



Article Air Quality Modeling with the Use of Regression Neural Networks

Szymon Hoffman *[®], Mariusz Filak and Rafał Jasiński

Faculty of Infrastructure and Environment, Czestochowa University of Technology, 69 Dabrowskiego St., 42-200 Czestochowa, Poland

* Correspondence: szymon.hoffman@pcz.pl

Abstract: Air quality is assessed on the basis of air monitoring data. Monitoring data are often not complete enough to carry out an air quality assessment. To fill the measurement gaps, predictive models can be used, which enable the approximation of missing data. Prediction models use historical data and relationships between measured variables, including air pollutant concentrations and meteorological factors. The known predictive air quality models are not accurate, so it is important to look for models that give a lower approximation error. The use of artificial neural networks reduces the prediction error compared to classical regression methods. In previous studies, a single regression model over the entire concentration range was used to approximate the concentrations of a selected pollutant. In this study, it was assumed that not a single model, but a group of models, could be used for the prediction. In this approach, each model from the group was dedicated to a different sub-range of the concentration of the modeled pollutant. The aim of the analysis was to check whether this approach would improve the quality of modeling. A long-term data set recorded at two air monitoring stations in Poland was used in the examination. Hourly data of basic air pollutants and meteorological parameters were used to create predictive regression models. The prediction errors for the sub-range models were compared with the corresponding errors calculated for one full-range regression model. It was found that the application of sub-range models reduced the modeling error of basic air pollutants.

Keywords: air quality; air monitoring; air pollutants; regression; prediction; prediction error; artificial neural networks

1. Introduction

Air pollution is considered one of the main factors affecting the human population and the environment. The World Health Organization estimates that many millions of people die prematurely due to poor air quality [1]. Once pollutants are emitted into the air, it is impossible to stop them. If pollutants get into the atmosphere, they contribute to the deterioration of air quality in the vicinity of the emission, but by spreading they can have negative effects hundreds and thousands of kilometers from the point of emission. Therefore, air pollution is treated as a global threat and emission reduction strategies are implemented in many countries. It should be highlighted that, in Poland, the PM_{10} (particulate matter) concentration still exceeds the permissible limits and causes premature deaths of over 40,000 people every year [2].

Air pollution causes negative changes in the human respiratory and circulatory systems, even when pollutant concentrations do not exceed permissible levels [3–8]. It may trigger various reactions of organisms, including mental health disorders [9,10]. It was also found that air pollution can negatively affect the economy [11–14]. High concentrations of air pollutants have a negative effect on plants, which is manifested in the reduction of crop yields in agriculture [15,16].

The control and reduction of anthropogenic emissions are now recognized as the keys to good global air quality. An important element of the control system is the assessment of



Citation: 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. https:// doi.org/10.3390/ijerph192416494

Academic Editor: Paul B. Tchounwou

Received: 2 November 2022 Accepted: 6 December 2022 Published: 8 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/). air quality. This task is performed using air quality monitoring. Air monitoring includes continuous measurements of air pollutants. The main air pollutants include O_3 , SO_2 , NO_x , CO and PM_{10} [17]. Concentrations of the mentioned pollutants can be measured with automatic air monitoring stations. These types of stations are often equipped with devices that also continuously measure meteorological data such as temperature, wind direction, wind speed and solar radiation. Measurement results at automatic air monitoring stations are recorded in the form of 1 h averages, e.g., hourly concentrations, hourly temperature, etc. In the EU, average hourly concentrations are the basis for calculating averages over longer periods of time that are required for air quality standards, such as 8 h, 24 h and annual concentrations [18]. The system acquires and collects measurement data of the air pollutants' concentration levels at many individual air monitoring stations, according to the standardized measurement methods [19–23]. The collected hourly concentrations in the air monitoring system constitute the basis for direct and indirect statistical evaluation of air quality in the zones represented with individual monitoring stations, in accordance with the procedures described in the relevant legal acts [18]. Correct assessment requires a high degree of completeness of the time series of concentrations obtained at the monitoring stations. Usually, the completeness should exceed 90% in an annual series of hourly measurements. Unfortunately, monitoring data are never 100% complete in annual terms, and often they do not even have the completeness required by the regulations. When there is a deficit of data in a series of measurements, then there is a need to complete the missing data.

Missing data can be supplemented by introducing modeled concentrations in the measurement gaps [24–27]. The time series data obtained with air monitoring have specific characteristics. All measurements, both of concentrations and meteorological parameters, are performed simultaneously and are recorded in similar time series. Therefore, methods based on auto-regression and regression can be used for modeling monitoring concentrations. If historical data from the selected air monitoring station are available, they can be used to explore the knowledge hidden in them. Autonomous models of this type can be accurate and have a very significant advantage: the approximation of concentrations does not require external data from outside the monitoring system [26,27]. In the first models, classical statistical regression techniques were used [28]. Classical methods are increasingly being replaced by methods that use machine learning artificial intelligence, including artificial neural networks (ANNs) [29–37]. ANN models enable a deeper exploration of knowledge hidden in historical data and, as a result, more accurate concentration predictions.

In regression modeling, it was found that the application of one neural network to the entire range of concentrations of the predicted pollutant resulted in different prediction accuracies in the concentration sub-ranges [38,39]. It was considered advisable to replace one neural network with several networks (sub-models), each of which would be adjusted to specific concentration sub-ranges [39]. The use of several sub-range models should improve the accuracy of the prediction. This paper presents an analysis that verifies the above thesis.

The main aim of the study was to improve the accuracy of prediction of air pollutant concentrations in neural regression models, using many predictive models created for various sub-ranges of air pollutant concentrations. The analyzed data came from two air monitoring stations in the Upper Silesian Region, Poland. The quality of prediction was assessed separately for the concentrations of six main air pollutants: O_3 , NO, NO₂, SO₂, PM₁₀ and CO. Multi-layer perceptrons with an identical architecture were used to model all pollutants. The predicted concentrations were compared with the observed ones to estimate the prediction error. The prediction errors were calculated for various sub-models, and, based on them, the prediction error was estimated over the entire concentration range. This error was compared with the approximation error obtained from a single model covering the entire concentration range.

2. Materials and Methods

2.1. Air Monitoring Data

Hourly data from two air monitoring stations located in a highly industrialized region of Poland were used for the analysis. The data were recorded in Zabrze and Złoty Potok in the years 2011–2016. The location of selected monitoring stations is shown in Figure 1.



Figure 1. Locations of the air monitoring stations Zabrze and Złoty Potok in Poland.

The city of Zabrze is one of the most polluted towns in Poland and throughout the EU. The station at Zabrze was an urban background monitoring site. Złoty Potok is a rural town, located outside the Upper Silesian Agglomeration. In Złoty Potok, there is a background monitoring station for the Upper Silesian Region. The data were provided by the Voivodeship Inspectorate of Environmental Protection in Katowice. Time series data, including 1 h average values of O₃, NO, NO₂, SO₂, PM₁₀ and CO concentrations, as well as meteorological data for temperature, wind speed, solar radiation and relative humidity, were recorded. The analyzed data sets are not publicly available. We received them on an individual request from the Voivodeship Inspectorate for Environmental Protection. Time series of concentrations are available on the website of the Chief Inspectorate of Environmental Protection: https://powietrze.gios.gov.pl/pjp/archives, accessed on 30 April 2022. This is only part of the data. Meteorological data are not available online. The data also include two variables describing the time: day and hour. These two variables were converted to numeric form following the procedure described in [39].

The following symbols were used to describe the time series (variables):

D—numeric day,

H-numeric hour,

 O_3 —hourly average of O_3 concentration ($\mu g/m^3$),

NO—hourly average of NO concentration ($\mu g/m^3$),

NO₂—hourly average of NO₂ concentration ($\mu g/m^3$),

SO₂—hourly average of SO₂ concentration (μ g/m³),

CO—hourly average of CO concentration (mg/m^3) ,

 PM_{10} —hourly average of PM_{10} concentration ($\mu g/m^3$),

WS—hourly average of wind speed (m/s),

T—hourly average of air temperature (°C),

I—hourly average of solar radiation intensity (W/m^2) ,

RH—hourly average of relative humidity (%). Descriptive statistics of the 6-year set of hourly concentrations of monitored air pollutants are presented in Table 1.

Table 1. Descriptive statistics of the set of hourly concentrations of monitored air pollutants,2011–2016.

