

# Article An Efficient Method for Capturing the High Peak Concentrations of PM<sub>2.5</sub> Using Gaussian-Filtered Deep Learning

Inchoon Yeo and Yunsoo Choi \*

Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX 77004, USA; iyeo@uh.edu \* Correspondence: ychoi23@central.uh.edu

Abstract: This paper proposes a deep learning model that integrates a convolutional neural network with a gate circulation unit that captures patterns of high-peak PM<sub>2.5</sub> concentrations. The purpose is to accurately predict high-peak PM2.5 concentration data that cannot be trained in general deep learning models. For the training of the proposed model, we used all available weather and air quality data for three years from 2015 to 2017 from 25 stations of the National Institute of Environmental Research (NIER) and the Korea Meteorological Administration (KMA) observatory in Seoul, South Korea. Our model trained three years of data and predicted high-peak PM2.5 concentrations for the year 2018. In addition, we propose a Gaussian filter algorithm as a preprocessing method for capturing high concentrations of PM2.5 in the Seoul area and predicting them more accurately. This model overcomes the limitations of conventional deep learning approaches that are unable to predict high peak PM2.5 concentrations. Comparing model measurements at each of the 25 monitoring sites in 2018, we found that the deep learning model with a Gaussian filter achieved an index of agreement of 0.73-0.89 and a proportion of correctness of 0.89–0.96, and compared to the conventional deep learning method (average POC = 0.85), the Gaussian filter algorithm (average POC = 0.94) improved the accuracy of high-concentration PM<sub>2.5</sub> prediction by an average of about 9%. Applying this algorithm in the preprocessing stage could be updated to predict the risk of high PM<sub>2.5</sub> concentrations in real time.

Keywords: air-quality; Gaussian filtering; deep learning; CNN; high peak forecasting of PM2.5

# 1. Introduction

The World Health Organization (WHO) guidelines for ensuring sufficient air quality recommend that concentrations of particulate matter with diameters of 2.5  $\mu$ m and smaller (PM<sub>2.5</sub>) and PM<sub>10</sub> should not exceed 24-h average concentration thresholds of 15  $\mu$ g/m<sup>3</sup> and 45  $\mu$ g/m<sup>3</sup>, respectively [1]. A close, quantitative relationship exists between exposure to high concentrations of small particulates (PM<sub>10</sub> and PM<sub>2.5</sub>), both daily and over time, and increased mortality or morbidity [1]. Actual recorded hourly peak concentrations frequently exceed thresholds recommended by the WHO. An average hourly PM concentration at or above the WHO threshold can contribute to adverse health effects [2–5].

To assess the acute health effects of ambient air pollution, most studies have used average daily concentrations as exposure variables. The average daily exposure of PM<sub>2.5</sub>, however, does not assume a constant level of exposure throughout the day, nor does it account for the difference between high and low levels of concentration. Therefore, examining the daily peak concentrations of air pollution, rather than the average daily concentrations of air pollution, may better reflect the biological mechanisms triggered by high-level exposure to pollutants [6]. Although the time-dependent nature of hourly concentrations, which show daily peak concentrations, may be a more accurate indicator of exposure than daily averages, few studies have examined the health effects stemming from exposure to hourly peak concentrations of air pollution. Existing studies, such as this one, indicate the importance of determining peak hourly concentrations of air pollution [7,8].

The geographical characteristics of East Asia, which contribute to an inflow of external air pollutants, significantly affect the overall concentrations of air pollutants in South



**Citation:** Yeo, I.; Choi, Y. An Efficient Method for Capturing the High Peak Concentrations of PM<sub>2.5</sub> Using Gaussian-Filtered Deep Learning. *Sustainability* **2021**, *13*, 11889. https:// doi.org/10.3390/su132111889

Academic Editors: Baihua Li and Fei Chao

Received: 12 August 2021 Accepted: 19 October 2021 Published: 27 October 2021

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



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Korea. Several studies have continuously examined the trends of air pollutant emissions in South Korea and foreign inflows, contributions by region and emission sectors, and the advancement of the emission calculation system [9–12]. From 2011 to 2014, when South Korea started calculating ultrafine dust emissions, domestic fine dust ( $PM_{10}$ ) and ultrafine dust ( $PM_{2.5}$ ) decreased overall. In 2015 and 2016, however, when scattering dust and biological combustion were included as an official source of emissions, the trend reversed, with significant increases in such emissions. Studies of air quality concentrations in South Korea have divided the sources of  $PM_{2.5}$  concentrations into those from abroad, such as long-distance pollutants, and those from domestic emissions. High concentrations of  $PM_{2.5}$  depend on the domestic geographic location, the economic and industrial structure, and the climatic factors of domestic emissions [13,14].

To address concerns about air quality and the risks posed by high emissions, air pollution researchers have begun to apply various machine learning and deep learning algorithms to various fields of atmospheric sciences to more accurately predict PM2.5 concentrations [7,15–17]; and several studies have improved numerical models [18]. Many approaches have applied deep learning algorithms and implemented convolutional neural network (CNN) models capable of learning joint features and classifiers. In addition to demonstrating a higher accuracy with large datasets, CNN models used for feature extraction are more efficient than other neural network models, particularly when multiple hidden layers are structured. These features have been responsible for significant advances in classification [19] and image processing in atmospheric sciences and other applications [20,21]. Long short-term memory (LSTM), an artificial recurrent neural network (RNN) that handles sequence modeling tasks, has shown great potential for executing various modeling tasks in the field of air pollution prediction [17,22]. To apply the LSTM method to the prediction of air pollution concentrations, Li et al. used LSTM with several machine learning techniques [22] and using previous air pollution concentrations; they also proposed an LSTM architecture that predicts PM<sub>2.5</sub> concentrations up to 24 h in advance.

Along with LSTM, a recently applied RNN architecture in the field of deep learning is the gated recurrent unit (GRU) [23]. The main difference between the GRU and LSTM architectures is that the GRU has no output gate. Therefore, as the GRU has fewer parameters than LSTM, its learning rate is faster than that of LSTM [24,25]. In fields such as air pollution concentration prediction, which uses time sequences, it is suitable to apply the classification or prediction of time series data using GRU networks among RNN architectures [7].

