Journal of Cleaner Production 237 (2019) 117729

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2.5

Jun Ma [a, b], Yuexiong Ding [b], Jack C.P. Cheng [a], Feifeng Jiang [c], Zhiwei Wan [d, *]

[a] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[b] Department of Research and Development, Big Bay Innovation Research and Development Limited, Hong Kong, China
[c] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China
[d] School of Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

ARTICLE INFO

Article history:
Received 7 May 2019
Received in revised form
17 July 2019
Accepted 20 July 2019
Available online 22 July 2019

Handling Editor: Prof. Jiri Jaromir Klemeš

Keywords:
Air quality
Deep learning
Geographic LSTM
PM2.5
Spatial extrapolation

ABSTRACT

Nowadays, real-time air pollution monitoring has been an important approach for supporting pollution control and reduction. However, due to the high construction cost and limited detection range of monitoring stations, not all the air pollutant concentrations in every corner can be monitored, and a whole picture of the spatial distribution of air pollution is usually lacked for comprehensive spatial analysis and air quality control. To address this problem, satellite remote sensing and spatial interpolation/extrapolation technologies have been commonly used in past research. However, the spatial distribution calculated by remote sensing techniques could be less accurate due to the limited amount of recorded data for testing and adjustments. Performance of traditional spatial interpolation/extrapolation techniques, such as Kriging and IDW, was limited by several subjective assumptions and pre-set formulations that are less suitable for non-linear real-world situations. As an alternative, machine learning and neural network-based methods have been proposed recently. However, most of these methods failed to well consider the long short temporal trend and spatial associations of air pollution simultaneously. To overcome these limitations, this paper proposed a newly designed spatial interpolation/extrapolation methodology namely Geo-LSTM to generate the spatial distribution of air pollutant concentrations. The model was developed based on the Long Short-Term Memory (LSTM) neural network to capture the long-term dependencies of air quality. A geo-layer was designed to integrate the spatial-temporal correlation from other monitoring stations. To evaluate the effectiveness of the proposed methodology, a case study in Washington state was conducted. The experimental results show that Geo-LSTM has a RMSE of 0.0437, and is almost 60.13% better than traditional methods like IDW.

1. Introduction

1.1. Background

Recent years have witnessed the unprecedented development speed of urbanization and industrialization. As an unavoidable result of the rapid development, the emission load of air pollutants, such as PM2.5, PM10, nitrogen oxides and carbon dioxide are increasing. These pollutants, conversely, have adverse effects on human health and natural environment (Yang and Liu, 2018). To mitigate the impacts, the live surveillance of air pollutants and the release of the real-time air quality information to the public has been an important approach (A et al., 2018). Numerous air quality monitoring stations have been built to monitor and predict air pollutant concentrations, and give early warnings.

However, due to the high cost of construction and the constraints of land use, the numbers and distributions of the air quality monitoring stations are always limited. They cannot observe the air pollutant concentrations of every corner due to their limited monitoring range. The observed data given by these observation points therefore cannot reflect the whole picture of the real-time spatial distribution of the air pollution in the region, which actually is of great practical significance for comprehensive spatial analysis and air pollution control (Tong et al., 2015).

* Corresponding author.
E-mail address: wanzhiwei0213@outlook.com (Z. Wan).

https://doi.org/10.1016/j.jclepro.2019.117729
0959-6526/

1.2. Literature review and research objective

Since it is unreasonable to allocate monitoring sites to cover every corner of the area, prediction of the air pollution in areas without monitoring stations has been an important issue in order to generate the spatial distribution picture. Currently, satellite remote sensing and spatial interpolation/extrapolation technologies are commonly applied to accomplish the task. Satellite remote sensing retrieves air quality using satellite data collected from space-based sensors by using electromagnetic radiation (Engel-Cox et al., 2004). It can give a full mapping of the atmosphere and air quality assessment without being in physical contact with the atmosphere. Several scholars have adopted satellite remote sensing to generate a spatial distribution of air pollutant concentrations. For example, Gupta et al. (2006) used one year of aerosol optical thickness (AOT) retrieved from the MODIS sensors, along with ground measurements of PM2.5 mass concentrations, to assess air quality over different locations across the global urban areas. Huang et al. (2019, p.) incorporated satellite data such as percentage of sunshine coverage and aerosol optical depth with annually and monthly land use regression models to estimate the spatial distribution of NO_2 and PM2.5 in Taiwan. Hirtl et al. (2014) developed fine resolved PM10 maps from MODIS AOT to improve the prediction accuracy for air pollution. Yang et al. (2019) also used MODIS AOT to estimate ground-level PM2.5 concentrations over a coastal region of China. However, despite the wide application of satellite data, it has some unavoidable limitations (Martin, 2008). Since satellite sensors mainly monitor the earth surface from the space, images they retrieved are easily influenced by the cloud in the troposphere. The resolution of the image therefore might be low and the data extracted from it could be inaccurate especially in cloudy days. The number of satellite data, on the other hand, is limited by the swath width and orbital geometry of sensors. Furthermore, since sensors are always moving around in the space, their recorded data for one region is not continuous. Time duration between two data or images could be hours, days, months or even years. Therefore, when using remote sensing techniques to present the spatial distribution of air pollution, the performance could be limited and inaccurate.

Considering the limitations mentioned above on satellite image-based approaches, other scholars started to study numerical spatial interpolation/extrapolation methods. Spatial interpolation or extrapolation is the procedure of estimating the value of observations of non-monitored locations with data from known and monitored points within the area (Deligiorgi and Philippopoulos, 2011). Currently, point-based interpolation/extrapolation such as the distance weighting, the nearest neighbor, spline interpolation and Kriging are the most frequently used methods (Lam, 1983). For example, Joseph et al. (2013) applied Inverse Distance Weighting (IDW), nearest neighbor, original Kriging as well as universal Kriging for 8-h ozone interpolation with thousands of observed data. Tong et al. (2015) analyzed the performance of different Kriging interpolation methods based on the air quality index in Wuhan. Chen and Liu (2012) estimated the spatial rainfall distribution using IDW in the middle of Taiwan.

However, although these interpolation/extrapolation methods have been widely used for air quality prediction, their prediction accuracy could be limited due to subjective assumptions and preset formulations. For example, the major calculation method of IDW is developed based on the self-determined power or exponential function, which is highly depended on the modeling experience of researchers and cannot be easily generalized (Lu and Wong, 2008a). Kriging methods and IDW both are developed based on the assumption that the study objective is static, which is not suitable to the real world (OLIVER and WEBSTER, 1990).

To overcome the limitations, recent studies have adopted machine learning and neural networks as an effective alternative. For examples, Liu et al. (2011) developed a novel spatial interpolation/extrapolation method based on integrated Radial Basis Function (RBF) neural network, and proved that their proposed approach provided a more detail spatial distribution than using the traditional methods for lead element content in the soil. Wahid et al. (2013) also adopted the RBF neural network to estimate the spatial distribution of ozone concentrations in the Sydney basin, Australia. Chronopoulos et al. (2008) applied Artificial Neural Network (ANN) models to estimate meteorological data values in areas with sparse meteorological stations and proved that ANN models provided advantages over more traditional models or methods. Pfeiffer et al. (2009) presented a method based on diffusive sampling measurements and ANN evaluation to calculate the average spatial distribution of air pollutant level.

However, despite the strong robustness, memory capacity, nonlinear mapping capacity and strong self-learning ability of neural networks, they have limitations. While meteorological data, such as air pollution, rainfall, temperature, and soil conditions, follow a temporal variation trend, most of the existing neural networks-based spatial interpolation methods did not consider the temporal trend and spatial associations together. Their results therefore failed to consider the lagged influence from historical time series and the supporting information from the surrounding stations. Rigol et al. (2001) described the spatial interpolation of daily minimum air temperature using a Back-Propagation Neural Network (BPNN) and took into consideration the temporal mechanism of the data. However, the BPNN model assumes consecutive time series data as independent and cannot capture the impact of the previous record on the next (Hochreiter and Schmidhuber, 1997).

