

Article



Deep Random Subspace Learning: A Spatial-Temporal Modeling Approach for Air Quality Prediction

Xiaotong Sun and Wei Xu *

School of Information, Renmin University of China, Beijing 100872, China; sun_xiao_tong@ruc.edu.cn

* Correspondence: weixu@ruc.edu.cn

Received: 30 June 2019; Accepted: 23 August 2019; Published: 18 September 2019



Abstract: Decrease in air quality is one of the most crucial threats to human health. There is an imperative and necessary need for more accurate air quality prediction. To meet this need, we propose a novel long short-term memory-based deep random subspace learning (LSTM-DRSL) framework for air quality forecasting. Specifically, we incorporate real-time pollutant emission data into the model input. We also design a spatial-temporal analysis approach to make good use of these data. The prediction model is developed by combining random subspace learning with a deep learning algorithm in order to improve the prediction accuracy. Empirical analyses based on multiple datasets over China from January 2015 to September 2017 are performed to demonstrate the efficacy of the proposed framework for hourly pollutant concentration prediction at an urban-agglomeration scale. The empirical results indicate that our framework is a viable method for air quality prediction. With consideration of the regional scale, the LSTM-DRSL framework performs better at a relatively large regional scale (around 200–300 km). In addition, the quality of predictions is higher in industrial areas. From a temporal point of view, the LSTM-DRSL framework is more suitable for hourly predictions.

Keywords: air quality prediction; random subspace learning; deep learning; spatial-temporal analysis; smart city

1. Introduction

The pervasiveness of poor air quality in both developing and developed countries has brought about a global threat, having huge negative impacts on the environment and health. According to the Health Effects Institute (HEI), more than 90% of the world's population, around 7 billion people, live in unhealthy air environments [1]. Many studies across the world seek ways to quantify the magnitude of health harm caused by air pollution through systematic scientific efforts (e.g., [2]). Based on World Health Organization (WHO) data, about 4.2 million deaths happen every year due to ambient air pollution. Premature diseases such as lung disease, heart disease, and stroke, etc., have been noted as being mainly caused by air pollutants. The direct reason is that minors and babies are exposed to air pollutants for prolonged periods [3,4]. Long-term exposure can also cause ill health, such as new-onset type 2 diabetes among adults [5]. To protect people from these adverse health impacts, people are encouraged by the American Lung Association to pay more attention to air quality forecasts and take timely precautions [6]. Air quality forecasts making predictions of air pollutant concentrations have become imperative and urgent necessities for air quality control. These forecasts play a role as an essential component of air pollutant control strategies implemented on a regional scale. For instance, an O₃ action plan employed in French Bouches du Rhône uses air quality forecasting as a tool to trigger an emission reduction strategies selection process [7].

In the past few decades, researchers have devoted their efforts to exploring trends in air pollutant concentrations. Existing prediction studies can be generally separated into three categories based on modeling methods, namely, numerical methods, statistical methods, and machine learning methods.

Numerical models for air quality prediction, also called chemical transport models, model the transport velocity, diffusion path, and possible chemical reactions during pollutant movements so that the concentration of pollutants can be predicted using mathematical algorithms [8]. The input matrices for these models commonly include source terms (historical concentrations of air pollutants) [9], meteorological conditions (wind direction, humidity, and temperature, etc.) [10], emission source parameters (location and height, etc.) [11], terrain elevations [12], and properties of blocks in the path of pollutant movements [13]. Various numerical models have been developed to improve forecasting processes, such as the Community Multiscale Air Quality (CAMQ) model, the Weather Research and Forecasting (WRF) model, and the WRF model coupled with Chemistry (WRF-Chem) [14–16]. Furthermore, the COPERNICUS project uses a multi-model ensemble approach to provide regional air quality forecasts since the results derived from such an approach are more robust and are of better quality [17]. Although nowadays numerical models are still popularly implemented in forecast systems by governments, in order to achieve high-quality prediction performance, a relevant high portion of input data requires more accuracy than the accuracy which is currently available [18]. In addition, most practical air environment situations are complex and hard to express mathematically. This is another significant limitation of the numerical models available at present [19].

Plenty of statistical approaches have been used in the air quality prediction field. Traditional statistical methods have been widely employed to deal with air quality prediction problems. Good predictive effects have been gained through modeling and understanding of data probability generation mechanisms [20]. These models are substantiated by validating correctness of the probability distribution hypothesis for data. However, the assumptions for data probability distribution also imply limitations to further development of the traditional statistic model for practical applications [21].

With the development of artificial intelligence and big data analytics, prediction methods based on machine learning technologies are becoming increasingly common. These kind of models directly explore complicated hidden patterns in data, requiring neither hypothetical distributions of variables or data nor an in-depth understanding of physical or chemical properties of air pollutants. The models' quality is basically judged by their predictive accuracy and effectiveness [22]. Commonly used machine learning algorithms include multiple linear regression (MLR), random forest (RF) [23], support vector regression (SVR) [24], artificial neural networks (ANN) [25], and so forth. Previous studies have found that machine learning methods achieve excellent performance due to the nonlinear relationships within data, meaning that these methods are better suited to parameter statistic models and need less training time than dispersion models [26–28]. Deep learning algorithms, as a relative newcomer, have obtained outstanding prediction or detection performances in various application domains such as speech recognition, natural language processing, and computer vision [29]. Some pioneering research has used deep learning techniques to address air quality prediction problems.

In this paper, we developed a long short-term memory-based deep random subspace learning (LSTM-DRSL) framework to achieve high forecast accuracy. The framework employs a deep learning long and short-term memory model and combines it with a random subspace learning approach. As discussed above, deep learning methods make a better fit than traditional numerical and statistic methods since forecasting accuracy becomes more important than interpretability [27]. The LSTM model is a popular variant of the recurrent neural network (RNN) method, a famous deep learning algorithm. It performs better than traditional RNN when facing gradient vanishing or gradient explosion problems [30]. Besides, when using LSTM to make predictions or detections, it considers the effect of previous values on the current one in the model calculation [31–33]. This feature makes LSTM one of the best suited models for air quality prediction problems since temporal dependence is a typical phenomenon observed in air pollutant concentration series. We further combined LSTM with RSL to obtain the proposed framework, in which individual LSTM models employing different subsets of features serve as base models. The RSL approach is incorporated here to achieve more generality. In previous research, preliminary frameworks of a simpler combination of deep learning methods and RSL have been applied in several domains, such as crop disease prediction [34], stock

price manipulation detection [35], and financial market prediction [36]. The results of performance evaluation in these studies show that these frameworks outperform all baseline models and are capable of dealing with prediction and detection problems. The main differences between the framework proposed in this paper and previous frameworks exist in the base models selected for RSL and the spatial-temporal feature engineering processes.

Spatial-temporal feature engineering for air quality prediction is designed to capture the internal dependence between input variables. While previous research has mainly focused on predicting air quality using time-series data, the LSTM-DRSL framework considers spatial dependences caused by air pollutant interactions and transmissions as well. The direction of pollutant transmission is affected by the relationship between relative positions of pollutant emission sites to local monitoring sites and meteorological conditions like wind direction. The spatial dependence caused by fluidity of air pollutants is featured in our proposed method through two steps: zoning and regional synthesizing.