In this paper, to improve the accuracy of  $PM_{2.5}$  prediction, we used a Gaussian filter that removes noise from the original image in the image-processing research field [26]; we propose a method of improving the prediction accuracy of daily maximum concentrations of  $PM_{2.5}$ , which is currently a limitation of the already developed deep learning prediction system [7,18]. We propose a method of capturing maximum daily concentrations of  $PM_{2.5}$  using a Gaussian filter, which will improve the prediction accuracy of the captured maximum daily concentrations of  $PM_{2.5}$ .

#### 2. Observations

#### 2.1. Preparing Data Used for Predicting Daily Maximum Concentrations of PM<sub>2.5</sub>

To predict  $PM_{2.5}$  concentrations, we chose Seoul, South Korea because of its high population density and automobile exhaust gas, which continue to rapidly elevate concentrations of air pollutants in the city. The  $PM_{2.5}$  concentration prediction method in Seoul can be applied to New York, Beijing, and Tokyo, which are similar to or larger in scale than Seoul. The Seoul area is divided into 25 administrative districts with observation stations from the National Institute of Environmental Research (NIER). The 25 NIER stations measure concentrations of carbon monoxide (CO), sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), ozone (O<sub>3</sub>), and  $PM_{2.5}$  in real-time and provide an hourly average of each measurement. Air pollution data in Seoul are collected by Air Korea (http://www.airkorea.or.kr/web, accessed on 12 August 2021). This study used not only the air pollution data, but also weather conditions that affect concentrations of  $PM_{2.5}$ . We used time average wind speed (m/s), wind direction (0–360°), temperature (°C), relative humidity (%), dew point temperature (°C), surface pressure (hPa), and precipitation (mm) obtained from the Korean Meteorological Administration (KMA) and measurement data from 2015 to 2017 to train a model that predicts  $PM_{2.5}$  concentrations per hour in 2018. When we were missing measurement data, we applied the K-nearest neighbor (KNN) imputation method [27], which estimates the values of missing data using k samples of spatially close proximity in the given data and then imputes each missing datum by computing the mean of its k neighbors.

We defined the hourly peak concentration of  $PM_{2.5}$  as the maximum concentration of  $PM_{2.5}$  in 24 h on a given day. We also used data from these stations to calculate the daily average and daily maximum concentrations of  $PM_{2.5}$ ,  $NO_2$ ,  $SO_2$ , and  $O_3$ .

We obtained the daily average and maximum concentrations of  $PM_{2.5}$  for 25 atmospheric monitoring stations in Seoul, shown in Figure 1, each separated from the next by a distance between 1 km and 5 km. We collected air samples from remote stations far from high-traffic areas, industrial resources, buildings or residential sources of coal, waste, or petroleum combustion, which we considered representative of the air in the surrounding area in which the stations are located.



Figure 1. Average and maximum concentrations of PM<sub>2.5</sub> in the Seoul area (in the year 2018).

Figure 1 represents  $PM_{2.5}$  concentrations that exceed 76 µg/m<sup>3</sup> (red dot line) (the WHO standard high concentration) [1], which correspond to "very poor". Figure 1 displays a boxplot of 2018  $PM_{2.5}$  concentration data measured at the 25 stations in Seoul. The x-axis represents the names of the 25 stations, and the y-axis represents the measured  $PM_{2.5}$  concentration, we used 8760 datasets (24 h \* 365 days) of the 2018 data. The box of each station represents 8760 median values, upper quartile values, and lower quartile values. As shown in Figure 1, the stations with the highest daily maximum concentrations of  $PM_{2.5}$  in Seoul were #221, #231, #241, and #251.

As  $PM_{2.5}$  is small in quantity but constitutes a large specific surface area, it easily absorbs various heavy metals and harmful air pollutants. Therefore, when  $PM_{2.5}$  penetrates deep into the respiratory tract of humans, it attaches to the lung tissue and causes respiratory diseases. As heavy metals contained in  $PM_{2.5}$  are absorbed into blood vessels and cause strokes or cardiovascular disease, it is also associated with an increase in mortality [28]. Therefore, we required a more accurate method of predicting the maximum concentrations of  $PM_{2.5}$ , which could contribute to mitigating the adverse effects on human health.

# 2.2. Correlation between Maximum Daily Concentrations of PM<sub>2.5</sub> and Traffic Volume, the Factory Area, and Population Density

We began this study by investigating the causes of high peaks of PM<sub>2.5</sub>. A majority of studies have identified sources of high-concentration fine dust in Korea as those emitted from abroad, that is, long-distance pollutants and domestic emissions. We found that concentration levels depend on climatic factors [13,14]. For example, fine dust and ultrafine dust originating from outside of Korea mainly flow into the country during the spring and the winter, when heating for cold weathers and northwesterly winds are frequent [29,30]. Although the domestic contribution of foreign fine dust varies depending on the season or air quality modeling conditions, it has been reported to be approximately 30% or more [10,30].

According to several studies, the effects of fine dust from domestic emissions in South Korea vary by sources, such as thermal power plants, industrial combustion, and road/non-road pollution sources [31–33]. This paper investigates the correlation between the traffic volume, the number of factories, and the population density in the Seoul area as factors affecting daily maximum concentrations of  $PM_{2.5}$  in the Seoul metropolitan area.