To overcome the limitations and fill the research gap, this paper proposes a neural network-based spatial-temporal interpolation/extrapolation model, namely Geo-LSTM, to estimate the air pollution at non-monitored locations with data from known and observed points. The proposed method is based on deep learning neural networks, and is better at learning long short temporal dependency from time series data than traditional machine learning algorithms and neural networks. The newly designed geo layer can help integrate the spatial correlation from other known points to interpolate the values in unknown points. Details of the proposed methodology are introduced in Section 2. To validate the effectiveness of the proposed model, a case study is presented in Section 3. Results are evaluated and discussed in Section 4. Section 5 concludes the work.

2. Methodology framework

Fig. 1 presents the methodology framework proposed in this paper. It is composed of three parts, including data preprocessing, modeling and interpolation. The first part includes data collection and data cleaning. The second part contains time series sample modeling and model training. Sample modeling is conducted first to take into consideration the temporal and spatial correlation among observations. Then the constructed samples are input into the geo Long Short-Term Memory (Geo-LSTM) model to train and test the interpolation performance of the model. Model structure and parameters are optimized, and performances of different interpolation models are compared. After that, the model can then be applied for spatial interpolation. Experiments and discussions on its temporal-spatial extrapolation performance are followed.

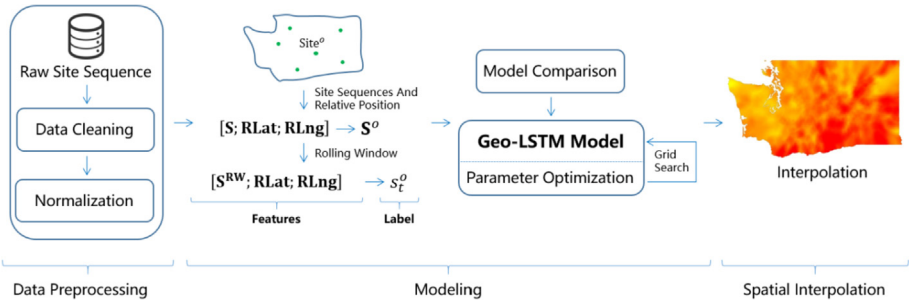

Fig. 1. Methodology framework.

2.1. Sample modeling

As shown in Fig. 1, after the raw data are collected and cleaned, the modeling process is conducted. The process contains two parts, including time series sample modeling and model training. Sample modeling transforms the data inputs into a more model-friendly form (Ma and Cheng, 2016a). Usually, the Rolling Window (RW) method is adopted for time series sample modeling. Given a time sequence $\mathbf{X} = [x_1, x_2, \ldots, x_{T-1}, x_T]$ with the time length as T, to predict the value at time t, RW will input not only the value at t-1 but also values at t-2, t-3, …, t-r into the model. The time series sample then can be formulated as Equation (1).

$$\{[x_{t-r}, \ldots, x_{t-2}, x_{t-1}] \rightarrow x_t\} \tag{1}$$

where r represents the rolling window size. Larger r means fewer time series samples but more training features, while smaller r means more time series samples but fewer training features (Ma and Cheng, 2016b).

Based on RW, sample modeling approach developed in this paper is as follows. Assume there are N monitoring stations within the study area. Observations of these N stations can be expressed as Equation (2).

$$\boldsymbol{S}_t^1 = [\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_t] \tag{2}$$

where $\boldsymbol{S}_t^1 \in \mathbb{R}^{t \times N}$, and $\mathbf{S}_t = [s_t^1, \ldots, s_t^N] \in \mathbb{R}^{1 \times N}$, which represents the observed value of the N stations at time t. s_t^j represents the observation value of the jth station at time t, $1 \le t \le T$, $1 \le j \le N$.

Then for a target point o without monitoring stations, since its air pollution concentration could be influenced by the current and previous values of the surrounding stations, the time series sample for modeling an unknown place o at time t using the RW method can be expressed as Equation (3).

$$\{[\mathbf{S}_{t-r+1}, \ldots, \mathbf{S}_t] \rightarrow s_t^o\} \tag{3}$$

Furthermore, as mentioned above, this paper takes into consideration the spatial mechanisms of air pollution as well. For the N monitoring stations, their latitude and longitude can be expressed as Equations (4) and (5).

$$\mathbf{Lat} = \left[lat^1, \ldots, lat^N\right] \in \mathbb{R}^{1 \times N} \tag{4}$$

$$\mathbf{Lng} = \left[lng^1, \ldots, lng^N\right] \in \mathbb{R}^{1 \times N} \tag{5}$$

where lat^j, and lng^j represent the latitude and longitude of station j, $1 \le j \le N$.

Since closer stations might have larger impacts on the target point while farther stations might have smaller impacts, directly calculating the distance between the known point and the unknown point might lose the direction information (Ma and Cheng, 2017). To address this problem, relative positions between points are considered in this paper. **RLat** is used to represent the vector of the relative latitudes and **RLng** represents the vector of the relative longitudes. Calculations of **RLat** and **RLng** are shown as follows:

$$\mathbf{RLat} = (lat^o - \mathbf{Lat}) \in \mathbb{R}^{1 \times N} \tag{6}$$

$$\mathbf{RLng} = (lng^o - \mathbf{Lng}) \in \mathbb{R}^{1 \times N} \tag{7}$$

where lat^o and lng^o are the latitude and longitude of the target point. After adding the spatial information, the inputs and outputs of the proposed model can be formulated as follow:

$$\left\{\left[\boldsymbol{S}_t^{t-r+1}, \mathbf{RLat}, \mathbf{RLng}\right] \rightarrow s_t^o\right\} \tag{8}$$

where $\boldsymbol{S}_t^{t-r+1} = [\mathbf{S}_{t-r+1}, \ldots, \mathbf{S}_t] \in \mathbb{R}^{r \times N}$, and r is the size of the rolling window. Note that when modeling the training and testing samples in the case study, the target point o will be set to the existing stations for performance comparison. In this case, the relative data in the inputs for station o would be removed, and resulting. $\boldsymbol{S}_t^{t-r+1} \in \mathbb{R}^{r \times (N-1)}$ and $\{\mathbf{RLat}, \mathbf{RLng}\} \in \mathbb{R}^{1 \times (N-1)}$

2.2. Geo-LSTM network structure

2.2.1. Geo-layer

Following the methodology framework, after the time series samples are constructed considering both temporal and spatial information using RW, they are input into the geo Long Short-Term Memory (Geo-LSTM) model for prediction and interpolation. Geo-LSTM is a special LSTM model which takes into consideration the spatial and temporal mechanisms of the input data at the same time. The network structure of the Geo-LSTM model used in this paper is presented in Fig. 2. It can be seen that compared with the

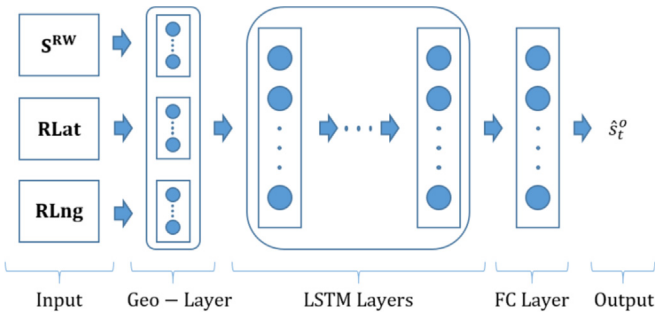

Fig. 2. Network structure of Geo-LSTM.

ordinary LSTM model, the proposed model has not only the input layer, LSTM layers, the fully connected (FC) layer and the output layer, but also a newly designed geo-layer.

