A Locally Weighted Neural Network Constrained by Global Training for Remote Sensing Estimation of $PM_{2.5}$

Tongwen Li, Huanfeng Shen, *Senior Member, IEEE*, Qiangqiang Yuan, *Member, IEEE*, and Liangpei Zhang, *Fellow, IEEE*

***Abstract*— Fine particulate matter ($PM_{2.5}$) pollution can cause serious public health problems worldwide. A novel geographically and temporally weighted neural network constrained by global training (GC-GTWNN) is proposed in this article for the remote sensing estimation of surface $PM_{2.5}$. The global neural network (NN) is trained to learn the overall effect of the influencing variables on surface $PM_{2.5}$, and the local geographically and temporally weighted NN (GTWNN) addresses the spatiotemporal heterogeneity of the relationship between $PM_{2.5}$ and the influencing variables. Specifically, a global NN is trained with all samples collected from the entire study domain and period. Then, initialized with the global NN, the GTWNN models are built for each location and time and fine-tuned via spatiotemporally localized samples. Meanwhile, the geographically weighted loss function is designed for GTWNN. The proposed GC-GTWNN modeling is tested with a case study across China, which integrates satellite aerosol optical depth, surface $PM_{2.5}$ measurements, and auxiliary variables. Cross-validation results indicate that a remarkable improvement is observed from the global NN to GC-GTWNN modeling (R^2 value increasing from 0.49 to 0.80), and GC-GTWNN modeling also notably outperforms the conventionally popular $PM_{2.5}$ estimation models.**

***Index Terms*— Geographically and temporally weighted neural network (GTWNN), global training constraint, remote sensing of $PM_{2.5}$.**

I. Introduction

THE $PM_{2.5}$, also known as fine particulate matter, denotes particulate matter with an aerodynamic diameter of no greater than 2.5 μm. $PM_{2.5}$ has gradually received attention because of its negative impacts on human health [1], [2]. Although $PM_{2.5}$ shows some positive effects on climate change [3], it also has great negative effects [4], [5]. Concerning the monitoring of $PM_{2.5}$, the fusion of satellite aerosol optical depth (AOD) products and surface $PM_{2.5}$ measurements is considered to be a promising approach [6]–[12]. In recent years, various AOD–$PM_{2.5}$ statistical models have been developed for the spatial estimation of surface $PM_{2.5}$ [13], [14]. In the following, two categories of AOD–$PM_{2.5}$ modeling, namely, global modeling and spatiotemporal modeling, are summarized to pave the way to our approach.

Assuming that the AOD–$PM_{2.5}$ relationship does not change in space and time, global modeling describes the AOD–$PM_{2.5}$ relationship by the constant model in the entire study domain and period. A common way of global modeling is to establish a constant model using all collected samples; in the early stage, it is often implemented on the basis of relatively simple regression techniques, such as linear regression [15], [16], multiple linear regression [17], and semiempirical model [18]. These models have gradually lost popularity because they show limited ability to address the complicated nonlinear relationship between surface $PM_{2.5}$ and the influencing variables. Afterward, machine learning algorithms, including neural network (NN) [19]–[21], random forest [22], support vector machine [23], and gradient boosting [24], have been growingly introduced to describe the nonlinear AOD–$PM_{2.5}$ relationship. AOD–$PM_{2.5}$ modeling with machine learning algorithms has been an encouraging tendency for surface $PM_{2.5}$ estimation. However, the global machine learning models rarely consider the spatiotemporal variability (or heterogeneity) of the AOD–$PM_{2.5}$ relationship and often fail to acquire accurate estimates for a local context.

On the contrary to global modeling, spatiotemporal modeling establishes the AOD–$PM_{2.5}$ relationship using location-specific and/or time-specific coefficients; that is, the AOD–$PM_{2.5}$ models vary in space and time. Thus, spatiotemporal modeling can address the spatiotemporal variability of the AOD–$PM_{2.5}$ relationship. Concerning spatiotemporal modeling, the most representative models include linear mixed-effects (LMEs) [25], [26], daily (or other time scales) geographically weighted regression (GWR) [27], [28], and geographically and temporally weighted regression (GTWR) [29], [30]. Therein, the LME model often adds a

Manuscript received November 25, 2020; revised April 4, 2021; accepted April 15, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB2102900 and in part by the Major Projects of Technological Innovation of Hubei Province under Grant 2019AAA046. *(Corresponding author: Huanfeng Shen.)*

Tongwen Li is with the School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai 519082, China (e-mail: litw8@mail.sysu.edu.cn).

Huanfeng Shen is with the School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China, also with the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China, and also with the Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, Wuhan 430079, China (e-mail: shenhf@whu.edu.cn).

Qiangqiang Yuan is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China, and also with the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: yqiang86@gmail.com).

Liangpei Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China, and also with the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TGRS.2021.3074569.

Digital Object Identifier 10.1109/TGRS.2021.3074569

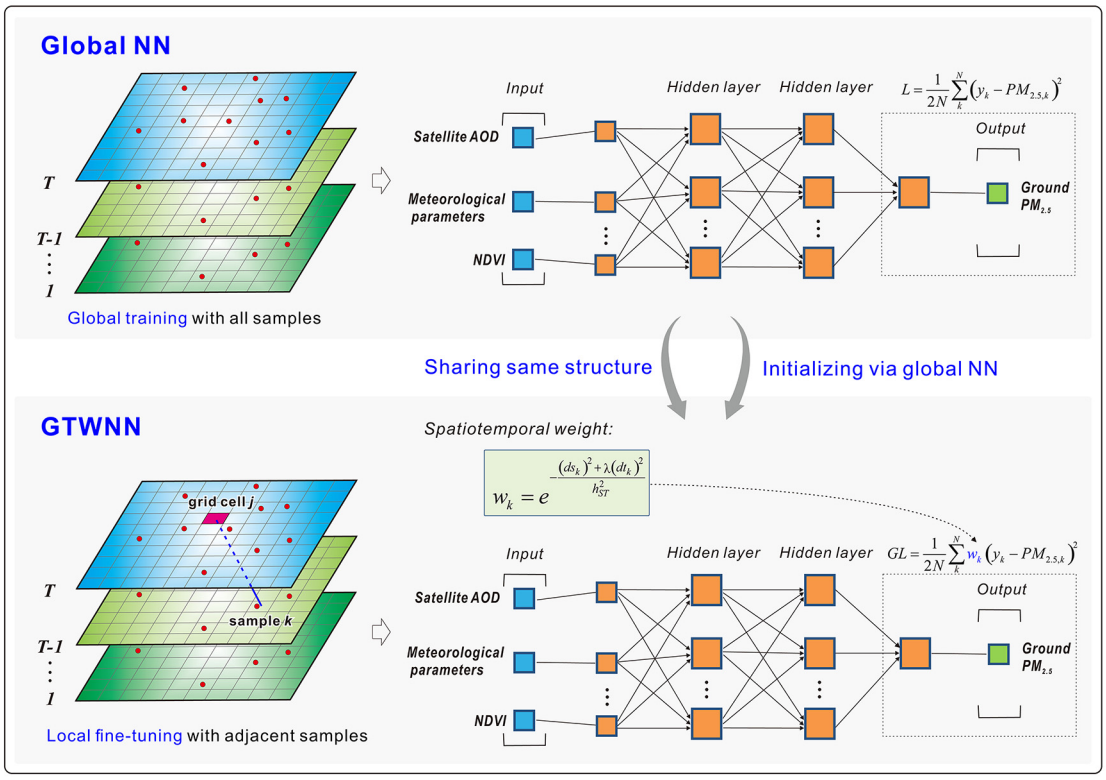


Fig. 1. Schematic of the GC-GTWNN model.

random effect for individual times to allow for the temporal variability. In addition, daily GWR and GTWR models are based on a local regression technique and simultaneously consider the spatial and temporal variability of the AOD–$PM_{2.5}$ relationship. These models have continuously received wide attention in the satellite-based $PM_{2.5}$ estimation due to the consideration of spatiotemporal variability. However, these models usually fit a linear relationship between surface $PM_{2.5}$ and the influencing variables for a certain time and/or location, and they thus may be hard to effectively address the nonlinear and complicated relationship between $PM_{2.5}$ and the influencing variables.

Overall, various $PM_{2.5}$ remote sensing estimation models have been developed to address the nonlinearity and the spatiotemporal variability of the AOD–$PM_{2.5}$ relationship. However, the nonlinear AOD–$PM_{2.5}$ relationship is adequately modeled in the global machine learning models, but not in spatiotemporal modeling. On the contrary, the spatiotemporal variability exists in the AOD–$PM_{2.5}$ relationship, which is considered by spatiotemporal modeling, while rarely considered in the machine learning models. Therefore, models that can simultaneously address the nonlinearity and the spatiotemporal variability are few; however, these models are continuously pursued in the remote sensing estimation of surface $PM_{2.5}$ [31], [32].

This study aims to address both the nonlinearity and the spatiotemporal variability by developing a geographically and temporally weighted neural network (GTWNN) model. This model represents the nonlinear AOD–$PM_{2.5}$ relationship using multilayer NNs and addresses the spatial and temporal variability of the AOD–$PM_{2.5}$ relationship based on local modeling technique. However, the establishment of this model may still encounter some challenges due to local modeling. First, the computational cost will be sharply enlarged compared with global NN modeling. Second, the data samples obtained from a local context are often limited to the NN training. Hence, a novel GTWNN constrained by global training (GC-GTWNN), in which the global NN training provides a favorable guide for the local GTWNN model training, is proposed. In this process, the GC-GTWNN model is trained with the input of the satellite AOD product, normalized difference vegetation index (NDVI) data, and meteorological parameters. This model was tested by a case study in China over one year, and daily estimates of surface $PM_{2.5}$ were effectively generated from satellite observations.