	Zabrze						Złoty Potok					
Descriptive Statistics	O ₃ µg/m ³	NO μg/m ³	NO ₂ μg/m ³	SO ₂ µg/m ³	ΡM ₁₀ μg/m ³	CO mg/m ³	O ₃ µg/m ³	NO μg/m ³	NO ₂ μg/m ³	SO ₂ µg/m ³	ΡM ₁₀ μg/m ³	CO mg/m ³
Minimum	0.7	0.0	1.0	0.1	0.1	0.07	1.1	0.0	0.0	0.0	1.0	-
Maximum	198.0	709.0	160.0	362.0	1187.0	9.02	212.6	50.0	82.0	116.0	298.0	-
Mean	42.3	10.2	24.4	16.8	51.0	0.60	59.5	1.1	9.6	7.4	27.8	-
Median	37.0	3.0	20.0	10.0	33.8	0.42	56.0	0.97	7.0	4.6	22.3	-
Standard deviation	32.4	24.2	16.8	19.4	61.1	0.59	31.0	1.8	7.9	8.1	20.7	-
Completeness	94.2	78.8	94.5	91.0	80.2	94.2	92.3	87.1	87.1	88.5	95.7	-

2.2. Regression Models

Multi-layer perceptrons with identical architecture were used to model all pollutants. The predicted concentrations were compared with the observed ones to estimate the prediction error. The prediction errors were calculated for various sub-models, and, based on them, the prediction error was estimated over the entire concentration range. This error was compared with the approximation error obtained from a single model covering the entire concentration range. For all air pollutants, similar perceptron models were created. The output of this was the concentration of the chosen air pollutant (explained variable), and the inputs were the date and hour, concentrations of other air pollutants and meteorological parameters (explanatory variables). Table 2 presents the predicted variables and predictors, with separate lines for individual models. For example, the following 11 input variables were used to model the ozone concentration in Złoty Potok: H, D, NO, NO₂, SO₂, CO, PM₁₀, WS, T, I and RH.

Table 2. Classification of predictors for individual models at both air monitoring stations. A + symbol means the variable was treated as an input in the model. A - symbol means no measurement.

	F 1 ' 1 Y ' 1 1	Explanatory Variables (Predictors)											
Locality	Explained Variable	Н	D	O ₃	NO	NO ₂	SO_2	СО	PM ₁₀	WS	Т	Ι	RH
	O3	+	+		+	+	+	+	+	+	+	+	_
Zabrze	NO	+	+	+		+	+	+	+	+	+	+	_
	NO ₂	+	+	+	+		+	+	+	+	+	+	_
	SO ₂	+	+	+	+	+		+	+	+	+	+	_
	CO	+	+	+	+	+	+		+	+	+	+	_
	PM_{10}	+	+	+	+	+	+	+		+	+	+	_
	O ₃	+	+		+	+	+	+	+	+	+	+	+
	NO	+	+	+		+	+	+	+	+	+	+	+
Złoty Potok	NO ₂	+	+	+	+		+	+	+	+	+	+	+
Złoty Potok	SO ₂	+	+	+	+	+		+	+	+	+	+	+
	CO	_	_	_	_	_	_	_	_	_	_	_	_
	PM ₁₀	+	+	+	+	+	+	+		+	+	+	+

For regression modeling, artificial neural networks in the form of multi-layer perceptrons were used. Each perceptron consisted of input neurons, with 10 neurons in one hidden layer and one output neuron. Figure 2 shows such a perceptron used to model O₃ concentrations. The Broyden–Fletcher–Goldfarb-Shanno (BFGS) algorithm was used in the learning process. The Broyden–Fletcher–Goldfarb–Shanno algorithm is used in problems related to numerical optimization [40]. The BFGS algorithm was developed on the basis of solutions proposed in 1970 by the four mathematicians mentioned in its name [41–44]. The algorithm uses an iterative method of solving unlimited non-linear optimization problems.



Figure 2. An architecture diagram of the multi-layer perceptron with ten neurons in a single hidden layer.

The learning process was always limited to 300 epochs. An epoch is a specialized expression related to the learning process of a neural network. In the network learning process, one epoch means a single learning cycle. In the learning process, the network repeats the cycles many times in order to minimize the learning error. The function of activating hidden and output neurons was a logistic function. The scales were initiated randomly. The network initialization was random (Gaussian). Modeling was performed using the Statistica program, version 13.3. Each prediction was performed 5 times. Repeating the training of the neural network, while maintaining the same parameters of the learning process, is a routine procedure. Each model's learning process is somewhat random and leads to a different network. The created networks have the same structure of neurons, but differ in terms of weights and the degree of activation of individual neurons. In general, they differ slightly in the modeling error. The most accurate of the 5 created models was selected for reporting. The other models were rejected. The sum of squares (SOS) was assumed as the error function. SOS is the sum of the squared distances of all predicted values from the actual values.

2.3. Preparation of Data for Modeling

The complete set of hourly air monitoring data from the 6-year measurement period should have included 52,608 cases (hourly observations). Prior to modeling, raw data from the air monitoring database were prepared by removing the cases where there were missing data. After removing the cases with missing data, a set of cases with complete data was obtained for further analysis. This set was called the full-range set. The full-range sets included 36,460 and 15,536 cases, for Zabrze and Złoty Potok stations, respectively. Before starting the learning process, each data set was divided into three subsets: the training subset consisted of 70% of the cases, the validation subset consisted of 15% of the cases.

Two approaches were adopted during modeling and, therefore, two groups of models were developed for each of the stations. In the first approach, all cases in the data set were sorted according to the real value of the predicted variable, from the lowest concentration to the highest. Then, a set of cases sorted in this way was divided into subsets. The entire set of cases was divided gradually into 2, then 4, and finally 8 subsets of the same size. The subsets prepared in this way differed in the ranges of the real concentration values of the predicted variable. For each of these subsets, regression models called RVS (Real Values Sorting) were created. As a result, for each set/subset, a separate predictive model was obtained, marked with the same symbol as the modeled set/subset. The name of the monitoring station (ZAB or ZP) was also added to the designations of sets and subsets. The full-range model was marked as RVS-1/1. As a result of the division of the full-range set, the following subsets and submodels were obtained:

Two sub-models (RVS-1/2, RVS-2/2), after division into two subsets;

Four sub-models (RVS-1/4, RVS-2/4, RVS-3/4, RVS-4/4), after division into four subsets;
Eight sub-models (RVS-1/8, RVS-2/8, RVS-3/8, RVS-4/8, RVS-5/8, RVS-6/8, RVS-7/8, RVS-8/8), after division into eight subsets.

In a situation where the real concentrations of the pollutant to be modeled were not known, the RVS-type models could not be used because it was not possible to sort and classify the cases into the real concentration sub-ranges. Therefore, RVS models reflect only the potential, not practical, opportunities to improve the quality of modeling through segmentation of the prediction process. If the real concentrations are not known and there is a need to predict them, then the RVS sub-range models are not available. In such a situation, a different approach can be proposed, which also uses division into sub-ranges, but then with sectoral modeling in designated sub-ranges. The most important step in this approach is the initial modeling of the concentrations of the selected pollutant in the full-range set, understood as a set of all cases containing complete data of all explanatory variables (model inputs) for the modeled pollutant. After initial modeling of the entire range of cases, predictive concentrations of the dependent variable are obtained. Then, the entire set of cases is sorted according to the increasing predictive concentrations of the modeled pollutant. The next step is to divide the sorted full-range set of cases into a specified number of equal sub-ranges. In this way, you can generate a model called PVS (Predicted Values Sorting). As a result of the division of the full-range set, the following subsets and submodels were obtained:

- Two sub-models (PVS-1/2, PVS-2/2), after division into two subsets;

Four sub-models (PVS-1/4, PVS-2/4, PVS-3/4, PVS-4/4), after division into four subsets;
Eight sub-models (PVS-1/8, PVS-2/8, PVS-3/8, PVS-4/8, PVS-5/8, PVS-6/8, PVS-7/8, PVS-8/8), after division into eight subsets.

The scheme of dividing the full-range set into sub-ranges is shown in Figures 3 and 4 for both monitoring stations.



Figure 3. The scheme of the division into sub-ranges for the Zabrze station.



Figure 4. The scheme of the division into sub-ranges for the Złoty Potok station.

Thanks to the divisions of the full-range set, it was possible to check how the modeling accuracy changed in the concentration sub-ranges, and whether the modeling carried out in the sub-ranges would improve the modeling quality in relation to the full-range modeling.

2.4. Assessment of the Approximation Error

To assess the accuracy of the obtained regression models, the MAE and RMSE values were used, which were calculated on the basis of the discrepancy between the actual and predicted values. The formulas for calculating individual errors are presented in Equations (1) and (2).

MAE—mean absolute error

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

RMSE—root-mean-squared error

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

where *n*—number of cases, *y*—predicted concentrations, *x*—real concentrations, *i*—the case number.

3. Results

For each pollutant, modeling errors were calculated in relation to the real pollutant concentrations (O_3 , NO, NO_2 , SO_2 , CO and PM_{10} for the Zabrze station; and O_3 , NO, NO_2 , SO_2 and PM_{10} for the Zloty Potok station). To assess the modeling accuracy, two error measures were calculated: MAE and RMSE.

3.1. The Results of the Modeling of O₃ Concentrations

Tables 3–6 show the results of O_3 concentration predictions obtained with the fullrange and sub-range models. The errors of the PVS models for the Zabrze monitoring station are presented in Table 3 and the Złoty Potok monitoring station in Table 4. Similar lists of errors for the RVS models are presented in Tables 5 and 6. The presented results show that the sub-range modeling errors changed for individual sub-ranges. Regardless of the number of sub-range models, the modeling error usually increased with increasing concentration values in the sub-ranges. A good way to assess the quality of the prediction is to compare the overall prediction error over the entire range of concentrations of the modeled pollutant, and not in individual sub-ranges. Therefore, the tables include average values for the entire ranges: "overall MAE" and "overall RMSE".

