

Hourly PM_{2.5} Concentration Monitoring With Spatiotemporal Continuity by the Fusion of Satellite and Station Observations

Jingan Wu¹, Tongwen Li², Chengyue Zhang, Qing Cheng³, and Huanfeng Shen¹, *Senior Member, IEEE*

Abstract—Hourly monitoring of ground-level fine particulate matter (PM_{2.5}) concentrations forms the basis to assess the short-term PM_{2.5} exposure and make rapid responses to pollution events. Satellite remote sensing and ground monitoring stations are able to measure hourly PM_{2.5} concentrations, but both of them have strengths and weaknesses: the former features wide spatial coverage, whereas displaying a discontinuous timeline as the retrievals have numerous gaps; conversely, the latter allows for temporally continuous monitoring, but with a limited spatial range around stations being reflected. Thus, efforts are required to map ground-level PM_{2.5} at an hourly scale with spatiotemporal continuity. In this article, we developed a framework to generate hourly seamless PM_{2.5} estimates by integrating the aforementioned two data sources with complementary spatiotemporal traits. The satellite-derived aerosol optical depth acquisitions are converted along with auxiliary predictors to retrieve ground-level PM_{2.5}, and then the missing gaps in the retrievals are filled by fusing the satellite-based retrievals and station-based measurements. Meanwhile, we proposed a promising approach to fill the gaps by combining an adapted spatiotemporal fusion model and an error correction method. The validity of the proposed method is confirmed by mapping hourly PM_{2.5} distributions for 2016 in the Wuhan urban agglomeration, China. The proposed reconstruction method achieved R^2 (root-mean-square error) of 0.87 (6.50 $\mu\text{g}/\text{m}^3$) and 0.82 (15.01 $\mu\text{g}/\text{m}^3$) in the area-based and point-based evaluation, respectively, indicating an excellent model performance. The presented framework maps hourly ground-level PM_{2.5} with spatiotemporal continuity and satisfactory accuracy, and represents an important step towards near real-time monitoring.

Index Terms—Data reconstruction, fine particulate matter (PM_{2.5}), near real-time monitoring, spatiotemporal continuity, spatiotemporal fusion.

Manuscript received April 6, 2021; revised June 30, 2021; accepted August 1, 2021. Date of publication August 10, 2021; date of current version August 25, 2021. This work was supported by the National Key R&D Program of China under Grant 2018YFA0605500. (*Corresponding author: Huanfeng Shen.*)

Jingan Wu and Chengyue Zhang are with the School of Resource and Environmental Sciences, Wuhan University, Wuhan 430072, China (e-mail: wujg@whu.edu.cn; cowherbseed@gmail.com).

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

Qing Cheng is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: qingcheng@whu.edu.cn).

Huanfeng Shen is with the School of Resource and Environmental Sciences and the Collaborative Innovation Center for Geospatial Technology, Wuhan University, Wuhan 430072, China (e-mail: shenhf@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3103020

I. INTRODUCTION

FINE particulate matter (particulate matter with an aerodynamic diameter of less than 2.5 μm , also known as PM_{2.5} for short) has been a focus of global public concern over the past decade, as a result of its adverse health effects [1]–[3] and close connection to climate change [4], [5]. The considerable attention has inspired the scientific community to investigate the PM_{2.5} spatiotemporal distributions. In recent years, there have been lots of studies to map ground-level PM_{2.5} distributions on the yearly, monthly, or daily basis [6]–[11], providing important guidance for environmental management and decision making of governmental agency. Currently, hourly monitoring of PM_{2.5} concentrations is in great demand [12], since it forms the basis to estimate short-term PM_{2.5} exposure and make rapid responses to serious pollution events. The generation of hourly PM_{2.5} estimates has thus been of great significance and worthy of investigation.

Owing to the spatially wide coverage and temporally repeated observation, satellite remote sensing is considered a very good choice to derive ground-level PM_{2.5} [13], [14]. Given the monitoring mission at an hourly scale, geostationary satellites do us a favor. For instance, the new generation geostationary meteorological satellite of Japan, Himawari-8 [15], was launched on July 2, 2015, and the satellite-derived hourly aerosol optical depth (AOD) product has been routinely released, which can be used as a major predictor to estimate PM_{2.5} concentrations [13], [16], [17]. The relationship between AOD and PM_{2.5} has been extensively investigated, and a number of models were developed, including simulation-based models [18], semiphysical models [19], [20], and statistical-based models [21]–[24]. Based on the presented models, some studies mapped and analyzed the hourly distribution of PM_{2.5} [25]. For example, Wang, *et al.* [12] applied an improved linear mixed-effect model to retrieve hourly PM_{2.5} concentrations over the Beijing-Tianjin-Hebei region in China; and Zeng, *et al.* [26] mapped the hourly PM_{2.5} distributions in Hebei, China, by using a vertical-humidity correction method.

Although the satellite-derived observations can be employed, they have a notable limitation for mapping hourly PM_{2.5} distributions. The satellite-measured AOD product usually suffers from numerous missing gaps for the following reasons. First, the optical instrument is incapable of working at night [27], resulting in the nighttime AOD completely undetectable. Second, cloud cover and high surface albedo (e.g., ice and snow) set barriers

to effectively retrieve AOD [28], and thus the daytime AOD is partially missing. Since the major predictor AOD has missing gaps, the satellite-derived $PM_{2.5}$ retrievals inherit this legacy and present a discontinuous timeline. To fill the missing gaps, some approaches have been proposed, even though not all of them specialize in the hourly mapping mission. By averaging the multiple daily estimates [8], [29], [30], gaps are likely to be filled in the yearly, seasonal, and monthly $PM_{2.5}$ maps. As for the daily $PM_{2.5}$ maps, some studies propose solutions by exploiting the spatiotemporal autocorrelation of $PM_{2.5}$ [31]–[33]. For instance, Kloog, *et al.* [34] used the spatial smoothing technique to predict $PM_{2.5}$ at the locations without AOD records. Xiao, *et al.* [35] presented a multiple imputation approach to fill the gaps in AOD product, in which the relationship was established between AOD and other predictors (e.g., meteorological parameters and model simulations) in a five-day time window. These approaches generally showed good performance in the daily-specific reconstruction task. However, since the assumption of $PM_{2.5}$ autocorrelation is only effective for a certain range of spatial and temporal coverage, these methods are unsuitable for the scenarios with very serious data incompleteness. The hourly mapping mission, unfortunately, features the abovementioned scenario, in which the satellite-derived $PM_{2.5}$ retrievals are likely to be seriously missing in the spatial coverage or consecutively missing in the temporal coverage [36]. Effective reconstruction methods, therefore, are being called for to fill the gaps in the satellite-based hourly $PM_{2.5}$ retrievals.

Besides satellite remote sensing, ground stations are another widely employed way of monitoring $PM_{2.5}$ [8]. The monitoring stations precisely record $PM_{2.5}$ concentrations with a consecutive timeline, but they are only able to reflect $PM_{2.5}$ in a limited spatial extent around stations. This just complements the weaknesses of satellite remote sensing. Given the complementary spatiotemporal features, it would be interesting to integrate the aforementioned two data sources to map hourly $PM_{2.5}$ concentrations. Thus, the main objective of this article is to present a framework to produce hourly seamless $PM_{2.5}$ distributions by integrating satellite-derived observations and station-based measurements. To be specific, the key is to fill the missing gaps in the satellite-derived $PM_{2.5}$ retrievals. The effectiveness of the proposed method is verified by a case study, in which the hourly seamless $PM_{2.5}$ estimates of 2016 in the Wuhan urban agglomeration (WUA), China, are generated and assessed by using Himawari-8 AOD product, meteorological parameters, moderate resolution imaging spectroradiometer (MODIS) normalized difference vegetation index (NDVI) data and station-measured $PM_{2.5}$ concentrations as input data.

II. STUDY REGION AND DATA

A. Study Region

The study area is the WUA, as shown in Fig. 1. The WUA, with the latitude range of $29^{\circ} 05' \sim 31^{\circ} 50' N$ and the longitude range of $112^{\circ} 30' \sim 116^{\circ} 10' E$, is located in the east of Hubei province, China [see Fig. 1(a)]. It comprises Wuhan and the eight cities around Wuhan, namely, Huanggang, Xiaogan, Ezhou, Huangshi, Xianning, Qianjiang, Xiantao, and Tianmen. As a core urban

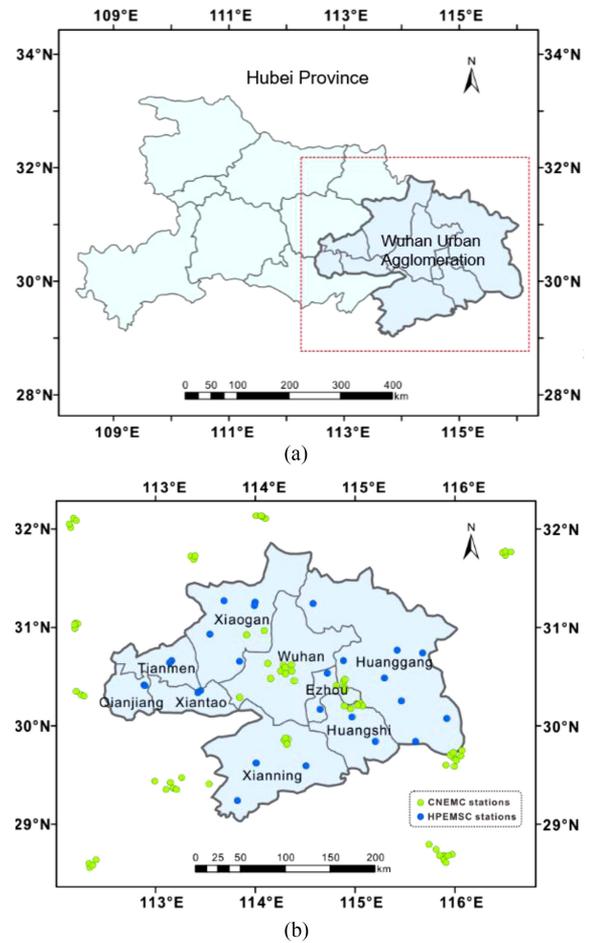


Fig. 1. Study region and spatial distribution of $PM_{2.5}$ monitoring stations. (a) Location of the WUA in Hubei province. (b) Spatial distribution of $PM_{2.5}$ monitoring stations in the WUA. CNEMC is the China National Environmental Monitoring Center, and HPEMSC is the Hubei Provincial Environmental Monitoring Station Center.

agglomeration in Central China, the WUA has a total area of 57800 km^2 , and 38 million permanent residents [37]. Although covering less than one-third of the area of Hubei province, the WUA contributes to more than half of the provincial gross domestic product [38]. Over the past decade, increasing investments and rapid infrastructure construction have taken place in the region, with Wuhan in particular attracting significant socio-economic development. However, serious air pollution comes hand in hand with these achievements [39]. It has been reported that the mean atmospheric $PM_{2.5}$ mass concentration for winter in Wuhan is about $160 \pm 50 \mu\text{g}/\text{m}^3$, which is much higher than the limit in the national standards for atmospheric environment quality ($75 \mu\text{g}/\text{m}^3$) for China's cities [40], [41]. Given the severe atmospheric situation this area is facing, we decided to take the WUA as our study area in this research.

