Estimation of Prediction Error in Regression Air Quality Models

Szymon Hoffman

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

Abstract: Combustion of energy fuels or organic waste is associated with the emission of harmful gases and aerosols into the atmosphere, which strongly affects air quality. Air quality monitoring devices are unreliable and measurement gaps appear quite often. Missing data modeling techniques can be used to complete the monitoring data. Concentrations of monitored pollutants can be approximated with regression modeling tools, such as artificial neural networks. In this study, a long-term set of data from the air monitoring station in Zabrze (Silesia, South Poland) was analyzed. Concentration prediction was tested for the main air pollutants, i.e., O₃, NO, NO₂, SO₂, PM₁₀, CO. Multilayer perceptrons were used to model the concentrations. The predicted concentrations were compared to the observed ones to evaluate the approximation accuracy. Prediction errors were calculated separately for the whole concentration range as well as for the specified concentration subranges. Some different measures of error were estimated. It was stated that the use of a single measure of the approximation accuracy may lead to incorrect interpretation. The application of one neural network to the entire concentration range results in different prediction accuracy in various concentration subranges. Replacing one neural network with several networks adjusted to specific concentration subranges should improve the modeling accuracy.

Keywords: air monitoring; air pollutants; air quality models; approximation error; concentration modeling; prediction; regression; multilayer perceptron; autonomous models

1. Introduction

The air pollutants can be gases or particles. The basic gas pollutants include O₃, NO, NO₂, SO₂, CO, and volatile organic compounds (VOC). Aerosol pollutants usually appear as airborne particles, i.e., very fine particles made up of either solid or liquid matter that can stay for a long time suspended in the air and spread with the wind [1]. They are called PM₁₀, PM₂.₅ or PM₁.₀, depending on the particle’s size. Air pollution comes from both natural and anthropogenic sources. In urban areas, the combustion of fuels, biomass, and organic waste is the main source of gaseous and particle pollution.

The impact of air pollution on the environment, economy, and human health are indisputable. It is also increasingly well documented in scientific reports. Air pollution is considered to be one of the most important factors influencing human health [2–5]. Polluted air can cause negative changes in living organisms, even when the concentrations do not exceed the permissible levels. Air pollution is linked to mental health disorders [6,7]. It has also been reported that air pollution can have negative economic effects related to lower employee productivity and labor supply [8–11]. The World Health Organization reported, that about 7 million people died in 2012 because of poor air quality [12]. This points to a significant global threat from air pollution. In many European countries particulate matter (measured as PM₁₀ or PM₂.₅), NO₂, and O₃ concentrations are still above acceptable limits [13,14].

Air pollution is an important social, economic, and health problem, especially in highly urbanized areas. The level of pollutant concentration in the air is standardized and
monitored all over the world. Air monitoring systems may differ from country to country, however, the basic measurement package includes monitoring of O$_3$, NO, NO$_2$, PM$_{10}$, SO$_2$, and CO concentration levels. The Polish air monitoring system is a part of the European Union system. The institution that controls the air monitoring in Poland is the Chief Inspectorate of Environmental Protection with its regional offices. This agency manages the entire domestic air monitoring system. 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. Poland has the reputation of a country with heavily polluted air. The air monitoring station in Zabrze is located in Upper Silesia—the highest urbanized area in Poland. This station ranks among the 50 most polluted stations in the EU. Measurement data from this station were selected for this research purpose.

The results of air monitoring are the basis for the air quality assessment in zones represented by individual measurement stations. The measured concentration levels are compared with the air quality standards. The occurrence of exceedances of the permissible concentration levels should lead to the implementation of programs and actions which improve the air quality. However, a formal assessment of the air quality is only possible if the completeness of the measurement data is sufficient (usually it should exceed 90% on an annual basis). Deficient measurement series hinder a correct air quality assessment [15].

The data, collected continuously by air monitoring systems as hourly concentrations, are never complete. The need to introduce the interpolated values into the measurement gaps arises when the completeness of the analyzed time-series is too low [16,17]. The simplest imputation technique is linear interpolation. However, this method is not accurate enough, in particular, it fails when monitoring is interrupted for longer periods. The methods that exploit the knowledge gathered in historical data at the same monitoring station are more accurate. The most popular are modeling techniques, which provide prediction without recourse to any data coming from outside of the monitoring system. Such models may be called autonomous models [18,19]. Concentrations of pollutants measured at air monitoring stations can be modeled using prediction techniques based on regression analysis, time-series analysis, or other statistical methods [20,21]. To approximate a specified pollutant concentration at a specified time some predictors are used. They could be concentrations of other pollutants and meteorological parameters as well. The predictors must be correlated in some way with the variable to be predicted.

A modeling method should ensure the highest possible prediction accuracy. When actual data are available, the approximation quality can be assessed by comparing the actual values with the modeled values. The comparison of the actual and predicted data enables the calculation of the modeling error. Various measures of the approximation error are used, each of them results from a specific statistical approach [22,23]. Some of them, like the mean absolute error, the root mean squared error, the mean squared error, the mean absolute relative error, or the Pearson’s correlation coefficient are commonly known in statistics. Some, such as Willmott’s indexes of agreement, are especially recommended to atmospheric sciences [24–26].

Formerly, traditional statistical techniques have been used to create regression models. Recently, regression neural networks have been becoming more and more popular. The neural models allow for deeper data mining and obtaining more accurate models. The publications provide values of modeling errors averaged over the entire concentration range of the modeled pollutant [16,27–29]. During the modeling of O$_3$ and CO concentrations, it was found that the error values may significantly differ in particular concentration ranges [30]. It seems advisable to perform tests on more pollutants and based on a dataset from a different air monitoring station.