As mentioned in the last section, the relative position between the known point and the unknown point will influence the strength of the spatial influence. Therefore, when using the information of monitored stations to predict the pollution concentration of the unknown point, not all the information should be considered and assigned the same weight. Closer stations might play a more important role while farther stations might be noisy data. To this end, the geo-layer is designed to select stations that have stronger correlations and drop those have weaker correlations. The graphical structure of the Geo-layer is shown in Fig. 3, and its calculation formulation is expressed in Equation (9).where $g_t^{in} = [\mathbf{S}_t, \mathbf{RLat}, \mathbf{RLng}]$, $\boldsymbol{R_A}$ is the activated matrix for indicating the valid inputs, and is calculated by the activation function $A_g(x)$ shown in Equations (10) and (11).

$$\boldsymbol{R_A} = A_g(\boldsymbol{R}) \tag{10}$$

$$A_g\left(\mathrm{R}^{\mathrm{j}}\right) = \begin{cases} 1, & if\ \mathrm{R}^{\mathrm{j}} \le K \\ 0, & otherwise \end{cases} \tag{11}$$

Note that $\boldsymbol{R} = [\mathrm{R}^1, \ldots, \mathrm{R}^{\mathrm{j}}, \ldots, \mathrm{R}^{\mathrm{N}}] \in \mathbb{R}^{1\times \mathrm{N}}$. It is calculated by the rank function shown in Equations (12) and (13).

$$\boldsymbol{R} = \mathrm{Rank}(\mathbf{RLat}, \mathbf{RLng}) \tag{12}$$

$$\mathrm{R}^{\mathrm{j}} = \begin{cases} inf, & if\ (\mathrm{Rlat}_{\mathrm{t}} = 0)\ and (\mathrm{Rlng}_{\mathrm{t}} = 0) \\ \mathrm{R}^{\mathrm{j}}, & otherwise \end{cases} \tag{13}$$

where Rank(x) is a rank function based on the shortest Euclidean distance. Equation (13) is used to avoid self-prediction in training.

After the preprocessing of the geo-layer, the LSTM layer will extract information from the input samples layer by layer and pass them on. Finally, the extracted information will be passed to the FC layer and the final predicted results will be output by the output layer. Expanded structure of the Geo-LSTM network is shown in Fig. 4.

2.2.2. LSTM unit

Furthermore, it can be seen from Fig. 4 that the information is passed on from one layer to the next and from one time point to the next through some blocks. These blocks are known as LSTM units. They can allow a value or gradient that flows into the unit to be preserved and subsequently retrieved at the required time step (Hochreiter and Schmidhuber, 1997), and therefore to overcome the exploding/vanishing gradient problem during the gradient propagation over layers and to learn from long-term dependencies (Bengio et al., 1994). Due to this advantage, recent studies on cleaner productions have implemented this technique in various

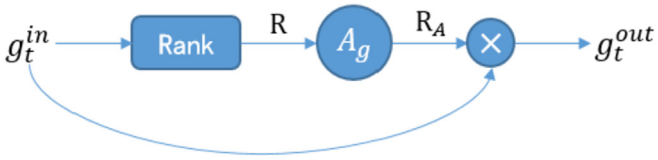

Fig. 3. Network structure of the geo-layer.

$$\boldsymbol{g_t^{out}} = \boldsymbol{R_A} * \boldsymbol{g_t^{in}} \tag{9}$$

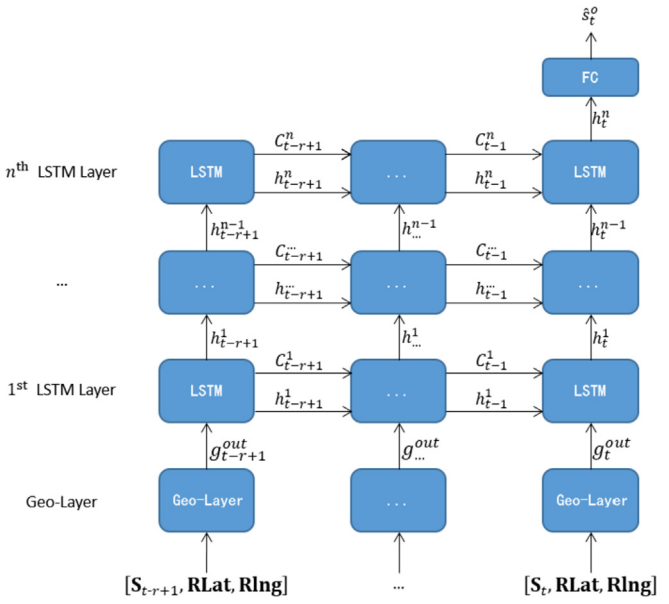

Fig. 4. Expanded network structure of Geo-LSTM.

topics. For example, Zhou et al. (2019) explored a multi-output LSTM neural network for regional multi-step-ahead air quality forecasts. QU et al. (2019) implemented LSTM to predict the air quality indicators at different heights and pressures in Beijing. Wen et al. (2019) combined convolutional neural networks with LSTM to promote the air quality predictions in China. Qi et al. (2019) integrated Graph Convolutional networks and LSTM (GC-LSTM) to model and forecast the spatiotemporal variation of PM2.5 concentrations. However, existing LSTM based forecasting methods are mainly designed for typical time series forecasting. They can only predict the air quality at where there exist historical records. Limited studies have explored the possibilities of implementing LSTM based algorithm on spatial interpolation/extrapolation.

The architecture of an LSTM unit is shown in Fig. 5. Inputs of the unit include the current input data x_t, the hidden state of the last unit h_{t-1} and the cell state of the last unit C_{t-1}. Outputs of the unit contain the hidden state h_t and the cell state C_t, which are controlled by the forget gate f_t, the input gate i_t, and the output gate o_t. The controlling process and the outputs of the unit can be mathematically formulated as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, \mathrm{x}_t] + b_f\right) \tag{14}$$

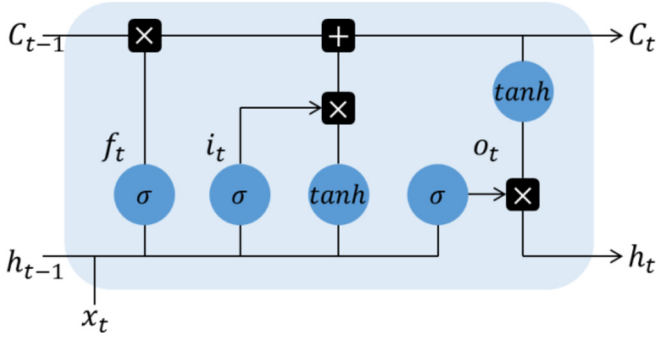

Fig. 5. Architecture of an LSTM unit.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (15)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (16)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (17)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (18)$$

$$h_t = o_t * \tanh(C_t) \quad (19)$$

where W and b are the coefficient and bias vector.

Overall, this paper designs a geo LSTM-based spatial interpolation (Geo-LSTM-SI) model to generate the spatial distribution for air pollutants. The model inputs consider not only the temporal changing trend of the pollutant but also the spatial mechanism of it. To validate the effectiveness of the proposed Geo-LSTM-SI model, a case study is then conducted.

3. Case study

3.1. Data collection and preprocessing

The State of Washington is selected as the study area, because this state is reported to have the worst air quality in the U.S. (FOX news, 2017). Hourly PM2.5 concentration data of 37 monitoring stations within the area are collected from the United States Environmental Protection Agency (EPA), covering from 1st January to 31st January 2017. Since four out of the 37 stations contain a large number of missing data, the remaining 33 stations without missing values are utilized in this paper. Data of each station can be considered as a time sequence with 31*24 = 744 observations. The distribution of the 33 monitoring stations is shown in Fig. 6. A summary of the PM2.5 data from these 33 stations is shown in Table 1.

3.2. Model optimization

Based on the collected PM2.5 concentration data, time series samples are constructed using the Rolling Window (RW) method. Then according to the methodology framework, these samples are used to train the geo Long Short-Term Memory (Geo-LSTM) model and interpolate the concentrations in unknown places. To achieve the best interpolation performance of the model, several parameters need to be determined at first, such as the rolling window size, the number of the top K stations, and the network structure and parameters. PM2.5 records of all the 33 stations are used. The way to model the time series samples follows Equation (8). Note that the

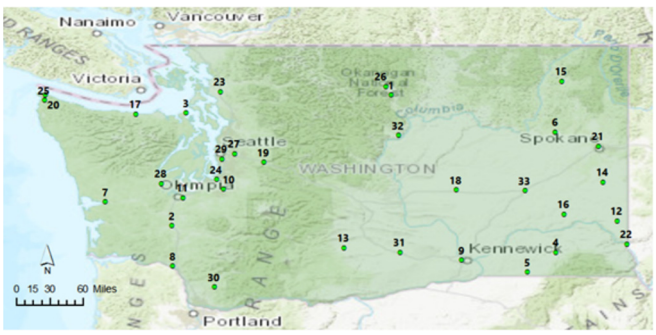

Fig. 6. Distribution of the 33 monitoring stations.

