Evaluation of LJ1-01 Nighttime Light Imagery for Estimating Monthly $PM_{2.5}$ Concentration: A Comparison With NPP-VIIRS Nighttime Light Data

Guo Zhang, Yingrui Shi, and Miaozhong Xu

***Abstract*—Air quality degradations caused by fine particulate matter ($PM_{2.5}$) can lead to various health problems, and accurate $PM_{2.5}$ data are critical for managing the environment and ensuring public health. Radiation signals collected by nighttime light (NTL) remote sensing satellites are influenced by $PM_{2.5}$ concentrations, and thus, incorporating NTL imagery in statistical models has been widely used to predict $PM_{2.5}$ concentrations. However, scarce work has been carried out with new-generation NTL data from the LJ1-01 satellite, which has a fine spatial resolution and wide measurement range. In this study, we integrated satellite observation data and meteorological data to construct five models based on the geographically weighted regression to validate the feasibility of LJ1-01/NPP-VIIRS in Moderate Resolution Imaging Spectroradiometer AOD-based $PM_{2.5}$ prediction in the Beijing–Tianjin–Hebei region. The models were validated by the cross-validation method. The results showed that the addition of NTL information could improve the performance of the $PM_{2.5}$ prediction model. The seasonal R^2 with NTL in AOD-$PM_{2.5}$ model have improved by 5.07%, 4.50%, 2.95%, and 2.56% in model fitting and 1.20%, 1.75%, 2.20%, and 4.41% in cross-validation. Furthermore, the LJ1-01 NTL data revealed additional details and improved the prediction accuracy, compared with the NPP-VIIRS in AOD-$PM_{2.5}$ model, the seasonal R^2 with LJ1-01 in AOD-$PM_{2.5}$ model increased by 1.16%, 1.79%, 0.76%, and 1.15% in model fitting and 1.04%, 0.85%, 0.78%, 1.37% in cross-validation. Thus, our findings indicate that LJ1-01 and NTL data have the potential for predicting $PM_{2.5}$ and that they could constitute a useful supplemental data source for estimating ground-level $PM_{2.5}$ distributions.**

***Index Terms*—Geographically weighted regression (GWR) model, LJ1-01, moderate resolution imaging spectroradiometer (MODIS) aerosol optical depth (AOD), NPP-VIIRS, particulate matter ($PM_{2.5}$) concentration, satellite observation.**

I. Introduction

IN RECENT decades, rapid economic development and industrial construction activities in China have caused severe environmental pollution problems, particularly in relation to air pollution [1]. Fine particulate matter ($PM_{2.5}$) is now a major air pollutant in many regions. Numerous studies have shown that the incidence of cardiovascular and respiratory diseases is closely associated with the timing and intensity of exposures to $PM_{2.5}$ [2], [3]. Long-term exposures can accelerate declines in lung function and increase mortality [4]–[6]. Therefore, accurate assessments of air pollution levels and characterization of the spatial and temporal changes in $PM_{2.5}$ concentrations are important for formulating effective pollution prevention and control measures; these issues have received considerable attention from current researchers in related fields [7].

Ground-based $PM_{2.5}$ monitoring networks can provide accurate and real information on $PM_{2.5}$ concentrations and components. However, such network sites are sparsely distributed, mostly in urban areas, which means that the spatial coverage of routine measurement data is limited, and observed data are only available at certain times and specific sites [8], [9]. In contrast, because of the large spatial coverage and reliability of repeated measurements, satellite remote sensing provides a potentially cost-effective method for assessing and predicting $PM_{2.5}$ concentrations, which can help to supplement the sparse distribution of the $PM_{2.5}$ ground-monitoring stations [10], [11]. The most popular satellite-derived product for estimating surface $PM_{2.5}$ concentrations is aerosol optical depth (AOD). AOD is indicative of the integrated light extinction of particles in the atmosphere [12].

Many satellite sensors such as moderate resolution imaging spectroradiometer (MODIS), multi-angle imaging spectroradiometer, and sea-viewing wide field-of-view sensor collect aerosol scattering and absorption information in the atmosphere [13], [14]. Among them, the potential of using MODIS-based AOD to derive surface $PM_{2.5}$ concentrations was demonstrated. The latest MODIS AOD products, Collection 6, was constructed from MODIS imagery via both the enhanced deep blue algorithm and the dark target algorithm, which is adaptable for both dark and bright surfaces. It has been validated by AOD observations from Aerosol Robotic Network sites in China and the results were satisfactory [10], [15], [16]. MODIS-retrieved data are obtained from spectral observations of visible and near-infrared wavelengths, and the corresponding AOD particle size range is 0.1–2 μm, which is very close to the $PM_{2.5}$ particle size range; thus, these data have provided an important theoretical basis for exploring the relationship between satellite observed AOD and $PM_{2.5}$ [17].

Manuscript received February 29, 2020; revised April 30, 2020 and May 27, 2020; accepted June 4, 2020. Date of publication June 18, 2020; date of current version July 5, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41901400, and in part by the China Postdoctoral Science Foundation under Grant 2018M632919, and in part by the National High Resolution Earth Observation Foundation under Grant 11-Y20A12-9001-17/18. *(Corresponding author: Miaozhong Xu.)*

The authors are with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: guozhang@whu.edu.cn; 2018206190030@whu.edu.cn; mzxu6319@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3002671

Different statistical models have been proposed to establish quantitative relationships for AOD-$PM_{2.5}$, e.g., linear mixed effects models [18], generalized linear regression [19], generalized additive models [20], and geographically weighted regression (GWR) models [10], geographically and temporarily weighted regression model [17], and two-step models [21]. A GWR model adopts a local regression model to embed the spatial position of data into the regression parameters, and it uses the local weighted least-square method to estimate the point-by-point parameters. The nonstationarity of spatial relations can be directly detected by the changes in the predicted values of parameters in each spatial position. This estimation method is simple and easy to operate. The obtained estimation parameters can be used for statistical tests, with good applicability to a wide range of applications. Jiang *et al*. [22] utilized a GWR model to integrate meteorological and geographical factors to explain the generation and dilution of $PM_{2.5}$ concentrations in the Yangtze River Delta, and their research pointed out the significance of using proper auxiliary variables in modeling the AOD–$PM_{2.5}$ relationships. Li *et al*. [23] used satellite-retrieved NDVI, AOD, and the nighttime light (NTL) information for $PM_{2.5}$ concentration predictions in a GWR model in the Northeastern United States. The results indicated that the combination of NTL imagery and NDVI is promising for providing additional information for $PM_{2.5}$ monitoring and prediction. Ma *et al*. [24] developed a national-scale GWR model to estimate $PM_{2.5}$ concentrations in China with fused satellite AOD as the primary predictor, meteorological, and land use information as the auxiliary variables. The results confirmed that satellite-derived AOD in conjunction with auxiliary information can be successfully applied to the extension of the ground $PM_{2.5}$ monitoring network in China, and this greatly improved the model performance.

The NTL remote sensing satellites have the capability to detect faint sources emitted on the Earth's surface, this ability allows for the capturing nighttime signals of light on the Earth's surface [25]. These NTL radiation signals are scattered by aerosol particles in the air such as $PM_{2.5}$. Therefore, the radiation signals collected by NTL remote sensing satellites will be influenced by $PM_{2.5}$ concentrations [26]. In addition, the NTL imagery is also shown to be a good indicator of urbanization and human activity. Previous research has demonstrated that population, economic growth, and urban expansion are the three main driving forces that impact $PM_{2.5}$ concentrations [27], [28]. Thus, the NTL data provided by NPP-VIIRS have been used extensively as proxies for the urban geographic footprint of human activities and is also for $PM_{2.5}$ concentration [29], [30]. The predecessor NTL imagery is mainly derived from the Suomi National Polar-orbiting Partnership Satellite (NPP-VIIRS). VIIRS has 22 channels, with a nominal spatial resolution of 375 m in the five imagery bands (I-bands) and 750 m in 16 moderate-resolution bands (M-bands), covering a spectral range from 0.41 to 12.01 μm. It uses a whiskbroom procedure, scanning the earth perpendicular to the track of the satellite. The unique Day–Night band (DNB) that measures radiances over a broadband spectrum from 0.4 to 0.9 μm with a nominal spatial resolution of 750 m is designed to detect radiance during the night. It has 32 symmetrical aggregation schemes on each side of the scan, resulting in similar pixel size (742 m) throughout the whole scan. The unique aggregating scheme in DNB also removes the "bowtie effect" (the pixel area overlap) completely. Consequently, the pixels in DNB NTL imagery have approximately the same size throughout the whole scan [31]. Furthermore, the onboard calibrations are also conducted on DNB NTL imagery to offer a series of high-quality imagery. Zhang and Hu [32] integrated MODIS AOD products, meteorological factors, and NPP-VIIRS NTL data as predictors to predict $PM_{2.5}$ concentrations, and the results indicated that NPP-VIIRS NTL data have potential for improving $PM_{2.5}$ concentrations prediction and serving as a useful supplemental data source for estimating ground-level $PM_{2.5}$ distributions [23]. However, the widespread application of NPP-VIIRS NTL data is mostly limited to projects at a moderate spatial resolution.

TABLE I
SENSOR PARAMETERS OF LJ1-01

Sensor Parameter	Value
Number of active detectors	2048 × 2048
Detector size	11 μm × 11 μm
Imaging mode	Standard (STD) mode High dynamic range (HDR) mode
Spectral range	460–980 nm
Resolution	129 m
Shutter type	Electronic rolling
Quantization bits	12-bit, processing to 15bit
Frame rate	24 fps @HDR mode 48 fps @STD mode
Noise Equivalent Radiance	$5e^{-5}$ W/m^2/sr

As a successor, the new generation NTL remote sensing satellite named LJ1-01 was successfully launched on June 2, 2018 by Wuhan University in China. It is a 6-kg cube scientific experimental satellite, equipped with a 129-m resolution NTL remote sensing sensor and navigation enhancement load. The frame push-broom imaging mode, the main imaging mode of the LJ1-01 camera, uses its camera system to capture the NTL on the ground. The sensor parameters are listed in Table I. Furthermore, DNB and LJ1-01 have similar spectral response function, their dominated outdoor lights for the broadband radiance are from pressure sodium lamps, fluorescent lamps, and light-emitting diode lamps. These lamps are primarily in the visible spectrum less than 0.65 μm by comparing the spectral radiances emitted from these three types of lamps with the spectral transmittance of in the LJ1-01 spectrum, they do not have the spectra overlapped with these gas absorption and water vapor spectra. Thus, water vapor and gas absorption has a negligible effect on the DNB and LJ1-01 radiance [33]. The outstanding features of LJ1-01 are also that:

1) have a broad spectral coverage (of 0.46–0.98 μm and half-width and half maxima of the spectral response function at 0.65 μm, Fig. 1);
2) use an electronic rolling shutter;
3) have two operation modes;
4) day and night imaging [31].