II. GC-GTWNN Modeling for Surface $PM_{2.5}$ Estimation

A novel GC-GTWNN model, which integrates surface $PM_{2.5}$ measurements, satellite AOD, meteorological parameters, and satellite NDVI, is developed to estimate the surface $PM_{2.5}$. The schematic of the GC-GTWNN model is shown in Fig. 1, which indicates that the local GTWNN model is established by the constraint of global NN training.

The GTWNN model (as shown at the bottom of Fig. 1) considers the spatiotemporal variability of the AOD–$PM_{2.5}$ relationship, and the general structure of the GTWNN model can be depicted as follows:

$$PM_{2.5} = f_j(\text{AOD, Meteorology, NDVI}) \quad (1)$$

where $f_j()$ means the location-time-specific NN for grid cell j, which varies in space and time to cope with the spatiotemporal variability of the AOD–$PM_{2.5}$ relationship. An NN model with one input layer, two hidden layers, and one output layer is shown in Fig. 1. The number of nodes in the input layer is determined by the inflecting variables, which include satellite AOD, meteorological parameters, and NDVI. The output layer owns only one node, that is, ground $PM_{2.5}$. Only those samples on day T and previous days are included for the training of $f_j()$. The observed sample near grid cell j is assumed to have a larger contribution to the estimation of $PM_{2.5}$ than those samples located farther to grid cell j. Specifically, the spatiotemporal weight for sample k can be expressed as

$$w_k = e^{-\frac{(\mathrm{ds}_k)^2+\lambda(\mathrm{dt}_k)^2}{h_{\mathrm{ST}}^2}} \tag{2}$$

where ds_k and dt_k refer to the spatial and temporal distances from sample k to prediction grid cell j, respectively; λ is a scale factor used to balance the effects of the spatial and temporal distances; h_{ST} is considered to be the bandwidth in the spatiotemporal weighting scheme [29], [33]. In the determination of the above-mentioned two parameters, there are two main schemes, namely, fixed bandwidth and adaptive bandwidth. The former has a constant distance, and the number of surrounding stations varies. In contrast, the latter maintains the same number of surrounding stations. In this study, the adaptive bandwidth scheme is used, and the nested-loop method is adopted to determine the best parameters in the process of site-based cross-validation (described in the following), that is, to achieve the best performance of cross-validation.

The spatiotemporal weights are incorporated into NN, and the geographically weighted loss function is designed for GTWNN, which embodies the novelty of GTWNN compared with the global NN, as shown in the following:

$$\mathrm{GL} = \frac{1}{2N}\sum_{k=1}^{N} w_k\left(y_k - \mathrm{PM}_{2.5,k}\right)^2 \tag{3}$$

where GL means the geographically weighted loss, and w_k is the spatiotemporal weight for sample k via the calculation of (2). y_k and $\mathrm{PM}_{2.5,k}$ refer to model the estimate and station measurements of sample k, respectively. Thus, the spatiotemporally closer samples show larger contributions to the estimation of $PM_{2.5}$. Afterward, the BP algorithm [34] is used to acquire the optimal parameter by minimizing the above geographically weighted loss function. The GTWNN model can achieve the desired result through the aforementioned process.

However, the implementation of GTWNN still encounters three main challenges. First, the design of GTWNN uses a local modeling strategy, whereas acquiring sufficient samples in a localized context for the training of GTWNN is often severe. Second, the computational cost will be enormously enlarged compared with global modeling because the GTWNN model is separately built for each location and time. Finally, GTWNN may produce unbalanced estimates for different locations and times because of the random initialization of NN.

Considering the above issues, a global NN training constraint (as shown at the top of Fig. 1) is introduced in the establishment of GTWNN (i.e., GC-GTWNN). Fig. 1 shows that a global NN model is established using all the samples (i.e., the entire year of samples collected in this study). The GTWNN models are then separately built for individual locations and times, and they are initialized via the global NN and fine-tuned with spatiotemporally localized samples. Notably, the global NN and local GTWNN models share the same network structure.

GC-GTWNN modeling can address the aforementioned challenges of GTWNN due to the following reasons. In the GC-GTWNN modeling framework, the global NN model learns the overall effect of input variables on $PM_{2.5}$ for the entire region and period, and the GTWNN merely needs to be fine-tuned to explain the temporal and spatial variations in the relationship between $PM_{2.5}$ and the influencing variables. The spatiotemporally localized samples are sufficient for the fine-tuning of the GTWNN model, and the amounts of iterations for the model training are avoided. Furthermore, all the GTWNN models are initialized with the same global NN model, which helps eliminate the unbalanced estimates.

Fig. 2 shows the procedure of GC-GTWNN modeling for remote sensing estimation of $PM_{2.5}$. First, surface $PM_{2.5}$ measurements, satellite observations, and meteorological data are collected, and the preprocessing is conducted to form a spatially and temporally consistent data set. On this basis, a global NN model (net_0) can be trained. Subsequently, initialized by the global NN (net_0), the GTWNN models are separately set up for individual locations and times and fine-tuned using spatiotemporally localized data samples. The geographically weighted loss is adopted as the optimization target during the fine-tuning process of GTWNN. Afterward, the GC-GTWNN model should be evaluated, and then daily estimates of ground $PM_{2.5}$ concentration can be generated.

The widely used tenfold site-based CV technique [35] is used to test the feasibility and prediction capability of the proposed GC-GTWNN model. All the grid cells containing monitoring stations are randomly and averagely partitioned into tenfolds. Onefold is used as the validation set, and ninefolds are adopted for model fitting. The same process will be repeated until each validation fold has been validated. Notably, the validation grid cells of monitoring stations are not used for both the global NN training and the GTWNN fine-tuning during the CV process of GC-GTWNN modeling. Meanwhile, the linear regression determination coefficient (R^2, unitless), the mean prediction error (MPE, unit: $\mu g/m^3$), the root-mean-square error (RMSE, unit: $\mu g/m^3$), and the relative prediction error (RPE, RMSE divided by the mean value of surface $PM_{2.5}$, unit: %) are calculated between model estimates and ground observations as the statistical indicators for accuracy evaluation.

Furthermore, the monitoring stations are not uniformly located in space, and the validation stations tend to be substantially close to the modeling stations. Hence, the site-based CV may face the risk of merely revealing the estimation accuracy near the modeling stations. A CV-based approach considering the nonuniform spatial distribution of monitoring

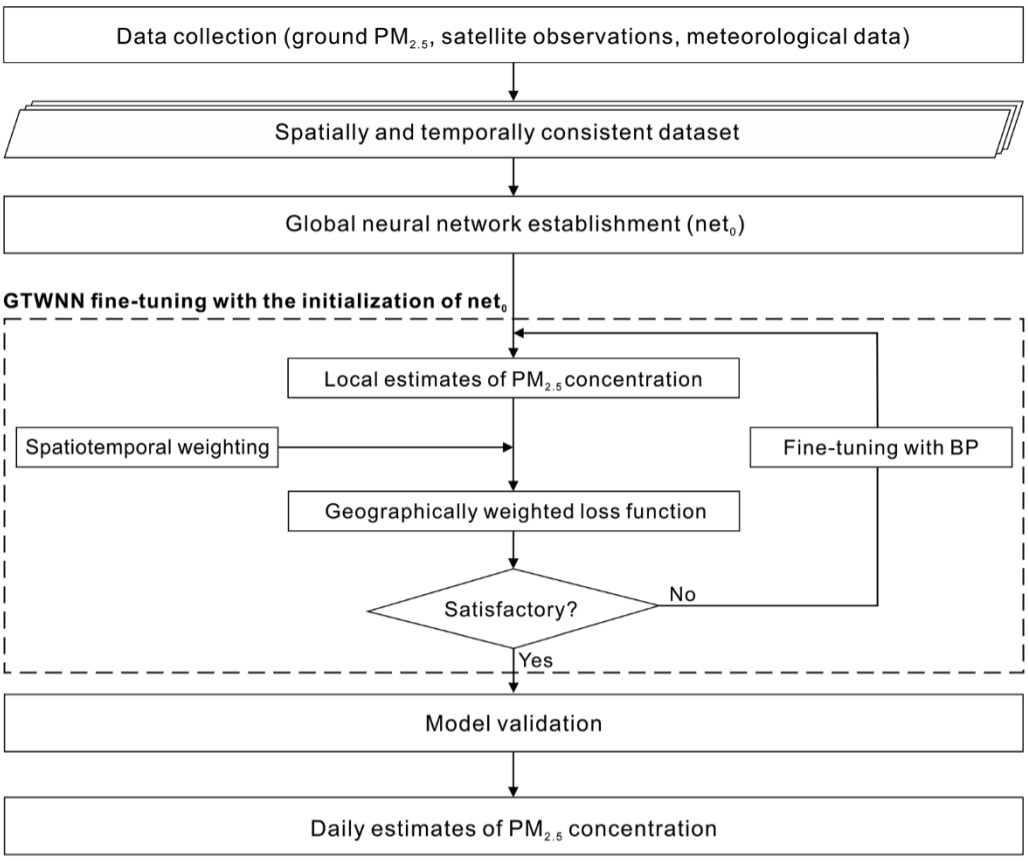


Fig. 2. Procedure for the proposed GC-GTWNN modeling.

stations (denoted as SDCV) is proposed in our previous study [36] to evaluate the model performance completely. SDCV incorporates the spatial distance between the modeling and validation stations into model validation and excludes modeling stations within a specific spatial distance. The details of SDCV can be found in [36]. SDCV is also adopted to assist in the performance evaluation of GC-GTWNN modeling, and the validation results are discussed in Section V-A.