B. Data Collection

1) *Ground-Level $PM_{2.5}$ Measurements:* Hourly ground-level $PM_{2.5}$ measurements for 2016 were collected from the China National Environmental Monitoring Center (CNEMC,

<http://www.cnemc.cn>) and the Hubei Provincial Environmental Monitoring Station Center (HPEMSC). PM_{2.5} measurements from 104 monitoring stations located in the spatial range of 28.4°~32.3°N and 112.0°~116.7°E were collected, including 78 CNEMC stations and 26 HPEMSC stations. According to the Chinese National Ambient Air Quality Standard (GB3905-2012), the ground-level PM_{2.5} concentrations are measured with beta attenuation monitors (or beta-gauges) or by the tapered element oscillating microbalance method, with an uncertainty of 0.75% for the hourly records [42]. Due to the instrumental malfunction and data transmission error, some of the hourly station records were unavailable in the collected measurements, resulting in the number of data records in a collected measurement being less than 104. To guarantee the estimation performance, we discarded the measurements with a significantly insufficient station number (i.e., <70). Finally, 8541 PM_{2.5} measurements were used for the whole year of 2016.

2) *Satellite-Based AOD Product*: The satellite-derived AOD product acquired by the advanced Himawari imager were considered in this article. The Himawari-8 satellite provides observations with a coverage of about one-third of the Earth, including East and Southeast Asia, and the AOD product is derived from the visible and infrared bands by referencing a look-up table based on an assumed spheroid-particle aerosol model [43]. The quality of the Himawari-8 AOD product has been confirmed [44]. The level-3 AOD product for 2016 was downloaded from the Japan Aerospace Exploration Agency P-Tree System (<http://www.eorc.jaxa.jp/ptree/>), with a spatial resolution of 0.05° and a temporal resolution of one hour. The recorded AOD is subject to quality assurance with four confidence levels (i.e., “very good,” “good,” “marginal,” and “no confidence”). In this article, only the AOD pixels with the highest confidence level (i.e., “very good”) were used, and the other pixels were identified as “missing.”

3) *Auxiliary Variables*: Since meteorological parameters have proven highly correlated with PM_{2.5} distributions [45], they were employed as auxiliary variables to estimate PM_{2.5}. Four kinds of meteorological variables were used in this article, namely, specific humidity (SH, kg/kg), air temperature at a 2-m height (TMP, K), wind speed at 10 m above ground (WS, m/s), and surface pressure (PS, kPa). They were downloaded from the Goddard Earth Observing System Data Assimilation System GOES-5 Forward Processing system (<https://gmao.gsfc.nasa.gov/>), with a spatial resolution of 0.25° latitude × 0.3125° longitude. In addition, the MODIS NDVI product was used to characterize the land covers, and it was obtained from the level 1 and atmosphere archive and distribution system website (<https://ladsweb.modaps.eosdis.nasa.gov/>), with a spatial resolution of 1 km and a temporal frequency of 16 days.

III. METHOD

The target of this article is to generate hourly PM_{2.5} estimates with spatiotemporal continuity. A two-stage operational framework is provided: the satellite-based AOD is converted along with other auxiliary variables to retrieve PM_{2.5} concentrations, which is referred to as “PM_{2.5} retrieval”; and the

satellite-based PM_{2.5} estimates from the first stage are fused with station-measured PM_{2.5} to fill the missing gaps, which is referred to as “PM_{2.5} reconstruction”. As mentioned earlier, it has been widely explored to estimate PM_{2.5} from AOD in the past studies [30], so we directly used our previous work, i.e., geo-intelligent deep belief network (Geoi-BDN), in the first stage [46]. The performance of the adopted PM_{2.5} retrieval model has been comprehensively verified and fully confirmed. Instead, the PM_{2.5} reconstruction remains a challenge, and thus we focused on the second stage and developed an approach to reconstruct the missing pixels. Accordingly, in the following, the first stage is introduced briefly in Section III-A, whereas the second stage is described in detail in Section III-B.

A. Retrieving Hourly PM_{2.5} From AOD Observations

In the first stage, the satellite-based AOD observations, meteorological parameters and NDVI serve as predictor variables to predict PM_{2.5} concentrations. Generally, the samples (station-measured PM_{2.5} concentrations vs. the values of predictor variables) are collected at the locations with station monitors to derive the relationship between PM_{2.5} and the predictors, and then the relationship is applied at the locations without station monitors to predict PM_{2.5} concentrations.

Before the derivation of the relationship, data preprocessing was carried out to establish a temporally and spatially consistent dataset. The auxiliary variables were reprojected and resampled to ensure the same projection and spatial resolution (0.05°) as Himawari-8 AOD observations. The PM_{2.5} concentrations recorded by multiple monitoring stations within a 0.05° grid cell were averaged. Then, AOD and auxiliary variables were extracted at the locations where station records were available, and the collected samples (station-based PM_{2.5} concentrations vs. the values of predictor variables) were used to model the relationship between PM_{2.5} and predictor variables. In this research, we adopted our previous work, the Geoi-BDN [46] to establish the relationship. This model applies a deep learning architecture to describe the complex nonlinear relationship between the variables [47], and it also incorporates the spatiotemporal autocorrelation of PM_{2.5} distributions to enhance the prediction performance. The modeling process between PM_{2.5} and various predictors is mathematically depicted as

$$PM_{2.5} = f(AOD, SH, TMP, WS, PS, NDVI, S_PM_{2.5}, T_PM_{2.5}, DIS) \quad (1)$$

where $f(\cdot)$ represents the nonlinear relationship between PM_{2.5} and the predictor variables. AOD, SH, TMP, WS, PS, and NDVI are the predictor variables described in Section II-B. $S_PM_{2.5}$ and $T_PM_{2.5}$ are the spatial and temporal terms, representing the spatiotemporal autocorrelation of PM_{2.5} concentrations, and they are derived by considering the ground-based measurements spatially and temporally close to the predicted values. Given that the station density has an effect on the prediction, an additional term DIS (the geographical distance) is used to reflect the heterogeneity of station distributions. The

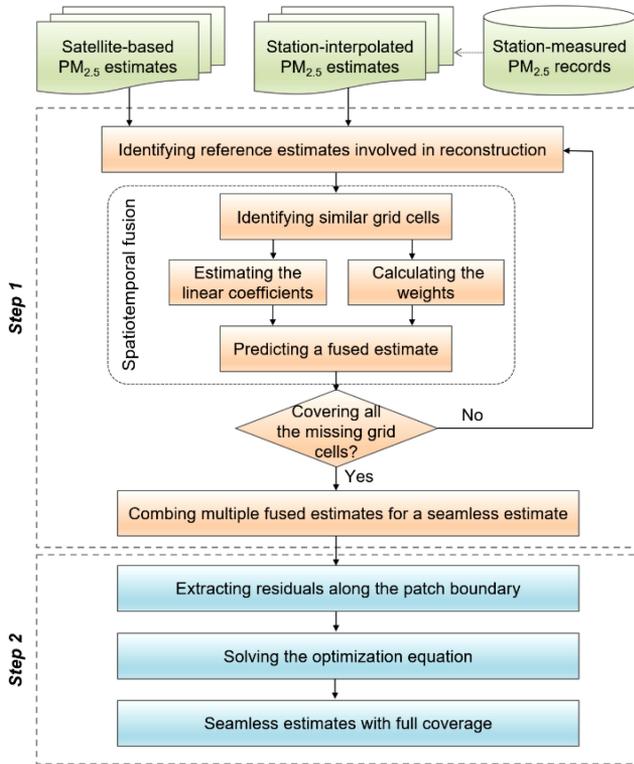


Fig. 2. Flowchart of the propose $PM_{2.5}$ reconstruction method.

term DIS is expressed as

$$DIS = \min \left(\frac{1}{ds_k} \right) \quad (2)$$

where ds_k represents the geographical distance between the given location and the k th nearby station in the prediction. A larger DIS value indicates denser station distribution around the given location. For more details of the Geoi-BDN model, we refer the readers to Li *et al.* [46]. In this article, a Geoi-DBN model with two restricted Boltzmann machine layers was used, and through the pretraining and fine-tuning, the relationship between $PM_{2.5}$ and the predictor variables were characterized. The model trained with collected sampled was then applied to the locations without monitoring stations to predict $PM_{2.5}$, and thus the satellite-based $PM_{2.5}$ estimates were produced.

B. Filling Missing Gaps in the Hourly $PM_{2.5}$ Estimates

The primary predictor variable AOD has numerous missing gaps, and the predicted $PM_{2.5}$ retrievals also show the same problem of data incompleteness. The second stage aims to reconstruct the missing grid cells in the hourly $PM_{2.5}$ retrievals, and it is technically divided into two steps, in which the first step fills the gaps by spatiotemporal fusion of satellite retrievals and station measurements and the second step reduces the fusion errors as more as possible. A flowchart of the proposed reconstruction method is illustrated in Fig. 2.

1) Spatiotemporal fusion of the two data sources: The $PM_{2.5}$ retrievals from the first stage successfully characterizes the local

spatial variations of $PM_{2.5}$ distributions, but it is unable to fully reflect the temporal variations due to the gaps, especially the consecutive gaps during the nighttime. So, they represent the data source with spatial superiority but temporal limitation. Meanwhile, we interpolate the station measurements to cover the study area by using the inverse distance weighting model [30], [46]. Due to the consecutive timeline of the ground-based measurements, the interpolated estimates are good at capturing temporal variations, but they cannot precisely map the spatial variations because of the interpolation uncertainty and artifacts. Thus, they represent the data source with temporal superiority, but spatial limitation. Given the contrasting but complementary traits, it would be a good idea to derive the $PM_{2.5}$ estimates with the enhanced spatiotemporal superiority by integrating the aforementioned two data sources. The spatiotemporal data fusion technique [48], as a research hotspot over the past decade, can help to achieve the goal. It can be used to fuse the two data sources with contrasting spatiotemporal characteristics and creates synthetic data with enhanced spatiotemporal resolutions [49]–[51]. Although great achievements have been made in developing spatiotemporal fusion models [48], few of them can be directly applied in this task, as the hourly $PM_{2.5}$ retrievals suffer from serious data incompleteness. In this article, we adapted the current work in the new scenario and proposed a spatiotemporal-fusion-based method for $PM_{2.5}$ reconstruction.

