Article

Monitoring of PM_{2.5} Concentrations by Learning from Multi-Weather Sensors

Yuexia Wang and Zhihuo Xu *

Radar Remote Sensing Group, School of Transportation, Nantong University, Nantong 226019, China; venus@ntu.edu.cn

* Correspondence: xuzhihuo@ntu.edu.cn

Received: 28 September 2020; Accepted: 24 October 2020; Published: 26 October 2020

Abstract: This paper aims to monitor the ambient level of particulate matter less than 2.5 μm (PM_{2.5}) by learning from multi-weather sensors. Over the past decade, China has established a high-density network of automatic weather stations. In contrast, the number of PM monitors is much smaller than the number of weather stations. Since the haze process is closely related to the variation of meteorological parameters, it is possible and promising to calculate the concentration of PM_{2.5} by studying the data from weather sensors. Here, we use three machine learning methods, namely multivariate linear regression, multivariate nonlinear regression, and neural network, in order to monitor PM_{2.5} by exploring the data of multi-weather sensors. The results show that the multivariate linear regression method has the root mean square error (RMSE) of 24.6756 $\mu\text{g}/\text{m}^3$ with a correlation coefficient of 0.6281, by referring to the ground truth of PM_{2.5} time series data; and the multivariate nonlinear regression method has the RMSE of 24.9191 $\mu\text{g}/\text{m}^3$ with a correlation coefficient of 0.6184, while the neural network based method has the best performance, of which the RMSE of PM_{2.5} estimates is 15.6391 $\mu\text{g}/\text{m}^3$ with the correlation coefficient of 0.8701.

Keywords: particulate matter; meteorological parameters; multivariate linear regression; multivariate nonlinear regression; neural network; machine learning

1. Introduction

Particulate matter (PM) is a kind of atmospheric aerosol, formatting as minute solid particles or liquid droplets suspended in air [1]. PM is mainly from anthropogenic origin, derived from industrial, home heating and cooking, and transportation sources while most natural sources are relatively less important [2]. PM less than 10 μm (PM₁₀) and PM less than 2.5 μm (PM_{2.5}) consist of a number of components such as sulfate, nitrate, ammonium, elemental carbon, organic carbon, and soil or dust particles. Fine particles, PM_{1.0} (aerodynamic diameter of less than 1.0 μm), carry toxic trace elements, like Se, S, V, Cu, Fe, Pb, As, Cd, Ni, Zn, Mn, etc. [3]. Scientific studies reveal that these PM increase the risk of anthroposphere, atmosphere, hydrosphere, biosphere, and lithosphere [2–4]. The negative impact of PM can be summarized as follows: first, it affects human health [5–7]. PM_{2.5} and PM₁₀ generally passes through the nose and throat and even enters the lungs; fine particles, PM_{1.0}, and smaller particles are able to penetrate into the human respiratory and circulation system, resulting in adverse health effects [4]. Second, PM episodes reduce visibility and lead to climate change [8]. PM is the main cause of reduced visibility (haze) in the world. It suppresses convection and precipitation by both radiative and micro-physical effects, changes lighting phenomenon in different regions, weakens the hydrological cycle, and leads to less fresh water and nutrient imbalance in coastal waters and large river basins [4]. Third, particle pollution and acid rain make lakes and streams acidic, damage sensitive forests, farm crops, stone, soil buildings and other materials, and deplete the nutrients in soil, affecting the diversity of ecosystems [9].

As particles pollution is becoming an increasingly severe problem, measurement and prediction of PM are crucial for environment protection and climate investigations. Since the 1970s, global aerosol monitoring programs using remote sensing tools have been launched to measure the level of component aerosol on the regional or global scale [10–15]. Launched on 13 October 2017, the European Space Agency (ESA) Sentinel-5 Precursor is a low-Earth orbit polar satellite designed to provide information and services on air quality, climate, and the ozone layer, as part of the Global Monitoring for Environment and Security Space Sub-Programme. The mission's payload is the tropospheric monitor, which measures key atmospheric constituents, including ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide, methane, nitrous oxide, and Ultraviolet aerosol index [16]. At the same time, ground based direct samplers were developed to accurately measure PM concentrations [17,18]. Zhang et al. [19] demonstrated how vertical wind shear affects ground-level PM_{2.5} by using radar wind profiler observations in Beijing, providing new insights into the role of vertical wind shear in modulating the variation of PM_{2.5}, which is worth considering in future air quality predictions. In addition, airborne observations were conducted to advance the remote sensing of aerosols. Knobelspiesse et al. [20] conducted one field campaign that used the airborne hyper angular rainbow polarimeter, the airborne multiangle spectro-polarimetric imager, the airborne spectrometer for planetary exploration, and the research scanning polarimeter to test new observation systems, develop new algorithms, and validate orbital observations. Wang et al. [21] developed an unmanned aerial vehicle PM monitoring system that can performed three-dimensional stereoscopic observation of PM_{2.5} and PM₁₀ in the atmosphere. These measurements significantly improve our understanding of the characteristics of PM.