The main aim of this analysis was the calculation of the approximation errors for the entire range of concentrations as well as for specified concentration subranges, separately for each pollutant. A long-term set of data was analyzed in the study. The data came from the air monitoring station in Zabrze (Silesia, South Poland). Concentration predictions
were tested for main air pollutants, i.e., O$_3$, NO, NO$_2$, SO$_2$, PM$_{10}$ and CO. The multilayer perceptrons were used to model the concentrations. The predicted concentrations were compared to the observed ones to evaluate the approximation accuracy. The prediction errors were calculated for the whole concentration range as well as for the specified concentration subranges separately.

2. Materials and Methods

2.1. Air Monitoring Data

The concentrations of basic air pollutants at the monitoring station in Zabrze, used in the analysis, were measured in the years 2011–2016. The station belongs to the type of urban background monitoring site. It is located at 34 Curie-Sklodowska Street in the city of Zabrze—a highly urbanized area of Upper Silesia, South Poland. The examined data were obtained from the Voivodeship Inspectorate of Environmental Protection in Katowice, which managed the air monitoring in the region during the mentioned period. The analyzed time series included the hourly concentrations of O$_3$, NO, NO$_2$, SO$_2$, PM$_{10}$, CO as well as meteorological data such as temperature, wind speed, and solar radiation.

The concentration data were measured according to standardized reference methods for monitoring stations in the EU [31]. The reference methods were described in the following EN standards:

- EN 14625:2012: Ambient Air—Standard method for the measurement of the concentration of ozone by ultraviolet photometry [32];
- EN 14211:2012: Ambient air—Standard method for the measurement of the concentration of nitrogen dioxide and nitrogen monoxide by chemiluminescence [33];
- EN 14212:2012: Ambient air—Standard method for the measurement of the concentration of sulfur dioxide by ultraviolet fluorescence [34];
- EN 14626:2012: Ambient Air—Standard method for the measurement of the concentration of carbon monoxide by non-dispersive infrared spectroscopy [35];
- EN12341:2014: Ambient Air—standard gravimetric measurement method for the determination of the PM$_{10}$ or PM$_{2.5}$ mass concentration of suspended particulate matter [36].

The devices measuring pollutants were placed in a thermostated kiosk. Each of them was equipped with auto-calibration systems. The measurements were carried out automatically.

The data subjected to the regression analysis consisted of 3 groups: time data, concentration data, and meteorological data. The following symbols were used to describe the variables:

D - day
H - hour
O$_3$ - hourly O$_3$ concentration, µg/m$^3$
NO - hourly NO concentration, µg/m$^3$
NO$_2$ - hourly NO$_2$ concentration, µg/m$^3$
SO$_2$ - hourly SO$_2$ concentration, µg/m$^3$
PM$_{10}$ - hourly concentration of PM$_{10}$, µg/m$^3$
CO - hourly CO concentration, mg/m$^3$
T - hourly mean temperature, °C
I - the hourly mean intensity of solar radiation, W/m$^2$
WS - hourly mean wind speed, m/s.

2.2. Transformation of Time Data

In the case of the date, the discrete form of this variable was changed to a cyclic form in which the same values were assigned to the same dates in different years. This procedure allowed for assigning higher values to the dates in the winter months, with the maximum equal 1.00 for 31 December, and lower values to dates in the summer months.
with the minimum of 0.00 for 2 July. Thus, in the period from 2 July to 31 December, the
date variable increases linearly by 0.005494 a day (in leap years by 0.005479) every day,
while after December 31, it decreases at the same rate for half a year, reaching 0.00 again on
2 July.

In the case of the variable describing the time of day, the minimum value of 0.00 was
assigned to 12 a.m., and the maximum value of 1.00 was assigned to 12 p.m. For hours
from 12 a.m. to 12 p.m., the variable value increases linearly from 0.00 to 1.00 by 0.08333
for each subsequent hour and then decreases by 0.08333 for subsequent hours from 12 p.m.
to 12 a.m.

2.3. Regression Models Concept

Artificial neural networks were used to build all regression models. Regression
relationships present in the data were used to predict the concentrations of pollutants. In
the specific case of the time series, the concentration of a selected pollutant was correlated
with the time data, concentrations of other pollutants, and meteorological data. The
knowledge hidden in the data can be used to make predictions according to the pattern
shown in Figure 1.

2.4. Artificial Neural Networks

All neural models used a multilayer perceptron with five neurons in a single hidden
layer (Figure 2). Such a relatively simple structure of a neural network allows for efficient
exploration of the knowledge hidden in the data [37]. Six perceptrons were created, one for
each pollutant as the output, and with 10 other variables as the inputs, The choice of the
input variables for each of the 6 models is presented in Table 1.

![Figure 1. The pattern of regression modeling shows the concept for O₃ prediction.](image1)

![Figure 2. An architecture diagram of the multilayer perceptron with five neurons in a single hidden layer.](image2)
Table 1. Classification of predictors for each of the 6 models.

| Predicted Variable (Output) | Predictors (Inputs) |
|-----------------------------|---------------------|
|                             | G  D  O3  NO  NO2  CO  SO2  PM10  WS  T  I  |
| O3                          | +  +  +  +  +  +  +  +  +  +  |
| NO                          | +  +  +  +  +  +  +  +  +  +  |
| NO2                         | +  +  +  +  +  +  +  +  +  +  |
| CO                          | +  +  +  +  +  +  +  +  +  +  |
| SO2                         | +  +  +  +  +  +  +  +  +  +  |
| PM10                        | +  +  +  +  +  +  +  +  +  +  |

The analysis was carried out using the Artificial Neural Network module in the Statistica program. During the neural network training, the analyzed set of data was randomly divided into three different subsets: the training subset (70% of cases), the verification subset (15% of cases), and the test subset (15% of cases). The BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm was used in the network learning process. The weights were initialized randomly before starting the network training. A logistic activation function was used in the neurons of the hidden layer as well as at the output. Each neural network was trained 5 times, and then the best model was selected for further analysis. The learning process was limited to 200 epochs. This number of epochs is sufficient to stabilize the modeling error. A learning rate of 0.1 was assumed. The random Gaussian initialization of the network was used. The sum of squares (SOS) was assumed as the error function.