Fig. 3(a) displays the schematic of the spatiotemporal fusion process. For simplicity, the satellite-retrieved data source and the station-interpolated data source are denoted as R and I , respectively. Specifically, supposing the $PM_{2.5}$ retrieval at t_p is full of missing gaps, and in this case, we would like to fill the gaps and produce a reconstructed estimate R_p . It is assumed that the $PM_{2.5}$ temporal variation can be locally described as a linear relationship. More clearly, the $PM_{2.5}$ concentration in a location (x, y) can be modeled as a linear transformation of the $PM_{2.5}$ at some point before t_p , which is mathematically depicted as

$$R_p(x, y) = a(x, y) \times R_k(x, y) + b(x, y) \quad (3)$$

where R_k represents the retrieved estimate at t_k , a time point before t_p , at which the $PM_{2.5}$ concentration is validly retrieved in (x, y) . a and b are the local linear coefficients accounting for the $PM_{2.5}$ variation from t_k to t_p . Given that $PM_{2.5}$ distributions show spatial autocorrelation, it is reasonable to take advantage of the neighboring grid cells with similar $PM_{2.5}$ concentrations to enhance the prediction robustness. Thus, an improved version of the prediction model in (3) can be expressed as

$$R_p^k(x, y) = \sum_{i=1}^N w(x_i, y_i) \times [a(x_i, y_i) \times R_k(x_i, y_i) + b(x_i, y_i)] \quad (4)$$

where (x_i, y_i) is the location of the i th neighboring similar grid cell, and N is the number of similar grid cells in the prediction. $R_k(x_i, y_i)$ is the known $PM_{2.5}$ concentration at (x_i, y_i) in the estimate R_k . $w(x_i, y_i)$ is the weight of the i th similar grid cell, which measures its contribution to the prediction. $R_p^k(x, y)$ is the predicted $PM_{2.5}$ at t_p , which uses the estimates at t_k as reference. In the right side of (4), only $R_k(x_i, y_i)$ is already know, so

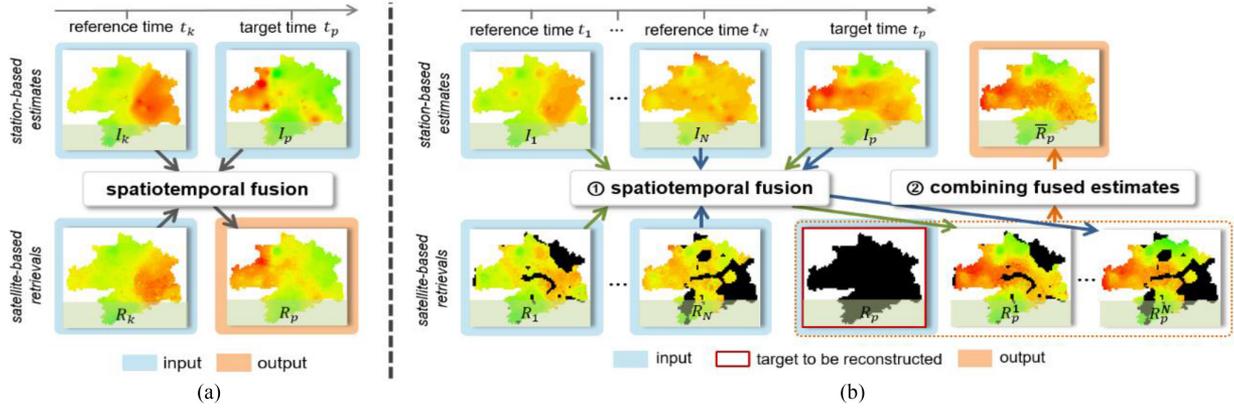


Fig. 3. Schematic of the proposed spatiotemporal fusion based reconstruction method (a) Schematic of a spatiotemporal fusion process. (b) Schematic of the proposed PM_{2.5} reconstruction method. The missing grid cells are marked in black. Please note that the arrows in the same color in (b) represent a fusion process.

we have to estimate the linear coefficients [i.e., $a(x_i, y_i)$ and $b(x_i, y_i)$] and the weights of similar grid cells [i.e., $w(x_i, y_i)$] to get the fusion result. The implementation details to identify the similar grid cells, estimate the linear coefficients, and calculate the weights are given later.

- 1) *Identifying the Locations of Similar Grid Cells:* For the location (x, y) , a local window centered at (x, y) is applied, and then a filtering procedure is used for selecting similar grid cells. A location (x_i, y_i) will be identified if the corresponding grid cell satisfies the following criteria simultaneously

$$|R_k(x, y) - R_k(x_i, y_i)| < d \quad (5)$$

$$|R_k(x_i, y_i) - I_k(x_i, y_i)| < \varepsilon \quad (6)$$

where (x_i, y_i) is the location of an eligible similar grid cell. In (5), the retrieved estimate R_k is exploited for screening, requiring the central grid cell and similar grid cell to have similar PM_{2.5} concentrations. In (6), the recorded PM_{2.5} concentrations in (x_i, y_i) are compared in the two estimates I_k and R_k , requiring that they do not have striking differences. This helps to exclude the grid cells with significant interpolation errors from the following prediction. d and ε are the thresholds to segment the similar grid cells from other potential candidates, and they were empirically set to 9 and 15 through a trial-and-error test in this article.

- 2) *Estimating the Local Linear Coefficients:* The linear coefficients account for the temporal variations of PM_{2.5} in the fusion model. As we mentioned before, the ground-based observations show a consecutive timeline and accordingly, the station-interpolated estimates can reflect the PM_{2.5} variations. We assume the linear coefficients derived from the interpolated data source can be applied to the retrieved data source, so the linear coefficients in the fusion model can be estimated. Technically, based on the locations identified, we can track two sets of similar grid cells from the interpolated estimates I_k and I_p , respectively. The two sets of similar grid cells are linearly regressed to obtain the

coefficients a and b , which characterize the PM_{2.5} variations from t_k and t_p . Following the previous studies [52], we applied a restricted least-squares regression model to estimate the coefficients.

- 3) *Calculating the Weights:* Although the selected grid cells show similar PM_{2.5} concentrations, there are still some differences. A grid cell with fewer differences offers more reliable information, and thus it should contribute more to the prediction. In the implementation, we evaluate the PM_{2.5} differences between the central grid cell and the similar grid cells through the estimate R_k , and their reciprocals are normalized to [0, 1] as the weight

$$D(x_i, y_i) = |R_k(x, y) - R_k(x_i, y_i)| \quad (7)$$

$$w(x_i, y_i) = [1/D(x_i, y_i)] / \sum_{i=1}^n [1/D(x_i, y_i)] \quad (8)$$

where $D(x_i, y_i)$ is the PM_{2.5} difference between the target grid cell and i th similar grid cell, and $w(x_i, y_i)$ represents the weight of the i th similar grid cell. After these steps, the missing grid cell can be predicted based on (4).

- 4) *Performing Spatiotemporal Fusion With Multiple Reference Estimates:* As shown in Fig. 3(a), if the retrieved estimate R_k does not have gaps, the fused result R_p^k will be seamless, presenting spatially continuous PM_{2.5} distributions at t_p . However, in real applications, it is uncommon to find the retrieved estimate R_k without any gaps. As a consequence, the fused estimate R_p^k cannot be seamless in most cases. To solve this problem, we came up with the strategy to perform the spatiotemporal fusion with multiple reference estimates, as shown in Fig. 3(b). A procedure is established to select a set of retrieved estimates as reference in the spatiotemporal fusion. Technically, a retrieved estimate with high quality should have a large coverage of valid grid cells, so as to ensure the spatial coverage that can be recovered in the fused result. In the implementation, the retrieved estimates before t_p are ordered chronologically and checked separately, and those with validly retrieved grid cells accounting for >

40% of the study area are selected as qualified in this article. The selection procedure will not end until all the missing locations can be recovered, and afterwards, n retrieved estimates $\{R_k|k = 1, 2, \dots, n\}$ are selected as the input of multiple loops of spatiotemporal fusion, as shown in Fig. 3(b). To fill all the missing grid cells in R_p , multiple fusion loops with different reference estimates are performed, and correspondingly, multiple fused results $\{R_p^k|k = 1, 2, \dots, n\}$ are generated. At a missing location (x, y) , it is likely that a subset of fused results $\{R_p^{l_1}, R_p^{l_2}, \dots, R_p^{l_m}\}$ ($m < n$) would have recovered grid cells, and they will be combined in a weighted manner to produce the final prediction. As reported in the previous studies [50], [52], a fused result is more reliable if less temporal variations occur from t_k and t_p . Thus, we use the interpolated pairs to measure the temporal variations and give the temporal weights. The weighting averaging of multiple fused estimates $\{R_p^{l_1}, R_p^{l_2}, \dots, R_p^{l_m}\}$ is expressed as follows:

$$\bar{R}_p(x, y) = \sum_{j=1}^{l_m} w_{Tj} \times R_p^j(x, y) \quad (9)$$

$$w_{Tj} = (1/S_j) / \sum_{j=1}^M (1/S_j) \quad (10)$$

where w_{Tj} denotes the temporal weight (w_T is used for representing ‘‘temporal weight’’ to distinguish it from the weight of similar grid cells) of the j th fused result R_p^j ($j \in \{1, 2, \dots, l_m\}$). S_j measures the temporal difference of PM_{2.5} concentrations from t_j to t_p , calculated as the mean absolute difference between I_j and I_p . As a larger difference indicates a smaller weight, the reciprocal of S_j is adopted and further normalized to $[0, 1]$ as the weight w_{Tj} . After averaging the fused grid cells in the same location, we finally fill all the missing grid cells in the estimate R_p and produce an estimate \bar{R}_p without gaps.