Table 1
A brief summary of the collected PM2.5 data.

Attribute	Value
Data content	PM2.5 concentrations
Location	Washington (state), the U.S.
Duration	2017/01/01–2017/01/31
Number of stations	33
Observations	744 records per station
1st quantile	3.3000
Mean	8.7547
Median	6.5000
3rd quantile	12.0000
Standard Deviation	7.4057
Unit	Micrograms/Cubic Meter

data of 31st January (last 24 h) are not used in this section in order to test the final interpolation performance. Assume the rolling window size r equals 24, which will be further justified in Section 3.2.1, then for each station, it will have 24 h * 30 days - 24 window size = 696 samples according to the way of modeling shown in Equation (8). 33 stations will have 22,968 samples. After shuffling these samples, 70% are used for training and the remained 30% for testing. This results in 16,078 training cases and 6890 testing cases. Grid search is used for parameter optimization. Root Mean Square Error (RMSE) and R^2 are adopted as indicators to evaluate the model performance with different values of parameters. Calculations of RMSE and R^2 are presented in Equations (20) and (21). Smaller value of RMSE and larger value of R^2 mean higher prediction accuracy.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (21)$$

3.2.1. Rolling window size and top K stations

As mentioned in the previous section, the value of rolling window size r will influence the number of temporal inputs, while the number of the top K stations will impact the number of spatial inputs. Therefore, these two values will indirectly influence the modeling performance. Too small r or K cannot provide enough information for learning and predicting, while too large r or K will introduce more noise, increase modeling complexity and slow down the training speed. Therefore, to ensure the effectiveness of the model, the optimal values of these two parameters need to be determined.

Following previous studies (Azzouni and Pujolle, 2017; Li et al., 2017; Ma and Cheng, 2016c; Salman et al., 2018), this paper sets epochs as 1000, batch size as one tenth of all the training samples, and learning rate as 0.001. RMSprop is adopted as the optimizer and the number of neurons of the FC layer is set as 1024. The number of LSTM layers is initialized as 1 and its number of neurons is 32. Candidates of K are {4, 8, 12, 16, 20} and r are {3, 6, 12, 24, 48}. RMSE is used to evaluate the model performance.

By using grid search, test results for r and K are shown in Table 2 and Fig. 7. It can be seen that when K = 16 and r = 24, the model has the lowest RMSE value of 0.0639. Therefore, for the following experiments, the number of the top K stations is set as 16 and the value of rolling window size is set as 24. Also, it can be noticed that when r < 24, the RMSE gets lower as r becomes larger, and when

Table 2
RMSE performance using different K and *r*.

K / *r*	4	8	12	16	20
3	0.0900	0.0781	0.0750	0.0728	0.0747
6	0.0806	0.0748	0.0717	0.0685	0.0684
12	0.0788	0.0715	0.0694	0.0653	0.0670
24	0.0761	0.0684	0.0644	0.0639	0.0644
48	0.0762	0.0700	0.0658	0.0648	0.0644

$r \geq 24$, RMSE increases as r becomes larger. This reflects the daily period of air quality trends. It is necessary to include 24 h to get enough information for the model to learn, but keep the inputs no more than 24 h to maintain a high level of cost-effectiveness for computation.

3.2.2. Network structure and parameters

Besides K and *r*, network structure and parameters will also influence the modeling performance a lot. In this paper, the number of LSTM layers and the number of neurons of each LSTM layer are optimized for the Geo-LSTM model while other parameters are set as the optimized values in the last section. Candidates of the number of LSTM layers are selected from {1, 2, 3, 4, 5}. Following the study of Li et al. (2017) and Zhou et al. (2019), the number of neurons of each LSTM layer is set as the same. Candidate values include {16, 32, 64, 128, 256}. RMSE is used to evaluate the model performance.

Grid search results for these two parameters are presented in Table 3. Similarly, to make the observation more straightforward, the values are visualized in Fig. 8. It can be observed from Table 3 and Fig. 8 that when the number of LSTM layers is 4 and the number of neurons is 128, the model has the lowest RMSE value of 0.0437. Therefore, in the following of the study, Geo-LSTM models with 4 LSTM layers and 128 neurons in each LSTM layer are constructed.

This experiment also reflected the advantages of deep learning.

Table 3
RMSE performance of the neural networks with different number of layers and neurons.

Layers / Nodes	1	2	3	4	5
16	0.0892	0.0771	0.0827	0.0763	0.0842
32	0.0639	0.0657	0.0609	0.0609	0.0611
64	0.0598	0.0553	0.0511	0.0498	0.0489
128	0.0548	0.0471	0.0449	0.0437	0.0449
256	0.0506	0.0453	0.0464	0.0447	0.0450

It can be seen that the RMSE of the shallow network and insufficient neurons (Layer = 1, Nodes = 16) can be higher than twice as much as the optimal value (Layer = 4, Nodes = 128). This, on another angle, shows the importance of parameter tuning for neural network-based models.

4. Results and discussion

4.1. Model comparison

4.1.1. Comparison of different models for the current hour

In this paper, a geo Long Short-Term Memory (Geo-LSTM) model is designed for spatial interpolation. To prove the effectiveness of the proposed model, interpolation performance of Geo-LSTM is compared with several other models, including (1) five machine learning methods, Support Vector Regression (SVR), LASSO Regression, Ridge Regression, Gradient Boosting Decision Tree (GBDT) and Random Forest (RF), (2) three neural networks, Artificial Neural Network (ANN), Recurrent Neural Network (RNN) and the ordinary LSTM network (without geo layer), and (3) three traditional interpolation approaches, Inverse Distance Weighting (IDW), Original Kriging (OK) and Cubic Spline (CS).

Besides, as a newly designed method for spatial interpolation of air quality, this paper considers not only the spatial correlation among the unknown position to monitoring stations, but also the

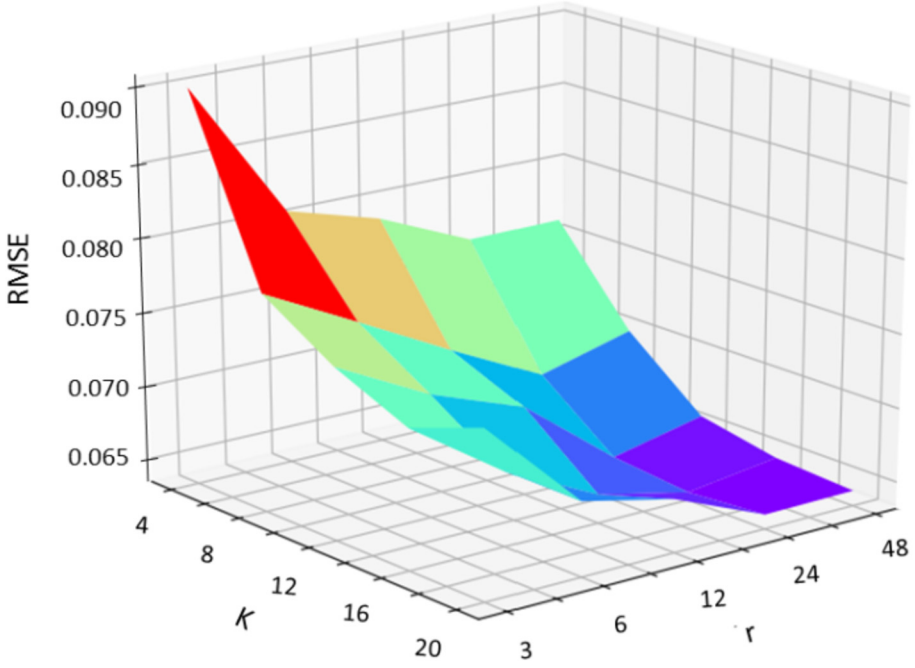

Fig. 7. Grid search results for the number of the top K stations and the value of rolling window size.

