$ —daily averaged concentration of particles up to 2.5 μm in size.

D—date in the numerical form.

2.3. Temporal Variable's Transformation

The numerical form of the date (D) was prepared in such a way that each date was replaced with a value from the range $0 \div 1$. For January 1 (the first day of the year), the value was set to 1, and for July 2 (the middle day of the year), the value was set to 0. During the first half of the year, the numerical value of the date decreases linearly from 1 to 0. During the second half of the year, the numerical value of the date increases linearly from 0 to almost 1.

The purpose of transforming the date into such a cyclic numerical form was to assign the same values to similar days in different years. The transformation also made it possible to maintain continuity when changing the year: 31 December and 1 January of the following year have almost identical values.

2.4. Data Preparation

Not all the air monitoring stations carried out measurements for the entire 12 years. Even if measurements were performed, their completeness was sometimes unsatisfactory. Therefore, only the annual series of measurements whose completenesses exceeded 80% were taken into account for the analysis. At lower completeness, the entire annual measurement series was removed from the analyzed set. The aim of this procedure was to ensure that the analyzed cases covered all seasons of a calendar year as evenly as possible. Cases without PM measurements were also removed from the set of analyzed data. Only those cases (days) for which both PM₁₀ and PM_{2.5} concentration values were known were left to train the network. As for the choice of the station, the principle was adopted that measurements of PM₁₀ and PM_{2.5} concentrations should be carried out for a minimum of 5 consecutive years. The processed data, prepared for the analysis in this way, have been attached as a Supplemental File.

Table 2 shows the completeness of the data series for individual air monitoring stations, after removing cases with missing data and after removing annual time series with completeness below 80%. Due to the fact that only complete cases were retained, the completeness of data for both PM₁₀ and PM_{2.5} pollutants was the same.

Table 2. Completeness of the annual series of 24 h PM₁₀ and PM_{2.5} concentrations for the years covered by the analysis, 2010–2021. Only values above 80% are shown.

Air Monitoring Station	Total Number of Observations (Cases)	Completeness of the Annual Series											
		2010 %	2011 %	2012 %	2013 %	2014 %	2015 %	2016 %	2017 %	2018 %	2019 %	2020 %	2021 %
Jaslo	2043	-	-	-	-	91.5	91.5	99.5	97.3	81.4	98.4	-	-
Katowice	2731	-	-	-	-	89.0	89.3	89.6	95.9	92.1	97.3	99.5	95.1
Koscierzyna	1709	-	-	-	-	90.1	84.4	98.4	95.6	99.5	-	-	-
Krakow	2102	-	-	-	-	91.0	96.7	97.3	94.2	97.0	99.5	-	-
Lodz	2813	-	-	-	-	91.0	99.5	99.5	99.7	95.1	99.7	96.7	89.0
Lublin	3229	-	-	-	96.7	90.4	100.0	100.0	100.0	97.0	100.0	100.0	100.0
Olsztyn	2438	-	-	-	-	-	95.9	95.6	94.0	89.0	97.3	97.3	98.4
Osieczow	3412	-	95.3	95.6	-	88.8	94.8	86.3	91.5	97.0	89.9	97.0	97.8
Puszcza Borecka	3413	-	-	92.9	91.8	87.1	94.8	91.3	95.1	97.5	96.2	93.4	94.2
Zielona Gora	3849	-	89.9	92.9	94.8	92.9	91.8	99.7	97.5	100.0	99.7	96.4	98.1
Zielonka	3767	96.4	100.0	89.1	99.2	89.0	-	89.1	87.1	98.4	96.4	92.9	93.7

Figure 2 presents a statistical description of the set of daily PM₁₀ and PM_{2.5} concentrations, measured at considered air monitoring stations, calculated only for complete cases. Among the selected stations, the highest average and maximum concentrations of suspended dust occurred at measuring stations located in large urban agglomerations, i.e., in Katowice, Krakow and Lodz. The lowest values were at air monitoring stations mainly located in rural areas, i.e., Puszcza Borecka, Zielonka, Osieczow and Olsztyn.

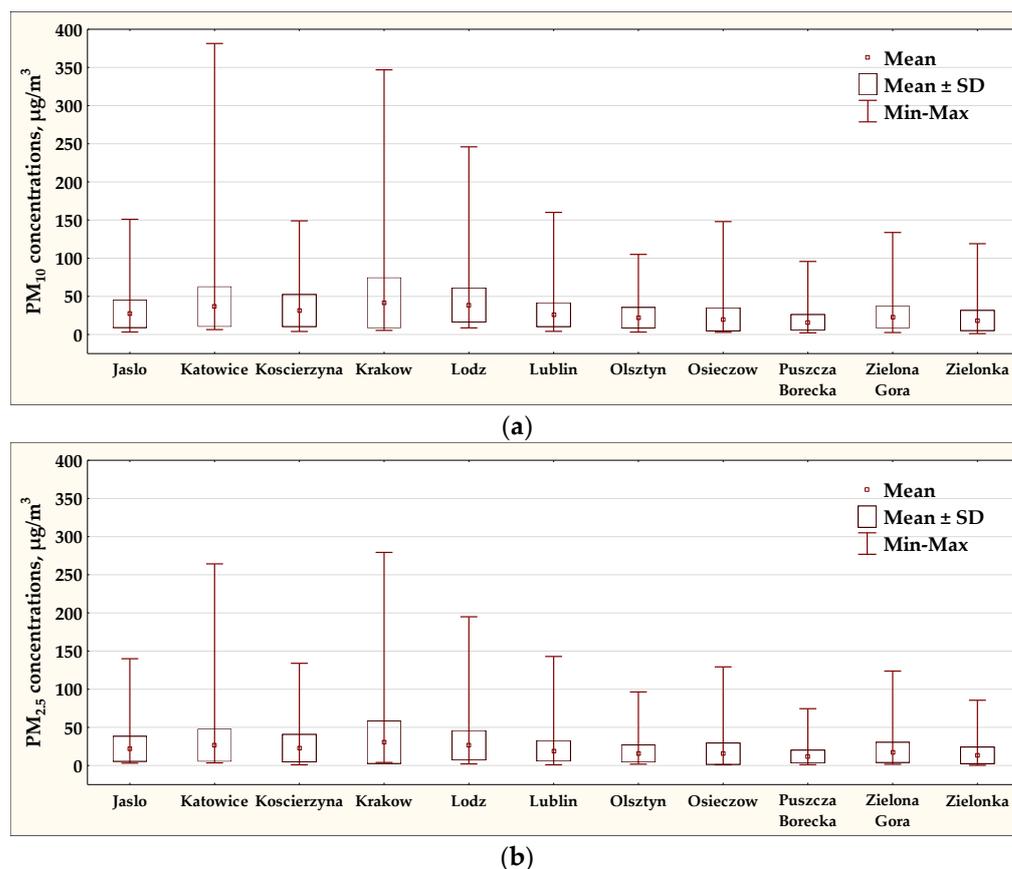


Figure 2. Graphical presentation of basic statistical parameters of PM₁₀ and PM_{2.5} concentrations from the considered air monitoring stations in 2010–2021: (a) PM₁₀ concentrations, (b) PM_{2.5} concentrations. Values calculated after removing cases with missing data and for the years included in the analysis.

2.5. Regression Models

For modeling, the Statistica version 13.3 program was used, together with the SANN (Statistica Artificial Neural Networks) extension subprogram, enabling the creation of neural networks [58]. A separate data set was prepared for each station. The modeling was performed separately for each of the 11 air monitoring stations. The multilayer perceptron (MLP) was adopted in the generated models. It was assumed that all models have the same architecture. All perceptrons have a single hidden layer of 10 neurons. Before training a neural network, a data set was randomly divided into three subsets: a training subset (50% of cases), a testing subset (25% of cases) and a validation subset (25% of cases). The number of all cases in the sets from various air monitoring stations is given in Table 2. The BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm was used in the network training process. This algorithm is intended for numerical optimization [59]. The mathematical basis of the algorithm was developed by the above-mentioned mathematicians in 1970 [60–63].

For all networks, the learning process was stopped after 300 epochs. A logistic activation function was assumed in the hidden neurons and a linear activation function in the output neurons. The network was initialized randomly using the Gaussian method. The initial weights followed a normal distribution with zero mean and unit variance. The sum of squares (SOS) was used as the error function in the network training process. The SOS is the sum of the squares of the differences between the predicted and observed concentration values. For each air monitoring station, the prediction was repeated 5 times, each time with different random assignment of cases to subsets and with different randomly chosen initial weights. The most precise of the 5 generated models was selected for each station, and the results of these most accurate models are presented in the Results section. The accuracy of the models was assessed by calculating 5 different error measures, MAE, RMSE, MARE, R²

and d. The mentioned error measures were described in the next subsection. The networks generated for the same monitoring stations differed slightly in the modeling errors. They had identical neuronal structures but differed in the weights and degrees of activation of individual neurons in the hidden layer. Statistica Neural Networks automatically scales input and output variables using a linear transformation to the interval [0, 1].

