Atmospheric Environment 286 (2022) 119257

Contents lists available at ScienceDirect

Atmospheric Environment

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

Synergistic data fusion of satellite observations and in-situ measurements for hourly $PM_{2.5}$ estimation based on hierarchical geospatial long short-term memory

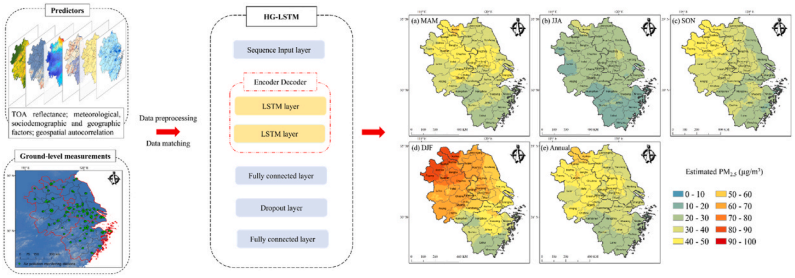
Xinyu Yu [a], Man Sing Wong [a,b,*], Chun-Ho Liu [c], Rui Zhu [a,b]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[b] Research Institute for Land and Space, The Hong Kong Polytechnic University, Hong Kong, China
[c] Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China

HIGHLIGHTS

- HG-LSTM was formed by integrating spatial autocorrelation with encoder-decoder LSTM.
- HG-LSTM performs better than other models in year-based and site-based CV tests.
- Estimated results are consistent with in-situ measurements, with an R^2 of 0.88.
- Hourly $PM_{2.5}$ values were generated with 2-km resolution seamlessly and accurately.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
$PM_{2.5}$
HG-LSTM
Geospatial autocorrelation
Encoder-decoder structure

ABSTRACT

$PM_{2.5}$ as a primary air pollutant has adverse effects on the environment and public health. The air quality monitoring stations are distributed sparsely and unevenly, making it difficult to provide continuous and precise regional measurements, which can be supplemented by satellite observations. However, most satellite-based approaches for air pollution estimation are difficult to extract the spatio-temporal dependencies effectively, leading to lower accuracy in long-term prediction and assessment of episodic changes. To fill this gap, a hierarchical geospatial long short-term memory method (HG-LSTM) by considering the geospatial autocorrelation was proposed for hourly $PM_{2.5}$ estimation with 2-km spatial resolution in Yangtze River Delta (YRD) urban agglomeration. The superior accuracy of the HG-LSTM is compared with other models via the site-based and year-based cross-validation (CV) tests, indicating geospatial autocorrelation exerts non-negligible impacts on the $PM_{2.5}$ estimation. The estimations are consistent with the in-situ observations with site-based CV R^2 of 0.88. The deviations less than 10 μ g/m^3 account for over 80%. The $PM_{2.5}$ spatiotemporal characteristics in the YRD reveal that $PM_{2.5}$ concentrations are higher in the morning and decline significantly in the afternoon. As well, elevated $PM_{2.5}$ values are accumulated in the northern regions of the study area. Although the prediction accuracy decreases as the augment of prediction timesteps, the results can still be useful to detect air pollution changes in the

* Corresponding author. Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China.
E-mail address: ls.charles@polyu.edu.hk (M.S. Wong).

https://doi.org/10.1016/j.atmosenv.2022.119257
Received 29 March 2022; Received in revised form 13 June 2022; Accepted 27 June 2022
Available online 30 June 2022

near future. Overall, the HG-LSTM model can estimate hourly $PM_{2.5}$ concentrations accurately and seamlessly, which is beneficial for air pollution monitoring and environmental protection strategy formation.

1. Introduction

$PM_{2.5}$ particles having a diameter of 2.5 μm or less constitute a major air pollutant in China, which is associated with various adverse impacts on human health (Raaschou-Nielsen et al., 2013; Kampa and Castanas, 2008). The air pollution study has attracted increasing attention, and it has also become popular to study how to improve the efficiency of air pollutants estimation and study the diffusion trend of air pollution as well as take effective environmental pollution control in advance to prevent serious air pollution incidents. High-resolution and wide-coverage $PM_{2.5}$ data enable epidemiologists to analyze the $PM_{2.5}$ impacts on human health more effectively (Di et al., 2019). Numerous ground-level monitoring stations were constructed to measure the air pollution concentrations, and the observations are generally regarded as the most reliable and precise concentrations of air pollutants at ground level (Huang et al., 2021). Nevertheless, air quality monitoring stations are sparsely and unevenly distributed, making it difficult to provide continuous and precise measurements at a large scale. By contrast, satellite-based remote sensing datasets provide spatially continuous observations, which can complement ground-level measurements (Lee et al., 2012). Therefore, satellite observations can be integrated with ground-level measurements, making full use of their advantages to acquire accurate air pollution estimation with high spatio-temporal resolution (Wong et al., 2011; Wei et al., 2019).

To date, $PM_{2.5}$ concentration prediction methods based on satellite observations are categorized into data-driven models and chemical transmission models. For chemical transmission models, the air pollution concentrations can be estimated by simulating the transit, emission, deposition, and various chemicals undergoing chemical transformations, including the Weather Research and Forecasting – Chemistry model (Sha et al., 2022), Community Multiscale Air Quality model (Wang et al., 2015) and the Goddard Earth Observing System chemical transport model (Li et al., 2020). For regions without in-situ air pollution measuring stations, chemical transmission models are frequently utilized to estimate the concentrations of air pollutants based on emission inventories and environmental conditions. Generally, a large amount of calculation is carried out based on certain theoretical assumptions and prior knowledge to derive the final predicted pollutant concentration. Moreover, the estimation accuracy is likely to hinge on the accuracy of emission inventories, while it is difficult to obtain accurate and timely emission inventories, may be resulting in deviations from the actual concentrations (Han et al., 2008; Isukapalli, 1999; Sun et al., 2013). Since data from various sensors are becoming increasingly available, estimating air pollution concentrations based on data-driven models has become a popular trend (Li and Cheng, 2021; Zang et al., 2021). For estimating $PM_{2.5}$ concentrations, non-linear and linear regression models were adopted for exploring the associations of the independent variables with air pollutants in earlier studies, including multivariate linear regression (MLR, Vlachogianni et al., 2011) and autoregressive integrated moving average models (ARIMA, Pozza et al., 2010). These methods have been widely used because of the interpretability of estimated parameters based on some prior assumptions and complex calculation processes. Whereas it is difficult to use environmental big data efficiently and investigate specific associations of air pollution with the impact factors. More importantly, these methods are more likely to result in loss of important features and accumulation of redundant information, leading to poor estimations if the environmental conditions have changed.

Recently, data derived from geospatial sources are being analyzed increasingly using deep learning and machine learning approaches for further untangling earth system problems (Reichstein et al., 2019). Satellite-derived aerosol optical depth (AOD) is commonly used for the estimations of particulate matter values (Guo et al., 2017; Hu et al., 2014). For instance, a neural network was employed to predict daily $PM_{2.5}$ concentrations based on aerosol optical depth (AOD) products by Di et al. (2016). Decision tree methods were also widely adopted to process air pollution problems. Random Forest (RF) was employed to explore the PM_{10} and $PM_{2.5}$ spatiotemporal distribution at a national scale, based on AOD data and other auxiliary datasets (Chen et al., 2018; F. Jiang et al., 2021). The RF approach for $PM_{2.5}$ estimation with a spatial resolution of 0.01°, and further quantified the associations of $PM_{2.5}$ with the impact factors in the study of Liu et al. (2018). Fan et al. (2020) developed a spatially local extreme gradient boosting for air pollutant estimation at urban scale. However, AOD products usually have a considerable amount of missing data due to cloud contamination and algorithm limitations, which tend to limit the prediction accuracy of $PM_{2.5}$. Li et al. (2022) used the multimodal AOD data fusion method to fill the missing values in the AOD product, allowing to derive near real-time $PM_{2.5}$ and PM_{10} estimation. Generally, extensive data preprocessing of the multi-source AOD product integration are required. Previous studies suggest that Top of atmosphere (TOA) reflectance has a significant correlation with AOD, and it is widely adopted to retrieve AOD (Torres et al., 2002). Some researchers have also demonstrated that it is feasible and reliable to predict $PM_{2.5}$ concentrations based on TOA reflectance (Shen et al., 2018; Yang et al., 2020). Additionally, the TOA reflectance generally has a higher spatial resolution than AOD products, thereby more detailed information of derived $PM_{2.5}$ concentrations at higher spatial resolution can be presented. Similarly, deep learning methods for air pollution estimation are extensively used. Xing et al. (2021) proposed temperature-based deep belief networks to predict daily concentrations with better performance. Wang et al. (2021a) used a long short-term memory (LSTM) method for capturing the $PM_{2.5}$ temporal dependencies, generating national $PM_{2.5}$ spatial distribution maps. However, spatiotemporal air pollution sequence estimation based on classic deep learning methods has to make improvements since the spatial information cannot be fully used. LSTM, which shows better performance in the estimation of time series sequence data than other deep learning algorithms, still lacks an understanding of geographical distance and spatial autocorrelation effects. The effectiveness of integrating spatial characteristics of $PM_{2.5}$ with LSTM for seamless and accurate $PM_{2.5}$ estimation has been presented by Wang et al. (2021b). Therefore, it is necessary to develop an air pollution estimation model that can automatically capture the representative spatial and temporal characteristics of air pollutants, to derive accurate and continuous regional air pollution estimations in a high spatiotemporal resolution.

To tackle these problems, this study aims to develop a hierarchical geospatial LSTM (HG-LSTM) model for $PM_{2.5}$ concentration estimation by synergistic data fusion of satellite datasets and ground-level measurements based on deep learning algorithms. The HG-LSTM was constructed by incorporating the geospatial autocorrelation factor in LSTM, by stacking LSTM layers according to the encoder-decoder structure, making up for the inherent defects of LSTM and enhancing the spatiotemporal sequence-to-sequence prediction capability. Comparative tests were designed following the year-based and site-based cross-validation (CV) schemes, and several popular models were also adopted to compare and assess the HG-LSTM performance. Results elucidate geospatial autocorrelation shows an important impact on $PM_{2.5}$ concentration estimation. The HG-LSTM can fully learn and explore the representative spatiotemporal features from the input sequences, and generate continuous and reliable spatiotemporal $PM_{2.5}$ distribution maps. Furthermore, our proposed model is capable of performing a multi-step prediction, which is beneficial to the government and relevant institutes

for formulating environmental protection regulations.