2) Optimization through error correction: After the spatiotemporal fusion, we can already obtain seamless PM_{2.5} estimates. It should be noted that, some satellite-derived estimates contain some validly retrieved grid cells before the spatiotemporal fusion, and these grid cells offer valuable information to reduce the uncertainty of the fused estimates. An error correction procedure was applied for optimizing the fused result. This procedure was originally presented by Pérez *et al.* [53] for seamless image cloning and has since been applied to recover cloudy pixels in remote sensing imagery [54], [55]. The error correction is performed in each gap region, and it achieves the optimization globally, rather than in a pixel-wise way. Specifically, we denote the spatial range of a gap region as Ω (the boundary of the gap is $\partial\Omega$), and its neighborhood with validly retrieved grid cells as Ω^* . Accordingly, as displayed in Fig. 4(a), the preliminary seamless estimate \bar{R}_p can be divided into two regions: a reconstructed patch $\bar{R}_p|_{\Omega}$ defined over Ω , which is full of fused grid cells, and its spatially adjacent region $R_p|_{\Omega^*}$ defined over Ω^* , which is full of validly retrieved grid cells. The reconstructed patch $\bar{R}_p|_{\Omega}$ is expected to be adjusted as an optimized patch $R_p'|_{\Omega}$ in

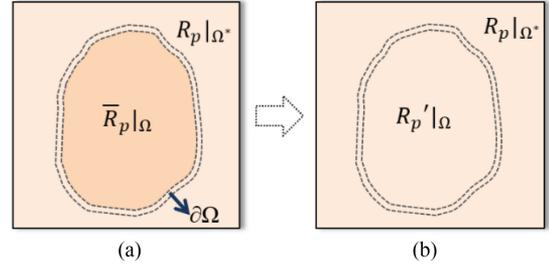


Fig. 4. Illustration of the error correction procedure. (a) Preliminary estimate from spatiotemporal fusion. (b) Seamless estimate after error correction.

Fig. 4(b), with the adjacent region $R_p|_{\Omega^*}$ serving as supporting information. To obtain the optimized patch $R_p'|_{\Omega}$, we employ the gradient of the reconstructed patch $\bar{R}_p|_{\Omega}$ as a vector field to guide the optimization, as well as a boundary condition to ensure the transition smoothness from the outside neighborhood $R_p|_{\Omega^*}$ to the recovered region $R_p'|_{\Omega}$. Mathematically, the optimization process is formulated as

$$\begin{aligned} R_p'|_{\Omega} &= \operatorname{argmin} \left(|\nabla R_p'|_{\Omega} - \nabla \bar{R}_p|_{\Omega}|^2 \right), \text{ with } R_p'|_{\partial\Omega} \\ &= R_p|_{\partial\Omega} \end{aligned} \quad (11)$$

where $\nabla = (\partial/\partial x, \partial/\partial y)$ is the gradient operator. To simplify the solving of (11), a residual term \widetilde{R}_p is defined to report the differences between the optimized patch $R_p'|_{\Omega}$ and the previously reconstructed patch \bar{R}_p over Ω

$$R_p'|_{\Omega} = \bar{R}_p|_{\Omega} + \widetilde{R}_p|_{\Omega}. \quad (12)$$

To obtain the residual term, we extract the residuals along the patch boundary, i.e., $(R_p - \bar{R}_p)$ defined over $\partial\Omega$, and the residual term can be regarded as the membrane interpolation of the residuals along the boundary, which is mathematically expressed as

$$\begin{aligned} \widetilde{R}_p|_{\Omega} &= \operatorname{argmin} \left(\Delta \widetilde{R}_p|_{\Omega} = 0 \right), \text{ with } \widetilde{R}_p|_{\partial\Omega} \\ &= (R_p - \bar{R}_p)|_{\partial\Omega} \end{aligned} \quad (13)$$

where $\Delta = (\partial^2/\partial x^2, \partial^2/\partial y^2)$ is the Laplacian operator. After getting the residual term $\widetilde{R}_p|_{\Omega}$, we can obtain the optimized patch $R_p'|_{\Omega}$ and produce the final reconstructed result.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The effectiveness of the Geoi-DBN model for retrieving PM_{2.5} concentrations from satellite-based AOD product has been investigated and verified in our previous work [46]. In this part, we focus on assessing the performance of the proposed reconstruction method. In Section IV-A and IV-B, the reconstruction performance is comprehensively evaluated from the area-based and point-based perspectives, respectively. The hourly consecutive PM_{2.5} estimates in one day are displayed in Section IV-C to demonstrate the superiority of the reconstructed estimates for revealing the diurnal PM_{2.5} change. Detailed discussions follow in Section IV-D to discuss the three key issues.

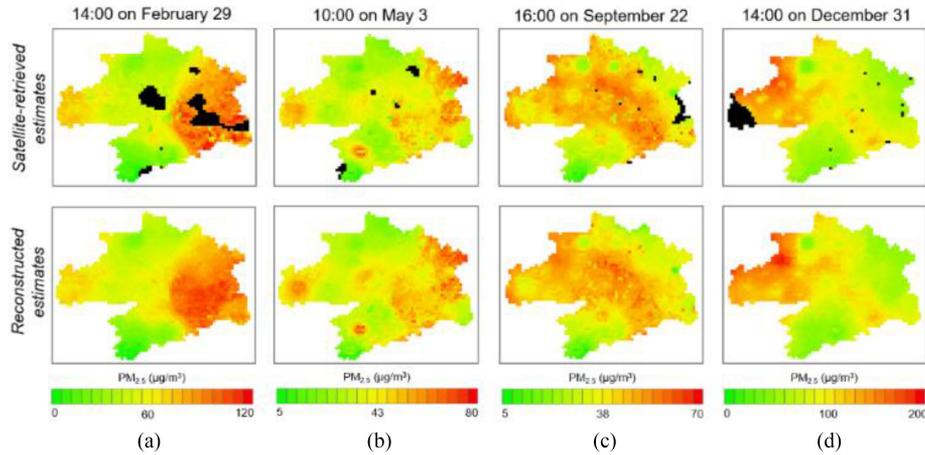


Fig. 5. Experimental results in the four validation tests. (a)–(d) Satellite-retrieved PM_{2.5} estimates and reconstructed PM_{2.5} results on 14:00 on February 29, 10:00 on May 3, 16:00 on September 22, and 14:00 on December 31, respectively. Please note that the satellite-retrieved estimates here are used as evaluation reference, because the reconstruction is performed on all the grid cells, including those that are not missing.

A. Area-Based Evaluation

The area-based evaluation scheme aims to investigate if the reconstructed results would be consistent with the satellite-retrieved PM_{2.5} estimates when the proposed method is carried out on the nonmissing grid cells in the satellite PM_{2.5} retrievals. The name “area-based evaluation” emphasizes the fact that the evaluation of the reconstruction performance in a continuous surface, rather than in a few discrete points. Due to the wide coverage of the evaluation reference, the area-based evaluation is able to reveal the ability of the reconstruction method for recovering PM_{2.5} spatial variation. Technically, a time point is identified at which the corresponding satellite-retrieved estimate shows high coverage. We pretend the corresponding estimate as completely missing and perform the proposed approach at this time point to produce a brand-new reconstructed estimate. It is noteworthy that all the grid cells in the reconstructed estimate are created from reconstruction, and they are independent from the corresponding satellite retrieval. The reconstructed result is then assessed against the correspondingly nonmissing grid cells in the satellite-retrieved estimate to reveal the reconstruction performance from the area-based perspective. In this article, we identified 321 time points to conduct the experiment, requiring that the effective-retrieved coverage rate in the estimate is over 70%. Please note that, due to the serious data incompleteness, chances are that the identified reference estimates are days apart from the target estimate in real applications, and thus the satellite-retrieved estimates with an interval of fewer than three days to the target estimate were rejected as reference in this validation experiment, so as not to result in an inflated accuracy. After the reconstruction, the produced results were visually and quantitatively assessed against the satellite-derived estimates based on the non-missing regions. A set of statistical measures were used to quantitatively assess the consistency between the reconstructed results and the satellite-retrieved estimates, including mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R^2). A new measure

Q is also defined to report the relative accuracy of reconstruction as follows:

$$Q = 1 - \left(\frac{1}{m} \times \sum_{i=1}^m \frac{|P_i - O_i|}{O_i} \right) \quad (14)$$

where O_i is the i th non-missing grid cell in the satellite-retrieved estimate, and P_i is the correspondingly reconstructed grid cell in the reconstructed result. m is the total number of grid cells involved in the evaluation. According to the definition, a Q value closer to 1 indicates a better reconstruction performance. The quantitative results were averaged amongst the 321 tests to allow a comprehensive assessment.

Fig. 5 shows the experimental results in four validation tests. The time points are 14:00 on February 29, 10:00 on May 3, 16:00 on September 22, and 14:00 on December 31, respectively, temporally covering different seasons in a year. The satellite-retrieved estimates in different seasons show varying PM_{2.5} levels and spatial patterns. By visually comparing the satellite-derived estimates and the reconstructed results, it can be found that they are significantly consistent and comparable, reflecting similar PM_{2.5} concentrations. The spatial patterns of PM_{2.5} are also successfully recovered in the reconstructed results. In addition, given the fact that PM_{2.5} changes dramatically with meteorological conditions in a short time lag, it is necessary to verify the ability of the proposed method to capture the dynamic changes. The experimental results of a series of consecutive hours on September 22 are illustrated in Fig. 6. As shown from the satellite-derived estimates, the PM_{2.5} peaked at 9:00 in the central and western part of the WUA, and since then it gradually dissipated to a low level until 17:00. Correspondingly, the reconstructed time series show similar spatial distributions and temporal dynamics to the satellite-retrieved data. These findings from visual aspect suggest the proposed approach is able to produce reconstructed results highly comparable to the satellite retrievals, and it shows strong ability to capture the

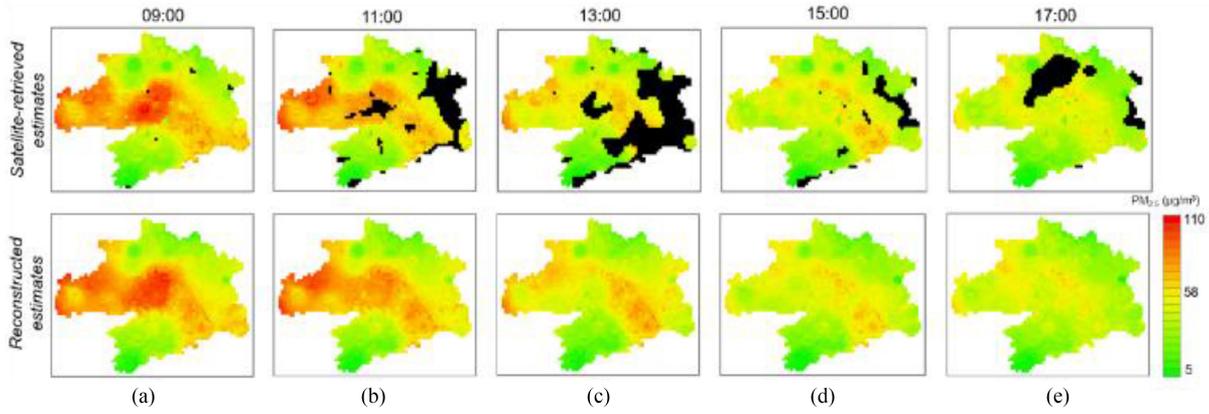


Fig. 6. Satellite-retrieved $PM_{2.5}$ estimates and reconstructed $PM_{2.5}$ results of the five daytime hours of (a) 9:00, (b) 11:00, (c) 13:00, (d) 15:00, and (e) 17:00 on September 22. Please note that the satellite-retrieved estimates here are used as evaluation basis, because the reconstruction is performed on all the grid cells, including those that are not missing.