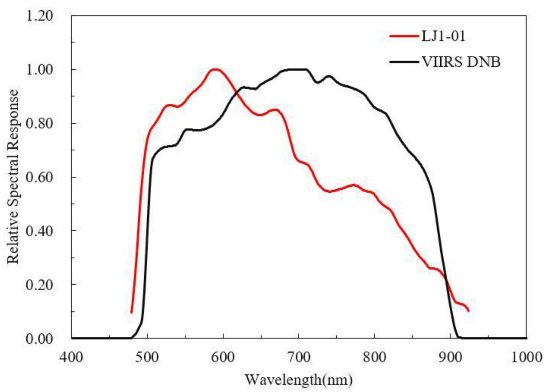


Fig. 1. Spectral response curves of LJ1-01 and VIIRS.

TABLE II
PARAMETER COMPARISONS OF THE LJ1-01 AND VIIRS DATA

Parameters	**LJ1-01**	**NPP-VIIRS**
Spatial resolution	130 m	750 m
Swath width	250 km	3060 km
Spectrum range	0.46–0.98 μm	0.5–0.9 μm
Radiometric resolution	14 bits	14 bits
Available period	2018–present	2012–present
Saturation	Not saturated	Not saturated
Revisit time	12h	15d
On-board calibration	Yes	Yes
Half width	0.65 μm	0.70 μm

The NTL imagery generated from the LJ1-01 satellite has not only enriched the available data for NTL remote sensing applications. These characteristics of the new generation NTL data also greatly improve the quality of NTL data, enhance detection capacity for artificial lighting and urban structure, bringing new possibilities and insights to the researchers working on urban environments. The parameters comparison for NPP-VIIRS and LJ1-01 data is shown in Table II.

This study developed GWR-based models to estimate $PM_{2.5}$ concentrations using MODIS AOD data, together with LJ1-01 NTL/NPP-VIIRS data, meteorological data, and geographic data. The objectives of this article were to investigate whether the NTL data can improve $PM_{2.5}$ concentration prediction accuracy and to validate the difference of LJ1-01/NPP-VIIRS in AOD-based $PM_{2.5}$ predict power. A cross-validation method was utilized to verify the models' performance. Temporal and spatial distributions of predicted surface $PM_{2.5}$ were then derived from the model to illustrate the variations of ground-level $PM_{2.5}$ concentrations within the study area.

II. METHODOLOGY

A detailed flowchart of the procedure with LJ1-01 in AOD-based $PM_{2.5}$ concentration prediction is shown in Fig. 2. The key steps of LJ1-01 NTL data for estimating monthly $PM_{2.5}$ concentration is divided into the following three parts.

1) The vegetation index is an important factor to estimate $PM_{2.5}$ concentration. Considering the similarity of the NDVI index and EVI index. We fit and validate the performance of these two vegetation indexes by the method of controlling variables, thus determine the appropriate vegetation index as the input of vegetation variables.
2) We construct five models, with the additional explanatory variables of only NTL (LJ1-01, NPP-VIIRS), only AOD, and NTL+AOD to predict $PM_{2.5}$ concentration at monitor sites. To test the effectiveness and reliability of the proposed method, a comparison with the contribution of LJ1-01/NPP-VIIRS in AOD-based $PM_{2.5}$ prediction is conducted, and the details are given in Section IV.
3) To further verify the potential and performance of LJ1-01 in AOD-based $PM_{2.5}$ concentration prediction, we compare the temporal and spatial distribution of predicted $PM_{2.5}$ concentrations at ground observation sites with the resolution of 750 × 750 m and 10 × 10 km. The spatial distribution and predict difference are elaborated in Section IV.

A. GWR Model

The GWR model is a spatial regression model, which changes the assumption that the objects are independent of each other in traditional econometric and statistical studies, and it gives full consideration to the spatial and temporal correlation of the data for each observation unit. The estimation results of the parameters change with the change in spatial position, which can be quantified to reflect the heterogeneity or nonstationarity relationship between the independent and dependent variables [10]. The GWR model assumes that the predictors–$PM_{2.5}$ relationship varies greatly with spatial locations in the study area, and it has been adapted to describe the unstable relations between $PM_{2.5}$ concentrations and AOD, as well as other geographical or meteorological factors [see (1)]. In this study, the constructed five models based on the GWR model were used to predict $PM_{2.5}$ concentrations. The five models and corresponding variables are shown in Table III

$$\begin{aligned}\mathrm{PM}_{2.5(i,j)} = {}& \beta_{0(i,j)} + \beta_{1(i,j)}\mathrm{Meteoro}_{(i,j)} \\ &+ \beta_{2(i,j)}\mathrm{Veg}_{(i,j)} + \beta_{3(i,j)}\mathrm{DEM}_{(i,j)} \\ &+ \beta_{4(i,j)}\mathrm{AOD}_{(i,j)} + \beta_{5(i,j)}\ln(\mathrm{NTL}_{(i,j)}) \quad (1)\end{aligned}$$

where $\mathrm{PM}_{2.5(i,j)}$ represents the $PM_{2.5}$ concentration of ground observation stations at position i on data j, $\beta_{0(i,j)}$ is a constant coefficient denoting the location-specific intercept at position i on data j, $\beta_{1(i,j)} - \beta_{3(i,j)}$ denotes the location-specific slope or coefficient of its corresponding auxiliary meteorological variables, vegetation index, and elevation variables, $\beta_{4(i,j)}$ denotes the slope of AOD, and $\beta_{5(i,j)}$ represents the slope of the logarithm of the NTL intensity.

B. Model Evaluation and Verification

Statistical indicators are commonly employed to evaluate a model's accuracy through comparing the fitted values with ground observed $PM_{2.5}$ concentration values, such as the R^2, root mean square error (RMSE) [24]. The R^2 denotes the agreement

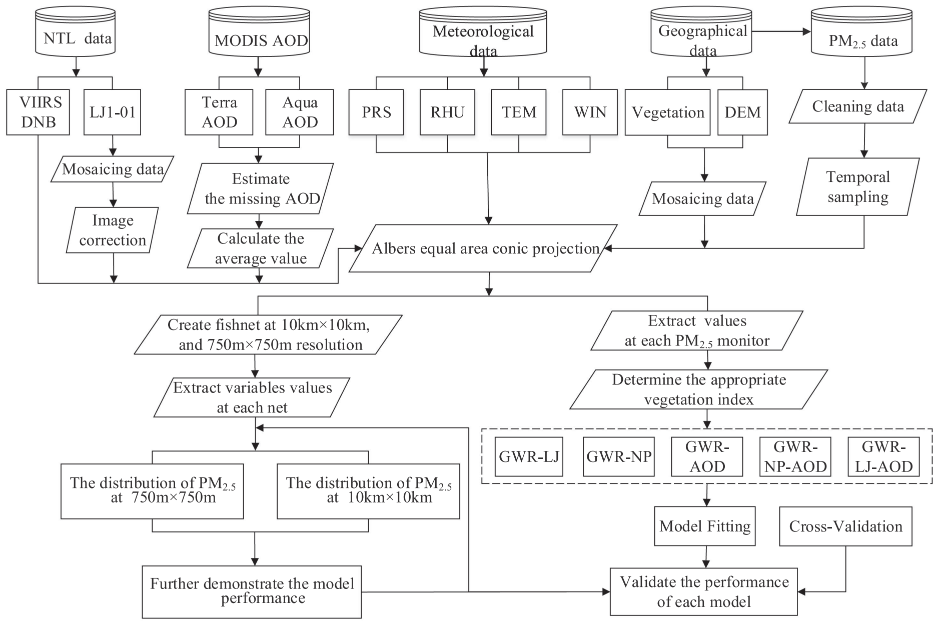


Fig. 2. Flowchart of the analysis of the performance of LJ1-01 NTL data in MODIS AOD-based monthly $PM_{2.5}$ concentration prediction.

TABLE III
DESCRIPTION OF THE FIVE SPECIFIC MODELS AND THE CORRESPONDING EXPLANATORY VARIABLES

Models	Explanatory variables
GWR_NP	Meteoro (RHU, TEM, WIN, PBLH), EVI, DEM, VIIRS DNB
GWR_LJ	Meteoro (RHU, TEM, WIN, PBLH), EVI, DEM, LJ1-01
GWR_AOD	Meteoro (RHU, TEM, WIN, PBLH), EVI, DEM, AOD
GWR_NP_AOD	Meteoro (RHU, TEM, WIN, PBLH), EVI, DEM, VIIRS DNB, AOD
GWR_LJ_AOD	Meteoro (RHU, TEM, WIN, PBLH), EVI, DEM, LJ1-01, AOD

degree of fitting between predicted and observed $PM_{2.5}$ concentrations. The RMSE indicates the prediction accuracy of the models. Smaller RMSE values are indicative of a higher accuracy at the $PM_{2.5}$ observation site. The formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\mathrm{PM}_{2.5(i)}^{\mathrm{obs}} - \mathrm{PM}_{2.5(i)}^{\mathrm{sat}}\right)^2}{\sum_{i=1}^{n}\left(\mathrm{PM}_{2.5(i)}^{\mathrm{obs}} - \overline{\mathrm{PM}_{2.5(i)}^{\mathrm{obs}}}\right)^2} \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{PM}_{2.5(i)}^{\mathrm{obs}} - \mathrm{PM}_{2.5(i)}^{\mathrm{sat}}\right)^2\right)} \tag{3}$$

where n denotes the number of ground observation stations, $\mathrm{PM}_{2.5(i)}^{\mathrm{obs}}$ and $\overline{\mathrm{PM}_{2.5(i)}^{\mathrm{obs}}}$ represent the $PM_{2.5}$ concentration and average $PM_{2.5}$ concentration of the ground observation stations, respectively, and $\mathrm{PM}_{2.5(i)}^{\mathrm{sat}}$ denotes the fitted $PM_{2.5}$ concentration of the models.

In addition, there may be an overfitting phenomenon in the process of model processing. This study used the method of 10-fold cross-validation to evaluate the performance of each model. The dataset was first broken into 10 folds with approximately 10% of the total data points in each fold. In each round of the cross-validation, the model was fitted with nine folds (90% of the total dataset) and one fold was predicted by using the fitted model. This step was repeated 10 times until every fold was tested. The R^2 and RMSE of the model fitted values and observed $PM_{2.5}$ concentrations values were utilized to evaluate the performance of the models.