As shown in Figure 2a, the area with the largest population density is #273 in southeastern Seoul. Not surprisingly, the daily maximum concentrations of PM<sub>2.5</sub> in #273 are higher than those in other regions; this area, however, does not have the highest daily maximum concentrations. The area is a representative suburb/residential community in Seoul, and although, administratively, it is home to a large number of residents, most of the residents commute to other areas. They travel to other areas between 9 a.m. and 7 p.m. and typically return after 7 p.m. In addition, the population density in areas #221, #231, #241, and #251, where the maximum daily concentrations of  $PM_{2.5}$  are high, is relatively low compared to that in the other areas. As shown in Figure 2b, the area with the largest daily average traffic volume is southeastern Seoul, and the station with the highest traffic volume is #261 (as shown in Figure 1), where the daily maximum concentrations of PM<sub>2.5</sub> are higher than in other areas but not the highest. In addition, the population density in areas #221, #231, #241, and #251, where the maximum daily concentrations of PM<sub>2.5</sub> are high, is relatively low compared to that in other areas. Figure 2c shows the distribution of factories in Seoul. Areas with a large number of factories are #221, #231, and 281, and as shown in Figure 1, these areas, compared to others in Seoul, have relatively high daily maximum concentrations of  $PM_{2.5}$ . They are located in the center of Seoul, home to a concentration of industries. Figure 2c, however, displays only factory locations in the Seoul metropolitan area, not factory locations outside of Seoul. As shown in Figure 2d, a major factory area outside of Seoul is concentrated adjacent to the sea, southwest of Seoul, thus calling for an investigation of air pollutants in factory areas outside of Seoul. Since this study used only air quality data for the Seoul metropolitan area instead of analyzing direct air quality data from outside of Seoul, we investigated data on wind directions at five stations (from among the 25) in the areas that registered the highest concentrations of daily PM<sub>2.5</sub> and the five that recorded the lowest.

In our predictions of the daily maximum concentrations of  $PM_{2.5}$  at the 25 stations in Seoul, we included wind direction information at each station (Figure 3). Figure 3a shows five stations with the lowest daily maximum concentrations of  $PM_{2.5}$  in Seoul. Most wind directions are easterly or westerly. Figure 3b, however, shows that most winds at the five stations with the highest daily maximum concentrations of  $PM_{2.5}$  in Seoul are southerly and southwesterly. These observations show that air pollutants from factory areas, concentrated in the southwestern part of Seoul, influence the maximum daily concentration of  $PM_{2.5}$ in Seoul.



(a) Distribution of population density in Seoul area





(b) Distribution of daily traffic volume in Seoul area



(c) Distribution of factory locations in Seoul area

(d) Location and size of factory area outside Seoul area

**Figure 2.** (a) Population density, (b) traffic volume, (c) distribution of factory areas in Seoul, and (d) the location and size of factory zones outside of Seoul.



concentrations of daily PM<sub>2.5</sub>

(b) Wind direction of the five stations with the highest concentrations of daily  $PM_{2.5}$ 

**Figure 3.** (a) The five stations with the highest daily maximum concentrations of  $PM_{2.5}$  and (b) the five stations with the lowest daily maximum concentrations of  $PM_{2.5}$  in the Seoul area.

#### 3. Materials and Methods

In general, when predicting future data using deep learning approaches such as DNN, CNN, and LSTM, researchers remove outliers, such as maximum and minimum values, during the learning process because a feature of deep learning enables it to learn the general trend of data, not specific data such as outliers [19]. In this paper, we used a deep learning method to predict the maximum concentrations of PM<sub>2.5</sub>, but after extracting only outliers, not general data, we identified the outlier values and predicted the maximum concentrations of PM<sub>2.5</sub> from the predicted outlier values.

#### 3.1. Outlier Extraction of PM<sub>2.5</sub> Data Using a Gaussian Filter

The Gaussian filter, a widely used tool in the image processing field, converts a smooth image by removing noise from the original image via alleviating the color change of abrupt pixels [34]. In general, pixels in an image have a greater weight the closer

they are to the current pixel and less weight the farther they are from the current pixel. As the Gaussian filter spatially changes an image smoothly, nearby pixels have similar values. Therefore, since the noise value has relatively little correlation with the values of neighboring pixels, it can be mitigated by a weighted average of neighboring pixel values. Using this characteristic of the Gaussian filter, it is possible to capture the pattern of daily maximum concentrations of  $PM_{2.5}$ .

The idea of the Gaussian filter algorithm applied in this study was to use the distribution of  $PM_{2.5}$  data as a point-spread function, carried out by convolution [26]. Before we performed convolution, however, we needed to make individual approximations to Gaussian functions. In the theory of Gaussian smoothing, the Gaussian distribution is non-zero everywhere, requiring an infinitely large convolution kernel; in practice, nevertheless, it roughly equals zero greater than 3 standard deviations from the mean, and we could truncate the kernel at this point. Since we used one-dimensional data, referred to as  $PM_{2.5}$  data, in this study, the Gaussian distribution was as follows:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$
(1)

where  $\sigma$  is the standard deviation of the distribution.

Figure 4 shows the convolution kernel of Gaussian approximation using a Gaussian filter, defined as an appropriate value for the  $PM_{2.5}$  data. If we applied the Gaussian filter, shown in Figure 4b, to the  $PM_{2.5}$  data in Figure 4a, we obtained the result shown in Figure 4c. By subtracting the obtained data from the original  $PM_{2.5}$  data, we obtained the pattern for the high peak of  $PM_{2.5}$  shown in Figure 4d, which displays the result of a comparison between the original  $PM_{2.5}$  data and the outlier high-peak  $PM_{2.5}$  data through a Gaussian filter.



**Figure 4.** (a) Observed PM2.5 data, (b) Gaussian filter, (c) Convoluted a Gaussian filter, and (d) Extracted the high-peak data.

Figure 5 shows the results of a comparison between the high peak data of  $PM_{2.5}$  and the original  $PM_{2.5}$  by applying a Gaussian filter to the daily  $PM_{2.5}$  data in Seoul. Figure 5 shows high accuracy for high peak data of 76 µg/m<sup>3</sup> or more, which satisfies the WHO standard guideline. Using the Gaussian filter, we preprocessed data that captures the high-peak pattern of daily maximum concentrations of  $PM_{2.5}$  for deep learning.



Figure 5. PM<sub>2.5</sub> data of the Seoul area and high-peak PM<sub>2.5</sub> data.

#### 3.2. Deep Learning Architecture

This study proposes an architecture that captures the characteristics of input data using a CNN deep learning model and trains time-sequence characteristics for the data trained by CNN using a GRU deep learning model. In order to capture high-peak  $PM_{2.5}$  data, we used a Gaussian filter as input data to a neural network along with weather data and air pollutants.