For each monitoring station, regression models were created whose output variable was the daily concentration of $PM_{2.5}$, while the input variables were introduced in one of the three variants:

Variant I—D—(MLP 1-10-1, Figure 3a);

Variant II— PM_{10} —(MLP 1-10-1, Figure 3b);

Variant III—D and PM_{10} —(MLP 2-10-1, Figure 3c).

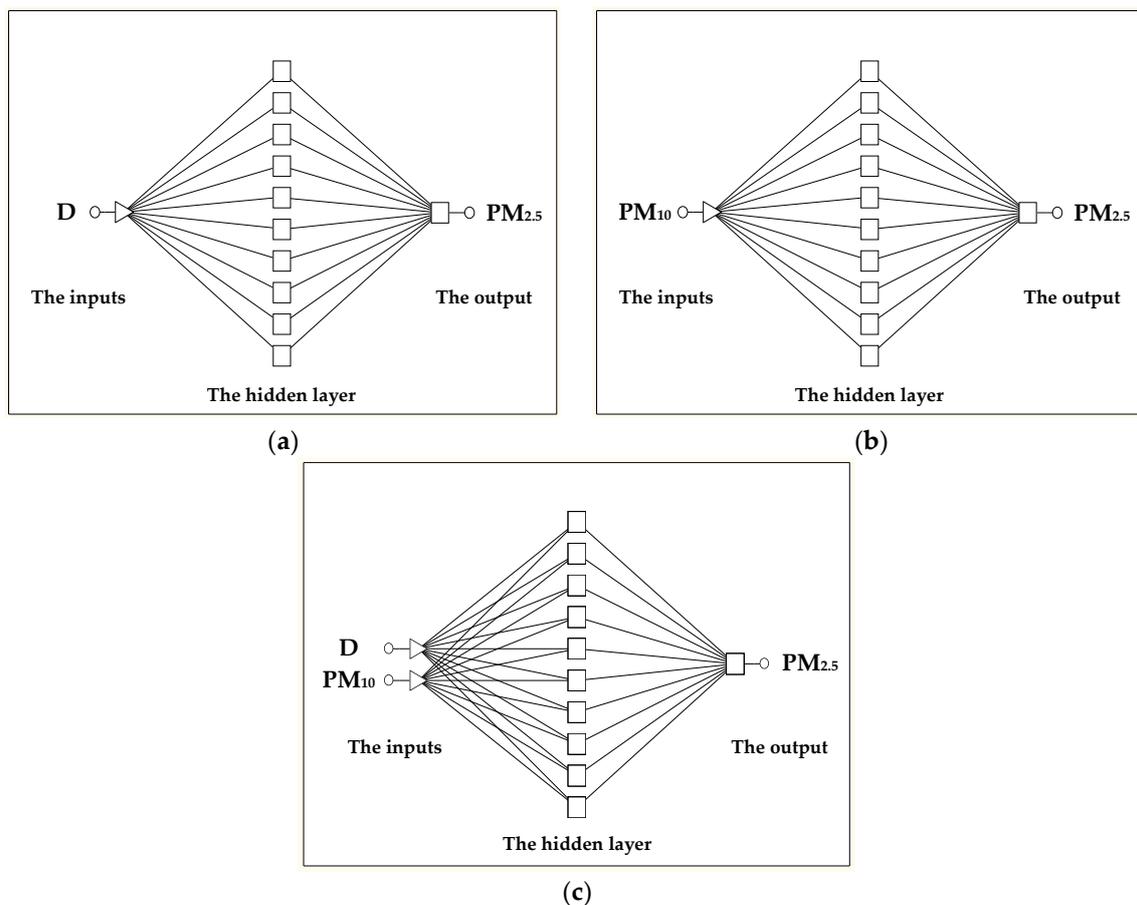


Figure 3. MLP architecture diagrams with 10 neurons in one hidden layer and 3 variants of predictors: (a) D; (b) PM_{10} ; (c) D, PM_{10} .

Figure 3 shows the MLP architectures for these different variants.

In addition to the perceptron models, simpler models were also created to compare accuracy. A linear regression model (LIN) and a naive mean model (MEAN) were generated. In the latter, the average value of $PM_{2.5}$ concentration was assumed for each station, calculated separately for each station from historical measurements.

2.6. Assessment of the Prediction Accuracy

To assess the accuracy of the obtained MLP artificial neural network models, the following error values were used: MAE (mean absolute error), RMSE (root mean squared error), MARE (mean absolute relative error), R^2 (coefficient of determination) and d (Willmott index of agreement). Index d is a special formula dedicated to air quality modeling [64]. These

values were calculated by comparing real 24 h PM_{2.5} concentration values with predicted values. The formulas for calculating the listed errors are given below in Equations (1)–(5).

MAE—Mean Absolute Error

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (1)$$

RMSE—Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (2)$$

MARE—Mean Absolute Relative Error

$$\text{MARE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \quad (3)$$

R²—Coefficient of Determination

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

d—Willmott Index of Agreement

$$d = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |x_i - \bar{x}|)^2} \quad (5)$$

where:

- n—number of cases;
- y—predicted concentrations;
- x—real concentrations;
- \bar{x} —arithmetic average of real concentrations;
- i—the case number.

2.7. Verification of Models

Approach 1

In order to check whether it is possible to predict PM_{2.5} concentrations in a later measurement period using neural networks trained on historical data, a trial modeling of daily PM_{2.5} concentrations was performed. Each trial prediction was made for a period of one month, using a neural network model trained on data recorded in the past at the same station. The test time series were selected at different time periods than those used to train the network, so the input data packets were completely unknown to the network. By comparing the course of actual and predicted concentrations for the selected periods, the usefulness of the obtained MLP neural network models for predicting PM_{2.5} concentrations in “new conditions” was tested. The MLP 2-10-1 models with numerical date D and PM₁₀ concentration as predictors were verified. The modeling errors were calculated for each chosen period.

Approach 2

This approach examined the “universality” of the obtained MLP models. For this purpose, a neural network trained at one of the stations was used to predict daily PM_{2.5} concentrations at other air monitoring stations. The models in Variant III (MLP 2-10-1) created for 3 different air monitoring stations (Kraków) were tested as reference ones. Then, the prediction quality of these models “fed” with input data from other stations previously unknown to these models was assessed. The modeling error was then calculated.

3. Results

3.1. Annual Courses of PM_{10} and $PM_{2.5}$ Concentrations

Based on many years of data recorded at the air monitoring stations, annual statistical patterns of $PM_{2.5}$ and PM_{10} concentrations were calculated for each station. The plots of these patterns are presented in Figure 4. At all the stations, the lowest concentrations of suspended dust occurred in the spring/summer periods. In typical winter months, the particulate matter concentrations reached their highest values.