III. STUDY AREA AND DATA

A. Ground Observation Data

The Beijing–Tianjin–Hebei (BTH) region was selected as the study area; this area is situated in northern China and extends over 113–120°E and 36–43°N (see Fig. 3). It includes Beijing, Tianjin, and eight neighboring cities in China's Hebei Province. There are 99 $PM_{2.5}$ monitoring stations. We selected the data from the $PM_{2.5}$ monitoring stations during the latest available period (366 days) from June 2018 to May 2019 for the analysis. Fig. 3 shows the $PM_{2.5}$ monitoring stations in the BTH region. Ground-level hourly $PM_{2.5}$ concentrations were download from China's air quality publishing platform.[1] Additional data for environmental evaluation points and comparison points in Beijing suburbs were obtained from the Beijing environmental monitoring center.[2] The $PM_{2.5}$ concentrations were measured by the Tapered Element Oscillating Microbalances method or beta attenuation method (β-gauge). These data had an uncertainty of less than 0.75%, with an accuracy reaching up to $\pm 1.5\ \mu$g/m^3 for

[1] [Online]. Available: http://106.37.208.233:20035/
[2] [Online]. Available: http://zx.bjmemc.com.cn/

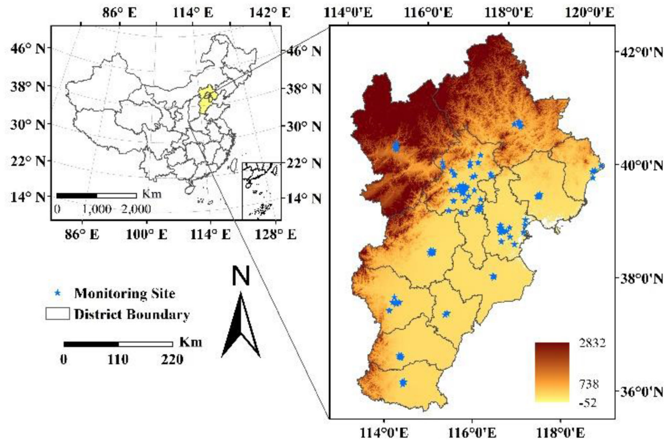


Fig. 3. Sketch map of the BTH region analyzed in this study.

the hourly average, so these data were accurate enough to serve as a ground truth for $PM_{2.5}$ concentrations [22]. In accordance with the study of Li *et al*. [23], $PM_{2.5}$ concentration values (<2 $\mu g/m^3$) less than the detection range were removed. The experiment of Fu *et al*. [12] and Wang *et al*. [33] showed that $PM_{2.5}$ concentration at nigh time (~1:00, local standard time) better represented daily-mean $PM_{2.5}$ concentration. Hence, the $PM_{2.5}$ concentration at night time 1:00 was utilized to represent the $PM_{2.5}$ concentration of 24-h mean.

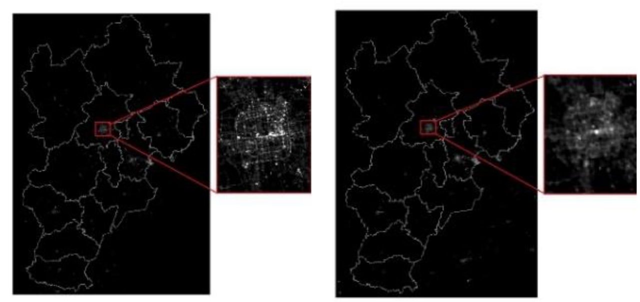
By referring to previous studies, we manually chose four factors as preliminarily auxiliary meteorological variables, and these factors included the temperature (TEM,°C), wind speed (WIN, m/s), relative humidity (RHU, %), and planetary boundary layer height (PBLH, m). The particle concentrations of $PM_{2.5}$ largely depend on meteorological conditions. A high surface temperature will accelerate the atmospheric vertical motion and the transportation of ground pollutants into higher places. Wind speed is an effective index for quantifying surface motions of airflow and affects the horizontal transport of ground pollutants. Relative humidity allows for a correction of the aerosol humidity in the atmosphere to achieve better matches with ground dry $PM_{2.5}$ concentrations. A high relative humidity largely enhances the size and light extinction of particles, which are comprised of sulfate, nitrate, and ammonium from coal and biomass burning, industrial, and vehicular sources. The PBLH is also taken as a meteorological variable because the height of the boundary layer determines the mixing space of pollutants, and the interactions between pollutants and the boundary layer bring about changes to the stability of the atmospheric boundary layer and the atmospheric allowable emissions [35]. These data were obtained from the ERA-Interim reanalysis data archive with a resolution of $0.125° \times 0.125°$ at the European Center for Medium-Range Weather Forecasts.[3]

B. NTL Imagery

In order to cover the entire BTH region, eight moonless and cloudless LJ1-01 images from August 20, 2018 to October 13, 2018 were obtained from the High-Resolution Earth Observation System of the Hubei Data and Application Center.[4] The detailed information of those images is shown in Table IV.

TABLE IV
DETAILED INFORMATION OF EIGHT LJ1-01 IMAGES COVERING THE BTH REGION

Orbit Number	Imaging Time (UTC)	Central Longitude/Latitude
6	2018-10-13T14:26:42	114.2828/39.1598
12	2018-10-13T14:27:12	113.6647/40.9635
24	2018-9-6 T14:22:40	116.5561/38.9776
28	2018-9-6 T14:23:0	116.1838/40.1951
38	2018-10-5 T14:30:30	114.8710/36.9424
42	2018-10-5 T14:30:50	114.4994/38.1526
44	2018-8-20 T14:7:52	118.4223/39.7870
57	2018-10-29T14:18:17	117.5947/41.5834



(a) (b)

Fig. 4. NTL images after radiometric correction and unit conversion in the BTH region. (a) LJ1-01. (b) NPP-VIIRS.

The LJ1-01 images were only processed by system geometry corrections and had a low positioning accuracy (ranging from 0.49 km to 0.93 km). Therefore, the released images should be geometrically corrected before analysis. Therefore, we carried out efficient geometric correction processing based on the research of Jiang *et al*. [35]. Because of the LJ1-01 imagery with the high image resolution, the road network was clear and identified as the ground control points (GCPs) (see Fig. 4). First, 30 evenly distributed GCPs were manually collected from road intersections. According to these GCPs, the geometric correction was then conducted. The monthly composites NPP-VIIRS day/night band data are filtered to exclude data impacted by stray light, lightning, lunar illumination, and cloud coverage [36]. The composite data encompasses a suite of average radiance composite imagery. It depicts persistent lights from towns, cities, and other sites spanning the globe from 75°N latitude to 65°S. The composite image in August, September, and October 2018 was derived from the website of EOG,[5] then the images average value of these three months is calculated. Due to the constraint of data, at a specific location, we assume that surface features do not change from June 2018 to May 2019. Consequently, the

[3][Online]. Available: https://apps.ecmwf.int/datasets/data/interim-full-daily/

[4][Online]. Available: http://59.175.109.173:8888/

[5][Online]. Available: https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html

intensity of NTL is treated as a constant for each location within VIIRS and LJ1-01 pixel but is varied spatially.

Images from both NPP-VIIRS and LJ1-01 were absolute radiation-corrected [36]. The radiance conversion formula of the LJ1-01 image was

$$\mathrm{L} = d^{3/2} * 10^{-10} \quad (4)$$

where *L* is the radiance value after absolute radiation correction in $\mathrm{Wm^{-2}sr^{-1}\mu m^{-1}}$ and *d* is the gray value of the image.

It is observed that the radiation unit of NPP-VIIRS is $\mathrm{nWcm^{-2}sr^{-1}}$, which is not consistent with that of LJ1-01. The reason for this fact is that the radiance of LJ1-01 is converted to the central wavelength, while that of NPP-VIIRS uses the full-band radiance. Therefore, the radiation unit of LJ1-01 is converted to that of VIIRS by (5), the NTL images after radiometric correction and unit conversion are shown in Fig. 4

$$L' = \mathrm{L} * 10^5 * w \quad (5)$$

where L' is the radiation value after unit conversion and *w* is the bandwidth of LJ1-01. The radiometric range of LJ1-01 is 460–980 nm, so w is equal to $5.2 * 10^{-7}$m.

C. MODIS AOD Data

The MODIS sensors on Terra and Aqua satellites have 36 visible to near-infrared spectral bands ranging from 0.4 to 14 μm, with a global revisit cycle of 1–2 days. Thus, MODIS provides comprehensive global information on Earth's atmospheric system. Furthermore, the characteristics of high temporal resolution, reasonable spatial resolution, and good accuracy of AOD products make it adaptable for characterizing earth surfaces [22]. In this study, the monthly Terra MODIS Level-2 10-km AOD product and Aqua MODIS Level-2 10-km AOD product were downloaded from the Atmosphere Archive & Distribution System of the National Aeronautics and Space Administration.[6] However, there are portions of the region at specific months where only one of the two retrievals is available. To overcome this shortcoming, we used a simple linear regression to estimate the missing AOD values according to the study of Zhang and Hu [32], in which the derived regression equations are as follows:

$$\tau_{\mathrm{AQUA}} = 0.8406 \times \tau_{\mathrm{TERRA}} + 0.0517 \quad (6)$$

$$\tau_{\mathrm{TERRA}} = 0.9137 \times \tau_{\mathrm{AQUA}} + 0.0621 \quad (7)$$

where τ is the AOD. The missing AOD values were estimated by using the above regression equations and then averaged. Consequently, each AOD grid cell contains a mean AOD value.

D. Geographic Data

A change in elevation will impact the spread of air pollutants due to the effect of gravity sedimentation, and a 90-m digital elevation model (DEM) of the SRTMDEM3 dataset was acquired from the Geospatial Data Cloud.[7] Studies have shown that $PM_{2.5}$ concentrations are also closely related to the vegetation index [37]. Specifically, high vegetation coverage reduces the entry of aerosols into the atmosphere and vegetation absorbs particles in the atmosphere. Currently, MODIS NDVI and MODIS EVI are the most widely applied vegetation index, and MODIS NDVI is often used to estimate $PM_{2.5}$ concentrations [21], [32]. However, Waring *et al.* [38] pointed out that MODIS EVI is a more recent development and continuation of NDVI and shows improvements from the NDVI formula and synthesis method. Zhuo *et al.* [39] suggested that NDVI is susceptible to large sources of error and uncertainty under variable atmospheric and canopy background conditions. Thus, this study compares the performance of these two vegetation indexes for the estimation of $PM_{2.5}$ concentrations, and the more appropriate vegetation index was selected as the auxiliary variable for $PM_{2.5}$ concentration estimations thereafter. The monthly synthesized NDVI and EVI products of MODIS (MOD13A3) were obtained from NASA,[8] and these data represent the vegetation coverage at spatial resolutions of 1000 m.