III. Test of GC-GTWNN

A. Case Study Design

China is selected as the study region for the testing of GC-GTWNN, and the study period was the entirety of 2015. The data collected for GC-GTWNN modeling include hourly $PM_{2.5}$ surface measurements, satellite AOD product, meteorological reanalysis data, and satellite NDVI, which are comprehensively described in Section III-B. Three different modeling schemes were implemented based on the data set for comparison purposes. First, global NN was established using all the collected samples. Second, the GC-GTWNN was built with the support of the global NN. Third, the GTWNN model, which differs from GC-GTWNN in that it was not fine-tuned on the basis of the global NN, was directly trained. GC-GTWNN was also compared with other widely used $PM_{2.5}$ estimation models (i.e., daily GWR [27], [28] and GTWR [29], [30]) to further test its effectiveness. Meanwhile, the aforementioned site-based CV and SDCV approaches were exploited to evaluate the performance of these models.

B. Data and Preprocessing

1) $PM_{2.5}$ Monitoring Data: The China National Environmental Monitoring Center (CNEMC, available at http://106.37.208.233:20035/) releases the $PM_{2.5}$ monitoring data at an hourly scale, with high reliability [37]. As shown in Fig. 3, approximately 1500 stations have been established for the monitoring of air quality in China by the end of 2015. The station network is not uniformly distributed, with dense stations located in east China and sparse stations in west China. The 24-h mean $PM_{2.5}$ based on the hourly $PM_{2.5}$ measurements is obtained for the calibration of the AOD–$PM_{2.5}$ relationship [38], and those days that have less than 18 h of valid measurements were excluded from the analysis.

2) Satellite AOD Product: The well-known Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard the Terra and Aqua satellites can offer global observations in one to two days [39]. MODIS AOD product has achieved considerable popularity in the large-scale monitoring of surface $PM_{2.5}$ [40], [41]. Collection 6 of AOD product (MxD04, x is O for terra AOD and Y for Aqua AOD) was acquired at https://ladsweb.modaps.eosdis.nasa.gov/. The data field with the name of "AOD_550_Dark_Target_Deep_Blue_Combined," which has a spatial resolution of 10 km, was extracted.

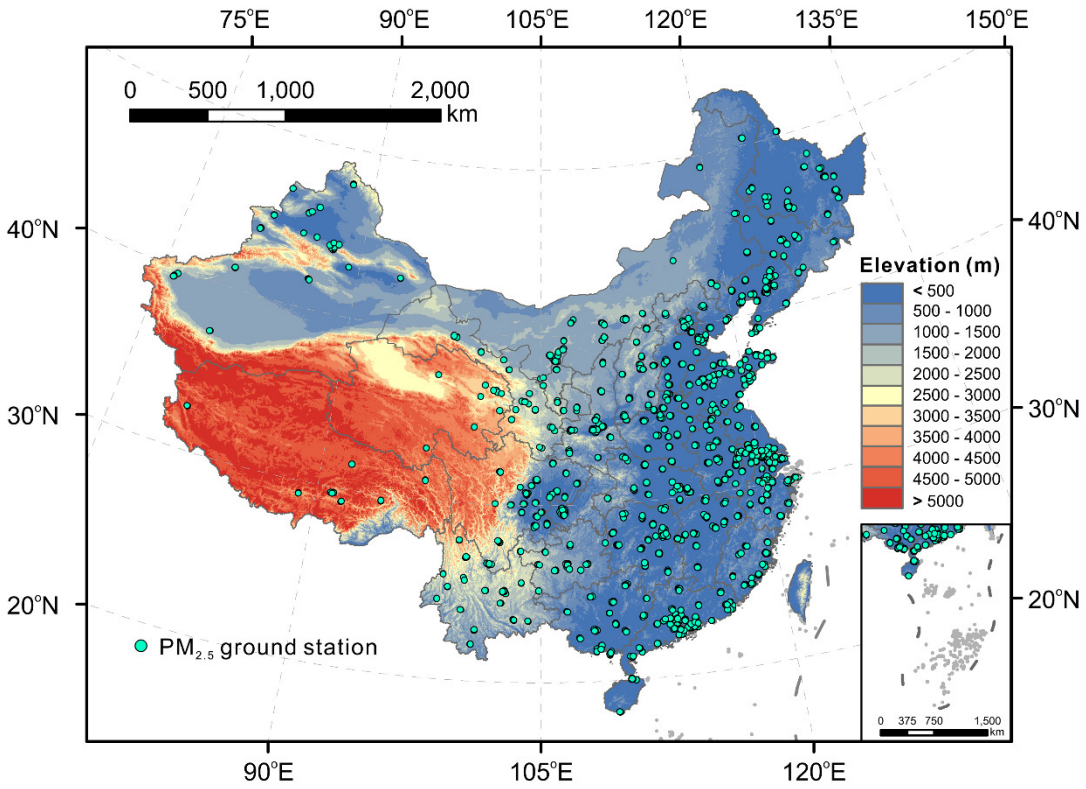


Fig. 3. Distribution of $PM_{2.5}$ stations across China.

The AOD data from Terra and Aqua MODIS were averaged in GC-GTWNN modeling to estimate the daily $PM_{2.5}$ concentrations.

3) Other Auxiliary Data: The meteorological parameters have been reported to show influences on the AOD–$PM_{2.5}$ relationship [11], [42], [43]. The meteorological variables in this study were acquired from the second Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) reanalysis data [44], whose spatial resolution is 0.5° latitude × 0.625° longitude. The relative humidity (unit: %), air temperature (unit: K), wind speed (unit: m/s), surface pressure (unit: Pa), and planetary boundary layer height (unit: m) were used for GC-GTWNN modeling. Additional details are found on the official website (http://gmao.gsfc.nasa.gov/GMAO_products/).

In addition to meteorological variables, the MODIS NDVI (MOD13) product, which has a spatial resolution of 1 km, was also incorporated as a supplementary predictor.

4) Data Preprocessing: A 0.1° × 0.1° grid was created for GC-GTWNN modeling, validating, and mapping. First, the reprojection was conducted on all the data to the WGS84 geographic coordinate system because the aforementioned data are collected from various sources with different coordinate systems. Meanwhile, the satellite AOD and NDVI products and meteorological data were resampled to 0.1° to match with the modeling grid. Multiple station measurements were averaged in each 0.1° grid for the $PM_{2.5}$ monitoring data. Finally, the satellite observations and meteorological data on the grid cells with surface $PM_{2.5}$ measurements were extracted to form the modeling sample set. After preprocessing, the correlation coefficient between satellite AOD and surface $PM_{2.5}$, which is equal to 0.48 and reports a similar correlation level with a previous national study on China [45], was calculated.

IV. Results and Analysis

A. GC-GTWNN Performance Evaluation

1) Overall Evaluation of GC-GTWNN: The global NN, the directly trained GTWNN, and the GC-GTWNN models were conducted, and the site-based CV results are shown in Fig. 4. The result reveals that the global NN obtained a frustrating performance, with R^2, RMSE, MPE, and RPE values of 0.49, 27.62 μg/m^3, 19.37 μg/m^3, and 50.08%, respectively. The global NN model considers the nonlinear relationship between surface $PM_{2.5}$ and predictors, but the nonconsideration of the spatiotemporal variability results in poor performance. The GTWNN model considers the spatial and temporal variability of the AOD–$PM_{2.5}$ relationship to overcome the drawback, and a remarkable improvement is reported, with R^2 value increasing by 0.29 (from 0.49 to 0.78) and RMSE value decreasing by 9.36 μg/m^3 (from 27.62 to 18.26 μg/m^3). Furthermore, the GC-GTWNN model achieves the most advantageous performance, and the CV R^2 and RMSE values are 0.80 and 17.44 μg/m^3, respectively. Compared with the directly trained GTWNN model, the advantage of the GC-GTWNN lies in the overall effect of the global sample set on the AOD–$PM_{2.5}$ relationship. The global training provides a favorable foundation for the fine-tuning of local models and results in improved performance. Meanwhile, another highly anticipated advantage for GC-GTWNN modeling is the improvement in efficiency, and GC-GTWNN takes

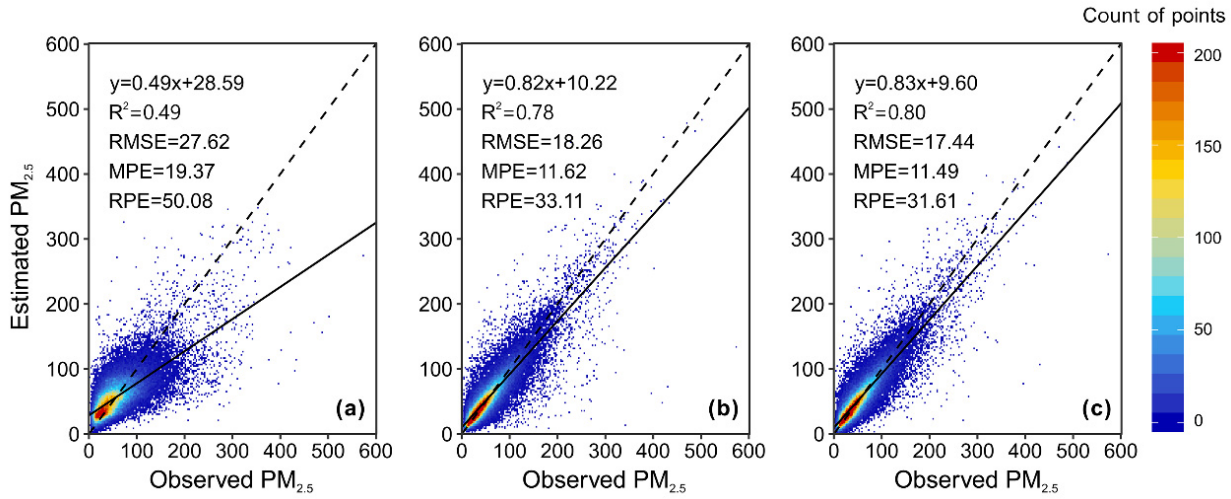


Fig. 4. Scatter plots for (a) global NN, (b) GTWNN, and (c) GC-GTWNN site-based CV results ($N = 66657$). $PM_{2.5}$ unit: $\mu g/m^3$. The dashed line stands for the $y = x$ reference line.