2. Study area and data collection

2.1. Study area

This study focuses on estimating hourly $PM_{2.5}$ concentrations over the Yangtze River Delta (YRD) region, where is a key economic zone with dense population and a large urban agglomeration in China. The boundary of the YRD region includes the mega-city Shanghai and 25 other well-industrialized and urbanized cities of more than 211,000 square kilometers. In this study, an expanded YRD urban agglomeration was defined by including Shanghai, and three provinces of Zhejiang, Jiangsu and Anhui (Fig. 1). Because of the rapid development of transportation, industry and urbanization in this region, air quality issues have attracted great concerns. Even though the air quality has been significantly improved in recent years as related regulations were introduced by the government and institutes, the recommended level of the World Health Organization (WHO) in most cities still cannot be satisfied.

2.2. Data collection

2.2.1. In-situ $PM_{2.5}$ observations

In our study, hourly $PM_{2.5}$ observations at ground level within the expanded YRD area in 2018 and 2019 can be downloaded from the China National Environmental Monitoring Center (CNEMC, http://www.cnemc.cn) and deemed as the ground truths. Fig. 1 depicted the geographical distribution of air pollutant measuring stations. Due to weather conditions or malfunction of monitoring devices, missing values of measurements are inescapable. Data from stations which missing records account for at least 30% of a year were removed. In terms of the missing values at remaining stations, linear interpolation method as a commonly used interpolation method for air pollution data (Zhao et al., 2019; Lu et al., 2021), was adopted to convert the missing data to expected values. After data preprocessing, more than 200 in-situ stations remained in the study area to provide ground truths.

2.2.2. Top of atmosphere (TOA) reflectance from AHI

The main predictor that can affect the accuracy of $PM_{2.5}$ estimation was TOA reflectance measured by the Advanced Himawari Imager (AHI), which helps observing the environment of the earth and providing rainstorm, typhoon and weather prediction, and other related applications. Generally, reflectance records from band 1 (0.46 μm), band 3 (0.64 μm) and band 6 (2.3 μm) are used to retrieve the AOD. In this study, the hourly reflectance measurements of band 1, band 3 and band 6 with 2-km spatial resolution were selected and acquired from the Japan Aerospace Exploration Agency Monitor P-Tree system (https://www.eorc.jaxa.jp/ptree/). Besides, solar azimuth angle (SOA), solar zenith angle (SOZ), satellite azimuth angle (SAA), satellite zenith angle (SAZ), these four observation angles were also taken into consideration since these parameters play important roles in retrieving AOD. The records of these variables were concatenated and deemed as main predictors in our study. The TOA reflectance data were collected under all-sky conditions. Since there are some missing values in cloud-removed reflectance, which may also affect the continuity of the regional estimated $PM_{2.5}$. Additionally, as suggested by previous studies, using original TOA reflectance to estimate $PM_{2.5}$ concentrations is a better option for non-linear deep learning and machine learning models, since they are capable of learning complex relationships between dependent and independent variables. In addition, better accuracy can be achieved when using original TOA reflectance for $PM_{2.5}$ estimation than cloud-removed TOA reflectance (Wang et al., 2021b; Yang et al., 2022). It should be noted that the AHI is a geostationary satellite and the coverage of the observation area varies in the time domain, resulting in non-negligible missing values. Therefore, the observations from 08:00 to 17:00 local time (UTC+8) were selected, corresponding to the daytime over the study region.

2.2.3. Auxiliary datasets

In addition to TOA reflectance from the AHI, other influencing factors like meteorological information, normalized difference vegetation index (NDVI), digital elevation model (DEM), and socio-demographic factors (i.e., population density and road density) were also included to improve the $PM_{2.5}$ estimation accuracy. The meteorological records can be acquired from the ERA-5 land, which is a reanalysis dataset (https://cds.climate.copernicus.eu/). Boundary layer height (BLH), surface pressure (SP), 2-m temperature (T2M), specific humidity (SH), 10-m east and north wind speed (U10M, V10M) with 0.25° × 0.25° spatial resolution and hourly temporal resolution. DEM data can be obtained at the Shuttle Radar Topography Mission (SRTM) (https://srtm.csi.cgiar.org/srtmdata/), which is a public and popular dataset with a 90-m spatial resolution and global coverage. The NDVI records were extracted from the MODIS product (MOD13A1, https://search.earthdata.nasa.gov/search). The spatial resolution is 500 m and the temporal resolution is 16 days. The population density was obtained based on the LandScan

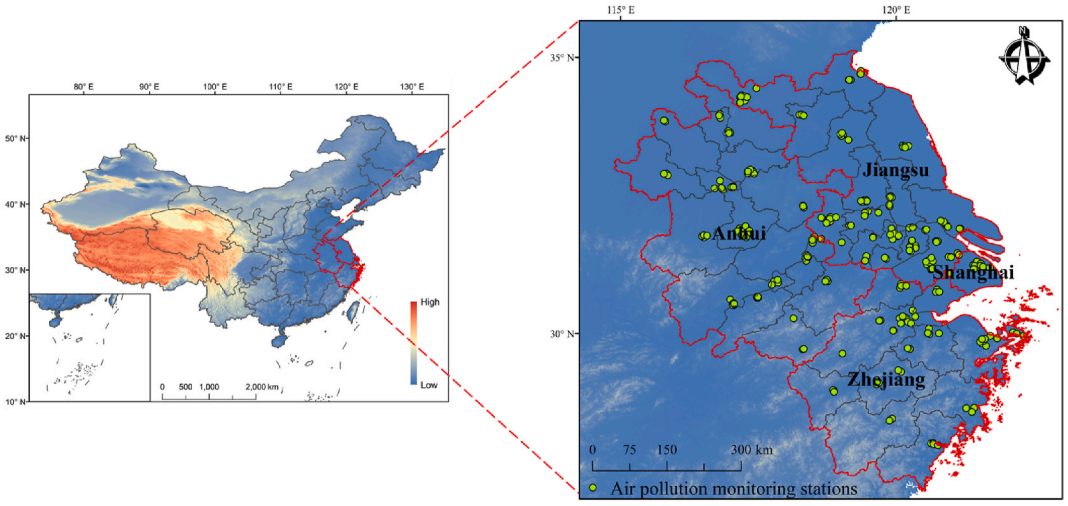


Fig. 1. Geographical distribution of air pollution monitoring stations and study region. The background maps display the elevation of the study area. Green dots denote the geospatial location of air pollution monitoring stations. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

global grid population distribution dataset (https://landscan.ornl.gov/landscan-datasets), and The road density was derived from road vector data provided by Open Street Map (https://www.openstreetmap.org/). Road density data were defined according to the total length of roads within each grid and the area of the grid.

2.3. Data preprocessing

As aforementioned, the auxiliary datasets extracted from satellite observations were from different sources, thereby resampling was indispensable to ensure consistency. The datasets were predefined to the same spatiotemporal resolution as that of the TOA reflectance by the bilinear interpolation approach. Subsequently, the whole study area was in geographic grids of about 2-km spatial resolution. Afterwards, the geospatial autocorrelation can be calculated grid by grid. Overall, the expanded YRD urban agglomeration in this study consists of 83,742 independent 0.02° grid cells. For each grid, all datasets were concatenated to estimate the surface $PM_{2.5}$, including TOA reflectance, meteorological variables, geographical data, socio-demographic factors, geospatial autocorrelation. The grid-based datasets were matched with the ground-level $PM_{2.5}$ measurements spatially and temporally. If there are several monitoring stations within a grid, the averaged in-situ $PM_{2.5}$ measurements were calculated and matched with other observations in the grid. The grids with air pollution measuring stations were extracted and employed for training and evaluating the estimated models.

3. Methodology

3.1. Long short-term memory

LSTM, which stems from the recurrent neural network (RNN), was further developed by Hochreiter and Schmidhuber (1997) to rectify the vanishing gradient problem. Different from common feedforward neural networks (i.e., Artificial Neural Network), LSTM has a set of recurrently connected blocks, which not only be used to process single input data but also sequences of data, like time-series datasets and video. A standard LSTM structure is composed of three gates, namely a forget gate, an output gate and an input gate (Fig. 2a). The forget gate (f_t) is used to decide whether to throw away unnecessary information. Then the input gate (i_t) is designed for updating the important values by a sigmoid function, and a hyperbolic tangent function is used for generating a new candidate values vector and storing them in the current cell state. Finally, filtering the output information by the output gate (o_t), simultaneously, the value was reshaped in the range of −1–1, then multiplied by the sigmoid function output via a hyperbolic tangent function. The expected output is finally obtained. The specific processes can be explained:

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

$$i_t = \sigma(W_i \cdot [h_{t-1},\ x_t] + b_i) \tag{2}$$

$$\widetilde{C}_t = tanh(W_C \cdot [h_{t-1},\ x_t] + b_C) \tag{3}$$

$$C_t = f_t{*}C_{t-1} + i_t{*}\widetilde{C}_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1},\ x_t] + b_o) \tag{5}$$

$$h_t = o_t{*}\tanh(C_t) \tag{6}$$

where o_t i_t, and f_t, represent the output gate layer, the input gate layer and the forget gate layer respectively; The input values at the t^{th} timestep was represented by x_t; h_{t-1} and h_t denote the output values at the $t-1^{th}$ and t^{th} timestep; $\sigma(\cdot)$ and $\tanh(\cdot)$ are sigmoid and hyperbolic tangent