Number of Sub-Ranges	Regression Model	Sub-Ranges of O ₃ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZAB	1.0–165.0	36,460	8.4	8.4	11.3	11.3
2	PVS-1/2-ZAB PVS-2/2-ZAB	1.0–37.4 37.4–165.0	18,230 18,230	6.1 10.4	8.25	8.5 13.4	10.94
4	PVS-1/4-ZAB PVS-2/4-ZAB PVS-3/4-ZAB PVS-4/4-ZAB	1.0–17.7 17.7–37.4 37.4–59.9 59.9–165.0	9115 9115 9115 9115 9115	3.5 8.45 9.84 10.3	8.02	5.0 10.8 12.7 13.2	10.43
8	PVS-1/8-ZAB PVS-2/8-ZAB PVS-3/8-ZAB PVS-4/8-ZAB PVS-5/8-ZAB PVS-6/8-ZAB PVS-7/8-ZAB PVS-8/8-ZAB	$\begin{array}{c} 1.0-8.6\\ 8.6-17.7\\ 17.7-27.5\\ 27.5-37.4\\ 37.4-47.6\\ 47.6-59.9\\ 59.9-79.5\\ 79.5-165.0\end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	1.98 4.81 7.47 9.4 9.91 9.98 9.78 10.81	8.02	2.8 6.3 9.6 11.9 12.7 12.9 12.7 13.7	10.34

Table 3. Values of approximation errors of O₃ concentrations in PVS models, Zabrze 2011–2016.

Table 4. Values of approximation errors of O₃ concentrations in PVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of O ₃ Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZP	1.6–147.5	15,536	8.41	8.41	10.7	10.7
2	PVS-1/2-ZP PVS-2/2-ZP	1.6–50.3 50.3–147.5	7768 7768	7.66 8.16	7.91	9.9 10.3	10.09
4	PVS-1/4-ZP PVS-2/4-ZP PVS-3/4-ZP PVS-4/4-ZP	1.6–34.8 34.8–50.3 50.3–68.7 68.7–147.5	3884 3884 3884 3884 3884	6.22 8.52 7.81 7.64	7.55	8.2 10.9 10.0 9.6	9.68
8	PVS-1/8-ZP PVS-2/8-ZP PVS-3/8-ZP PVS-4/8-ZP PVS-5/8-ZP PVS-6/8-ZP PVS-7/8-ZP PVS-8/8-ZP	$\begin{array}{c} 1.6-25.3\\ 25.3-34.8\\ 34.8-42.9\\ 42.9-50.3\\ 50.3-58.4\\ 58.4-68.7\\ 68.7-89.9\\ 90.0-147.5\end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	5.46 8.26 8.57 8.69 7.86 7.57 7.60 7.36	7.67	7.2 10.6 10.9 11.3 10.1 9.6 9.7 9.3	9.84

Division into the sub-ranges generally improved the accuracy of prediction, especially in the case of RVS models. The exceptions were the eight-sub-range PVS models for both monitoring stations. In the case of Zabrze, the same average value of MAE was recorded in the eight-sub-range models as in the four-sub-range models ($8.02 \ \mu g/m^3$). In the case of Zloty Potok, even an increase in the overall values of MAE and RMSE errors was observed in the eight-sub-range models compared to the overall error of the four-sub-range models. In the case of RVS models, a decrease in the overall values of MAE and RMSE was always observed with an increase in the number of sub-ranges. The comparison showed that dividing the area into sub-ranges and separately modeling these sub-ranges improved the overall quality of the prediction, but that having too many sub-ranges and sub-models may be ineffective.

Number of Sub-Ranges	Regression Model	Sub-Ranges of O ₃ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZAB	0.7–198.0	36,460	8.4	8.4	11.3	11.3
2	RVS-1/2-ZAB RVS-2/2-ZAB	0.7–37.0 37–198.0	18,230 18,230	4.8 8.8	6.8	6.3 11.4	8.9
4	RVS-1/4-ZAB RVS-2/4-ZAB RVS-3/4-ZAB RVS-4/4-ZAB	0.7–15.0 15.0–37.0 37.0–62.0 62.0–198.0	9115 9115 9115 9115 9115	2.2 4.6 5.1 8.6	5.1	2.8 5.6 6.1 11.1	6.4
8	RVS-1/8-ZAB RVS-2/8-ZAB RVS-3/8-ZAB RVS-4/8-ZAB RVS-5/8-ZAB RVS-6/8-ZAB RVS-7/8-ZAB RVS-8/8-ZAB	$\begin{array}{c} 0.7-6.0\\ 6.0-15.0\\ 15.0-26.0\\ 26.0-37.0\\ 37.0-49.0\\ 49.0-62.0\\ 62.0-82.0\\ 82.0-198.0\end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	1.0 1.8 2.6 2.8 2.8 3.1 4.2 8.3	3.3	1.2 2.2 3.0 3.2 3.3 3.6 5.1 10.8	4.0

Table 5. Values of approximation errors of O₃ concentrations in RVS models, Zabrze 2011–2016.

Table 6. Values of approximation errors of O₃ concentrations in RVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-ranges of O ₃ Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZP	1.1–162	15,536	8.38	8.38	10.7	10.7
2	RVS-1/2-ZP RVS-2/2-ZP	1.1–51.0 51.0–162.0	7768 7768	5.97 7.05	6.51	7.6 8.9	8.25
4	RVS-1/4-ZP RVS-2/4-ZP RVS-3/4-ZP RVS-4/4-ZP	1.1–32.3 32.3–51.0 51.0–72.0 72.0–162.0	3884 3884 3884 3884 3884	4.06 3.85 4.20 6.56	4.67	5.1 4.7 5.1 8.4	5.84
8	RVS-1/8-ZP RVS-2/8-ZP RVS-3/8-ZP RVS-4/8-ZP RVS-5/8-ZP RVS-6/8-ZP RVS-7/8-ZP RVS-8/8-ZP	$\begin{array}{c} 1.1-21.1\\ 21.1-32.3\\ 32.3-42.0\\ 42.0-51.0\\ 51.0-60.4\\ 60.4-72.0\\ 72.0-90.3\\ 90.4-162.0\end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	2.98 2.48 2.27 2.17 2.28 2.56 3.68 6.06	3.06	3.6 3.0 2.7 2.6 2.6 3.1 4.5 7.9	3.74

The values of "overall MAE" and "overall RMSE" depending on the number of sub-ranges are presented in Figures 5–8.



Figure 5. Overall MAE and RMSE values for O₃ concentration prediction in PVS models depending on the number of created sub-models, Zabrze.



Figure 6. Overall MAE and RMSE values for O₃ concentration prediction in PVS models depending on the number of created sub-models, Złoty Potok.



Figure 7. Overall MAE and RMSE values for O₃ concentration prediction in RVS models depending on the number of created sub-models, Zabrze.



Figure 8. Overall MAE and RMSE values for O₃ concentration prediction in RVS models depending on the number of created sub-models, Złoty Potok.

3.2. The Results of the Modeling of NO Concentrations

Tables 7–10 show the results of NO concentration predictions obtained with the fullrange and sub-range models. The errors of the PVS models for the Zabrze monitoring station are presented in Table 7 and the Złoty Potok monitoring station in Table 8. The corresponding errors for the RVS models are presented in Tables 9 and 10. The division into sub-ranges generally improved the accuracy of prediction, especially in the case of RVS models. In general, as the number of sub-models increased, the overall measures of modeling error decreased. The exceptions were the eight-sub-range PVS models for the monitoring station at Złoty Potok. At this station, the overall MAE value for the eightsub-range sub-models (0.429 μ g/m³). In the case of RVS models, a decrease in the overall values of MAE and RMSE errors was always observed with an increase in the number of sub-ranges. In the Zloty Potok station, for the real concentrations in sub-ranges with a range of $0.0-0.0 \ \mu g/m^3$ and $1.0-1.0 \ \mu g/m^3$, RVS sub-models could not be created due to the lack of variability in these concentration ranges. Therefore, the overall error for this group of sub-models was not estimated. Figures 9–12 show the overall values of MAE and RMSE graphically.