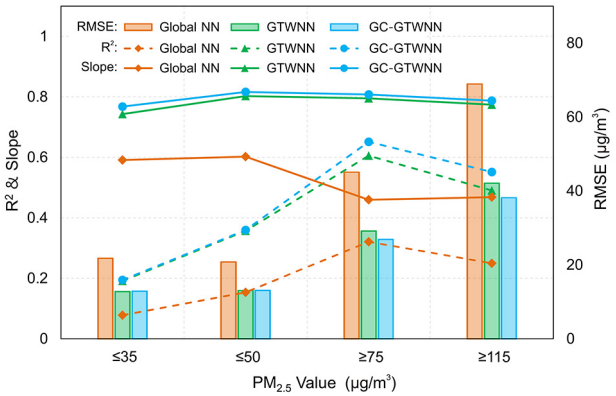


Fig. 5. Model performance for low and high values of $PM_{2.5}$ concentration.

less than a third as long as the directly trained GTWNN. This result means that realizing the timely remote sensing monitoring of $PM_{2.5}$ via GC-GTWNN modeling while improving the estimation accuracy of $PM_{2.5}$ is possible.

2) Evaluation of GC-GTWNN for Low and High $PM_{2.5}$ Values: $PM_{2.5}$ remote sensing estimation models are liable to overestimate the low values of $PM_{2.5}$ and underestimate the high $PM_{2.5}$ values. Four cases were conducted to evaluate the model performance under the conditions of low and high values. These cases include ≤ 35, ≤ 50, ≥ 75, and ≥ 115 $\mu g/m^3$, and their results are exhibited in Fig. 5. This figure shows that the global NN model performs the worst for all four cases due to the nonconsideration of spatiotemporal variability of the AOD–$PM_{2.5}$ relationship. The comparison between GTWNN and GC-GTWNN is summarized as follows. First, GC-GTWNN modeling merely reports a slight advantage over the GTWNN model for low values of $PM_{2.5}$ in the evaluation of R^2 and RMSE metrics, whereas it has a linear regression slope closer to 1. Second, a relatively more considerable superiority is observed when comparing the GC-GTWNN model with the GTWNN model considering high $PM_{2.5}$ values. The R^2, RMSE, and slope values for GC-GTWNN modeling are 0.65, 26.85 $\mu g/m^3$, and 0.81, respectively, in the case of ≥ 75 $\mu g/m^3$. Meanwhile, the corresponding metrics for the GTWNN model are 0.61, 29.11 $\mu g/m^3$, and 0.79. Thus, the GC-GTWNN model suffers from a lesser degree of overestimation and/or underestimation for low/high values of surface $PM_{2.5}$ concentrations.

3) Temporal Evaluation of GC-GTWNN: Table I presents the seasonal statistics of site-based CV results for the global NN, GTWNN, and GC-GTWNN models. Among the three models, the global NN model performs the worst for the four seasons, with R^2 values of 0.33, 0.39, 0.44, and 0.51 for spring (March to May), summer (June to August), autumn (September to November), and winter (December to February), respectively. The seasonal RMSE values of the global NN are 25.85, 20.05, 26.74, and 35.65 $\mu g/m^3$. A notable improvement is observed from the global NN to GTWNN for all the four seasons, with R^2 value increasing by 0.39 (from 0.33 to 0.72), 0.26 (from 0.39 to 0.65), 0.30 (from 0.44 to 0.74), and 0.29 (from 0.51 to 0.80), respectively. The GC-GTWNN model demonstrated its superiorities in CV performance for all the four seasons, and the seasonal R^2 (RMSE, slope) values are 0.74 (15.83 $\mu g/m^3$, 0.79), 0.68 (13.99 $\mu g/m^3$, 0.76), 0.76 (17.40 $\mu g/m^3$, 0.79), and 0.82 (21.82 $\mu g/m^3$, 0.84), respectively. From another perspective, winter reported the highest R^2 value among the four seasons, which indicates the good fit of the models in winter. However, winter also has the highest RMSE values, which can be attributed to the high levels of $PM_{2.5}$ concentration in winter.

Fig. 6 exhibits the daily mean observations and model estimates of $PM_{2.5}$ based on the CV results. First, the minimum and maximum values of daily mean $PM_{2.5}$ observations are 17.33 and 152.68 $\mu g/m^3$, respectively. Meanwhile, as observed in the gray shading, the $PM_{2.5}$ values present apparent spatial variations (large standard deviation) within China. Second, the global-NN-based estimates present notable deviations against the surface $PM_{2.5}$ observations, whereas the GTWNN and the GC-GTWNN estimates show substantially similar trends to surface $PM_{2.5}$ measurements, as indicated

TABLE I

SEASONAL STATISTICS OF SITE-BASED CV RESULTS FOR THE GLOBAL NN, GTWNN, AND GC-GTWNN MODELS. THE EQUATION STANDS FOR THE FIT LINE BETWEEN MODEL ESTIMATES AND SURFACE OBSERVATIONS

Season	Global NN		GTWNN		GC-GTWNN	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
Spring	0.33	25.85	0.72	16.33	0.74	15.83
	$y = 0.41x + 33.62$		$y = 0.77x + 12.09$		$y = 0.79x + 11.45$	
Summer	0.39	20.05	0.65	14.66	0.68	13.99
	$y = 0.48x + 26.01$		$y = 0.71x + 11.96$		$y = 0.76x + 10.23$	
Autumn	0.44	26.74	0.74	18.37	0.76	17.40
	$y = 0.40x + 29.80$		$y = 0.78x + 11.52$		$y = 0.79x + 10.59$	
Winter	0.51	35.65	0.80	22.94	0.82	21.82
	$y = 0.52x + 32.05$		$y = 0.84x + 13.00$		$y = 0.84x + 12.33$	

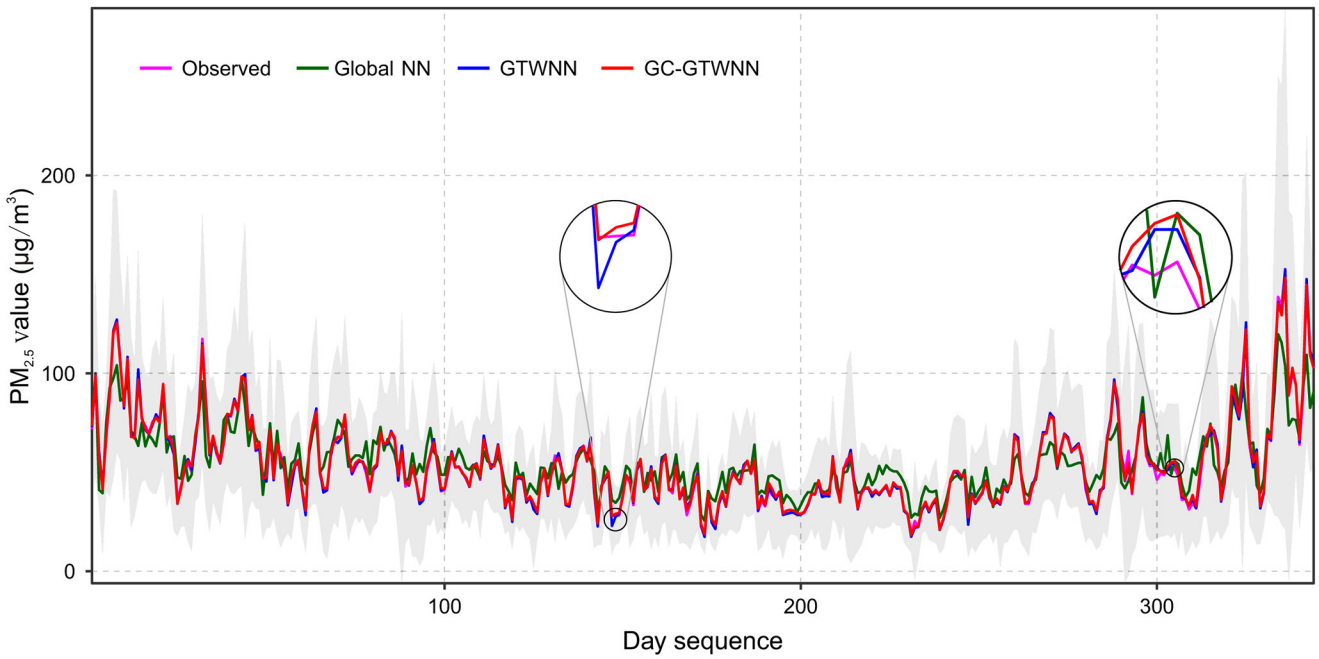


Fig. 6. Daily mean observations and model estimates of surface $PM_{2.5}$ concentration in 2015. Gray shading: standard deviation of surface $PM_{2.5}$ measurements. Black circle: zoomed-in view map.

in Fig. 6. Furthermore, the GC-GTWNN estimates are frequently found to be closer to surface $PM_{2.5}$ measurements (left-hand zoomed-in map). In contrast, the GTWNN model sometimes performs slightly better than the GC-GTWNN model (right-hand zoomed-in map); however, this is relatively rare. Meanwhile, the statistics confirm that the GC-GTWNN estimates ($R^2 = 0.996$, RMSE $= 1.43$ μg/m^3) have some advantages over GTWNN estimates ($R^2 = 0.994$, RMSE $= 1.62$ μg/m^3) compared with surface $PM_{2.5}$.