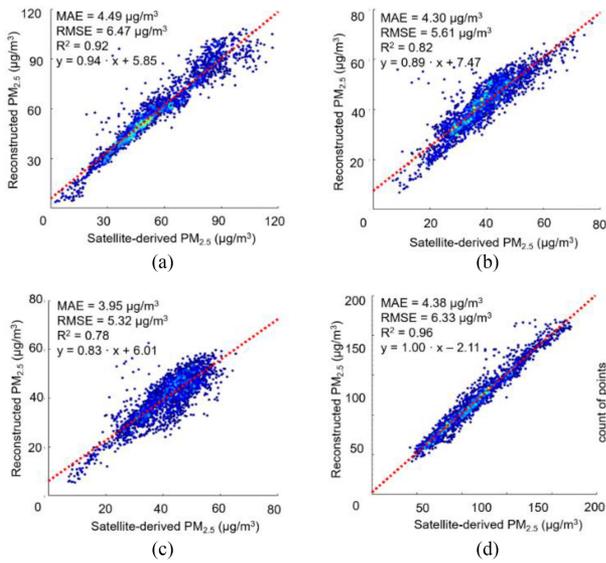


Fig. 7. Scatter plots of the reconstructed $PM_{2.5}$ results against the satellite-derived estimates. (a)–(d) Quantitative evaluation results for the four tests in Fig. 5.

$PM_{2.5}$ concentrations, spatial distributions, and temporal dynamics.

Fig. 7 illustrates the quantitative assessment results in the four validation tests in Fig. 5. The reconstructed grid cells are compared with the correspondingly retrieved grid cells to map the scatter plots and generate the quantitative descriptions in terms of MAE, RMSE, R^2 , and linear fitting coefficients. Generally speaking, a good performance of the proposed reconstruction method can be observed. The differences are generally small between the reconstructed $PM_{2.5}$ and the satellite-retrieved $PM_{2.5}$, with RMSE (MAE) being $5.32\text{--}6.47 \mu\text{g}/\text{m}^3$ ($3.95\text{--}4.49 \mu\text{g}/\text{m}^3$) in the four tests. Meanwhile, the R^2 values are $0.79\text{--}0.96$, indicating a significant correlation level. Interestingly, it can be found that both the R^2 and RMSE values are higher in winter than in the other seasons (i.e., R^2 : 0.96 on December 31 versus 0.82 on May 3; RMSE: $6.33 \mu\text{g}/\text{m}^3$ on December 31 versus $5.61 \mu\text{g}/\text{m}^3$

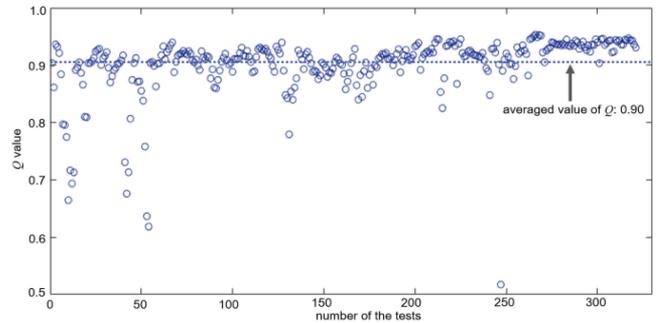


Fig. 8. Quantitative measure Q values in the 321 tests. The blue circles represent the Q values, and the dashed line indicates the averaged value.

on May 3). This can be attributed to the fact that winter has higher levels of $PM_{2.5}$ concentration than the other seasons [56]. To allow for comprehensive assessment, the quantitative measure values were averaged among the 321 tests. On average, the RMSE (MAE) value is $6.50 \mu\text{g}/\text{m}^3$ ($4.80 \mu\text{g}/\text{m}^3$) and the R^2 is 0.87 , indicating that this approach is capable of producing the reconstructed results highly consistent to the satellite-retrieved $PM_{2.5}$ concentrations. Fig. 8 shows the measured Q values in the 321 tests. Among the 321 tests, the averaged Q value reaches up to 0.90 , and 295 tests have a Q value of more than 0.85 , accounting for about $\sim 92\%$ in terms of the number of tests. This fully demonstrates the robustness of the proposed reconstruction method.

B. Point-Based Evaluation

The point-based evaluation scheme using the station measurements as reference aims to explore the consistency between the reconstructed results and the ground-based measurements. As the ground stations are located discretely, this scheme is referred to as “point-based evaluation” in this article. Technically, all the ground-based records located in the missing gaps of the year-long satellite-retrieved estimates are used as the evaluation reference. When a station record is adopted as reference, the information measured by this station is excluded from the

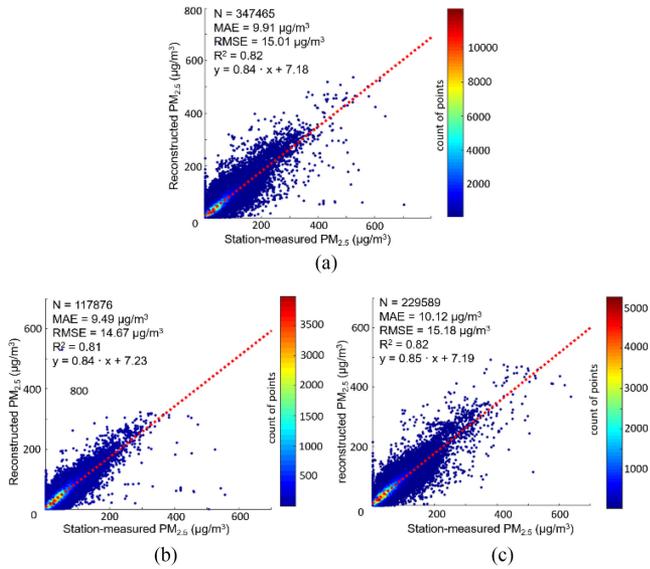


Fig. 9. Scatter plots of point-based samples in (a) the whole year, (b) in the daytime period, and (c) in the nighttime period.

reconstruction, i.e., the interpolation processes do not involve any records from this station. We fuse the satellite-derived estimates and the remaining ground-based records to reconstruct the missing gaps. The recovered PM_{2.5} over the validation station is extracted and paired with the corresponding PM_{2.5} record as a validation sample. In this experiment, we conducted the tests based on the station-measured records in the gap regions, and collected 347 465 validation samples in total. These validation samples are used to map the scatter plots and derive the statistical descriptions to assess the reconstruction performance.

Fig. 9(a) shows the scatter plot of the collected validation samples (the reconstructed PM_{2.5} against the station-measured PM_{2.5}) throughout the whole year, along with quantitative descriptions in terms of MAE, RMSE, R^2 , and linear fitting coefficients. Generally speaking, the RMSE (MAE) value is 15.01 $\mu\text{g}/\text{m}^3$ (9.91 $\mu\text{g}/\text{m}^3$) and the R^2 value is 0.82, indicating that the reconstructed PM_{2.5} present satisfactory correlation with the station-measured PM_{2.5}. On the other hand, the slope in the linear fitting is under 1, and the intercept is over 0. Although this suggests the proposed method tends to overestimate the low-level PM_{2.5} concentrations and underestimate the high-level PM_{2.5} concentrations, this problem is common to the satellite-based PM_{2.5} studies [6], [8]. In addition, given the fact that the nighttime (0:00–8:00 and 18:00–23:00) and daytime (9:00–17:00) periods have a different situation of data incompleteness, the reconstruction performance is assessed during the two periods separately. Fig. 9(b) and (c) shows the quantitative results of the daytime and nighttime samples. It can be observed that even though the nighttime estimates are completely missing, the consistency between the recovered results and the station records does not significantly vary during the two periods (RMSE: 14.67 $\mu\text{g}/\text{m}^3$ versus 15.18 $\mu\text{g}/\text{m}^3$; R^2 : 0.81 versus 0.82). The reason why the proposed approach can achieve a satisfactory nighttime performance is that the spatiotemporal

fusion technique incorporates the PM_{2.5} change into the model by employing the interpolated source as the guidance. According to the comparable quantitative results, we can conclude that the proposed method is almost free from the effect of the daytime and the nighttime periods, presenting a satisfactory performance in the two separate periods.