Table 7. Values of approximation errors of NO concentrations in PVS models, Zabrze 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZAB	0.0–709.0	36,460	3.74	3.74	8.26	8.26
2	PVS-1/2-ZAB PVS-2/2-ZAB	0.0–3.0 3.0–709.0	18,230 18,230	0.52 6.22	3.37	0.66 11.70	6.18
4	PVS-1/4-ZAB PVS-2/4-ZAB PVS-3/4-ZAB PVS-4/4-ZAB	0.0–1.1 1.1–3.0 3.0–8.0 8.0–709.0	9115 9115 9115 9115 9115	0.23 0.34 0.94 9.55	2.76	0.32 0.42 1.17 15.37	4.32
8	PVS-1/8-ZAB PVS-2/8-ZAB PVS-3/8-ZAB PVS-4/8-ZAB PVS-5/8-ZAB PVS-6/8-ZAB PVS-7/8-ZAB PVS-8/8-ZAB	$\begin{array}{c} 0.0 - 1.0 \\ 1.0 - 1.1 \\ 1.1 - 2.0 \\ 2.0 - 3.0 \\ 3.0 - 4.7 \\ 4.7 - 8.0 \\ 8.0 - 19.3 \\ 19.4 - 709.0 \end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	0.23 0.01 0.15 0.19 0.39 0.77 2.16 13.86	2.22	0.32 0.01 0.21 0.26 0.47 0.92 2.64 20.25	3.14

Table 8. Values of approximation errors of NO concentrations in PVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZP	0.1–49.8	15,536	0.459	0.459	0.789	0.789
2	PVS-1/2-ZP PVS-2/2-ZP	0.1–0.6 0.6–49.8	7768 7768	0.324 0.535	0.430	0.425 0.979	0.702
4	PVS-1/4-ZP PVS-2/4-ZP PVS-3/4-ZP PVS-4/4-ZP	0.1-0.4 0.4-0.6 0.6-1.1 1.1-49.8	3884 3884 3884 3884	0.301 0.319 0.331 0.762	0.428	0.373 0.434 0.501 1.344	0.663
8	PVS-1/8-ZP PVS-2/8-ZP PVS-3/8-ZP PVS-4/8-ZP PVS-5/8-ZP PVS-6/8-ZP PVS-7/8-ZP PVS-8/8-ZP	$\begin{array}{c} 0.1 - 0.3 \\ 0.3 - 0.4 \\ 0.4 - 0.5 \\ 0.5 - 0.6 \\ 0.6 - 0.8 \\ 0.8 - 1.1 \\ 1.1 - 1.8 \\ 1.8 - 49.8 \end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	$\begin{array}{c} 0.261 \\ 0.302 \\ 0.311 \\ 0.357 \\ 0.350 \\ 0.316 \\ 0.386 \\ 1.150 \end{array}$	0.429	0.326 0.389 0.398 0.497 0.497 0.517 0.586 1.859	0.634

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZAB	0.0–709.0	36,460	3.74	3.74	8.26	8.26
2	RVS-1/2-ZAB RVS-2/2-ZAB	0.0–3.0 3.0–709.0	18,230 18,230	0.52 6.22	3.37	0.66 11.70	6.18
4	RVS-1/4-ZAB RVS-2/4-ZAB RVS-3/4-ZAB RVS-4/4-ZAB	0.0–1.1 1.1–3.0 3.0–8.0 8.0–709.0	9115 9115 9115 9115 9115	0.23 0.34 0.94 9.55	2.76	0.32 0.42 1.17 15.37	4.32
8	RVS-1/8-ZAB RVS-2/8-ZAB RVS-3/8-ZAB RVS-4/8-ZAB RVS-5/8-ZAB RVS-6/8-ZAB RVS-7/8-ZAB RVS-8/8-ZAB	$\begin{array}{c} 0.0{-}1.0\\ 1.0{-}1.1\\ 1.1{-}2.0\\ 2.0{-}3.0\\ 3.0{-}4.7\\ 4.7{-}8.0\\ 8.0{-}19.3\\ 19.4{-}709.0\end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	$\begin{array}{c} 0.23 \\ 0.01 \\ 0.15 \\ 0.19 \\ 0.39 \\ 0.77 \\ 2.16 \\ 13.86 \end{array}$	2.22	0.32 0.01 0.21 0.26 0.47 0.92 2.64 20.25	3.14

Table 9. Values of approximation errors of NO concentrations in RVS models, Zabrze 2011–2016.

Table 10. Values of approximation errors of NO concentrations in RVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZP	0.0–50.0	15,536	0.430	0.430	0.746	0.746
2	RVS-1/2-ZP RVS-2/2-ZP	0.0–0.8 0.8–50.0	7768 7768	0.184 0.436	0.310	0.234 0.908	0.571
4	RVS-1/4-ZP RVS-2/4-ZP RVS-3/4-ZP RVS-4/4-ZP	0.0–0.3 0.3–0.8 0.8–1.0 1.0–50.0	3884 3884 3884 3884	0.054 0.064 0.037 0.760	0.229	0.088 0.081 0.053 1.411	0.408
8	RVS-1/8-ZP RVS-2/8-ZP RVS-3/8-ZP RVS-4/8-ZP RVS-5/8-ZP RVS-6/8-ZP RVS-7/8-ZP RVS-8/8-ZP	$\begin{array}{c} 0.0-0.0\\ 0.0-0.3\\ 0.3-0.6\\ 0.6-0.8\\ 0.8-1.0\\ 1.0-1.0\\ 1.0-2.0\\ 2.0-50.0\end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	- 0.069 0.048 0.041 0.045 - 0.148 0.991	-	- 0.097 0.060 0.049 0.056 - 0.206 1.752	-



Figure 9. Overall MAE and RMSE values for NO concentration prediction in PVS models depending on the number of created sub-models, Zabrze.



Figure 10. Overall MAE and RMSE values for NO concentration prediction in PVS models depending on the number of created sub-models, Złoty Potok.



Figure 11. Overall MAE and RMSE values for NO concentration prediction in RVS models depending on the number of created sub-models, Zabrze.



Figure 12. Overall MAE and RMSE values for NO concentration prediction in RVS models depending on the number of created sub-models, Złoty Potok.

3.3. The Results of the Modeling of NO₂ Concentrations

Tables 11–14 show the results of NO₂ concentration predictions obtained with the full-range and sub-range models. The errors of the PVS models for the Zabrze monitoring station are presented in Table 11 and the Złoty Potok monitoring station in Table 12. The corresponding errors for the RVS models are presented in Tables 13 and 14. The division into sub-ranges generally improved the accuracy of prediction, especially in the case of RVS models. In general, as the number of sub-models increased, the overall measures of modeling error decreased. The exceptions were the eight-sub-range PVS models for both monitoring stations. The overall MAE and RMSE values for the four-sub-range sub-models were higher than the overall MAE and RMSE values for the four-sub-range sub-models. In the case of RVS models, a decrease in the overall values of MAE and RMSE

errors was always observed with an increase in the number of sub-ranges. In the case of PVS models, it appeared that having too many sub-ranges and sub-models could degrade the quality of the modeling. Figures 13–16 show the overall values of MAE and RMSE graphically.

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO ₂ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZAB	3.6–134.4	36,460	5.3	5.3	7.4	7.4
2	PVS-1/2-ZAB PVS-2/2-ZAB	3.6–20.6 20.6–134.4	18,230 18,230	3.13 7.29	5.21	4.2 9.6	6.89
4	PVS-1/4-ZAB PVS-2/4-ZAB PVS-3/4-ZAB PVS-4/4-ZAB	3.6–12.2 12.3–20.6 20.6–32.8 32.8–134.4	9115 9115 9115 9115 9115	2.29 3.85 5.89 8.54	5.14	3.0 5.0 7.6 11.1	6.66
8	PVS-1/8-ZAB PVS-2/8-ZAB PVS-3/8-ZAB PVS-4/8-ZAB PVS-5/8-ZAB PVS-6/8-ZAB PVS-7/8-ZAB PVS-8/8-ZAB	3.6–9.2 9.2–12.2 12.3–16 16–20.6 20.6–26.2 26.2–32.8 32.8–41.7 41.7–134.4	4558 4558 4558 4558 4558 4558 4558 4558	1.80 2.88 3.43 4.33 5.71 6.33 7.65 9.47	5.20	2.4 3.7 4.4 5.6 7.3 8.2 9.9 12.3	6.72

 Table 11. Values of approximation errors of NO2 concentrations in PVS models, Zabrze 2011–2016.

Table 12. Values of approximation errors of NO₂ concentrations in PVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO ₂ Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZP	1.6-52.9	15,536	1.919	1.919	2.65	2.65
2	PVS-1/2-ZP PVS-2/2-ZP	1.6–6.7 6.7–52.9	7768 7768	1.268 2.435	1.852	1.69 3.24	2.463
4	PVS-1/4-ZP PVS-2/4-ZP PVS-3/4-ZP PVS-4/4-ZP	1.6-4.7 4.7-6.7 6.7-10.8 10.8-52.9	3884 3884 3884 3884	0.989 0.989 1.852 2.915	1.686	1.33 1.33 2.39 3.78	2.207
8	PVS-1/8-ZP PVS-2/8-ZP PVS-3/8-ZP PVS-4/8-ZP PVS-5/8-ZP PVS-6/8-ZP PVS-7/8-ZP PVS-8/8-ZP	$\begin{array}{c} 1.63.7\\ 3.74.7\\ 4.75.6\\ 5.66.7\\ 6.78.2\\ 8.210.8\\ 10.816.0\\ 16.152.9\end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	0.812 1.273 1.366 1.671 1.786 1.972 2.361 3.246	1.811	1.08 1.66 1.73 2.17 2.26 2.55 3.06 4.22	2.342

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO ₂ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZAB	1.2–145	36,460	5.31	5.31	7.49	7.49
2	RVS-1/2-ZAB RVS-2/2-ZAB	1.2–20.0 20.0–145	18,230 18,230	2.37 6.47	4.42	2.97 8.84	5.91
4	RVS-1/4-ZAB RVS-2/4-ZAB RVS-3/4-ZAB RVS-4/4-ZAB	1.2–11.5 11.5–20.0 20.0–33.0 33.0–145.0	9115 9115 9115 9115 9115	1.56 1.82 2.73 6.61	3.18	1.91 2.19 3.27 9.03	4.10
8	RVS-1/8-ZAB RVS-2/8-ZAB RVS-3/8-ZAB RVS-4/8-ZAB RVS-5/8-ZAB RVS-6/8-ZAB RVS-7/8-ZAB RVS-8/8-ZAB	$\begin{array}{c} 1.28.0\\ 8.011.5\\ 11.515.2\\ 15.220.0\\ 20.026.0\\ 26.033.0\\ 33.043.1\\ 43.1145.0\end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	$1.03 \\ 0.86 \\ 0.93 \\ 1.12 \\ 1.40 \\ 1.68 \\ 2.44 \\ 6.84$	2.04	1.26 1.00 1.07 1.32 1.64 1.97 2.87 9.44	2.57