Real-time pollutant emission data are incorporated into the prediction framework to improve the prediction accuracy. Normally, the accuracy of the forecast increases as the data become more comprehensive and the quality of the data becomes better [37,38]. Although the emission data in our study are limited to stationary sources such as heating plants, it is reasonable to suppose that the incorporation of these data can improve the prediction performance and that replacing the estimation of emissions with the real-time observations can enhance the forecast ability. Moreover, the addition of real-time pollutant emission information into the prediction model can help with potential simulation of air planning effects [38]. Air planning refers to taking a series of structural measures to improve air quality and is another important activity in this domain in addition to forecasting. It is necessary to analyze what would happen when these structural measures are applied. Previous studies have shown that emission reduction policy and emission source relocation are effective in air quality control [39]. Taking the real-time emission data as part of the model input is helpful for decision-makers to estimate the performance of related strategies based on model output and make air planning decisions. However, as far as we know, there is still no research that includes real-time hourly emission data, other than emission inventories, in air quality forecasts. To mitigate and control the inaccuracy caused by the errors and hysteresis contained in raw data, real-time emission data is first employed in our prediction framework.

The main contributions of our study are the following. First, we propose a novel LSTM-DRSL framework using multiple data sources and spatial-temporal features to obtain air quality predictions. Second, we incorporate pollutant emission data into the model and design a spatial-temporal analysis approach for them. Third, a LSTM model combined with random subspace learning is developed and adopted in our framework to predict air pollution concentrations. Fourth, we demonstrate the effectiveness of our proposed methods by systematical comparative experiments on practical data.

2. Method

2.1. Overview

The method we propose in this study focuses on tackling the problem of lacking prediction accuracy. To address this issue, we introduce an important input data dimension, pollutant emission, into the prediction model. The extraction of emission features is based on spatial-temporal dependencies. The temporal autocorrelations existing in air pollutant concentrations are effectively captured by applying LSTM as the base model. According to the characteristics of prediction tasks, a data mining model which we called deep random subspace learning is developed to make good use of the features in a big data context. The DRSL model, which integrates the deep learning algorithms of LSTM and ensemble learning methods is represented as LSTM-DRSL below. The framework of the proposed method is shown in Figure 1.



Figure 1. The air quality forecasting framework. Legend: MODIS, moderate-resolution imaging spectroradiometer.

2.2. Data Collection and Preprocessing

During data collection, we collected datasets for the required variables of our prediction model from different data sources. The formation and propagation of air pollutants are complicated processes and have significant correlations with environmental circumstances. To achieve accurate predictions, it is crucial to contain all the information of the propagation process in the input to the prediction model as much as possible. We approached responsible government departments to get approval for obtaining data through the application programming interface (API) in their systems, while other data were collected from public government websites, such as Earthdata, via which data are regularly released by NASA (https://earthdata.nasa.gov/). The collected data have been classified into five types of variables, as illustrated in Table 1.

Туре	Variables	Observations	Data Source	
Air Quality	Ground pollutant measurement (GPM)	Hourly concentrations of PM _{2.5} , PM ₁₀ , CO, NO, NO ₂ , NO _X , O ₃ , SO ₂	National air quality monitoring network	
	Atmospheric air quality (AAQ)	Aerosol optical depth, total ozone burden	MODIS	
Meteorology	Surface meteorological measurement (SMM)	Hourly atmospheric pressure (hpa), humidity (%), temperature (°C), wind speed (m/s), wind direction (deg)	Automatic weather monitoring system	
	Atmospheric meteorology (AM)	Atmospheric stability, moisture, atmospheric temperature, atmospheric water vapor	MODIS	
Emission	Pollutant emission (PE) Hourly emissions of SO ₂ , NO ₂ particles (kg/h); hourly benchmark gas flow (m ³)		National key monitored enterprise	

Table 1. Independent variables. Legend: $PM_{2.5}$, particulate matter with an aerodynamic diameter of 2.5 µm or less; PM_{10} , particulate matter with an aerodynamic diameter of 10 µm or less.

Ground pollutant measurement variables (GPM). China established a national-scale ground monitoring network for typical pollutants in late 2012, since the air quality problem in various regions has become severe [40]. The network consists of more than 1400 monitor stations covering 300 cities. Each station site monitors multiple air pollutants including PM_{2.5} (particulate matter with an aerodynamic diameter of

 $2.5 \,\mu m$ or less), NO₂, and SO₂, etc. The concentration data of these pollutants are generated at an hourly frequency. Surface meteorological measurement variables (SMM). As stated in previous literature [41], pollutant concentrations are strongly associated with meteorological conditions as the factors affecting physical processes of pollutant generation and dispersion are associated with meteorological variables like temperature. Therefore, it is important to integrate various meteorological variables within the prediction model. Atmospheric air quality variables (AAQ) and atmospheric meteorology variables (AM). Both AAQ and AM are derived from satellite data collected by sensors like the moderate-resolution imaging spectroradiometer (MODIS). These two types of variables have been widely used to estimate air quality (e.g., PM_{2.5} exposures) in previously unmonitored areas [42]. Among these remote sensing variables, the aerosol optical depth (AOD) has been proven to be one of the most valuable predictors [43]. Pollutant emission variables (PE). It is universally acknowledged that pollutant emissions are one of the major air pollution sources. China launched an initiative in the year 2013 to nationally monitor crucial industrial enterprises' emissions, which required the polluting enterprises to self-monitor and publish their data to a government platform. According to the relevant regulations from the Ministry of Ecology and Environment, the hourly average of an enterprise's emissions should be obtained through automatic monitoring equipment and should be published in real time. It should be noted that the emission data used in our study can only express the emission situation of "point" sources due to the data acquisition limitation.

In summation, there are five kinds of independent variables in the prediction framework, i.e., GPM, SMM, AAQ, AM, and PE. From the view of connotation, these variables can be grouped into three categories, namely, air quality, meteorology, and emission (see Table 1). For each individual dataset from a different source, the data can be transformed into time series by clustering data based on spatial information first regardless of the type of variable and the initial form of the data. The extraction of spatial information and the unification of data forms lay the foundation for data preprocessing.

During data preprocessing, we first integrated the multisource data from all the original datasets by making full use of time and space information, as shown in Table 1. The integration process helps obtain a unified representation of input variables that can improve the quality of data mining. The GPM data and SMM data were first integrated based on the latitude and longitude value of the monitoring sites as well as the data's acquisition time. For the sites serving as monitoring stations for both air pollution and weather conditions, we were able to simply put these two types of data together in one sample. We constructed samples for other sites based on the proximity principle, which used the SMM data collected by the nearest automatic weather monitoring station. Remote sensing data were the average values assigned to each 1 km × 1 km space grid in the area; we combined them with the GPM and SMM data in a way that adopted the AAQ and AM data of the grid occupied by the station site. For the monitoring sites that fell on the grid boundary, we calculated the average of the AAQ and AM variable values of all the grids sharing their edge or vertex and took the average as the corresponding values for the site.

We then imputed the missing data for the sample set constructed by the integration process so that substantial bias created by missing information could be avoided and the analysis of data be more efficient. The data form for the samples were unified as time series. The imputation values were determined only by the values before and after missing data in the temporal dimension. Our fill values were calculated by adhering to the following formula in the data imputation process:

$$Input = \alpha_1 avg_h + \alpha_2 (\beta_1 V_{forw} + \beta_2 V_{backw})$$

where *Input* is the imputation value of the missing data, avg_h is the average of all the valid data collected at h o'clock, which is also the time of situation that is supposed to be reflected by the missing data, V_{forw} and V_{backw} are the nearest valid values after and before the missing value, β_1 and β_2 are the weights determined by the ratio of distance between the missing value and the nearest one (before or after) and the distance between the two nearest valid values, respectively, before and after the missing one, and α is the parameter representing the weight of the corresponding component in the addition

operation which are experimentally determined to optimize the filling effect. Adhering to previous research, we set $\alpha_1 = \alpha_2 = 0.5$ in our experiments [44].