4) Spatial Evaluation of GC-GTWNN: In addition, the spatial statistic was conducted to provide a complete evaluation of the models. The time series of surface $PM_{2.5}$ observations and model CV estimates were collected for each grid cell containing $PM_{2.5}$ stations, and the R^2 and RMSE values were calculated. Fig. 7 shows the boxplots of R^2 and RMSE values of all the grid cells for the models. The mean values of R^2 and RMSE for the global NN model, which performs the poorest among the three models, are 0.42 and 25.00 μg/m^3,

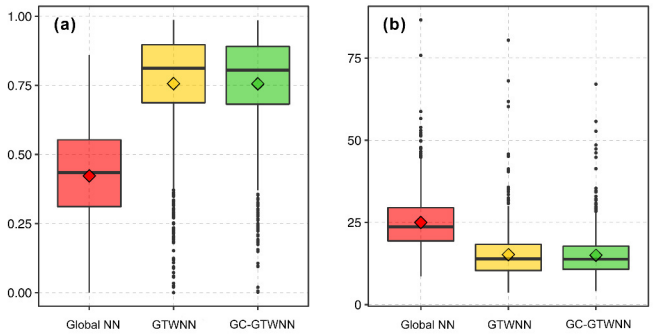


Fig. 7. Boxplots of spatial (a) R^2 and (b) RMSE (unit: μg/m^3) for the models.

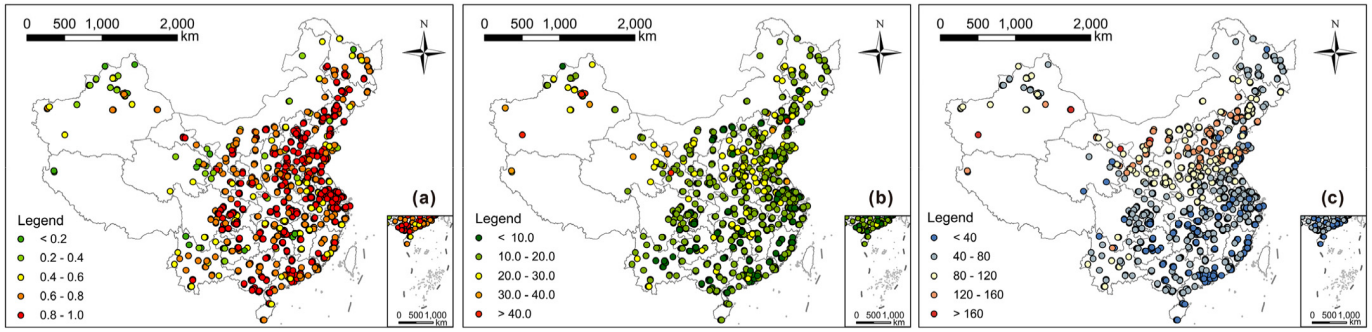


Fig. 8. Spatial evaluation of the GC-GTWNN model. (a) R^2. (b) RMSE (unit: μg/m^3). (c) Match-up samples.

respectively. Considering the GTWNN and GC-GTWNN models, the mean R^2 values of GTWNN and GC-GTWNN are 0.76 and 0.76, respectively. The RMSE values of the GTWNN range from 3.64 to 80.44 μg/m^3, with a mean value of 15.21 μg/m^3; the minimum, maximum, and mean values of RMSE for GC-GTWNN are 4.14, 67.08, and 15.01 μg/m^3, respectively. Meanwhile, 66 grid cells have a low R^2 value of less than 0.4 for the GTWNN model, whereas merely 53 grid cells for GC-GTWNN modeling. This condition indicates that GC-GTWNN can mitigate the extremely poor estimation of $PM_{2.5}$ concentrations to some degree.

We also mapped the spatial distribution of R^2, RMSE, and match-up samples for each grid cell, which are shown in Fig. 8. Eastern China shows higher R^2 and lower RMSE values, whereas western China reports worse performance, which may be due to the lower monitoring station density in western China. Another interesting finding is that southern China suffers from more missing satellite data, and fewer match-up samples are found, whereas the situation in northern China is better.

B. Spatial Distribution of $PM_{2.5}$ Derived From GC-GTWNN

The daily estimates of surface $PM_{2.5}$ concentration were obtained on the basis of the GC-GTWNN model. Considering sampling bias due to satellite missing data, a merging strategy proposed in our previous study for the mapping of annual and seasonal $PM_{2.5}$ distributions [19] was adopted. As shown in Fig. 9, severe pollution is revealed by the distributions of $PM_{2.5}$ across China. The spatial mean value of $PM_{2.5}$ is 44.8 μg/m^3, which is 28.1% higher than level 2 of the Chinese National Ambient Air Quality Standards (CNAAQS). Meanwhile, a spatial variation is revealed by the Hu Line (or Heihe–Tengchong Line) [46]. The west of the Hu line has a mean $PM_{2.5}$ level of 43.9 μg/m^3, whereas the east of the Hu Line shows a higher level of 46.4 μg/m^3, which is consistent with the degree of economic development and urbanization in China.

Spatially, heavy pollution hotspots with high levels of $PM_{2.5}$ have been clustered in China, including North China Plain (Beijing, Tianjin, Hebei, Henan, Anhui, Shandong, and Jiangsu), Xinjiang Autonomous Region, and Sichuan Basin. The aggregations of $PM_{2.5}$ may be attributed to high emissions, unfavorable meteorological influences, and/or topographic conditions [8], [29], [33]. Seasonally, winter reports the heaviest $PM_{2.5}$ pollution, followed by spring and autumn. Summer is found to be the cleanest season in one year. Therefore, the above findings reveal that GC-GTWNN modeling helps characterize the spatial and temporal patterns of $PM_{2.5}$ levels beyond ground station measurements alone.

V. Discussion

A. Comparison With Other Models

Considering the nonuniform distribution of monitoring stations, SDCV incorporates the spatial distance between the validation and modeling stations into model validation. For a given distance (d), the modeling station with a distance to the closest validation station of less than d is excluded in the model establishment [36]. The distances ranging from 0 to 200 km with a step of 10 km are selected for the evaluation of models. Meanwhile, GC-GTWNN modeling is compared with the widely used daily GWR and GTWR models and the previous geographically and temporally weighted generalized regression neural network (GTW-GRNN) [33]. Fig. 10 shows the SDCV performance of the models. All models generally report a downward trend in performance with the increasing distance (adjacent modeling stations are excluded). Especially in the 0–30-km phase, the models experience a dramatic performance decrease because they are spatiotemporally local models and are strongly dependent on the adjacent monitoring stations.

The comparison between the models follows. First, $d = 0$ km refers to the site-based CV. Considering spatiotemporal heterogeneity of the AOD–$PM_{2.5}$ relationship, GWR obtains an acceptable performance with R^2 and RMSE values of 0.72 and 20.54 μg/m^3, respectively, which outperforms the global NN model ($R^2 = 0.49$, RMSE = 27.62 μg/m^3). Benefitting from the use of temporal dependence, GTWR is more advantageous than GWR, with R^2 and RMSE values of 0.73 and 20.26 μg/m^3, respectively. These findings demonstrate that space–time regression models achieved relatively satisfactory results. Through nonlinear modeling of NN, the GTW-GRNN model reports some advantages, with R^2 and RMSE values of 0.79 and 17.81 μg/m^3, respectively. Meanwhile, GC-GTWNN notably outperforms the GWR and GTWR models. Compared with the GTWR model, the R^2 value is increased by 0.07 (from 0.73 to 0.80), and the RMSE value is decreased by 2.82 μg/m^3 (from 20.26 to 17.44 μg/m^3). The models displayed similar decreasing