Regarding predictions of PM concentrations, some methods have been developed. For example, a Bayesian based regression model by Mølgaard et al. [22], a Gaussian process regression method by Reggente et al. [23], and a multiple linear regression model have been applied for forecasting the fine particle concentrations based on measurements of nitrogen oxides. Using the raw data of air temperature, relative humidity and PM, Wang et al. [21] recently applied a multiple linear regression model, support vector machine, and random forest for correction of PM_{2.5} and obtained satisfactory results. To reduce bias and improve non-reference monitoring data used for community air monitoring studies, Commodore et al. [24] proposed one nonlinear statistical model as an example of instrument evaluation prior to assessment of non-reference monitoring measures of PM. Based on satellite observations and ground monitor data, a geographically weighted regression model was applied to estimate global PM_{2.5} [25]. Donkelaar et al. [26] developed one geoscience-derived approach to estimate of PM_{2.5} composition over North America from 2000 to 2016. Hammer et al. estimated global PM_{2.5} concentrations and trends in the period of 1998–2018 by using satellite observations, chemical transport models, and ground-based monitoring [27].

In fact, the haze pollution process is closely related to the evolution of meteorological parameters [28]. First, PM concentrations are the main cause of reduced visibility with light extinction effects, which causes the haze phenomenon to arise. Visibility is a measure of the clearness of the atmosphere [29]. Regional haze reduces visibility by the presence of particles suspended in the atmosphere and is usually expressed in terms of light extinction or haze index. PM scatters and absorbs light substantially. It has been shown that particles of different size and chemical composition can affect visibility very differently. The fine sulfate and nitrate particles are the main contributors to haze, especially in the presence of water vapor. The contributions of elemental and organic carbon are also important in visibility degradation while most other particles are relatively less important. During PM episodes, the atmospheric visibility is less than 10 km due to the presence of PM in the atmosphere. Second, wind speed and wind direction of automatic weather stations (AWS) are generally used to assess how well mass transport is being characterized [30]. In general, the probability of PM episodes is low with high speed wind. Previous work has demonstrated that, if relative humidity (RH) is higher than 80%, the probability of PM episodes is also low. It is inferred that the high temperature will make the water of surface evaporate up into the air, which indirectly affects the

PM cycle. In addition, the visibility also decreases in the presence of rain. Thus, it is necessary to supplement the visibility observation using precipitation measurements, in order to decide whether PM episodes really happen depending on the visibility measurements.

As addressed above, since the meteorological parameters can be affected by PM episodes, it is possible and promising to measure the concentration of PM by studying the data from weather sensors, even though AWS were not originally designed for PM observation. There is a high density distribution network of AWS in China, and these AWSs work 24/7 under all weather conditions. Because of the absence of dense urban PM monitoring networks, values observed at a ‘central monitor’ were frequently considered to be representative for ambient pollutant levels within a metropolitan area. With recent growth of the high density network of AWS, we have an opportunity to measure PM more accurately based on data from AWS. However, little work has been done to calculate PM concentration by using meteorological parameters. Our previous study using hidden Markov models to quantify PM concentrations have yielded some encouraging results [31]. In this paper, we aim to use three machine learning methods, namely multivariate linear regression, multivariate nonlinear regression and neural network, to retrieve PM concentrations by learning from the data of multi-weather sensors.

2. Materials and Methods

2.1. Materials

Observations of PM_{2.5} and meteorological parameters were collected from January 2014 to June 2014 at the National Xiamen weather station, Fujian, China. The PM monitoring station was installed in the same standard observation site as the automatic multi-weather sensors. The meteorological parameters and PM_{2.5} data were collected at a frequency of once per hour and processed by using the world meteorological data quality control standard. The main meteorological parameters are visibility, wind direction, wind speed, temperature, relative humidity, atmospheric pressure, and hourly rainfall rate. Figure 1 shows variations of PM_{2.5} concentrations and meteorological parameters at the National Xiamen weather station during the coordinated observation period. For better visualization, we divided the PM_{2.5} data and meteorological parameters into two dimensions according to 24 h per day, as shown in Figure 2.

In order to study the relationship between PM_{2.5} and meteorological parameters, linear regressions were performed by using Pearson’s linear correlation. Pearson’s correlation coefficient is widely used to measure the degree of linear correlation between two quantitative variables [32]. Given N samples of two variables x and y , the coefficient r_{xy} is calculated as

$$r_{xy} = \frac{N \sum_{i=1}^N (x_i y_i) - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\sqrt{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \sqrt{N \sum_{i=1}^N y_i^2 - (\sum_{i=1}^N y_i)^2}} \quad (1)$$

where x_i and y_i are the i th sample points.