In addition, we performed reprocessing steps like data cleansing and data standardization. We conducted data cleansing by identifying the irregular and abnormal data caused by fortuitous equipment malfunctions and then corrected them according to relevant monitoring standards. During data standardization, the effect of dimensions was removed through defining a specific base dimension for each variable and transforming them into a unified base dimension.

2.3. Emission Features Based on Spatial-Temporal Analysis

Spatial dependence in a variety of monitoring data denotes the circumstance in which observations associated with one site depend on those at surrounding sites. For instance, particulate matter (PM_{2.5}) concentrations at a particular position will subsequently rise as an outcome of the wind blowing particles from a nearby region to the local area. Considering the latent regional interactions observed in multipoint monitoring, there is a critical need to quantify and model the indeterminate spatial dependence. In this study, we proposed an elementary design to incorporate spatial information and the internal dependence on them into the forecasting framework.

Determining and partitioning the neighboring region of a site is a principal consideration. The distance between the nearby pollutant emission sites and the target monitoring site plays a crucial role in the process since it is believed to have a substantial impact on the spatial dependence between sites. Generally, the distances between encompassing sites and the central site vary enormously among the different sites. As shown in Figure 2a, three circles employ different diameters but the same center to determine and separate the surrounding area of the site located at the center. The size of diameters should range within the distances of any two monitoring sites. The three diameters applied in this study were 1 km, 10 km, and 100 km, respectively. The diameter values were set based on two principles. The first one referred to previous research. Based on this principle, we set the minimum diameter as 1 km, since the maximum ground concentrations of air pollutants calculated in most studies are less than 1 km. The second principle was to ensure that there were emission sources in each subzone. For each monitoring site, around 60% of the distances between the emission sources and the site were less than 10 km and the maximum distance was less than 100 km. To take into account the orientational relationship of sites and the influence of the wind direction, we further separated each circular area into eight equal pieces by four lines passing through the common center. All in all, the marked neighboring regions (rounded by the outermost circle) were ultimately divided into several parts where closer regions had a finer granularity and farther regions had a coarser granularity. It should be noted that the partitioning process has practical significance since it leads to an upper bound to both the number of variables and the number of corresponding model parameters in prediction models. This process helps control the total input data amount and reduce the training time.

The variable data of pollutant emission sites falling within the same part of the surrounding regions were be aggregated in different ways considering the nature of the data. The emission sites were projected onto the corresponding piece of neighboring regions based on their geographic coordinates. For PE data, we summed the emission volume for each pollutant and took the sum as the representation of the discharge load of pollutants in certain regions. Eventually, each part of the neighboring region obtained values for a set of PE variables. Every center site added 24 extra sets of PE variable values representing the 24 parts of its neighboring regions, respectively, in the sample. As a result, the spatial dependence and relevant spatial information of pollutant emissions were integrated into the prediction model.

Temporal dependence refers to the impact of past situations on the status quo. This dependency has been broadly considered in prediction models employed to different domains, including agriculture, environment, finance, and so forth [36,45]. We prepared the data in chronological order so that the prediction model could mine the dependence through time-series data. We took the variable *T*, representing the length of the time series, as the key parameter to enhance the predictive capability.

That is, when forecasting the air quality at time t, the sample included the variable data at times from t - 1 to t - T (as illustrated in Figure 2b). In agreement with previous research, we set T = 3 in this study [44]. For each time slice, the spatial information from nearby stations within the circle of the outermost diameter was added on the basis of temporal dependence. In addition, this kind of dependency is believed to exist among the panel data of air pollutant concentration and meteorological conditions as well, since these factors vary by time and affect their future values by influencing the propagation process of pollutants through air. Hence, temporal dependence was also considered for the remaining variables, including GPM, SMM, AAQ, and AM.



Figure 2. Illustration of spatial-temporal feature engineering. (**a**) Spatial dependence; (**b**) temporal dependence.

2.4. Modeling and Prediction

The architecture for the LSTM network applied in our study is shown in Figure 3. As can be observed from the figure, each dotted box represents an individual module which transforms input into output. Repeating these modules in the form of a chain enables LSTM to learn long-term dependencies. The memory cell layer, the major difference between LSTM and traditional RNN, acts as a conveyor belt to connect the information, meaning the LSTM can remember information for long periods of time. The introduction of memory cells ameliorates the gradient training through using the cell to determine the extent of the absorption of previously acquired knowledge and the extent of updating the hidden state. Gate mechanisms are designed to regulate the ability to remove or add information, and an LSTM here has three of these gates (denoted G_i , G_f , and G_o).



Figure 3. The long short-term memory network.

In a traditional RNN network, as introduced by Elman (1990), the hidden state is simply updated from the previous state H_t through the adoption of new input information I_{t+1} as shown in the formula given below:

$$H_{t+1} = f(H_t, I_{t+1})$$
(1)

Yet, with the addition of gate mechanisms built on the linearly-connected memory cells, the update process becomes more complicated for the hidden layer. To explain the process clearly, we basically split the updating course into three phases.

When a new sample, I_t , is put into an LSTM model, the most recent information contained in the input sample will first be mixed with the information accumulated in the previous state of the hidden layer H_{t-1} to estimate a value for the memory cell unit, \widetilde{M}_t , containing all the latest information, which can be calculated as

$$\overline{M_t} = \tanh(U_M I_t + V_M H_{t-1}) \tag{2}$$

where *U* and *V* represent the coefficient matrices of the input state and last hidden state optimized by training, separately. The gate mechanism facilitating decision-making function depends on three vectors: $G_{i,t}$, $G_{f,t}$, and $G_{o,t}$. These gating vectors are iterated in every step. $G_{f,t}$ and $G_{i,t}$ are used to control the extent to which the memory cells should be erased or updated. These two vectors, as input gate and forget gate, respectively, have entries ranging from 0 to 1. On the basis of these two gates at time *t*, we can determine the ultimate value of M_t as follows.

$$M_t = G_{f,t} \odot M_{t-1} + G_{i,t} \odot M_t, \tag{3}$$

where the operation \odot denotes elementwise multiplication between the memory cell states and the gate vectors. Subsequently, the hidden state H_t is able to be obtained by M_t and the output gate $G_{o,t}$, as represented in the following:

$$H_t = G_{o,t} \odot \tanh M_t \tag{4}$$

The transition equations of those three gates mentioned above are the following:

$$G_{i,t} = \sigma(U_i X_t + V_i H_{t-1} + P_i M_{t-1}),$$
(5)

$$G_{f,t} = \sigma \Big(U_f X_t + V_f H_{t-1} + P_f M_{t-1} \Big), \tag{6}$$

$$G_{o,t} = \sigma (U_o X_t + V_o H_{t-1} + P_o M_t),$$
(7)

where *P* represents the coefficient matrices of the memory cell state and σ denotes the logistic sigmoid function f_{σ} , i.e.,

$$f_{\sigma}(x) = \frac{1}{1 + e^{-x}}.$$

So far, the LSTM units at time *t* containing a hidden state H_t , a memory cell C_t , and three gates including an output gate $G_{o,t}$, a forget gate $G_{f,t}$, and an input gate $G_{i,t}$ have all been presented.