 $\label{eq:Table 13. Values of approximation errors of NO_2 \ concentrations in RVS \ models, Zabrze \ 2011-2016.$

Table 14. Values of approximation errors of NO₂ concentrations in RVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of NO ₂ Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZP	0.4.0-60.3	15,536	1.919	1.919	2.672	2.672
2	RVS-1/2-ZP RVS-2/2-ZP	0.4–7.0 7.0–60.3	7768 7768	0.894 2.164	1.529	1.110 2.985	2.047
4	RVS-1/4-ZP RVS-2/4-ZP RVS-3/4-ZP RVS-4/4-ZP	0.4-4.0 4.0-7.0 7.0-11.1 11.1-60.3	3884 3884 3884 3884	0.593 0.586 0.891 2.529	1.150	0.726 0.706 1.088 3.426	1.487
8	RVS-1/8-ZP RVS-2/8-ZP RVS-3/8-ZP RVS-4/8-ZP RVS-5/8-ZP RVS-6/8-ZP RVS-7/8-ZP RVS-8/8-ZP	$\begin{array}{c} 0.4 & -3.0 \\ 3.0 & -4.0 \\ 4.0 & -5.4 \\ 5.4 & -7.0 \\ 7.0 & -8.7 \\ 8.7 & -11.1 \\ 11.1 & -16.6 \\ 16.6 & -60.3 \end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	$\begin{array}{c} 0.427 \\ 0.304 \\ 0.283 \\ 0.332 \\ 0.438 \\ 0.631 \\ 1.107 \\ 2.661 \end{array}$	0.773	0.521 0.354 0.349 0.408 0.501 0.736 1.332 3.544	0.968



Figure 13. Overall MAE and RMSE values for NO₂ concentration prediction in PVS models depending on the number of created sub-models, Zabrze.



Figure 14. Overall MAE and RMSE values for NO₂ concentration prediction in PVS models depending on the number of created sub-models, Złoty Potok.



Figure 15. Overall MAE and RMSE values for NO₂ concentration prediction in RVS models depending on the number of created sub-models, Zabrze.



Figure 16. Overall MAE and RMSE values for NO₂ concentration prediction in RVS models depending on the number of created sub-models, *Z*_i toty Potok.

3.4. The Results of the Modeling of SO₂ Concentrations

Tables 15–18 show the results of SO₂ concentration predictions obtained with the full-range and sub-range models. The errors of the PVS models for the Zabrze monitoring station are presented in Table 15 and the Złoty Potok monitoring station in Table 16. The corresponding errors for the RVS models are presented in Tables 17 and 18. The division into sub-ranges generally improved the accuracy of prediction, especially in the case of RVS models. In general, as the number of sub-models increased, the overall measures of modeling error decreased. The exceptions were the four-sub-range and eight-sub-range PVS sub-models for the monitoring station at Zabrze.

Number of Sub-Ranges	Regression Model	Sub-Ranges of SO ₂ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZAB	1.3–321.9	36,460	5.25	5.25	8.1	8.1
2	PVS-1/2-ZAB PVS-2/2-ZAB	1.3–10.6 10.6–321.9	18,230 18,230	2.57 7.78	5.17	3.7 10.9	7.30
4	PVS-1/4-ZAB PVS-2/4-ZAB PVS-3/4-ZAB PVS-4/4-ZAB	1.3–6.4 6.4–10.6 10.6–21.3 21.3–321.9	9115 9115 9115 9115 9115	1.81 3.30 5.88 9.69	5.17	2.6 4.6 8.1 13.3	7.15
8	PVS-1/8-ZAB PVS-2/8-ZAB PVS-3/8-ZAB PVS-4/8-ZAB PVS-5/8-ZAB PVS-6/8-ZAB PVS-7/8-ZAB PVS-8/8-ZAB	$\begin{array}{c} 1.3-4.9\\ 4.9-6.4\\ 6.4-8.2\\ 8.2-10.6\\ 10.6-14.4\\ 14.4-21.3\\ 21.3-34.0\\ 34.0-321.9\end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	1.49 2.19 2.85 3.85 5.46 6.69 7.08 11.82	5.18	2.2 3.0 3.9 5.3 7.5 9.1 9.3 15.9	7.02

 Table 15. Values of approximation errors of SO2 concentrations in PVS models, Zabrze 2011–2016.

Table 16. Values of approximation errors of SO₂ concentrations in PVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of SO ₂ Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZP	0.8–74.2	15,536	1.920	1.920	3.133	3.133
2	PVS-1/2-ZP PVS-2/2-ZP	0.8–4.0 4.0–74.2	7768 7768	0.935 2.899	1.917	1.302 4.298	2.800
4	PVS-1/4-ZP PVS-2/4-ZP PVS-3/4-ZP PVS-4/4-ZP	0.8–2.8 2.8–4.0 4.0–7.2 7.2–74.2	3884 3884 3884 3884	0.823 1.135 1.710 3.836	1.876	1.099 1.597 2.395 5.389	2.620
8	PVS-1/8-ZP PVS-2/8-ZP PVS-3/8-ZP PVS-4/8-ZP PVS-5/8-ZP PVS-6/8-ZP PVS-7/8-ZP PVS-8/8-ZP	0.8-2.2 2.2-2.8 2.8-3.3 3.3-4.0 4.0-5.1 5.1-7.2 7.2-11.8 11.8-74.2	1942 1942 1942 1942 1942 1942 1942 1942	$\begin{array}{c} 0.699\\ 0.855\\ 1.008\\ 1.246\\ 0.057\\ 2.024\\ 3.108\\ 5.072 \end{array}$	1.759	$\begin{array}{c} 0.899\\ 1.131\\ 1.379\\ 1.778\\ 0.080\\ 2.813\\ 4.545\\ 6.681\end{array}$	2.413

Number of Sub-Ranges	Regression Model	Sub-Ranges of SO ₂ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZAB	0.1–362	36,460	5.26	5.26	8.15	8.15
2	RVS-1/2-ZAB RVS-2/2-ZAB	0.1–10.0 10.0–362.0	18,230 18,230	1.58 7.08	4.33	1.95 10.33	6.14
4	RVS-1/4-ZAB RVS-2/4-ZAB RVS-3/4-ZAB RVS-4/4-ZAB	0.1–5.0 5.0–10 10.0–22.3 22.4–362.0	9115 9115 9115 9115 9115	0.85 1.11 2.66 8.11	3.18	1.03 1.32 3.21 11.64	4.30
8	RVS-1/8-ZAB RVS-2/8-ZAB RVS-3/8-ZAB RVS-4/8-ZAB RVS-5/8-ZAB RVS-6/8-ZAB RVS-7/8-ZAB RVS-8/8-ZAB	$\begin{array}{c} 0.1-3.2\\ 3.2-5.0\\ 5.0-7.0\\ 7.0-10\\ 10.0-14.9\\ 14.9-22.3\\ 22.4-36.0\\ 36.0-362.0\\ \end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	$\begin{array}{c} 0.55 \\ 0.41 \\ 0.48 \\ 0.73 \\ 1.16 \\ 1.82 \\ 3.02 \\ 9.84 \end{array}$	2.25	$\begin{array}{c} 0.64 \\ 0.48 \\ 0.56 \\ 0.85 \\ 1.36 \\ 2.14 \\ 3.58 \\ 14.21 \end{array}$	2.98

Table 17. Values of approximation errors of SO₂ concentrations in RVS models, Zabrze 2011–2016.

Table 18. Values of approximation errors of SO₂ concentrations in RVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of SO ₂ Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, µg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZP	0.0-85.0	15,536	1.96	1.96	3.186	3.186
2	RVS-1/2-ZP RVS-2/2-ZP	0.0 - 4.0 4.0 - 85.0	7768 7768	0.63 2.68	1.66	0.780 4.116	2.45
4	RVS-1/4-ZP RVS-2/4-ZP RVS-3/4-ZP RVS-4/4-ZP	0.0–2.1 2.1–4.0 4.0–7.4 7.4–85.0	3884 3884 3884 3884	0.38 0.36 0.73 3.45	1.23	0.455 0.461 0.891 5.048	1.71
8	RVS-1/8-ZP RVS-2/8-ZP RVS-3/8-ZP RVS-4/8-ZP RVS-5/8-ZP RVS-6/8-ZP RVS-7/8-ZP RVS-8/8-ZP	$\begin{array}{c} 0.0{-}1.8\\ 1.8{-}2.1\\ 2.1{-}3.0\\ 3.0{-}4.0\\ 4.0{-}5.0\\ 5.0{-}7.4\\ 7.4{-}12.0\\ 12.0{-}85.0\end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	$\begin{array}{c} 0.26 \\ 0.04 \\ 0.19 \\ 0.23 \\ 0.34 \\ 0.58 \\ 1.01 \\ 4.05 \end{array}$	0.84	$\begin{array}{c} 0.333\\ 0.063\\ 0.245\\ 0.289\\ 0.385\\ 0.683\\ 1.216\\ 5.909 \end{array}$	1.14

At this station, the overall MAE values (5.18 μ g/m³ for the eight-sub-range submodels, and 5.17 μ g/m³ for the four-sub-range sub-models) were comparable to the overall MAE value for the two-sub-range sub-models (5.17 μ g/m³). In the case of RVS models, a decrease in the overall values of MAE and RMSE was always observed with an increase in the number of sub-ranges. In the case of PVS models, it appeared that having too many sub-ranges and sub-models could degrade the quality of the modeling. Figures 17–20 show the overall values of MAE and RMSE graphically.