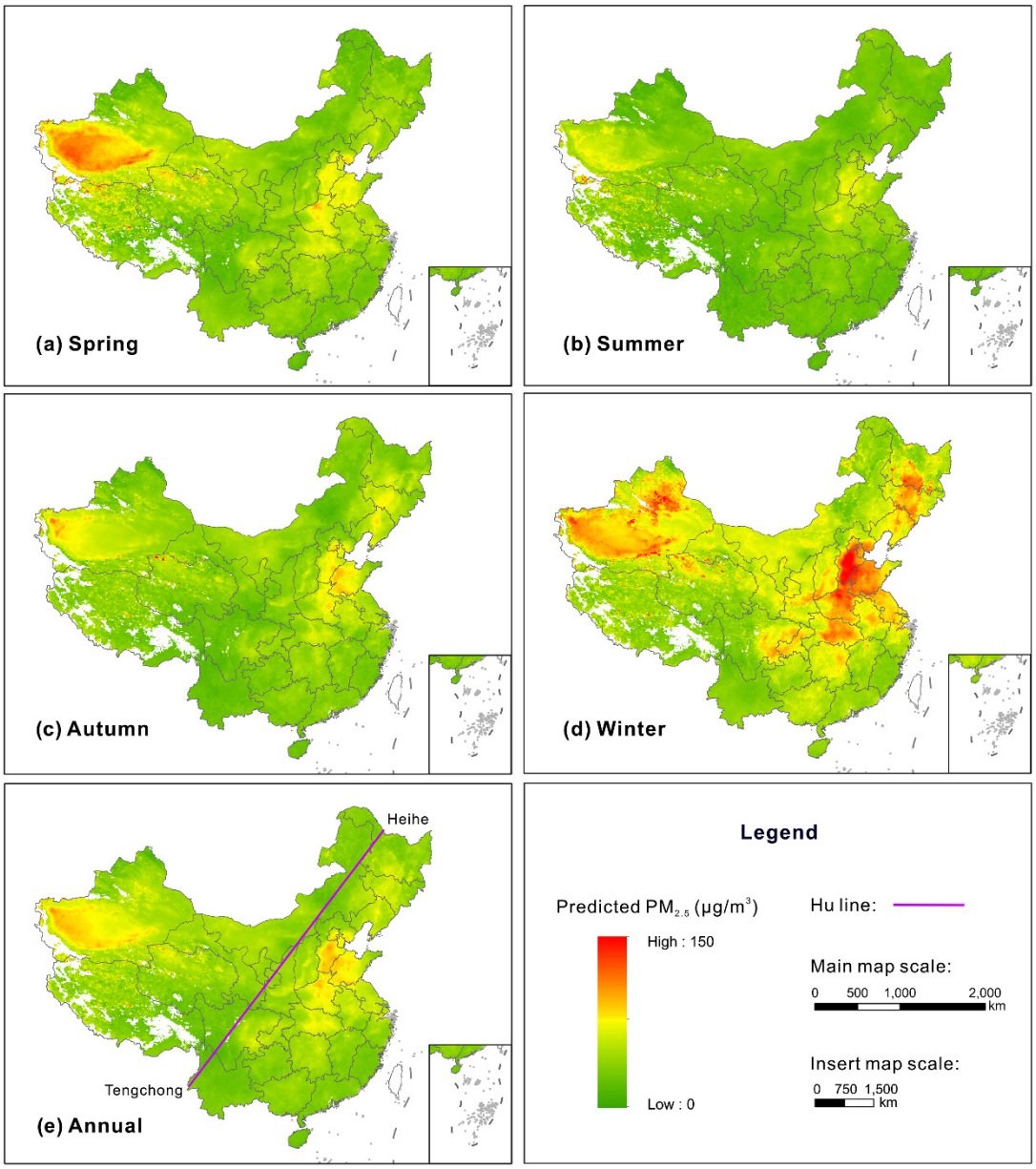


Fig. 9. Spatial distribution of surface $PM_{2.5}$ in China predicted by GC-GTWNN. (a) Spring. (b) Summer. (c) Autumn. (d) Winter. (e) Annual.

patterns in performance with the increasing distance from 0 to 200 km. However, the GC-GTWNN model performs the best among all the distances. The SDCV performance comparison with other popular models confirmed the superiority of GC-GTWNN for the spatial estimation of surface $PM_{2.5}$ from remote sensing observations.

In addition, we have compared the GC-GTWNN model with the recent models used to estimate surface $PM_{2.5}$ from remote sensing data, such as deep residual neural network (ResNet) [47] and extremely randomized trees (Extra-trees) [48]. The site-based CV results are shown in Table II, which demonstrates that the proposed GC-GTWNN model achieves better estimation performance. This condition may be due to that the GC-GTWNN model can effectively handle the spatial and temporal heterogeneity of the AOD–$PM_{2.5}$ relationship.

TABLE II
SITE-BASED CV RESULTS OF GC-GTWNN AND OTHER MODELS

Models	R^2	RMSE	MPE	RPE
GC-GTWNN	0.80	17.44	11.49	31.61
ResNet	0.77	18.57	12.42	33.66
Extra-trees	0.77	18.44	13.11	33.42

B. Configuration of GC-GTWNN Modeling

Two main categories of parameters must be configured in GC-GTWNN modeling. One category is the NN structure, and the other is the spatiotemporal weighting parameters. For the first category, the global NN and the GC-GTWNN share the same structure, which contains one input layer (seven nodes

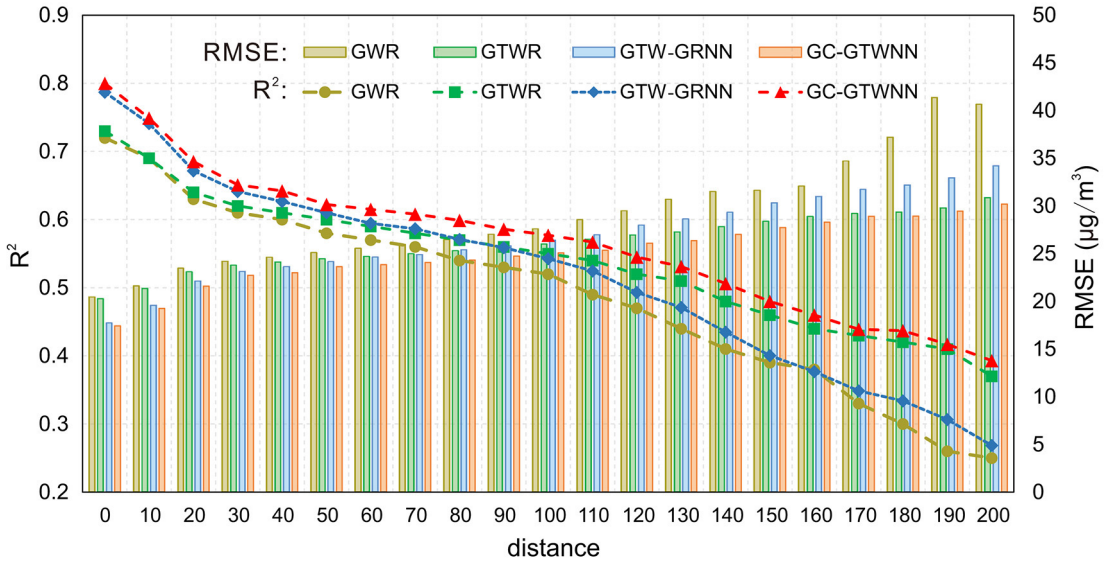


Fig. 10. SDCV performance of the models.

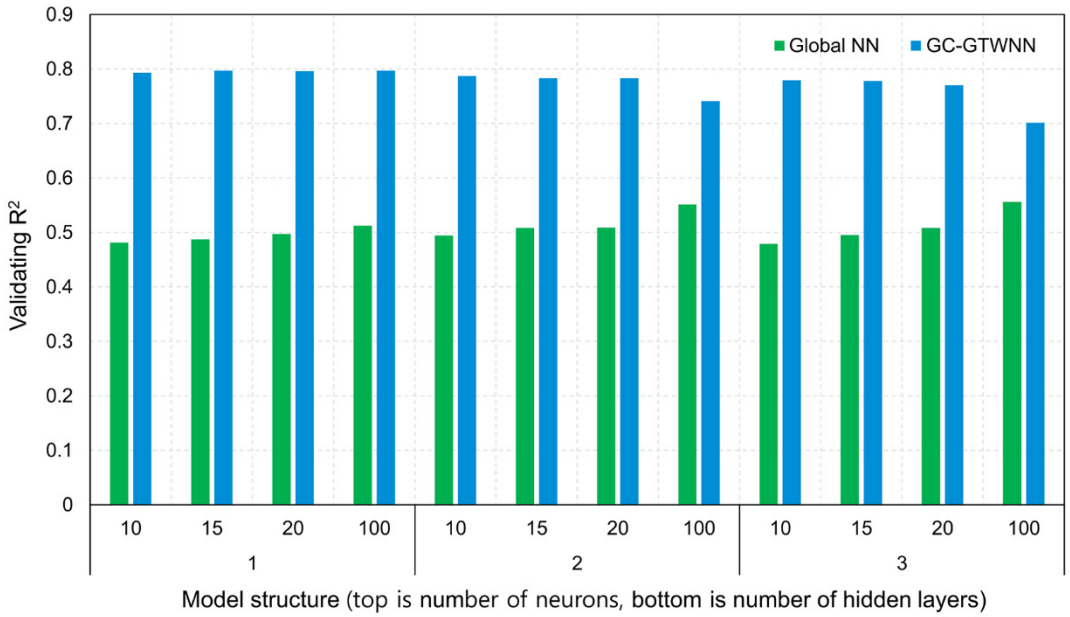


Fig. 11. Validation R^2 with different hidden layers and nodes.

for the input satellite observations and meteorological data) and one output layer (one node for surface $PM_{2.5}$). Concerning hidden layers, it is interesting to find that with more hidden layers, the global NN model can achieve a better performance, over a certain range. However, GC-GTWNN stands in the opposite direction (see Fig. 11). Consequently, one hidden layer with 15 nodes was selected in this study considering the prediction capability and computational complexity of GC-GTWNN modeling. In addition, the iteration epochs for global NN training, GTWNN fine-tuning, and GTWNN direct training were set as 800, 120, and 800, respectively. Besides, the adaptive bandwidth regime, in which the number of nearest neighbors remains constant, whereas the distance varies, was adopted in this study for the configuration of spatiotemporal weighting parameters. A CV approach was used to select the values of λ and h_{ST} [33] and designated as 8E4 and 2, respectively, via a nested-loop process.

The model fitting of the GC-GTWNN model was conducted, and the R^2 and RMSE values are 0.97 and 6.21 μg/m^3, respectively. The overfitting phenomenon exists in our model, with the site-based CV R^2 and RMSE values of 0.80 and 17.44 μg/m^3, respectively. However, the overfitting phenomenon is inevitable to a certain extent for the space–time weighted models, because a monitoring station locates at the position to be estimated, and its weight is maximal (weight = 1) in the model fitting, whereas no monitoring stations locate at the position to be estimated in the model validation. Hence, the imbalance exists in the model fitting and model validation. For example, several previous papers [29], [30], [49], [50] used the GTWR model

to estimate surface $PM_{2.5}$, and they all reported overfitting problems.

C. "Global-to-Local" Modeling Framework

A "global-to-local" framework is proposed in this study to estimate surface $PM_{2.5}$ from satellite observations. First, a global NN is established to model the relationship between surface $PM_{2.5}$, satellite AOD, and other auxiliary factors. Then, the GTWNN model is designed to address the spatiotemporal heterogeneity of this relationship. Indeed, the contribution of GC-GTWNN modeling may not only lie in the success of satellite-based $PM_{2.5}$ estimation but perhaps more importantly embody in proposing a "global-to-local" space–time modeling framework.