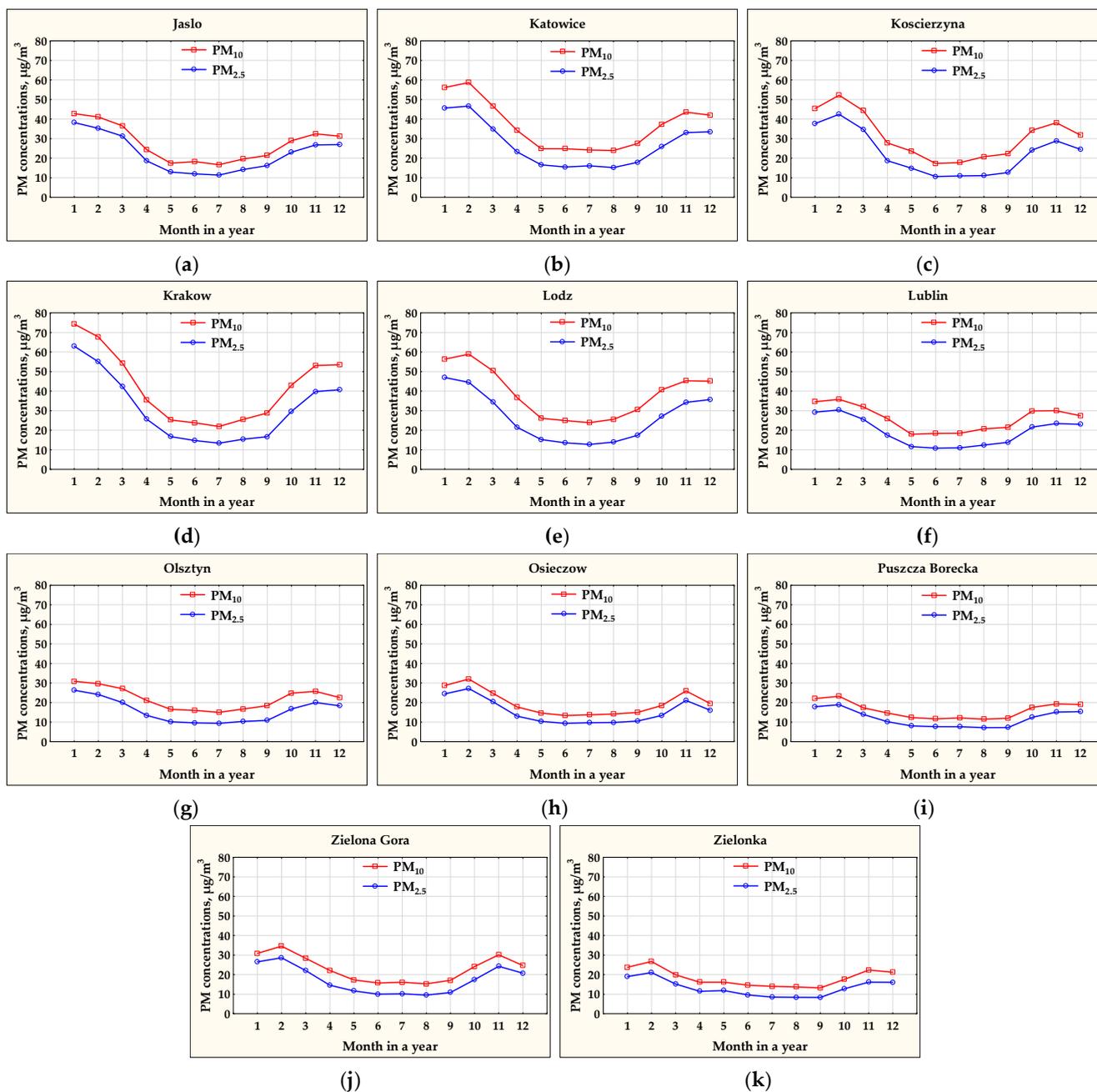


Figure 4. Annual changes in PM_{10} and $PM_{2.5}$ concentrations at the monitoring stations: (a) Jaslo, (b) Katowice, (c) Koscierzyna, (d) Krakow, (e) Lodz, (f) Lublin, (g) Olsztyn, (h) Osieczow, (i) Puszcza Borecka, (j) Zielona Gora, (k) Zielonka.

3.2. Correlations of Variables

A correlation analysis of the input variables, $PM_{2.5}$, PM_{10} concentrations and date (D), was performed at each of the air monitoring stations. This analysis was performed to compare potential predictors of $PM_{2.5}$ concentrations. The values of the Pearson correlation coefficients are presented in Table 3. The Pearson correlation coefficient ranges from -1 to 1 ; the closer to 1 or -1 , the stronger the correlation between the variables. A positive sign means a positive correlation and a negative sign means a negative correlation. A coefficient value close to zero means a complete lack of correlation between the variables. At all the stations, the strongest correlations in relation to the $PM_{2.5}$ concentration occurred for the PM_{10} concentration, much lower for the time variable D. These predictors were used to generate models I, II and III (see Section 2.5).

Table 3. Pearson's correlation coefficient for the input variables at individual air monitoring stations, 24 h average values, 2010–2021.

Air Monitoring Station	Variable	D	PM_{10}	$PM_{2.5}$
Jaslo	D	1.0000		
	PM_{10}	0.4201	1.0000	
	$PM_{2.5}$	0.4768	0.9735	1.0000
Katowice	D	1.0000		
	PM_{10}	0.3842	1.0000	
	$PM_{2.5}$	0.4450	0.9639	1.0000
Koscierzyna	D	1.0000		
	PM_{10}	0.4441	1.0000	
	$PM_{2.5}$	0.5069	0.9487	1.0000
Krakow	D	1.0000		
	PM_{10}	0.4689	1.0000	
	$PM_{2.5}$	0.5048	0.9792	1.0000
Lodz	D	1.0000		
	PM_{10}	0.4755	1.0000	
	$PM_{2.5}$	0.5661	0.9339	1.0000
Lublin	D	1.0000		
	PM_{10}	0.3319	1.0000	
	$PM_{2.5}$	0.4553	0.9646	1.0000
Olsztyn	D	1.0000		
	PM_{10}	0.3432	1.0000	
	$PM_{2.5}$	0.4519	0.9493	1.0000
Osieczow	D	1.0000		
	PM_{10}	0.3219	1.0000	
	$PM_{2.5}$	0.3443	0.9852	1.0000
Puszczka Borecka	D	1.0000		
	PM_{10}	0.3478	1.0000	
	$PM_{2.5}$	0.4329	0.9611	1.0000
Zielona Gora	D	1.0000		
	PM_{10}	0.3725	1.0000	
	$PM_{2.5}$	0.4434	0.9487	1.0000
Zielonka	D	1.0000		
	PM_{10}	0.2662	1.0000	
	$PM_{2.5}$	0.3176	0.9387	1.0000

3.3. Results of Predicting $PM_{2.5}$ Concentrations

Modeling errors were calculated by comparing the predicted concentrations to the actual $PM_{2.5}$ concentrations. Statistical prediction errors were calculated for each model variant separately for each air monitoring station. To evaluate the modeling accuracy,

the five error measures defined by Formulas (1)–(5) were calculated. A summary of the prediction error values is presented in Table 4. The MAE and RMSE values are also presented graphically in Figure 5.

Table 4. Values of modeling errors of PM_{2.5} concentrations in MEAN model, LINEAR model with PM₁₀ as a predictor and MLP models with D and PM₁₀ as predictors.

Air Monitoring Station	Regression Model	Explanatory Variable (Predictors)	MAE $\mu\text{g}/\text{m}^3$	RMSE $\mu\text{g}/\text{m}^3$	MARE	R ²	d
Jaslo	MEAN	-	11.43	16.55	0.6893	0.0000	0.0000
	LIN	PM ₁₀	2.43	3.79	0.1400	0.9477	0.9864
	MLP 1-10-1	D	9.85	15.74	0.5515	0.2373	0.6155
	MLP 1-10-1	PM ₁₀	2.36	3.80	0.1321	0.9491	0.9867
	MLP 2-10-1	D, PM ₁₀	2.04	3.47	0.1148	0.9576	0.9890
Katowice	MEAN	-	12.90	23.28	0.4161	0.0000	0.3162
	LIN	PM ₁₀	3.87	5.73	0.1691	0.9292	0.9814
	MLP 1-10-1	D	11.50	18.95	0.5165	0.2236	0.5760
	MLP 1-10-1	PM ₁₀	3.84	5.81	0.1683	0.9274	0.9805
	MLP 2-10-1	D, PM ₁₀	3.35	5.18	0.1509	0.9422	0.9848
Koscierzyna	MEAN	-	13.18	18.04	0.9755	0.0000	0.0000
	LIN	PM ₁₀	3.74	5.70	0.2668	0.9000	0.9732
	MLP 1-10-1	D	10.01	14.97	0.6693	0.3050	0.6807
	MLP 1-10-1	PM ₁₀	3.67	5.67	0.2675	0.9014	0.9735
	MLP 2-10-1	D, PM ₁₀	3.15	4.89	0.2339	0.9266	0.9805
Krakow	MEAN	-	17.96	28.20	0.7839	0.0000	0.1307
	LIN	PM ₁₀	3.91	5.69	0.1576	0.9588	0.9894
	MLP 1-10-1	D	14.56	23.83	0.6539	0.2780	0.6492
	MLP 1-10-1	PM ₁₀	3.80	5.63	0.1497	0.9598	0.9897
	MLP 2-10-1	D, PM ₁₀	3.38	5.26	0.1340	0.9648	0.9910
Lodz	MEAN	-	13.11	21.32	0.4614	0.0000	0.3499
	LIN	PM ₁₀	4.86	6.94	0.2269	0.8722	0.9648
	MLP 1-10-1	D	10.05	15.63	0.4580	0.3522	0.7073
	MLP 1-10-1	PM ₁₀	4.74	6.86	0.2183	0.8754	0.9659
	MLP 2-10-1	D, PM ₁₀	3.76	5.54	0.1821	0.9188	0.9783
Lublin	MEAN	-	9.50	15.38	0.4408	0.0000	0.3713
	LIN	PM ₁₀	2.65	3.58	0.1692	0.9304	0.9817
	MLP 1-10-1	D	7.91	11.84	0.5503	0.2369	0.6047
	MLP 1-10-1	PM ₁₀	2.62	3.56	0.1640	0.9310	0.9820
	MLP 2-10-1	D, PM ₁₀	1.79	2.59	0.1161	0.9635	0.9905
Olsztyn	MEAN	-	7.77	12.02	0.5267	0.0000	0.3059
	LIN	PM ₁₀	2.51	3.58	0.1793	0.9012	0.9734
	MLP 1-10-1	D	7.02	9.98	0.6197	0.2315	0.6118
	MLP 1-10-1	PM ₁₀	2.42	3.52	0.1683	0.9049	0.9742
	MLP 2-10-1	D, PM ₁₀	1.72	2.52	0.1301	0.9511	0.9872
Osieczow	MEAN	-	8.64	15.89	0.4456	0.0000	0.3434
	LIN	PM ₁₀	1.70	2.43	0.1457	0.9705	0.9925
	MLP 1-10-1	D	8.18	12.95	0.7748	0.1621	0.5094
	MLP 1-10-1	PM ₁₀	1.58	2.31	0.1302	0.9733	0.9932
	MLP 2-10-1	D, PM ₁₀	1.48	2.22	0.1239	0.9754	0.9937
Puszcza Borecka	MEAN	-	6.35	9.96	0.4845	0.0000	0.3940
	LIN	PM ₁₀	1.62	2.36	0.1692	0.9237	0.9798
	MLP 1-10-1	D	5.52	7.79	0.6995	0.2188	0.5910
	MLP 1-10-1	PM ₁₀	1.61	2.34	0.1702	0.9255	0.9801
	MLP 2-10-1	D, PM ₁₀	1.38	2.06	0.1458	0.9420	0.9849
Zielona Gora	MEAN	-	9.33	15.50	0.4389	0.0000	0.3829
	LIN	PM ₁₀	2.84	4.21	0.2107	0.9001	0.9732
	MLP 1-10-1	D	7.95	11.76	0.5897	0.2252	0.5708
	MLP 1-10-1	PM ₁₀	2.78	4.17	0.2058	0.9023	0.9734
	MLP 2-10-1	D, PM ₁₀	2.45	3.76	0.1812	0.9207	0.9787
Zielonka	MEAN	-	7.87	12.75	0.5603	0.0000	0.3856
	LIN	PM ₁₀	2.51	3.77	0.2376	0.8811	0.9675
	MLP 1-10-1	D	7.11	10.12	0.9410	0.1295	0.4835
	MLP 1-10-1	PM ₁₀	2.50	3.77	0.2445	0.8815	0.9676
	MLP 2-10-1	D, PM ₁₀	2.38	3.63	0.2385	0.8901	0.9701