The time series of the 6-year measurement period included 52,608 hourly cases. Only 25,523 of them were the cases with no missing data for any variable and they were used to train the networks. When the best network was chosen, the approximation errors were calculated for the entire range of concentrations as well as for several concentration subranges. The errors and their variability in different subranges were analyzed further.

2.5. Estimation of Prediction Error

The values of the prediction errors were estimated based on divergences between the predicted concentrations (model outputs) \( y_i \) and the actual concentrations \( x_i \). Seven different categories of approximation error were calculated for each regression model. The corresponding formulas were listed below, where \( n \)—number of observations, \( \bar{y} \)—average value in the set of predicted concentrations, \( \bar{x} \)—average value in the set of observed concentrations:

**MAE**—mean absolute error:

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

**MSE**—mean squared error:

\[
MSE = \frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n} 
\] (2)

**RMSE**—root mean squared error:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} 
\] (3)

**MARE**—mean absolute relative error:

\[
MARE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} 
\] (4)
3. Results

The approximation errors were calculated for 6 air pollutants: O$_3$, NO, NO$_2$, SO$_2$, PM$_{10}$, CO. The hourly concentrations of these pollutants were modeled. The prediction was performed using multilayer perceptrons. For each of the pollutants, one, the most accurate of the created neural networks, was chosen. The modeled concentration sets were compared to the actual concentrations to assess the prediction error. The modeling errors of the hourly concentrations were averaged over the entire 6-year period (2011–2016). The seven different measures of the prediction error were calculated, listed above, separately for each pollutant. The errors were estimated for the entire range of observed concentrations as well as for several concentration subranges. The error values are shown in Tables 2–7, separately for each pollutant. The tables also show the number of observations as well as the mean observed and predicted concentrations in the various ranges of concentrations.

Table 2. Values of approximation errors calculated for different subranges and the entire range of O$_3$ concentrations (hourly data, Zabrze 2011–2016); O$_3$$_{obs}$ means real O$_3$ concentration, and O$_3$$_{pred}$ means predicted O$_3$ concentration.

| Subranges of O$_3$$_{obs}$ Concentrations g/m$^3$ | Number of Observations | O$_3$$_{obs}$ Average Concentration in the Subrange g/m$^3$ | O$_3$$_{pred}$ Average Concentration in the Subrange g/m$^3$ | MAE $\mu$g/m$^3$ | $MSE$ $(\mu$g/m$^3)^2$ | $RMSE$ $\mu$g/m$^3$ | MARE | r | d | $d_1$
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 ÷ 20 | 7881 | 8.9 | 13.6 | 6.7 | 86.4 | 9.3 | 1.082 | 0.542 | 0.602 | 0.445|
| 20 ÷ 40 | 5626 | 29.6 | 32.2 | 8.7 | 125.4 | 11.2 | 0.306 | 0.351 | 0.530 | 0.388|
| 40 ÷ 60 | 5201 | 49.4 | 47.3 | 9.3 | 139.6 | 11.8 | 0.190 | 0.357 | 0.521 | 0.373|
| 60 ÷ 80 | 3429 | 68.7 | 63.6 | 12.0 | 221.5 | 14.9 | 0.175 | 0.346 | 0.458 | 0.327|
| 80 ÷ 100 | 1834 | 88.9 | 83.7 | 11.9 | 242.6 | 15.6 | 0.135 | 0.354 | 0.443 | 0.323|
| 100 ÷ 120 | 905 | 108.6 | 101.7 | 12.9 | 266.0 | 16.3 | 0.118 | 0.279 | 0.419 | 0.307|
| 120 ÷ 140 | 443 | 127.9 | 117.7 | 14.3 | 324.0 | 18.0 | 0.112 | 0.208 | 0.339 | 0.236|
| 140 ÷ 160 | 144 | 147.9 | 134.8 | 14.4 | 312.4 | 17.7 | 0.097 | 0.356 | 0.385 | 0.267|
| 160 ÷ 180 | 59 | 166.7 | 145.2 | 21.4 | 511.5 | 22.6 | 0.128 | 0.389 | 0.275 | 0.153|
| 180 ÷ 200 | 5 | 188.9 | 151.1 | 37.8 | 1539.8 | 39.2 | 0.199 | −0.337 | 0.230 | 0.137|
| 0 ÷ 200 | 25,523 | 42.3 | 42.3 | 9.2 | 148.2 | 12.2 | 0.480 | 0.927 | 0.961 | 0.817|

3.1. Modelling of O$_3$ Concentrations

The prediction results of the O$_3$ concentrations were presented in Table 2. The error measures, such as the $MAE$, $MSE$, and $RMSE$, behave alike. They achieve the minimum values for the first O$_3$ concentration subrange (0–20 $\mu$g/m$^3$), which is the most numerous one. Moving to higher concentration subranges the modeling precision decreases alongside the decrease of the number of observations. The opposite changes are shown by the mean
absolute relative error (MARE). The \( r, d, d_1 \) measures have much lower values in the specified subranges than in the entire concentration range.

3.2. Modelling of NO Concentrations

The prediction results of the NO concentrations are presented in Table 3. The values of MAE, MSE, and RMSE behave alike. They achieve the minimum values for the first NO concentration subrange (0–20 \( \mu g/m^3 \)) which is also the most numerous subrange. These errors gradually increase, with minor variations, as we move to higher concentration subranges. In the same direction, the number of observations in the subranges decreases. The opposite changes are shown by values of MARE, \( r, d, d_1 \). In the case of \( r, d, d_1 \) the decreasing of values may be interpreted as a loss of the modeling accuracy. Similar to the ozone \( r, d, d_1 \) have much lower values in the specified subranges than in the entire concentration range.