E. Data Integration

The data used in this study ($PM_{2.5}$ concentrations, MODIS AOD, meteorological data, DEM data, NTL data, and vegetation coverage data) have different temporal and spatial resolutions. To create a synchronized dataset for modeling, validation, and analysis, all collocated data were first integrated to the average value of a month to reduce the impact of time mismatches and noise errors. Then, the data were reprojected by using the Albers equal-area conic projection. After that, all data were resampled to a spatial resolution of 750×750 m and 10×10 km, the values of each pixel ware extracted from the clipped administrative boundaries in the BTH region. However, the coverage of a pixel is large at the resolution of 10×10 km, many sites may be located in one grid. Therefore, if a single geographical grid contain more than one $PM_{2.5}$ site, the average $PM_{2.5}$ concentration of these sites was taken as the true $PM_{2.5}$ concentration. Finally, the normalized MODIS AOD, NTL data, vegetation index, and other auxiliary variables were spatially matched to each grid for model fitting and validation.

IV. Results

A. Descriptive Statistics

Fig. 5 shows the histograms of all related variables expressed in frequency distributions. The descriptive statistics [e.g., mean, standard deviation (SD), maximum (Max), and minimum (Min)] for the variables are summarized in Table V. As shown in Fig. 5, except for the temperature variable, all other variables presented an approximate unimodal lognormal distribution pattern. The annual average $PM_{2.5}$ value from July 2018 to May 2019 was 56.81 μg/m^3, with an SD of 14.20 μg/m^3. MODIS AOD had an overall mean value of 0.49, with an SD of 0.27. NDVI had an overall mean value of 0.49, with an SD of 0.27. PBLH had a mean value of 171.62 m, with an SD of 86.15 m. The frequency distribution histograms of AOD, NDVI, and PBLH were similar

[6] [Online]. Available: http://ladsweb.nascom.nasa.gov/

[7] [Online]. Available: http://www.gscloud.cn/

[8] [Online]. Available: http://ladsweb.nascom.nasa.gov/

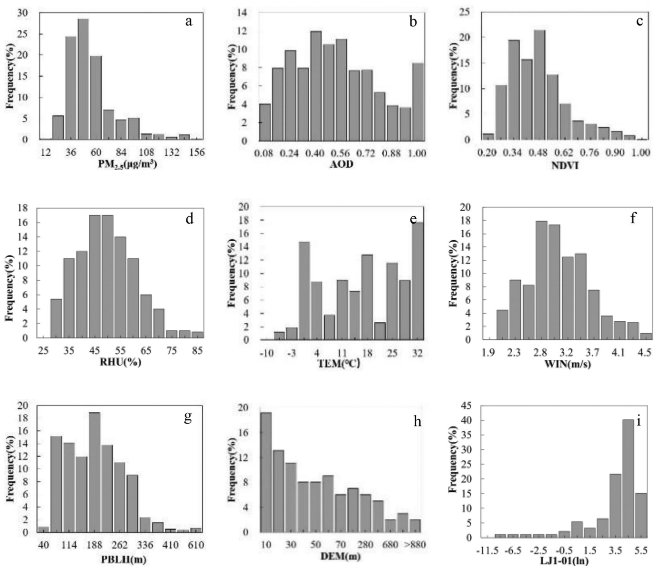


Fig. 5. Histograms of all related variables in this study. (a) $PM_{2.5}$. (b) MODIS AOD. (c) NDVI. (d) RHU. (e) TEM. (f) WIN. (g) PBLH. (h) DEM. (i) LJ1-01.

TABLE V
DESCRIPTIVE STATISTICS FOR THE ANNUALLY RELATED VARIABLES

Yearly (N = 1188)	Mean	SD	Max	Min
$PM_{2.5}$ (μg/m^3)	56.81	14.20	187.46	14.10
RHU (%)	47.19	4.06	83.78	27.48
TEM (°C)	12.39	2.11	30.13	-12.04
WIN (m/s)	2.93	0.33	3.98	2.40
DEM (m)	110.57	200.92	983.00	0.00
PBLH (m)	171.62	86.15	608.47	36.44
AOD	0.49	0.27	1.00	0.00
NDVI	0.44	0.15	1.00	0.00
EVI	0.39	0.17	1.00	0.00
LJ1-01(ln)	2.79	2.61	5.19	-11.00
DNB(ln)	2.84	1.18	4.73	-1.31

to that of $PM_{2.5}$ concentrations. TEM (12.39±2.11 °C) showed seasonal aggregation trends in winter, summer, autumn, and spring. LJ1-01 (2.79±2.61) and NPP-VIIRS (2.84±1.18) NTL brightness values showed irregular distributions, where 90% of the pixels were distributed in the low and medium brightness areas and a small portion of the pixel brightness values fluctuated in the high-value range.

The seasonal changes for these variables were also analyzed. The mean $PM_{2.5}$ concentrations during the winter (82.91±34.71 μg/m^3) were higher than those during the summer (40.05± 9.71 μg/m^3). The mean $PM_{2.5}$ concentrations in spring (47.58±12.56 μg/m^3) and autumn (56.72±26.81 μg/m^3) were close to each other and slightly higher than those in summer. The mean AOD values were much higher in the spring (0.52±0.22) and summer (0.59±0.31) than those in the fall (0.38±0.23) and winter (0.46±0.27); the same trend also existed in the NDVI and EVI data. As for WIN, the highest mean value was observed in the spring (4.84±0.40 m/s), while the lowest mean value was observed in the summer (1.94±0.41 μg/m^3). Furthermore, on account of the summer's intense solar radiation and high frequency of heavy rainfall, the highest mean values of RHU (83.78.57±12.98%), TEM (30.13±1.76°C), and PBLH (342.99±51.63 m) all occurred in summer.

TABLE VI
PEARSON'S CORRELATION COEFFICIENTS FOR THE $PM_{2.5}$ CONCENTRATIONS AND PREDICTOR VARIABLES

Predictors	Pearson	p value
TEM	0.592**	0.015
DEM	-0.293*	0.039
AOD	0.762**	0.000
NDVI	-0.283**	0.046
EVI	-0.329*	0.020
WIN	-0.545**	0.007
RHU	0.536**	0.000
PBLH	-0.831**	0.000
LJ1-01	0.197	0.168
DNB	0.165	0.241

$* p < 0.05$; $** p < 0.01$.

TABLE VII
FOUR MODELS AND CORRESPONDING EXPLANATORY VARIABLES

Model	Explanatory variables
GWR_LJ_EVI	EVI, DEM, AOD, LJ1-01
GWR_LJ_NDVI	NDVI, DEM, AOD, LJ1-01
GWR_NP_EVI	EVI, DEM, AOD, VIIRS DNB
GWR_NP_NDVI	NDVI, DEM, AOD, VIIRS DNB

B. Determine the Appropriate Vegetation Index

Correlation analysis was used to study the correlations between $PM_{2.5}$ concentrations and all predictors, where the influence of all predictors on $PM_{2.5}$ concentrations was expressed by Pearson correlation coefficients; the strength of the correlations was assessed with *p*-values from statistical tests. The results presented in Table VI show that TEM, AOD, and RHU had significant positive correlations with $PM_{2.5}$ concentrations, while DEM, NDVI, EVI, WIN, and PBLH had significant negative correlations with $PM_{2.5}$ concentrations. Most variables' effects on $PM_{2.5}$ concentrations were significant at the 0.01 level. However, the LJ1-01 and VIIRS DNB NTL data appeared to show less of an influence. Because NTL is closely related to the intensity of human activity and population density [40], it characterized regional aggregations and discrete distributions of different regions, in which the results for the NTL are often sparser than those for other variables. The sensitivity of $PM_{2.5}$ to changes in NTL was low [21].

In this part, we used the method of control variables to construct four models, which were GWR-LJ-EVI, GWR-LJ-NDVI, GWR-NP-EVI, and GWR-NP-NDVI model, respectively. The detailed information of the model and explanatory variables are shown in Table VII. To avoid the potential problem of strong multicollinearity among the predictor variables, variance

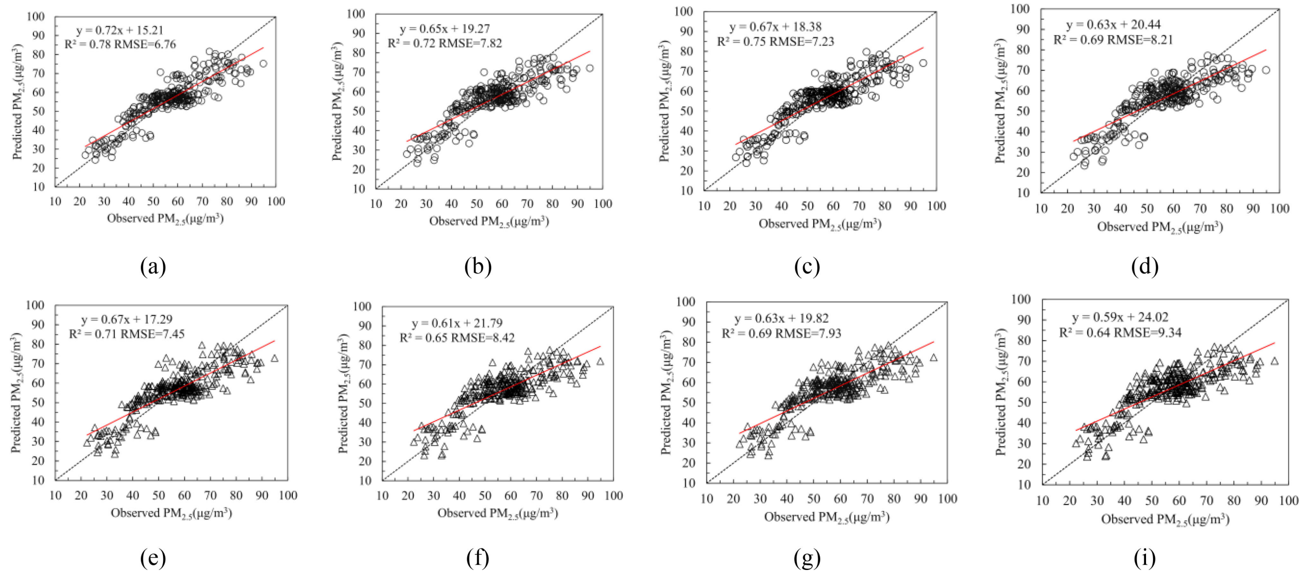


Fig. 6. Fitted curves of observed monthly $PM_{2.5}$ concentrations, with the fitted monthly $PM_{2.5}$ concentrations and cross-verified monthly $PM_{2.5}$ concentrations (the number of points is 297), where x-axis represents the ground observed $PM_{2.5}$ concentrations, y-axis represents the fitted $PM_{2.5}$ concentration obtained from constructed models. (a)–(d) Fitting results for the GWR-LJ-EVI, GWR-LJ-NDVI, GWR-NP-EVI, and GWR-NP-NDVI models, respectively. (e)–(i) Results for the model cross-validation research. The solid and dashed lines in the figure are fitted regression curves and 1:1 reference curves, respectively.

inflation factor (VIF) was examined. The result showed that the VIF between the predictors in the GWR model was all less than 7.5. We determined the optimal vegetation index through the performance of model fitting [see Fig. 6(a)–(d)] and cross-validation [see Fig. 6(e)–(i)]. As can be seen from the figures, compared with the fitting results for the GWR-LJ-NDVI model (R^2 of 0.72), the GWR-LJ-EVI model (R^2 of 0.78) showed a significant improvement, and the GWR-NP-EVI model (R^2 of 0.75) had a better performance than the GWR-NP-NDVI model (R^2 of 0.69).