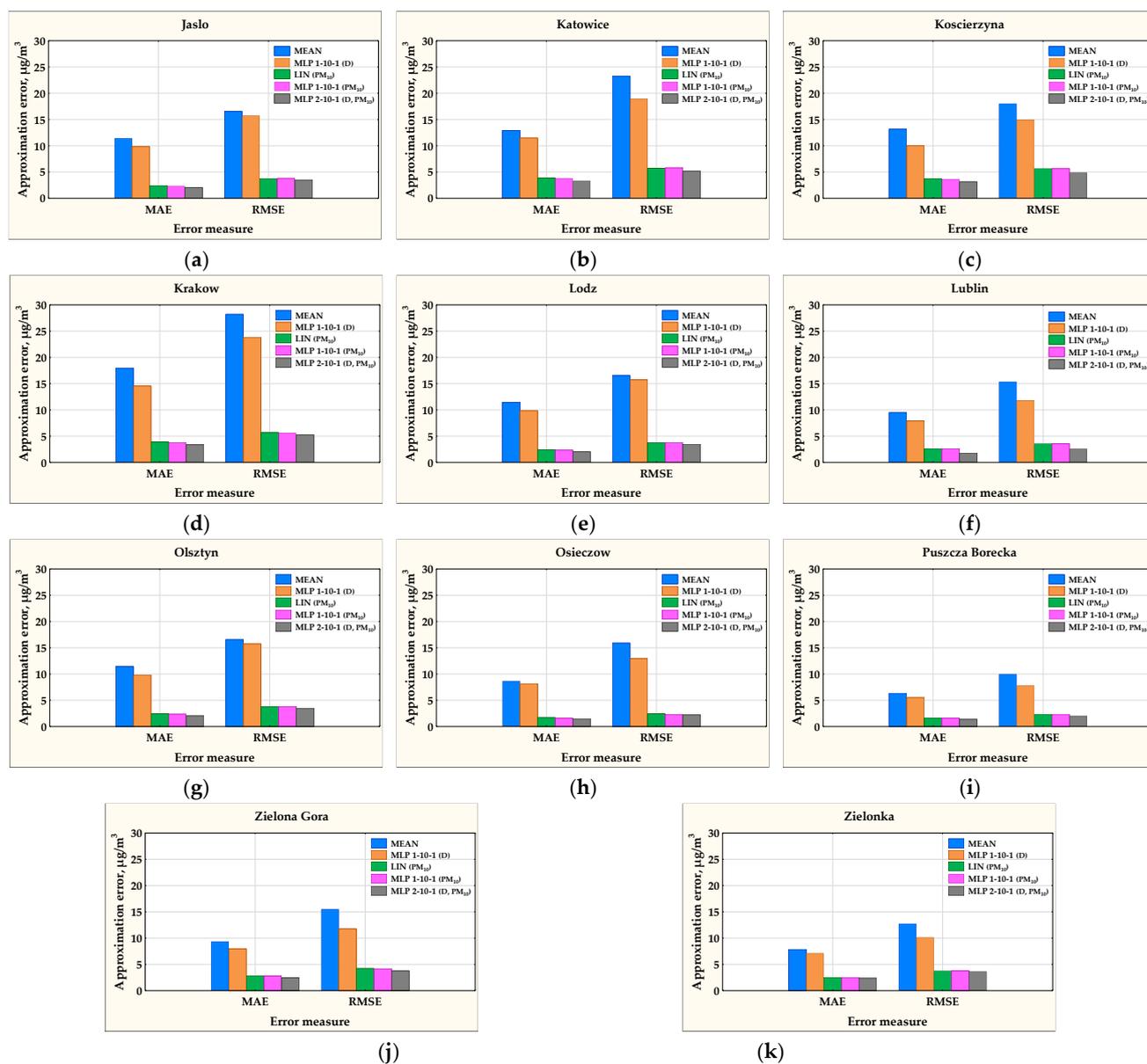


Figure 5. MAE and RMSE values for approximating PM_{2.5} concentrations in MEAN model, LINEAR model with PM₁₀ as predictor and MLP models with D and PM₁₀ predictors: (a) Jaslo, (b) Katowice, (c) Koscierzyna, (d) Krakow, (e) Lodz, (f) Lublin, (g) Olsztyn, (h) Osieczow, (i) Puszcza Borecka, (j) Zielona Gora, (k) Zielonka.

For all the stations, the largest modeling errors were achieved for the simplest neural models (Variant I), in which the only predictor was the time variable D (day). A very significant increase in the quality of modeling was noted for models in Variant II, in which the second predictor was the PM₁₀ variable, strongly correlated with the explained variable PM_{2.5}. These results are not surprising because PM₁₀ is potentially the most important explanatory variable for PM_{2.5}.

All naive MEAN models showed significantly lower accuracy than the MLP models. In turn, the LIN models were only slightly less accurate than the corresponding MLP models. To compare the predicted and observed PM_{2.5} concentration values at different air monitoring stations, the corresponding scatterplots are shown in Figure 6. The results are presented for models with two predictors (Variant III: MLP 2-10-1). The perfect fit lines (red lines: $y = x$) and regression lines (black lines) are also shown in the scatterplots, as well as linear regression equations and determination coefficients.