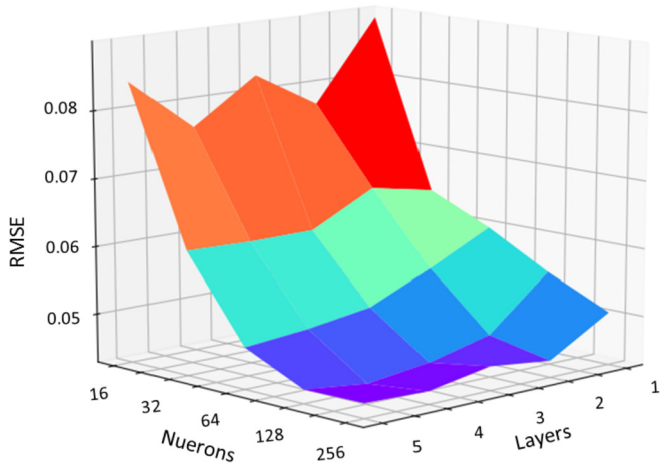

Fig. 8. Grid search results for the number of LSTM layers and the number of neurons of each LSTM layer.

temporal mechanism of air pollution. To prove the necessity of considering the temporal mechanism, two groups of comparison experiments are conducted. One is setting $r = 1$, which means that excluding the temporal mechanism but using only spatial interpolation like traditional ways. In other words, assume we were interpolating the PM2.5 value of a point at time t_{exp}, then in the first experiment we will only use the data from other monitoring stations at time t_{exp}.

The other experiment sets $r = 24$, which considered the historical influence on spatial interpolation. Still, samples of all the 33 stations are used. 70% of them are used for training and the remained 30% are used as testing sets. Note that since the r values are different, the number of training and testing cases in these two experiments are a bit different. The first experiment has 16,609 training cases and 7118 test cases, while the second experiment has 1,6078 training cases and 6890 test cases. Parameters of SVR, LASSO, Ridge, GBDT, RF, ANN, RNN and the ordinary LSTM are optimized using similar grid search techniques and cross-validation (Cheng and Ma, 2015a; Jun and Cheng, 2017). Interpolation formulas of IDW and OK are set as $S_o^t = \sum_{i=1}^{n} \lambda_i S_i^t$, where λ_i is the coefficient of the ith monitoring stations. Power value of IDW is set as 1. Estimation formula of CS, on the other hand, is set as $S_o^t = f_t(Lng_o, Lat_o)$. RMSE and R^2 are used to evaluate the model performance.

The first group sets the rolling window size as 1, which means the temporal correlation among observations is excluded. The comparison results of these models are shown in Table 4. Compared with machine learning methods and neural networks, the three traditional approaches have higher RMSE values and lower R^2 values. This is caused by the pre-set assumption of these models. For example, IDW directly resorts to the inverse of the distance to each known points when assessing weights without adjustment (Lu and Wong, 2008b). OK sets the same expected value and variance for all the points in the space (Zhang et al., 2015). CS can only achieve great interpolation performance when all the points are on the surface of the space and the surface has the smallest second-order curvature (Greco and Cuomo, 2013). However, in the real world, the spatial correlation could be dynamic. The fixed assumption is not suitable for all the cases, and therefore their performance is limited.

Table 4
Comparison results when $r = 1$.

Model type	Model	RMSE	R^2
Traditional interpolation methods	IDW	0.1096	0.1444
	OK	0.1089	0.1547
	CS	0.1440	0.0353
Machine learning methods	SVR	0.0951	0.3552
	LASSO	0.0938	0.3735
	Ridge	0.0938	0.3736
	GBDT	0.0677	0.6732
	RF	0.0776	0.5713
Neural networks	ANN	0.1085	0.1594
	RNN	0.0851	0.4845
	Ordinary LSTM	0.0848	0.4802
	Geo-LSTM	0.0756	0.5927

Machine learning methods and neural networks, on the other hand, can non-linearly model the relationship according to the inputs and outputs to fit the real situation (Cheng and Ma, 2015b). Their interpolation accuracy therefore has a higher potential to get better. It also can be observed from Table 4 that among all the models, the two tree-like models, including GBDT and RF, have relatively low values of RMSE and high values of R^2. Especially for GBDT, it has the lowest RMSE value of 0.0677 and highest R^2 value of 0.6732. Its performance is better than the proposed Geo-LSTM model. This is because this group of experiments does not take into consideration the temporal correlation among observations. The ability of the proposed model in utilizing temporal correlation therefore is not fully utilized.

To take advantage of the temporal correlation, the second group of contrast experiments is conducted. This group sets the rolling window size as 24, which is the optimal value identified in previous sections. The comparison results are shown in Table 5. RMSE and R^2 improvements of each model compared with Table 4 are also presented. Note that the three traditional interpolation approaches are not included in this group since they cannot utilize the historical time series data as a part of inputs. It can be observed from Table 5 that compared with other models, the three temporal neural networks, including RNN, ordinary LSTM and Geo-LSTM have lower RMSE values, higher R^2 values and higher improvements. This is because these three models can estimate the temporal correlation among air pollution and use it for interpolation. Tree-like models, which have lower RMSE values when $r = 1$, however, have very limited improvements when $r = 24$. This is because they are not good at modeling the temporal correlation and dependency. Furthermore, compared with the ordinary LSTM model, the proposed Geo-LSTM model has the highest RMSE improvements over 42%. This suggests that it is necessary to consider both the spatial and temporal correlation among air pollutant concentrations in order to improve the interpolation accuracy. Also note that compared with traditional methods like IDW in Table 4, the performance of Geo-LSTM in Table 5 is almost 60.13% better in RMSE (0.0437 over 0.1096).

4.1.2. Prediction performance for the next 24 hours

Furthermore, in many practical applications, interpolating the spatial distribution of air pollution for the current moment may not be satisfying enough. Predicting multi-step-ahead can sometimes provide more information for air pollution control and prevention. However, most recent studies only focused on the interpolation for the current moment (Deligiorgi and Philippopoulos, 2011; Tong et al., 2015). Few have tried to predict the spatial distribution in the next several hours in one single model. Most traditional spatial interpolation methods such as IDW, OK and CS, also are not equipped with this ability. The Geo-LSTM model designed in this paper, on the other hand, can help on this research gap. It can not only utilize the information of historical observations of monitoring stations to interpolate the spatial distribution of air pollutant in the current hour, but also interpolate the distribution in the next hours

Table 5
Comparison results when $r = 24$.

Model type	Model	RMSE	R^2	RMSE improvement*	R^2 improvement*
Machine learning methods	SVR	0.0891	0.4348	6.37%	22.40%
	LASSO	0.0915	0.4034	2.41%	7.99%
	Ridge	0.0917	0.4010	2.22%	7.35%
	GBDT	0.0642	0.7066	5.24%	4.96%
	RF	0.0757	0.5920	2.44%	3.62%
Neural networks	ANN	0.1065	0.1636	1.85%	2.63%
	RNN	0.0625	0.7219	23.53%	49.02%
	Ordinary LSTM	0.0532	0.7983	37.26%	66.23%
	Geo-LSTM	0.0437	0.8637	42.15%	45.72%

*The improvements are calculated based on the comparison with the values in Table 4 when $r = 1$.

by modifying the output neurons in the last layer. Of course, in order to evaluate the performance of the single Geo-LSTM model multi-step-ahead strategy (marked as strategy-1), we compared it with the one Geo-LSTM model one step strategy (marked as strategy-2). Details are as follows.