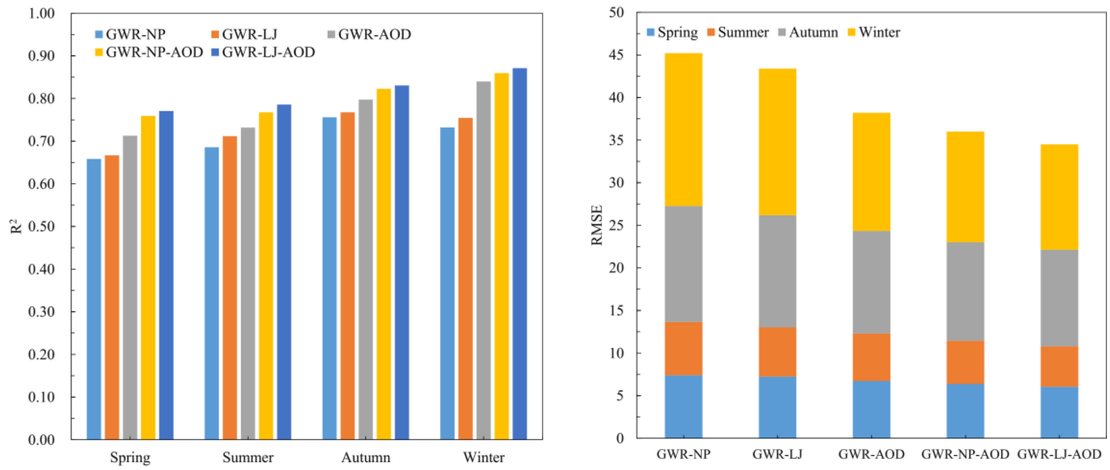
The 10-fold cross-validation results were shown in Fig. 6(e)–(i), the cross-validation R^2 and RMSE of the GWR-LJ-EVI model were 0.71 and 7.45 μg/m^3 [see Fig. 6(e)], and the GWR-LJ-NDVI model were 0.65 and 8.42 μg/m^3 [see Fig. 6(f)], respectively. Compared with the GWR-LJ-NDVI model, the cross-validation R^2 of the GWR-LJ-EVI model improved by 6%, while the RMSE decreased by 0.97 μg/m^3. Moreover, the fitting slope of the GWR-LJ-EVI model is 6% higher than that of the GWR-LJ-NDVI model in the cross-validation. As can be seen from Fig. 6(g) and (i), the cross-validation R^2 of the GWR-NP-EVI model had improved by 5%, and the RMSE decreased by about 1.41 μg/m^3; additionally, the slope increased by 4%, compared with the GWR-NP-NDVI model. The results indicated that the predicted $PM_{2.5}$ concentrations with EVI as the predictive variable were closer to the actual observed $PM_{2.5}$ concentrations than that with NDVI as the predictive variable. In other words, although there wase slight overfitting of the models, the EVI index showed a better performance than the NDVI index in estimating $PM_{2.5}$ concentrations.

Some studies have indicated [38], [39], [41] that NDVI has shortcomings in terms of atmospheric corrections and soil background treatments in low-vegetation cover areas. When the vegetation coverage ratio is less than 15% or more than 80%, the sensitivity of NDVI to vegetation coverage decreases. The EVI is an extension and improvement of the NDVI, which not only takes into account the influence of the soil background, but also makes further corrections to the atmospheric and saturation problems, and thus, it has a higher sensitivity and superiority in estimating the vegetation coverage [41]. The experiments in this study confirmed the validity and effectiveness of the EVI in AOD-based $PM_{2.5}$ prediction.

C. Validation of the Contributions of LJ1-01/NPP-VIIRS in AOD-Based $PM_{2.5}$ Prediction

Fig. 6 compares the ability of the different vegetation index to estimate the annual $PM_{2.5}$ concentrations, and the results clearly indicated that the performance of estimating $PM_{2.5}$ concentration with EVI as an auxiliary variable was superior to that of NDVI. However, the scatter points and fitted curves corresponding to the model fitted $PM_{2.5}$ concentrations and the observed $PM_{2.5}$ concentrations were similar, which resulted in difficulties in observing detailed seasonal changes. Therefore, this section validates the seasonal performance of LJ1-01/NPP-VIIRS DNB in AOD-based $PM_{2.5}$ prediction with the vegetation index of EVI.

Fig. 7 shows the R^2 histogram of and RMSE cumulative histogram of observed monthly $PM_{2.5}$ concentrations, with the fitted monthly $PM_{2.5}$ concentrations of GWR-NP model, GWR-LJ model, GWR-AOD model, GWR-NP-AOD model, and GWR-LJ-AOD model. As shown in the figure, the prediction abilities of the five models were stable. According to the determination coefficient R^2 and the root-mean-squared prediction error RMSE, GWR-NP, and GWR-LJ performed the worst (R^2 of 0.66 and 0.67 in spring, 0.69 and 0.71 in summer, 0.76 and 0.77 in autumn, 0.73 and 0.75 in winter). The GWR-AOD model



(a) (b)

Fig. 7. R^2 (a) and RMSE (b) of observed monthly $PM_{2.5}$ concentrations, with the fitted monthly $PM_{2.5}$ concentrations of GWR-NP model, GWR-LJ model, GWR-AOD model, GWR-NP-AOD model, and GWR-LJ-AOD model in four seasons.

with the AOD as predict variable has modest improvement (R^2 of 0.71 in spring, 0.73 in summer, 0.80 in autumn, and 0.84 in winter) compared with the NTL as an auxiliary in GWR model.

By adding both the NTL and AOD to the GWR model, the prediction performance of the GWR-LJ-AOD (R^2 of 0.77 in spring, 0.79 in summer, 0.83 in autumn, 0.87 in winter) and GWR-NP-AOD model (R^2 of 0.76 in spring, 0.77 in summer, 0.82 in autumn, 0.86 in winter) have been greatly improved. The GWR-LJ-AOD model was improved by 5.80%, 5.39%, 3.33%, and 3.13%, and the GWR-NP-AOD model was improved by 4.63%, 3.60%, 2.57%, and 1.99% compared with GWR-AOD model in four seasons.

The fitting effect GWR-NTL-AOD model (the average R^2 of the GWR-NP-AOD and GWR-LJ-AOD model) was better than that of the GWR-NTL model (the average R^2 of the GWR-NP and GWR-LJ model), with R^2 increases by 10.22%, 7.79%, 6.50%, and 12.21% in four seasons. Furthermore, the superiority of the GWR-NTL-AOD model over GWR-AOD model and GWR-NTL model can also be proven by the RMSE in Fig. 7(b). Compared with the GWR-NTL model, the RMSE of the GWR-NTL-AOD model decreased by 1.10, 1.11, 1.94, and 4.89 $\mu g/m^3$, respectively. The RMSE of the GWR-NTL-AOD model is reduced by 0.51, 0.66, 0.59, and 1.19 $\mu g/m^3$ compared to the GWR-AOD model.

The fitting R^2 of the GWR-LJ-AOD model in Fig. 8 is slightly higher for the GWR-NP-AOD model. The superiority of the GWR-LJ-AOD model to the GWR-NP-AOD model in estimating $PM_{2.5}$ concentrations is also confirmed by those RMSEs and associated slopes in Fig. 8. The seasonal slope of GWR-LJ-AOD has increased by 2%, 1%, 1.6%, and 1.3% with comparison to the GWR-NP-AOD model. For low $PM_{2.5}$ concentrations (less than 50 $\mu g/m^3$), all models overestimated the $PM_{2.5}$ concentrations. When the $PM_{2.5}$ concentrations were greater than 60 $\mu g/m^3$, underestimates existed in all models, and the underestimates became more serious as the observed $PM_{2.5}$ concentration increased. This may be because the model parameters were determined using mainly $PM_{2.5}$ concentrations smaller than 120 $\mu g/m^3$, and high $PM_{2.5}$ values got less weight [21].

Table VIII shows the results for the seasonal cross-validation of the GWR-NP model, GWR-LJ model, GWR-AOD model, GWR-NP-AOD model, and GWR-LJ-AOD model. Slight overfitting occurred in all models because the R^2 values of model fitting and cross-validation were all less than 1, and the R^2 values of cross-validation were smaller than those for the model fitting. The performances of these models for each season were similar. The GWR-LJ-AOD model and GWR-NP-AOD model performed best (the R^2 of 0.72 and 0.71 in spring, 0.73 and 0.72 in summer, 0.79 and 0.79 in autumn, 0.85 and 0.83 in winter), while the GWR-AOD came third (the R^2 of 0.71, 0.71, 0.77, and 0.80 in four seasons). The GWR-NP and GWR-LJ had the lowest corrections (the R^2 of 0.63 and 0.65 in spring, 0.67 and 0.69 in summer, 0.72 and 0.73 in autumn, 0.73 and 0.74 in winter). The better prediction performance of GWR-NTL-AOD indicates that the NTL data are helpful for more accurate predicting the $PM_{2.5}$ concentration.

Furthermore, the R^2 of the GWR-LJ-AOD model is higher than that of GWR-NP-AOD, with the 1.16%, 1.79%, 0.76%, 1.15% improvement in model fitting, and with the 1.04%, 0.85%, 0.78%, 1.37% improvement in cross-validation. The GWR-LJ model is superior to GWR-NP, with the R^2 improved by 0.87%, 2.60%, 1.14%, and 2.22% in model fitting and 1.75%, 2.52%, 0.58%, and 0.96% in cross-validation. The performance of the models can also be verified by the change in RMSE. In general, the GWR model with LJ1-01 performs better than that with VIIRS DNB. The better estimation ability of LJ1-01 in AOD-based $PM_{2.5}$ prediction indicates that the new generation LJ1-01 NTL data in the GWR model is conducive to a more accurate $PM_{2.5}$ concentration prediction.