Figure 17. Overall MAE and RMSE values for SO₂ concentration prediction in PVS models depending on the number of created sub-models, Zabrze.



Figure 18. Overall MAE and RMSE values for SO₂ concentration prediction in PVS models depending on the number of created sub-models, Złoty Potok.



Figure 19. Overall MAE and RMSE values for SO₂ concentration prediction in RVS models depending on the number of created sub-models, Zabrze.



Figure 20. Overall MAE and RMSE values for SO₂ concentration prediction in RVS models depending on the number of created sub-models, Złoty Potok.

3.5. The Results of the Modeling of PM₁₀ Concentrations

Tables 19–22 show the results of PM_{10} concentration predictions obtained in the fullrange and sub-range models. The errors of the PVS models for the Zabrze monitoring station are presented in Table 19 and the Złoty Potok monitoring station in Table 20. The corresponding errors for the RVS models are presented in Tables 21 and 22. The division into sub-ranges generally improved the accuracy of prediction, especially in the case of RVS models. In general, as the number of sub-models increased, the overall measures of modeling error decreased. The exceptions were the eight-sub-range PVS models for the monitoring station at Złoty Potok. The overall MAE and RMSE values for the eight-subrange models were higher than the overall MAE and RMSE values for the four-sub-range models and even the two-sub-range models. In the case of RVS models, a decrease in the overall values of MAE and RMSE was always observed with an increase in the number of sub-ranges. In the case of PVS models, it appeared that having too many sub-ranges and submodels could degrade the quality of the modeling. Figures 21–24 show the overall values of MAE and RMSE graphically.

Number of Sub-Ranges	Regression Model	Sub-Ranges of PM ₁₀ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZAB	8.2–980.5	36,460	11.83	11.83	18.7	18.7
2	PVS-1/2-ZAB PVS-2/2-ZAB	8.2–33.1 33.1–980.5	18,230 18,230	6.53 16.99	11.76	8.79 25.22	17.00
4	PVS-1/4-ZAB PVS-2/4-ZAB PVS-3/4-ZAB PVS-4/4-ZAB	8.2–22.1 22.1–33.1 33.1–55.6 55.6–980.5	9115 9115 9115 9115 9115	5.52 7.42 10.01 23.07	11.51	7.6 9.7 13.2 32.6	15.76
8	PVS-1/8-ZAB PVS-2/8-ZAB PVS-3/8-ZAB PVS-4/8-ZAB PVS-5/8-ZAB PVS-6/8-ZAB PVS-7/8-ZAB PVS-8/8-ZAB	8.2–17.9 17.9–22.1 22.1–26.9 26.9–33.1 33.1–41.2 41.2–55.6 55.6–90.1 90.2–980.5	4558 4558 4558 4558 4558 4558 4558 4558	4.47 6.39 6.72 8.24 9.18 11.11 14.67 28.79	11.20	5.9 8.8 8.9 10.7 12.1 14.4 19.0 39.7	14.93

Table 19. Values of approximation errors of PM_{10} concentrations in PVS models, Zabrze 2011–2016.

Table 20. Values of approximation errors of PM_{10} concentrations in PVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of PM ₁₀ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZP	5.1-120.4	15,536	6.726	6.726	9.10	9.10
2	PVS-1/2-ZP PVS-2/2-ZP	5.1–22.7 22.7–120.4	7768 7768	4.638 8.541	6.589	6.11 11.18	8.646
4	PVS-1/4-ZP PVS-2/4-ZP PVS-3/4-ZP PVS-4/4-ZP	5.1–18.0 18.0–22.7 22.7–30.9 30.9–120.4	3884 3884 3884 3884	3.964 5.230 6.602 9.972	6.442	5.12 6.76 8.57 12.70	8.285
8	PVS-1/8-ZP PVS-2/8-ZP PVS-3/8-ZP PVS-4/8-ZP PVS-5/8-ZP PVS-6/8-ZP PVS-7/8-ZP PVS-8/8-ZP	5.1-15.7 $15.7-18.0$ $18.0-20.2$ $20.2-22.7$ $22.7-25.9$ $25.9-30.9$ $30.9-40.5$ $40.5-120.4$	1942 1942 1942 1942 1942 1942 1942 1942	3.406 4.634 4.963 5.613 6.751 7.202 8.027 12.516	6.639	4.39 6.01 6.43 7.20 8.74 9.28 10.23 15.88	8.519

Number of Sub-Ranges	Regression Model	Sub-Ranges of PM ₁₀ Concentrations, µg/m ³	Number of Observations	MAE, µg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZAB	0.1–1145.0	36,460	11.84	11.84	18.8	18.8
2	RVS-1/2-ZAB RVS-2/2-ZAB	0.1–33.0 33.0–1145.0	18,230 18,230	4.66 15.20	9.93	5.8 23.8	14.77
4	RVS-1/4-ZAB RVS-2/4-ZAB RVS-3/4-ZAB RVS-4/4-ZAB	0.1–20.0 20.0–33.0 33.0–57.1 57.1–1145.0	9115 9115 9115 9115 9115	3.09 2.97 5.02 20.23	7.83	3.8 3.5 6.1 30.4	10.96
8	RVS-1/8-ZAB RVS-2/8-ZAB RVS-3/8-ZAB RVS-4/8-ZAB RVS-5/8-ZAB RVS-6/8-ZAB RVS-7/8-ZAB RVS-8/8-ZAB	$\begin{array}{c} 0.1{-}14.0\\ 14.0{-}20.0\\ 20.0{-}26.0\\ 26.0{-}33.0\\ 33.0{-}42.5\\ 42.5{-}57.1\\ 57.1{-}91.1\\ 91.1{-}1145.0\\ \end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	2.26 1.49 1.51 1.74 2.30 3.51 6.99 26.32	5.76	2.8 1.7 1.8 2.0 2.7 4.1 8.4 38.0	7.68

 $\label{eq:Table 21. Values of approximation errors of PM_{10} \mbox{ concentrations in RVS models, Zabrze 2011–2016.}$

Table 22. Values of approximation errors of PM_{10} concentrations in RVS models, Złoty Potok 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of PM ₁₀ Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZP	3.0–138.0	15,536	6.88	6.88	9.34	9.34
2	RVS-1/2-ZP RVS-2/2-ZP	3.0–22.8 22.8–138.0	7768 7768	2.99 7.11	5.05	3.64 9.75	6.70
4	RVS-1/4-ZP RVS-2/4-ZP RVS-3/4-ZP RVS-4/4-ZP	3.0–16.0 16.0–22.8 22.8–32.9 32.9–138.0	3884 3884 3884 3884	1.88 1.62 2.29 8.07	3.46	2.31 1.90 2.71 11.00	4.48
8	RVS-1/8-ZP RVS-2/8-ZP RVS-3/8-ZP RVS-4/8-ZP RVS-5/8-ZP RVS-6/8-ZP RVS-7/8-ZP RVS-8/8-ZP	$\begin{array}{c} 3.0{-}12.1\\ 12.1{-}16.0\\ 16.0{-}19.0\\ 19.0{-}22.8\\ 22.8{-}27.0\\ 27.0{-}32.9\\ 32.9{-}43.0\\ 43.0{-}138.0 \end{array}$	1942 1942 1942 1942 1942 1942 1942 1942	$1.43 \\ 0.84 \\ 0.89 \\ 1.04 \\ 1.46 \\ 2.50 \\ 7.88$	2.11	$ 1.76 \\ 0.99 \\ 0.97 \\ 1.03 \\ 1.21 \\ 1.69 \\ 2.94 \\ 10.71 $	2.66



Figure 21. Overall MAE and RMSE values for PM_{10} concentration prediction in PVS models depending on the number of created sub-models, Zabrze.



Figure 22. Overall MAE and RMSE values for PM_{10} concentration prediction in PVS models depending on the number of created sub-models, Złoty Potok.



Figure 23. Overall MAE and RMSE values for PM_{10} concentration prediction in RVS models depending on the number of created sub-models, Zabrze.



Figure 24. Overall MAE and RMSE values for PM_{10} concentration prediction in RVS models depending on the number of created sub-models, Złoty Potok.

3.6. The Results of the Modeling of CO Concentrations

CO concentrations were not monitored at the Złoty Potok station, so the analysis was carried out only using monitoring data from Zabrze.