In the CNN model used in this paper, each layer extracted features of the input received from the previous layer and output them as the input of the next layer. Thus, neurons in each layer could extract features from inputs and then combine them with features in subsequent layers. The combined output became the convolution results, to which the activation function was applied. The convolution results had the same features as other instances of the time series, and multiple features could be extracted from each instance [35]. This study, however, did not use a pooling layer that uses local averaging to reduce the output sensitivity.

The PM<sub>2.5</sub> concentration prediction method using the conventional CNN or CNN/GRU removes information about peak concentrations, a phenomenon essential for predicting PM<sub>2.5</sub> concentrations [7]. The PM<sub>2.5</sub> data used in this study, however, were not the typical data of PM<sub>2.5</sub> per hour gathered for 24 h a day, but the results of preprocessing the maximum concentration data of PM<sub>2.5</sub> using a Gaussian filter. Therefore, it solves the problem of the existing CNN or CNN/GRU algorithm, from which the data of maximum concentrations were removed. As the kernel size (convolution window) that we used in the CNN model was  $2 \times 1$ , the convolution of the two-hour continuous input functions

occurred in the first layer. Then, we applied the results of the convolution operation to the activation function, which, in this paper, was ReLU. The CNN/GRU deep learning architecture proposed here connected the final output part of the CNN model to the input of the GRU model. If the deep learning architecture was configured in this way, the GRU model could use the features trained by the CNN. Because CNN processes all inputs and outputs independently, they may not sufficiently generate accurate predictions when previous information, such as time series, is required. Therefore, RNN models combined with CNN models allowed us to use previous information as an input, resulting in more accurate predictions [7,25,36]. RNN, however, cannot reset internal state information that affects subsequent observations. Therefore, to solve this shortcoming of RNN, we used GRU [7,23].

The last layer of the proposed CNN/GRU deep learning architecture was the output layer, and the predicted values of high-peak  $PM_{2.5}$  concentrations became the final output.

Figure 6 illustrates the structure of a neuron (expressed as  $x_1, x_2, ..., x_n$ ) with n inputs. As  $W_n$  represented the weight of each neuron connected to the next layer,  $W_i$  was the weight of the *ith* neuron. Computation of neuron  $x_n$  used the  $W_n$  of the weight component, as shown in Figure 6, and when the computed value reached a threshold, it was also converted to a specific value using an activation function. In Figure 6, the activation function was denoted as f(x). To maintain nonlinearity with respect to the computational results of input features and weights in a neural network, it must have an activation function. In our proposed CNN/GRU model, we used ReLU, defined by Equation (2).



$$f(x) = \max(0, x) \tag{2}$$

Figure 6. Operation of neurons in one convolution layer in the CNN/GRU model.

We used the results of applying the convolution operation and activation function of the first layer of CNN as input to the second layer. In the same way, we used the results of the calculations for the second layer as input to the third layer. As a result, our proposed CNN/GRU model had a four-layer CNN, each layer having 32 filters (activated by ReLU). Limiting the number of layers to four would neither increase the training time nor lead to overfitting. Conversely, using fewer than four layers would reduce accuracy because it limits feature extraction. Therefore, we found four CNN layers to be optimal. We used the features map obtained via the four CNN layers as input to the three-layer GRU layer, along with 64 filters, as shown in Figure 7. Once the operation of the last GRU layer ended, a high-peak PM<sub>2.5</sub> concentration prediction value was obtained through a fully connected hidden layer model with 256 nodes. We used the Keras library and the TensorFlow backend to implement the development environment [37].



**Figure 7.** Schematic of the structure of the CNN/GRU deep learning architecture used to predict high-peak PM<sub>2.5</sub> concentrations.

### 3.3. Model Training and Prediction

The CNN/GRU model proposed here needed to convert both air pollutant and meteorological data obtained from the KMA station into a format that the CNN/GRU model could understand. Table 1 lists air pollutant and meteorological data from 00:00 on 1 January 2015 to 23:00 on 31 December 2017.

Table 1. Input feature description: a sample of the organization of all inputs, column-wise.

	Wind Speed (miles/hour)	Wind Direction (Degrees Compass)	Temperature	Relative Humidity (percent)	 NO <sub>x</sub> (ppb)	Preprocessed PM <sub>2.5</sub> Using Gaussian Filter
1 January 2015 0:00						
1 January 2015 1:00						
:						
31 December 2017 23:00						

As observed data often do not exist for various reasons (e.g., a power outage or system error), the next step was to impute these missing values. If values were left without replacing them, properly training the neural network model would have been virtually impossible. Thus, after replacing nonexistent values using the KNN imputation method, we normalized each input value.

We sorted the data by input and output functions, as shown in Figure 8. Each row contained a day and each column the time of day. For example, the first 24 columns had a 24-h wind speed, and the next 24 columns had a 24-h wind direction. The current 24-h observation data constituted the input function. In addition, the high peak data of the  $PM_{2.5}$  preprocessed with the Gaussian filter was included. The maximum power function for that day had the maximum concentration value for  $PM_{2.5}$  on the next day. Once the model was trained, the maximum of  $PM_{2.5}$  on the next day (2 January 2018) was the predicted high peak value of  $PM_{2.5}$ .

	Input Feature (normalized)												output target		
	Wind speed		Wind direction							PM <sub>2.5</sub> using gaussi an filtered		Predicted high-pe ak PM <sub>2.5</sub>			
Hours	0	1		23	0	1		23	0	1		23			
02-Jan-2018		01-Jan-2018									02-Jan-2018				
03-Jan-2018															
•															
•															
n <sub>th</sub> day	(n-1) <sub>th</sub> day									n <sub>th</sub> day					

**Figure 8.** Training set for the deep learning model: An example of how the input and output features are arranged to train the model.

To predict high-peak PM<sub>2.5</sub> concentrations on a particular day, we needed a training set of input and output functions from the previous day. For example, to predict the high-peak PM<sub>2.5</sub> concentration on an *nth* day, we trained the input data of the  $(n-1)_{th}$  day. In addition, to predict the high peak PM<sub>2.5</sub> concentration for the entire year of 2018, we used SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>, and pretreated PM<sub>2.5</sub> for three years from 2015 to 2017; and to model the time series characteristics of high-peak PM<sub>2.5</sub> concentrations, we used observations of SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>, and PM<sub>2.5</sub> during each hour (recorded by NIER). We also used several meteorological variables that impact air pollutant concentrations, such as wind speed, humidity, and wind direction. In this study, after training three years of data from 2015 to 2017, high-peak PM<sub>2.5</sub> concentrations for the next 24 h were predicted for each date in 2018.