The "global-to-local" framework first models the global relationship using all collected samples, and then fine-tunes the local estimates based on a localized modeling technique. This framework demonstrates the potential for deep learning considering spatiotemporal variability. Deep learning models work effectively in approximating complicated relationships [51], [52], but are difficult to be trained in a local context for addressing spatiotemporal variability. Indeed, GC-GTWNN modeling with more hidden layers and neurons demonstrates the possibility to train a deep learning model considering spatiotemporal variability (Fig. 11). Thus, this framework provides some inspirations for combining deep learning and spatiotemporal characteristics. Furthermore, the proposed modeling framework may have considerable potentials in other applications, such as deep-learning-based remote sensing missing data reconstruction [53], [54]. In such a process, a deep learning model is usually trained on an abundant sample set, and then the missing information is recovered on the target image. However, useful information on the target image, which can embody the self-features of the target, is often neglected. Whether the fine-tuning for the trained deep learning model via useful information can improve the reconstruction accuracy deserves further attention.

D. Limitations

GC-GTWNN modeling may still encounter challenges in some cases; for instance, only a short time (e.g., a few days) of observations can be collected, and establishing the global NN models to provide the baseline of fine-tuning is difficult. At this time, the observations from other times (e.g., other years) or regions may be useful for pretraining of the NN models, which are popular in machine learning fields. Meanwhile, the validated results of the directly trained GTWNN also show remarkable potentials to work without the support of the global NN. Moreover, the satellite AOD product used in this study has a coarse resolution and may lose its popularity in the fine-scale monitoring of $PM_{2.5}$ pollution. However, this study aims to develop GC-GTWNN modeling for remote sensing estimation of $PM_{2.5}$. The proposed approach can be used for generating fine-scale $PM_{2.5}$ data from high-resolution satellite AOD observations in future studies.

This study does not effectively deal with the scale effect, that is, the mismatch between point-based measurements of the ground monitoring stations and satellite-based area observations. A previous study [55] has pointed out that the spatial representation of surface $PM_{2.5}$ monitoring station is generally approximately 0.5–16 km^2 with the most possible values of 2–3 km^2. However, this issue is common in this field, which still needs further research, and is also one of the key directions in our future studies.

Furthermore, more influencing factors will be considered in future studies. For instance, a previous study [56] indicated that the secondary formation could be a very important source for aerosol particles, and introducing solar radiation data will help consider this issue.

VI. Conclusion

Remote sensing estimation of surface $PM_{2.5}$ has been broadly studied in recent years. However, the nonlinearity and the spatiotemporal variability of the AOD–$PM_{2.5}$ relationship are rarely simultaneously addressed. Therefore, this study proposed a novel GC-GTWNN for the remote sensing estimation of surface $PM_{2.5}$. First, a global NN is established to learn the overall effect of input variables on ground $PM_{2.5}$. Initialized by the global NN, the GTWNN models are established and fine-tuned with spatiotemporally localized data samples. A case study across China showed that the CV R^2 and RMSE values of GC-GTWNN are 0.80 and 17.44 $\mu g/m^3$, respectively, which reports a significant improvement over the global NN ($R^2 = 0.49$, RMSE $= 27.62$ $\mu g/m^3$). Meanwhile, a comparison with the traditionally popular $PM_{2.5}$ estimation models also indicated that GC-GTWNN modeling achieved remarkable advantages for the estimation of $PM_{2.5}$ from satellite observations. Furthermore, the daily estimates of GC-GTWNN were effectively generated, and the national annual $PM_{2.5}$ exceeded level 2 of CNAAQS (i.e., 35 $\mu g/m^3$) by 28.1%. These results reveal the reliability of the proposed approach for satellite-based $PM_{2.5}$ estimation in this study. This approach will be of considerable value for the large-scale monitoring of $PM_{2.5}$ pollution.

Acknowledgment

This work was mainly conducted at Wuhan University. In this study, the satellite AOD and NDVI products are downloaded from Level 1 and Atmosphere Archive and Distribution System (LAADS), surface $PM_{2.5}$ measurements are obtained from the China National Environmental Monitoring Center (CNEMC), and MERRA-2 reanalysis data are available from MERRA-2 team. They acknowledge the data providers for freely releasing the above-mentioned data.

References

[1] Y. Yang *et al.*, "Ambient fine particulate pollution associated with diabetes mellitus among the elderly aged 50 years and older in China," *Environ. Pollut.*, vol. 243, pp. 815–823, Dec. 2018.

[2] A. J. Cohen *et al.*, "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the global burden of diseases study 2015," *Lancet*, vol. 389, no. 10082, pp. 1907–1918, May 2017.

[3] Y. Sun and C. Zhao, "Influence of saharan dust on the large-scale meteorological environment for development of tropical cyclone over north atlantic ocean basin," *J. Geophys. Res., Atmos.*, vol. 125, no. 23, Dec. 2020.

[4] M. Zhang *et al.*, "Haze events at different levels in winters: A comprehensive study of meteorological factors, aerosol characteristics and direct radiative forcing in megacities of north and central China," *Atmos. Environ.*, vol. 245, Jan. 2021, Art. no. 118056.

[5] L. Yu, M. Zhang, L. Wang, Y. Lu, and J. Li, "Effects of aerosols and water vapour on spatial-temporal variations of the clear-sky surface solar radiation in China," *Atmos. Res.*, vol. 248, Jan. 2021, Art. no. 105162.

[6] R. V. Martin, M. Brauer, A. van Donkelaar, G. Shaddick, U. Narain, and S. Dey, "No one knows which city has the highest concentration of fine particulate matter," *Atmos. Environ., X*, vol. 3, Jul. 2019, Art. no. 100040.

[7] R. M. Hoff and S. A. Christopher, "Remote sensing of particulate pollution from space: Have we reached the promised land?" *J. Air Waste Manage. Assoc.*, vol. 59, no. 6, pp. 645–675, Jun. 2009.

[8] T. Li, H. Shen, Q. Yuan, X. Zhang, and L. Zhang, "Estimating ground-level $PM_{2.5}$ by fusing satellite and station observations: A geo-intelligent deep learning approach," *Geophys. Res. Lett.* , vol. 44, pp. 11985–11993, Dec. 2017.

[9] X. Yan, Z. Zang, N. Luo, Y. Jiang, and Z. Li, "New interpretable deep learning model to monitor real-time $PM_{2.5}$ concentrations from satellite data," *Environ. Int.*, vol. 144, Nov. 2020, Art. no. 106060.

[10] T. Xue, Y. Zheng, T. Zhu, and Q. Zhang, "Spatiotemporal continuous estimates of $PM_{2.5}$ concentrations in China, 2000-2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations," *ISEE Conf. Abstr.*, vol. 2018, no. 1, pp. 345–357, Sep. 2018.

[11] C. Zheng *et al.*, "Analysis of influential factors for the relationship between $PM_{2.5}$ and AOD in Beijing," *Atmos. Chem. Phys.*, vol. 17, pp. 13473–13489, 2017.

[12] X. Su, L. Wang, M. Zhang, W. Qin, and M. Bilal, "A high-precision aerosol retrieval algorithm (HiPARA) for advanced himawari imager (AHI) data: Development and verification," *Remote Sens. Environ.*, vol. 253, Feb. 2021, Art. no. 112221.

[13] G. Zhang, X. Rui, and Y. Fan, "Critical review of methods to estimate $PM_{2.5}$ concentrations within specified research region," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 9, p. 368, Sep. 2018.

[14] Y. Chu *et al.*, "A review on predicting ground $PM_{2.5}$ concentration using satellite aerosol optical depth," *Atmosphere*, vol. 7, no. 10, p. 129, Oct. 2016.

[15] R. B. A. Koelemeijer, C. D. Homan, and J. Matthijsen, "Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over europe," *Atmos. Environ.*, vol. 40, no. 27, pp. 5304–5315, Sep. 2006.

[16] J. A. Engel-Cox, C. H. Holloman, B. W. Coutant, and R. M. Hoff, "Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality," *Atmos. Environ.*, vol. 38, no. 16, pp. 2495–2509, May 2004.

[17] P. Gupta and S. A. Christopher, "Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach," *J. Geophys. Res.*, vol. 114, no. D14, 2009.

[18] J. Tian and D. Chen, "A semi-empirical model for predicting hourly ground-level fine particulate matter ($PM_{2.5}$) concentration in southern ontario from satellite remote sensing and ground-based meteorological measurements," *Remote Sens. Environ.*, vol. 114, no. 2, pp. 221–229, Feb. 2010.

[19] T. Li, H. Shen, C. Zeng, Q. Yuan, and L. Zhang, "Point-surface fusion of station measurements and satellite observations for mapping $PM_{2.5}$ distribution in China: Methods and assessment," *Atmos. Environ.*, vol. 152, pp. 477–489, Mar. 2017.

[20] Q. Di, I. Kloog, P. Koutrakis, A. Lyapustin, Y. Wang, and J. Schwartz, "Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States," *Environ. Sci. Technol.*, vol. 50, pp. 4712–4721, May 2016.

[21] P. Gupta and S. A. Christopher, "Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach," *J. Geophys. Res.*, vol. 114, no. D20, 2009.

[22] C. Brokamp, R. Jandarov, M. Hossain, and P. Ryan, "Predicting daily urban fine particulate matter concentrations using a random forest model," *Environ. Sci. Technol.*, vol. 52, no. 7, pp. 4173–4179, Apr. 2018.

[23] K. de Hoogh, H. Héritier, M. Stafoggia, N. Kánzli, and I. Kloog, "Modelling daily $PM_{2.5}$ concentrations at high spatio-temporal resolution across switzerland," *Environ. Pollut.*, vol. 233, pp. 1147–1154, Feb. 2018.