Strategy-2 trains multiple Geo-LSTM models. For example, to interpolate the PM2.5 concentration values in unknown places at time t, a model based on $\{[S^{RW}; RLat; RLng] \rightarrow s_t^o\}$ will be trained. Mark this model as M_t. To interpolate the spatial distribution at time t+1, another model based on $\{[S^{RW}; RLat; RLng] \rightarrow s_{t+1}^o\}$ will be built, which is marked as M_{t+1}. Therefore, to present the spatial distribution in the next 24 h, strategy-2 will train 25 Geo-LSTM models, includingM_t, M_{t+1}, …, M_{t+24}. Each model has the same network structure as optimized in previous sections but different optimal parameters optimized by its own training labels.

Strategy-1, on the other hand, trains only one model based on $\{[S^{RW}; RLat; RLng] \rightarrow [s_t^o, s_{t+1}^o, \ldots, s_{t+24}^o]\}$ to interpolate the spatial distributions in the next 24 h. This model is marked as M_{multi}. It has the same network structure and parameters of the model optimized above but 25 neurons in its output layer. Each neuron will give an output for one moment.

Interpolation performances of these two approaches are shown in Fig. 9. The results are evaluated using R-Square value. LOWESS means the locally weighted linear regression method for line smoothing. (1) It can be seen from Fig. 9 that, overall, the performance of both strategies drops a bit as the prediction time span becomes larger. This is reasonable since as the time span becomes larger, more uncertainty would be included and the prediction becomes more difficult. (2) In addition, the blue lines in Fig. 9 represent the performance of Strategy-1, while the orange lines are those of Strategy-2. It can be seen that the orange line is above the blue line in most cases. We calculated the average R-Square value of these two strategies, and found Strategy-1 has an average R-Square of 0.8580, which is 0.77% lower than the 0.8639 average value of Strategy-2. This is also understandable because for each different moment (t + i), Strategy-2 would train each with a more appropriate set of parameters based on the inputs and the labels. Spatial interpolation for each moment therefore can achieve higher prediction accuracy. Strategy-1, on the other hand, focuses on a global set of parameters instead of the hourly optimization. Its prediction accuracy for each moment therefore could be relatively lower. However, considering that Strategy-1 trained only 1 model while Strategy-2 trained 25 models, the total time of Strategy-1 spent in model training is almost 1/25 of Strategy-2. In this case, 0.77% performance reduction becomes acceptable.

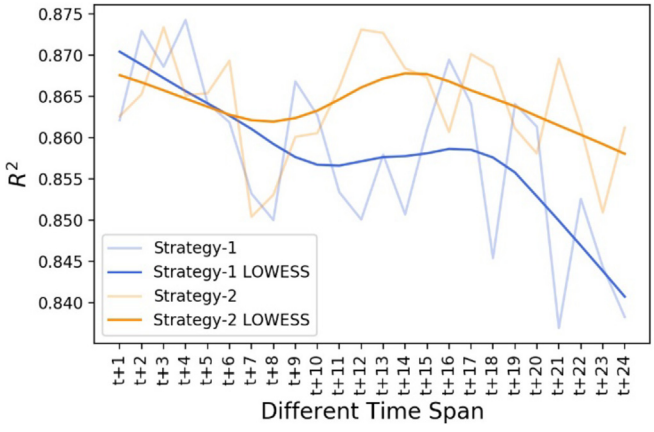

Fig. 9. Performance comparison between Strategy-1 and Strategy-2 for the next 24 h.

4.2. Discussion on extrapolated results

4.2.1. Extrapolation for the current hour

After the model optimization and comparison, the proposed Geo-LSTM model is used to extrapolate the spatial distribution of PM 2.5 concentrations in the study area. Note that interpolation and extrapolation are similar things but different ranges. Interpolation means only predict the unknown places within the spatial range of the known points, while extrapolation will additionally predict the values outside the maximum/minimum latitude and longitude of the known points. In order to draw the whole map within Washington, extrapolation is applied.

Firstly, extrapolation result of the model for the PM2.5 concentration distribution at 0am, 31st January 2017 is presented in Fig. 10. Yellow means high PM2.5 concentrations and purple means

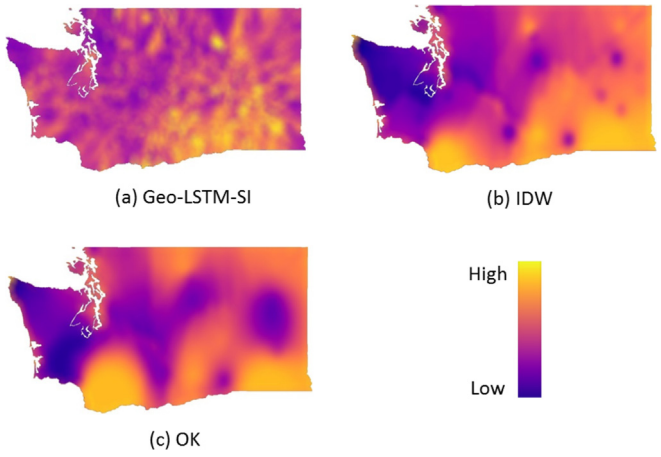

Fig. 10. Extrapolation performance of (a) Geo-LSTM-SI, (b) IDW and (c) OK for the current hour.

low concentrations. For comparison, extrapolation results of two traditional methods, IDW and OK are also presented. The data used for training and testing are the same, and no other supplementary materials were additionally included. The studied area is divided into 200 × 100 pixels with ocean area later removed, and each pixel will be predicted use these three methods. It can be observed from Table 4 that when $r = 1$, the interpolation accuracy of the proposed Geo-LSTM model is over 30% higher than that of IDW and OK. Fig. 10 further presents that compared with IDW and OK, the proposed model can present more details of the spatial distribution. Extrapolation result of IDW can only present the extreme points in places with monitoring stations. In unknown areas, the interpolated/extrapolated concentrations do not change too much. This is because in most cases, IDW takes monitored stations as extreme points. Extrapolation result given by OK is rougher and almost cannot tell details of the distribution. Its interpolation accuracy is also unsatisfied such as in the southwest and Middle East. The reason behind the better performance of Geo-LSTM are quite understandable. It can non-linearly and flexibly capture the long-short dependency of the historical data, while IDW and OK are only utilizing the information at one moment based on their preset spatial assumptions. This further proves the priority and effectiveness of the proposed Geo-LSTM model in spatial interpolation/extrapolation.

4.2.2. Extrapolation for the next 24 hours

Furthermore, the spatial distributions of PM2.5 concentrations for the 24 h of 31st January are also calculated. For a higher accuracy, the Strategy-2 in Section 4.1.2 was implemented in this experiment. For illustration purpose, extrapolation results of every 2 h are presented in Fig. 11. Based on these results, several interesting phenomena are worth discussion:

As shown in Fig. 11, the Midwestern area which is in a long-stripe shape has a low concentration of PM2.5 during the whole 24 h. By referring the topographic and land cover map in Fig. 12 (a) and (d), it can be inferred that the low PM2.5 concentration level in this area might be caused by following three reasons. Firstly, this area is a part of the Cascade Mountains. Due to the complicated topography, few people live here. Consequently, there are fewer human activities and fewer pollution emissions. Secondly, due to the high elevation, air pollution produced by surrounding areas cannot disperse to this area. Strong wind in the mountain top also can blow away pollution. Thirdly, high forest coverage in this area means the strong ability of filtering particle matters out of air.

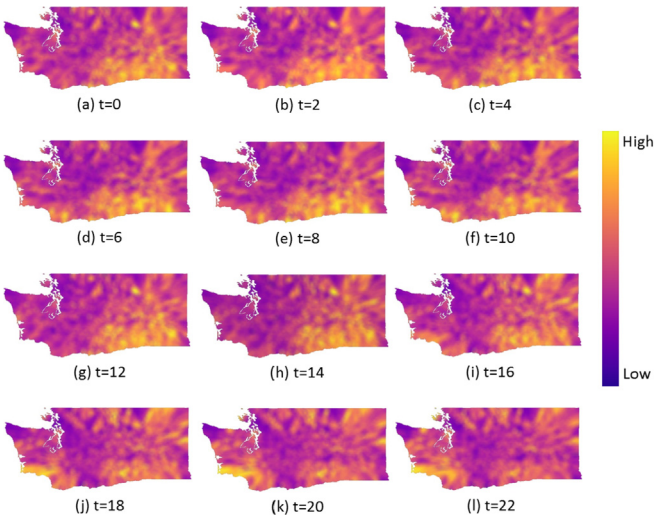

Fig. 11. Interpolation performance of Geo-LSTM-SI for the 24 h of 31st January.