The results were reported in Table 1. PM_{2.5} has a high correlation with visibility, wind direction, wind speed, and relative humidity, and a low correlation with air temperature, atmospheric pressure, and rainfall rate. Therefore, we use the meteorological parameters with high correlation coefficients, namely visibility, wind direction, wind speed and relative humidity, in order to develop the multivariate regression model.

Interestingly, we also found that the performance of neural network based method can be improved by using these meteorological parameters, even though relative humidity, atmospheric pressure, and rainfall rate have low correlation with PM_{2.5}. Thus, all seven meteorological parameters were used in the neural network method.

Figure 1. This figure shows temporal variations of PM_{2.5} concentrations and meteorological parameters at the National Xiamen weather station during the period of January 2014–June 2014.

Figure 2. This figure shows two-dimensional temporal variations of PM_{2.5} concentrations and meteorological parameters by 24 h per day in the period of January 2014–June 2014.

Table 1. Pearson’s linear correlation coefficient between PM_{2.5} and meteorological parameters.

Visibility	Wind Direction	Wind Speed	Relative Humidity	Temperature	Atmospheric Pressure	Rainfall Rate
−0.5639	0.2830	−0.2839	0.1201	−0.0424	0.0828	−0.0308

2.2. Machine Learning Methods

2.2.1. Multivariate Linear Regression

Denote the i th observation of PM_{2.5}, visibility, wind direction, wind speed, and relative humidity as y_i , X_i^1 , X_i^2 , X_i^3 , and X_i^4 , respectively, we predict the PM_{2.5} via the model as

$$\hat{y} = a_0 + \sum_{k=1}^4 X_i^k a_k \quad (2)$$

where a_0 is the intercept. We include a_0 in the vector of coefficients $\mathbf{a}$, Equation (1) can be written in vector form as an inner product

$$\hat{y} = \mathbf{X}^T \mathbf{a} \quad (3)$$

The optimal vector of coefficients $\hat{\mathbf{a}}$ can be generated by minimizing the distance between the predictions and the ground truth data. Using Euclidean distance, the solution of the vector of coefficients can be formulated as

$$\hat{\mathbf{a}} = \arg \min_a \|\mathbf{y} - \mathbf{X}^T \mathbf{a}\|_2^2 \quad (4)$$

where $\mathbf{y}$ is a vector of PM_{2.5} data in the training set.

The least squares method was applied to fit the above model, and the solution is given by using the Moore–Penrose inverse operation as

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

2.2.2. Multivariate Nonlinear Regression

The physical principle behind our proposed nonlinear regression model is that the value of atmospheric optical visibility decays as an exponential function with increasing of PM concentrations. In addition, as the wind speed increases, it becomes easier for the particles matter to disperse and the concentration becomes smaller. Figure 3 shows this physical nonlinear relationship. The nonlinear regression model is given by

$$\hat{y} = \sum_{k=1}^4 \gamma_k \exp(-\beta_k X_i^k) \quad (6)$$

where γ_k and β_k are the k th coefficients of the model.

In this approach, we pick the coefficients γ and β to minimize the cost function as residual sum of squares as

$$f(\theta) = \sum_{i=1}^N \left(y_i - \sum_{k=1}^4 \gamma_k e^{(-\beta_k X_i^k)} \right)^2 \quad (7)$$

where N is total number of data samples, the parameters $\theta = (\gamma, \beta)$.

The above problem was solved by using the Nelder–Mead simplex method [33,34]. One vector of the parameters θ represents one simplex. The major procedures of the Nelder–Mead simplex method include order, reflection, expansion, contraction, and shrink operation. Algorithm 1 summarizes the detailed procedures for fitting the multivariate nonlinear regression model.

Figure 3. This figure shows nonlinear distribution of $PM_{2.5}$ concentrations with visibility and wind speed, respectively.

Algorithm 1: Nelder–Mead simplex method for multivariate nonlinear regression

Data: $PM_{2.5}$ data, visibility, wind direction and speed, and relative humidity data.

Initialization: set three vertices of simplex θ_1, θ_2 , and θ_3 .