As for the training, we adjusted parameters M, N, and P through the backpropagation through time (BPTT) method (see Appendix A) to minimize the loss function constructed on mean squared error (MSE):

$$L_{MSE} = \frac{1}{N} \sum_{t=t_0}^{t+N} ||F_t - TrueV_t||_2^2$$
(8)

where *N* is the batch size employed in the mini-batch BPTT training, $TrueV_t$ represents the ground-truth value of the prediction target at time *t*, and *F*_t denotes the predictive value obtained from the equation, as in

$$F_t = f_F(H_t) = W_F H_t + b_F \tag{9}$$

9 of 21

where f_F represents the activation function in the forecast layer, W_F represents the weight matrix, and b_F represents the bias term. The squared Euclidean norm, expressed as $\|\cdot\|$ in Formula (9), represents the prediction error of each timestamp.

We then developed a data mining model integrating the LSTM model and random subspace ensemble (RSE) method, referred to as the LSTM-DRSL. The RSE approach was introduced into the air quality prediction task since the total number of constructed features (N_F) reached several hundred and RSE achieves great performance in modeling high-dimensional data. In our proposed framework, we employed the bootstrap method to construct random spaces by randomly sampling n_f emission features ($n_f \leq N_F$) during the feature selection process. It is worth noting that the number of emission features in each random subspace has been confirmed to be sufficiently large to demonstrate the usefulness of the pollutant emission. We repeated the sampling process N times so that we could obtain N random subspaces. For each random subspace, we combined emission features with air quality and meteorology features to train an LSTM predictor. Then, the predictions resulting from the base LSTM classifiers served as inputs to the stack classifier. The training process of LSTM-DRSL is interpreted detailly in the following pseudocode (Algorithm (1)).

Algorithm 1: Learning LSTM-DRSL through BPTT

```
Input: Training samples S;
    Output: Weight matrices U_i, V_i, P_i, U_f, V_f, P_f, U_o, V_o, P_o, and W_F for N base models respectively (N is the
    number of random subspaces); stack model StackM;
1: Initialize U_i, V_i, P_i, U_f, V_f, P_f, U_o V_o, P_o, and W_F randomly;
2: Set prediction time window T_w;
3: Sort input samples in chronological order from t_0 to t_T;
4: Set time stamp t = t_o + T_w;
5: Initialize the serial number of random subspaces n = 0;
6: // Build random subspaces
7: while n \le N do
8:
     Randomly sample n_f of emission features and integrate them with other
9:
     features into S_n;
10: n = n + 1;
11: end while
12: n = 0;
13: // Training the base LSTM models and the stack model
14: while not converge do
15: while t \le t_T do
16:
        while n \le N do
17:
          Compute F_t based on S_n; (Formula (1)~(7), (9))
          Compute L_{MSE} at t; (Formula (8))
18:
19:
          Update weight matrices for model PreM<sub>n</sub>; (BPTT)
20:
          Obtain prediction value Pre<sub>n</sub>;
21:
          n = n + 1;
22:
        end while
        Train the stack model StackM(Pre_n) (n \in \{n | 0 \le n \le N\});
23:
24:
        t = t + 1;
25: end while
```

26: end while

3. Experiment

3.1. Data Description

We collected observations of the major air pollutants including $PM_{2.5}$, PM_{10} , CO, NO, NO₂, NO_X, O₃, and SO₂. These data were generated from monitoring networks set up by the Chinese government.

A dataset containing 451,509 records was obtained from an air quality monitoring system through an authorized API. The data range from January 2013 to September 2017. Other related data from data sources including MODIS and automatic weather monitoring systems were matched to the air quality observations by time and location in the preprocessing progress. After that, we excluded samples from January to December in 2013 since the meteorological data collection mechanism had not been established by then, and 275,362 records remained. Data representing pollutant emission conditions were also incorporated as supplemental spatial information. Pollutant emission data were regularly reported by 179 national crucial monitoring enterprises. However, the enterprise monitoring system was not established until 2015. We dropped the data collected before the monitoring system was operated and the total amount of records further decreased to 212,424. The fluctuation of air pollutant concentration data is shown above (Figure 4).



Figure 4. The fluctuation of air pollutant concentration trends. This figure shows the hourly concentration fluctuation for eight major pollutants at one monitoring site in Shenyang (from January 1st 2017 to March 31st 2017).

3.2. Evaluation Metrics

A ten-fold cross-validation technique was employed in our study. The original training dataset was randomly partitioned into ten subsets, each of which contained approximately one-tenth of the training data. This was referred to as the cross-validation process and in it we used nine subsets as training data and the remaining subset as testing data. The process was then repeated 10 times, during which every subset was taken as testing data once. Each round of trial yielded performance results. The average values of each metric were used as the estimates of the predictive capacity of algorithm.

When evaluating models, the data was first sorted into chronological order and then the last 20% were used as the hold-out test set. We trained each model on the remaining 80% of the data and applied the 10-fold cross-validation to verify the effectiveness of the model. The cross-validation process also helped tune models' parameters and complete model selection.

We calculated statistical indicators including root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) between observations in the hold-out test set and predictions for them to assess the prediction accuracy of models for the entire study period.

We performed five sets of experiments to methodically appraise the efficacy of our proposed framework. Each set of comparative experiments were designed to assess the effectiveness of one innovative aspect of our method. The first set experiments were set to compare the performance of the LSTM algorithm which we applied in this work to the benchmark machine learning models discussed above (in Section 1). It is worth noting that the results of the experiments were also an estimation of the usefulness of considering time dependency, since LSTM is the only algorithm for which its the computational processes include time-relevant mechanisms. In the second comparative experiment set, we made air quality forecasts using and without using the emission features and assessed the improvement when adding feature dimensions. As shown in Section 2.3, the pollutant emission variables were reorganized based on the spatial environment before being incorporated into the models. Accordingly, the experiments assessed the validity of spatial information as well. The third set of experiments compared the predictive performance before and after combining random subspace ensemble learning with LSTM models. In the fourth set of experiments, we tested the performance of the proposed framework across variable regions. The input data was separated by sub-regions so we could examine whether the framework achieved the same performance given consideration of the regional variations. The fifth set of experiments were designed to see the capability of the framework from a temporal point of view, in which we observed if the prediction results aggregated per season

4. Results and Analysis

were in the same quality.

4.1. Comparison with Baseline Models

In this subsection, a few machine learning technologies which are commonly used in previous research were employed as benchmark methods. As introduced in Section 1, MLR is one of the most widespread models in the air quality prediction field. ANN, SVR, and RF are also popular machine learning techniques. They usually achieve better forecast performance through exploiting the latent relationship between independent variables and dependent variables.

The features data were organized in a time sequential way so that the input for time-independent models could be equivalent to those for time-dependent ones. The input for time-independent models turned into a collection of sample data from t–T to T, where t represents the time that the forecast value corresponds to and T indicates the length of the time window as introduced in Section 2.3 (according to pre-experiment results, T = 3 here [44]).