Table 3. Values of approximation errors calculated for different subranges and the entire range of NO\(_{obs}\) concentrations (hourly data, Zabrze 2011–2016), NO\(_{obs}\) means real NO concentration, and NO\(_{pred}\) means predicted NO concentration.

| Subranges of NO\(_{obs}\) Concentrations \( \mu g/m^3 \) | Number of Observations | NO\(_{obs}\) Average Concentration in the Subrange \( \mu g/m^3 \) | NO\(_{pred}\) Average Concentration in the Subrange \( \mu g/m^3 \) | MAE \( \mu g/m^3 \) | MSE \( (\mu g/m^3)^2 \) | RMSE \( \mu g/m^3 \) | MARE | \( r \) | \( d \) | \( d_1 \) |
|------------------------------------------------------|-------------------------|-------------------------------------------------------------|-------------------------------------------------------------|----------------|----------------|----------------|------|------|------|------|
| 0 \( \div \) 20                                    | 22,339                  | 4.1                                                         | 4.6                                                         | 2.3            | 17.8           | 4.2            | 0.731 * | 0.707 | 0.802 | 0.651 |
| 20 \( \div \) 40                                   | 1679                    | 28.1                                                        | 27.8                                                        | 9.2            | 138.6          | 11.8           | 0.330   | 0.351 | 0.521 | 0.373 |
| 40 \( \div \) 60                                   | 670                     | 48.3                                                        | 42.4                                                        | 14.2           | 324.4          | 18.0           | 0.291   | 0.299 | 0.400 | 0.285 |
| 60 \( \div \) 80                                   | 329                     | 68.3                                                        | 55.9                                                        | 19.9           | 591.0          | 24.3           | 0.289   | 0.053 | 0.264 | 0.192 |
| 80 \( \div \) 100                                  | 170                     | 88.3                                                        | 77.2                                                        | 26.2           | 1013.8         | 31.8           | 0.297   | 0.235 | 0.251 | 0.178 |
| 100 \( \div \) 120                                 | 110                     | 109.5                                                       | 91.0                                                        | 30.1           | 1320.1         | 36.3           | 0.275   | 0.193 | 0.222 | 0.159 |
| 120 \( \div \) 140                                 | 73                      | 128.9                                                       | 116.3                                                       | 38.8           | 1310.4         | 36.2           | 0.224   | 0.200 | 0.221 | 0.164 |
| 140 \( \div \) 160                                 | 41                      | 148.3                                                       | 122.3                                                       | 34.4           | 1719.0         | 41.5           | 0.233   | 0.222 | 0.194 | 0.134 |
| 160 \( \div \) 180                                 | 28                      | 170.0                                                       | 164.8                                                       | 31.3           | 1301.0         | 36.1           | 0.184   | 0.137 | 0.208 | 0.138 |
| 180 \( \div \) 200                                 | 28                      | 189.2                                                       | 168.1                                                       | 33.9           | 1640.7         | 40.5           | 0.178   | –0.323 | 0.117 | 0.085 |
| 200 \( \div \) 572                                 | 56                      | 286.4                                                       | 248.9                                                       | 51.1           | 3927.5         | 62.7           | 0.196   | 0.883 | 0.892 | 0.689 |
| 0 \( \div \) 572                                   | 25,523                  | 10.2                                                        | 10.1                                                        | 3.9            | 71.7           | 8.5            | 0.676 * | 0.933 | 0.965 | 0.829 |

* cases with NO\(_{obs}\) = 0 were removed to calculate MARE.

3.3. Modelling of NO\(_2\) Concentrations

The prediction results of the NO\(_2\) concentrations are presented in Table 4. The MAE, MSE, and RMSE achieve the minimum values for the first NO\(_2\) concentration subrange (0–20 \( \mu g/m^3 \)). This subrange is also the most numerous one. Moving to higher concentrations, the MAE, MSE, and RMSE values in subranges increase. It means that modeling accuracy decreases. In the same direction, the number of observations in the subranges decreases. The opposite changes are shown by values of other errors: MARE, \( r, d, d_1 \). The decreasing values of \( r, d, d_1 \) may be interpreted as the loss of modeling accuracy. \( R, d, d_1 \) show much lower values in the specified subranges than in the entire concentration range.

3.4. Modelling of SO\(_2\) Concentrations

The prediction results of the SO\(_2\) concentrations are presented in Table 5. The error measures, such as MAE, MSE, and RMSE, achieve the minimum values for the first SO\(_2\) concentration subrange (0–10 \( \mu g/m^3 \)). This subrange is also the most numerous one. Moving to the higher concentration subranges the prediction precision decreases, and the number of observations also decreases. The opposite changes are shown by the MARE, \( r, d, d_1 \). When the subrange width is bigger, the values of \( r, d, d_1 \) may increase. In the wide subranges, 100–200 \( \mu g/m^3 \) and 200–308 \( \mu g/m^3 \) these values are bigger than in the
narrow subranges, for example, 90–100 µg/m³. The highest values of these measures are observed for the entire concentration range.

Table 4. Values of approximation errors calculated for different subranges and the entire range of NO₂,obs concentrations (hourly data, Zabrze 2011–2016). NO₂,obs means real NO₂ concentration, and NO₂,pred means predicted NO₂ concentration.

| Subranges of NO₂,obs Concentrations g/m³ | Number of Observations | NO₂,obs Average Concentration in the Subrange g/m³ | NO₂,pred Average Concentration in the Subrange g/m³ | MAE µg/m³ | MSE (µg/m³)² | RMSE µg/m³ | MARE r d d₁ |
|----------------------------------------|------------------------|-----------------------------------------------|-----------------------------------------------|------------|--------------|------------|-------------|
| 0–20                                   | 12,501                 | 11.5                                          | 13.8                                          | 3.8        | 26.1         | 5.1        | 0.405       |
| 20–40                                  | 8902                   | 28.4                                          | 28.8                                          | 6.1        | 60.4         | 7.8        | 0.217       |
| 40–60                                  | 3119                   | 47.5                                          | 42.6                                          | 8.3        | 104.0        | 10.2       | 0.174       |
| 60–80                                  | 768                    | 67.6                                          | 56.1                                          | 13.8       | 273.4        | 16.5       | 0.203       |
| 80–100                                 | 177                    | 86.9                                          | 66.9                                          | 21.6       | 614.4        | 24.8       | 0.247       |
| 100–120                                | 44                     | 107.8                                         | 79.9                                          | 28.4       | 1121.2       | 33.5       | 0.264       |
| 120–140                                | 12                     | 128.8                                         | 78.0                                          | 50.9       | 3318.7       | 57.6       | 0.396       |
| 0–140                                  | 25,523                 | 24.2                                          | 24.3                                          | 5.6        | 62.5         | 7.9        | 0.304       |