- 1 **Order:** sort the three vertices of simplex $\theta_1, \theta_2, \theta_3$ by the simplexes's function values that satisfy $f(\theta_1) \leq f(\theta_2) \leq f(\theta_3)$.
 - 2 Compute the centroid of simplex $\theta_o = 0.5(\theta_1 + \theta_2)$.
 - 3 **Reflection:** calculate the reflected simplex as $\theta_r = \theta_o + (\theta_o - \theta_3)$.
 - 4 **if** $f(\theta_1) \leq f(\theta_r) < f(\theta_2)$ **then**
 - 5 $\theta_3 = \theta_r$;
 - 6 go to step 1.
 - 7 **else if** $f(\theta_r) < f(\theta_1)$ **then**
 - 8 **Expansion:**
 - 9 calculate the expanded simplex as $\theta_e = \theta_o + 2(\theta_r - \theta_o)$;
 - 10 **if** $f(\theta_e) < f(\theta_r)$ **then**
 - 11 $\theta_3 = \theta_e$;
 - 12 go to step 1.
 - 13 **else**
 - 14 $\theta_3 = \theta_r$;
 - 15 go to step 1.
 - 16 **else if** $f(\theta_r) \geq f(\theta_2)$ **then**
 - 17 **Contraction:**
 - 18 calculate the contracted simplex as $\theta_c = \theta_o + 0.5(\theta_3 - \theta_o)$;
 - 19 **if** $f(\theta_c) < f(\theta_3)$ **then**
 - 20 $\theta_3 = \theta_c$;
 - 21 go to step 1.
 - 22 **Shrink:**
 - 23 $\theta_2 = \theta_1 + 0.5(\theta_2 - \theta_1)$;
 - 24 $\theta_3 = \theta_1 + 0.5(\theta_3 - \theta_1)$;
 - 25 go to step 1.
-

2.2.3. Neural Network

Neural networks are particularly well suited to dealing with nonlinear fitting problems, due to the fact that enough elements (called neurons) can fit any data with arbitrary precision [35,36]. A multilayer perception (MLP) network [37] is applied to explore the nonlinear regression for PM_{2.5}. Figure 4 shows a conceptualized structure of a two-layer feed-forward network that is used for predicting PM concentrations.

Figure 4. This figure shows a conceptualized diagram of a two-layer feed-forward network that is used for predicting PM concentrations.

The proposed two-layer feed-forward neural network includes a sigmoid hidden layer and an affine transformation output layer. Assuming that the number of entries in the hidden layer is m_h , the input–output function can be formulated as

$$f_{\theta}(\mathbf{X}) = b + v \operatorname{sigmoid}(c + \mathbf{W}\mathbf{x}) \quad (8)$$

where $\mathbf{x}$ is defined above, b is the output offset, v is the weight vector for the output layer, $\operatorname{sigmoid}(x) = 1/(1 + e^{-x})$, c is the offset vector for the hidden layer, and $\mathbf{W}$ is the weight matrix for the hidden layer, and the parameters $\theta = (b, c, v, \mathbf{W})$. As mentioned above, all seven meteorological parameters were input in the neural network. Thus, the dimension of the parameters can be determined as $c \in \mathbb{R}^{m_h}$, $v \in \mathbb{R}^{m_h}$, $\mathbf{W} \in \mathbb{R}^{m_h \times 7}$.

To avoid over-fitting the data, we applied the Bayesian regularization method to train the network [38,39]. Firstly, the sum of squares of the vector weights is defined as

$$E_W = \alpha \left(\sum_{i=1}^{m_h} \sum_{k=1}^7 W_{ik}^2 + \sum_{i=1}^{m_h} v_i \right)^2 \quad (9)$$

where α is the parameters of the function.

In addition, the sum of squared errors is given

$$E_D = \beta \sum (\mathbf{y} - (b + v \operatorname{sigmoid}(c + \mathbf{W}\mathbf{x})))^2 \quad (10)$$

where β is the parameters of the error function.

From the perspective of Bayesian framework, the weights of neural network are considered as stochastic variables. According to Bayes' rule, the probability density function of the parameters for a neural network M can be formulated as

$$p(\theta | D, \alpha, \beta, M) = \frac{p(D | \theta, \alpha, \beta, M)p(\theta | \alpha, \beta, M)}{p(D | \alpha, \beta, M)} \quad (11)$$

Specifically, the likelihood function does not depend on the regularizer α once the parameters θ is known, and the prior function does not depend on the parameter β that regularizes the data term [38]. Therefore, the above equation can be simplified as

$$p(\theta | D, \alpha, \beta, M) = \frac{p(D | \theta, \beta, M)p(\theta | \alpha, M)}{p(D | \alpha, \beta, M)} \quad (12)$$

Assuming that the prior of the parameters θ and the training data are Gaussian distributed, the probability density functions can be represented

$$p(D | \theta, \beta, M) = \frac{\exp(-\beta E_D)}{(\pi/\beta)^{N_D/2}} \quad (13)$$

where N_D is the total number of training data samples.

In addition,

$$p(\theta | \alpha, M) = \frac{\exp(-\alpha E_W)}{(\pi/\alpha)^{N_W/2}} \quad (14)$$

where N_W is the number of the weights.

The optimal parameters of the neural network can be obtained by maximizing the posterior probability (Equation (12)). Substituting Equations (13) and (14) into Equation (12), we can obtain

$$p(\theta | D, \alpha, \beta, M) \propto \frac{1}{Z(\alpha, \beta)} \exp(-\alpha E_W - \beta E_D) \quad (15)$$

where $\frac{1}{Z(\alpha, \beta)}$ is the normalization factor. According to the above derivation, maximizing the above posterior probability is equivalent to minimized the regularized cost function as

$$J(\theta) = \alpha E_W + \beta E_D \quad (16)$$

Here, we use one approach of Gauss–Newton approximation for Bayesian regularization [39]. The more details of Bayesian regularization for neural network can also be found in [40–43].