C. Mapping of the Hourly PM_{2.5} Estimates

Hourly PM_{2.5} estimates on February 26, 2016, are presented to visualize the performance of the proposed method. Fig. 10(a) shows the satellite-derived hourly PM_{2.5} estimates in the WUA on February 26, together with the ground station measurements. As mentioned before, the optical instrument, limited by the physical properties, could not observe the Earth's surface at night, resulting in the nighttime PM_{2.5} completely undetectable. The daytime PM_{2.5} shows a better condition. The PM_{2.5} distributions over almost the entire WUA are revealed from 13:00 to 15:00, but the missing gaps present in the southern part of the WUA from 9:00 to 12:00. The worst situation during the daytime period occurs at 9:00, with most areas in Wuhan, Ezhou, Huangshi, and Xianning unobserved. Although fitting the station records over the effectively retrieved regions, the satellite-derived PM_{2.5} estimates suffer from the missing gaps, which pose barriers to the hourly monitoring goal. Fig. 10(b) shows the seamless estimates reconstructed by the proposed method. It can be seen that the missing PM_{2.5} concentrations are successfully recovered, and the reconstructed PM_{2.5} in the boundary of the gaps show good adjacency to the effectively-retrieved PM_{2.5} in the neighborhood, without noticeable artifacts. In addition, the reconstructed PM_{2.5} concentrations also fits the station records, if they have monitoring stations nearby. The seamless estimates with full coverage show a significant superiority in revealing the spatial distribution and diurnal evolution of PM_{2.5} concentrations. In this case on February 26, the PM_{2.5} stayed at a relatively high level from 0:00 to 7:00, then declined gradually from 8:00 to 15:00, and finally remained stable at a low level after 16:00. Also, in the southeast part of the WUA, the satellite-retrieved estimates were missing from 10:00 to 12:00, while the reconstructed estimates indicate relatively high concentrations, which demonstrate the potential of the proposed method to reveal pollution events in the missing gaps. In summary, with the seamless PM_{2.5} estimates produced by the proposed method, the hourly monitoring of PM_{2.5} can be realized.

D. Discussion

1) *Justifying the Basic Assumption of Spatiotemporal Fusion*: The spatiotemporal fusion of station-interpolated PM_{2.5} and satellite-derived PM_{2.5} plays a vital role in the reconstruction process. It is assumed in the spatiotemporal fusion that the two data sources have a similar coefficient set describing the PM_{2.5} temporal change. The interpolated data source with a continuous timeline is used to model the PM_{2.5} temporal variation and the derived model is then applied to the satellite-retrieved data source to get the prediction. This section aims to justify the assumption of spatiotemporal fusion.

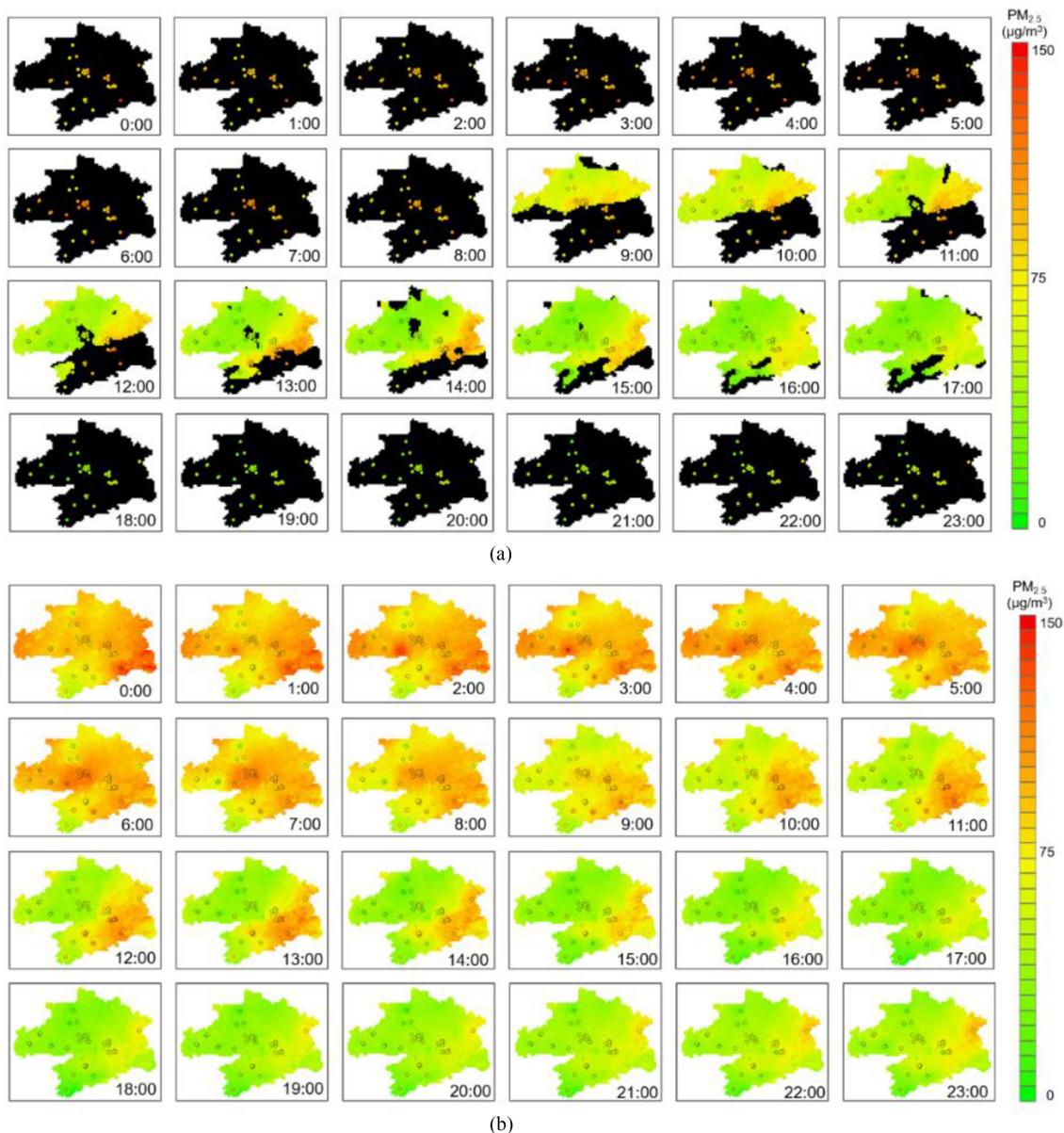


Fig. 10. Hourly PM_{2.5} estimates on February 26, 2016. (a) Satellite-derived estimates. (b) Reconstructed seamless estimates.

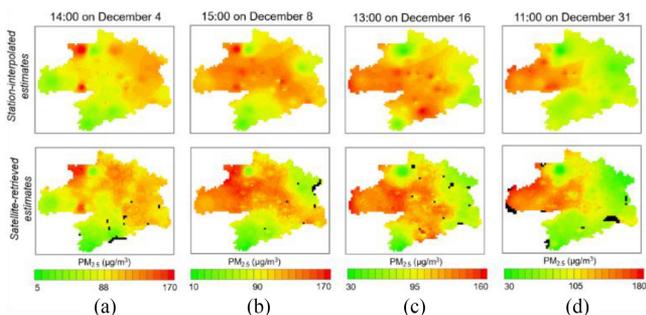


Fig. 11. Four pairs of PM_{2.5} estimate in December, 2016, used to warrant the assumption of spatiotemporal fusion. (a) 14:00 on December 4. (b) 15:00 on December 8. (c) 13:00 on December 16. (d) 11:00 on December 31.

Fig. 11 displays four pairs of PM_{2.5} estimates (i.e., satellite-retrieved estimates and station-interpolated estimates) in December 2016. Generally, the interpolated estimates roughly characterize the PM_{2.5} distributions but suffer from interpolation smoothness, while the satellite retrievals present clearer spatial variations of PM_{2.5}. Any two pairs are selected at random, with the former as the reference and the latter as the target. All the locations of grid cells are checked to compare the linear coefficients from the two data sources. Specifically, for a given grid cell location, a 5×5 local window is applied to the two interpolated estimates, and the grid cells inside the window at both the reference and the target time are linearly regressed to obtain the coefficient sets on the interpolated data source. Similarly, the local window is also applied to the two retrieved estimates to obtain the coefficient set on the retrieved data source.

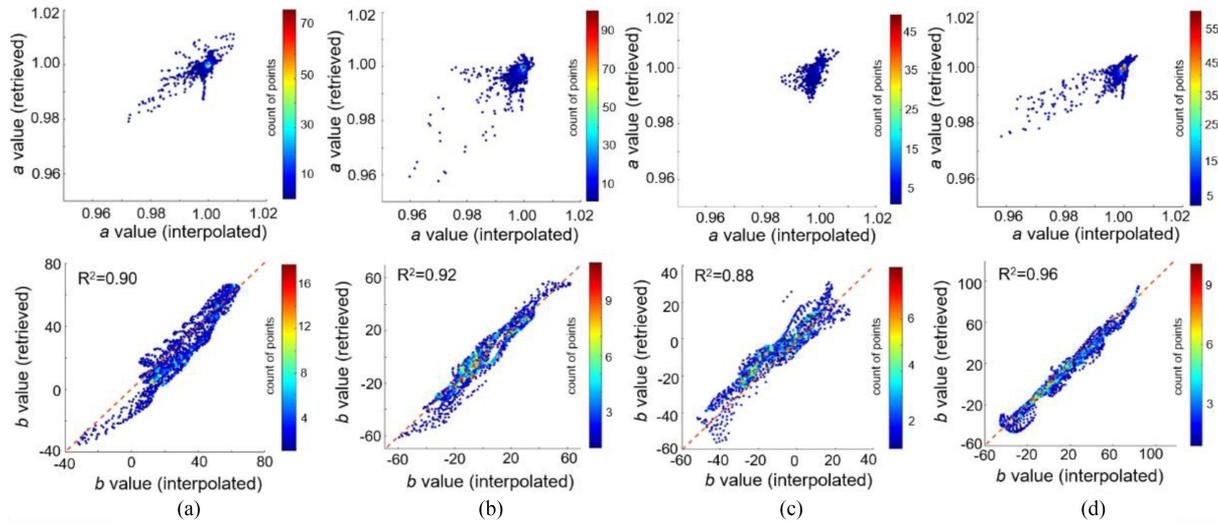


Fig. 12. Scatter plots for the estimated coefficient sets from two data sources. (a)–(d) Correspond to the four tests in Table I.

TABLE I
REFERENCE TIME, TARGET TIME, AND ACQUISITION INTERVALS
IN THE FOUR VALIDATION TESTS

Test	Reference time	Target time	Acquisition interval
#1	14:00, Dec. 4	15:00, Dec. 8	~ 4 days
#2	15:00, Dec. 8	13:00, Dec. 16	~ 8 days
#3	13:00, Dec. 16	11:00, Dec. 31	~ 15 days
#4	14:00, Dec. 4	11:00, Dec. 31	~ 27 days

The two coefficient sets a and b are compared at all the locations over the study area, and if they are generally similar, the basic assumption is justified. Accordingly, we performed four tests with by using different data pairs (see Table I), and the time intervals last 4–27 days in these tests. Fig. 12 shows the scatter plots of the derived linear coefficients in the four tests. It can be found that the coefficients a from the two data sources vary in the vicinity of 1, with most of the cases falling in the range of [0.98, 1.01]. The highly fixed range guarantees consistency of slope coefficient from the two data sources. In terms of the intercept coefficient b which plays a major role in the temporal model, the two data sources show good correlation ($R^2 > = 0.88$), and the scatter points are distributed around the 1:1 line (i.e., the orange dashed line). Please note that even though the time interval increases from several days to nearly four weeks, the derived coefficient sets still show good consistency. Thus, the aforementioned assumption of spatiotemporal fusion can be confirmed.