#### 4. Results and Discussion

We evaluated the performance of the model using a Gaussian filter to predict the maximum concentrations of  $PM_{2.5}$  and trained the model from 2015 to 2017 for 25 stations located in the Seoul area. While existing studies predicting concentrations of  $PM_{2.5}$  have used  $PM_{2.5}$  concentration data as one of the input data [16,17,38], this study, however, after preprocessing  $PM_{2.5}$  using a Gaussian filter to predict daily maximum concentrations of  $PM_{2.5}$ , used high-peak data of captured  $PM_{2.5}$  concentrations as one of the input data. The model then generated predictions for maximum  $PM_{2.5}$  concentrations for the entire year of 2018 and compared them to field measurements. To evaluate the hourly performance of the model, we used the index of agreement (IOA), and to evaluate the daily performance based on the daily maximum  $PM_{2.5}$ , we used categorical statistics [39,40].

Figure 9 shows the IOA for the hourly prediction of  $PM_{2.5}$  concentrations generated by the three deep learning models we tested. The blue line in the figure represents the  $PM_{2.5}$  data from the 25 stations in Seoul using the CNN model. In most cases, the average IOA of each station did not exceed 0.8. The orange dotted line shows the results of the CNN/GRU model, which could train the time-dependent information of  $PM_{2.5}$ . The figure shows that the average IOA of each station was higher than that of the CNN, but in most cases, it did not exceed 0.8. The green line shows that at most stations, the maximum daily concentration value of  $PM_{2.5}$  was preprocessed with a Gaussian filter and then trained with the CNN/GRU. The figure also shows the performance of the model as a result of training by capturing the high peak of  $PM_{2.5}$ . Thus, to accurately predict high-peak  $PM_{2.5}$ concentrations, we recommend using the GRU, which trains time information along with the CNN. In addition, high peaks of  $PM_{2.5}$  could not be accurately predicted with the daily data of  $PM_{2.5}$  alone. Therefore, to preprocess a maximum concentration pattern, using a Gaussian filter is an effective approach to learning data.



**Figure 9.** Results of the IOA for the predictions of high-peak PM<sub>2.5</sub> concentrations generated by the three deep learning approaches.

To compare the prediction results of high-peak  $PM_{2.5}$  concentrations with and without a Gaussian filter in more detail, we selected the four stations with the largest high-peak  $PM_{2.5}$  concentrations (#221, #231, #241, #251) from the 25 Seoul observation stations.

Figure 10 displays a graph of the predicted high-peak concentration values of  $PM_{2.5}$  in 2018 using the CNN model without the Gaussian filter for the four selected stations. We used CNN, the most common model used to predict  $PM_{2.5}$  concentrations, to predict high-peak concentration values of  $PM_{2.5}$ . As  $PM_{2.5}$  concentrations rapidly increased, however, the prediction accuracy was low for the high-peak concentrations.

Figure 11 shows the results of predictions using the CNN/GRU model without Gaussian filtering for high-peak  $PM_{2.5}$  concentration data exceeding the threshold of 76 µg/m<sup>3</sup>, recommended by the WHO, from the selected four stations in the Seoul area. Although these results were slightly more accurate than those presented in Figure 10, their accuracy was not high. For example, for station #231, the results of predicting high-peak  $PM_{2.5}$  concentrations using the CNN model without a Gaussian filter was 0.67, and the result using the CNN/GRU model without a Gaussian filter was 0.71. In contrast, the CNN/GRU model using a Gaussian filter was 0.82. Both CNN and CNN/GRU models produced poor results for the predictions of high-peak  $PM_{2.5}$  concentrations above a threshold of 76 µg/m<sup>3</sup>.

In this paper, we proposed a method of using outliers as input data to overcome the problem of generalization, a characteristic of deep learning. In order to use outliers as input data, we preprocessed the data using a Gaussian filter to extract only outliers, the maximum concentration values per day, from the general  $PM_{2.5}$  data. Using the preprocessed high-peak  $PM_{2.5}$  data as the input, we trained the data by applying a CNN/GRU deep learning model that produced slightly more accurate predictions than the CNN model. The graph in Figure 12 displays the values predicted by CNN/GRU after training on high-peak  $PM_{2.5}$  concentration data preprocessed with a Gaussian filter.

The results in Figure 12 exhibit a high prediction accuracy for high-peak PM<sub>2.5</sub> concentration data above the threshold of 76  $\mu$ g/m<sup>3</sup>, recommended by the WHO. Even though the same CNN/GRU method was used, the predictions of high-peak PM<sub>2.5</sub> concentrations using a Gaussian filter, compared with the predictions in Figure 11, were more accurate than they were using CNN and CNN/GRU. Table 2 shows the IOA, the *r*-value, and the mean absolute error (MAE) of the high-peak PM<sub>2.5</sub> concentration predictions for each of the 25 stations in Seoul generated by the CNN, the CNN/GRU, and the CNN/GRU models using a Gaussian filter.



(a) Observations and predictions of daily maximum concentration of  $PM_{2.5}$  for #221 station in 2018



(b) Observations and predictions of daily maximum concentration of  $PM_{2.5}$  for #231 station in 2018



(c) Observations and predictions of daily maximum concentration of  $PM_{2.5}$  for #241 station in 2018



(d) Observations and predictions of daily maximum concentration of  $PM_{2.5}$  for #251 station in 2018

**Figure 10.** Observations and predictions for daily maximum concentrations of PM<sub>2.5</sub> in CNN model for the year 2018.



(a) Observations and predictions of daily maximum concentration of  $PM_{2.5}$  for #221 station in 2018



(b) Observations and predictions of daily maximum concentration of  $PM_{2.5}$  for #231 station in 2018



(c) Observations and predictions of daily maximum concentration of PM2.5 for #241 station in 2018



(d) Observations and predictions of daily maximum concentration of  $PM_{2.5}$  for #251 station in 2018

Figure 11. Observations and predictions for daily maximum concentrations of  $PM_{2.5}$  in CNN/GRU model for the year 2018.



