ACS EARTH AND SPACE CHEMISTRY

http://pubs.acs.org/journal/aesccq

Article

# Integrating Fixed Monitoring Systems with Low-Cost Sensors to Create High-Resolution Air Quality Maps for the Northern China Plain Region

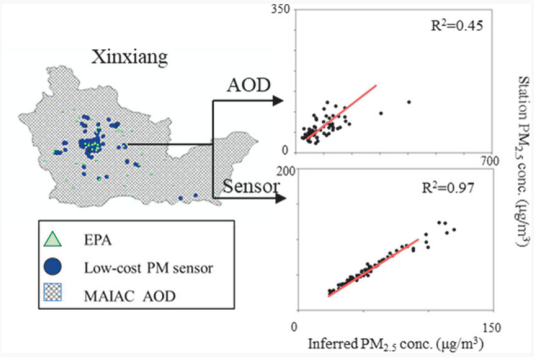
Chun-Ying Chao, Huang Zhang, Melanie Hammer, Yu Zhan, David Kenney, Randall V. Martin, and Pratim Biswas*

Cite This: *ACS Earth Space Chem.* 2021, 5, 3022−3035

Read Online

ACCESS | Metrics & More | Article Recommendations

**ABSTRACT:** To address and remediate severe particulate matter (PM) pollution in the North China Plain (NCP), many studies have traced pollution sources by using fixed air quality monitoring stations. However, these fixed monitors have high maintenance costs that make it economically infeasible to construct spatially dense networks for air quality measurement. Alternatively, using satellite monitoring systems and a low-cost air quality sensor network can greatly increase the spatiotemporal resolution of the ground-level PM concentration data for a given region. This study comprehensively examines the performance of China's EPA monitoring stations (CN-EPA), low-cost PM sensor networks and satellite aerosol optical depth (AOD) measurements. The goal is to improve the spatiotemporal resolution of ground-level PM concentration data for Xinxiang, a typical industrial city in the NCP. The inferred results show that low-cost PM sensors demonstrate high linearity with CN-EPA data sets for $PM_{2.5}$ concentrations with an $R^2$ value of 0.82. The $PM_{2.5}$ concentration inferred from the AOD retrievals demonstrates a moderate correlation with fixed monitoring stations with an $R^2$ value of 0.53. To evaluate the impact of human activities on air pollution, four traditional Chinese festivals, Chinese New Year, Tomb Sweeping Day, Ghost Festival, and Moon Festival, are chosen to observe the PM distribution in Xinxiang. Heat-maps of the ground-level $PM_{2.5}$ concentration reveal pollution hotspots in areas of high population density. Cross-validation is employed to evaluate the accuracy of the created pollution maps. The results demonstrate that pollution maps that were interpolated from data measured by CN-EPA data sets have the smallest root mean squared error (RMSE). Finally, our results show that low-cost PM sensor data can be integrated with traditional fixed air quality measurements to provide more detailed information about emission sources on pollution maps in urban and rural areas.




**KEYWORDS:** *air quality, $PM_{2.5}$, North China Plain, AOD, low-cost sensor, anthropogenic air pollution, Chinese New Year*



## I. INTRODUCTION

The North China Plain (NCP) suffers from air pollution from anthropogenic activities.[1−4] Severe haze events with extremely high particulate matter (PM) mass concentrations, ranging from 100 to 1000 $\mu g \cdot m^{-3}$, happen frequently in this region.[1] Research has demonstrated that a high PM concentration can cause cardiovascular, cerebrovascular, and respiratory problems.[5] The World Health Organization recently stated that exposure to ambient air pollution causes over 4 million deaths every year and that PM is a major source of ambient air pollution.[6] To reduce the adverse impacts of PM on human health and mitigate environmental concerns, national governments have routinely monitored and regulated air quality. The Chinese Ministry of Environmental Protection measures PM concentration by gravimetry, $\beta$-ray attenuation monitoring (BAM), and tapered element oscillating microbalance monitoring (TEOM).[3,7,8] Many studies use data from these three reference methods to evaluate health impacts, derive haze mechanisms, and improve the accuracy and quality of measurement techniques in the NCP area.[9−13] The reference methods yield high quality $PM_{2.5}$, fine particles with diameters that are below 2.5 $\mu$m, data that is comparable across all three methods, but these standard instruments can be deployed only sparsely due to their high cost and maintenance requirements.[14−16]

**Special Issue:** Mario Molina Memorial

**Received:** June 29, 2021
**Revised:** September 26, 2021
**Accepted:** September 28, 2021
**Published:** October 18, 2021





https://doi.org/10.1021/acsearthspacechem.1c00174
*ACS Earth Space Chem.* 2021, 5, 3022−3035

The air quality monitoring system of the Chinese EPA (CN-EPA) is sparse with only 1.05 monitors per million people and 1.53 monitors per 10 000 $km^2$ in 2014.[17] Most air quality monitors are primarily located in big cities. Thus, pollution mapping based on CN-EPA measurements can demonstrate a general trend, but it cannot accurately reveal pollutant concentrations that vary sharply over short distances.[18,19]

CN-EPA stations are usually deployed in urban areas. However, satellite remote sensing systems observing global distribution of aerosols can address the paucity of PM concentration measurements in rural areas.[20] Remote sensing has been used to predict $PM_{2.5}$ concentrations at local, regional, and global scale.[21] Since the early 2000s, the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Multiangle Imaging Spectroradiometer (MISR) instruments on the Terra and Aqua satellites have been applied to retrieve aerosol optical depth (AOD).[22−25] AOD is a measure of the aerosols (e.g., urban haze, smoke particles, desert dust, sea salt) distributed within a column of air from the Earth's surface to the top of the atmosphere.[26] Many studies convert AOD data to surface PM concentration via semiempirical models, simulation-based methods, and multivariate statistical regression.[27−29] To predict surface PM concentration from AOD, several factors are considered, including the AOD-PM retrieval algorithm, the column's vertical structure, the composition and size distribution of the particulate matter, and the water content of the aerosol.[20,30,31] In addition, to develop simple empirical relationships between AOD and $PM_{2.5}$ a chemical transport model can also be used to calculate the coincident ratio of AOD and $PM_{2.5}$, including aerosol composition, relative humidity, and the vertical structure of aerosol extinction.[32,33] In this research, satellite AOD data from the Multiangle Implementation of Atmospheric Correction (MAIAC) system is used to estimate $PM_{2.5}$. The satellite remote sensing system can observe the AOD in unpopulated and remote areas that conventional monitoring sites do not cover. However, remote-sensing-based $PM_{2.5}$ concentrations may have higher spatiotemporal uncertainties in their AOD values.