[24] Z.-Y. Chen *et al.*, "Extreme gradient boosting model to estimate $PM_{2.5}$ concentrations with missing-filled satellite data in China," *Atmos. Environ.*, vol. 202, pp. 180–189, Apr. 2019.

[25] H. J. Lee, Y. Liu, B. A. Coull, J. Schwartz, and P. Koutrakis, "A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations," *Atmos. Chem. Phys.*, vol. 11, no. 15, pp. 7991–8002, Aug. 2011.

[26] Z. Ma, Y. Liu, Q. Zhao, M. Liu, Y. Zhou, and J. Bi, "Satellite-derived high resolution $PM_{2.5}$ concentrations in yangtze river delta region of China using improved linear mixed effects model," *Atmos. Environ.*, vol. 133, pp. 156–164, May 2016.

[27] Z. Ma, X. Hu, L. Huang, J. Bi, and Y. Liu, "Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing," *Environ. Sci. Technol.*, vol. 48, no. 13, pp. 7436–7444, Jul. 2014.

[28] W. Song, H. Jia, J. Huang, and Y. Zhang, "A satellite-based geographically weighted regression model for regional $PM_{2.5}$ estimation over the pearl river delta region in China," *Remote Sens. Environ.*, vol. 154, pp. 1–7, Nov. 2014.

[29] Q. He and B. Huang, "Satellite-based mapping of daily high-resolution ground $PM_{2.5}$ in China via space-time regression modeling," *Remote Sens. Environ.*, vol. 206, pp. 72–83, Mar. 2018.

[30] Y. Bai, L. Wu, K. Qin, Y. Zhang, Y. Shen, and Y. Zhou, "A geographically and temporally weighted regression model for ground-level $PM_{2.5}$ estimation from satellite-derived 500 m resolution AOD," *Remote Sens.*, vol. 8, no. 3, p. 262, Mar. 2016.

[31] Q. Xiao *et al.*, "Full-coverage high-resolution daily $PM_{2.5}$ estimation using MAIAC AOD in the Yangtze River Delta of china," *Remote Sens. Environ.*, vol. 199, pp. 437–446, Sep. 2017.

[32] Q. Xiao, H. H. Chang, G. Geng, and Y. Liu, "An ensemble machine-learning model to predict historical $PM_{2.5}$ concentrations in China from satellite data," *Environ. Sci. Technol.*, vol. 52, no. 22, pp. 13260–13269, Nov. 2018.

[33] T. Li, H. Shen, Q. Yuan, and L. Zhang, "Geographically and temporally weighted neural networks for satellite-based mapping of ground-level $PM_{2.5}$," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 178–188, Sep. 2020.

[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[35] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of K-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.

[36] T. Li, H. Shen, C. Zeng, and Q. Yuan, "A validation approach considering the uneven distribution of ground stations for satellite-based $PM_{2.5}$ estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1312–1321, 2020.

[37] H. Fan, C. Zhao, and Y. Yang, "A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018," *Atmos. Environ.*, vol. 220, Jan. 2020, Art. no. 117066.

[38] H. Shen, T. Li, Q. Yuan, and L. Zhang, "Estimating regional ground-level $PM_{2.5}$ directly from satellite top-of-atmosphere reflectance using deep belief networks," *J. Geophys. Res., Atmos.*, vol. 123, no. 24, p. 13, Dec. 2018.

[39] L. A. Remer *et al.*, "The MODIS aerosol algorithm, products, and validation," *J. Atmos. Sci.*, vol. 62, no. 4, pp. 947–973, 2005.

[40] Y. Wang, Q. Yuan, T. Li, H. Shen, L. Zheng, and L. Zhang, "Evaluation and comparison of MODIS collection 6.1 aerosol optical depth against AERONET over regions in China with multifarious underlying surfaces," *Atmos. Environ.*, vol. 200, pp. 280–301, Dec. 2019.

[41] Y. Wang, Q. Yuan, T. Li, H. Shen, L. Zheng, and L. Zhang, "Large-scale MODIS AOD products recovery: Spatial-temporal hybrid fusion considering aerosol variation mitigation," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 1–12, Nov. 2019.

[42] K. Zhang, C. Zhao, H. Fan, Y. Yang, and Y. Sun, "Toward understanding the differences of $PM_{2.5}$ characteristics among five China urban cities," *Asia–Pacific J. Atmos. Sci.*, vol. 56, no. 4, pp. 493–502, Nov. 2020.

[43] Z. Chen *et al.*, "Influence of meteorological conditions on $PM_{2.5}$ concentrations across China: A review of methodology and mechanism," *Environ. Int.*, vol. 139, Jun. 2020, Art. no. 105558.

[44] A. Molod, L. Takacs, M. Suarez, and J. Bacmeister, "Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2," *Geosci. Model. Develop.*, vol. 8, no. 5, pp. 1339–1356, May 2015.

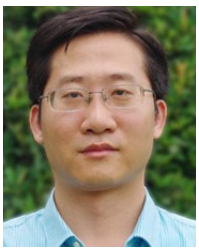
[45] X. Fang, B. Zou, X. Liu, T. Sternberg, and L. Zhai, "Satellite-based ground $PM_{2.5}$ estimation using timely structure adaptive modeling," *Remote Sens. Environ.*, vol. 186, pp. 152–163, Dec. 2016.

[46] H. Hu, "The distribution, regionalization and prospect of China's population," *Acta Geographica Sinica*, vol. 2, pp. 139–145, Jun. 1990.

[47] L. Li *et al.*, "Ensemble-based deep learning for estimating $PM_{2.5}$ over california with multisource big data including wildfire smoke," *Environ. Int.*, vol. 145, Dec. 2020, Art. no. 106143.
[48] J. Wei *et al.*, "Reconstructing 1-km-resolution high-quality $PM_{2.5}$ data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112136.
[49] Q. He and B. Huang, "Satellite-based high-resolution $PM_{2.5}$ estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model," *Environ. Pollut.*, vol. 236, pp. 1027–1037, May 2018.
[50] Y. Guo, Q. Tang, D.-Y. Gong, and Z. Zhang, "Estimating ground-level $PM_{2.5}$ concentrations in Beijing using a satellite-based geographically and temporally weighted regression model," *Remote Sens. Environ.*, vol. 198, pp. 140–149, Sep. 2017.
[51] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[52] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111716.
[53] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 36–38, Dec. 2017.
[54] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
[55] X. Shi, C. Zhao, J. H. Jiang, C. Wang, X. Yang, and Y. L. Yung, "Spatial representativeness of $PM_{2.5}$ concentrations obtained using observations from network stations," *J. Geophys. Res., Atmos.*, vol. 123, no. 6, pp. 3145–3158, Mar. 2018.
[56] C. Zhao, Y. Li, F. Zhang, Y. Sun, and P. Wang, "Growth rates of fine aerosol particles at a site near Beijing in June 2013," *Adv. Atmos. Sci.*, vol. 35, no. 2, pp. 209–217, Feb. 2018.

Tongwen Li received the B.S. degree in geographic information system and the Ph.D. degree in cartography and geographic information engineering from Wuhan University, Wuhan, China, in 2015 and 2020, respectively.

He is an Assistant Professor with the School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai, China. He has published more than 20 articles in peer-reviewed international journals. His research interests include atmospheric remote sensing (such as $PM_{2.5}$ and ozone) and machine learning in environmental remote sensing.



Huanfeng Shen (Senior Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2007, he joined the School of Resource and Environmental Sciences, Wuhan University, where he is a Luojia Distinguished Professor. He is a "Chang-Jiang Scholar" Distinguished Professor appointed by the Ministry of Education of China with the School of Resource and Environmental Sciences, Wuhan University. He has authored more than 100 research articles. His research interests include image quality improvement, remote sensing mapping and application, data fusion and assimilation, and regional and global environmental change.

Dr. Shen is a member of the Editorial Board of the *Journal of Applied Remote Sensing*. He has been supported by several talent programs, such as the Youth Talent Support Program of China in 2015, the China National Science Fund for Excellent Young Scholars in 2014, and the New Century Excellent Talents by the Ministry of Education of China in 2011.

Qiangqiang Yuan (Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006 and 2012, respectively.

In 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is a Professor. He has published more than 90 research articles, including more than 70 peer-reviewed articles in international journals. His research interests include image reconstruction, remote sensing image processing and application, and data fusion.

Dr. Yuan was a recipient of the Youth Talent Support Program of China in 2019, the Top-Ten Academic Star of Wuhan University in 2011, and the recognition of Best Reviewers of the IEEE GRSL in 2019. In 2014, he received the Hong Kong Scholar Award from the Society of Hong Kong Scholars and the China National Postdoctoral Council. He is an Associate Editor of five international journals and has frequently served as a referee for more than 40 international journals for remote sensing and image processing.

Liangpei Zhang (Fellow, IEEE) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the Remote Sensing Program in China. He has published more than 500 research articles and five books. He holds 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a fellow of the Institution of Engineering and Technology (IET), an Executive Member (Board of Governor) of the China National Committee of International Geosphere–Biosphere Programme, Executive Member of the China Society of Image and Graphics, and so on. He was a recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS). He received the best reviewer awards from the IEEE GRSS for his service to the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS) in 2012 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL) in 2014. His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) Student Paper Contest in recent years. He regularly serves as the Co-Chair of the series SPIE conferences on multispectral image processing and pattern recognition, conference on Asia remote sensing, and many other conferences. He is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He was the General Chair for the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) and the Guest Editor of JSTARS. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor for the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, and the *Journal of Remote Sensing*, and a Guest Editor for the *Journal of Applied Remote Sensing* and the *Journal of Sensors*. He also serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.