The performance of the GWR-LJ-AOD model and GWR-NP-AOD model in autumn and winter were much better than that in the spring and summer. According to the results of

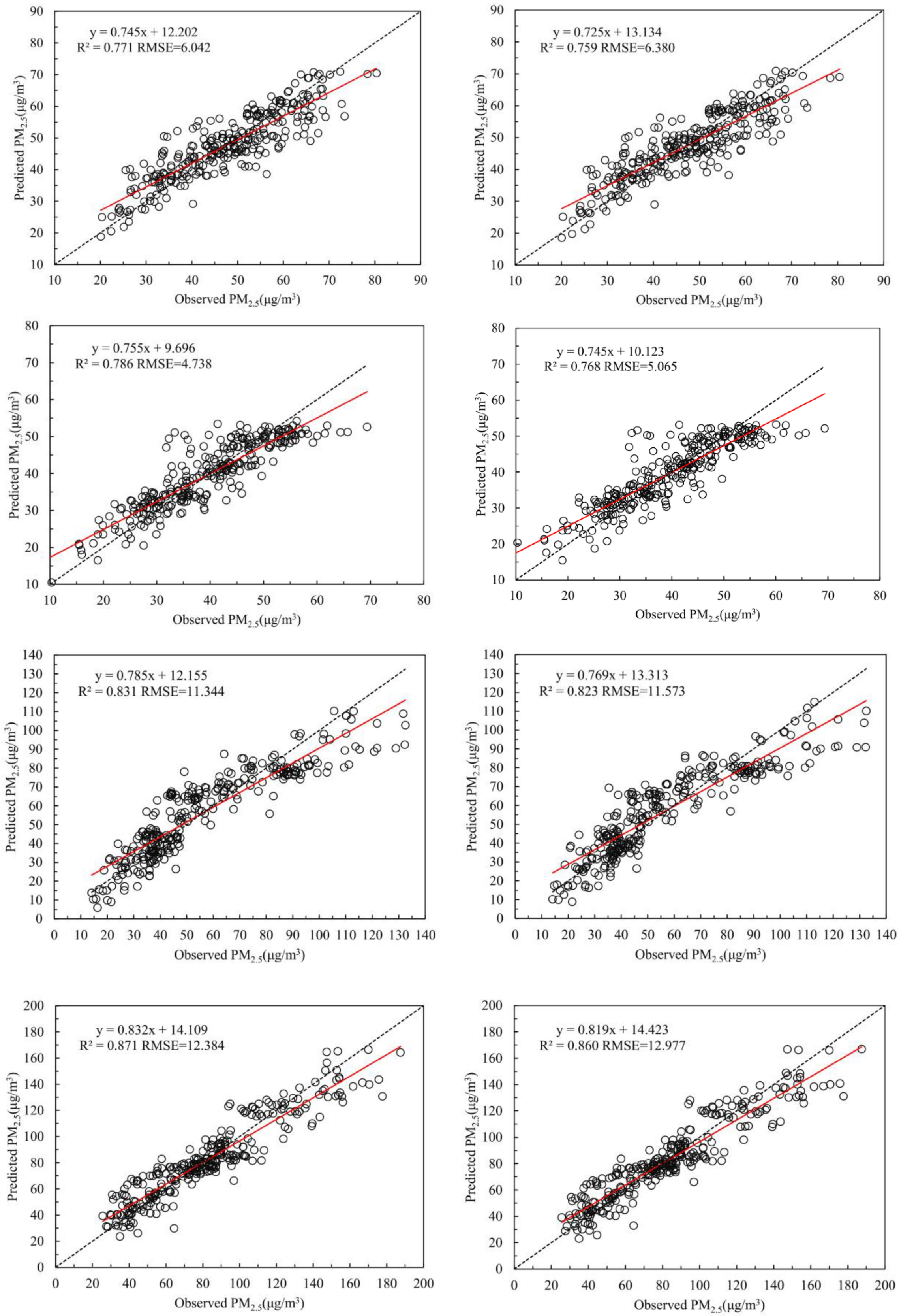


Fig. 8. Comparison of prediction performances for seasonal $PM_{2.5}$ concentrations derived with the (a) GWR-LJ-AOD (the left column) and (b) GWR-NP-AOD model (the right column). From top to bottom, the figures represent spring, summer, autumn, and winter, respectively. The number of points is 297. The *x*-axis represents the ground observed $PM_{2.5}$ concentrations, *y*-axis represents the fitted $PM_{2.5}$ concentration obtained from constructed models. The solid and dashed lines in the figure are fitted regression curves and 1:1 reference curves, respectively.

TABLE VIII
TENFOLD CROSS-VALIDATION RESULTS FOR THE CONSTRUCTED FIVE MODELS

Season	GWR-NP		GWR-LJ		GWR-AOD		GWR-NP-AOD		GWR-LJ-AOD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Spring	0.63	7.79	0.65	7.61	0.71	6.75	0.71	6.67	0.72	6.33
Summer	0.67	5.60	0.69	5.31	0.71	5.14	0.72	5.06	0.73	4.74
Autumn	0.72	14.69	0.73	13.90	0.77	12.89	0.79	12.65	0.79	12.12
Winter	0.73	22.03	0.74	18.57	0.80	15.34	0.83	14.51	0.85	13.51

model fitting and cross-validation, this is probably because of dense vegetation and low wind speed in autumn, leading to the accumulation of aerosol particles in the BTH region, thus enabling the model to best predict $PM_{2.5}$ at this time. Winter is the urban heating period, the main contribution to elevated $PM_{2.5}$ concentrations is from coal and biomass combustion for heating. This information might be not captured by MODIS AOD, whereas NTL as an indicator of human activity can reflect the effect of heating to some extent. Thus, NTL data might introduce into the model's additional $PM_{2.5}$ emission sources missed by MODIS AOD, which could lead to predictions of $PM_{2.5}$ that are more accurate [32]. In spring, the drier climate, greater wind speeds, and lower amounts of rainy weather were conducive to the formation of dust conditions, which resulted in the diffusion of aerosol particles and decreases in $PM_{2.5}$ concentrations. In summer, the increase in temperature, decrease in atmospheric stability, and high frequency of heavy rainfall were conducive to the diffusion, wet settlement, and dilution of atmospheric pollutants [23]. In those conditions, $PM_{2.5}$ concentrations in spring and summer were lower than those in other seasons, and the predictive performance for $PM_{2.5}$ concentrations was reduced.

D. Temporal and Spatial Distribution of Predicted $PM_{2.5}$ Concentrations

Fig. 9 showed the seasonal mean spatial pattern of the estimated $PM_{2.5}$ concentrations from June 2008 to May 2019, the $PM_{2.5}$ concentrations were calculated by using the GWR-LJ-AOD and GWR-NP-AOD model with a 10 $\times$ 10 km and 750 $\times$ 750 m resolution around each of the $PM_{2.5}$ sites in the BTH region. To examine the spatial and seasonal prediction accuracy of the model, the ground $PM_{2.5}$ measurements were also shown in Fig. 9.

Clearly, the temporal and spatial patterns of $PM_{2.5}$ concentrations were very similar, with low values in the northwest and high values in the southeast, i.e., a significant north-to-south increasing gradient. These continuous $PM_{2.5}$ concentration surfaces with 750 $\times$ 750 m depict more detailed information and structural change than the resolution of 10 $\times$ 10 km and the stationary monitoring sites. However, the $PM_{2.5}$ concentrations obtained by the GWR-NP-AOD model were lower than those attained by the GWR-LJ-AOD model, especially when $PM_{2.5}$ concentrations were higher.

The reason may be that the GWR-LJ-AOD model, which integrates LJ1-01 data with more accurate artificial lighting details, can more clearly reflect human activities and capture the status of pollution sources, which improves model performance and more accurately reflects the spatial distribution of $PM_{2.5}$ concentration.

Spatially, the $PM_{2.5}$ regional pollution differences in the BTH region were greatly affected by topographical dynamics and population [20], and the boundary of the polluted region was basically consistent with the direction of the Yan Mountains in the north and the Taihang Mountains in the west. The Taihang Mountains and Yan Mountains in the northwest region have a certain blocking and weakening effect on cold air activity, causing the stagnation of airflow, and the concentration of pollutants and water vapor. Pollutants along the leeward slope of the Taihang Mountains cannot diffuse easily. Furthermore, the suburbs and mountainous areas west and north of the Taihang Mountains have a low population density, low levels of social and economic activities, and few pollution sources, which resulted in lower $PM_{2.5}$ concentrations in this area. The highest population density, traffic density, industrial production, and combustion sources in the eastern plain region contributed to the high $PM_{2.5}$ concentrations in this region. This region includes the highly urbanized and polluted Haidian, Daxing, Shunyi, and Tongzhou districts in Beijing, where the $PM_{2.5}$ concentrations were 73.69, 74.86, 74.08, and 76.25 $\mu g/m^3$. The $PM_{2.5}$ concentrations were all higher than 80 $\mu g/m^3$ in central and southern Hebei. For example, the $PM_{2.5}$ concentration of Shijiazhuang was 84.98 $\mu g/m^3$, Hengshui was 82.51 $\mu g/m^3$, Xingtai was 84.97 $\mu g/m^3$, and Handan was 86.10 $\mu g/m^3$.