Besides remote sensor monitoring, increasing the number of ground monitors yields better air quality data.[19,34−36] Low-cost PM sensors have been shown as a potential method to supplement conventional PM concentration measurement and create a high-resolution pollution map.[15,37−39] These sensors use simple nephelometers or optical particle counters to convert optical signals to PM concentrations.[15] Besides their purchase price, low-cost PM sensors are small, lightweight, and have minimal power consumption. With these advantages, low-cost PM sensors have been used to identify pollution hot spots, track the sources of pollutants, and collect exposure data.[34] However, due to their design, low-cost PM sensors may have higher uncertainty than BAM and TEOM. The performance of low-cost PM sensors is affected by temperature, humidity, and ambient pressure.[40,41] Also, the accuracy of the pollution map depends highly on the sensor positions, and sampling at fixed locations may not be applicable for dynamic unpredictable sources.[37] Thus, low-cost PM sensor networks have to be calibrated with reference monitor methods to increase the accuracy of the pollution map.

Inferred PM concentrations from different measurements may improve the spatiotemporal resolution of sampling.[4,30,37,38,42,43] However, few studies have made estimates from ground-based sampling and satellite remote sensing in the North China Plain. Xu et al. measured ambient fine particulate matter over 336 cities in China from January 2016 to December 2017 by using the reference methods to find the correlation between PM concentration and health effects, but they did not create a high spatial resolution pollution map of each city.[16] Wu et al. observed AOD in Wuhan via MODIS in 2017.[44] They used the long short-term memory (LSTM) neural network method to improve the accuracy of $PM_{2.5}$ concentration values and compared them with values from 10 fixed monitoring stations in Wuhan. Their result from the neural network method, however, could not provide reliable maximum and minimum PM concentrations due to the algorithm's performance limitations and the small number of data points.[44] Thus, collecting sampling data from different methods can improve the prediction of PM concentration.[27,43] Previous studies of the NCP have not dealt with the limited spatiotemporal resolution in rural areas. Most studies have evaluated PM concentration via CN-EPA data sets and provided a general trend for PM pollution.

To fill these research gaps, data from two measurement approaches, CN-EPA data sets and low-cost sensor networks are used to map PM levels with a high spatiotemporal resolution for Xinxiang, a prefecture-level city in the northern Henan province in the NCP. Data from both low-cost sensors and satellite AOD measurements were compared to fixed CN-EPA data sets. Further, inferred $PM_{2.5}$ data was used to create pollution maps via an Inverse Distance Weighting (IDW) interpolation methodology.

## II. MATERIALS AND METHODS

**2.1. Data Sources and Their Spatial Distributions.** In this study, we investigate $PM_{2.5}$ concentrations in Xinxiang, a typical industrial city in northern Henan province. The data sets used are the Environmental Protection Administration in China (CN-EPA) data set, the (satellite) AOD data set, and the low-cost PM sensor data set. Xinxiang lies between 113°23′ and 114°59′E and 34°53′ and 35°50′N, and its total area is 8291 $km^2$. The monthly average temperature and humidity of Xinxiang are shown in Figure 1a,b. The humid, north subtropical monsoon climate averages around 30 °C in summer and 5 °C in winter with an annual average temperature of 20 °C.

The spatial distribution of measurement sites is shown in Figure 2. The red circles in Figure 2a are the high population areas in Xinxiang. The largest red circle in the center of Xinxiang is in downtown, the most populated area in the study. In Figure 2b, the 48 CN-EPA monitoring stations are marked by green triangles. The hourly average $PM_{2.5}$ concentrations were recorded at routine sampling locations in Xinxiang. The hourly $PM_{2.5}$ concentrations were measured by BAM or TEOM, methods which both have demonstrated good agreement with standard gravimetric measurement in previous studies.[19,31,36,45] Both BAM and TEOM can continuously monitor airborne particle concentration in real-time and provide reliable $PM_{2.5}$ levels. The AOD data set, marked by the black cross in Figure 2b, is obtained from the satellite data algorithm, MAIAC. Clearly, the satellite measurements have the greatest coverage. The black circles in Figure 2b represent data from 144 low-cost light scattering PM sensors. The low-cost PM sensor used in this study is the XHAQSN-808 model (Table 1). The measured PM concentration range is from 10 to 1000 $\mu g/m^3$, and the sensors can operate between −20 to 55 °C, and relative humidity ranges from 15 to 90%. The