The forecast of the PM_{2.5} concentration, which has become a top public concern in recent years, was taken as an example to evaluate the validity of all models mentioned above. Although not shown in tabular forms in this paper, the predictive effectiveness for other pollutants exhibited the same trend. As illustrated in Table 2, the LSTM algorithm outperformed all the baseline models. The results demonstrate that LSTM attained 5.07%, 6.56%, and 9.60% improvement over the best baseline model RF in RMSE, MAE, and MAPE, respectively. Additionally, the average improvement rate between the LSTM model and the time-independent models achieved 10.43%, 11.49%, and 15.27%. By applying algorithms containing sequence learning mechanisms, air pollutant concentration can be predicted with an around 75% precision on the hold-out test dataset. This suggests that including temporal analysis within the prediction process leads to improvements in prediction accuracy.

Table 2. Performance of prediction models using only non-spatial features. Legend: RMSE, root-mean-square error; MAE, mean absolute error; MAPE, mean absolute percentage error; MLR, multiple linear regression; ANN, artificial neural networks; SVR, support vector regression; RF, random forest.

Predictors	RMSE	MAE	MAPE (%)
MLR	14.761	12.709	32.523
ANN	13.686	11.894	29.132
SVR	13.454	11.817	28.685
RF	12.896	11.273	27.508
LSTM	12.241	10.534	24.867

We also performed a Wilcoxon signed-rank test to assess whether there was a statistically significant change in the pair of values predicted by the matched pair of models. Each pair of models contained the LSTM model and one of the baseline models. As shown in Table 3, differences existed between the above two types of models and they were all significant. Since improvement can be realized by replacing the predicting model with LSTM, the test results further demonstrate the validity of the proposed methods.

Table 3. Wilcoxon signed-rank test results on prediction performances.

H_0 : There are No Significant Differences in the Sample Means				
LSTM				
0.00 ***				
0.00 ***				
0.00 ***				
0.00 ***				

*** p – value ≤ 0.01 .

4.2. Incremental Effect of Combined Spatial Features

To test the incremental effect of combining spatial emission features, we added spatialized variables to prediction frameworks based on different models and then predicted concentrations for different air pollutants. First, we compared the performance of all five models using and not using emission feature sets to evaluate the incremental effect of the combined spatial features. In this experiment, $PM_{2.5}$ was also taken as an example and the best-performance model was selected for further experiments. The results of this pre-experiment are recorded in Figure 5. Second, we applied the previous best-performance model, i.e., the LSTM model (as shown in Figure 5), to all the major pollutants so we could estimate the generality of our prediction framework to different pollutants. The performance of LSTM on eight pollutants is shown in Figure 6.





Figure 5. Performances of prediction models using different baselines.



Figure 6. Performances of prediction models for different air pollutants.

Comparing the performance of all models shown in Table 2 and Figure 5, it is evident that the performance was enhanced when spatial features were incorporated, achieving average improvements by 6.079%, 6.179%, and 9.274% on each evaluation, respectively. The LSTM model attained the best predictive capacity with an around 80% accurate rate. However, for the MLR model, the performance

of adding spatial emission features was somewhat worse, which demonstrated that more features do not necessarily lead to effectiveness. Instead, the performance on hold-out test datasets may be decreased when overfitting occurs.

We trained predictors for each of the eight major air pollutants correspondingly. The results applying the best performance model (LSTM) are presented in Figure 6. The average MAPE for each pollutant is 20.908%, with the minimum and maximum values being 17.327% (CO) and 23.902% (PM₁₀). The average improvements of RMSE, MAE, and MAPE are 8.069%, 8.235% and 11.965%. The positive influence that spatial information can contribute to predictive precision was further confirmed by these evaluation results.

From the performance of the two different feature sets above in the comparison experiments, it can be seen that putting in spatial features is essential for air quality prediction process. The best prediction performance was observed for the LSTM-based predictor whose input incorporated spatial features. Moreover, we note that the spatial analysis is universally effective for all models and contexts. This further implies the potential for the spatial analysis approach to be applied to other advanced techniques or research problems.

4.3. Overall Improvement with Random Subspace Ensemble

Random subspace ensemble learning was utilized in our framework to enhance the air pollutant concentration prediction by building subspaces and training models for each subspace. According to the principle of random subspace ensemble learning introduced in Section 2.4, various but a sufficient number of features should be selected to develop subspaces to seek a higher generalization ability. The feature selection process can be illustrated as a bias-variance tradeoff which influences the effectiveness of the random subspace ensemble.

The random subspace ensemble learning mechanism was employed in the LSTM-based framework (the best predictors demonstrated in Section 4.2) with spatial information added. The evaluations detailed in Table 4 reveal the overall performance of the prediction framework for each pollutant. The predictive results were effectively improved after adding random subspace ensemble learning. The RMSE, MAE, and MAPE of major pollutant concentrations decreased by 4.501%, 4.763%, and 5.124% on average. The greatest increase occurred when predicting PM_{10} concentrations, for which the MAPE was decreased by 6.774%, followed by 5.815% for $PM_{2.5}$ and 11.815% for SO₂. The highest accuracy appeared in CO concentration prediction, with an accuracy of around 82%. The mean MAPE of CO prediction results dropped from a previous 19.777% to 17.327%. Therefore, we can conclude from the third set of experiments that random subspace ensemble learning can help improve the performance of air quality prediction.

TASKS	RMSE		MAE		MAPE (%)	
mono	LSTM	LSTM-DRSL	LSTM	LSTM-DRSL	LSTM	LSTM-DRSL
PM _{2.5}	11.138	10.537	9.585	9.094	21.295	20.057
PM_{10}	18.197	17.252	15.812	14.899	23.902	22.283
NO	0.664	0.625	0.596	0.564	20.006	18.957
NO ₂	7.40	7.198	6.328	6.128	20.106	19.141
NO _X	5.91	5.655	5.129	4.884	22.790	21.799
SO ₂	4.454	4.265	3.846	3.665	21.787	20.527
СО	0.138	0.133	0.120	0.115	17.327	16.683
O ₃	8.955	8.562	7.723	7.344	19.931	19.029
Average Improvements		4.501%		4.763%		5.124%

Table 4. Performances using LSTM combined with random subspace ensemble.

4.4. Performance Comparison with Consideration of Spatial Variations

We verified the quality of the predictions at a smaller regional scale to see how our framework performs with consideration of spatial variations. Liaoning province is a main industry base in China. The urban agglomeration of central Liaoning, which suffers from severe air pollution, is comprised of eight cities, i.e., Shenyang, Anshan, Fushun, Benxi, Yingkou, Liaoyang, Tieling, and Fuxin. We applied our framework to predict $PM_{2.5}$ concentrations for each city in this urban agglomeration and compared the overall performance to previous prediction performance (at a larger regional scale). The performance of prediction models for $PM_{2.5}$ concentrations in these cities are contrasted with each other in Figure 7.



Figure 7. Performances of forecasting PM_{2.5} in sub-regions.

As shown in the following figure, the overall RMSE, MAE, and MAPE values for these eight individual cities are higher than those values for a larger regional scale (RMSE = 10.537, MAE = 9.094, MAPE = 20.057, as shown in Table 4). We suggest that the combination of RSL is a possible reason for our framework performing better at the larger regional scale.

In addition, for each evaluation metric, the comparison among these sub-regions show a similar result. It can be seen in the figure that the southern cities of the urban agglomeration, such as Yingkou, Benxi, and Anshan, have smaller forecasting errors, while the northern cities including Tieling and Shenyang experience lower accuracy. Through further analysis, we found that most steel plants and petrochemical factories in this area are in the southern region, which has a plain terrain, and the northern region has a higher terrain and is mostly mountainous areas. We suppose that the difference in factory pollution emission between south and north may be a contributing factor to this comparative result.