(a) Observations and predictions based on a Gaussian filter of daily maximum concentration of PM2.5 for #221 stations in 2018



(b) Observations and predictions based on a Gaussian filter of daily maximum concentration of PM<sub>2.5</sub> for #231 stations in 2018



(c) Observations and predictions based on a Gaussian filter of daily maximum concentration of PM2.5 for #241 stations in 2018



(d) Observations and predictions based on a Gaussian filter of daily maximum concentration of PM<sub>2.5</sub> for #251 stations in 2018

**Figure 12.** Observations and predictions for daily maximum concentrations of  $PM_{2.5}$  in CNN/GRU model with a Gaussian filter for the year 2018.

Station #		CNN		C	CNN/GR	U	CNN/GRU w/Gaussian Filters					
	IOA	r	MAE	IOA	r	MAE	IOA	r	MAE			
121	0.68	0.55	0.11	0.75	0.64	0.09	0.87	0.75	0.08			
123	0.62	0.44	0.11	0.76	0.67	0.09	0.83	0.71	0.08			
131	0.66	0.55	0.08	0.71	0.59	0.08	0.78	0.74	0.06			
141	0.64	0.45	0.12	0.71	0.52	0.12	0.79	0.68	0.09			
142	0.61	0.46	0.09	0.70	0.62	0.08	0.76	0.65	0.08			
151	0.80	0.68	0.09	0.79	0.67	0.09	0.89	0.83	0.07			
152	0.75	0.63	0.09	0.70	0.63	0.08	0.78	0.68	0.08			
161	0.71	0.58	0.09	0.78	0.67	0.08	0.81	0.75	0.07			
171	0.70	0.52	0.11	0.77	0.65	0.09	0.81	0.70	0.09			
181	0.71	0.58	0.09	0.77	0.72	0.08	0.83	0.76	0.07			
191	0.71	0.62	0.09	0.78	0.70	0.08	0.77	0.73	0.07			
201	0.69	0.54	0.09	0.81	0.69	0.08	0.83	0.71	0.08			
212	0.75	0.59	0.11	0.77	0.65	0.09	0.80	0.64	0.09			
221	0.65	0.48	0.11	0.67	0.56	0.10	0.74	0.62	0.09			
231	0.67	0.49	0.12	0.71	0.63	0.10	0.82	0.74	0.09			
241	0.67	0.55	0.10	0.74	0.66	0.09	0.79	0.71	0.08			
251	0.69	0.51	0.11	0.78	0.68	0.09	0.77	0.71	0.08			
261	0.70	0.55	0.10	0.71	0.64	0.09	0.86	0.79	0.08			
262	0.70	0.61	0.09	0.75	0.59	0.09	0.80	0.66	0.09			
273	0.63	0.57	0.10	0.74	0.69	0.08	0.74	0.65	0.08			
274	0.72	0.57	0.10	0.74	0.65	0.09	0.88	0.78	0.07			
281	0.67	0.52	0.11	0.75	0.62	0.10	0.89	0.81	0.07			
291	0.60	0.44	0.10	0.74	0.58	0.09	0.86	0.75	0.07			
301	0.67	0.53	0.11	0.76	0.63	0.09	0.73	0.66	0.09			
311	0.66	0.49	0.10	0.74	0.58	0.10	0.80	0.70	0.07			

**Table 2.** IOA, *r*-value, and mean absolute error (MAE) of high-peak PM<sub>2.5</sub> concentration predictions for the 25 stations in Seoul generated by the CNN, CNN/GRU, and CNN/GRU using a Gaussian filter.

We also evaluated the performance of the model by daily maximum values. To do so, we used categorical statistics [39,40] and divided pairs of observations and predictions as follows:

- 1. *N<sub>a</sub>*, number of days when an observation was below the threshold and a prediction was above.
- 2.  $N_b$ , number of days when both observations and predictions were above the threshold.
- 3.  $N_c$ , number of days when both observations and predictions were below the threshold.
- 4. *N<sub>a</sub>*, number of days when an observation was above the threshold and a prediction was below.

After categorizing the observations and predictions, we defined the following metrics based on the following: the hit rate (HIT), which represented the capability of the model to correctly forecast an extreme event (i.e., an event that exceeded the threshold of 76  $\mu$ g/m<sup>3</sup>); the false alarm rate (FAR), which represented times when the model falsely forecasted an extreme event; the equitable threat score (ETS), which defined the skill of the model on a scale of -1 to 1, with 1 indicating that the model was skillful; and the proportion of

correctness (POC), which defined the times the model was able to correctly predict the occurrence of an event (both exceedances and non-exceedances).

$$HIT = \frac{N_b}{N_b + N_d},\tag{3}$$

$$FAR = \frac{N_a}{N_a + N_b},\tag{4}$$

$$ETS = \frac{N_b - N_r}{N_a + N_b + N_d - N_r},\tag{5}$$

where

$$N_r = \frac{(N_a + N_b) \times (N_b + N_d)}{N_a + N_b + N_c + N_d}$$
(6)

$$POC = \frac{N_b + N_c}{N_a + N_b + N_c + N_d} \tag{7}$$

High-peak PM<sub>2.5</sub> concentrations were predicted by the deep learning model trained by the CNN/GRU architecture. As input data for the model, the high-peak PM<sub>2.5</sub> concentrations were preprocessed using a Gaussian filter. Figure 13 presents the POC, the ratio of the accuracy predictions of the occurrence of an event exceeding the threshold of 76  $\mu$ g/m<sup>3</sup> for each of the 25 stations in Seoul. In most cases, the accuracy of the POC values exceeded 0.93, and stations #221, #231, #241, and #251, with low IOAs (predictive accuracy) with high-peak PM<sub>2.5</sub> concentrations, also showed POC values close to 0.95.



Figure 13. POC for CNN/GRU using a Gaussian filter for the 25 stations in Seoul.