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

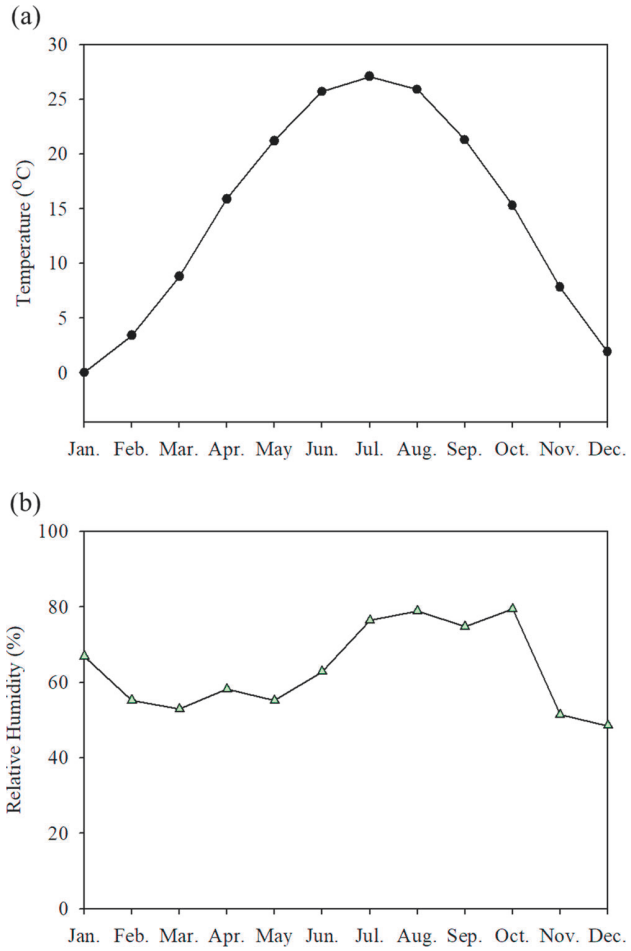



**Figure 1.** (a) Monthly average temperature and (b) monthly average relative humidity in Xinxiang during 2017.

detection limit of hourly PM concentration is 5 $\mu g/m^3$ and the resolution is 0.01 $\mu g/m^3$.[40] The low-cost PM sensor data sets in this study include measurement results for the whole year of 2017. The low-cost PM sensor spatial distribution in downtown Xinxiang is shown in Figure 2c.

The spatial distributions are different in each data set, and they are more heterogeneous for CN-EPA and AOD data sets than low-cost PM sensor data sets. CN-EPA monitoring stations are mainly located in schools and administrative buildings in every district. Low-cost PM sensors are mainly deployed in high population density areas and clustered around the commercial district in the central area of Xinxiang. Both instruments have lower density in the agricultural (rural) areas in the east of Xinxiang. The satellite remote sensing observes the AOD over most parts of Xinxiang, including the business districts and agricultural (rural) areas.

Statistics for each data set are provided in Table 2 and Figure 3. For the CN-EPA and low-cost PM sensor data sets, $PM_{2.5}$ mass concentrations are in units of $\mu g/m^3$. The data for the satellite remote sensing of the AOD is dimensionless. The temporal resolution of the CN-EPA and low-cost PM sensor data sets is 1 h, but for the AOD the temporal resolution is 1 day. The maximum, minimum, and arithmetic mean of the CN-EPA data sets are 396.79, 23.94, and 72.86 $\mu g/m^3$, respectively; for the low-cost PM sensor data sets they are 301.70, 22.78, and 74.66 $\mu g/m^3$; for the AOD data sets they

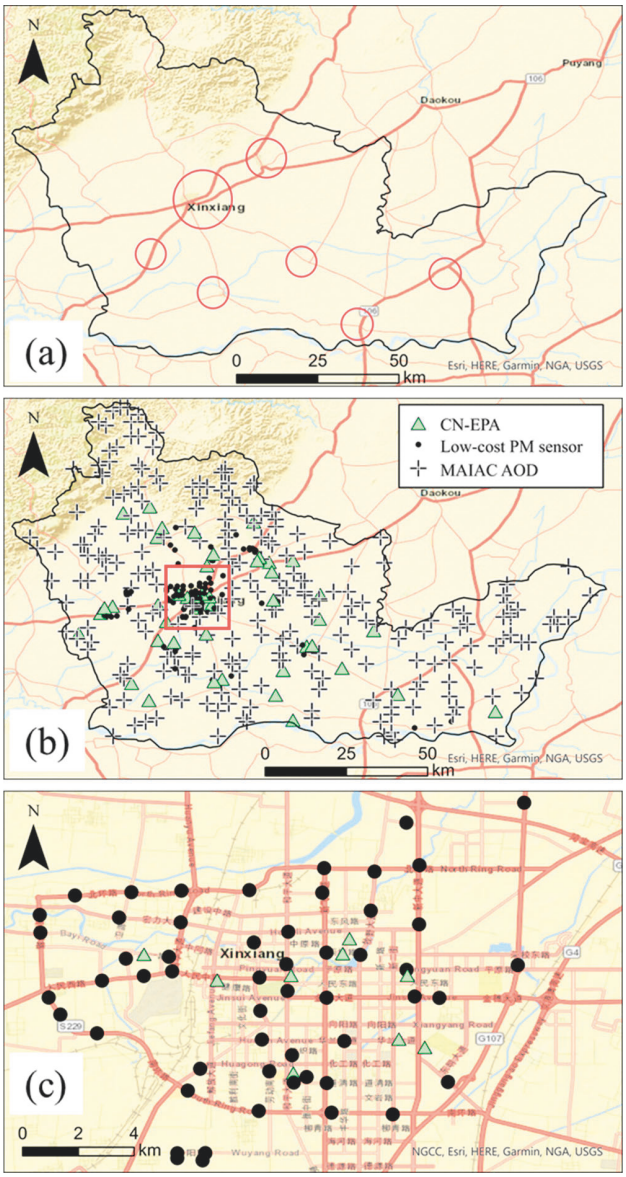



**Figure 2.** (a) High-population areas in Xinxiang. (b) Distribution of CN-EPA stations, low-cost PM sensors, and MAIAC AOD in Xinxiang. (c) Distribution of low-cost PM sensors and CN-EPA stations in downtown Xinxiang.

**Table 1. Characteristics of XHAQSN-808**

| instrument model | XHAQSN-808 |
|---|---|
| $PM_{2.5}$ measurement | method: laser light scattering<br>range: 10−1000 $\mu g/m^3$ |
| temporal resolution | 1 min |
| power supply | municipal power supply (220 V), or solar powered (12 V) |
| boundary dimension | 220 × 220 × 350 mm |
| work environment | temperature: −20−55 °C; RH: 15%−90% |
| communication | GPRS, Wi-Fi, Bluetooth |
| battery | lead-acid battery |
| working hours | 720 h without external power |
| weight | 2.5 kg |

are 4.00, 0.18, and 1.06, respectively. The maximum value of the CN-EPA data is much higher than the maximum value of


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

Table 2. Sampling Period and Statistical Information of $PM_{2.5}$ Dataset for Monitoring Station, Low-Cost PM Sensor, and MAIAC AOD in 2017

| | monitoring station | low-cost PM sensor | MAIAC AOD |
|---|---|---|---|
| sampling days | 365 days | 365 days | 280 days |
| maximum value | 396.79 $\mu g/m^3$ | 301.70 $\mu g/m^3$ | 4 |
| minimum value | 23.94 $\mu g/m^3$ | 22.78 $\mu g/m^3$ | 0.176 |
| averaged value | 72.86 $\mu g/m^3$ | 74.66 $\mu g/m^3$ | 1.056 |

low-cost PM sensor data because the CN-EPA data sets have more data from the industrial districts. The minimum and mean values of CN-EPA data sets and low-cost PM sensor data sets are similar because many sampling locations are in the business districts and have similar pollution sources.