4.5. Performance Comparison with Consideration of Temporal Variations

We examined the capability of the LSTM-DRSL framework from a temporal point of view by segregating the $PM_{2.5}$ forecasting results per season. The seasons were characterized following the classification of the China Meteorological Administration as winter (November, December, January, February, and March), spring (April and May), summer (June, July, and August), and autumn (September. and October). A characteristically seasonal variation of $PM_{2.5}$ concentrations can be observed from Figure 8. In this figure, we also used dark blue bars to present the MAE of the prediction model and drew a polyline based on the right axis to better compare the MAE of daily predictions to hourly predictions.



Figure 8. Performances of forecasting PM_{2.5} in different seasons.

As shown in Figure 8, $PM_{2.5}$ levels peaked in winter for the seasonal cycle due to increased fuel burning for heat. However, the model usually performed better in winter, followed by autumn, spring, and summer. We think that the incorporation of pollutant emission information on heating plants is a possible reason for our framework to perform best in winter and worst in summer.

When comparing between the MAE in Figure 8a,b, the quality of daily predictions is poorer than hourly predictions. We believe that the application of LSTM accounts for the better performance of the hourly concentration prediction. The reason to consider the LSTM model as a top priority is that the main difference between the daily and hourly prediction models is the T parameters of the LSTM models (T = 1 in daily predictions and T = 3 in hourly predictions). The parameter T represents a time interval limitation. When the time interval between the time point of the historical data and prediction is smaller than the T parameter, it can be incorporated into an input for a prediction framework. Since the advantage of the LSTM model is learning long-term dependence, it cannot exhibit its advantage when the parameter value is small.

5. Conclusions

In recent years, air pollution issues worldwide have become increasingly serious, resulting in many adverse health outcomes [46]. This study focused on air pollution predictions and proposed a LSTM-DRSL framework to estimate future concentrations of major air pollutants. The input variables were gathered from multiple sources including remote sensing satellites as well as mainstream monitoring networks. Feature engineering based on spatial-temporal analysis is designed to capture the dependence of air conditions on nearby pollutant emissions and the influence from a previous state to the present state. The LSTM-DRSL model was developed to make concentration predictions

with higher stability and suitability. We employed our LSTM-DRSL framework to forecast air quality and conducted systematical comparative experiments to demonstrate the effectiveness with regard to each aspect.

In the empirical analysis, we carried out five sets of comparative experiments on real-world datasets to assess the effectiveness of our framework. The results of the first-set experiments showed that LSTM significantly outperformed all the benchmark machine learning models. We could draw the conclusion from these results that the introduction of LSTM models to air pollution forecasts is valuable and applicable. The second set of experimental results revealed improvements with the addition of spatial features in both baseline models and the LSTM model from which we could see the validity of spatial feature engineering. In the third set of experiments, the forecast results evinced great enhancements after combing random subspace learning technique with the LSTM model, which achieved the best performance compared with the above two experiments. Overall, the completed results of the first three comparative experiments displayed the competence of the LSTM-DRSL framework in air pollutant concentration prediction based on spatial-temporal feature engineering and a combination of new techniques. However, there are still some limitations that should be noted in the application of our framework. The first two limitations were revealed by the fourth set of experiments. The results demonstrated that the effectiveness of the LSTM-DRSL framework cannot act completely at a smaller regional scale. Additionally, the framework performed a little poorer in non-factory-intensive areas. This limitation was also disclosed in the fifth set of experiments. Finally, the ability of our framework to deal with temporal variations, which was tested in the last set of experiments, was worse when predicting daily pollutant concentrations. In other words, the LSTM-DRSL performs better in solving problems which have a small unit of time or long-term time dependency.

In future work, in addition to further research regarding computational costs, the sources of input variables for air pollutant emissions can be further enriched, especially for mobile pollution sources such as motor vehicle exhaust. Another possible direction of future study is applying newly developed techniques for time sequence problems to existing prediction frameworks. Such work is worth the research effort since techniques are continuously being developed and bring new opportunities.

Author Contributions: Conceptualization, W.X. and X.S.; methodology, W.X. and X.S.; validation, X.S.; formal analysis, X.S.; investigation, X.S.; resources, W.X.; data curation, X.S.; writing—original draft preparation, X.S.; revision, W.X. and X.S.; supervision, W.X.

Funding: This work was supported in part by the National Science and Technology Major Project (Grant No. 2017YFC0212501), the National Natural Science Foundation of China (Grant No. 71771212, U1711262), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075, 15YJA630068), and Undergraduate Teaching Reform Project at Renmin University of China (Research on the training mode of top-notch innovative talents under interdiscipline subject)

Acknowledgments: The authors would like to thank Prof. Guojun Song at Renmin University of China for his valuable suggestions, and thank the editors and the anonymous reviewers for their valuable comments and suggestions which have helped immensely in improving the quality of this paper.

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

Appendix A

Backpropagation Through Time (BPTT)

The basic formulas of the LSTM model applied in our study are the following (also seen in Section 2.4):

$$F_t = f_F(H_t) = W_F H_t + b_F$$
$$H_t = G_{o,t} \odot \tanh M_t$$
$$M_t = G_{f,t} \odot M_{t-1} + G_{i,t} \odot \widetilde{M}_t$$
$$\widetilde{M}_t = \tanh(U_C I_t + V_C H_{t-1})$$

The equations for the input, forget, and output gates are demonstrated as:

$$G_{i,t} = \sigma(U_iI_t + VH_{t-1} + P_iM_{t-1})$$

$$G_{f,t} = \sigma(U_fI_t + V_fH_{t-1} + P_fM_{t-1})$$

$$G_{o,t} = \sigma(U_oI_t + V_oH_{t-1} + P_oM_t)$$

The loss is defined as MSE, given by:

$$L_{MSE} = \|F_t - TrueV_t\|_2^2$$

Each training sample contains observation data acquired during the time window. The sample can be regarded as a time-series sequence whose length is T (as introduced in Section 2.3). As a result, the error of the sample at time t is accumulated from the error of observation at time t - T to that at time t.

We calculated the gradient of the loss function with regard to each parameter of the model. Like the summation operation of loss, the gradient is summed up at each timestamp from t - T to t for each training example.

The chain rule of differentiation is employed in the gradient calculation process. First, we take partial derivation with respect to F_t , and we can obtain:

$$\delta F_t = \frac{\partial L_{MSE}}{\partial F_t} = 2(F_t - TrueV_t)$$

Then, we calculate the derivation with respect to W_F , H_t , and b_F separately:

$$\begin{split} \delta W_F &= \frac{\partial L_{MSE}}{\partial W_F} = \frac{\partial L_{MSE}}{\partial F_t} \frac{\partial F_t}{\partial W_F} = \delta F_t H_t \\ \delta b_F &= \frac{\partial L_{MSE}}{\partial b_F} = \frac{\partial L_{MSE}}{\partial F_t} \frac{\partial F_t}{\partial b_F} = \delta F_t \\ \delta H_t &= \frac{\partial L_{MSE}}{\partial H_t} = \frac{\partial L_{MSE}}{\partial F_t} \frac{\partial F_t}{\partial H_t} = \delta F_t W_F \end{split}$$