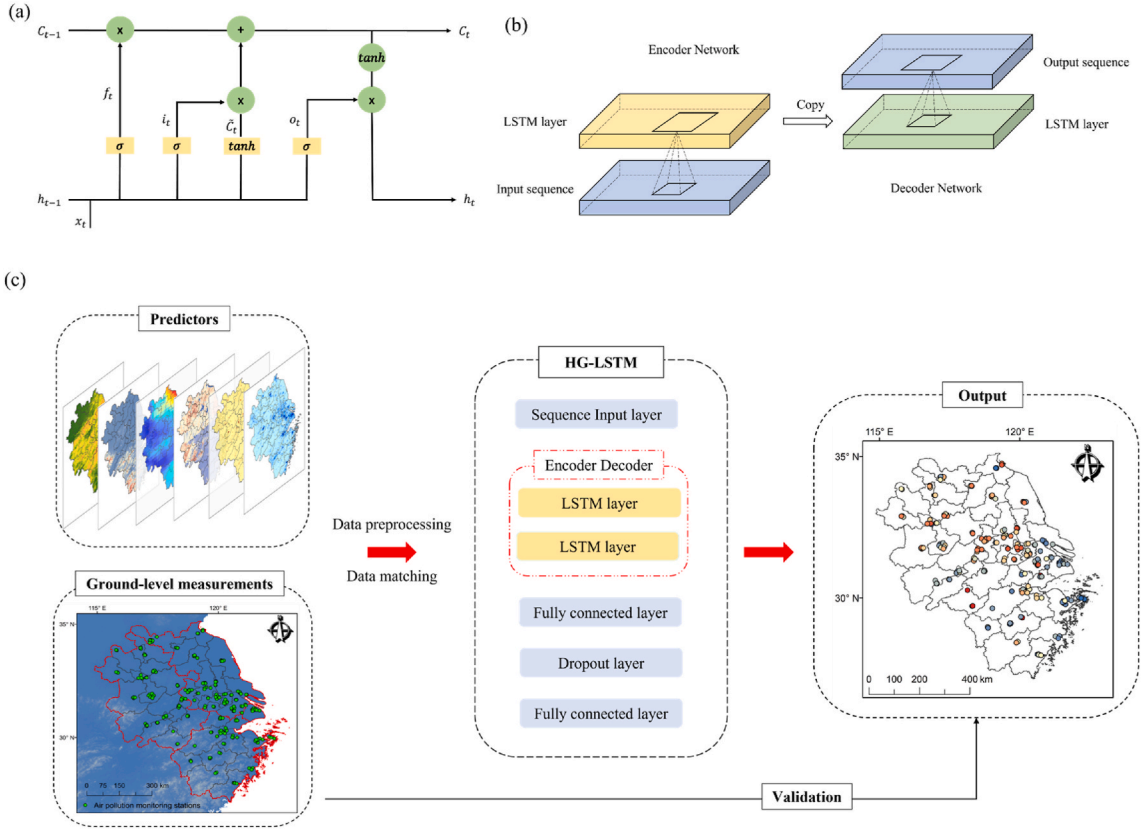


Fig. 2. (a) The architecture of an LSTM cell; (b) Encoder-decoder LSTM network structure; (c) The framework of HG-LSTM for the estimation of $PM_{2.5}$ concentrations.

function; W and b denote the weights and offsets of the corresponding layer; $\widetilde{C}_t$ and C_t are candidate values stored to cell state at the t^{th} time-step and cell state. Thanks to the special architecture, it is beneficial for learning the long-term dependencies from the input sequences.

3.2. Construction of the hierarchical geospatial long short-term memory

It is customary to note that the distribution pattern of $PM_{2.5}$ values has noticeable spatiotemporal dependencies. Specifically, at a given location and time step, $PM_{2.5}$ values of surrounding areas and measurements in the past hours or days at the same area tend to have a similar distribution pattern (Wang et al., 2021b; Li et al., 2017). LSTM has a prominent performance in processing time-series data because of its specialized structure. Nonetheless, spatial dependence is commonly ignored to some extent. Therefore, geospatial autocorrelation was considered to be introduced when training the LSTM for estimating air pollution. For each grid at a given timestep, the geospatial autocorrelation can be represented as:

$$GA = \frac{\sum_{i=1}^{n} w_{i,j} PM_{2.5,i}}{\sum_{i=1}^{n} w_{i,j}} \tag{7}$$

$$w_{i,j} = \frac{1}{d_{i,j}^2} \tag{8}$$

where GA denotes geospatial autocorrelation. It is used as an input variable and concatenated with other predictors to estimate hourly $PM_{2.5}$ values. i means the grid with $PM_{2.5}$ monitoring stations. $w_{i,j}$ is the inverse distance squared weight matrix for calculating the geospatial autocorrelation, which was constructed based on the distance between air pollution monitoring stations. $PM_{2.5,i}$ is the ground-level observations. $d_{i,j}$ is the spatial distance between grid i where a monitoring station exists and target grid j. n refers to the number of nearby stations to be considered when estimating the geospatial autocorrelation that was set to 7 according to the experiment results shown in Table S1. The optimal input variable selection scheme was also implemented and more details are discussed in the supplementary material (Table S2).

Additionally, in this study, LSTM layers were stacked according to the encoder-decoder structure, which is capable of providing sequence-to-sequence accurate prediction results (Lyu et al., 2020). The perspective of the encoder-decoder model construction is to create a vector of fixed length by encoding the input sequence, learning effective features from the input sequence, then decode the fix-length vector and get the corresponding predicted sequence. It is worthy mentioning that the length of the input sequence can be different from that of the output sequence, which can be applied to address multi-step prediction problems. In this study, a hierarchical geospatial long short-term memory (HG-LSTM) can be developed based on considering geospatial autocorrelation and stacking LSTM based on encoder-decoder architecture to predict $PM_{2.5}$ concentration in the following timestep. As shown in Fig. 2b, the encoder-decoder structure has an encoder network and a decoder network. The encoder network is implemented to learn the representative features from the input sequence. For the decoder network, which is constructed to generate the forecasted $PM_{2.5}$ values at the next timestep sequentially. The overall HG-LSTM model structure for $PM_{2.5}$ concentration prediction is displayed in Fig. 2c. The model consists of an input sequence layer and the size is the same as the dimensions of inputs. Then the encoder-decoder network follows and it comprises two LSTM layers, which have 200 cells and Rectified Linear Unit (ReLU) is the activation function. Finally, the first dense layer's cell size sets to 100 and the size of the last one is the same as the dimension of outputs for the two fully connected layers. A dropout layer is involved with a dropout rate of 0.3.

3.3. Evaluation criteria of HG-LSTM performance

To assess the effectiveness and reliability of the HG-LSTM model, two different cross-validation schemes were presented to assess the HG-LSTM performance, namely year-based and site-based CV (Liu et al., 2020). The datasets were separated into two groups by the calendar year in the year-based CV scheme. Samples in the first group and last group were used for training and testing, respectively. The training process was repeated two times to ensure each of the two groups was used exactly once as testing data. It means that the datasets in 2018 and 2019 were alternately used as the training and testing set to avoid breaking the time continuity and long-term features, because the HG-LSTM estimation accuracy is contingent upon learning long-term features from time-series datasets. Besides, to reveal the HG-LSTM spatial performance, the site-based CV method was included, and the 10-fold CV was employed to assess the performance of HG-LSTM, which is also beneficial for avoiding overfitting. To be specific, the datasets were ordered chronologically and divided into ten equal-sized subsamples based on the station locations. A single subsample was retained for testing and other sub-groups were for training. This training and testing process was repeated 10 times to guarantee that every sub-group was adopted once for testing during the processing period.

With regard to implementing a comprehensive evaluation of the HG-LSTM performance, contrast tests were constructed for evaluating the HG-LSTM effectiveness by comparing with several popular models which were widely adopted in air pollution estimation, including RF, Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN) (Zamani Joharestani et al., 2019; Maleki et al., 2019; Stafoggia et al., 2019). Random search hyperparameter optimization scheme was used to tune the contrast model hyperparameters, and detailed setups of the model hyperparameters can be found in Table S3. Additionally, multivariate linear regression (MLR) was also employed to provide a baseline reference. In addition, traditional LSTM without considering geospatial autocorrelation and encoder-decoder structure was also involved. The coefficient of determination R-squared (R^2), mean absolute error (MAE), absolute error (AE) and root-mean-square error (RMSE) were used to assess and compare the estimation accuracy between different models. Accuracy evaluation criteria can be defined as the following formulas:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{9}$$

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{11}$$

$$AE = |\hat{y}_i - y_i| \tag{12}$$

where $\hat{y}_i$ and y_i denote estimated results and ground-level $PM_{2.5}$ observations, respectively; The total number of samples is denoted by n; $\bar{y}_i$ represents averaged $PM_{2.5}$ measurements at ground level.

4. Results and discussion

4.1. Evaluation of the HG-LSTM performance for $PM_{2.5}$ estimation

To verify the HG-LSTM performance, contrast tests were implemented to compare the estimation accuracy of the HG-LSTM and other popular algorithms based on site-based and year-based CV tests. Table 1 shows the estimated results of different algorithms. As demonstrated, on account of the monotonous structure and limited learning capability, MLR elucidates the lowest accuracy with an R^2 of around 0.2. ANN performs better with a year-based CV R^2 of 0.26, RMSE of 29.02 μg/m^3,

Table 1
Contrast test results based on different models.

Model	Year-based CV (N = 781,100)			Site-based CV (N = 153,300)		
	RMSE	MAE	R^2	RMSE	MAE	R^2
MLR	31.71	22.13	0.221	28.30	20.49	0.228
ANN	29.02	19.93	0.264	27.11	19.06	0.312
RF	24.83	17.76	0.384	24.06	16.29	0.438
XGBoost	24.14	17.51	0.417	24.89	17.84	0.379
LSTM	19.62	15.38	0.583	17.26	13.53	0.686
HG-LSTM	11.85	7.67	0.867	11.34	7.31	0.878

and MAE of 19.93 μg/m^3 and RMSE of 27.11 μg/m^3, MAE of 19.06 μg/m^3, and site-based CV R^2 of 0.31. RF and XGBoost have comparable estimation results of year-based contrast tests, while RF demonstrates better performance in site-based validation tests, which tends to be attributed to the similar structure constructed based on decision-tree ensemble machine learning algorithms. Traditional LSTM presents much better accuracy of $PM_{2.5}$ concentration estimation with RMSE of 17.26 μg/m^3, site-based CV R^2 of 0.69 and MAE of 13.53 μg/m^3, indicating that LSTM is capable of processing time series datasets with better accuracy. The estimation results based on HG-LSTM are superior to other models, with RMSE of 11.34 μg/m^{3}site-based CV, R^2 of 0.88, and MAE of 7.31 μg/m^3. The results imply that taking the geospatial autocorrelation of $PM_{2.5}$ concentrations into consideration and stacking the encoder-decoder structure can ameliorate the estimation accuracy, as well as affirm the effectiveness and feasibility of our proposed model.