**2.2. Equations for Retrieving and Cross-Validating Data.** We used linear regression to calibrate the PM concentration of the low-cost PM sensor and the AOD data sets. For the CN-EPA and low-cost PM sensor data sets, the low-cost PM sensors were paired with CN-EPA sites within a 500 m radius[46−48] and the hourly PM concentrations of low-cost PM sensors were correlated with the PM concentrations of CN-EPA sites. To calibrate the AOD data set, the CN-EPA and the low-cost PM sensor data sets were converted from the hourly data to daily data. For the daily CN-EPA and the low-cost PM sensor data sets, daily average PM concentrations were the average of all valid data for every location each day. For the AOD data set, we used daily AOD across Xinxiang. Although satellite AOD is retrieved daily, under some conditions (e.g., dense cloud cover, snow, bright land surface) the data are not valid.[49−51] To calibrate the MAIAC AOD to ground-level PM concentration, we matched the AOD data set with CN-EPA sites within a 500 m radius and correlated the daily CN-EPA data set with the AOD data set. The correlation equation for the linear regression is

$$PM_{2.5(ref)} = \alpha PM_{2.5(pre)} + \beta \quad (1)$$

where $PM_{2.5(ref)}$ are the $PM_{2.5}$ concentrations measured by the fixed monitoring sites, $PM_{2.5(pre)}$ are either the $PM_{2.5}$ concentrations measured by the low-cost PM sensors or the inferred $PM_{2.5}$ concentrations from the AOD, $\alpha$ is the slope of the line, and $\beta$ is the $y$-intercept.

Data sets were interpolated by inverse distance weighting (IDW), a geostatistical method for estimating unsampled points by a weighted average of the values of the sampled points. This method has been widely used to interpolate air pollution data.[52] Generally, IDW uses the following equation

$$\hat{v} = \frac{\sum_{i=1}^{n} \frac{1}{d_i^p} v_i}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \quad (2)$$

where $\hat{v}$ is the PM concentration to be estimated, $v_i$ is the measured PM concentration, $d_i^p$ are the distances from the measured points to the estimated points, and $p$ is the power of the estimated points. The lower the exponent is, the more uniform the interpolation is and the more neighbors that are incorporated into it. The higher the exponent is, the more discontinuous the interpolation is and the fewer the neighbors that are incorporated into it. In most situations, the power of the distance is 2.

The basic assumption of IDW is that concentrations that are spatially close to one another are more alike than those that are farther apart. In this study, the IDW geostatistical analyst in ArcGIS 10.6.1 was used to interpolate among four data sets from each day: the CN-EPA data sets, the calibrated low-cost PM sensor data sets, AOD derived $PM_{2.5}$ data sets and the combined $PM_{2.5}$ data sets which combined the CN-EPA data sets and the low-cost PM sensor data set together. The IDW geostatistical analyst predicted a value for any unsampled location and created the pollution map of Xinxiang.

Cross-validation is widely used to evaluate how well a model predicts values at unknown locations. We used leave-one-out cross-validation (LOOCV) to evaluate the performance of the results from IDW. LOOCV uses all the data to estimate trends and autocorrelation models. LOOCV omits one data point and calculates the value at the location of the omitted point by using the remaining data points. Then, it compares the predicted and the actual values. This procedure is repeated for each point. For example, assume that a data set includes $n$ data points and that we use $(x_i, y_i, PM_i)$ to represent each point, where $x$ is longitude, $y$ is latitude, and PM is the sampled PM concentration. For the first point, LOOCV removes data points $(x_i, y_i, PM_i)$, uses $n - 1$ data points to interpolate and predict the value of $(x_i, y_i, PM_{i,pre})$, and then compares the values of $PM_i$ and $PM_{i,pre}$. The procedure is repeated for a second point, and so on. The error and root-mean-square error (RMSE) are calculated by eqs 3 and 4

$$\text{error} = PM_i - PM_{i,pre} \quad (3)$$

$$\text{rmse} = \sqrt{\frac{\sum_{i=1}^{n} (PM_i - PM_{i,pre})^2}{n}} \quad (4)$$

## III. RESULTS AND DISCUSSION

**3.1. Correlations among the AOD, Monitoring Sites, and Low-Cost PM Sensor Data Sets.** Figures 4 and 5 show daily average and monthly average values for the three data sets. As can be seen from Figure 4b,c, the CN-EPA data sets and the low-cost PM sensor data sets show similar trends. In winter, the $PM_{2.5}$ concentrations are much higher than in summer, ranging between 90 to 180 $\mu g/m^3$ when the monthly average temperature is below 10 °C. Winter air particulate concentrations are affected by several causes, including different emission sources, winter monsoons, temperature, and atmospheric pressure.[53,54] Previous studies suggested that emissions from human activities are a major reason for $PM_{2.5}$ pollution in the NCP. The main source of emissions from human activities is incomplete combustion of coal in simple residential heating stoves.[55] These relatively high emissions include many products of incomplete combustion, such as $PM_{2.5}$ and carbon monoxide.[56]

In Figure 4a and Figure 5, the AOD data sets do not show a similar trend to those of the CN-EPA data sets and the low-cost PM sensor data sets. In summer, especially in July and August, the AOD data sets demonstrate higher variations than in spring and winter. The AOD value is not always strongly correlated with the ground $PM_{2.5}$ concentration because AOD represents light extinction by aerosol in the atmospheric column above the earth's surface.[20,31] Also, AOD is affected by the vertical structure, composition, size distribution, and water content of the atmospheric aerosols.[57−59] During summer, the variable relative humidity in Xinxiang introduced day-to-day unpredictability, so we divided the AOD data sets into months or seasons to infer the $PM_{2.5}$ concentration.