The PM2.5 concentration in the southeastern area keeps at a high level. We think the high PM2.5 in this district is more from the natural phenomena instead of inappropriate human activities. The reason behind is that this district includes large areas of semiarid steppe and a few truly arid deserts. In addition, it can be seen from Fig. 12 (b) and (c) that this district has limited precipitation but relatively high wind speed. Strong wind and limited precipitation make the particulate matter from the deserts easily fly around. This results in the high level PM2.5 concentration there.

Fig. 11 also shows that in the western area, the PM 2.5 concentration is low before 16 o'clock but becomes high after 18 o'clock, especially in the southwestern area. The western area is the major area for human activities in the state. It has the city of Seattle and the city of Portland. The most important reason behind this phenomenon is the evening peak. More traffic during 18–22 o'clock caused more air pollutants.

To sum, the experiment conducted in this section shows that the proposed Geo-LSTM-SI model can not only present the details of the spatial distribution of PM2.5 concentrations for a moment, but also estimate the changing trend of a period. This could offer much useful information for city planners and governors.

5. Conclusions

To conclude, this paper proposes a geo Long Short-Term Memory (Geo-LSTM) model to interpolate the spatial distribution of air pollutants. Compared with traditional spatial interpolation methods and other machine learning-based interpolation models, the proposed model can not only non-linearly learn from the long-term dependencies of temporal sequence, but also take into consideration the spatial-temporal mechanism of air pollutants. Compared with other existing LSTM based methods, the novelty of this study lies on the proposed Geo-layer and the way of sample modeling. This helped the method be able to predicted the air quality at where there have no station records, and therefore facilitate spatial interpolation/extrapolation.

To verify the effectiveness of the proposed model, a case study is conducted in the State of Washington, US. Spatial distribution of PM2.5 concentration is interpolated. Experimental results show that:

- Compared with traditional spatial interpolation approaches such as IDW and Kriging, machine learning models, and other neural networks, the proposed Geo-LSTM model exhibits higher interpolation R-Square value and lower RMSE for spatial distribution.
- When excluding the temporal correlation among observations ($r = 1$), ensemble tree learning models such as GBDT and RF have lower RMSE values for spatial interpolation. However, when considering the temporal mechanism, the proposed model can achieve lower RMSE values and improve the interpolation accuracy by more than 40%.
- Optimization on algorithm parameters and network structures are very important for neural network-based models. According to our experiment, a proper network structure and algorithm setting can help improve the model performance by more than 50%.
- When predicting multi-step-ahead PM2.5, although separately training each model for each time step has a slightly lower RMSE and higher R-Square, it costs much more training time than using one Geo-LSTM model but modifying the output layer.
- Compared with traditional interpolation methods, the spatial distribution of PM2.5 concentrations given by Geo-

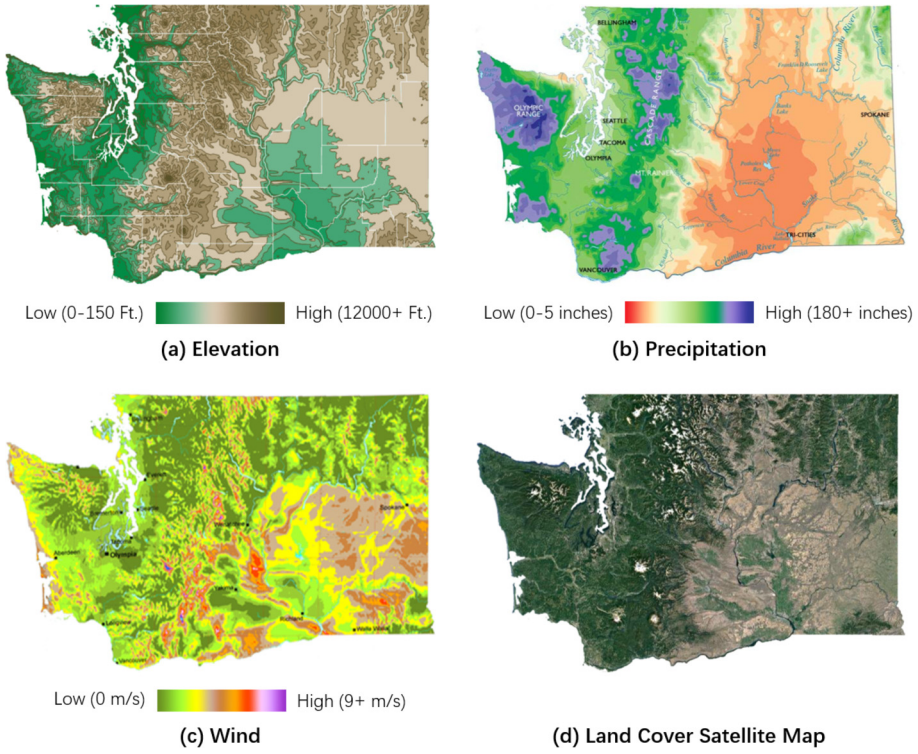

Fig. 12. Supplemental information on Washington state (Falchi et al., 2016; Geology.com, 2019; NatureMapping, 2019).

LSTM can uncover more details of the distribution and estimate the changing trend of the concentration.

- The case study conducted in Washington state shows that the Cascade Range district has a low PM2.5 concentration during the whole day, while the southeast district keeps a much higher level of PM2.5 pollution. The reasons behind may result from natural phenomena such as elevation, land cover, precipitation and wind. Also, the high PM2.5 in the western part of Washington during the night may result from the dense human activity, and this district could be the focus of air treatment for city planners.

Still, this study and the proposed method can be further improved. Due to data availability, only historical air quality concentrations were implemented for interpolation. However, air pollutant concentrations are influenced by multiple variables such as demography and meteorology. Future work therefore can include more relevant data and design a multivariate interpolation model. In addition, compared with traditional interpolation methods like IDW and Kriging, the proposed deep learning-based method is relatively data intensive and calculation expensive. Geo-LSTM requires enough historical time series as the inputs, and plenty of training time to obtain an optimal result. While methods like IDW only require the data at one moment, and no need for pre-training. Therefore, another direction of further research can focus on the efficiency of the model training with limited inputs.

References

A, V., P, G., R, V., K p, S., 2018. DeepAirNet: applying recurrent networks for air quality prediction. Procedia Comput. Sci. Int. Conf. Comput. Intell. Data Sci. 132, 1394–1403. In: https://doi.org/10.1016/j.procs.2018.05.068.

Azzouni, A., Pujolle, G., 2017. A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction. ArXiv170505690 Cs.

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. 5, 157–166. https://doi.org/10.1109/72.279181.

Chen, F.-W., Liu, C.-W., 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy Water Environ. 10, 209–222. https://doi.org/10.1007/s10333-012-0319-1.

Cheng, J.C.P., Ma, L.J., 2015a. A non-linear case-based reasoning approach for retrieval of similar cases and selection of target credits in LEED projects. Build. Environ. 93, 349–361. https://doi.org/10.1016/j.buildenv.2015.07.019.

Cheng, J.C.P., Ma, L.J., 2015b. A data-driven study of important climate factors on the achievement of LEED-EB credits. Build. Environ. 90, 232–244. https://doi.org/10.1016/j.buildenv.2014.11.029.

Chronopoulos, K.I., Tsiros, I.X., Dimopoulos, I.F., Alvertos, N., 2008. An application of artificial neural network models to estimate air temperature data in areas with sparse network of meteorological stations. J. Environ. Sci. Health Part A 43, 1752–1757. https://doi.org/10.1080/10934520802507621.

Deligiorgi, D., Philippopoulos, K., 2011. Spatial interpolation methodologies in urban air pollution modeling: application for the greater area of metropolitan athens, Greece. Adv. Air Pollut. 341–362. https://doi.org/10.5772/17734.