It can also be found that for all models, the estimation accuracy of site-based CV tests is higher than that of year-based CV tests, except XGBoost. Two possible reasons are accounting for the issue. One of the possible reasons is the discrepancy in the amount of training and testing datasets between year-based and site-based CV contrast tests. In addition, for year-based CV tests, as the training datasets are completely independent of the testing datasets, adequate features cannot be explored, which was also suggested by other researchers (Wang et al., 2021b; Wei et al., 2019). Furthermore, ANN, RF and XGBoost are classical machine learning methods, whereas LSTM and HG-LSTM can be classified as deep learning algorithms. Comparing the estimation results of all models, the $PM_{2.5}$ concentrations estimated by machine learning algorithms exhibit worse estimation accuracy than deep learning methods. The possible reason is that traditional machine learning models are confined to processing massive amounts of data and time series datasets. In contrast, LSTM and HG-LSTM can learn representative features from sequence inputs and acquire expected outputs. Additionally, after integrating the geospatial autocorrelation with LSTM, our proposed model can both explore and analyze the spatial and temporal features from the time series sequence, thereby the estimation accuracy can be substantially enhanced. Overall, the estimation results demonstrate that the geospatial autocorrelation makes paramount contributions to promoting hourly $PM_{2.5}$ estimation accuracy.

To provide a comprehensive and quantitative assessment, the discrepancies between ground-level measurements and estimated results based on different models in year-based and site-based CV tests were calculated. The values of AE were divided into eight classes, namely [0, 10], (10, 20], (20, 40], (40, 60], (60, 80], (80, 100], (100, 200] and (200, 300]. The frequencies of AE were plotted and displayed in Fig. 3. The latter three classes account for a small percentage for all models and are difficult to distinguish in the figure, thereby the values have been labeled for better comparison. As depicted, the majority of the deviations between in-situ measurements and estimated results are smaller than 10 μg/m^3 for the two CV tests, taking up from around 30%–80%. Especially for the HG-LSTM model, the error in the first range makes up around 80% of the total. Followed by the second class (10, 20] and the third class (20, 40], the results estimated by different models constitute from around 10% to nearly 30%, except HG-LSTM, whose errors located in the third range are less than 5%. For biases larger than 40 μg/m^3 but smaller than 60 μg/m^3 between ground truth and estimations derived by MLR and the machine learning models, the percentages exceed 5%. In terms of LSTM and HG-LSTM, the errors in this range are just small fractions, accounting for about 2% and 1%, respectively. Analogously, the percentages of the discrepancies in the range from 60 to 80 μg/m^3 are around 2% for MLR and the machine learning models, but for LSTM and HG-LSTM, they are less than 1%. Other than MLR in the site-based CV tests, the percentages of errors in the sixth and seventh classes are below 1% for all models. For all models, the percentages of the biases greater than 200 μg/m^3 but smaller than 300 μg/m^3 are under 0.1%, for LSTM and HG-LSTM, the numbers of the errors account for less than 0.01%. By comparing the AE distributions of LSTM and HG-LSTM models, it is obvious that the percentage of errors less than 10 μg/m^3 increases and the proportion of other classes decreases after considering the effects of geospatial autocorrelation. It demonstrates that on the basis of the temporal dependence, introducing geospatial autocorrelation is beneficial for improving the estimation accuracy. As shown in Fig. 3, the errors of site-based and year-based CV tests are comparable, but slight differences can be found in the estimations of the machine learning models. In conclusion, HG-LSTM and LSTM have better

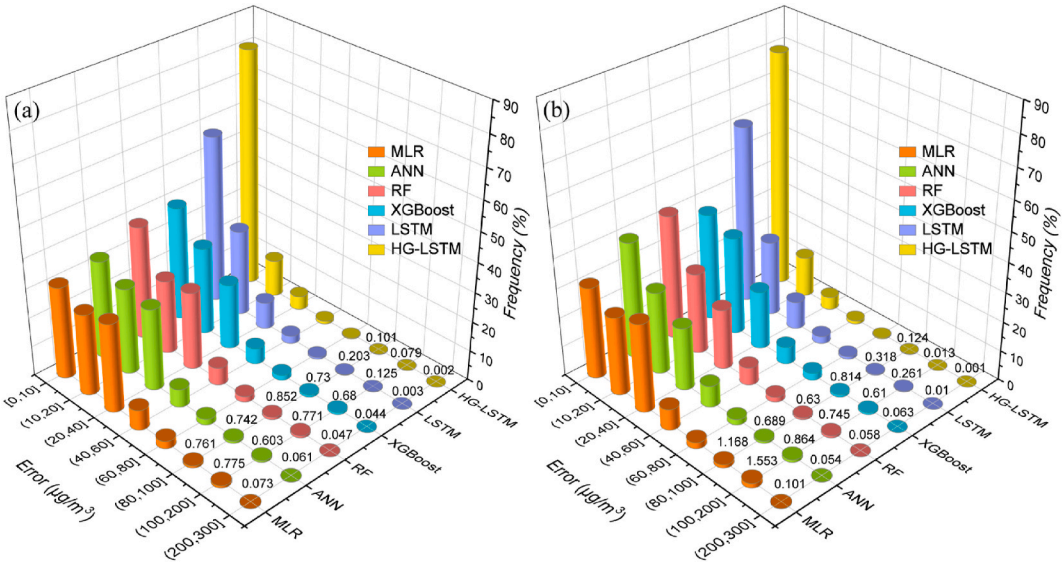


Fig. 3. Comparison of AE distribution frequencies between HG-LSTM and other models based on (a) year-based CV test results; (b) site-based CV test results.

performance for processing time series prediction. In addition, a large proportion of the deviations of the HG-LSTM are less than 10 μg/m^3, and a small portion of the errors are greater than 40 μg/m^3. The results illustrate that geospatial autocorrelation carries much weight in the $PM_{2.5}$ estimation. Additionally, integrating geospatial autocorrelation with LSTM can compensate for the inherent shortcomings of LSTM in air pollution estimation, which can enhance the accuracy and reliability by taking into account both adjacency effects and temporal dependence.

4.2. Comparison between the estimations of the HG-LSTM performance and in-situ measurements

Apart from the estimation of hourly $PM_{2.5}$ concentrations, different temporal scales including daily and monthly average estimations of the HG-LSTM model and the ground-level measurements, were calculated and analyzed to make a comprehensive understanding of the HG-LSTM model estimation performance. The density scatter plots of three temporal scales were displayed in Fig. 4. As depicted, the daily average estimation results of year-based CV tests derived by the HG-LSTM model show a better consistency than hourly estimation, with MAE of 3.89 μg/m^3, RMSE of 5.61 μg/m^3 and R^2 of 0.94. A higher R^2 and lower errors can be observed in the monthly mean estimated $PM_{2.5}$ values in Fig. 4c. Fig. 4d and e are the scatter plots of the estimated results based on site-based CV tests, and the accuracy of site-based CV contrast tests is better than that of year-based CV. RMSE, $R^2$and MAE of site-based CV at daily temporal resolution are 4.65 μg/m^3, 0.96 and 3.22 μg/m^3, respectively. Monthly average estimations of site-based CV have the best accuracy with MAE of 1.89 μg/m^3, RMSE of 2.30 μg/m^3 and R^2 of 0.98.

It can be found that for both year-based and site-based CV tests, the daily and monthly average $PM_{2.5}$ estimations have much better accuracy than hourly estimations. It in turn means the HG-LSTM shows a better accuracy at a coarser temporal resolution as averaged values can effectively remove random errors to improve the estimation accuracy (Wang et al., 2021b). Notably, the monthly average $PM_{2.5}$ estimations of site-based CV tests are extremely consistent with the observations at ground level ($R^2 = 0.98$). The results further validate the reliability of the HG-LSTM approach for $PM_{2.5}$ estimation at multiple temporal resolutions.

Additionally, the scatter plots of hourly $PM_{2.5}$ values derived from the contrast models were also displayed in Figures S1 – 2. The slopes of the linear fit in MLR, ANN, RF and XGBoost are smaller than 0.5, suggesting that the higher values (>100 μg/m^3) estimated by these models were significantly underestimated and lower values (<20 μg/m^3) were overestimated. Compared to the statistical and machine learning-based models, LSTM has better performance, while the remarked underestimation and overestimation still can be observed. The possible reason for the underestimation or overestimation may be due to inadequate samples of higher or lower $PM_{2.5}$ values, thereby the estimated models cannot fully learn and capture the representative features from the inputs. In addition, the biases between ground-level measurements and satellite-based observations also contribute to the estimated errors (Wang et al., 2021b). After considering geospatial autocorrelation, the underestimation and overestimation can be mitigated effectively, and the estimated results are more consistent with the ground truth.

Furthermore, the comparison between in-situ measurements and estimated results of the HG-LSTM model at a daily resolution for several typical stations in 2019 was plotted and displayed in Fig. 5. The selected stations are located in the megacity Shanghai, provincial capitals, boundary areas, suburban cities, well-industrialized and urbanized cities, as well, the coordinates of the stations have been added in Fig. 5. As depicted, the estimations are in overall remarkable agreement with the ground truths, specifically, RMSE, R^2 and MAE values vary in the range of 3.73–7.93 μg/m^3, 0.85–0.97 and 2.76–6.65 μg/m^3. Pronounced discrepancies can be found when higher or lower values exist in the time-series measurements, which is also suggested by other researchers (Tian et al., 2021; Wang et al., 2021b). As shown in Fig. 5, the values of $PM_{2.5}$ larger than 400 μg/m^3 can be observed usually at the beginning or end of the year. Whereas in the middle of the year, $PM_{2.5}$ concentrations are generally lower with fewer fluctuations commonly below 100 μg/m^3, where the HG-LSTM model exhibits larger prediction errors. Moreover, the maximum daily average $PM_{2.5}$ concentrations for most stations are under 300 μg/m^3, except a few stations located in well-industrialized areas. It implies that $PM_{2.5}$ concentrations have a significant geospatial autocorrelation. In summary, the estimations of the HG-LSTM model exhibit a prominent consistency with the ground-level measurements.