#### 5. Conclusions

To improve the prediction accuracy of high-peak PM<sub>2.5</sub> concentration data, this paper proposed a method of preprocessing PM<sub>2.5</sub> concentration data using a Gaussian filter and used the high-peak concentration as input data for a deep learning model. Using a Gaussian filter to preprocess the PM<sub>2.5</sub> concentration data, we were able to overcome the limitations of existing deep learning models, which consider high-peak PM<sub>2.5</sub> concentration data as outliers and exclude them from training. To train the data, we applied a CNN/GRU deep learning model with a Gaussian filter, which produced slightly more accurate results than a CNN model without a Gaussian filter or a CNN/GRU model without a Gaussian filter. After we applied the CNN/GRU to train the high-peak PM<sub>2.5</sub> concentration data preprocessed with a Gaussian filter, improvement in the accuracy of predictions approached 0.93. We used CNN to extract features of the data and GRUs to track changes in predicted values resulting from temporal changes. To predict high-peak PM<sub>2.5</sub> concentrations in Seoul, South Korea, we used three years of weather observations and chemical variables from 2015 to 2017 to train a model to predict maximum PM<sub>2.5</sub> concentrations for the next 24 h. After evaluating the full 2018 model, we found that

the deep learning method predicted maximum concentrations with sufficient accuracy (IOA = 0.73–0.89, POC = 0.89–0.96) by modeling the relationship between local weather and species concentrations. Neither the CNN nor the CNN/GRU-based prediction system predicted high-peak PM<sub>2.5</sub> concentrations as accurately as the prediction system using a Gaussian filter. Thus, the latter could also be used as an effective implementation tool in field measurement locations. The ability to predict high concentrations of PM<sub>2.5</sub>. Therefore, if the method of PM<sub>2.5</sub> concentration data preprocessed with a Gaussian filter is integrated as input into the PM<sub>2.5</sub> concentration prediction model, we could extend the function of the prediction system and achieve the accurate prediction of high-concentration PM<sub>2.5</sub> cases.

**Author Contributions:** Conceptualization, I.Y. and Y.C.; methodology, I.Y.; software, I.Y.; validation, I.Y. and Y.C.; formal analysis, I.Y.; investigation, I.Y.; resources, I.Y.; data curation, I.Y.; writing—original draft preparation, I.Y.; writing—review and editing, Y.C.; visualization, I.Y.; supervision, Y.C.; project administration, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the High Priority Area Research Seed Grant of the University of Houston.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

**Acknowledgments:** This study was supported by the High Priority Area Research Seed Grant of the University of Houston.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

# References

- 1. WHO. Ambient (Outdoor) Air Pollution. 2018. Available online: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (accessed on 22 September 2021).
- 2. Brunekreef, B.; Holgate, S.T. Air pollution and health. *Lancet* 2002, 360, 1233–1242. [CrossRef]
- 3. Delfino, R.J. Epidemiologic evidence for asthma and exposure to air toxics: Linkages between occupational, indoor, and community air pollution research. *Environ. Health Perspect.* **2002**, *110*, 573–589. [CrossRef] [PubMed]
- 4. Lanki, T.; Hoek, G.; Timonen, K.L.; Peters, A.; Tiittanen, P.; Vanninen, E.; Pekkanen, J. Hourly variation in fine particle exposure is associated with transiently increased risk of ST segment depression. *Occup. Environ. Med.* **2008**, *65*, 782–786. [CrossRef]
- Lin, H.; Liu, T.; Xiao, J.; Zeng, W.; Guo, L.; Li, X.; Xu, Y.; Zhang, Y.; Chang, J.J.; Vaughn, M.G.; et al. Hourly peak PM<sub>2.5</sub> concentration associated with increased cardiovascular mortality in Guangzhou, China. *J. Expo. Sci. Environ. Epidemiol.* 2016, 27, 333–338. [CrossRef]
- 6. Ebisu, K.; Berman, J.D.; Bell, M.L. Exposure to coarse particulate matter during gestation and birth weight in the U.S. *U.S. Environ. Int.* **2016**, *94*, 519–524. [CrossRef] [PubMed]
- Yeo, I.; Choi, Y.; Lops, Y.; Sayeed, A. Efficient PM<sub>2.5</sub> forecasting using geographical correlation based on integrated deep learning algorithms. *Neural Comput. Appl.* 2021, 1–17. [CrossRef]
- 8. Sayeed, A.; Lops, Y.; Choi, Y.; Jung, J.; Salman, A.K. Bias correcting and extending the PM forecast by CMAQ up to 7 days using deep convolutional neural networks. *Atmos. Environ.* **2021**, 253, 118376. [CrossRef]
- 9. Lim, J.H.; Kwak, K.K.; Kim, J.; Jang, Y.K. Analysis of Annual Emission Trends of Air Pollutants by Region. *J. Korean Soc. Atmos. Environ.* **2018**, *34*, 76–86. (In Korean with English abstract) [CrossRef]
- Kim, J. Assessment and Estimation of Particulate Matter Formation Potential and Respiratory Effects from Air Emission Matters in Industrial Sectors and Cities/Regions. *J. Korean Soc. Environ. Eng.* 2017, *39*, 220–228, (In Korean with English abstract). [CrossRef]
- 11. Choe, J.-I.; Lee, Y.S. A Study on the Impact of PM<sub>2.5</sub> Emissions on Respiratory Diseases. *J. Environ. Policy Adm.* **2015**, 23, 155. (In Korean with English abstract) [CrossRef]
- 12. Kim, Y. Air Pollution in Seoul Caused by Aerosols. J. Korean Soc. Atmos. Environ. 2006, 22, 535–553. (In Korean with English abstract)
- 13. Kim, J.; Kang, S. Analysis of Factors Influencing PM10 Pollution in Korea. *Proc. J. Korean Soc. Environ. Econ.* **2018**, 2018, 779–791. (In Korean with English abstract)
- 14. Park, S.; Shin, H. Analysis of the Factors Influencing PM<sub>2.5</sub> in Korea: Focusing on Seasonal Factors. *J. Environ. Policy Adm.* **2017**, 25, 227–248. (In Korean with English abstract) [CrossRef]