Tables 23 and 24 show the results of PM_{10} concentration predictions obtained with the full-range and sub-range models. The errors in the PVS models for the Zabrze monitoring station are presented in Table 23. The errors in the RVS models are presented in Table 24. The division into sub-ranges improved the accuracy of prediction in the case of RVS models. In general, as the number of sub-models increased, the overall measures of modeling error decreased. The PVS models showed slight changes in accuracy. The MAE level did not change much. The overall MAE value for the eight-sub-range models was higher than the overall MAE value for the four-sub-range models and equal to the MAE value for the

two-sub-range models. In the case of RVS models, a decrease in the overall values of MAE and RMSE was always observed with an increase in the number of sub-ranges. In the case of PVS models, it appeared that having too many sub-ranges and sub-models could degrade the quality of the modeling. Figures 25 and 26 show the overall values of MAE and RMSE graphically.

Number of Sub-Ranges	Regression Model	Sub-Ranges of CO Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	PVS-1/1-ZAB	0.16-8.28	36,460	0.119	0.119	0.202	0.202
2	PVS-1/2-ZAB PVS-2/2-ZAB	0.16-0.42 0.42-8.28	18,230 18,230	0.071 0.162	0.117	0.155 0.230	0.193
4	PVS-1/4-ZAB PVS-2/4-ZAB PVS-3/4-ZAB PVS-4/4-ZAB	0.16-0.30 0.30-0.42 0.42-0.69 0.69-8.28	9115 9115 9115 9115 9115	0.066 0.074 0.111 0.210	0.115	0.196 0.098 0.149 0.285	0.182
8	PVS-1/8-ZAB PVS-2/8-ZAB PVS-3/8-ZAB PVS-4/8-ZAB PVS-5/8-ZAB PVS-6/8-ZAB PVS-7/8-ZAB PVS-8/8-ZAB	$\begin{array}{c} 0.16-0.26\\ 0.26-0.30\\ 0.30-0.35\\ 0.35-0.42\\ 0.42-0.52\\ 0.52-0.69\\ 0.69-1.03\\ 1.03-8.28\\ \end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	0.072 0.069 0.065 0.079 0.099 0.121 0.153 0.275	0.117	$\begin{array}{c} 0.198 \\ 0.177 \\ 0.085 \\ 0.106 \\ 0.132 \\ 0.158 \\ 0.203 \\ 0.365 \end{array}$	0.178

Table 23. Values of approximation errors of CO concentrations in PVS models, Zabrze 2011–2016.

Table 24. Values of approximation errors of CO concentrations in RVS models, Zabrze 2011–2016.

Number of Sub-Ranges	Regression Model	Sub-Ranges of CO Concentrations, µg/m ³	Number of Observations	MAE, μg/m ³	Overall MAE, μg/m ³	RMSE, μg/m ³	Overall RMSE, μg/m ³
1	RVS-1/1-ZAB	0.1–9.0	36,460	0.118	0.118	0.198	0.198
2	RVS-1/2-ZAB RVS-2/2-ZAB	0.1–0.4 0.4–9.0	18,230 18,230	0.047 0.160	0.104	0.059 0.262	0.161
4	RVS-1/4-ZAB RVS-2/4-ZAB RVS-3/4-ZAB RVS-4/4-ZAB	0.1-0.3 0.3-0.4 0.4-0.7 0.7-9.0	9115 9115 9115 9115 9115	0.034 0.030 0.055 0.196	0.078	0.042 0.035 0.066 0.274	0.104
8	RVS-1/8-ZAB RVS-2/8-ZAB RVS-3/8-ZAB RVS-4/8-ZAB RVS-5/8-ZAB RVS-6/8-ZAB RVS-7/8-ZAB RVS-8/8-ZAB	$\begin{array}{c} 0.1 - 0.2 \\ 0.2 - 0.3 \\ 0.3 - 0.3 \\ 0.3 - 0.4 \\ 0.4 - 0.5 \\ 0.5 - 0.7 \\ 0.7 - 1.1 \\ 1.1 - 9.0 \end{array}$	4558 4558 4558 4558 4558 4558 4558 4558	0.027 0.014 0.014 0.018 0.025 0.038 0.077 0.223	0.055	0.033 0.016 0.017 0.021 0.030 0.045 0.093 0.309	0.070



Figure 25. Overall MAE and RMSE values for CO concentration prediction in PVS models depending on the number of created sub-models, Zabrze.



Figure 26. Overall MAE and RMSE values for CO concentration prediction in RVS models depending on the number of created sub-models, Zabrze.

4. Summary and Discussion

Table 25 shows the percentage changes in the overall values of MAE and RMSE obtained by modeling the concentrations in sub-ranges, calculated in relation to the error values of the corresponding full-range models.

	Number of Sub-Models	Zabrze				Zloty Potok			
Air Pollutant		PVS Sub-Models		RVS Sub-Models		PVS Sub-Models		RVS Sub-Models	
		ΔΜΑΕ	ΔRMSE	ΔΜΑΕ	ΔRMSE	ΔΜΑΕ	ΔRMSE	ΔΜΑΕ	ΔRMSE
		%	%	%	%	%	%	%	%
O ₃	2	-2.1	-2.9	-19.2	-21.3	-5.9	-5.6	-22.3	-23.0
	4	-4.9	-7.4	-39.5	-43.4	-10.2	-9.5	-44.3	-45.5
	8	-4.9	-8.2	-60.7	-64.1	-8.7	-8.0	-63.5	-65.1
NO	2	-4.6	-22.9	-9.9	-25.2	-6.4	-11.0	-27.9	-23.5
	4	-7.7	-36.9	-26.1	-47.7	-6.7	-16.0	-46.8	-45.3
	8	-8.3	-41.9	-40.6	-62.0	-6.5	-19.7	-	-
NO ₂	2	-1.7	-7.5	-16.8	-21.2	-3.5	-7.0	-20.3	-23.4
	4	-3.0	-10.5	-40.2	-45.3	-12.1	-16.6	-40.1	-44.4
	8	-1.9	-9.8	-61.6	-65.7	-5.6	-11.5	-59.7	-63.8
SO ₂	2	-1.5	-10.2	-17.7	-24.7	-0.1	-10.6	-15.4	-23.2
	4	-1.6	-12.0	-39.5	-47.3	-2.3	-16.4	-37.2	-46.2
	8	-1.4	-13.6	-57.2	-63.5	-8.4	-23.0	-57.2	-64.2
PM ₁₀	2	-0.5	-8.9	-16.1	-21.3	-2.0	-5.0	-26.6	-28.2
	4	-2.7	-15.5	-33.9	-41.7	-4.2	-9.0	-49.6	-52.0
	8	-5.3	-20.0	-51.3	-59.1	-1.3	-6.4	-69.3	-71.5
СО	2	-2.4	-4.8	-11.8	-18.9	-	_	-	-
	4	-3.5	-10.1	-33.4	-47.3	-	-	-	-
	8	-2.2	-12.1	-53.6	-64.5	-	-	-	-

Table 25. The percentage changes in the overall values of MAE and RMSE obtained by modeling the concentrations in sub-ranges, in relation to the MAE and RMSE values of the corresponding full-range models.

In the case of RVS models, each division into narrower concentration sub-ranges and the development of appropriate sub-range models resulted in a significant reduction in the overall value of the modeling error. The division into sub-models always improved the accuracy of predictions in the case of RVS models. When the number of sub-models increased, the overall measures of modeling error decreased. A significant improvement in the quality of modeling was achieved at both air monitoring stations. Modeling errors could be reduced by more than 60% using eight sub-models. However, it should be emphasized that the RVS models are not of great practical importance because their use is related to knowledge of the real concentrations of the pollutants. Moreover, once the concentration values are known, there is no need to perform modeling. The importance of the RVS models was that they allowed us to assess the potential for improving the quality of the modeling.

The division into sub-models generally improved the accuracy of the PVS models; however, the decrease in modeling error was not as great as in the RVS models. Moreover, quite often, after splitting the full-range set into eight sub-ranges and running eight sub-models, the MAE and RMSE values could be higher than in the sub-models created after division into only four sub-ranges. Such an effect was found for O_3 , NO, NO₂ and PM₁₀ in Złoty Potok, and NO₂, SO₂ and CO in Zabrze. The probable cause of the deterioration in the quality of prediction in some eight sub-range PVS models was the error in the classification of cases into individual sub-ranges. The classification was made on the basis of the predicted concentration values obtained as a result of the preliminary prediction. With an increasing number of sub-ranges, the sub-ranges of concentrations became narrower and the number of misclassified cases also increased. The number of misclassified cases became so large that it increased the mean prediction error in the sub-range.

The PVS models always showed a lower accuracy than the RVS models. This is understandable, as the PVS models required a preliminary prediction, which introduced an additional error. In conclusion, having too many sub-ranges and sub-models can degrade the quality of modeling for the PVS models.

31 of 33

After division into sub-ranges (two sub-ranges, four sub-ranges, eight sub-ranges), the error in the models for the highest sub-ranges was always very large compared to the error in the models for the lower sub-ranges. There are two reasons for this effect. The first is that the width of the highest sub-range was always greater than that of the models of the lower sub-ranges. The second reason was the need to predict extremely high concentrations. The approximation of such unusual concentrations is always burdened with a higher error.