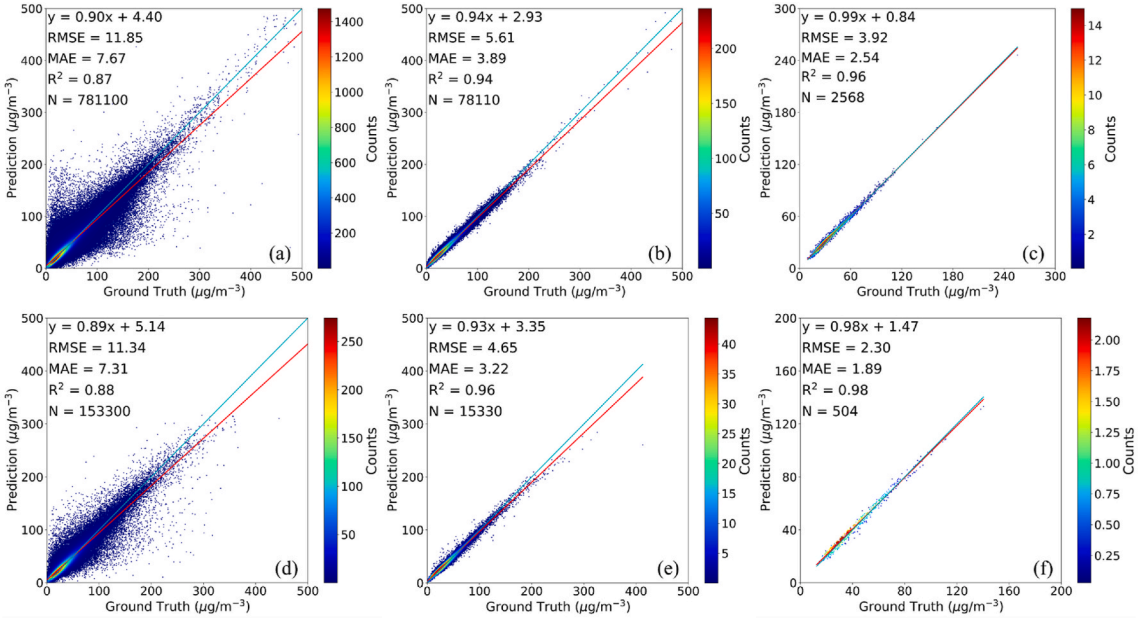


Fig. 4. The scatter plots of $PM_{2.5}$ concentrations estimated by the HG-LSTM based on year-based (2019) and site-based CV (2018–2019) tests. (a) Hourly, (b) Daily average, (c) Monthly average $PM_{2.5}$ concentrations of the year-based CV; (d) Hourly, (e) Daily average, (f) Monthly $PM_{2.5}$ concentrations of the site-based CV.

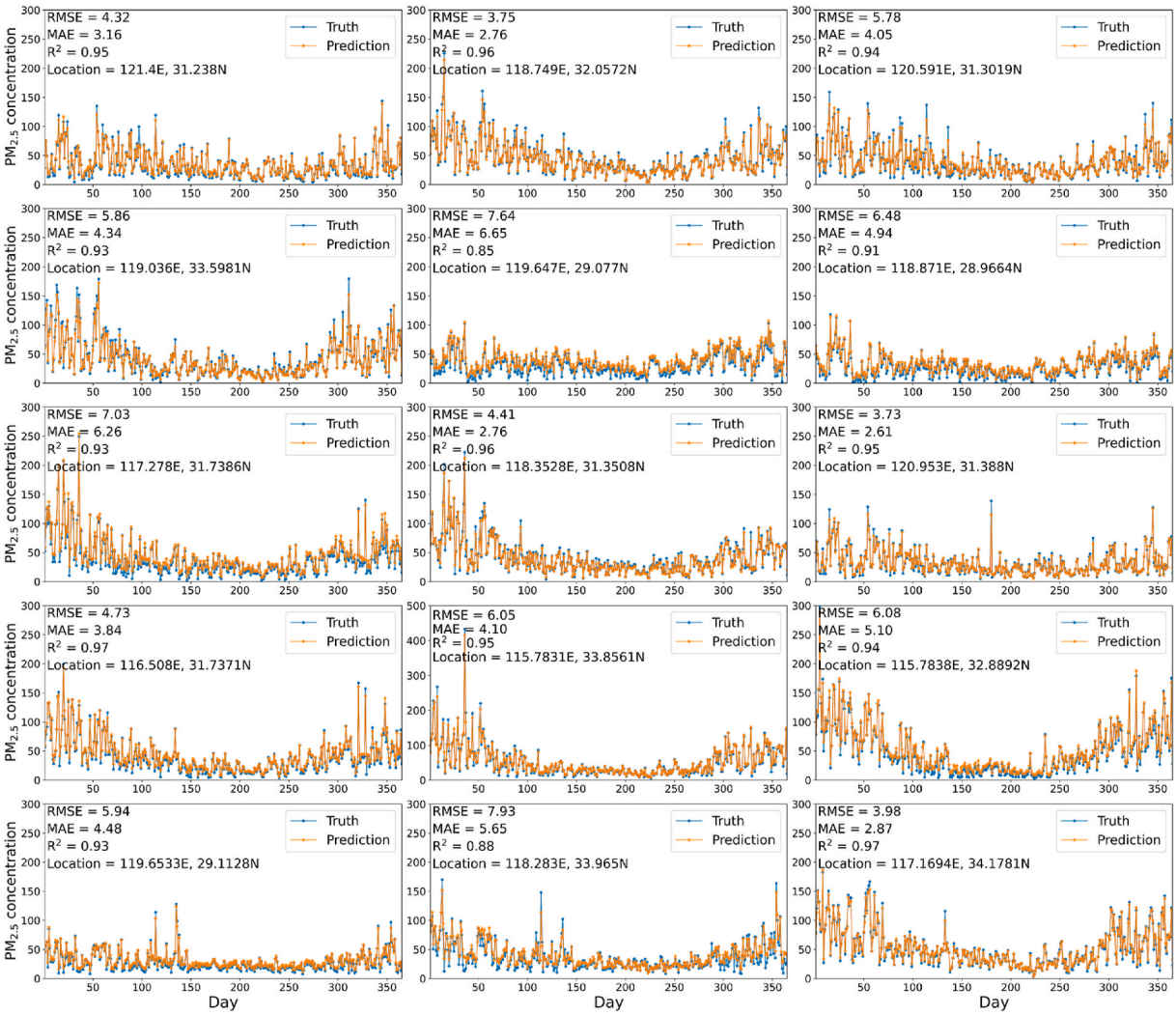


Fig. 5. Comparison of daily average $PM_{2.5}$ concentrations between estimations of the HG-LSTM and ground-level measurements in 2019. Orange lines denote the estimated results and blue lines are the ground-level measurements. The y-axis and x-axis represent, respectively, $PM_{2.5}$ concentrations, and the day of the year. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.3. *Estimated $PM_{2.5}$ spatiotemporal distribution patterns*

In addition to statistical estimation indicators, the spatial and temporal distribution patterns should be analyzed and discussed. The geographical distribution of mean estimated $PM_{2.5}$ values for each hour (00:00 UTC ~ 09:00 UTC, 08:00–17:00 local time correspondingly) in 2019 was displayed in Fig. 6. In the early morning, the highest estimated $PM_{2.5}$ values were observed and the lowest estimations are at 16:00 local time. The possible reason is that $PM_{2.5}$ concentrations from vehicles are greater in the morning peak hours. On account of the dissipation and diffusion characteristics, the $PM_{2.5}$ concentrations decrease gradually in the afternoon. As depicted, the polluted areas are in the northern regions of the YRD where the peaked concentrations are more than 70 $\mu g/m^3$. After that, there is a significant decline in the afternoon. The lowest hourly mean $PM_{2.5}$ values are in the southern area, and the city of Lishui in Zhejiang province with estimated $PM_{2.5}$ values of less than 10 $\mu g/m^3$. Lower $PM_{2.5}$ values can be discovered in the coastal regions, which is likely caused by the meteorological effects and geographical conditions (Xu et al., 2020).

To further display the predictability of the HG-LSTM model, the seasonal and annual average $PM_{2.5}$ concentrations distribution maps were derived (Fig. 7). Using a coarser time scale is beneficial to analyzing and exploring the spatio-temporal characteristics of $PM_{2.5}$ concentrations because there are more distinct spatiotemporal distribution patterns. From a temporal perspective, the most polluted season is winter, and summer has the cleanest air quality. In addition to the effects of vehicle emission and biomass burning, the changes in meteorological conditions between different seasons contribute to the distinct differences in the four seasons (Qiu et al., 2013; Hou and Wu, 2016). As shown in Fig. 7e, the annual mean $PM_{2.5}$ concentrations are smaller than 50 $\mu g/m^3$ in most regions. From a spatial perspective, similar to the hourly average $PM_{2.5}$ spatial distribution (Fig. 6), estimations of the northern regions at lower elevations are higher than those in the southern areas at higher elevations. High $PM_{2.5}$ concentration clusters can be observed in Fuyang, Bozhou, Huaibei, Suzhou and Xuzhou cities, as well. The elevated concentrations seem to extend to the surrounding cities in Anhui and Jiangsu province in winter. Significantly low-concentration clusters can be found in southern coastal cities and western areas with mountains in the study area.