Table 5. Values of approximation errors calculated for different subranges and the entire range of SO₂,obs concentrations (hourly data, Zabrze 2011–2016). SO₂,obs means real SO₂ concentration, and SO₂,pred means predicted SO₂ concentration.

| Subranges of SO₂,obs Concentrations g/m³ | Number of Observations | SO₂,obs Average Concentration in the Subrange g/m³ | SO₂,pred Average Concentration in the Subrange g/m³ | MAE µg/m³ | MSE (µg/m³)² | RMSE µg/m³ | MARE r d d₁ |
|------------------------------------------|------------------------|-----------------------------------------------|-----------------------------------------------|------------|--------------|------------|-------------|
| 0–10                                     | 12,842                 | 5.2                                          | 7.5                                          | 3.0        | 18.0         | 4.2        | 0.813       |
| 10–20                                    | 5652                   | 14.0                                         | 14.2                                         | 4.8        | 39.5         | 6.3        | 0.344       |
| 20–30                                    | 2952                   | 24.3                                         | 22.9                                         | 7.1        | 78.8         | 8.9        | 0.292       |
| 30–40                                    | 1825                   | 34.3                                         | 31.4                                         | 11.9       | 119.5        | 10.9       | 0.252       |
| 40–50                                    | 1029                   | 44.2                                         | 39.6                                         | 10.1       | 170.9        | 13.1       | 0.229       |
| 50–60                                    | 566                    | 54.0                                         | 47.7                                         | 12.4       | 235.1        | 15.3       | 0.229       |
| 60–70                                    | 348                    | 64.3                                         | 55.9                                         | 14.1       | 306.9        | 17.5       | 0.220       |
| 70–80                                    | 216                    | 74.6                                         | 64.1                                         | 16.7       | 422.1        | 20.5       | 0.224       |
| 80–90                                    | 136                    | 83.8                                         | 69.6                                         | 19.4       | 567.0        | 23.8       | 0.232       |
| 90–100                                   | 91                     | 94.5                                         | 83.4                                         | 19.4       | 575.2        | 24.0       | 0.206       |
| 100–200                                  | 214                    | 127.6                                        | 107.8                                        | 27.5       | 1383.8       | 37.2       | 0.212       |
| 200–300                                  | 12                     | 227.2                                        | 195.2                                        | 34.9       | 1624.4       | 40.3       | 0.149       |
| 0–308                                    | 25,523                 | 17.4                                         | 17.3                                         | 5.4        | 72.5         | 8.5        | 0.549       |

3.5. Modelling of PM₁₀ Concentrations

The prediction results of the PM₁₀ concentrations are presented in Table 6. The MAE, MSE, and RMSE achieve the minimum values for the second PM₁₀ concentration subrange (20–40 µg/m³). This subrange is also the most numerous one. Moving from the second subrange to the higher subranges these errors increase. It means that modeling accuracy decreases. In the same direction, the number of observations in the subranges decreases. The values of the MARE gradually decrease as we move to the higher concentration subranges. The decreasing of the r, d, d₁ values is observed up to the subrange (140–160 µg/m³). For the wider subranges, i.e., for the subranges above 200 µg/m³, the values of r, d, d₁ are higher, which can be explained by the effect of the range extension. The highest values of these three measures were estimated for the entire concentration range (0–1000 µg/m³).
3.6. Modelling of CO Concentrations

The prediction results of the CO concentrations in subranges are presented in Table 7. The MAE, MSE, and RMSE achieve the minimum values for the first CO concentration subrange (0–1 mg/m$^3$). This is also the most numerous subrange. Moving from this subrange to the higher subranges these errors increase up to the subrange (4–5 mg/m$^3$). In the same direction, the number of observations in the subranges decreases quickly. For the higher subranges, the uptrends are disrupted. Up to this subrange (4–5 mg/m$^3$) also the decreasing of the $r$, $d$, $d_1$ values is observed. For the higher subranges, the trend is disrupted. The highest values of $r$, $d$, $d_1$ were estimated for the entire concentration range (0–9 mg g/m$^3$).

Table 6. Values of approximation errors calculated for different subranges and the entire range of PM$^{10,\text{obs}}$ concentrations (hourly data, Zabrze 2011–2016), PM$^{10,\text{obs}}$ means real PM$^{10}$ concentration, and PM$^{10,\text{pred}}$ means predicted PM$^{10}$ concentration.