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

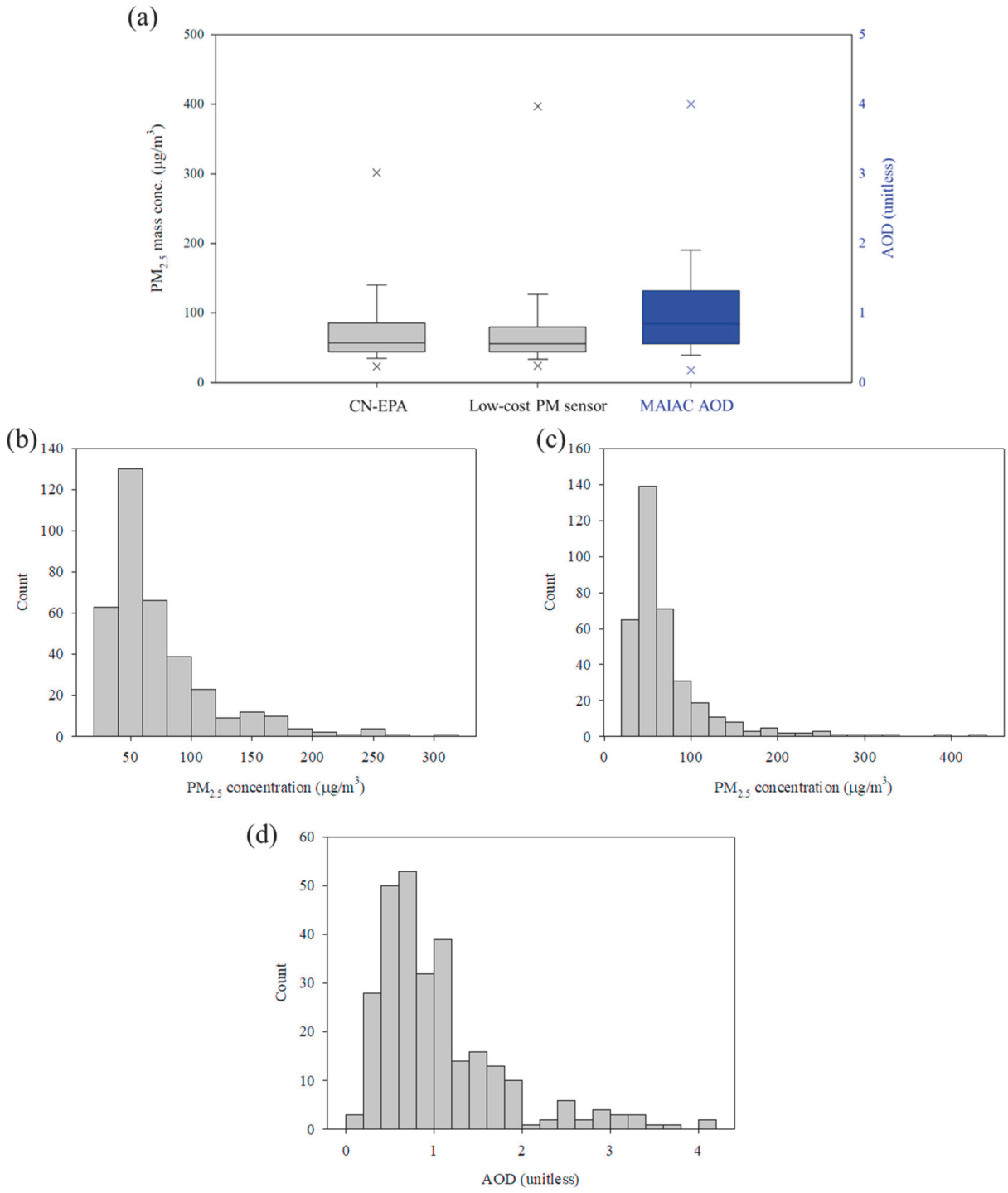



**Figure 3.** (a) Boxplots of MAIAC AOD and $PM_{2.5}$ concentrations measured by CN-EPA and low-cost PM sensors in Xinxiang during 2017. (b) Histogram of CN-EPA, (c) histogram of low-cost PM sensor, and (d) histogram of MAIAC AOD.

**3.2. Correlation between AOD Data Set and Monitoring Sites Data Sets.** Figure 6 shows the correlation between the AOD and CN-EPA data set determined by linear regression. There were 4494 paired data points. However, Figure 6 shows an $R^2$ value of 0.15 between the AOD data set and the CN-EPA data set. This $R^2$ value is at the lower end of literature reported values where $R^2$ values ranged from 0.01 to 0.64.[37,42] The reasons for the weak correlation between the AOD and the surface $PM_{2.5}$ concentrations are discussed in Section 3.1. Because the AOD-$PM_{2.5}$ relationships are time-dependent,[20] the monthly average ratio of $PM_{2.5}$ to AOD was used to reduce variations in the $PM_{2.5}$ concentration value caused by applying linear regression from the whole year data set in the daily AOD data set. Table 3 shows the monthly $PM_{2.5}$/AOD ratios calculated for the MAIAC AOD. The daily AOD data set was converted to $PM_{2.5}$ concentrations via