3. Results

3.1. Models Training

After solving Equation (4) by least squares, the multivariate linear regression model was determined to be

$$\begin{aligned} \text{PM}_{2.5} = & 113.93 - 0.0024 \times \text{Visibility} + 0.051 \times \text{Wind direction} \\ & - 2.962 \times \text{Wind speed} - 0.5336 \times \text{RH} \end{aligned} \quad (17)$$

Next, we describe the training process of the multivariate nonlinear regression model. As shown in Figure 5, the cost function value decreased rapidly during the first 300 iterations by using Algorithm 1.

When the number of iterations reached 900, the performance of the algorithm approached saturation. Finally, the multivariate nonlinear regression model was completely fitted after 1081 iterations as

$$\begin{aligned} \text{PM}_{2.5} = & 68.7672e^{(-0.0001 \times \text{Visibility})} + 8.3784e^{(0.0029 \times \text{Wind direction})} \\ & - 0.0536e^{(0.7548 \times \text{Wind speed})} + 7.3861e^{(-0.256 \times \text{RH})} \end{aligned} \quad (18)$$

In the training of neural network models, validation is not required due to the use of the Bayesian regularization method. Therefore, the PM data and meteorological parameters were randomly divided into two sets as: 60% for training, and 40% for a completely independent test. The number of these datasets is 70% of all data, meaning that 42% of the data has been used for training. Validation is often considered as a form of regularization to meet the balance between under-fitting and over-fitting. Interestingly, the Bayesian regularization method has its own form of validation built into the approach [38,39], so this paradigm disables validation of the dataset, since the purpose of checking validation is to see if the error on the validation set gets better or worse as training progresses. The error of Bayesian regularization is based not only on how the model behaves on the dataset, but also on the size of the weights in the hidden layers. The larger the weights, the larger the error. Thus, throughout the training process, the hidden layer may never be allowed to explore larger weights, even if larger weights may result in a global minimum.

Figure 5. This figure shows variations of the cost function value during the training of the multiple nonlinear regression model, using Algorithm 1. The value of the cost function is calculated by using Equation (7).

We averaged the training and test performance over 100 experiments on models with the different numbers of hidden neurons, and the results are shown in Figure 6. These results support the issues mentioned above. While increasing the number of hidden neurons can improve the performance of training, it does not reduce the error for testing. Therefore, by considering the training and test performance together, we set the number of hidden neurons as 16, which can produce satisfactory performance.

Figure 6. This figure shows variations of the training performance (a) and the test performance (b) with respect to the number of hidden neurons.

3.2. Predictions of $\text{PM}_{2.5}$ Concentrations

Using these three trained models, the meteorological parameters were input to obtain the estimated PM concentrations, which were shown in Figures 7 and 8, respectively. From the plots, it can be seen that both linear and nonlinear multivariate regression based methods can estimate the slow part of the PM changes, but the details of the rapid changes cannot be estimated precisely. In contrast, the neural network-based method, with good nonlinear learning capability, captured the changes in PM concentration more completely and accurately.

To further the performance of the three machine learning algorithms, the Pearson's linear correlation was used for performing linear regressions between the output of models and ground truth data. The results were reported in Figure 9 and Table 2. The results showed that the multivariate linear regression method has the root mean square error (RMSE) of $24.6756 \mu\text{g}/\text{m}^3$ with a correlation coefficient of 0.6281, by referring to the ground truth of PM time series data; and the multivariate nonlinear regression method has the RMSE of $24.9191 \mu\text{g}/\text{m}^3$ with a correlation coefficient of 0.6184,

while the neural network based method has the best performance, of which the RMSE of PM estimates is $15.6391 \mu\text{g}/\text{m}^3$ with a better correlation coefficient of 0.8701.

Table 2. PM_{2.5} prediction performances of three different models.

	Linear Regression	Nonlinear Regression	Neural Network
Averaged RMSE ($\mu\text{g}/\text{m}^3$)	24.6756	24.9191	15.6391
Correlation coefficient	0.6281	0.6184	0.8701

(a) Linear regression

(b) Non-linear regression

(c) Neural network

Figure 7. This figure shows the PM_{2.5} concentrations estimated by using the different machine learning methods.

Figure 8. This figure shows comparisons of PM_{2.5} concentrations estimated using different machine learning methods.

Figure 9. Cont.

(b) Non-linear regression

(c) Neural network

Figure 9. Linear regression lines with red color between estimated $\text{PM}_{2.5}$ concentrations and the $\text{PM}_{2.5}$ observation (reference) data. The spaces between the two green lines are the 95% prediction interval.