| Subranges of PM$^{10,\text{obs}}$ Concentrations g/m$^3$ | Number of Observations | PM$^{10,\text{obs}}$ Average Concentration in the Subrange g/m$^3$ | PM$^{10,\text{pred}}$ Average Concentration in the Subrange g/m$^3$ | MAE µg/m$^3$ | MSE (µg/m$^3$)$^2$ | RMSE µg/m$^3$ | MARE | $r$ | $d$ | $d_1$ |
|-------------------------------------------------|-------------------------|------------------------------------------------------------|-------------------------------------------------------------|-------------|-----------------|----------------|-------|-----|-----|-----|
| 0 ÷ 20                                          | 6112                    | 13.2                                                       | 21.8                                                        | 8.7         | 114.1           | 10.7           | 0.951 | 0.277 | 0.395 | 0.282 |
| 20 ÷ 40                                         | 8859                    | 28.8                                                       | 30.5                                                        | 6.9         | 89.3            | 9.5             | 0.243 | 0.374 | 0.562 | 0.426 |
| 40 ÷ 60                                         | 4468                    | 48.3                                                       | 45.0                                                        | 11.6        | 209.4           | 14.5            | 0.239 | 0.318 | 0.458 | 0.327 |
| 60 ÷ 80                                         | 2146                    | 68.9                                                       | 63.9                                                        | 16.5        | 412.1           | 20.3            | 0.240 | 0.272 | 0.370 | 0.259 |
| 80 ÷ 100                                        | 1111                    | 88.8                                                       | 84.2                                                        | 19.0        | 579.6           | 24.1            | 0.214 | 0.227 | 0.319 | 0.233 |
| 100 ÷ 120                                       | 770                     | 108.9                                                      | 102.1                                                       | 22.2        | 781.4           | 28.0            | 0.203 | 0.178 | 0.264 | 0.191 |
| 120 ÷ 140                                       | 565                     | 128.8                                                      | 122.5                                                       | 25.4        | 1013.4          | 31.8            | 0.197 | 0.123 | 0.220 | 0.163 |
| 140 ÷ 160                                       | 361                     | 149.8                                                      | 141.2                                                       | 29.3        | 1396.7          | 37.4            | 0.195 | –0.020 | 0.155 | 0.120 |
| 160 ÷ 180                                       | 254                     | 169.2                                                      | 161.3                                                       | 33.3        | 1676.0          | 40.9            | 0.197 | 0.128 | 0.180 | 0.127 |
| 180 ÷ 200                                       | 198                     | 188.5                                                      | 178.3                                                       | 33.3        | 1846.3          | 43.0            | 0.176 | 0.313 | 0.218 | 0.168 |
| 200 ÷ 400                                       | 560                     | 262.7                                                      | 243.4                                                       | 42.8        | 2903.0          | 53.9            | 0.165 | 0.662 | 0.777 | 0.571 |
| 400 ÷ 600                                       | 78                      | 481.7                                                      | 445.5                                                       | 81.7        | 10,393.8        | 102.0           | 0.169 | 0.550 | 0.626 | 0.449 |
| 600 ÷ 800                                       | 34                      | 689.1                                                      | 649.2                                                       | 77.0        | 10,356.4        | 101.8           | 0.114 | 0.425 | 0.494 | 0.362 |
| 800 ÷ 1000                                      | 6                       | 884.2                                                      | 791.5                                                       | 92.7        | 11,590.8        | 107.7           | 0.101 | 0.607 | 0.503 | 0.374 |
| 0 ÷ 1000                                        | 25,523                  | 51.1                                                       | 51.3                                                        | 12.3        | 363.6           | 19.1            | 0.404 | 0.948 | 0.973 | 0.818 |

Table 7. Values of approximation errors calculated for different subranges and the entire range of CO$^{\text{obs}}$ concentrations (hourly data, Zabrze 2011–2016), CO$^{\text{obs}}$ means real CO concentration, and CO$^{\text{pred}}$ means predicted CO concentration.

| Subranges of CO$^{\text{obs}}$ Concentrations mg/m$^3$ | Number of Observations | CO$^{\text{obs}}$ Average Concentration in the Subrange mg/m$^3$ | CO$^{\text{pred}}$ Average Concentration in the Subrange mg/m$^3$ | MAE mg/m$^3$ | MSE (mg/m$^3$)$^2$ | RMSE mg/m$^3$ | MARE | $r$ | $d$ | $d_1$ |
|-------------------------------------------------|-------------------------|------------------------------------------------------------|-------------------------------------------------------------|-------------|-----------------|----------------|-------|-----|-----|-----|
| 0 ÷ 1                                           | 21,935                  | 0.426                                                      | 0.447                                                        | 0.093       | 0.017           | 0.131           | 0.256 | 0.806 | 0.892 | 0.713 |
| 1 ÷ 2                                           | 3303                    | 1.536                                                      | 1.433                                                        | 0.260       | 0.112           | 0.335           | 0.174 | 0.804 | 0.878 | 0.677 |
| 2 ÷ 3                                           | 511                     | 2.391                                                      | 2.236                                                        | 0.355       | 0.196           | 0.443           | 0.148 | 0.523 | 0.646 | 0.464 |
| 3 ÷ 4                                           | 168                     | 3.424                                                      | 3.056                                                        | 0.599       | 0.712           | 0.844           | 0.179 | 0.482 | 0.445 | 0.317 |
| 4 ÷ 5                                           | 55                      | 4.321                                                      | 3.699                                                        | 0.870       | 1.740           | 1.319           | 0.206 | 0.439 | 0.286 | 0.247 |
| 5 ÷ 6                                           | 16                      | 5.396                                                      | 5.270                                                        | 0.596       | 0.487           | 0.698           | 0.110 | 0.899 | 0.706 | 0.473 |
| 6 ÷ 7                                           | 16                      | 6.601                                                      | 6.307                                                        | 0.696       | 0.692           | 0.832           | 0.106 | 0.448 | 0.471 | 0.324 |
| 7 ÷ 8                                           | 25                      | 7.431                                                      | 7.298                                                        | 0.421       | 0.266           | 0.516           | 0.056 | –0.012 | 0.388 | 0.301 |
| 8 ÷ 9                                           | 5                       | 8.402                                                      | 7.788                                                        | 0.614       | 0.439           | 0.662           | 0.073 | 0.444 | 0.417 | 0.262 |
| 0 ÷ 9                                           | 25,523                  | 0.615                                                      | 0.615                                                        | 0.121       | 0.039           | 0.197           | 0.244 | 0.947 | 0.972 | 0.832 |
3.7. The Comparison of Real and Predicted Concentrations

The scatterplots of the observed and the predicted concentration values for O$_3$, NO, NO$_2$, SO$_2$, PM$_{10}$ and CO are presented in Figure 3. The scatterplots show that neural prediction models underestimate the values for the higher concentrations. The same assessment results from comparing the averages of the actual and the predicted concentrations in the specified subranges for the pollutants (Tables 2–7). This effect applies to all the studied pollutants.