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

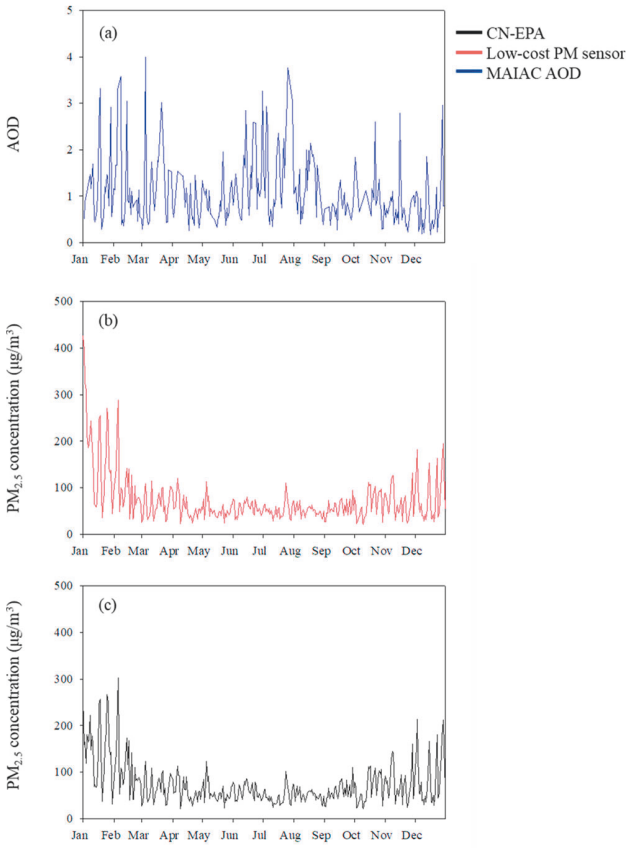



**Figure 4.** Daily average value of (a) MAIAC AOD, (b) $PM_{2.5}$ concentration measured by CN-EPA data sets, and (c) $PM_{2.5}$ concentration measured by low-cost PM sensors in 2017.

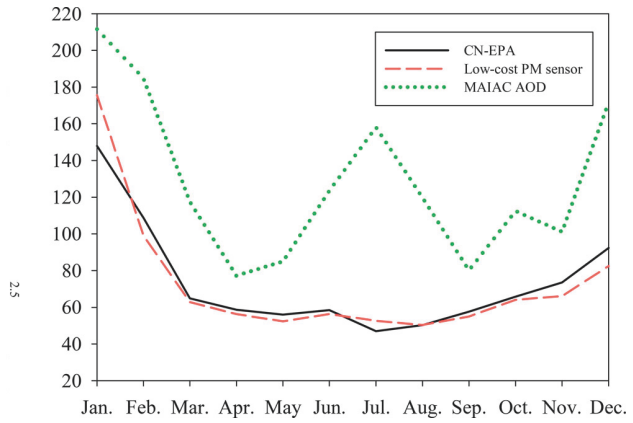



**Figure 5.** Monthly average of ground-level $PM_{2.5}$ concentration converted by MAIAC AOD and measured by CN-EPA and low-cost PM sensors in Xinxiang during 2017.

multiple monthly $PM_{2.5}$/AOD ratios. The monthly $PM_{2.5}$/AOD ratios for the MAIAC AOD ranged from 28.51 to 127.97.

Figure 7 shows the results of the inferred $PM_{2.5}$ values which are converted from the monthly $PM_{2.5}$/AOD ratios. The correlation between the inferred $PM_{2.5}$ values from the MAIAC AOD data set and the CN-EPA data set is moderate with an $R^2$ value of 0.53. The $R^2$ value from the monthly ratio is much

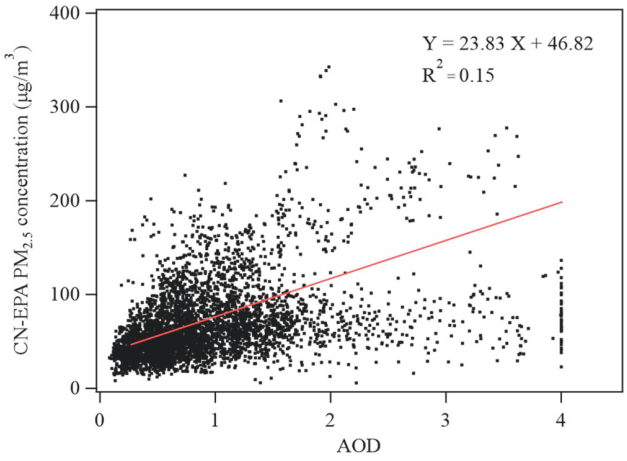



**Figure 6.** Correlation between MAIAC AOD and $PM_{2.5}$ concentration measured by CN-EPA in 2017.

**Table 3. $PM_{2.5}$/AOD Ratios from the Dataset of Satellite Product and CN-EPA**

| | $PM_{2.5}$/AOD ratio (MAIAC AOD) |
|---|---|
| Jan. | 127.97 |
| Feb. | 89.71 |
| Mar. | 55.78 |
| Apr. | 70.65 |
| May | 64.65 |
| Jun. | 43.79 |
| Jul. | 28.51 |
| Aug. | 41.45 |
| Sep. | 76.54 |
| Oct. | 65.84 |
| Nov. | 95.79 |
| Dec. | 119.34 |

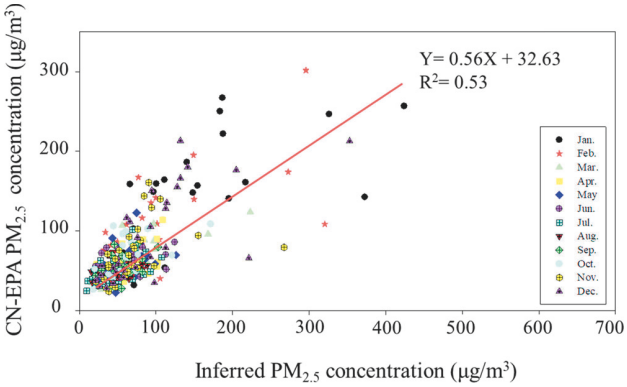



**Figure 7.** Correlation between measured and inferred $PM_{2.5}$ concentration in Xinxiang for different months during 2017.

higher than the value from directly using the daily data, which is 0.15. The monthly $PM_{2.5}$/AOD ratios reduce the time-dependent variations of linear regression. Thus, using the monthly ratios for MAIAC AOD are more suitable to estimate ground level $PM_{2.5}$.