4. Discussion

This paper attempts to estimate the concentration of $\text{PM}_{2.5}$ from meteorological parameters using three machine learning models to answer the question of whether it can be estimated and with what accuracy.

From Table 1, the correlation between $PM_{2.5}$ concentration and visibility is -0.5639 . The correlation coefficients between the estimated $PM_{2.5}$ value and the $PM_{2.5}$ reference value are 0.6281 and 0.6184 , respectively, after multiple linear and nonlinear regression models (see Table 2). This indicates that the accuracy of $PM_{2.5}$ estimation by these two models does not improve much. The main reason is that the nonlinear relationship between PM and meteorological parameters such as visibility, wind speed, wind direction, and humidity is complicated.

The estimation accuracy of the $PM_{2.5}$ is greatly improved by the neural network model, with a correlation coefficient of 0.8701 , which is better than our previous results by using the hidden Markov models [31]. The results demonstrate the ability of the neural network model to learn nonlinear relationships.

It is very interesting to note that, although the correlations between $PM_{2.5}$ and atmospheric pressure, rainfall rate, and temperature are very low (see Table 1), the use of these meteorological parameters is critical to the performance of the neural network. Therefore, we further conducted extensive experiments to investigate the effects of using different meteorological parameter inputs on the RMSE and correlation coefficients of the three machine learning model estimates.

As shown in Table 3, the performance of the neural network approach increases significantly with increasing meteorological parameters. The performance of the linear regression model increases slightly with increasing meteorological parameters; however, the performance of the nonlinear regression model decreases considerably.

Table 3. Effect of using different meteorological parameter inputs on the averaged RMSE ($\mu\text{g}/\text{m}^3$) and correlation coefficients of three machine learning model outputs.

Parameters	Linear Regression	Nonlinear Regression	Neural Network
visibility + wind + RH	24.6756/0.6281	24.9191/0.6184	20.4548/0.7643
visibility + wind + RH + temperature	24.6670/0.6284	27.5901/0.5107	17.2791/ 0.8387
visibility + wind + RH + temperature + air pressure	24.5770/0.6319	30.1810/0.3093	17.1101/ 0.8460
visibility + wind + RH + temperature + air pressure + rainfall	24.5141/0.6343	29.0009/ 0.4132	15.6391/ 0.8701

In machine learning, a very important issue is the problem of under-fitting and over-fitting of data. Unfortunately, the minimization of cost function for multivariate regression models suffers from poor generalization [40]. Therefore, the performance of the multivariate regression models for PM concentration prediction is limited. Our future work will introduce a regularization approach to further improve the performance of the multiple regression model. In contrast, Bayesian regularization has some promising advantages for the training of neural network models [40]. Bayesian regularization does not require a validation dataset; the method itself presents an evaluation of the evidence, which keeps the model from being overtrained. In addition, the Bayesian regularization method introduces Occam's razor that automatically penalizes overly complex models, so the model is difficult to be overfitted [38,39]. One limitation of this study is that the PM and meteorological data are six-month time series. Our future work will collect data for longer periods to further evaluate the performance of the model, and consider other types of neural network models to further improve the accuracy of PM concentration predictions.

5. Conclusions

This paper demonstrates the potential of using multi-weather sensors to monitor $PM_{2.5}$ concentrations. The accuracy of $PM_{2.5}$ concentrations has been studied by using a comparison of three classical machine learning methods. The results show that the neural network-based approach outperforms both multivariate linear and nonlinear regression approaches, with encouraging results, a root mean square error of $15.6391 \mu\text{g}/\text{m}^3$ and a correlation coefficient of 0.8701 . This study means that we can estimate the PM concentrations in real time from the high density network of automatic weather stations. Machine learning methods using data of these automated weather stations can provide new insights into mapping PM concentrations, and may make a valuable contribution to our understanding

of the particles distribution and its cycle. Our future work will include acquiring more data and using other types of neural network models to further improve the accuracy of PM predictions.

Author Contributions: Conceptualization, Z.X. and Y.W.; methodology, Z.X. and Y.W.; writing—editing and revisions, Z.X. and Y.W.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Grant No. 42005100 and 61801247, and the Natural Science Foundation of Jiangsu Province of China Grant No. BK20180945.