As H_t depends on $G_{o,t}$ and M_t , we use the chain rule again. So we have:

$$\delta G_{o,t} = \frac{\partial L_{MSE}}{\partial G_{o,t}} = \frac{\partial L_{MSE}}{\partial H_t} \frac{\partial H_t}{\partial G_{o,t}} = \delta H_t \tanh M_t$$

$$\delta M_t = \frac{\partial L_{MSE}}{\partial M_t} = \frac{\partial L_{MSE}}{\partial H_t} \frac{\partial H_t}{\partial G_{o,t}} = \delta H_t G_{o,t} (1 - \tanh^2(M_t))$$

Also, $G_{o,t}$ depends on U_o , V_o , and P_o , so as we can obtain δU_o , δV_o , and δP_o from the following:

$$\begin{split} \delta U_o &= \frac{\partial L_{MSE}}{\partial U_o} = \frac{\partial L_{MSE}}{\partial G_{o,t}} \frac{\partial G_{o,t}}{\partial U_o} = \delta G_{o,t} (1 - G_{o,t}) G_{o,t} I_t \\ \delta V_o &= \frac{\partial L_{MSE}}{\partial V_o} = \frac{\partial L_{MSE}}{\partial G_{o,t}} \frac{\partial G_{o,t}}{\partial V_o} = \delta G_{o,t} (1 - G_{o,t}) G_{o,t} H_{t-1} \\ \delta P_o &= \frac{\partial L_{MSE}}{\partial P_o} = \frac{\partial L_{MSE}}{\partial G_{o,t}} \frac{\partial G_{o,t}}{\partial P_o} = \delta G_{o,t} (1 - G_{o,t}) G_{o,t} M_t \end{split}$$

For M_t , partial derivations with respect to $G_{i,t}$, $G_{f,t}$, and \widetilde{M}_t are needed:

$$\begin{split} \delta G_{i,t} &= \frac{\partial L_{MSE}}{\partial G_{i,t}} = \frac{\partial L_{MSE}}{\partial M_t} \frac{\partial M_t}{\partial G_{i,t}} = \delta M_t M_{t-1} \\ \delta G_{f,t} &= \frac{\partial L_{MSE}}{\partial G_{f,t}} = \frac{\partial L_{MSE}}{\partial M_t} \frac{\partial M_t}{\partial G_{f,t}} = \delta M_t \widetilde{M}_t \\ \delta \widetilde{M}_t &= \frac{\partial L_{MSE}}{\partial \widetilde{M}_t} = \frac{\partial L_{MSE}}{\partial M_t} \frac{\partial M_t}{\partial \widetilde{M}_t} = \delta M_t G_{i,t} \end{split}$$

Then, the gradients of the MSE loss with respect to parameters *U*, *V*, and *P* can be gained:

$$\begin{split} \delta U_{S} &= \frac{\partial L_{MSE}}{\partial U_{S}} = \frac{\partial L_{MSE}}{\partial G_{S,t}} \frac{\partial G_{S,t}}{\partial U_{S}} = \delta G_{S,t} (1 - G_{S,t}) G_{S,t} I_{t} \\ \delta V_{S} &= \frac{\partial L_{MSE}}{\partial V_{S}} = \frac{\partial L_{MSE}}{\partial G_{S,t}} \frac{\partial G_{S,t}}{\partial V_{S}} = \delta G_{S,t} (1 - G_{S,t}) G_{S,t} H_{t-1} \\ \delta P_{S} &= \frac{\partial L_{MSE}}{\partial P_{S}} = \frac{\partial L_{MSE}}{\partial G_{S,t}} \frac{\partial G_{S,t}}{\partial P_{S}} = \delta G_{S,t} (1 - G_{S,t}) G_{S,t} M_{t} \end{split}$$

where *S* denotes the type of gates that the parameter effects $S \in \{i, f\}$.

Finally, the partial differential equations for U_C and V_C are as follows:

$$\delta U_{C} = \frac{\partial L_{MSE}}{\partial U_{C}} = \frac{\partial L_{MSE}}{\partial \widetilde{M}_{t}} \frac{\partial \widetilde{M}_{t}}{\partial U_{C}} = \delta \widetilde{M}_{t} \left(1 - \widetilde{M}_{t}^{2}\right) I_{t}$$
$$\delta V_{C} = \frac{\partial L_{MSE}}{\partial V_{C}} = \frac{\partial L_{MSE}}{\partial \widetilde{M}_{t}} \frac{\partial \widetilde{M}_{t}}{\partial V_{C}} = \delta \widetilde{M}_{t} \left(1 - \widetilde{M}_{t}^{2}\right) H_{t-1}$$

At this point, the gradients of the loss with respect to all parameters (U, V_i , P_i , U_f , V_f , P_f , $U_o V_o$, and P_o and U_C , V_C , W_F , and b_F) have obtained their calculation values. Then the mini-batch gradient descent (MBGD) is used to optimize parameters (according to Formula (8) in Section 2.4). Note that the gradients only depend on current values of terms on the right-hand side in the equations.

References

- Health Effects Institute, State of Global Air 2018. Available online: https://www.stateofglobalair.org/archives, 2018 (accessed on 21 March 2019).
- Kloog, I.; Ridgway, B.; Koutrakis, P.; Coull, B.A.; Schwartz, J.D. Long- and short-term exposure to PM2.5 and mortality: Using novel exposure models. *Epidemiology* 2013, 24, 555–561. [CrossRef] [PubMed]
- Ebenstein, A.; Fan, M.; Greenstone, M.; He, G.; Zhou, M. New evidence on the impact of sustained exposure to air pollution on life expectancy from china's huai river policy. *Proc. Natl. Acad. Sci. USA* 2013, 110, 12936–12941. [CrossRef] [PubMed]
- 4. Schwartz, J. Particulate air pollution and chronic respiratory disease. Environ. Res. 1993, 62, 7–13. [CrossRef]
- 5. Chan, C.K.; Yao, X. Air pollution in mega cities in china. Atmos. Environ. 2008, 42, 1–42. [CrossRef]
- American Lung Association, State of the Air 2018. Available online: https://www.lung.org/assets/documents/ healthy-air/state-of-the-air/sota-2018-full.pdf (accessed on 21 March 2019).
- 7. Ferretti, V.; Montibeller, G. Key challenges and meta-choices in designing and applying multi-criteria spatial decision support systems. *Decis. Support Syst.* **2016**, *84*, 41–52. [CrossRef]
- 8. Zhu, S.; Lian, X.; Liu, H.; Hu, J.; Wang, Y.; Che, J. Daily air quality index forecasting with hybrid models: A case in China. *Environ. Pollut.* **2017**, *231*, 1232–1244. [CrossRef] [PubMed]
- 9. Xu, B.; Lin, H.; Chiu, L.; Hu, Y.; Zhu, J.; Hu, M.; Cui, W. Collaborative virtual geographic environments: A case study of air pollution simulation. *Inform. Sci.* **2011**, *181*, 2231–2246. [CrossRef]
- Werner, M.; Kryza, M.; Ojrzynska, H.; Skjoth, C.A.; Walaszek, K.; Dore, A.J. Application of WRF-Chem to forecasting PM10 concentration over Poland. *Int. J. Environ. Pollut.* 2015, *58*, 280–292. [CrossRef]
- 11. Beevers, S.D.; Nutthida, K.; Williams, M.L.; Carslaw, D.C. One way coupling of CMAQ and a road source dispersion model for fine scale air pollution predictions. *Atmos. Environ.* **2012**, *59*, 47–58. [CrossRef]
- 12. Abdul-Wahab, S.; Sappurd, A.; Al-Damkhi, A. Application of California Puff (CALPUFF) model: A case study for Oman. *Clean Technol. Environ. Policy* **2011**, *13*, 177–189. [CrossRef]
- Tartakovsky, D.; Broday, D.M.; Stern, E. Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain. *Environ. Pollut.* 2013, 179, 138–145. [CrossRef]
- Byun, D.; Schere, K.L. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.* 2006, *59*, 51–77. [CrossRef]
- 15. Done, J.; Davis, C.A.; Weisman, M. The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.* **2004**, *5*, 110–117. [CrossRef]