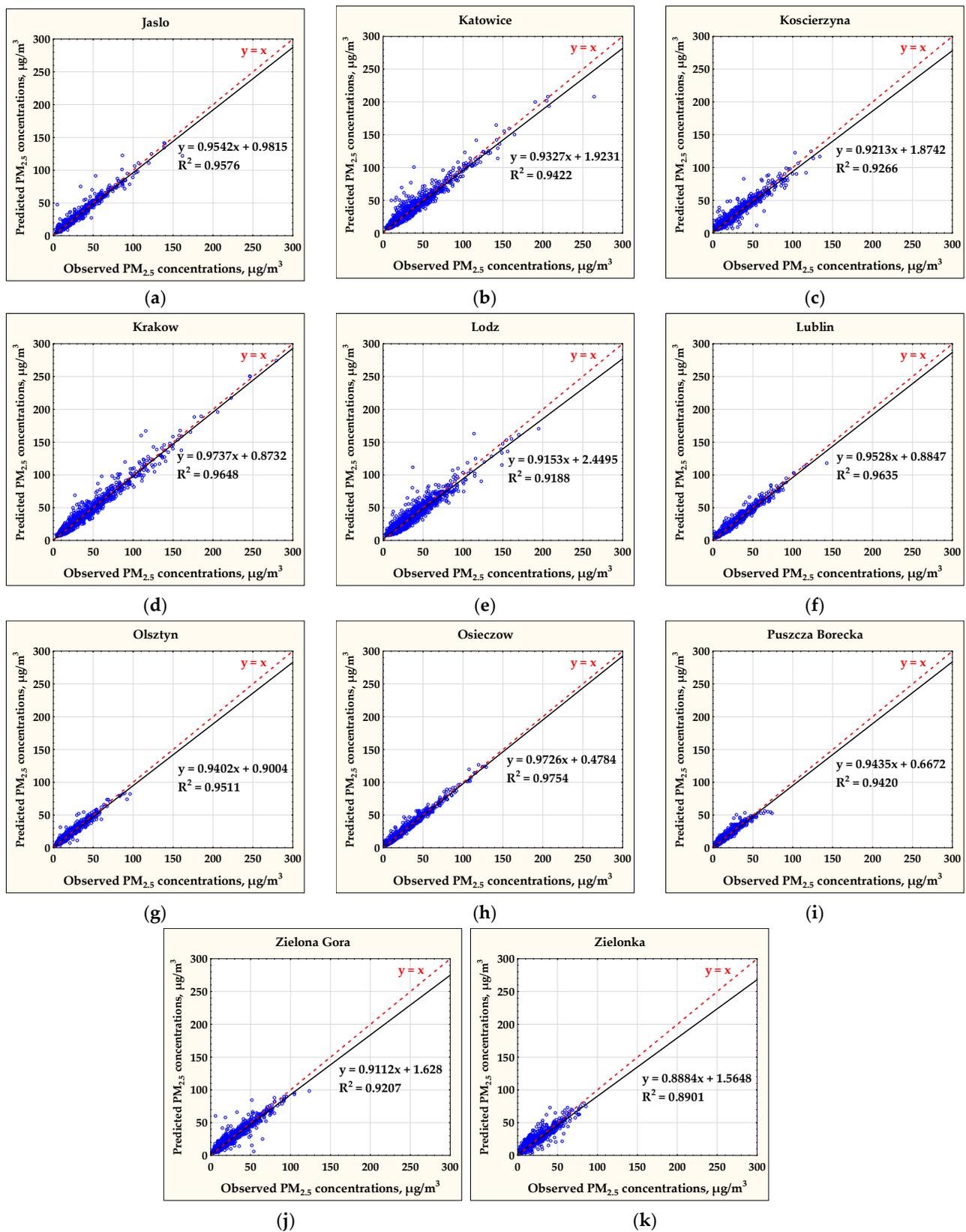


Figure 6. Scatterplots of predicted and observed PM_{2.5} concentrations for the MLP 2-10-1 models with D and PM₁₀ predictors: (a) Jaslo, (b) Katowice, (c) Koscierzyna, (d) Krakow, (e) Lodz, (f) Lublin, (g) Olsztyn, (h) Osieczow, (i) Puszcza Borecka, (j) Zielona Gora, (k) Zielonka.

3.4. Verification of the Models

Approach 1

The usefulness of the considered ANN models was verified by computing trial forecasts for a period of one month, using models trained on data recorded in the past at the same station. Figure 7 shows predicted $PM_{2.5}$ concentration courses for eight different locations and for different months in 2022. In each case, the network dedicated to a given station was used, i.e., the network trained on data from the same station. The graphs show actual courses and predicted $PM_{2.5}$ concentration courses obtained using MLP 2-10-1 models with numeric date D and PM_{10} concentration as predictors (the most accurate model). For each selected period, the R^2 value was calculated as a measure of the modeling accuracy.

Approach 2

In order to test the “universality” of the obtained MLP models, neural networks trained on data from one station were used to predict daily $PM_{2.5}$ concentrations at other air monitoring stations. The models created for three air monitoring stations, Krakow, Osieczow and Olsztyn, were tested in this way, and the corresponding networks for Variant III (MLP 2-10-1) were used. For each monitoring station, the prediction capability of each of these three “foreign” models was assessed. The modeling errors were calculated for all the implementation cases, and the results are presented in Tables 5–7.

Table 5. Prediction errors of 24 h $PM_{2.5}$ concentrations at individual stations. Applied model: MLP 2-10-1 trained on data from Krakow, 2010–2021.

Air Monitoring Station	Regression Model	Explanatory Variable (Predictors)	MAE $\mu\text{g}/\text{m}^3$	RMSE $\mu\text{g}/\text{m}^3$	MARE	R^2	d
Jaslo	MLP 2-10-1	D, PM_{10}	3.01	4.53	0.1463	0.9533	0.9796
Katowice	MLP 2-10-1	D, PM_{10}	3.40	5.38	0.1481	0.9401	0.9845
Koscierzyna	MLP 2-10-1	D, PM_{10}	3.30	5.10	0.2514	0.9204	0.9785
Krakow	MLP 2-10-1	D, PM_{10}	3.38	5.26	0.1340	0.9648	0.9910
Lodz	MLP 2-10-1	D, PM_{10}	4.13	6.21	0.2072	0.9046	0.9730
Lublin	MLP 2-10-1	D, PM_{10}	2.11	2.88	0.1369	0.9597	0.9877
Olsztyn	MLP 2-10-1	D, PM_{10}	2.04	2.84	0.1676	0.9408	0.9829
Osieczow	MLP 2-10-1	D, PM_{10}	2.17	3.18	0.1786	0.9707	0.9852
Puszcza Borecka	MLP 2-10-1	D, PM_{10}	1.76	2.35	0.2461	0.9307	0.9784
Zielona Gora	MLP 2-10-1	D, PM_{10}	2.63	3.95	0.1913	0.9203	0.9746
Zielonka	MLP 2-10-1	D, PM_{10}	2.66	3.81	0.3325	0.8800	0.9663

Table 6. Prediction errors of 24 h $PM_{2.5}$ concentrations at individual stations. Applied model: MLP 2-10-1 trained on data from Osieczow, 2010–2021.

Air Monitoring Station	Regression Model	Explanatory Variable (Predictors)	MAE $\mu\text{g}/\text{m}^3$	RMSE $\mu\text{g}/\text{m}^3$	MARE	R^2	d
Jaslo	MLP 2-10-1	D, PM_{10}	2.19	3.81	0.1215	0.9503	0.9872
Katowice	MLP 2-10-1	D, PM_{10}	4.24	7.76	0.1823	0.8886	0.9675
Koscierzyna	MLP 2-10-1	D, PM_{10}	3.84	6.25	0.2864	0.9158	0.9720
Krakow	MLP 2-10-1	D, PM_{10}	4.84	9.34	0.1790	0.9006	0.9712
Lodz	MLP 2-10-1	D, PM_{10}	5.69	8.43	0.2825	0.8826	0.9551
Lublin	MLP 2-10-1	D, PM_{10}	2.27	3.36	0.1514	0.9532	0.9853
Olsztyn	MLP 2-10-1	D, PM_{10}	2.26	3.47	0.1720	0.9315	0.9779
Osieczow	MLP 2-10-1	D, PM_{10}	1.48	2.22	0.1239	0.9754	0.9937
Puszcza Borecka	MLP 2-10-1	D, PM_{10}	1.48	2.32	0.1611	0.9366	0.9823
Zielona Gora	MLP 2-10-1	D, PM_{10}	2.61	4.03	0.1999	0.9150	0.9768
Zielonka	MLP 2-10-1	D, PM_{10}	2.53	4.09	0.2490	0.8842	0.9667