Acknowledgments: The authors would like to thank Xiamen weather station for providing the meteorological parameters and PM data. The authors are also deeply grateful to the editor and all three anonymous reviewers for their helpful comments and insightful suggestions that greatly improve the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mogireddy, K.; Devabhaktuni, V.; Kumar, A.; Aggarwal, P.; Bhattacharya, P. A new approach to simulate characterization of particulate matter employing support vector machines. *J. Hazard. Mater.* **2011**, *186*, 1254–1262. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Jo, H.Y.; Kim, C.H. Identification of long-range transported haze phenomena and their meteorological features over Northeast Asia. *J. Appl. Meteorol. Climatol.* **2013**, *52*, 1318–1328. [\[CrossRef\]](#)
3. Lee, B.K.; Park, G.H. Characteristics of heavy metals in airborne particulate matter on misty and clear days. *J. Hazard. Mater.* **2010**, *184*, 406–416. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Kadiyala, A.; Kumar, A. Development and application of a methodology to identify and rank the important factors affecting in-vehicle particulate matter. *J. Hazard. Mater.* **2012**, *213*, 140–146. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Araujo, L.N.; Belotti, J.T.; Alves, T.A.; Tadano, Y.D.S.; Siqueira, H.V. Ensemble method based on Artificial Neural Networks to estimate air pollution health risks. *Environ. Model. Softw.* **2020**, *123*, 104567. [\[CrossRef\]](#)
6. Tadano, Y.D.S.; Siqueira, H.V.; Alves, T.A. Unorganized machines to predict hospital admissions for respiratory diseases. In Proceedings of the 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Cartagena, Colombia, 2–4 November 2016; pp. 1–6.
7. Jerrett, M.; Turner, M.C.; Beckerman, B.S.; Pope, C.A., III; Van Donkelaar, A.; Martin, R.V.; Serre, M.; Crouse, D.; Gapstur, S.M.; Krewski, D.; et al. Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates. *Environ. Health Perspect.* **2017**, *125*, 552–559. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Harrison, R.M. Airborne particulate matter. *Philos. Trans. R. Soc. A* **2020**, *378*, 20190319. [\[CrossRef\]](#)
9. Yang, X.; Li, Z. Increases in thunderstorm activity and relationships with air pollution in southeast China. *J. Geophys. Res. Atmos.* **2014**, *119*, 1835–1844. [\[CrossRef\]](#)
10. Levy, R.C.; Pinker, R.T. Remote sensing of spectral aerosol properties: A classroom experience. *Bull. Am. Meteorol. Soc.* **2007**, *88*, 25–30. [\[CrossRef\]](#)
11. Delp, W.W.; Singer, B.C. Wildfire smoke adjustment factors for low-cost and professional PM_{2.5} monitors with optical sensors. *Sensors* **2020**, *20*, 3683. [\[CrossRef\]](#)
12. Franklin, M.; Kalashnikova, O.V.; Garay, M.J.; Fruin, S. Characterization of subgrid-scale variability in particulate matter with respect to satellite aerosol observations. *Remote Sens.* **2018**, *10*, 623. [\[CrossRef\]](#)
13. Shin, M.; Kang, Y.; Park, S.; Im, J.; Yoo, C.; Quackenbush, L.J. Estimating ground-level particulate matter concentrations using satellite-based data: A review. *GIScience Remote Sens.* **2020**, *57*, 174–189. [\[CrossRef\]](#)
14. Ma, X.; Huang, Z.; Qi, S.; Huang, J.; Zhang, S.; Dong, Q.; Wang, X. Ten-year global particulate mass concentration derived from space-borne CALIPSO lidar observations. *Sci. Total Environ.* **2020**, *721*, 137699. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Christopher, S.; Gupta, P. Global Distribution of Column Satellite Aerosol Optical Depth to Surface PM_{2.5} Relationships. *Remote Sens.* **2020**, *12*, 1985. [\[CrossRef\]](#)
16. Veefkind, J.; Aben, I.; McMullan, K.; Forster, H.; De Vries, J.; Otter, G.; Claas, J.; Eskes, H.; De Haan, J.; Kleipool, Q.; et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* **2012**, *120*, 70–83. [\[CrossRef\]](#)