- 15. Biancofiore, F.; Busilacchio, M.; Verdecchia, M.; Tomassetti, B.; Aruffo, E.; Bianco, S.; Di Carlo, P. Recursive neu-ral network model for analysis and forecast of PM10 and PM2. 5. *Atmos. Pollut. Res.* **2017**, *8*, 652–659. [CrossRef]
- 16. Pak, U.; Ma, J.; Ryu, U.; Ryom, K.; Juhyok, U.; Pak, K.; Pak, C. Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Sci. Total Environ.* **2020**, *699*, 10. [CrossRef] [PubMed]
- Kim, H.S.; Park, I.; Song, C.H.; Lee, K.; Yun, J.W.; Kim, H.K.; Jeon, M.; Lee, J.; Han, K.M. Development of a daily PM10 and PM<sub>2.5</sub> prediction system using a deep long short-term memory neural network model. *Atmos. Chem. Phys. Discuss.* 2019, 19, 12935–12951. [CrossRef]
- Marloes, E.; Rob, B.; Kees de, H.; Tom, B.; Giulia, C.; Marta, C.; Christophe, D.; Audrius, D.; Evi, D.; Audrey de, N.; et al. Development of Land Use Regression Models for PM<sub>2.5</sub>, PM<sub>2.5</sub> Absorbance, PM10 and PMcoarse in 20 Eu-ropean Study Areas; Results of the ESCAPE Project. *Environ. Sci. Technol.* 2012, 46, 11205–11215.
- 19. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.* **1995**, 3361, 1995.
- Scarpa, G.; Gargiulo, M.; Mazza, A.; Gaetano, R. A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. *Remote Sens.* 2018, 10, 236. [CrossRef]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural. Inf.* Process. Syst. 2012, 25, 1097–1105. [CrossRef]
- 22. Li, X.; Peng, L.; Yao, X.; Cui, S.; Hu, Y.; You, C.; Chi, T. Long short-term memory neural network for air pollutant concentra-tion predictions: Method development and evaluation. *Env. Pollut.* **2017**, *231*, 997–1004. [CrossRef]
- 23. Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, CoRR. *arXiv Prepr.* **2014**, arXiv:1406.1078.
- 24. Gers, F.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. In Proceedings of the ICANN'99, IEEE, London, UK, 7–10 September 1999; pp. 850–855.
- Ravanelli, M.; Brakel, P.; Omologo, M.; Bengio, Y. Light Gated Recurrent Units for Speech Recognition. *IEEE Trans. Emerg. Top. Comput. Intell.* 2018, 2, 92–102. [CrossRef]
- Huang, Y.; Qu, G.; Wei, Z. A new image restoration method by Gaussian smoothing with L1 norm regularization. In Proceedings of the 2012 5th International Congress on Image and Signal Processing, Chongqing, China, 16–18 October 2012; pp. 343–346. [CrossRef]
- 27. Batista, G.E.; Monard, M.C. A Study of K-Nearest Neighbor as an Imputation Method. *Soft Comput. Syst. Des. Manag. Appl.* 2002, 87, 251–260.
- Franklin, M.; Koutrakis, P.; Schwartz, J. The Role of Particle Composition on the Association between PM<sub>2.5</sub> and Mortality. *Epidemiology* 2008, 19, 680–689. [CrossRef] [PubMed]
- Cho, S.-H.; Kim, P.-R.; Han, Y.-J.; Kim, H.-W.; Yi, S.-M. Characteristics of Ionic and Carbonaceous Compounds in PM<sub>2.5</sub> and High Concentration Events in Chuncheon, Korea. J. Korean Soc. Atmos. Environ. 2016, 32, 435–447. (In Korean with English abstract) [CrossRef]
- Bae, C.; Yoo, C.; Kim, B.-U.; Kim, H.; Kim, S. PM<sub>2.5</sub> Simulations for the Seoul Metropolitan Area: (III) Application of the Modeled and Observed PM2.5 Ratio on the Contribution Estimation. *J. Korean Soc. Atmos. Environ.* 2017, 33, 445–457. (In Korean with English abstract) [CrossRef]
- Yu, G.-H.; Lee, B.-J.; Park, S.-S.; A Jung, S.; Jo, M.R.; Lim, Y.J.; Kim, S. A Case Study of Severe PM<sub>2.5</sub> Event in the Gwangju Urban Area during February 2014. *J. Korean Soc. Atmos. Environ.* 2019, 35, 195–213. [CrossRef]
- 32. Lee, H.-J.; Jeong, Y.; Kim, S.-T.; Lee, W.-S. Atmospheric Circulation Patterns Associated with Particulate Matter over South Korea and Their Future Projection. *J. Clim. Chang. Res.* **2018**, *9*, 423–433, (In Korean with English abstract). [CrossRef]
- Jeon, B.-I. Meteorological Characteristics of the Wintertime High PM<sub>10</sub> Concentration Episodes in Busan. J. Environ. Sci. Int. 2012, 21, 815–824. (In Korean with English abstract) [CrossRef]
- Wang, M.; Zheng, S.; Li, X.; Qin, X. A new image denoising method based on Gaussian filter. In Proceedings of the 2014 International Conference on Information Science, Electronics and Electrical Engineering, Sapporo, Japan, 26–28 April 2014; Volume 1, pp. 163–167.
- 35. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113. [CrossRef] [PubMed]
- 36. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [CrossRef] [PubMed]
- 37. Chollet, F. Keras. 2015. Available online: https://keras.io (accessed on 22 September 2021).
- Pérez, P.; Trier, A.; Reyes, J. Prediction of PM<sub>2.5</sub> concentrations several hours in advance using neural networks in Santiago, Chile. *Atmos. Environ.* 2000, 34, 1189–1196. [CrossRef]
- Eder, B.; Kang, D.; Mathur, R.; Yu, S.; Schere, K. An operational evaluation of the Eta–CMAQ air quality forecast model. *Atmos. Environ.* 2006, 40, 4894–4905. [CrossRef]
- Chai, T.; Kim, H.C.; Lee, P.; Tong, D.; Pan, L.; Tang, Y.; Stajner, I. Evaluation of the United States National Air Quality Forecast Capability experimental real-time predictions in 2010 using Air Quality System ozone and NO<sub>2</sub> measurements. *Geosci. Model Dev.* 2013, 6, 1831–1850. [CrossRef]