According to the spatiotemporal $PM_{2.5}$ estimations, some policy implications and promotion for pollution control can be provided and discussed. First, public transportation and renewable energy should be encouraged with the purpose of reducing the air pollution especially in the morning peak hours. Analogously, anthropogenic biomass burning is supposed to be strictly controlled especially in the winter. It can be found that the northern regions of the YRD experienced severe $PM_{2.5}$

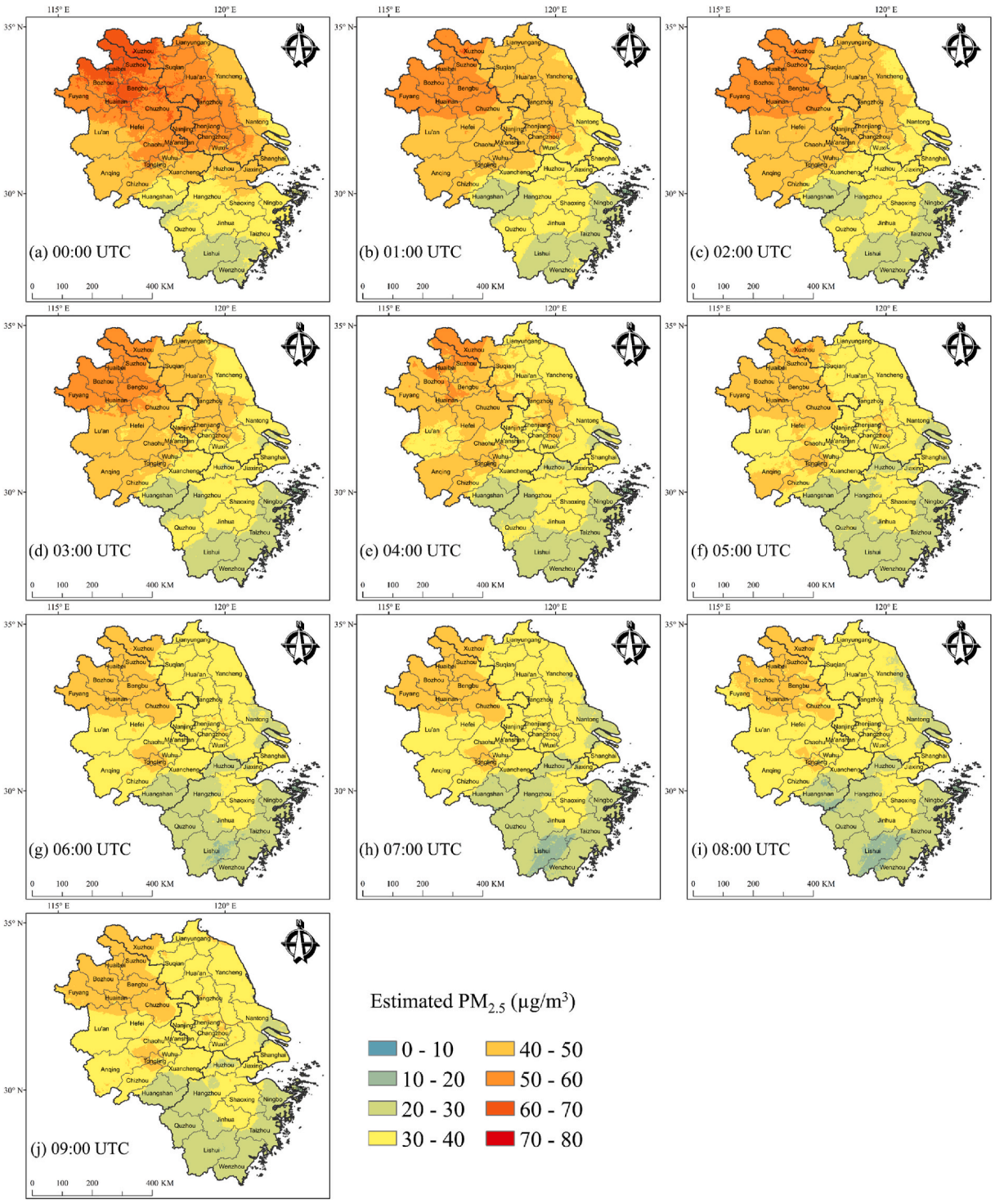


Fig. 6. Geospatial distribution of the hourly mean $PM_{2.5}$ concentrations in 2019 derived by the HG-LSTM. (a) ~ (j): 00:00–09:00 UTC.

pollution, thereby more efforts and rigorous pollution control policies should be implemented in these areas. Finally, closely cooperation between governments is also important for regional air pollution mitigation.

4.4. Application of hourly $PM_{2.5}$ estimations during a severe pollution episode

The $PM_{2.5}$ concentrations estimated by the HG-LSTM approach achieve a satisfactory accuracy with high spatial resolution. Therefore, based on the hourly $PM_{2.5}$ estimations, the spatial evolution and transportation of $PM_{2.5}$ during a severe pollution event in the YRD region can be illustrated. The spatial distribution of the estimated $PM_{2.5}$ on 25th January 2019 accompanied by the ground-level measurements was displayed in Fig. 8. This pollution event mainly erupted in the Anhui province and lasted throughout the morning. In the morning, extensive pollution can be observed in Fuyang, Bozhou, Huainan, Hefei and Chaohu cities due to the relatively high industrialization level. Higher $PM_{2.5}$ values gradually spread to the surrounding cities, leading to an increase in the pollution event. The pollution then dissipated gradually

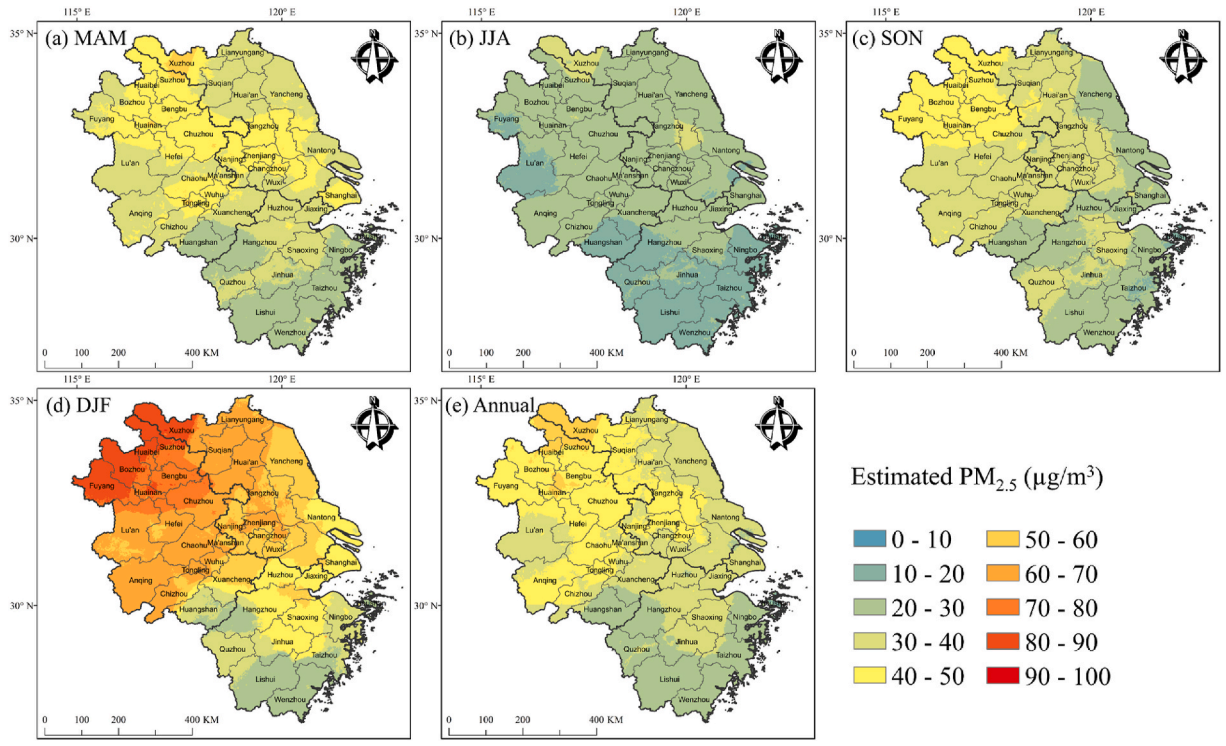


Fig. 7. Geospatial distribution of seasonal and yearly average $PM_{2.5}$ concentrations in 2019 derived by the HG-LSTM. (a) MAM; (b) JJA; (c) SON; (d) DJF; (e) Annual.

in the northwestern regions of the YRD from 06:00–09:00 UTC. Another significant pollution can be found at the border of Hangzhou and Shaoxing cities, which evolved similarly with an increase in the morning but a decrease in the afternoon. An overall decline dominated in Jiangsu province, especially the eastern border cities. Lower $PM_{2.5}$ values mainly accumulated in the southwestern cities of Zhejiang province, remaining at a stable level during the pollution event. The spatial distribution of estimated results shows a good agreement with that of ground-level measurements during the pollution period. As well, the estimated results can reveal the evolution of the pollution event more clearly than in-situ measurements.

4.5. *Multi-step prediction accuracy assessment*

As aforementioned, the HG-LSTM model exhibits satisfactory accuracy for one-step $PM_{2.5}$ concentration prediction. To further evaluate the HG-LSTM performance, multi-step prediction tests of $PM_{2.5}$ concentrations were conducted. By virtue of the HG-LSTM structure, it can be applied for sequence-to-sequence prediction and the length of the input sequence can be different from that of the output sequence. The multi-step prediction tests were implemented based on five different schemes in which the values of independent variables of the current timestep were adopted for $PM_{2.5}$ concentration prediction of the following three, five, seven, ten and twenty timesteps (t + n hours) according to year-based and site-based CV tests in Section 4.1. RMSE, MAE and R^2 were employed to evaluate the multi-step prediction performance of HG-LSTM (Table 2). It can be found that the prediction accuracy declines significantly when the prediction intervals become larger. RMSE, $R^2$and MAE values vary in the range of 14.56–24.84 μg/m^3, 0.46–0.75, 10.42–18.11 μg/m^3, respectively. Moreover, the estimation accuracy of site-based CV tests is superior to that of year-based CV tests, except for the $PM_{2.5}$ concentration prediction of the next 20 h. It should be mentioned that the estimated results based on our proposed model become less accurate and reliable when the number of timesteps increases, which is also presented by other researchers (Kow et al., 2020; Yang et al., 2019). The reasons could be attributed to the following points. Firstly, owing to the error accumulation issue, it is related to the prediction deviations with the augment of prediction timesteps (Hewamalage et al., 2021). Secondly, $PM_{2.5}$ prediction is affected by various external factors (Ma et al., 2020; T. Jiang et al., 2021). For instance, meteorological factors are regarded as commonly related external factors, which are difficult to estimate the unexpected changes, thereby becoming non-negligible obstacles for the long-term $PM_{2.5}$ prediction. In conclusion, the HG-LSTM model can provide accurate and reliable short-term $PM_{2.5}$ predictions, and the long-term prediction results can be used as a reference for future changes of air pollutants and relevant policy formulation.