17. Mei, H.; Han, P.; Wang, Y.; Zeng, N.; Liu, D.; Cai, Q.; Deng, Z.; Wang, Y.; Pan, Y.; Tang, X. Field evaluation of low-cost particulate matter sensors in Beijing. *Sensors* **2020**, *20*, 4381. [\[CrossRef\]](#)
18. Zheng, M.; Yan, C.; Zhu, T. Understanding sources of fine particulate matter in China. *Philos. Trans. R. Soc. A* **2020**, *378*, 20190325. [\[CrossRef\]](#)
19. Zhang, Y.; Guo, J.; Yang, Y.; Wang, Y.; Yim, S.H. Vertica Wind Shear Modulates Particulate Matter Pollutions: A Perspective from Radar Wind Profiler Observations in Beijing, China. *Remote Sens.* **2020**, *12*, 546. [\[CrossRef\]](#)
20. Knobelspiesse, K.; Barbosa, H.M.; Bradley, C.; Bruegge, C.; Cairns, B.; Chen, G.; Chowdhary, J.; Cook, A.; Di Noia, A.; van Diedenhoven, B.; et al. The Aerosol Characterization from Polarimeter and Lidar (ACEPOL) airborne field campaign. *Earth Syst. Sci. Data* **2020**, *12*, 2183–2208. [\[CrossRef\]](#)
21. Wang, T.; Han, W.; Zhang, M.; Yao, X.; Zhang, L.; Peng, X.; Li, C.; Dan, X. Unmanned Aerial Vehicle-Borne Sensor System for Atmosphere-Particulate-Matter Measurements: Design and Experiments. *Sensors* **2020**, *20*, 57. [\[CrossRef\]](#)
22. Mølgaard, B.; Hussein, T.; Corander, J.; Hämeri, K. Forecasting size-fractionated particle number concentrations in the urban atmosphere. *Atmos. Environ.* **2012**, *46*, 155–163. [\[CrossRef\]](#)
23. Reggente, M.; Peters, J.; Theunis, J.; Van Poppel, M.; Rademaker, M.; Kumar, P.; De Baets, B. Prediction of ultrafine particle number concentrations in urban environments by means of Gaussian process regression based on measurements of oxides of nitrogen. *Environ. Model. Softw.* **2014**, *61*, 135–150. [\[CrossRef\]](#)
24. Commodore, S.; Metcalf, A.; Post, C.; Watts, K.; Reynolds, S.; Pearce, J. A Statistical Calibration Framework for Improving Non-Reference Method Particulate Matter Reporting: A Focus on Community Air Monitoring Settings. *Atmosphere* **2020**, *11*, 807. [\[CrossRef\]](#)
25. Van Donkelaar, A.; Martin, R.V.; Brauer, M.; Hsu, N.C.; Kahn, R.A.; Levy, R.C.; Lyapustin, A.; Sayer, A.M.; Winker, D.M. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol.* **2016**, *50*, 3762–3772. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Van Donkelaar, A.; Martin, R.V.; Li, C.; Burnett, R.T. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol.* **2019**, *53*, 2595–2611. [\[CrossRef\]](#)
27. Hammer, M.S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A.M.; Hsu, N.C.; Levy, R.C.; Garay, M.; Kalashnikova, O.; Kahn, R.A.; et al. Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018). *Environ. Sci. Technol.* **2020**, *54*, 7879–7890. [\[CrossRef\]](#)
28. Dawson, J.P.; Bloomer, B.J.; Winner, D.A.; Weaver, C.P. Understanding the meteorological drivers of US particulate matter concentrations in a changing climate. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 521–532. [\[CrossRef\]](#)
29. Odman, M.T.; Hu, Y.; Unal, A.; Russell, A.G.; Boylan, J.W. Determining the sources of regional haze in the southeastern United States using the CMAQ model. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 1731–1743. [\[CrossRef\]](#)
30. Barker, H. Isolating the industrial contribution of PM_{2.5} in Hamilton and Burlington, Ontario. *J. Appl. Meteorol. Climatol.* **2013**, *52*, 660–667. [\[CrossRef\]](#)
31. Xu, M.; Wang, Y.X. Quantifying PM_{2.5} concentrations from multi-weather sensors using hidden Markov models. *IEEE Sens. J.* **2015**, *16*, 22–23. [\[CrossRef\]](#)
32. Patten, M.L.; Newhart, M. *Understanding Research Methods: An Overview of the Essentials*; Taylor & Francis: Abingdon, UK, 2017.
33. Powell, M.J. On search directions for minimization algorithms. *Math. Program.* **1973**, *4*, 193–201. [\[CrossRef\]](#)
34. Lagarias, J.C.; Reeds, J.A.; Wright, M.H.; Wright, P.E. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.* **1998**, *9*, 112–147. [\[CrossRef\]](#)
35. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.
36. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: <http://www.deeplearningbook.org> (accessed on 28 September 2020).
37. Haykin, S. *Neural Networks and Learning Machine*; Pearson Education, Inc.: Upper Saddle River, NJ, USA, 2009.
38. MacKay, D.J. Bayesian interpolation. *Neural Comput.* **1992**, *4*, 415–447. [\[CrossRef\]](#)
39. Foresee, F.D.; Hagan, M.T. Gauss–Newton approximation to Bayesian learning. In Proceedings of the IEEE International Conference on Neural Networks (ICNN'97), Houston, TX, USA, 12 June 1997; Volume 3, pp. 1930–1935.

40. Burden, F.; Winkler, D. Bayesian regularization of neural networks. In *Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 23–42.
41. Kayri, M. Predictive abilities of bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: A comparative empirical study on social data. *Math. Comput. Appl.* **2016**, *21*, 20. [[CrossRef](#)]
42. Park, J.G.; Jo, S. Approximate Bayesian MLP regularization for regression in the presence of noise. *Neural Netw.* **2016**, *83*, 75–85. [[CrossRef](#)]
43. Sariev, E.; Germano, G. Bayesian regularized artificial neural networks for the estimation of the probability of default. *Quant. Financ.* **2020**, *20*, 311–328. [[CrossRef](#)]

© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).