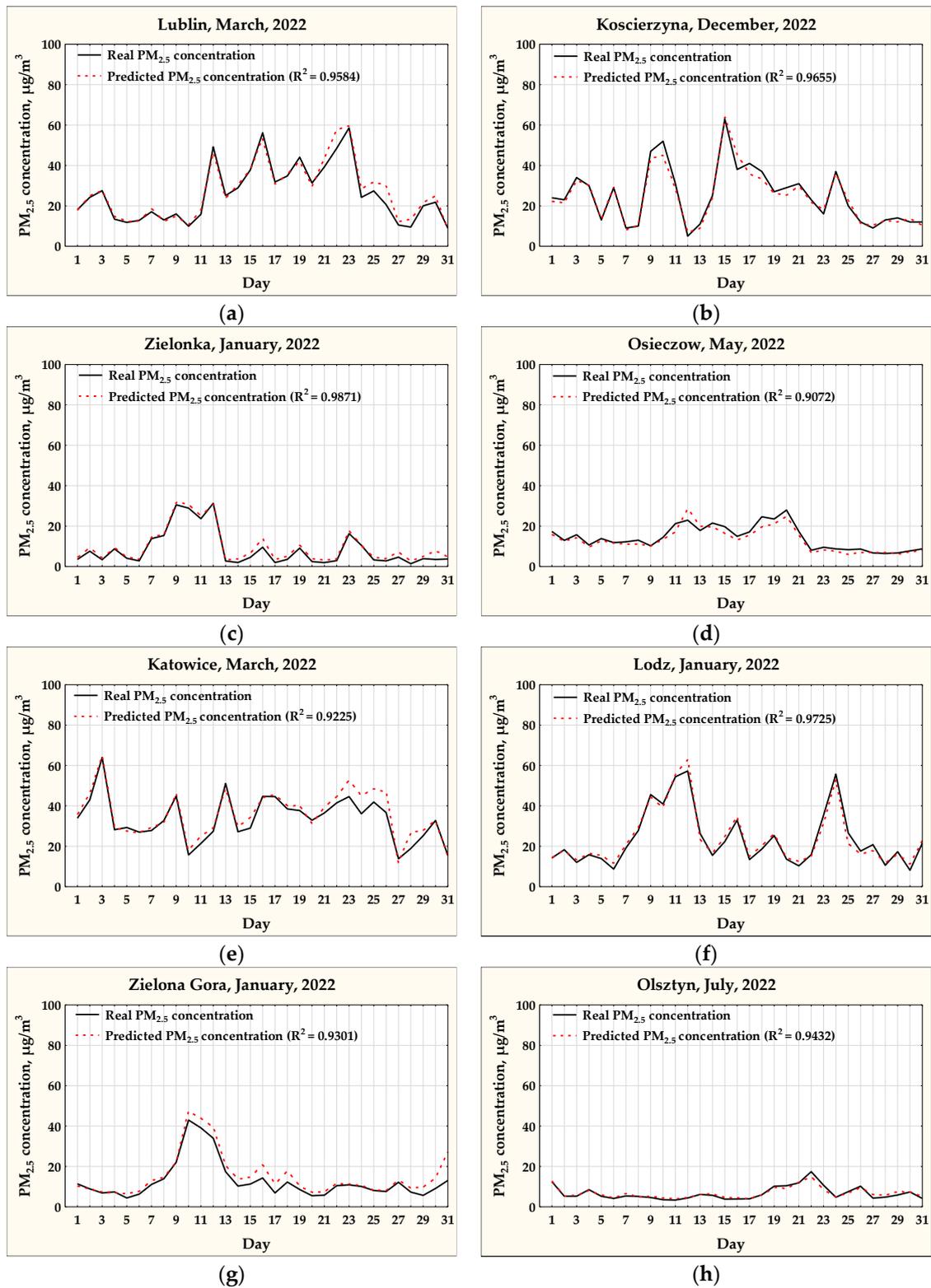


Figure 7. Example graphs of observed and modeled daily PM_{2.5} concentrations in selected months of 2022 using models of neural networks in Variant III (MLP 2-10-1), trained at the same station on the data from 2010 to 2021: (a) Lublin, March 2022; (b) Koscierzyzna, December 2022; (c) Zielonka, January 2022; (d) Osieczow, May 2022; (e) Katowice, March 2022; (f) Lodz, January 2022; (g) Zielona Gora, January 2022; (h) Olsztyn, July 2022.

Table 7. Prediction errors of 24 h PM_{2.5} concentrations at individual stations. Applied model: MLP 2-10-1 trained on data from Olsztyn, 2010–2021.

Air Monitoring Station	Regression Model	Explanatory Variable (Predictors)	MAE $\mu\text{g}/\text{m}^3$	RMSE $\mu\text{g}/\text{m}^3$	MARE	R ²	d
Jaslo	MLP 2-10-1	D, PM ₁₀	3.31	5.76	0.1572	0.9134	0.9661
Katowice	MLP 2-10-1	D, PM ₁₀	4.61	11.34	0.1630	0.7371	0.9061
Koscierzyna	MLP 2-10-1	D, PM ₁₀	3.48	5.97	0.2360	0.8925	0.9692
Krakow	MLP 2-10-1	D, PM ₁₀	6.03	15.86	0.1598	0.7119	0.8815
Lodz	MLP 2-10-1	D, PM ₁₀	4.35	8.22	0.1910	0.8226	0.9472
Lublin	MLP 2-10-1	D, PM ₁₀	2.14	3.76	0.1247	0.9281	0.9788
Olsztyn	MLP 2-10-1	D, PM ₁₀	1.72	2.52	0.1301	0.9511	0.9872
Osieczow	MLP 2-10-1	D, PM ₁₀	2.06	3.63	0.1433	0.9500	0.9811
Puszczka Borecka	MLP 2-10-1	D, PM ₁₀	1.48	2.17	0.1594	0.9360	0.9832
Zielona Gora	MLP 2-10-1	D, PM ₁₀	2.54	4.07	0.1761	0.9097	0.9741
Zielonka	MLP 2-10-1	D, PM ₁₀	2.46	3.87	0.2405	0.8788	0.9678

4. Summary and Discussion

The RMSE modeling errors compared to some PM_{2.5} and PM₁₀ concentration statistics are presented in Table 8.

Table 8. The RMSE modeling errors and some statistics on PM_{2.5} and PM₁₀ concentrations for considered air monitoring stations.

Air Monitoring Station	PM ₁₀ , $\mu\text{g}/\text{m}^3$		PM _{2.5} , $\mu\text{g}/\text{m}^3$		PM _{2.5} /PM ₁₀ Ratio, %	r-Pearson PM _{2.5} /PM ₁₀	RMSE, $\mu\text{g}/\text{m}^3$	
	Mean	SD	Mean	SD			Variant I (MLP 1-10-1)	Variant III (MLP 2-10-1)
Jaslo	27.2	18.3	21.9	16.6	0.79	0.9735	15.74	3.47
Katowice	36.6	26.5	26.7	21.5	0.71	0.9639	18.95	5.18
Koscierzyna	31.4	21.1	22.8	18.0	0.70	0.9487	14.97	4.89
Krakow	41.7	33.2	30.6	28.0	0.70	0.9792	23.83	5.26
Lodz	38.8	22.5	26.5	19.4	0.66	0.9339	15.63	5.54
Lublin	26.0	15.8	19.2	13.6	0.72	0.9646	11.84	2.59
Olsztyn	22.0	13.6	15.7	11.4	0.70	0.9493	9.98	2.52
Osieczow	19.9	15.3	15.5	14.1	0.74	0.9852	12.95	2.22
Puszczka Borecka	16.0	10.3	11.7	8.6	0.71	0.9611	7.79	2.06
Zielona Gora	23.0	14.7	17.2	13.3	0.72	0.9487	11.76	3.76
Zielonka	18.4	13.3	13.3	10.9	0.71	0.9387	10.12	3.63

The modeling of PM_{2.5} concentrations using only the time variable D (Variant I) is burdened with significant prediction errors. However, such models have certain advantages over completely simple naive models such as mean models. In the naive mean model, all modeling results are the same and equal to the mean. The RMSE error of such modeling is equal to the standard deviation (SD) in the set of actual concentrations (the same formula). For models of Variant I (MLP 1-10-1), the RMSE values at individual stations were in the range of 7.79–23.8 $\mu\text{g}/\text{m}^3$, while the standard deviations in the sets of actual PM_{2.5} concentrations were in the range of 8.6–28.0 $\mu\text{g}/\text{m}^3$, depending on the station. SD always achieved values definitely higher than RMSE. The prediction day is always known and available, so a model in Variant I can be easily created. The time variable D brings information about the time of year to the models, and this is sufficient to improve the modeling quality compared to the mean model.

The implementation of PM₁₀ concentration as an input significantly improved the quality of modeling at all stations (Variants II and III). RMSE values for models in Variant II were in the range of 2.31–6.86 $\mu\text{g}/\text{m}^3$ and in Variant III in the range of 2.06–5.54 $\mu\text{g}/\text{m}^3$. It should be emphasized that in the conditions occurring in Poland, the PM₁₀ concentrations

are usually strongly correlated with the $PM_{2.5}$ concentrations (correlation coefficients ranged from 0.934 (Lodz) to 0.985 (Osieczow)). The values of other error measures also confirm that the Variant III models are the most accurate. Therefore, models MLP 2-10-1 can be recommended for use in practice.

Practical modeling needs may result from various reasons. There may be a need to supplement missing $PM_{2.5}$ concentrations in data collected at air monitoring stations. The completeness of the $PM_{2.5}$ concentration time series is required to assess air quality and is also the basis for assessing mortality rates, such as the number of premature deaths. Completing missing data on the concentrations of pollutants such as $PM_{2.5}$ may be helpful in air quality management and environmental policy at local and regional levels.

Verification of the models in approach 1 showed the usefulness of the models in situations where historical data are available at the monitoring station. The results prove that models trained on historical data can be used to predict concentrations in periods other than those covered by measurements. In the episodes shown, which were randomly selected, the determination coefficients usually exceed 0.9 (Figure 7), which can be considered a good prediction quality. The prediction accuracy is only slightly inferior to the prediction accuracy determined for the tested measurement period.

Verification of the models in approach 2 was performed in order to test the “universality” of the obtained MLP models. Neural networks trained on data from one station were used to predict daily $PM_{2.5}$ concentrations at other air monitoring stations. Three different models were tested. In each case, the relatively good quality of modeling at “foreign” monitoring stations was confirmed. The determination coefficients range from 0.880 to 0.971 for the Krakow model (Table 5), from 0.883 to 0.975 for the Osieczow model (Table 6) and from 0.712 to 0.951 for the Olsztyn model (Table 7). Each of the tested models retained the ability to make reasonable predictions, although the accuracies of some models were clearly worse than others. However, it was found that the models have a certain universality. This means that “foreign” models can be used for modeling, but the lower accuracy of such models should be taken into account.

The presented research addresses the possibility of modeling the 24 h $PM_{2.5}$ concentrations, the so-called daily concentrations. Similar studies were previously carried out to check the possibility of modeling $PM_{2.5}$ concentrations averaged over 1 h measurement periods, i.e., for the so-called hourly concentrations [54]. In the previously studied 1 h models, an additional time variable—hour—had to be taken into account. Hourly concentrations are characterized by much greater variability than daily concentrations, which is why they are more difficult to model. However, the addition of the H (hour) variable enabled reasonably accurate modeling of hourly $PM_{2.5}$ concentrations. It was stated that neural regression models trained on the data from past years can be successfully used to model the current $PM_{2.5}$ concentrations. The results presented in this study confirm this conclusion.