Engel-Cox, J.A., Hoff, R.M., Haymet, A.D.J., 2004. Recommendations on the use of satellite remote-sensing data for urban air quality. J. Air Waste Manag. Assoc. 54, 1360–1371. https://doi.org/10.1080/10473289.2004.10471005.

Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C.C., Elvidge, C.D., Baugh, K., Portnov, B.A., Rybnikova, N.A., Furgoni, R., 2016. The new world atlas of artificial night sky brightness. Sci. Adv. 2 e1600377.

FOX news, 2017. Washington's Air Quality Worst in the Nation. https://q13fox.com/2017/08/04/washingtons-air-quality-worst-in-the-nation/. accessed 6.19.19.

Geologycom, 2019. Geology and Earth Science News [WWW Document]. URL. https://geology.com. accessed 6.19.19.

Greco, L., Cuomo, M., 2013. B-Spline interpolation of Kirchhoff-Love space rods. Comput. Methods Appl. Mech. Eng. 256, 251–269. https://doi.org/10.1016/j.cma.2012.11.017.

Gupta, P., Christopher, S.A., Wang, J., Gehrig, R., Lee, Y., Kumar, N., 2006. Satellite remote sensing of particulate matter and air quality assessment over global cities. Atmos. Environ. 40, 5880–5892. https://doi.org/10.1016/j.atmosenv.2006.03.016.

Hirtl, M., Mantovani, S., Krüger, B.C., Triebnig, G., Flandorfer, C., Bottoni, M., Cavicchi, M., 2014. Improvement of air quality forecasts with satellite and ground based particulate matter observations. Atmos. Environ. 84, 20–27. https://doi.org/10.1016/j.atmosenv.2013.11.027.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, C.-S., Lin, T.-H., Hung, H., Kuo, C.-P., Ho, C.-C., Guo, Y.-L., Chen, K.-C., Wu, C.-F., 2019. Incorporating satellite-derived data with annual and monthly land use regression models for estimating spatial distribution of air pollution. Environ. Model. Softw 114, 181–187. https://doi.org/10.1016/j.envsoft.2019.01.010.

Joseph, J., Sharif, H.O., Sunil, T., Alamgir, H., 2013. Application of validation data for

assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents. Environ. Pollut. 178, 411–418. https://doi.org/10.1016/j.envpol.2013.03.035.
Jun, M.A., Cheng, J.C.P., 2017. Selection of target LEED credits based on project information and climatic factors using data mining techniques. Adv. Eng. Inf. 32, 224–236. https://doi.org/10.1016/j.aei.2017.03.004.
Lam, N.S.-N., 1983. Spatial interpolation methods: a review. Am. Cartogr. 10, 129–150. https://doi.org/10.1559/152304083783914958.
Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., Chi, T., 2017. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ. Pollut. 231, 997–1004. https://doi.org/10.1016/j.envpol.2017.08.114.
Liu, S., Zhang, Y., Ma, P., Lu, B., Su, H., 2011. A novel spatial interpolation method based on the integrated RBF neural network. Procedia Environ. Sci. 10, 568–575. https://doi.org/10.1016/j.proenv.2011.09.092.
Lu, G.Y., Wong, D.W., 2008a. An adaptive inverse-distance weighting spatial interpolation technique. Comput. Geosci. 34, 1044–1055. https://doi.org/10.1016/j.cageo.2007.07.010.
Lu, G.Y., Wong, D.W., 2008b. An adaptive inverse-distance weighting spatial interpolation technique. Comput. Geosci. 34, 1044–1055. https://doi.org/10.1016/j.cageo.2007.07.010.
Ma, J., Cheng, J.C.P., 2016a. Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. Appl. Energy 183, 182–192. https://doi.org/10.1016/j.apenergy.2016.08.079.
Ma, J., Cheng, J.C.P., 2016b. Identifying the influential features on the regional energy use intensity of residential buildings based on Random Forests. Appl. Energy 183, 193–201. https://doi.org/10.1016/j.apenergy.2016.08.096.
Ma, J., Cheng, J.C.P., 2016c. Data-driven study on the achievement of LEED credits using percentage of average score and association rule analysis. Build. Environ. 98, 121–132. https://doi.org/10.1016/j.buildenv.2016.01.005.
Ma, J., Cheng, J.C.P., 2017. Identification of the numerical patterns behind the leading counties in the U.S. local green building markets using data mining. J. Clean. Prod. 151, 406–418. https://doi.org/10.1016/j.jclepro.2017.03.083.
Martin, R.V., 2008. Satellite remote sensing of surface air quality. Atmos. Environ. 42, 7823–7843. https://doi.org/10.1016/j.atmosenv.2008.07.018.
NatureMapping, 2019. General Landcover of Washington State [WWW Document]. URL. http://naturemappingfoundation.org/natmap/maps/maps/landcover.html. accessed 6.19.19.
OLIVER, M.A., WEBSTER, R., 1990. Kriging: a method of interpolation for geographical information systems. Int. J. Geogr. Inf. Syst. 4, 313–332. https://doi.org/10.1080/02693799008941549.
Pfeiffer, H., Baumbach, G., Sarachaga-Ruiz, L., Kleanthous, S., Poulida, O., Beyaz, E., 2009. Neural modelling of the spatial distribution of air pollutants. Atmos. Environ. 43, 3289–3297. https://doi.org/10.1016/j.atmosenv.2008.05.073.
Qi, Y., Li, Q., Karimian, H., Liu, D., 2019. A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. Sci. Total Environ. 664, 1–10. https://doi.org/10.1016/j.scitotenv.2019.01.333.
QU, Y., QIAN, X., SONG, H., HE, J., LI, J., XIU, H., 2019. Machine-learning-based model and simulation analysis of PM2.5 concentration prediction in Beijing. Chin. J. Eng. 1–9.
Rigol, J.P., Jarvis, C.H., Stuart, N., 2001. Artificial neural networks as a tool for spatial interpolation. Int. J. Geogr. Inf. Sci. 15, 323–343. https://doi.org/10.1080/13658810110038951.
Salman, A.G., Heryadi, Y., Abdurahman, E., Suparta, W., 2018. Single Layer & Multi-Layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting. Procedia Comput. Sci., the 3rd International Conference on Computer Science and Computational Intelligence (ICCSCI 2018) : Empowering Smart Technology in Digital Era for a Better Life, vol. 135, pp. 89–98. In: https://doi.org/10.1016/j.procs.2018.08.153.
Tong, Y., Yu, Y., Hu, X., He, L., 2015. Performance analysis of different kriging interpolation methods based on air quality index in Wuhan. In: 2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP). Presented at the 2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP). IEEE, Wuhan, China, pp. 331–335. https://doi.org/10.1109/ICICIP.2015.7388192.
Wahid, H., Ha, Q.P., Duc, H., Azzi, M., 2013. Neural network-based meta-modelling approach for estimating spatial distribution of air pollutant levels. Appl. Soft Comput. 13, 4087–4096. https://doi.org/10.1016/j.asoc.2013.05.007.
Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., Chi, T., 2019. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. Sci. Total Environ. 654, 1091–1099. https://doi.org/10.1016/j.scitotenv.2018.11.086.
Yang, T., Liu, W., 2018. Does air pollution affect public health and health inequality? Empirical evidence from China. J. Clean. Prod. 203, 43–52. https://doi.org/10.1016/j.jclepro.2018.08.242.
Yang, L., Xu, H., Jin, Z., 2019. Estimating ground-level PM2.5 over a coastal region of China using satellite AOD and a combined model. J. Clean. Prod. 227, 472–482. https://doi.org/10.1016/j.jclepro.2019.04.231.
Zhang, L., Lu, Z., Wang, P., 2015. Efficient structural reliability analysis method based on advanced Kriging model. Appl. Math. Model. 39, 781–793. https://doi.org/10.1016/j.apm.2014.07.008.
Zhou, Y., Chang, F.-J., Chang, L.-C., Kao, I.-F., Wang, Y.-S., 2019. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. J. Clean. Prod. 209, 134–145. https://doi.org/10.1016/j.jclepro.2018.10.243.