Fig. 9 shows the strong seasonal variations in the spatial distributions of $PM_{2.5}$ concentrations. Winter was the most polluted season, in which the average predicted $PM_{2.5}$ concentration was 71.34 $\mu g/m^3$. The average $PM_{2.5}$ concentrations in many areas were higher than 60 $\mu g/m^3$ at this time. In the winter, temperature inversions occur frequently and the dispersion of pollutants is limited in the shallow mixing layer. Furthermore, the smoke and dust released by centralized coal-fired power plants during the heating period contribute to the accumulation of fine particles, which leads to high $PM_{2.5}$ concentration on the ground. The summer with mostly flourishing vegetation, mainly sunny weather, a higher boundary layer, and enhanced mixing in the developing atmospheric boundary was the cleanest season, in which the average predicted $PM_{2.5}$ concentration was approximately 36.48 $\mu g/m^3$. The average predicted concentrations for spring and autumn were between those for summer and winter. In spring, the climate in the BTH area is dry, and local dust emissions and subsequent transportation have a large influence

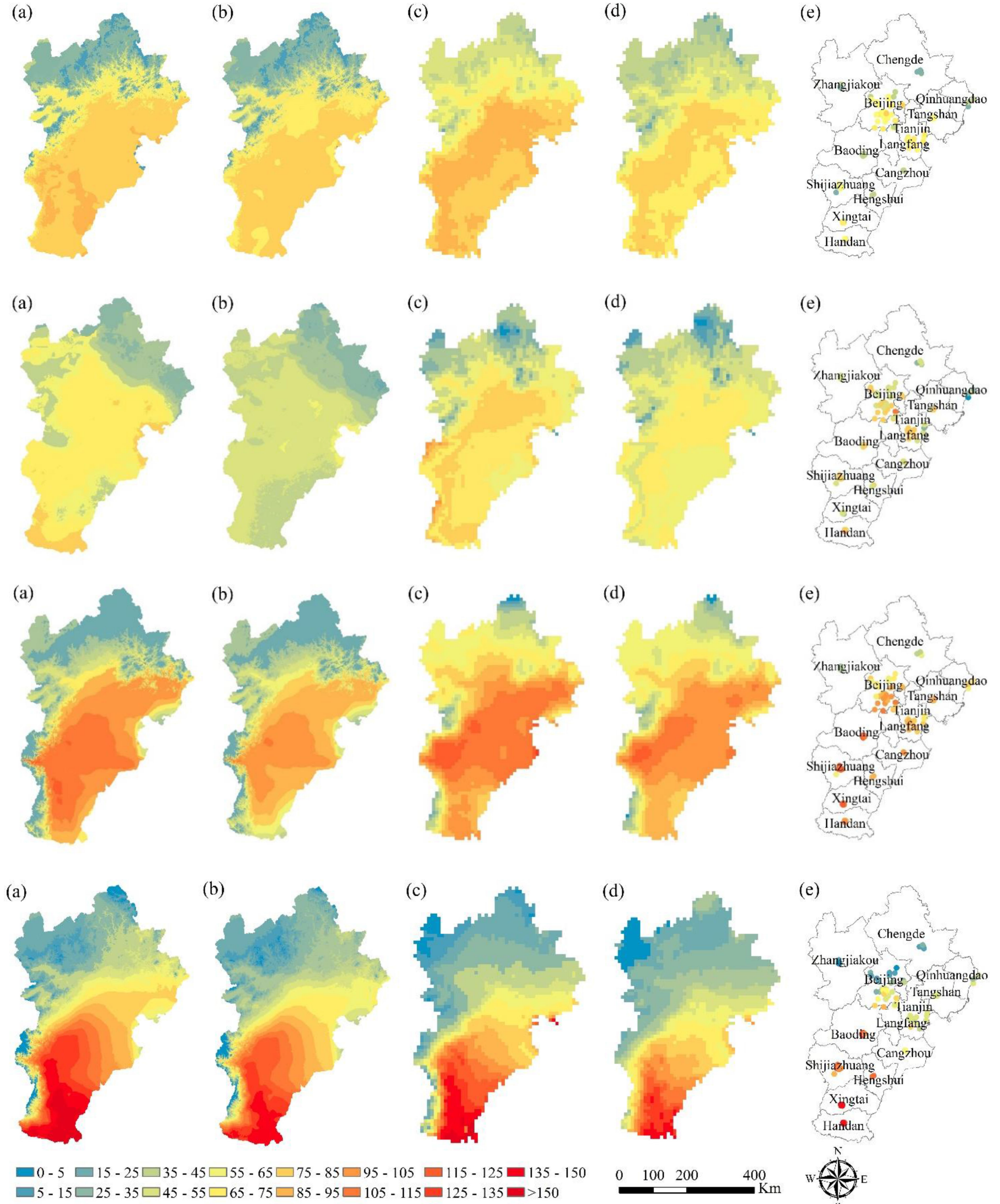


Fig. 9. Distribution of seasonal PM2.5 concentrations from the (a) GWR-LJ-AOD model at a 750 × 750 m resolution. (b) GWR-NP-AOD model at a resolution of 750 × 750 m. (c) GWR-LJ-AOD model at a 10 × 10 km resolution. (d) GWR-NP-AOD model at a resolution of 10 × 10 km. (e) Ground stationary monitoring sites. From top to bottom, the figure represents the results from spring, summer, autumn, and winter, respectively.

on urban pollution. Its average predicted $PM_{2.5}$ concentration was 37.86 μg/m^3, which was slightly less than the average predicted $PM_{2.5}$ concentration in autumn (44.57 μg/m^3). The predicted annual $PM_{2.5}$ concentration was 48.82 μg/m^3, a value far higher than the 35 μg/m^3 standard for $PM_{2.5}$ in the Chinese National Ambient Air Quality Standards (gb3095-2012) [42].

V. Discussion

Because of the sparse distribution of stationary air quality monitoring sites, satellite data with a wide spatial coverage are becoming an important supplementary tool for estimating $PM_{2.5}$ concentrations in various regions including urban areas and areas impacted by human activities. On the basis of the influence of several factors including meteorological and geographic factors on $PM_{2.5}$ concentrations, this study selected the most suitable vegetation index to predict $PM_{2.5}$ concentrations in the BTH region, where $PM_{2.5}$ pollution is problematic. The selected data were used to explore the ability of the new generation LJ1-01 NTL satellite product to help estimate $PM_{2.5}$ concentrations in comparison to traditional NPP-VIIRS NTL data. The results showed that the EVI, as an auxiliary variable in the GWR model, was better to evaluate of $PM_{2.5}$ concentrations than the NDVI, and subsequently, the EVI rather than the NDVI was used as the vegetation variable input into the GWR model in further analyses. In addition, the results demonstrated that LJ1-01 NTL data had better predictive power than NPP-VIIRS NTL data in AOD-based $PM_{2.5}$ concentration prediction. This advantage was partly due to the wide DN radiometric range and the spatial detail reduction ability satellite. However, there are still some uncertainties associated with the implementation process.

We used the monthly average values of Aqua MODIS and Terra MODIS to represent the true aerosol status and the mean AOD distribution of each grid cell. However, there were some areas during one specific month when only one of the two retrievals was available, thus resulting in a significant amount of missing data for the average of MODIS AOD. While the linear regression model, to a large extent, eased the absence of the AOD values, uncertainty still existed because of the appearance of clouds or high-reflectivity surface covers such as snow and ice in some regions, so it was difficult to accurately fit the changes in $PM_{2.5}$ concentrations with AOD. In this study, there were AOD anomalies and AOD deletions in Chengde and Zhangjiakou almost every month, so we used the surrounding pixel interpolation method to obtain the AOD values for missing points in evaluations of $PM_{2.5}$ concentrations. The determination of AOD outliers and data interpolation brought about great uncertainty during the model fitting. Furthermore, although the MODIS AOD products showed good performance in evaluating the $PM_{2.5}$ concentrations, the utilization of the AOD products was associated with some limitations. For example, the satellite can only capture the change values at one certain time (satellite overpass time), so the established model was specific to a specific time. The diurnal variability of $PM_{2.5}$ and AOD can be significant [20], so the prediction model for $PM_{2.5}$ concentrations in other time periods needs to consider more possible predictors to make up for the impact caused by the limited timeliness of AOD data.

In this study, LJ1-01 data, as an important information source for assessing the NTL intensity, were used to study the potential for estimating $PM_{2.5}$ concentrations. The experimental results showed that LJ1-01 NTL data had a better ability to help predict $PM_{2.5}$ concentrations than NPP-VIIRS NTL data. The high accuracy of the $PM_{2.5}$ concentration assessment benefited from the spatial characteristics of LJ1-01 NTL data. On the one hand, LJ1-01 data have a high spatial resolution and a broad spectral coverage, thereby LJ1-01 data can provide more variability and capture finer spatial details in anthropogenic lighting. The improvements achieved with the LJ1-01 data makes the spatial distribution pattern of human activities more consistent with the reference data. Thus, the LJ1-01 NTL data can precisely map the radiation signals to help evaluate $PM_{2.5}$ concentrations.

However, it is also well known that there are still some issues for the utilization of LJ1-01 data. First, the high resolution of LJ1-01 data limited the coverage of a single imaging, its swath width is only 250 km and the revisit time is 15 days. Furthermore, most of the data cover the region from tropical to subtropical regions, with few goes to the high latitude. These limitations pose a challenge for temporal sampling, application scope, and large-scale observations at the same time. Additionally, LJ1-01 imagery is affected by moonlight, which will bring about great uncertainty in remote sensing applications. Wang *et al.* [43] have developed a nighttime shortwave radiative transfer model capability in the UNifed and Linearized Radiative Transfer Model (UNL-VRTM). In the UNL-VRTM, a spectrally resolved moon-phase-based lunar irradiance model was added to calculate lunar irradiance at the top of the atmosphere. After learning the downward lunar radiance at the TOA, the radiative transfer in the atmosphere and the interaction of the moonlight with earth surfaces are treated in the similar manner as for the shortwave radiative transfer at daytime in UNL-VRTM. Thus, moonlight can be approximated as a collimated and directional source when it reaches TOA and can be treated in the similar way as is done for the solar due to the similarity of the polarization characteristics and the comparable solid angle of the moon to that of the Sun. However, the incorporating model still has several limitations. First, when the moonlight is treated as a collimated beam of electromagnetic plane waves as for direct-sun shortwave radiation, the difference of the moonlight polarization properties to that of the sun is ignored. Second, the uncertainty of the model is in the range of 5%-10% on account of the inability to account for temporal variations of the solar spectrum and libration effects on the lunar surface albedo. Future research will aim to apply the model to remove the effects of moonlight in LJ1-01 NTL images, and further treat the limitations in the model, providing more accurate knowledge.

The spatial and temporal mismatches of meteorological data, AOD, $PM_{2.5}$ concentration data, and geographical data will also bring about uncertainty for evaluation the $PM_{2.5}$ concentrations. For example, MODIS AOD represents a 10-km resolution, the DEM data resolution is 90 m, and the meteorological data with a resolution of $0.125° \times 0.125°$. Resampling of images with different resolutions, matching of point value data, and continuous remote sensing spatial coverage data will affect the results. In addition, some auxiliary variables in past studies were not added to the model as independent variables because of data and resource limitations, such as wind vector data in the four directions, traffic data, and demographic and economic statistics data. If the above factors are taken into account or more appropriate predictors are added to the model, the stability and accuracy of the model for predicting spatiotemporal changes will be improved. In future research, more data will be collected to help predict $PM_{2.5}$ concentrations and improve predict performance.

VI. Conclusion

This study explored the potential for the use of LJ1-01 NTL data in AOD-based $PM_{2.5}$ concentration predictions. In this study, five GWR-based models were established to illustrate the nonstatic spatial variation relationships of predictors such as the EVI and LJ1-01 NTL data with the $PM_{2.5}$ concentrations. Ground observation data for $PM_{2.5}$ concentrations were used to verify the accuracy of the model predictions. The established models solved the problem in which low spatial resolution NTL remote sensing data cannot accurately evaluate $PM_{2.5}$ concentrations. The conclusions are as follows.