The main trend of research is the search for new, more and more accurate methods of modeling air pollution concentrations. Our research went in the opposite direction towards finding models that were as simple and universal as possible and could be used at most air monitoring stations. The only condition is to monitor PM_{10} concentrations. In our models, PM_{10} concentration is the primary predictor. We have shown that models trained at other air monitoring stations, the so-called foreign models, can be successfully used to approximate $PM_{2.5}$ concentrations at a selected station. This is a novelty in modeling $PM_{2.5}$ concentrations. This modeling method provides new possibilities in air quality assessment. Approximate concentrations of the $PM_{2.5}$ fraction may be used to calculate mortality rates and other public health effects.

We are conscious of the limitations of the proposed methodology. The resulting models are accurate. The precision of these models can be improved, for example, by including additional predictors, such as concentrations of other pollutants, meteorological parameters and others. However, our goal was not to create the most accurate model possible. We

were looking for models that were as simple as possible and highly accessible. We also wanted to test the possibility of building universal models.

Future research may aim to look for models that fit specific data. Such research may lead to the use of more complex modeling tools, which have been described in many publications [51–53,65]. Research can also be conducted on segment modeling. Since differences in modeling accuracy were found in different concentration subranges [66,67], the improvement of modeling quality was tested by replacing a single model with a group of models dedicated to specific subranges of pollutant concentrations [68]. Promising results were obtained for segmented modeling.

The accuracy of the models can be increased by including other predictors that may influence PM concentration levels. These may be meteorological parameters affecting the emission of pollutants or the spread of pollutants in the air. Future research may also aim to find a more universal model that combines historical knowledge from measurements at various air monitoring stations.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/en17092202/s1>.

Author Contributions: Conceptualization, S.H.; methodology, S.H. and R.J.; software, R.J. and J.B.; validation, R.J.; formal analysis, S.H. and R.J.; resources, R.J.; data curation, R.J.; writing—original draft preparation, S.H.; writing—review and editing, S.H., R.J. and J.B.; visualization, R.J.; project administration, S.H.; funding acquisition, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the statute subvention of the Czestochowa University of Technology Faculty of Infrastructure and Environment BS/PB-400-301 and Faculty of Electrical Engineering BS/PB-3-300-301.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to commercial restrictions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gurjar, B.R.; Molina, L.T.; Ojha, C.S.P. *Air Pollution: Health and Environmental Impacts*; CRC Press: Boca Raton, FL, USA, 2010.
2. Kumar, P. *Airborne Particles: Origin, Emissions and Health Impacts*; Nova Science Publisher's, Inc.: Hauppauge, NY, USA, 2017.
3. Hoffmann, B.; Roebbel, N.; Gumy, S.; Forastiere, F.; Brunekreef, B.; Jarosinska, D.; Walker, K.D.; van Erp, A.M.; O'Keefe, R.; Greenbaum, D.; et al. A joint workshop report of ERS, WHO, ISEE and HEI. *Eur. Respir. J.* **2020**, *56*, 2002575. [[CrossRef](#)] [[PubMed](#)]
4. Pandya, S.; Gadekallu, T.R.; Maddikunta, P.K.R.; Sharma, R. A Study of the Impacts of Air Pollution on the Agricultural Community and Yield Crops (Indian Context). *Sustainability* **2022**, *14*, 13098. [[CrossRef](#)]
5. Wei, W.; Wang, Z. Impact of Industrial Air Pollution on Agricultural Production. *Atmosphere* **2021**, *12*, 639. [[CrossRef](#)]
6. Agathokleous, E.; Frei, M.; Knopf, O.M.; Muller, O.; Xu, Y.; Nguyen, T.H.; Gaiser, T.; Liu, X.; Liu, B.; Saitanis, C.J.; et al. Adapting crop production to climate change and air pollution at different scales. *Nat. Food* **2023**, *4*, 854–865. [[CrossRef](#)] [[PubMed](#)]
7. Chang, T.; Zivin, J.G.; Gross, T.; Neidell, M. Particulate Pollution and the Productivity of Pear Packers. *Am. Econ. J. Econ. Policy* **2016**, *8*, 141–169. [[CrossRef](#)]
8. Graff-Zivin, J.; Neidell, M. The Impact of Pollution on Worker Productivity. *Am. Econ. Rev.* **2012**, *102*, 3652–3673. [[CrossRef](#)] [[PubMed](#)]
9. Hanna, R.; Oliva, P. The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City. *J. Public Econ.* **2015**, *122*, 68–79. [[CrossRef](#)]
10. Aragon, F.; Miranda, J.; Oliva, P. Particulate Matter and Labor Supply: The Role of Caregiving and Non-linearities. *J. Environ. Econ. Manag.* **2017**, *86*, 295–309. [[CrossRef](#)]
11. Conti, S.; Ferrara, P.; D'Angiolella, L.S.; Lorelli, S.C.; Agazzi, G.; Fornari, C.; Cesana, G.; Mantovani, L.G. The economic impact of air pollution: A European assessment. *Eur. J. Public Health* **2020**, *30* (Suppl. 5), ckaa165.084. [[CrossRef](#)]
12. Vallero, D.A. *Fundamentals of Air Pollution*, 4th ed.; Academic Press: Cambridge, MA, USA, 2008.
13. Martinez, J. Great Smog of London. Encyclopedia Britannica, Article History. 27 February 2024. Available online: <https://www.britannica.com/event/Great-Smog-of-London> (accessed on 14 March 2024).
14. Maesano, I. The Air of Europe: Where Are We Going? *Eur. Respir. Rev.* **2017**, *26*, 170024. [[CrossRef](#)]