Figure 3. Scatterplots of observed and predicted concentration values; (a) scatterplot for O$_3$; (b) scatterplot for NO; (c) scatterplot for NO$_2$; (d) scatterplot for SO$_2$; (e) scatterplot for PM$_{10}$; (f) scatterplot for CO; the explicit form of the regression equations are included in the legends; a red line for equations $y = x$; a blue line for regression equations.
4. Discussion

The error measures based on differences between the sets of real and predicted concentrations, such as MAE, MSE, and RMSE, behave in like manner. They achieve the minimum values for the most numerous concentration subranges. For most pollutants, measured at the station Zabrze, it is the first subrange with the lowest concentrations. The only exception is the PM$_{10}$, for which the second subrange (20–40 µg/m$^3$) is the most numerous one. A similar effect was noted in the previous studies for a dataset from another air monitoring station in Lodz (central Poland) [30]. In that work, the error values for two pollutants O$_3$ and CO were calculated. The most numerous subranges showed the lowest values of MAE, MSE, and RMSE: the third concentration subrange of O$_3$ (40–60 µg/m$^3$) and the concentration subrange (0.3–0.4 mg/m$^3$) of CO. Both in the earlier and the current studies, the values of the remaining measures of modeling accuracy, i.e., MARE, $r$, $d$, and $d_1$, did not always reach optimal values in the same subranges. For all the pollutants, the mean absolute relative error (MARE) is extremely high in the first subrange of the actual concentrations. In the subsequent subranges, the values of MARE decrease rapidly. The MARE value can be misleading at very low concentrations of the pollutants, close to 0.0 µg/m$^3$ (or 0.0 mg/m$^3$ for CO). For the zero concentration, it is not feasible to calculate the MARE. For the concentrations close to 0 µg/m$^3$, the MARE values can be extremely high, regardless of the differences between the actual and predicted values. This measure cannot be recommended for assessing the accuracy of models in the subranges of the lowest concentrations. The formulas for $r$, $d$, and $d_1$ measures refer to the averages and the distance of the individual measurements from the averages. Generally, these measures have much lower values in the narrow subranges than in the entire concentration range. Comparing values of $r$, $d$, and $d_1$ for the entire range model with the error values for subranges can lead to false conclusions, for example, that modeling over the entire concentration range is more accurate than modeling in subranges. In any concentration subranges, the model does not reach values of the coefficients $r$, $d$, $d_1$ obtained for the entire range. The value of $r$, calculated for the entire concentration ranges of different pollutants, is in the scope 0.88–0.95, for all the pollutants. Usually, such values are considered to exhibit good modeling quality. However, a deeper analysis of the $r$ values in the subranges leads to completely different conclusions. For example, the $r$ value for the entire range of NO concentrations equals 0.933, but for the subrange (180–200 µg/m$^3$) it equals −0.323 (Table 3). The negative sign means a negative correlation of the actual and the predicted concentrations in this subrange. This result shows that it is necessary to carefully draw conclusions about the accuracy of the models based on the value of $r$. The accuracy measures such as $r$, $d$, $d_1$ fail when assessing the prediction precision, but these measures can be used for a comparison of precisions of models created for different air pollutants.

The errors based on differences between the actual and the predicted concentrations, like MAE, MSE, RMSE, show better prediction quality. However, the neural networks adapted to the whole range of concentrations cannot predict with the same quality in different concentration subranges. Moving on to less numerous concentration subranges, the modeling precision falls due to fewer training cases. This is due to the specific nature of machine learning. The adaptation process is predominated by the cases from the most numerous concentration ranges. The MSE formula is based on the sum of squared errors of the individual cases. The sum of squares (SOS) is also assumed as the error function during the neural network adaptation process. Thus, the neural network adaptation process leads to the minimization of the MSE error as well as related errors, i.e., RMSE and MAE. The advantage of measures such as MAE, MSE, RMSE is the ability to reflect real modeling accuracies in different concentration subranges.

In publications on modeling the air pollutant concentrations, the dominant approach is to create a single model that works across the entire concentration range. Among the publications, there are those in which only one measure of prediction accuracy is used, for example, MAE [38] or $R^2$ [39]. This approach seems risky, especially when using measures
such as $R^2$, which are highly dependent on the concentration range. It is advisable to use several different error measures to avoid misinterpretation.

5. Conclusions

Based on the results of the analysis, the following conclusions were drawn:

1. The use of a single measure of the approximation accuracy may lead to incorrect interpretation, especially when the measures refer in their formulas to the averages and the distances from averages. Such measures include Pearson’s correlation coefficient $r$ and Willmott’s indexes of agreement $d$, $d_1$.

2. Measures like $MAE$, $MSE$, $RMSE$ properly reflect the difficulties in modeling the concentrations in the entire range as well as in different subranges of concentrations.

3. The average relative errors like $MARE$ cannot be recommended for assessing the accuracies of autonomous models.

4. The application of one neural network to the entire concentration range results in different prediction accuracy in various concentration subranges. It is advisable to replace one neural network with several networks (submodels) adapted to specific concentration subranges. An entire-range model can be used initially to obtain predicted concentrations, and then the initially predicted concentration values can be classified into individual subranges and modeled more precisely in submodels.

Funding: The scientific research was funded by the statute subvention of the Czestochowa University of Technology, Faculty of Infrastructure and Environment, BS/PB-400-301/21.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The author would like to thank the Voivodeship Inspectorate of Environmental Protection in Katowice for providing the regional air monitoring data which were used in the analysis.