1) The annual prediction R^2 with EVI variable all increased by 6%, the slope had improved by 6% and 5% compared with the variable of NDVI.
2) The seasonal R^2 with NTL in AOD-$PM_{2.5}$ model have improved by 5.07%, 4.50%, 2.95%, 2.56% in model fitting and 1.20%, 1.75%, 2.20%, 4.41% in cross-validation.
3) Compared with the NPP-VIIRS in AOD-$PM_{2.5}$ model, the seasonal R^2 with LJ1-01 in AOD-$PM_{2.5}$ model increased by 1.16%, 1.79%, 0.76%, 1.15% in model fitting and 1.04%, 0.85%, 0.78%, 1.37% in cross-validation.

The findings of this study prove that EVI is a more appropriate vegetation variable in $PM_{2.5}$ concentration in the GWR model. LJ1-01 data have better performance and potential for evaluating $PM_{2.5}$ concentrations, and accurately reflect the spatial distribution of $PM_{2.5}$ concentration. In the future, more predictors will be integrated into the GWR model to further improve the accuracy of $PM_{2.5}$ concentrations.

References

[1] P. Xu, Y. Chen, and X. Ye, "Haze, air pollution, and health in China," *Lancet*, vol. 382, 2013, Art. no. 2067.
[2] F. Dominici *et al.*, "Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases," *J. Amer. Med. Assoc.*, vol. 295, pp. 1127–1134, 2006.
[3] C. A. Pope *et al.*, "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution," *J. Amer. Med. Assoc.*, vol. 287, pp. 1132–1141, 2002.
[4] 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, pp. 7436–7444, 2014.
[5] S. Weichenthal *et al.*, "Long-term exposure to fine particulate matter: Association with nonaccidental and cardiovascular mortality in the agricultural health study cohort," *Environ. Health Perspectives*, vol. 122, pp. 609–615, 2014.
[6] D. W. Dockery *et al.*, "An association between air pollution and mortality in six U.S. cities," *N. Engl. J. Med.*, vol. 329, pp. 1753–1759, 1993.
[7] M. B. Rice *et al.*, "Short-term exposure to air pollution and lung function in the Framingham Heart Study," *Amer. J. Respir. Crit. Care Med.*, vol. 188, pp. 1351–1357, 2013.
[8] 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, pp. 847–855, 2010.
[9] 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, pp. 645–675, 2012.
[10] W. You, Z. Zang, L. Zhang, Y. Li, X. Pan, and W. Wang, "National-scale estimates of ground-level $PM_{2.5}$ concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD," *Remote Sens.*, vol. 8, 2016, Art. no. 184.
[11] Y. Liu, C. J. Paciorek, and P. Koutrakis, "Estimating regional spatial and temporal variability of $PM_{2.5}$ concentrations using satellite data, meteorology, and land use information," *Environ. Health Perspectives*, vol. 117, pp. 886–892, 2009.
[12] D. Fu *et al.*, "Mapping nighttime $PM_{2.5}$ from VIIRS DNB using a linear mixed-effect model," *Atmos. Environ.*, vol. 178, pp. 214–222, 2018.
[13] Q. Xiao *et al.*, "Evaluation of VIIRS, GOCI, and MODIS collection 6 AOD retrievals against ground sunphotometer observations over East Asia," *Atmos. Chem. Phys.*, vol. 16, pp. 1255–1269, 2016.
[14] Y. Qi, J. Ge, and J. Huang, "Spatial and temporal distribution of MODIS and MISR aerosol optical depth over northern China and comparison with AERONET," *Chin. Sci. Bull.*, vol. 58, pp. 2497–2506, 2013.
[15] K. R. Kumar *et al.*, "Long-term (2003–2013) climatological trends and variations in aerosol optical parameters retrieved from MODIS over three stations in South Africa," *Atmos. Environ.*, vol. 95, pp. 400–408, 2014.
[16] Y. Xie, Y. Wang, K. Zhang, W. Dong, B. Lv, and Y. Bai, "Daily estimation of ground-level $PM_{2.5}$ concentrations over Beijing using 3 km resolution MODIS AOD," *Environ. Sci. Technol.*, vol. 49, pp. 12280–12288, 2015.
[17] 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, 2016, Art. no. 262.
[18] Z. Ma, Y. Liu, Q. Zhao, M. Liu, Y. Zhou, and J. Bi, "Satellite-derived high resolution PM concentrations in Yangtze river delta region of China using improved linear mixed effects model," *Atmos. Environ.*, vol. 133, pp. 156–164, 2016.
[19] Y. Yu *et al.*, "Short-term effects of ambient air pollutants and myocardial infarction in Changzhou, China," *Environ. Sci. Pollut. Res.*, vol. 25, pp. 22285–22293, 2018.
[20] B. Zou, J. Chen, L. Zhai, X. Fang, and Z. Zheng, "Satellite based mapping of ground $PM_{2.5}$ concentration using generalized additive modeling," *Remote Sens.*, vol. 9, 2016, Art. no. 1.
[21] K. Zhang, G. de Leeuw, Z. Yang, X. Chen, X. Su, and J. Jiao, "Estimating spatio-temporal variations of $PM_{2.5}$ concentrations using VIIRS-derived AOD in the Guanzhong Basin, China," *Remote Sens.*, vol. 11, 2019, Art. no. 2679.
[22] M. Jiang, W. Sun, G. Yang, and D. Zhang, "Modelling seasonal GWR of daily $PM_{2.5}$ with proper auxiliary variables for the Yangtze river delta," *Remote Sens.*, vol. 9, 2017, Art. no. 346.
[23] X. Li, C. Zhang, W. Li, and K. Liu, "Evaluating the use of DMSP/OLS nighttime light imagery in predicting $PM_{2.5}$ concentrations in the northeastern United States," *Remote Sens.*, vol. 9, 2017, Art. no. 620.
[24] 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, pp. 7436–7444, 2014.
[25] R. Li, X. Liu, and X. Li, "Estimation of the $PM_{2.5}$ pollution levels in Beijing based on nighttime light data from the defense meteorological satellite program-operational linescan system," *Atmosphere*, vol. 6, pp. 607–622, 2015.
[26] X. Li, F. Chen, and X. Chen, "Satellite-observed nighttime light variation as evidence for global armed conflicts," *IEEE J. Sel. Topic Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2302–2315, Oct. 2013.
[27] L. Han, W. Zhou, W. Li, and L. Li, "Impact of urbanization level on urban air quality: A case of fine particles ($PM_{(2.5)}$) in Chinese cities," *Environ. Pollut.*, vol. 194, pp. 163–170, 2014.
[28] G. Lin *et al.*, "Spatio-temporal variation of $PM_{2.5}$ concentrations and their relationship with geographic and socioeconomic factors in China," *Int. J. Environ. Res. Public Health*, vol. 11, pp. 173–186, 2014.
[29] M. L. Imhoff, W. T. Lawrence, D. C. Stutzer, and C. D. Elvidge, "A technique for using composite DMSP/OLS 'city lights' satellite data to accurately map urban areas," *Remote Sens. Environ.*, vol. 61, pp. 361–370, 1997.
[30] C. D. Elvidge *et al.*, "The Nightsat mission concept," *Int. J. Remote Sens.*, vol. 28, pp. 2645–2670, 2007.

[31] J. Wang *et al.*, "Detecting nighttime fire combustion phase by hybrid application of visible and infrared radiation from Suomi NPP VIIRS," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111466.

[32] X. Zhang and H. Hu, "Improving satellite-driven $PM_{2.5}$ models with VIIRS nighttime light data in the Beijing–Tianjin–Hebei region, China," *Remote Sens.*, vol. 9, 2017, Art. no. 908.

[33] J. Wang, C. Aegerter, X. Xu, and J. Szykman, "Potential application of VIIRS day/night band for monitoring nighttime surface $PM_{2.5}$ air quality from space," *Atmos. Environ.*, vol. 124, pp. 56–63, 2018.

[34] G. Zhang, L. Li, Y. Jiang, X. Shen, and D. Li, "On-orbit relative radiometric calibration of the night-time sensor of the LuoJia1-01 satellite," *Sensors*, vol. 18, 2018, Art. no. 4225.

[35] W. Jiang *et al.*, "Potentiality of using LuoJia1-01 nighttime light imagery to investigate artificial light pollution," *Sensors*, vol. 18, 2018, Art. no. 2900.

[36] X. Li, Z. Liu, X. Chen, and J. Sun, "Assessing the ability of LuoJia1-01 imagery to detect feeble nighttime lights," *Sensors*, vol. 19, 2019, Art. no. 3708.

[37] K. Shi, C. Huang, B. Yu, B. Yin, Y. Huang, and J. Wu, "Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas," *Remote Sens. Lett.*, vol. 5, pp. 358–366, 2014.

[38] R. H. Waring, N. C. Coops, W. Fan, and J. M. Nightingale, "MODIS enhanced vegetation index predicts tree species richness across forested eco regions in the contiguous USA," *Remote Sens. Environ.*, vol. 103, pp. 218–226, 2006.

[39] L. Zhuo, J. Zheng, X. Zhang, J. Li, and L. Liu, "An improved method of night-time light saturation reduction based on EVI," *Int. J. Remote Sens.*, vol. 36, pp. 4114–4130, 2015.

[40] P. Sutton, D. Roberts, C. Elvidge, and K. Baugh, "Census from heaven: An estimate of the global human population using night-time satellite imagery," *Int. J. Remote Sens.*, vol. 22, pp. 3061–3076, 2010.

[41] A. R. Huete, "A soil adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, pp. 295–309, 1988.

[42] Chinese National Ambient Air Quality Standards, GB 3095-2012. [Online]. Available: http://kjs.mep.gov.cn/hjbhbz/bzwb/dqhjbh/dqhjzlbz/201203/t20120302_224165.htm. Accessed on: January 12, 2016.

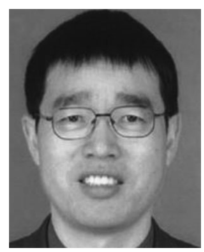
[43] J. Wang *et al.*, "Development of a nighttime shortwave radiative transfer model for remote sensing of nocturnal aerosols and fires from VIIRS," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111727.

Guo Zhang received the B.E. and Ph.D.E. degrees in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2000 and 2005, respectively.

His doctoral dissertation concerned the rectification for high-resolution remote sensing image under lack of ground control points. Since 2005, he has been with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, where he became a Professor in 2011. His research interests include space photogrammetry (including the processing and applications for optical load, SAR load, video load, and laser load).

Yingrui Shi received bachelor's degree from Sun Yat-sen University, Guangzhou, China, in 2018. She is currently working toward the master's degree at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in the field of cartography and geographic information systems.

Her research interests include the application of remote sensing, spatial big data, and remote sensing images processing.



Miaozhong Xu received the B.S. degree from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1989, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, in 2004.

He is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China. His research interests include high-resolution image processing, photogrammetry, and cartography.