5. Conclusions

Monitoring data are often not complete enough to carry out an air quality assessment. To fill the measurement gaps, predictive models are used. Such models often use archival measurement data from air monitoring systems. This is the best source of knowledge about the relationships between measured variables (concentrations and meteorological parameters). There is a need to model the missing concentrations as accurately as possible. The use of artificial neural networks reduces the prediction error compared to classical regression methods. In previous studies, a single regression model over the entire concentration range was used to approximate the concentrations of a selected pollutant. In this study, it was assumed that not a single model, but a group of models, could be used for the prediction. In this approach, each model from the group is dedicated to a different sub-range of the concentration of a modeled pollutant. The aim of this analysis was to check whether this approach would improve the quality of modeling.

The aim of the analysis was not to create the most up-to-date models based on possible new data. Once trained, a model using historical data (e.g., from 2011 to 2016) should also be able to predict concentrations for current data. This feature of the model's "generalization of acquired knowledge" was tested during the learning process on the cases from the test subset, and also on the cases from the validation subset after the network training was completed. The performed validation should show that the model can approximate the target value using data independent of the learning process.

Air monitoring data from the period 2011–2016 allowed us to verify the possibility of improving the accuracy of modeling by carrying out modeling in subsets. A similar analysis can be carried out using data from other monitoring stations, or data from Zabrze and Złoty Potok stations from a different period. However, the selected set of cases should cover a measurement period of several years, so that the recorded cases correspond to different meteorological situations and different ranges of concentrations of monitored pollutants.

The most important conclusions that resulted from the conducted analysis are as follows:

- 1. Modeling segmentation, consisting in prediction in the sub-ranges of concentrations of the modeled pollutant, allowed for a higher overall modeling accuracy.
- 2. For RVS models, segmentation of the modeling process guarantees a significant increase in modeling accuracy compared to a model based on the full-range of concentrations.
- 3. In the case of PVS models, segmentation of the modeling process allows to reduce overall prediction errors. However, the number of concentration sub-ranges cannot be too large. When predicting with 8 submodels, the modeling accuracy may be lower than when predicting with 4 submodels.

The authors are aware that the article has its shortcomings and limitations. The analysis was carried out using historical data from only two air monitoring stations. Stations from the same region of Poland were selected. We are convinced that a similar analysis of completely different data, e.g., from other regions of Poland or other countries/continents, would confirm the formulated conclusions. It can be assumed with high probability that the concept is valid and should yield similar results for other air monitoring data sets. The effect was confirmed for six air pollutants. Not all possible air pollutants, e.g., PM_{2.5}, were included. The authors are convinced that a similar effect could be obtained for other monitored pollutants.

Author Contributions: Conceptualization, S.H.; methodology, S.H. and M.F.; software, M.F.; validation, M.F.; formal analysis, M.F.; investigation, S.H. and M.F.; resources, M.F.; data curation, M.F and R.J.; writing—original draft preparation, S.H. and M.F.; writing—review and editing, S.H; visualization, S.H. and M.F.; supervision, S.H. and 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-400-301/22.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

- 1. World Health Organization. 7 Million Premature Deaths Annually Linked to Air Pollution. Available online: www.who.int/ mediacentre/news/releases/2014/air-pollution/en (accessed on 29 April 2021).
- 2. European Environment Agency. *Air Quality in Europe-2020 Report. No. 12/2018*; Publications Office of the European Union: Luxembourg, 2020.
- Widziewicz, Y.; Rogula-Kozlowska, W.; Loska, K.; Kociszewska, K.; Majewski, G. Health Risk Impacts of Exposure to Airborne Metals and Benzo(a)Pyrene during Episodes of High PM10 Concentrations in Poland. *Biomed. Environ. Sci.* 2018, 31, 23–36. [CrossRef] [PubMed]
- 4. Trojanowska, M.; Świetlik, R. Heavy metals cadmium, nickel and arsenic environmental inhalation hazard of residents of Polish cities. *Environ. Med.* **2012**, *15*, 33–41.
- 5. Gurjar, B.R.; Molina, L.T.; Ojha, C.S.P. *Air Pollution: Health and Environmental Impacts;* CRC Press: Boca Raton, FL, USA, 2010. [CrossRef]
- 6. 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]
- 7. Maesano, I. The Air of Europe: Where Are We Going? Eur. Respir. Rev. 2017, 26, 170024. [CrossRef] [PubMed]
- 8. 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]
- 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]
- 10. Kim, Y.; Manley, J.; Radoias, V. Air Pollution and Long Term Mental Health. Atmosphere 2020, 11, 1355. [CrossRef]
- 11. 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]
- 12. Graff-Zivin, J.; Neidell, M. The Impact of Pollution on Worker Productivity. Am. Econ. Rev. 2012, 102, 3652–3673. [CrossRef]
- 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]
- 14. 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]
- 15. 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]
- 16. Wei, W.; Wang, Z. Impact of Industrial Air Pollution on Agricultural Production. Atmosphere 2021, 12, 639. [CrossRef]
- 17. Vallero, D.A. Fundamentals of Air Pollution; Academic Press: London, UK, 2014.
- EU. 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 L 152, 11.6.2008, pp. 1–44), Current Consolidated Version: 18 September 2015. Available online: https://eur-lex.europa.eu/eli/dir/2008/50/oj (accessed on 20 December 2021).
- 19. *EN 14625:2012;* Ambient Air—Standard Method for the Measurement of the Concentration of Ozone by Ultraviolet Photometry. European Committee for Standardization: Brussels, Belgium, 2012.
- 20. *EN 14211:2012;* Ambient Air—Standard Method for the Measurement of the Concentration of Nitrogen Dioxide and Nitrogen Monoxide by Chemiluminescence. European Committee for Standardization: Brussels, Belgium, 2012.
- 21. EN 14212:2012; Ambient Air—Standard Method for the Measurement of the Concentration of Sulphur Dioxide by Ultraviolet Fluorescence. European Committee for Standardization: Brussels, Belgium, 2012.

- 22. EN 14626:2012; Ambient Air—Standard Method for the Measurement of the Concentration of Carbon Monoxide by Non dispersive Infrared Spectroscopy. European Committee for Standardization: Brussels, Belgium, 2012.
- 23. *EN* 12341:2014; Ambient Air—Standard Gravimetric Measurement Method for the Determination of the PM10 or PM2.5 Mass Concentration of Suspended Particulate Matter. European Committee for Standardization: Brussels, Belgium, 2014.
- Plaia, A.; Bondi, A.L. Single Imputation Method of Missing Values in Environmental Pollution Data Sets. Atmos. Environ. 2006, 40, 7316–7330. [CrossRef]
- 25. Gentili, S.; Magnaterra, L.; Passerini, G. An introduction to the statistical filling of environmental data time series. In *Handling Missing Data: Applications to Environmental Analysis*; Latini, G., Passerini, G., Eds.; Wit Press: Southampton, UK, 2004.
- Hoffman, S. Environmental Engineering; Pawłowski, L., Dudzińska, M.R., Pawłowski, A., Eds.; Taylor & Francis Group: London, UK, 2007; pp. 349–353.
- 27. Hoffman, S. Approximation of Imission Level at Air Monitoring Stations by Means of Autonomous Neural Models. *Environ. Prot. Eng.* **2012**, *38*, 109–119. [CrossRef]
- 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]
- 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]
- 30. 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]
- 31. 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.
- Karppinen, A.; Kukkonen, J.; Elolähde, T.; Konttinen, M.; Koskentalo, T.; Rantakrans, E. A Modelling System for Predicting Urban Air Pollution: Comparison of Model Predictions with the Data of an Urban Measurement Network in Helsinki. *Atmos. Environ.* 2000, 34, 3735–3743. [CrossRef]
- Nagendra, S.S.M.; Khare, M. Modelling Urban Air Quality Using Artificial Neural Network. *Clean. Technol. Environ. Policy* 2005, 7, 116–126. [CrossRef]
- 34. Maleki, H.; Sorooshian, A.; Goudarzi, G.; Baboli, Z.; Birgani, Y.T.; Rahmati, M. Air pollution prediction by using an artificial neural network model. *Clean. Technol. Environ. Policy* **2019**, *21*, 1341–1352. [CrossRef] [PubMed]
- Fallahizadeh, S.; Kermani, M.; Esrafili, A.; Asadgol, Z.; Gholami, M. The effects of meteorological parameters on PM10: Health impacts assessment using AirQ+ model and prediction by an artificial neural network (ANN). *Urban Clim.* 2021, 38, 100905. [CrossRef]
- 36. 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]
- 37. 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]
- Hoffman, S. Assessment of Prediction Accuracy in Autonomous Air Quality Models. Desalination Water Treat. 2015, 57, 1322–1326. [CrossRef]
- 39. Hoffman, S. Estimation of Prediction Error in Regression Air Quality Models. Energies 2021, 14, 7387. [CrossRef]
- 40. Fletcher, R. Practical Methods of Optimization, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2000.
- 41. Broyden, C.G. The convergence of a class of double-rank minimization algorithms. *J. Inst. Math. Its Appl.* **1970**, *6*, 76–90. [CrossRef]
- 42. Fletcher, R. A New Approach to Variable Metric Algorithms. Comput. J. 1970, 13, 317–322. [CrossRef]
- 43. Goldfarb, D. A Family of Variable Metric Updates Derived by Variational Means. Math. Comput. 1970, 24, 23–26. [CrossRef]
- 44. Shanno, D.F. Conditioning of quasi-Newton methods for function minimization. Math. Comput. 1970, 24, 647–656. [CrossRef]