After understanding the importance of time in the AOD-$PM_{2.5}$ relationships, we observed the AOD-$PM_{2.5}$ relationship in different seasons. Figure 8 demonstrates the correlations between the MAIAC AOD data sets and the CN-EPA data sets in different seasons. The ratios for the MAIAC AOD have a


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

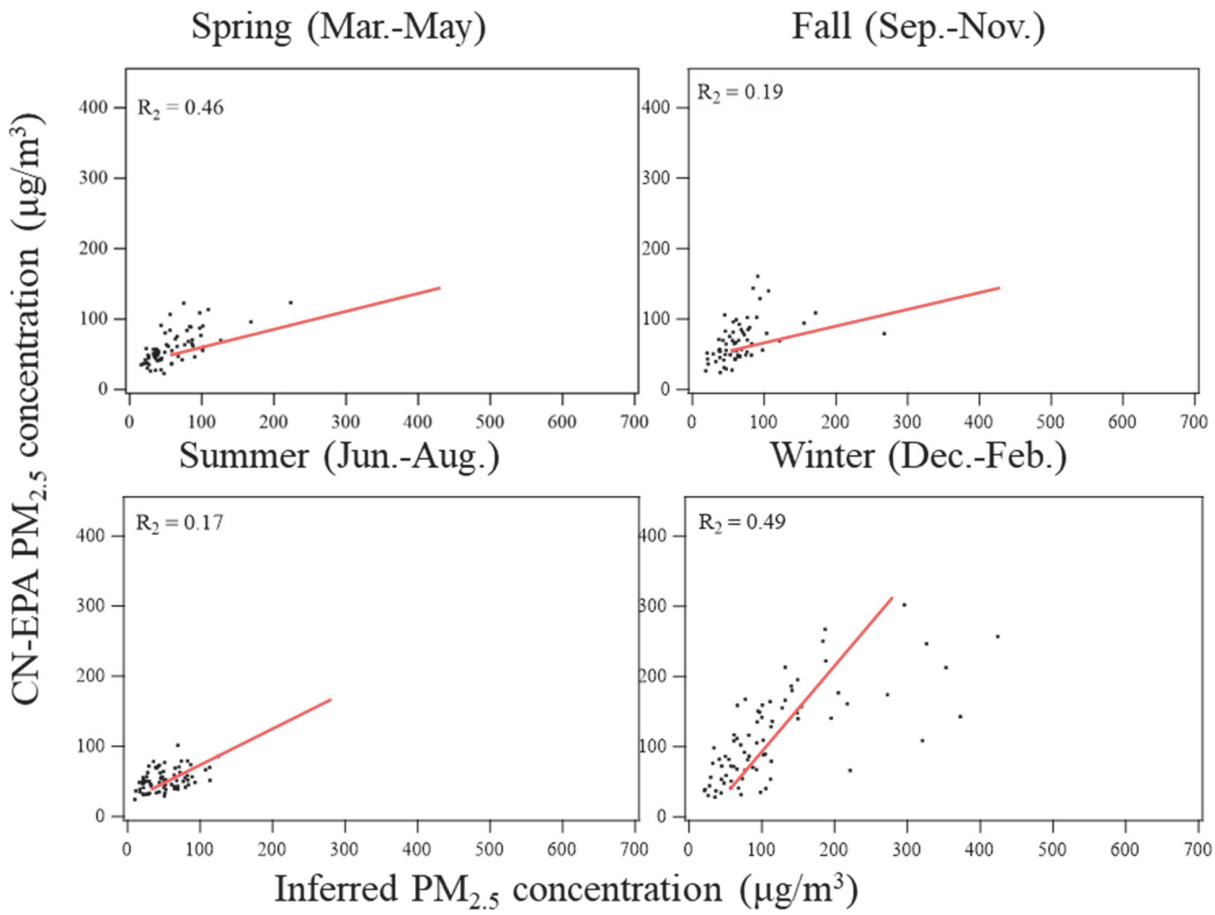



**Figure 8.** Correlation between measured and inferred $PM_{2.5}$ concentration in Xinxiang for different seasons during 2017.

higher relationship in the cold weather of spring and winter. In warmer weather, for example, summer and fall, the correlations are weak. Previous studies have similarly found that in the North China Plain the correlation is lower in summer and higher in winter. The variations in the inferred $PM_{2.5}$ concentrations were caused by weather conditions, including the cloud fraction and relative humidity.[42,58,60]

**3.3. Correlation between Low-Cost PM Sensor Data Sets and Monitoring Site Data Sets.** The hourly $PM_{2.5}$ concentrations obtained by low-cost PM sensors were correlated with CN-EPA sites within a 500 m radius. The calibration procedure includes three steps: first, designating the low-cost PM sensors within 500 m radius of CN-EPA sites as pairs; second, setting low-cost PM sensor measurement as the *x*-value and CN-EPA measurement as the *y*-value; third, using linear regression to yield the regression equation. There were 22 590 paired data points for CN-EPA and low-cost PM sensor data sets. In the linear regression, the *x*-variable is the low-cost PM sensor data and the *y*-variable is the CN-EPA data. Figure 9a illustrates the hourly $PM_{2.5}$ concentrations in the low-cost PM sensor data and the CN-EPA data in 2017. The calibration used eq 5, and the $R^2$ value is 0.82. The standard error for slope and intercept are 0.002 and 0.22. This $R^2$ value and low standard error of slope and intercept demonstrate a strong correlation, indicating that the low-cost PM sensor data are accurate for the reported sampling period. RMSEs were used to evaluate errors in the regression. The RMSE of Figure 9a is 21.39, indicating that in some situations the low-cost sensor data might not be represented accurately by the CN-EPA site data. For example, the data in the red circle indicated high response in EPA monitor sites but low responses in low-cost PM sensors; this means at the same time these pair might measure the same pollution events.

$$PM_{2.5,\text{calibrated}} = (0.86 \pm 0.002)PM_{2.5,\text{sensor}} + (8.69 \pm 0.22) \quad (5)$$