5. Limitations and implications for future work

HG-LSTM was developed by integrating geospatial autocorrelation with LSTM which can effectively enhance $PM_{2.5}$ estimation accuracy. Hourly $PM_{2.5}$ values with 2-km spatial resolution over the YRD can be generated accurately and seamlessly. Nevertheless, there are several limitations that need to be improved in our future work. First, linear interpolation was used to fill the missing values of ground-level $PM_{2.5}$ measurements in our study, aiming to ensure temporal continuity for HG-LSTM estimation. Bai et al. (2020) suggested that the statistical interpolation methods are difficult to accurately recover the missing $PM_{2.5}$ values, which may induce certain biases of the predicted $PM_{2.5}$ values. Therefore, more reliable interpolation approaches should be developed and used to reduce the uncertainties of the final estimated results in our future work. As suggested by previous studies, the calculation strategies of geospatial autocorrelation and the density of in-situ monitoring stations are critical for air pollution estimations (Bai et al., 2022; Wang et al., 2022). Although geospatial autocorrelation exerts vital impacts on improving $PM_{2.5}$ estimated accuracy, there may be distortions in the spatial distribution of $PM_{2.5}$ estimated results to some extent, especially in the regions with spare ground-level monitoring stations. Therefore, we will spare no effort working on exploring the optimal full-coverage $PM_{2.5}$ mapping strategy based on spatiotemporal analysis, as well as multiscale and multisource data fusion.

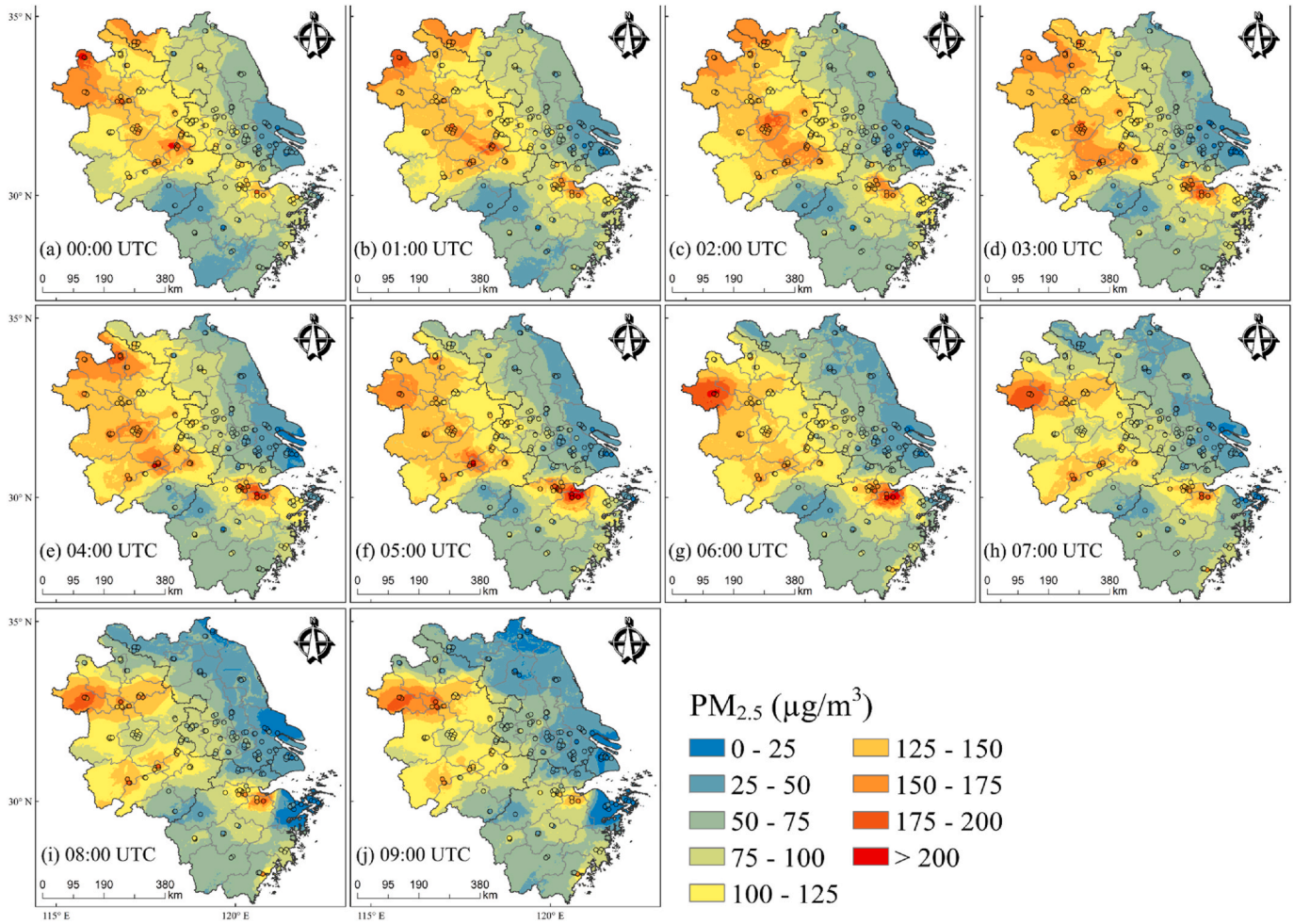


Fig. 8. An example of $PM_{2.5}$ spatial evolution during a severe pollution episode, 2019. (a)–(j): 00:00–09:00 UTC.

Table 2
Evaluation of multi-step prediction performance of HG-LSTM.

timestep (n hours)	Year-based CV			Site-based CV		
	RMSE	MAE	R^2	RMSE	MAE	R^2
3	16.85	11.59	0.710	14.56	10.42	0.751
5	19.24	12.16	0.646	18.96	11.93	0.685
7	21.87	15.14	0.595	20.78	13.09	0.622
10	23.31	16.31	0.536	22.32	15.64	0.557
20	24.28	17.55	0.477	24.84	18.11	0.460

6. Conclusion

This study presents a multi-disciplinary approach for air pollution estimation. Because of the spatiotemporal autocorrelation of the $PM_{2.5}$ distribution pattern, this study proposes a hierarchical geospatial LSTM (HG-LSTM) model by adding geospatial autocorrelation to LSTM with the encoder-decoder structure for hourly $PM_{2.5}$ concentration estimation in the YRD region on the basis of in-situ measurements and satellite observations. Contrast tests were implemented according to year-based and site-based CV tests and the performance was evaluated by comparing with other popular models. The results imply the HG-LSTM can effectively capture the representative spatiotemporal features from time series datasets after considering the geospatial autocorrelation effects, and the estimations are highly consistent with in-situ measurements for both year-based and site-based CV tests ($R^2 = 0.87$). It implies that spatial autocorrelation is critical in estimating $PM_{2.5}$ values and improves the estimation accuracy. Additionally, the spatiotemporal distribution pattern in the YRD was explored and displayed. From the temporal perspective, higher concentrations can be observed in the early morning, a declining trend in the afternoon is followed. The variations of $PM_{2.5}$ values in different seasons are significant. Winter has the highest levels followed by autumn and spring, whereas summer shows the best air quality. In terms of the spatial perspective, northern regions have larger estimated $PM_{2.5}$ values than western and southern cities in the study area. Based on the multi-step prediction tests, it can be found that the HG-LSTM model can provide accurate and reliable short-term prediction results, as well as a reference for long-term air pollutant changes. In conclusion, the HG-LSTM method is capable of estimating regional $PM_{2.5}$ values accurately and seamlessly with a high temporal and spatial resolution, and the results also illustrate the potential use for reliable estimations of other air pollutants and air quality, which is beneficial for obtaining a comprehensive understanding of air pollutant distribution and relevant policy formulation to abate the adverse effects of air pollution on human health and the environment.

CRediT authorship contribution statement

Xinyu Yu: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. **Man Sing Wong:** Conceptualization, Formal analysis, Methodology, Validation, Supervision, Funding acquisition, Writing – review & editing. **Chun-Ho Liu:** Formal analysis, Validation, Supervision, Funding acquisition, Writing – review & editing. **Rui Zhu:** Formal analysis, Validation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank the funding support from the General Research Fund (Grant No. 15603920 and 15609421), and Collaborative Research Fund (Grant No. C7064-18 GF), the Hong Kong Research Grants Council, China. M. S. Wong thanks the support from the Research Institute for Land and Space (project ID 1-CD81), The Hong Kong Polytechnic University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2022.119257.

References

Bai, K., Li, K., Guo, J., Yang, Y., Chang, N.B., 2020. Filling the gaps of in situ hourly $PM_{2.5}$ concentration data with the aid of empirical orthogonal function analysis constrained by diurnal cycles. Atmos. Meas. Tech. 13 (3), 1213–1226.

Bai, K., Li, K., Guo, J., Chang, N.B., 2022. Multiscale and multisource data fusion for full-coverage $PM_{2.5}$ concentration mapping: can spatial pattern recognition come with modeling accuracy? ISPRS J. Photogrammetry Remote Sens. 184, 31–44.

Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L.D., et al., 2018. Spatiotemporal patterns of PM_{10} concentrations over China during 2005–2016: a satellite-based estimation using the random forests approach. Environ. Pollut. 242, 605–613.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 50 (9), 4712–4721.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., et al., 2019. An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909.

Fan, Z., Zhan, Q., Yang, C., Liu, H., Bilal, M., 2020. Estimating $PM_{2.5}$ concentrations using spatially local xgboost based on full-covered SARA AOD at the urban scale. Rem. Sens. 12 (20), 3368.