Conflicts of Interest: The author declares 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. Kumar, P. Airborne Particles: Origin, Emissions and Health Impacts; Nova Science Publisher’s, Inc.: Hauppauge, NY, USA, 2017.
2. Hoffmann, B.; Roebbel, N.; Gumy, S.; Forastiere, F.; Brunekeef, B.; Jarosinska, D.; Walker, K.D.; Van Erp, A.M.; O’Keefe, R.; Greenbaum, D.; et al. Air Pollution and Health: Recent Advances in Air Pollution Epidemiology to Inform the European Green Deal: A joint workshop report of ERS, WHO, ISEE and HEL. Eur. Respir. J. 2020, 56, 2002575. [CrossRef] [PubMed]
3. Gurjar, B.R.; Molina, L.T.; Ojha, C.S.P. Air Pollution: Health and Environmental Impacts; CRC Press: Boca Raton, FL, USA, 2010. [CrossRef]
4. Adamkiewicz, G.; Liddie, J.; Gaffin, J.M. The Respiratory Risks of Ambient/Outdoor Air Pollution. Clin. Chest Med. 2020, 41, 809–824. [CrossRef]
5. Finicelli, M.; Squillaro, T.; Galderisi, U.; Peluso, G. Crossroads Between the Exposure to Environmental Particulate Pollution and the Obstructive Pulmonary Disease. Int. J. Mol. Sci. 2020, 21, 7221. [CrossRef] [PubMed]
6. 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]
7. Kim, Y.; Manley, J.; Radiois, V. Air Pollution and Long Term Mental Health. Atmosphere 2020, 11, 1355. [CrossRef]
8. 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]
9. Graff-Zivin, J.; Neidell, M. The Impact of Pollution on Worker Productivity. Am. Econ. Rev. 2012, 102, 3652–3673. [CrossRef]
10. 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]
11. 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]
12. 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).
13. Maesano, I. The Air of Europe: Where Are We Going? *Eur. Respir. Rev.* 2017; 26, 170024. [CrossRef]
14. European Environment Agency. *Air Quality in Europe–2020 Report*. No. 12/2018; Publications Office of the European Union: Luxembourg, 2020.
15. 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 30 September 2021). (In Polish)
16. Plaia, A.; Bondi, A.L. Single Imputation Method of Missing Values in Environmental Pollution Data Sets. *Atmos. Environ.* 2006, 40, 7316–7330. [CrossRef]
17. Gentili, S.; Magnaterra, L.; Passerini, G. *Handling Missing Data: Applications to Environmental Analysis*; Latini, G., Passerini, G., Eds.; Wit Press: Southampton, UK, 2004.
18. Hoffman, S. *Environmental Engineering*; Pawłowski, L., Dudzińska, M.R., Pawłowski, A., Eds.; Taylor & Francis Group: London, UK, 2007; pp. 349–353.
19. Hoffman, S. Approximation of Emission Level at Air Monitoring Stations by Means of Autonomous Neural Models. *Environ. Prot. Eng.* 2012, 38, 109–119. [CrossRef]
20. Millionis, A.E.; Davies, T.D. Regression and Stochastic Models for Air Pollution—I. Review, Comments and Suggestions. *Atmos. Environ.* 1994, 28, 2801–2810. [CrossRef]
21. 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]
22. Venkatram, A. Computing and Displaying Model Performance Statistics. *Atmos. Environ.* 2008, 42, 6862–6868. [CrossRef]
23. Mouton, A.M.; De Baets, B.; Goethals, P.L.M. Ecological Relevance of Performance Criteria for Species Distribution Models. *Ecol. Model.* 2010, 221, 1995–2002. [CrossRef]
24. Willmott, C.J. Some Comments on the Evaluation of Model Performance. *Bull. Am. Meteorol. Soc.* 1982, 63, 1309–1313. [CrossRef]
25. Willmott, C.J.; Ackleson, S.G.; Davis, R.E.; Feddema, J.J.; Klink, K.M.; Legates, D.R.; O’Donnell, J.; Rowe, C.M. Statistics for the Evaluation and Comparison of Models. *J. Geophys. Res.* 1985, 90, 8995–9005. [CrossRef]
26. Willmott, C.J.; Robeson, S.M.; Matsuura, K. A Refined Index of Model Performance. *Int. J. Climatol.* 2011, 32, 2088–2094. [CrossRef]
27. 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. [CrossRef]
28. 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]
29. Nagendra, S.S.M.; Khare, M. Modelling Urban Air Quality Using Artificial Neural Network. *Clean. Technol. Environ. Policy* 2005, 7, 116–126. [CrossRef]
30. Hoffman, S. Assessment of Prediction Accuracy in Autonomous Air Quality Models. *Desalination Water Treat.* 2015, 57, 1322–1326. [CrossRef]
31. The European Parliament and The Council of the European Union. 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. *Off. J. Eur. Union* 2008, 152, 1–44.
32. EN 14625:2012; Ambient Air—Standard Method for the Measurement of the Concentration of Ozone by Ultraviolet Photometry.
33. EN 14211:2012; Ambient Air—Standard Method for the Measurement of the Concentration of Nitrogen Dioxide and Nitrogen Monoxide by Chemiluminescence.
34. EN 14212:2012; Ambient Air—Standard Method for the Measurement of the Concentration of Sulphur Dioxide by Ultraviolet Fluorescence.
35. EN 14626:2012; Ambient Air—Standard Method for the Measurement of the Concentration of Carbon Monoxide by Non-dispersive Infrared Spectroscopy.
36. EN 12341:2014; Ambient Air—Standard Gravimetric Measurement Method for the Determination of the PM10 or PM2.5 Mass Concentration of Suspended Particulate Matter.
37. Hoffman, S. *Application of Neural Networks in Regression Modelling of Air Pollution Concentrations*; Wydawnictwa Politechniki Częstochowskiej: Częstochowa, Poland, 2004. (In Polish)
38. Trenchevski, A.; Kalendar, M.; Gjoreski, H.; Efrosheva, D. Prediction of Air Pollution Concentration Using Weather Data and Regression Models. In *Proceedings of the 8th International Conference on Applied Innovations in IT, (ICAIIIT)*; Issue 1Siemens, E., Mylnikov, L., Eds.; Anhalt University of Applied Sciences; Perm National Research Polytechnic University: Koethen, Germany, 2020; Volume 8.
39. 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]