- 16. Grell, G.A.; Peckham, S.E.; Schmitz, R.; McKeen, S.A.; Frost, G.; Skamarock, W.C.; Eder, B. Fully coupled "online" chemistry within the WRF model. *Atmos. Environ.* **2005**, *39*, 6957–6975. [CrossRef]
- 17. Global Forecast Plots-Copernicus. Available online: https://atmosphere.copernicus.eu/global-forecast-plots (accessed on 20 July 2019).
- 18. Taheri Shahraiyni, H.; Sodoudi, S. Statistical modeling approaches for PM₁₀ prediction in urban areas; A review of 21st-century studies. *Atmosphere* **2016**, *7*, 15. [CrossRef]
- 19. Wang, Y.; Xu, W. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decis. Support Syst.* **2018**, *105*, 87–95. [CrossRef]
- 20. Lv, M.; Li, Y.; Chen, L.; Chen, T. Air quality estimation by exploiting terrain features and multi-view transfer semi-supervised regression. *Inform. Sci.* **2019**, *483*, 82–95. [CrossRef]
- 21. Bai, L.; Wang, J.; Ma, X.; Lu, H. Air pollution forecasts: An overview. *Int. J. Environ. Res. Public Health* **2018**, 15, 780. [CrossRef]
- 22. Yang, C.S.; Wei, C.P.; Yuan, C.C.; Schoung, J.Y. Predicting the length of hospital stay of burn patients: Comparisons of prediction accuracy among different clinical stages. *Decis. Support Syst.* **2010**, *50*, 325–335. [CrossRef]
- 23. Yu, R.; Yang, Y.; Yang, L.; Han, G.; Move, O. RAQ—A random forest approach for predicting air quality in urban sensing systems. *Sensors* **2016**, *16*, 86. [CrossRef]
- 24. Lin, K.P.; Pai, P.F.; Yang, S.L. Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms. *Appl. Math. Comput.* **2011**, *217*, 5318–5327. [CrossRef]
- 25. Wang, P.; Liu, Y.; Qin, Z.; Zhang, G. A novel hybrid forecasting model for PM10 and SO2 daily concentrations. *Sci. Total Environ.* **2015**, *505*, 1202–1212. [CrossRef] [PubMed]
- 26. Wang, J.; Zhang, X.; Guo, Z.; Lu, H. Developing an early-warning system for air quality prediction and assessment of cities in China. *Expert Syst. Appl.* **2017**, *84*, 102–116. [CrossRef]
- 27. Rahman, N.H.A.; Lee, M.H.; Latif, M.T. Artificial neural networks and fuzzy time series forecasting: An application to air quality. *Qual. Quant.* **2015**, *49*, 1–15. [CrossRef]
- 28. Meissner, M.; Schmuker, M.; Schneider, G. Optimized Particle Swarm Optimization (OPSO) and its application to artificial neural network training. *BMC Bioinform.* **2006**, *7*, 125. [CrossRef]
- 29. Huang, X.; Qi, J.; Sun, Y.; Zhang, R.; Zheng, H.T. CARL: Aggregated search with context-aware module embedding learning. *arXiv* **2019**, arXiv:1908.03141.
- Sak, H.; Senior, A.; Beaufays, F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Proceedings of the International Speech Communication Association, Singapore, 14–18 September 2014.
- 31. Kraus, M.; Feuerriegel, S. Decision support from financial disclosures with deep neural networks and transfer learning. *Decis. Support Syst.* **2018**, *104*, 38–48. [CrossRef]
- 32. Mahmoudi, N.; Docherty, P.; Moscato, P. Deep neural networks understand investors better. *Decis. Support Syst.* 2018, 112, 23–34. [CrossRef]
- 33. Graves, A. Generating sequences with recurrent neural networks. arXiv 2013, arXiv:1308.0850.
- 34. Xu, W.; Wang, Q.; Chen, R. Spatio-temporal prediction of crop disease severity for agricultural emergency management based on recurrent neural networks. *GeoInformatica* **2018**, *22*, 363–381. [CrossRef]
- 35. Wang, Q.; Xu, W.; Huang, X.; Yang, K. Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. *Neurocomputing* **2019**, *347*, 46–58. [CrossRef]
- 36. Wang, Q.; Xu, W.; Zheng, H. Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing* **2018**, *299*, 51–61. [CrossRef]
- 37. Grolinger, K.; L'Heureux, A.; Capretz, M.A.; Seewald, L. Energy forecasting for event venues: Big data and prediction accuracy. *Energy Build.* **2016**, *112*, 222–233. [CrossRef]
- 38. Monteiro, A.; Lopes, M.; Miranda, A.I.; Borrego, C.; Vautard, R. Air pollution forecast in Portugal: A demand from the new air quality framework directive. *Int. J. Environ. Pollut.* **2005**, *25*, 4–15. [CrossRef]
- Wang, S.; Zhao, M.; Xing, J.; Wu, Y.; Zhou, Y.; Lei, Y.; He, K.; Fu, L.; Hao, J. Quantifying the air pollutants emission reduction during the 2008 olympic games in beijing. *Environ. Sci. Technol.* 2010, 44, 2490–2496. [CrossRef] [PubMed]
- 40. Liu, M.; Bi, J.; Ma, Z. Visibility-based PM2.5 concentrations in China: 1957–1964 and 1973–2014. *Environ. Sci. Technol.* **2017**, *51*, 13161–13169. [CrossRef] [PubMed]

- 41. Chuang, M.T.; Zhang, Y.; Kang, D. Application of WRF/Chem-MADRID for real-time air quality forecasting over the Southeastern United States. *Atmos. Environ.* **2011**, *45*, 6241–6250. [CrossRef]
- 42. Van Donkelaar, A.; Martin, R.V.; Park, R.J. Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *J. Geophys. Res. Atmos.* **2006**, *111*, D21201. [CrossRef]
- 43. Hsu, A.; Reuben, A.; Shindell, D.; Sherbinin, A.; Levy, M. Toward the next generation of air quality monitoring indicators. *Atmos. Environ.* **2013**, *80*, 584–590. [CrossRef]
- Sun, X.; Xu, W.; Jiang, H. Spatial-temporal prediction of air quality based on recurrent neural networks. In Proceedings of the Hawaii International Conference on System Sciences, Big Island, HI, USA, 8–11 January 2019.
- 45. Zheng, Y.; Yi, X.; Li, M.; Li, R.; Shan, Z.; Chang, E.; Li, T. Forecasting fine-grained air quality based on big data. In Proceedings of the 21th SIGKDD conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 2267–2276.
- 46. World Health Organization, Declaration of the Sixth Ministerial Conference on Environment and Health. Available online: http://www.euro.who.int/en/media-centre/events/events/2017/06/sixth-ministerialconference-on-environment-and-health/documentation/declaration-of-the-sixth-ministerial-conferenceon-environment-and-healthCopenhagen (accessed on 21 March 2019).



© 2019 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/).