15. Tiotiu, A.I.; Novakova, P.; Nedeva, D.; Chong-Neto, H.J.; Novakova, S.; Steiropoulos, P.; Kowal, K. Impact of Air Pollution on Asthma Outcomes. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6212. [[CrossRef](#)] [[PubMed](#)]
16. Brito, F.F.; Gimeno, P.M.; Sánchez, J.F.; García, J.A.L.; Arias, T.A.; Ardanaz, J.M.U. Air Pollution and Asthma. In *The Dangers of Allergic Asthma*; García-Menaya, J.M., Ed.; Nova Science Publisher's, Inc.: Hauppauge, NY, USA, 2023. [[CrossRef](#)]
17. Kusumawardani, I.A.J.D.; Indraswari, G.; Komalasari, N.L.G.Y. Air Pollution and Lung Cancer. *J. Respirasi* **2023**, *9*, 150–158. [[CrossRef](#)]
18. Berg, C.D.; Schiller, J.H.; Boffetta, P.; Cai, J.; Connolly, C.; Kerpel-Fronius, A.; Kitts, A.B.; Lam, D.C.; Mohan, A.; Myers, R.; et al. Air Pollution and Lung Cancer: A Review by International Association for the Study of Lung Cancer Early Detection and Screening Committee. *J. Thorac. Oncol.* **2023**, *18*, 10. [[CrossRef](#)] [[PubMed](#)]
19. Brook, R.D.; Rajagopalan, S.; Pope, C.A., 3rd; Brook, J.R.; Bhatnagar, A.; Diez-Roux, A.V.; Holguin, F.; Hong, Y.; Luepker, R.V.; Mittleman, M.A.; et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation* **2010**, *121*, 2331–2378. [[CrossRef](#)] [[PubMed](#)]
20. Münzel, T.; Hahad, O.; Daiber, A.; Lelieveld, J. Luftverschmutzung und Herz-Kreislauf-Erkrankungen [Air pollution and cardiovascular diseases]. *Herz* **2021**, *46*, 120–128. (In German) [[CrossRef](#)]
21. de Bont, J.; Jaganathan, S.; Dahlquist, M.; Persson, Å.; Stafoggia, M.; Ljungman, P. Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses. *J. Intern. Med.* **2022**, *291*, 779–800. [[CrossRef](#)] [[PubMed](#)]
22. Li, J.; Xin, Y. Air Pollution and Cardiovascular Diseases. *J. Am. Coll. Cardiol.* **2023**, *81*, e97. [[CrossRef](#)] [[PubMed](#)]
23. Peterson, B.S.; Rauh, V.A.; Bansal, R.; Hao, X.; Toth, Z.; Nati, G.; Walsh, K.; Miller, R.L.; Arias, F.; Semanek, D.; et al. Effects of Prenatal Exposure to Air Pollutants (Polycyclic Aromatic Hydrocarbons) on the Development of Brain White Matter, Cognition, and Behavior in Later Childhood. *JAMA Psychiatry* **2015**, *72*, 531–540. [[CrossRef](#)] [[PubMed](#)]
24. Kim, Y.; Manley, J.; Radoias, V. Air Pollution and Long Term Mental Health. *Atmosphere* **2020**, *11*, 1355. [[CrossRef](#)]
25. Calderón-Garcidueñas, L.; Ayala, A. Air Pollution, Ultrafine Particles, and Your Brain: Are Combustion Nanoparticle Emissions and Engineered Nanoparticles Causing Preventable Fatal Neurodegenerative Diseases and Common Neuropsychiatric Outcomes? *Environ. Sci. Technol.* **2022**, *56*, 6847–6856. [[CrossRef](#)]
26. Peters, R.; Ee, N.; Peters, J.; Booth, A.; Mudway, I.; Anstey, K.J. Air Pollution and Dementia: A Systematic Review. *J. Alzheimers Dis.* **2019**, *70*, S145–S163. [[CrossRef](#)] [[PubMed](#)]
27. Clean Air Act. UK Public General Acts, 5 July 1956. Available online: <https://www.legislation.gov.uk/ukpga/Eliz2/4-5/52/enacted> (accessed on 14 March 2024).
28. Knox, A.; Evans, G.J.; Lee, C.J.; Brook, J.R. Air Pollution Monitoring and Sustainability. In *Encyclopedia of Sustainability Science and Technology*; Meyers, R.A., Ed.; Springer: New York, NY, USA, 2012. [[CrossRef](#)]
29. Spandana, G.; Shanmughasundram, R. Design and Development of Air Pollution Monitoring System for Smart Cities. In Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 14–15 June 2018; pp. 1640–1643. [[CrossRef](#)]
30. Xing, Y.F.; Xu, Y.H.; Shi, M.H.; Lian, Y.X. The impact of PM_{2.5} on the human respiratory system. *J. Thorac. Dis.* **2016**, *8*, E69–E74. [[PubMed](#)]
31. Liu, G.; Li, Y.; Zhou, J.; Xu, J.; Yang, B. PM_{2.5} deregulated microRNA and inflammatory microenvironment in lung injury. *Environ. Toxicol. Pharmacol.* **2022**, *91*, 103832. [[CrossRef](#)] [[PubMed](#)]
32. Behinaein, P.; Hutchings, H.; Knapp, T.; Okereke, I.C. The growing impact of air quality on lung-related illness: A narrative review. *J. Thorac. Dis.* **2023**, *15*, 5055–5063. [[CrossRef](#)] [[PubMed](#)]
33. European Environment Agency. *Air Quality in Europe-2020 Report*. No. 12/2018; Publications Office of the European Union: Luxembourg, 2020.
34. World Health Organization. New WHO Global Air Quality Guidelines Aim to Save Millions of Lives from Air Pollution. 2021. Available online: <https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution> (accessed on 14 March 2024).
35. EN 12341:2014; Ambient Air—Standard Gravimetric Measurement Method for the Determination of the PM₁₀ or PM_{2.5} Mass Concentration of Suspended Particulate Matter. iTech, Inc.: Newark, DE, USA, 2014.
36. Hammit, J.K.; Morfeld, P.; Tuomisto, J.T.; Erren, T.C. Premature Deaths, Statistical Lives, and Years of Life Lost: Identification, Quantification, and Valuation of Mortality Risks. *Risk Anal.* **2020**, *40*, 674–695. [[CrossRef](#)] [[PubMed](#)]
37. Ministry of Climate and Environment (Polish Government). Regulation on the Evaluation of Levels of Substances in the Air. 11 December 2020. Available online: <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20200002279> (accessed on 12 March 2024). (In Polish)
38. Milionis, A.E.; Davies, T.D. Regression and Stochastic Models for Air Pollution-I. Review, Comments and Suggestions. *Atmos. Environ.* **1994**, *28*, 2801–2810. [[CrossRef](#)]
39. Manly, B.F.J. *Statistics for Environmental Science and Management*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2001.
40. Peng, G.; Leslie, L.M.; Shao, Y. *Environmental Modeling and Prediction*; Springer: Berlin/Heidelberg, Germany, 2002.
41. Plaia, A.; Bondi, A.L. Single Imputation Method of Missing Values in Environmental Pollution Data Sets. *Atmos. Environ.* **2006**, *40*, 7316–7330. [[CrossRef](#)]
42. Gardner, M.W.; Dorling, S.R. Artificial Neural Networks (the Multilayer Perceptron)-A Review of Applications in the Atmospheric Sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [[CrossRef](#)]

43. Dorling, S.R.; Gardner, M.W. Statistical Surface Ozone Models: An Improved Methodology to Account for Non-linear Behaviour. *Atmos. Environ.* **2000**, *34*, 21–34.
44. Hoffman, S. Short-Time forecasting of atmospheric NO_x concentration by neural networks. *Environ. Eng. Sci.* **2006**, *23*, 603–609. [[CrossRef](#)]
45. Gentili, S.; Magnaterra, L.; Passerini, G. *Handling Missing Data: Applications to Environmental Analysis*; Latini, G., Passerini, G., Eds.; Wit Press: Southampton, UK, 2004.
46. Hoffman, S. Missing data completing in the air monitoring systems by forward and backward prognosis methods. *Environ. Protec. Eng.* **2006**, *32*, 25–29.
47. Hoffman, S. Treating missing data at air monitoring stations. In *Environmental Engineering*; Pawłowski, L., Dudzińska, M., Pawłowski, A., Eds.; Taylor & Francis Group: London, UK, 2007; pp. 349–353.
48. Hoffman, S. Approximation of Imission Level at Air Monitoring Stations by Means of Autonomous Neural Models. *Environ. Prot. Eng.* **2012**, *38*, 109–119. [[CrossRef](#)]
49. Lin, W.C.; Tsai, C.F. Missing value imputation: A review and analysis of the literature (2006–2017). *Artif. Intell. Rev.* **2020**, *53*, 1487–1509. [[CrossRef](#)]
50. Shams, S.R.; Jahani, A.; Kalantary, S.; Moeinaddini, M.; Khorasani, N. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO₂ concentration. *Urban Clim.* **2021**, *37*, 100837. [[CrossRef](#)]
51. Rijal, N.; Gutta, R.T.; Cao, T.; Lin, J.; Bo, Q.; Zhang, J. Ensemble of Deep Neural Networks for Estimating Particulate Matter from Images. In Proceedings of the IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 27–29 June 2018; pp. 733–738. [[CrossRef](#)]
52. Chae, S.; Shin, J.; Kwon, S.; Lee, S.; Kang, S.; Lee, D. PM₁₀ and PM_{2.5} real-time prediction models using an interpolated convolutional neural network. *Sci. Rep.* **2021**, *11*, 11952. [[CrossRef](#)] [[PubMed](#)]
53. Zhang, B.; Rong, Y.; Yong, R.; Qin, D.; Li, M.; Zou, G.; Pan, J. Deep learning for air pollutant concentration prediction: A review. *Atmos. Environ.* **2022**, *290*, 119347. [[CrossRef](#)]
54. Hoffman, S.; Jasiński, R. The Use of Multilayer Perceptrons to Model PM_{2.5} Concentrations at Air Monitoring Stations in Poland. *Atmosphere* **2023**, *14*, 96. [[CrossRef](#)]
55. Duan, J.; Chen, Y.; Fang, W.; Su, Z. Characteristics and Relationship of PM, PM₁₀, PM_{2.5} Concentration in a Polluted City in Northern China. *Procedia Eng.* **2015**, *102*, 1150–1155. [[CrossRef](#)]
56. Colangeli, C.; Palermi, S.; Bianco, S.; Aruffo, E.; Chiacchiaretta, P.; Di Carlo, P. The Relationship between PM_{2.5} and PM₁₀ in Central Italy: Application of Machine Learning Model to Segregate Anthropogenic from Natural Sources. *Atmosphere* **2022**, *13*, 484. [[CrossRef](#)]
57. Chief Inspectorate of Environmental Protection (Poland)—Measurement Data Bank. Available online: <https://powietrze.gios.gov.pl/pjp/archives> (accessed on 12 March 2024).
58. Statistica. Electronic Textbook, 1984–2017, Available in the STATISTICA 13.3 Program.
59. Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2000.
60. Broyden, C.G. The convergence of a class of double-rank minimization algorithms. *J. Inst. Math. Its Appl.* **1970**, *6*, 76–90. [[CrossRef](#)]
61. Fletcher, R. A New Approach to Variable Metric Algorithms. *Comput. J.* **1970**, *13*, 317–322. [[CrossRef](#)]
62. Goldfarb, D. A Family of Variable Metric Updates Derived by Variational Means. *Math. Comput.* **1970**, *24*, 23–26. [[CrossRef](#)]
63. Shanno, D.F. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* **1970**, *24*, 647–656. [[CrossRef](#)]
64. Willmott, C.J. On the validation of models. *Phys. Geogr.* **1981**, *2*, 184–194. [[CrossRef](#)]
65. Li, X.; Peng, L.; Hu, Y.; Shao, J.; Chi, T. Deep learning architecture for air quality predictions. *Environ. Sci. Pollut. Res.* **2016**, *23*, 22408–22417. [[CrossRef](#)] [[PubMed](#)]
66. Hoffman, S. Assessment of Prediction Accuracy in Autonomous Air Quality Models. *Desalination Water Treat.* **2015**, *57*, 1322–1326. [[CrossRef](#)]
67. Hoffman, S. Estimation of Prediction Error in Regression Air Quality Models. *Energies* **2021**, *14*, 7387. [[CrossRef](#)]
68. Hoffman, S.; Filak, M.; Jasiński, R. Air Quality Modeling with the Use of Regression Neural Networks. *Int. J. Environ. Res. Public Health* **2022**, *19*, 16494. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.