2) *Comparison of the Two Evaluation Schemes*: In this article, we assessed the performance of the proposed method from multidimensional perspectives. Two schemes were adopted, namely, the point-based evaluation and the area-based evaluation. Table II gives a brief comparison between the two schemes. The point-based scheme takes ground-based PM_{2.5} observations monitored by stations as evaluation reference. This scheme is widely used [6], [56], as the ground-based observations are highly precise to record PM_{2.5} concentrations, but it should be

highlighted that it has intrinsic shortcomings. First, the reported precision show dependence on the locations of the monitoring stations, and it would vary even significantly at different test sites. Second, since only the station locations are involved, this scheme is unable to reveal the model ability to predict spatial variations of the PM_{2.5} concentrations over the study area. On the contrary, the area-based scheme shows contrasting traits and makes a good supplement to the assessment process [52]. This scheme aims to check if the reconstructed seamless results are consistent with the satellite-based PM_{2.5} retrievals when the proposed method is implemented on the non-missing grid cells in the satellite retrievals. Technically, we carried out the proposed method on the grid cells that are not missing in the satellite-based PM_{2.5} estimates, and compare the reconstructed PM_{2.5} grid cells with the effectively retrieved grid cells to assess the reconstruction accuracy. The reported precision would be spatially representative due to the wide spatial coverage of evaluation reference. Also, the ability to predict the spatial variations of PM_{2.5} concentrations can be assessed. The synergistic use of the two schemes in this article helps to confirm the effectiveness of the validation tests.

3) *Effect of Station Density on the Reconstruction Performance*: The density and distribution of the monitoring stations have an effect on the quality of the interpolated estimates, resulting in a follow-up influence on the reconstruction performance. This section aims to investigate the effect. The collected samples in the point-based evaluation were grouped according to the station locations, and the R^2 values of the grouped samples were calculated and mapped, as shown in Fig. 13. The spatial pattern of R^2 reveals the effect. The stations clustering in the urban areas of Wuhan, Ezhou, and Xianning (see the blue dashed circles) show higher R^2 values of over 0.80, while those distributed sparsely over rural areas show lower values of under 0.65. This implies the reconstruction performance has an association with the station density, and better performance can be expected over regions with denser monitoring stations. It should be noted that,

TABLE II
COMPARISON BETWEEN THE TWO EVALUATION SCHEMES

	Evaluation reference	Advantage	Disadvantage
Point-based	Ground-based PM _{2.5} measurements	The evaluation reference (i.e., ground-based observations) is highly precise and has little uncertainty.	Only a few locations with station monitors are used to assess the accuracy. The reported precision significantly depends on the station locations, and it would vary at different test sites.
Area-based	Satellite-based PM _{2.5} retrievals	The satellite-based retrievals have wider coverage as compared with ground-based observations. The reported precision would be representative due to the wide coverage of evaluation basis.	The evaluation reference (i.e., satellite-based retrievals) has some uncertainty due to retrieval errors.

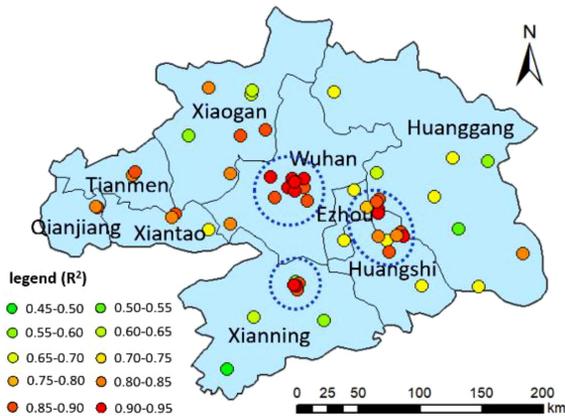


Fig. 13. R^2 values of the samples grouped by the station locations.

even though the effect exists, among the 54 stations in the WUA, 38 stations (i.e., accounting for 70% of the stations) distributed over the WUA show a strong correlation ($R^2 > 0.70$), while only one station in Xianning near the boundary of the WUA shows a weak correlation ($R^2 = 0.49$). Thus, the proposed reconstruction strategy is generally applicable for the WUA with a high density of monitoring stations.

As the importance of air quality is increasingly realized, the government is launching plans to enhance the air quality monitoring network, and more monitoring stations will be established. Denser distribution of stations is expected to further improve the reconstruction performance of the proposed method. For the regional-scale applications with insufficient monitoring stations, given that air quality models can simulate PM_{2.5} concentrations based on emissions inventories and chemical transport processes [57]–[59], a potential solution is to replace the interpolated estimates with the model simulations. In this case, the proposed method can be expected to work over regions with sparse monitoring stations.

V. CONCLUSION

The generation of hourly seamless PM_{2.5} estimates over the WUA has been performed by jointly using the station-measured PM_{2.5}, satellite-derived AOD, and auxiliary predictor variables. The Geoi-DBN model is applied to retrieve PM_{2.5} from AOD observations, and then a data reconstruction method is proposed to fill the missing gap in the retrieved estimates. Specifically, the developed reconstruction method successfully integrates the

advantages of the two input components, i.e., the continuous timeline of the station measurements and the spatial coverage of the satellite-retrieved estimates, and it produces the full-coverage PM_{2.5} estimates at an hourly frequency. Evaluation of the model performance has been conducted from multidimensional perspectives. The overall values of MAE, RMSE, and R^2 are $9.91 \mu\text{g}/\text{m}^3$, $15.01 \mu\text{g}/\text{m}^3$, and 0.82 in the point-based evaluation and $4.80 \mu\text{g}/\text{m}^3$, $6.50 \mu\text{g}/\text{m}^3$, and 0.87 in the area-based evaluation, indicating an excellent model performance. Besides, although the nighttime retrievals are completely missing, the reconstruction method can obtain a comparable performance in the nighttime to that in the daytime.

The presented method for generating hourly seamless PM_{2.5} estimates meets the demand for spatially and temporal continuous monitoring of PM_{2.5} distributions. It forms the basis for the target of the near real-time monitoring. Future work will be focusing on extending the framework to the regional scale for full-coverage estimation by combining the station measurements, satellite observations, and model simulations.

ACKNOWLEDGMENT

We acknowledge all the institutions for providing the data free of charge. In addition, the constructive comments and suggestions from the handling editor and anonymous reviewers are gratefully acknowledged.

REFERENCES

- [1] F. Dominici *et al.*, “Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases,” *JAMA*, vol. 295, no. 10, pp. 1127–1134, 2006.
- [2] J. D. Sacks *et al.*, “Particulate matter-induced health effects: Who is susceptible?,” *Environ. Health Perspectives*, vol. 119, no. 4, pp. 446–454, 2011.
- [3] S. M. Bartell, J. Longhurst, T. Tjoa, C. Sioutas, and R. J. Delfino, “Particulate air pollution, ambulatory heart rate variability, and cardiac arrhythmia in retirement community residents with coronary artery disease,” *Environ. Health Perspectives*, vol. 121, no. 10, pp. 1135–1141, 2013.
- [4] C. Zhao, X. Tie, and Y. Lin, “A possible positive feedback of reduction of precipitation and increase in aerosols over Eastern Central China,” *Geophysical Res. Lett.*, vol. 33, no. 11, 2006, Art. no. L11814.
- [5] J. - J. Cao *et al.*, “Winter and summer PM_{2.5} chemical compositions in fourteen Chinese cities,” *J. Air Waste Manage. Assoc.*, vol. 62, no. 10, pp. 1214–1226, 2012.
- [6] 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, 2014.
- [7] B. Lv, Y. Hu, H. H. Chang, A. G. Russell, and Y. Bai, “Improving the accuracy of daily PM_{2.5} distributions derived from the fusion of ground-level measurements with aerosol optical depth observations, a case study in North China,” *Environ. Sci. Technol.*, vol. 50, no. 9, pp. 4752–4759, 2016.