Guo, J., Xia, F., Zhang, Y., Liu, H., Li, J., Lou, M., et al., 2017. Impact of diurnal variability and meteorological factors on the $PM_{2.5}$-AOD relationship: implications for $PM_{2.5}$ remote sensing. Environ. Pollut. 221, 94–104.

Han, Z., Ueda, H., An, J., 2008. Evaluation and intercomparison of meteorological predictions by five MM5-PBL parameterizations in combination with three land-surface models. Atmospheric Environment 42 (2), 233–249.

Hewamalage, H., Bergmeir, C., Bandara, K., 2021. Recurrent neural networks for time series forecasting: current status and future directions. Int. J. Forecast. 37 (1), 388–427.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hou, P., Wu, S., 2016. Long-term changes in extreme air pollution meteorology and the implications for air quality. Sci. Rep. 6 (1), 1–9.

Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Al-Hamdan, M.Z., Crosson, W.L., et al., 2014. Estimating ground-level $PM_{2.5}$ concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. Rem. Sens. Environ. 140, 220–232.

Huang, G., Li, X., Zhang, B., Ren, J., 2021. $PM_{2.5}$ concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition. Sci. Total Environ. 768, 144516.

Isukapalli, S.S., 1999. Uncertainty Analysis of Transport-Transformation Models. Rutgers The State University of New Jersey-New Brunswick.

Jiang, F., Zhang, C., Sun, S., Sun, J., 2021. Forecasting hourly PM2.5 based on deep temporal convolutional neural network and decomposition method. Appl. Soft Comput. 107988.

Jiang, T., Chen, B., Nie, Z., Ren, Z., Xu, B., Tang, S., 2021. Estimation of hourly full-coverage $PM_{2.5}$ concentrations at 1-km resolution in China using a two-stage random forest model. Atmos. Res. 248, 105146.

Kampa, M., Castanas, E., 2008. Human health effects of air pollution. Environ. Pollut. 151 (2), 362–367.

Kow, P.Y., Wang, Y.S., Zhou, Y., Kao, I.F., Issermann, M., Chang, L.C., Chang, F.J., 2020. Seamless integration of convolutional and back-propagation neural networks for regional multi-step-ahead $PM_{2.5}$ forecasting. J. Clean. Prod. 261, 121285.

Lee, H.J., Coull, B.A., Bell, M.L., Koutrakis, P., 2012. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient $PM_{2.5}$ concentrations. Environ. Res. 118, 8–15.

Li, K., Bai, K., Li, Z., Guo, J., Chang, N.B., 2022. Synergistic data fusion of multimodal AOD and air quality data for near real-time full coverage air pollution assessment. J. Environ. Manag. 302, 114121.

Li, T., Cheng, X., 2021. Estimating daily full-coverage surface ozone concentration using satellite observations and a spatiotemporally embedded deep learning approach. Int. J. Appl. Earth Obs. Geoinf. 101, 102356.

Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017. Estimating ground-level $PM_{2.5}$ by fusing satellite and station observations: a geo-intelligent deep learning approach. Geophys. Res. Lett. 44 (23), 11–985.

Li, S., Chen, L., Huang, G., Lin, J., Yan, Y., Ni, R., et al., 2020. Retrieval of surface $PM_{2.5}$ mass concentrations over North China using visibility measurements and GEOS-Chem simulations. Atmospheric Environment 222, 117121.

Liu, R., Ma, Z., Liu, Y., Shao, Y., Zhao, W., Bi, J., 2020. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: a machine learning approach. Environ. Int. 142, 105823.

Liu, Y., Cao, G., Zhao, N., Mulligan, K., Ye, X., 2018. Improve ground-level $PM_{2.5}$ concentration mapping using a random forests-based geostatistical approach. Environ. Pollut. 235, 272–282.

Lu, X., Sha, Y.H., Li, Z., Huang, Y., Chen, W., Chen, D., et al., 2021. Development and application of a hybrid long-short term memory–three dimensional variational technique for the improvement of $PM_{2.5}$ forecasting. Sci. Total Environ. 770, 144221.

Lyu, P., Chen, N., Mao, S., Li, M., 2020. LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion. Process Saf. Environ. Protect. 137, 93–105.

Ma, J., Ding, Y., Cheng, J.C., Jiang, F., Gan, V.J., Xu, Z., 2020. A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant $PM_{2.5}$ prediction. Sustain. Cities Soc. 60, 102237.

Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Birgani, Y.T., Rahmati, M., 2019. Air pollution prediction by using an artificial neural network model. Clean Technol. Environ. Policy 21 (6), 1341–1352.

Pozza, S.A., Lima, E.P., Comin, T.T., Gimenes, M.L., Coury, J.R., 2010. Time series analysis of $PM_{2.5}$ and $PM_{10-2.5}$ mass concentration in the city of Sao Carlos, Brazil. Int. J. Environ. Pollut. 41 (1–2), 90–108.

Qiu, H., Yu, I.T.S., Wang, X., Tian, L., Tse, L.A., Wong, T.W., 2013. Season and humidity dependence of the effects of air pollution on COPD hospitalizations in Hong Kong. Atmos. Environ. 76, 74–80.

Raaschou-Nielsen, O., Andersen, Z.J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., et al., 2013. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol. 14 (9), 813–822.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204.

Sha, T., Ma, X., Wang, J., Tian, R., Zhao, J., Cao, F., Zhang, Y.L., 2022. Improvement of inorganic aerosol component in $PM_{2.5}$ by constraining aqueous-phase formation of sulfate in cloud with satellite retrievals: WRF-Chem simulations. Sci. Total Environ. 804, 150229.

Shen, H., Li, T., Yuan, Q., Zhang, L., 2018. Estimating regional ground-level $PM_{2.5}$ directly from satellite top-of-atmosphere reflectance using deep belief networks. J. Geophys. Res. Atmos. 123 (24), 13–875.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., De Hoogh, K., De'Donato, F., et al., 2019. Estimation of daily PM_{10} and $PM_{2.5}$ concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. Environ. Int. 124, 170–179.

Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., Liu, S., 2013. Prediction of 24-hour-average $PM_{2.5}$ concentrations using a hidden Markov model with different emission distributions in Northern California. Sci. Total Environ. 443, 93–103.

Tian, H., Zhao, Y., Luo, M., He, Q., Han, Y., Zeng, Z., 2021. Estimating $PM_{2.5}$ from multisource data: a comparison of different machine learning models in the Pearl River Delta of China. Urban Clim. 35, 100740.

Torres, O., Bhartia, P.K., Herman, J.R., Sinyuk, A., Ginoux, P., Holben, B., 2002. A long-term record of aerosol optical depth from TOMS observations and comparison to AERONET measurements. J. Atmos. Sci. 59 (3), 398–413.

Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., Kukkonen, J., 2011. Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM_{10} in Athens and Helsinki. Sci. Total Environ. 409 (8), 1559–1571.

Wang, Y., Yuan, Q., Zhu, L., Zhang, L., 2022. Spatiotemporal estimation of hourly 2-km ground-level ozone over China based on Himawari-8 using a self-adaptive geospatially local model. Geosci. Front. 13 (1), 101286.

Wang, L., Wei, Z., Wei, W., Fu, J.S., Meng, C., Ma, S., 2015. Source apportionment of $PM_{2.5}$ in top polluted cities in Hebei, China using the CMAQ model. Atmospheric Environment 122, 723–736.

Wang, Z., Zhou, Y., Zhao, R., Wang, N., Biswas, A., Shi, Z., 2021a. High-resolution prediction of the spatial distribution of $PM_{2.5}$ concentrations in China using a long short-term memory model. J. Clean. Prod. 297, 126493.

Wang, B., Yuan, Q., Yang, Q., Zhu, L., Li, T., Zhang, L., 2021b. Estimate hourly $PM_{2.5}$ concentrations from Himawari-8 TOA reflectance directly using geo-intelligent long short-term memory network. Environ. Pollut. 271, 116327.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., Cribb, M., 2019. Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach. Rem. Sens. Environ. 231, 111221.

Wong, M.S., Nichol, J., Lee, K.H., Lee, B.Y., 2011. Monitoring 2.5 μm particulate matter within urbanized regions using satellite-derived aerosol optical thickness, a study in Hong Kong. Int. J. Rem. Sens. 32 (23), 8449–8462.

Xing, H., Wang, G., Liu, C., Suo, M., 2021. $PM_{2.5}$ concentration modeling and prediction by using temperature-based deep belief network. Neural Network. 133, 157–165.

Xu, Y., Xue, W., Lei, Y., Huang, Q., Zhao, Y., Cheng, S., et al., 2020. Spatiotemporal variation in the impact of meteorological conditions on $PM_{2.5}$ pollution in China from 2000 to 2017. Atmospheric Environment 223, 117215.

Yang, L., Xu, H., Yu, S., 2020. Estimating $PM_{2.5}$ concentrations in Yangtze River Delta region of China using random forest model and the Top-of-Atmosphere reflectance. J. Environ. Manag. 272, 111061.

Yang, M., Fan, H., Zhao, K., 2019. $PM_{2.5}$ prediction with a novel multi-step-ahead forecasting model based on dynamic wind field distance. Int. J. Environ. Res. Publ. Health 16 (22), 4482.

Yang, Q., Yuan, Q., Li, T., 2022. Ultrahigh-resolution PM2.5 estimation from top-of-atmosphere reflectance with machine learning: theories, methods, and applications. Environ. Pollut. 119347.

Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S., 2019. $PM_{2.5}$ prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. Atmosphere 10 (7), 373.

Zang, Z., Guo, Y., Jiang, Y., Zuo, C., Li, D., Shi, W., Yan, X., 2021. Tree-based ensemble deep learning model for spatiotemporal surface ozone (O_3) prediction and interpretation. Int. J. Appl. Earth Obs. Geoinf. 103, 102516.

Zhao, J., Deng, F., Cai, Y., Chen, J., 2019. Long short-term memory-Fully connected (LSTM-FC) neural network for $PM_{2.5}$ concentration prediction. Chemosphere 220, 486–492.