- [8] 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," *Atmospheric Environ.*, vol. 152, pp. 477–489, 2017.
- [9] Z. Fan, Q. Zhan, C. Yang, H. Liu, and M. Bilal, "Estimating PM_{2.5} concentrations using spatially local xgboost based on full-covered SARA AOD at the urban scale," *Remote Sens.*, vol. 12, no. 20, 2020, Art. no. 3368.
- [10] A. Mhawish *et al.*, "Estimation of high-resolution PM_{2.5} over the indogangetic plain by fusion of satellite data, meteorology, and land use variables," *Environ. Sci. Technol.*, vol. 54, no. 13, pp. 7891–7900, 2020.
- [11] Y. Shi, M. Bilal, H. C. Ho, and A. Omar, "Urbanization and regional air pollution across South Asian developing countries – a nationwide land use regression for ambient PM_{2.5} assessment in Pakistan," *Environ. Pollut.*, vol. 266, 2020, Art. no. 115145.
- [12] W. Wang, F. Mao, L. Du, Z. Pan, W. Gong, and S. Fang, "Deriving hourly PM_{2.5} concentrations from himawari-8 AODs over Beijing–Tianjin–Hebei in China," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 858.
- [13] Y. Che, Y. Xue, J. Guang, L. She, and J. Guo, "Evaluation of the AVHRR deepblue aerosol optical depth dataset over mainland China," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 74–90, 2018.
- [14] F. Yao, J. Wu, W. Li, and J. Peng, "A spatially structured adaptive two-stage model for retrieving ground-level PM_{2.5} concentrations from VIIRS AOD in china," *ISPRS J. Photogramm. Remote Sens.*, vol. 151, pp. 263–276, 2019.
- [15] X. Yang, C. Zhao, N. Luo, W. Zhao, W. Shi, and X. Yan, "Evaluation and comparison of himawari-8 L2 V1.0, V2.1 and MODIS C6.1 aerosol products over Asia and the Oceania regions," *Atmospheric Environ.*, vol. 220, 2020, Art. no. 117068.
- [16] A. Retalis and N. Sifakis, "Urban aerosol mapping over Athens using the differential textural analysis (DTA) algorithm on MERIS-ENVISAT data," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 17–25, 2010.
- [17] K. Themistocleous and D. G. Hadjimitsis, "Development of an image based integrated method for determining and mapping aerosol optical thickness (AOT) over urban areas using the darkest pixel atmospheric correction method, RT equation and GIS: A case study of the limassol area in cyprus," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 1–10, 2013.
- [18] A. van Donkelaar *et al.*, "Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application," *Environ. Health Perspectives*, vol. 118, no. 6, pp. 847–855, 2010.
- [19] C. Lin, Y. Li, Z. Yuan, A. K. H. Lau, C. Li, and J. C. H. Fung, "Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM_{2.5}," *Remote Sens. Environ.*, vol. 156, pp. 117–128, 2015.
- [20] Y. Zhang and Z. Li, "Remote sensing of atmospheric fine particulate matter (PM_{2.5}) mass concentration near the ground from satellite observation," *Remote Sens. Environ.*, vol. 160, pp. 252–262, 2015.
- [21] Y. Wu *et al.*, "Synergy of satellite and ground based observations in estimation of particulate matter in eastern china," *Sci. Total Environ.*, vol. 433, pp. 20–30, 2012.
- [22] N. Benas, A. Beloconi, and N. Chrysoulakis, "Estimation of urban PM₁₀ concentration, based on MODIS and MERIS/AATSR synergistic observations," *Atmospheric Environ.*, vol. 79, pp. 448–454, 2013.
- [23] X. Hu *et al.*, "Estimating ground-level PM_{2.5} concentrations in the southeastern U.S. using geographically weighted regression," *Environ. Res.*, vol. 121, pp. 1–10, 2013.
- [24] W. You, Z. Zang, L. Zhang, Z. Li, D. Chen, and G. Zhang, "Estimating ground-level PM₁₀ concentration in northwestern china using geographically weighted regression based on satellite AOD combined with CALIPSO and MODIS fire count," *Remote Sens. Environ.*, vol. 168, pp. 276–285, 2015.
- [25] W. Fan, K. Qin, Y. Cui, D. Li, and M. Bilal, "Estimation of hourly ground-level PM_{2.5} concentration based on himawari-8 apparent reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 76–85, Jan. 2021.
- [26] Q. Zeng *et al.*, "Satellite-based estimation of hourly PM_{2.5} concentrations using a vertical-humidity correction method from Himawari-AOD in Hebei," *Sensors*, vol. 18, no. 10, 2018, Art. no. 3456.
- [27] D. Fu *et al.*, "Mapping nighttime PM_{2.5} from VIIRS DNB using a linear mixed-effect model," *Atmospheric Environ.*, vol. 178, pp. 214–222, 2018.
- [28] H. Xu *et al.*, "A consistent aerosol optical depth (AOD) dataset over mainland china by integration of several AOD products," *Atmospheric Environ.*, vol. 114, pp. 48–56, 2015.
- [29] G. Geng *et al.*, "Estimating long-term PM_{2.5} concentrations in china using satellite-based aerosol optical depth and a chemical transport model," *Remote Sens. Environ.*, vol. 166, pp. 262–270, 2015.
- [30] Y. Zheng, Q. Zhang, Y. Liu, G. Geng, and K. He, "Estimating ground-level PM_{2.5} concentrations over three megalopolises in china using satellite-derived aerosol optical depth measurements," *Atmospheric Environ.*, vol. 124, pp. 232–242, 2016.
- [31] A. C. Just *et al.*, "Using high-resolution satellite aerosol optical depth to estimate daily PM_{2.5} geographical distribution in mexico city," *Environ. Sci. Technol.*, vol. 49, no. 14, pp. 8576–8584, 2015.
- [32] H. Hu *et al.*, "Satellite-based high-resolution mapping of ground-level PM_{2.5} concentrations over east china using a spatiotemporal regression Kriging model," *Sci. Total Environ.*, vol. 672, pp. 479–490, 2019.
- [33] Q. Xiao *et al.*, "Evaluation of gap-filling approaches in satellite-based daily PM_{2.5} prediction models," *Atmospheric Environ.*, vol. 244, 2021, Art. no. 117921.
- [34] I. Kloog, P. Koutrakis, B. A. Coull, H. J. Lee, and J. Schwartz, "Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements," *Atmospheric Environ.*, vol. 45, no. 35, pp. 6267–6275, 2011.
- [35] 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, 2017.
- [36] B. Wang, Q. Yuan, Q. Yang, L. Zhu, T. Li, and L. Zhang, "Estimate hourly PM_{2.5} concentrations from himawari-8 TOA reflectance directly using geo-intelligent long short-term memory network," *Environ. Pollut.*, vol. 271, 2021, Art. no. 116327.
- [37] C. Zeng *et al.*, "Exploration on the spatial spillover effect of infrastructure network on urbanization: A case study in Wuhan urban agglomeration," *Sustain. Cities Soc.*, vol. 47, 2019, Art. no. 101476.
- [38] L. Zhang, M. Zhang, and Y. Yao, "Mapping seasonal impervious surface dynamics in wuhan urban agglomeration, china from 2000 to 2016," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 70, pp. 51–61, 2018.
- [39] L. Guo, J. Luo, M. Yuan, Y. Huang, H. Shen, and T. Li, "The influence of urban planning factors on PM_{2.5} pollution exposure and implications: A case study in china based on remote sensing, LBS, and GIS data," *Sci. Total Environ.*, vol. 659, pp. 1585–1596, 2019.
- [40] L. Wang, W. Gong, J. Li, Y. Ma, and B. Hu, "Empirical studies of cloud effects on ultraviolet radiation in central china," *Int. J. Climatol.*, vol. 34, no. 7, pp. 2218–2228, 2014.
- [41] L. Wang, W. Gong, X. Xia, J. Zhu, J. Li, and Z. Zhu, "Long-term observations of aerosol optical properties at wuhan, an urban site in central china," *Atmospheric Environ.*, vol. 101, pp. 94–102, 2015.
- [42] J. Engel-Cox, N. T. Kim Oanh, A. van Donkelaar, R. V. Martin, and E. Zell, "Toward the next generation of air quality monitoring: Particulate matter," *Atmospheric Environ.*, vol. 80, pp. 584–590, 2013.
- [43] S. Fukuda *et al.*, "New approaches to removing cloud shadows and evaluating the 380nm surface reflectance for improved aerosol optical thickness retrievals from the GOSAT/TANSO-Cloud and aerosol imager," *J. Geophys. Res. Atmos.*, vol. 118, no. 24, pp. 13,520–13,531, 2013.
- [44] L. Gao, L. Chen, C. Li, J. Li, H. Che, and Y. Zhang, "Evaluation and possible uncertainty source analysis of JAXA himawari-8 aerosol optical depth product over China," *Atmospheric Res.*, vol. 248, 2021, Art. no. 105248.
- [45] Q. Yang, Q. Yuan, T. Li, H. Shen, and L. Zhang, "The relationships between PM_{2.5} and meteorological factors in china: Seasonal and regional variations," *Int. J. Environ. Res. Public Health*, vol. 14, no. 12, 2017, Art. no. 1510.
- [46] 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, no. 23, pp. 11,985–11,993, 2017.
- [47] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019.
- [48] X. Zhu, F. Cai, J. Tian, and T. K.-A. Williams, "Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 527.
- [49] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.
- [50] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, 2010.

- [51] J. Wu, Q. Cheng, H. Li, S. Li, X. Guan, and H. Shen, "Spatiotemporal fusion with only two remote sensing images as input," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6206–6219, Oct. 2020.
- [52] Q. Cheng, H. Liu, H. Shen, P. Wu, and L. Zhang, "A spatial and temporal nonlocal filter-based data fusion method," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4476–4488, Aug. 2017.
- [53] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, 2003.
- [54] C. H. Lin, K. H. Lai, Z. B. Chen, and J. Y. Chen, "Patch-based information reconstruction of cloud-contaminated multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 163–174, Jan. 2014.
- [55] H. Shen, J. Wu, Q. Cheng, M. Aihemaiti, C. Zhang, and Z. Li, "A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 862–874, Mar. 2019.
- [56] 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, 2018.
- [57] T. Xue *et al.*, "Fusing observational, satellite remote sensing and air quality model simulated data to estimate spatiotemporal variations of PM_{2.5} exposure in China," *Remote Sens.*, vol. 9, no. 3, 2017, Art. no. 221.
- [58] Z. Song *et al.*, "Diurnal and seasonal variability of PM_{2.5} and AOD in north china plain: Comparison of MERRA-2 products and ground measurements," *Atmospheric Environ.*, vol. 191, pp. 70–78, 2018.
- [59] L. He, A. Lin, X. Chen, H. Zhou, Z. Zhou, and P. He, "Assessment of MERRA-2 surface PM_{2.5} over the Yangtze river basin: Ground-based verification, spatiotemporal distribution and meteorological dependence," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 460.



Chengyue Zhang received the M.S. degree in surveying and mapping engineering in 2018 from Wuhan University, Wuhan, China, where he is currently working toward the Ph.D. degree in cartography and geographic information engineering.

His current main research interests include are in urban big data and complex networks.



Qing Cheng received the B.S. degree in the geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2010 and 2015, respectively.

She is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. Her research interests include remote sensing data reconstruction, data fusion, and urban remote sensing.



Jingan Wu received the B.S. degree in geographic information system from Anhui University, Hefei, China, in 2015, and the Ph.D. degree in cartography and geographic information engineering from Wuhan University, Wuhan, China, in 2021.

He is currently a Postdoctoral Research Assistant with the School of Geospatial Engineering and Science, Sun Yat-sen University, Guangzhou, China. His research interests include spatiotemporal data fusion and missing information reconstruction of remote sensing images.



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 currently an assistant professor with the School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou, China. He has authored or coauthored more than 20 papers 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 (SRES), Wuhan University, where he is currently a Luojia Distinguished Professor and an Associate Dean with the SRES. He was or is the PI of two projects supported by National Key Research and Development Program of China, and six projects supported by National Natural Science Foundation of China. He has authored more than 100 research papers in peer-reviewed international journals. His research interests include remote sensing image processing, multisource data fusion, and intelligent environmental sensing.

Dr. Shen is a Council Member of China Association of Remote Sensing Application, Education Committee Member of Chinese Society for Geodesy Photogrammetry and Cartography, and Theory Committee Member of Chinese Society for Geospatial Information Society. He is currently a member of the Editorial Board of *Journal of Applied Remote Sensing and Geography and Geo-information Science*.