The low-cost PM sensors do not always perform stably over a long period.[36] In Figure 9b, we showed the correlations of the low-cost PM sensor data sets and the CN-EPA data sets in different seasons. To compare with the AOD data set, the daily average data was used. The seasonal calibration equation results are plotted in Figure 9. From spring to winter, the $R^2$ values are 0.91, 0.69, 0.89, and 0.85. Seasonal variations in $R^2$ values were also found in the previous studies.[61] The reasons include seasonal differences, meteorological factors (wind speed, temperature, relative humidity, and air mass), and specific events (e.g., fireworks). From June to August 2017, the daily average relative humidity (RH) was above 75% on 47 days and above 90% on 5 days. The working environment of RH for the XHAQSN-808 model is from 15% to 90%. The high RH can lead to a failure of the circuits of the sensors and cause biased measurements.[62,63] Additionally, the RMSEs from spring to winter are 7.17, 7.23, 9.46, and 24.62. Higher RMSEs in fall and winter were obtained than for spring and summer. The possible reason causing the high RMSE might be due to anthropogenic activities. In Xinxiang, there are many households heating with coal stoves in winter and these stoves cause


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

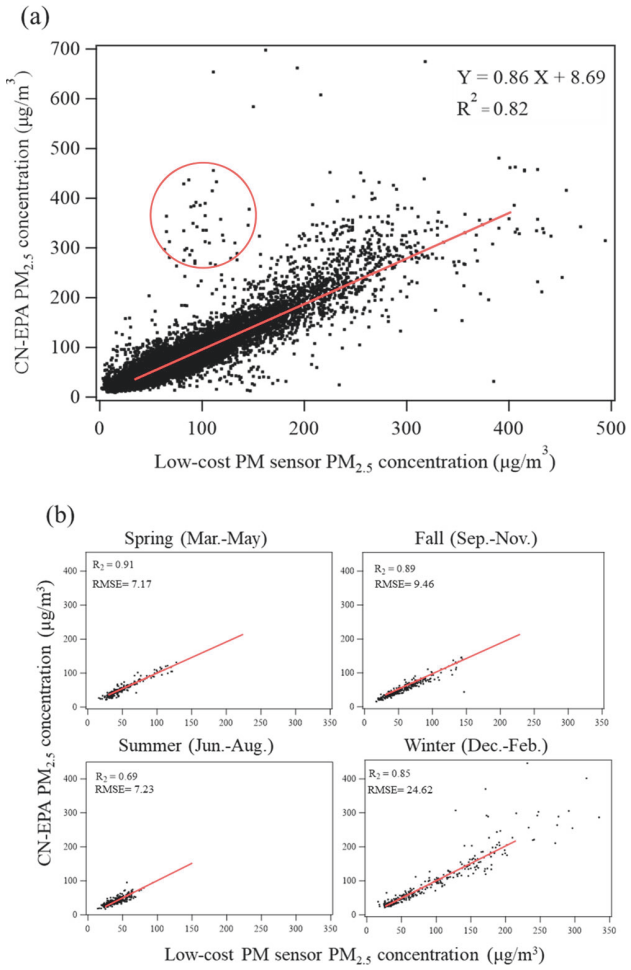



**Figure 9.** Correlation between $PM_{2.5}$ concentration measured by CN-EPA and low-cost PM sensors in Xinxiang for (a) whole data sets and (b) different seasons during 2017.

high emissions of $PM_{2.5}$ in high population areas.[55] These events could cause differences in CN-EPA measurements and low-cost sensor measurements.

**3.4. Cross-Validation Results.** We performed the cross-validations for seven data sets: (1) the CN-EPA data sets, (2) the low-cost sensor data sets, (3) the AOD data set, (4) the low-cost sensor combining CN-EPA data set, (5) the low-cost sensor combining AOD data set, (6) the CN-EPA combining AOD data set, and (7) the CN-EPA combining low-cost sensor and AOD data sets. $PM_{2.5}$ concentration data obtained by the low-cost PM sensors were calibrated according to eq 5. The $PM_{2.5}$/AOD ratios from Table 3 were used to infer the AOD data set to ground-level $PM_{2.5}$ concentrations. The size of the data sets and cross-validation results are provided in Tables 4 and 5. For the CN-EPA data set, there were 18 and 24 data points during Chinese New Year and Tomb Sweeping Day, whereas there were 46 and 47 data points during Ghost Festival and National Day. Similarly, there was fewer low-cost PM sensor data point during Chinese New Year and Tomb Sweeping Day than Ghost Festival and National Day. For AOD data sets, the data point number were similar from 54 to 67 data points in different sampling events. The LOOCV method is used to calculate the RMSEs and average absolute errors during Chinese New Year, Tomb Sweeping Day, Ghost Festival, and National Day in 2017.

The LOOCV results are shown in Table 6. The RMSEs of the AOD and CN-EPA data sets have smaller values for each event than other data sets. The RMSEs of low-cost sensor data sets are small except for the Chinese New Year Days (59.93). This high value means that there might be a local high concentration event that happened during the sampling period. The LOOCV method compares the values of the prediction with the value of measurements from the data set. The results of lower RMSEs show that the predictions are similar to the measurements from the data set. Thus, to evaluate whether the data set can accurately reflect $PM_{2.5}$ concentrations, the CN-EPA data sets were combined with AOD and low-cost sensor data sets to observe the change of the RMSEs and mean absolute error. Although the AOD data sets provide moderate RMSEs and low mean absolute errors when used by itself, after combining with CN-EPA data sets the mean absolute error increased from 7 to 12 on the Chinese New Year Day and the RMSEs increased in all events except the Tomb Sweeping Day. It means that the AOD derived $PM_{2.5}$ has differences with CN-EPA measurements and they may not be suitable to use for ground-level $PM_{2.5}$ concentration predictions. On the other hand, the RMSEs and mean absolute errors decreased after the low-cost PM sensor data set was combined with the CN-EPA data sets.

The difference might be caused by the different spatial distributions, data continuities, and size of the data sets. The monitoring sites for CN-EPA data sets are more homogeneously distributed than those of the low-cost PM sensor data sets and combined $PM_{2.5}$ data sets. The low-cost PM sensor data sets are clustered in densely overpopulated areas instead of rural areas. The AOD data sets are homogeneously distributed in urban and rural areas. The heterogeneity of the data sets can cause prediction errors and lead to larger RMSEs in cross-validation. In addition to the spatial distribution of data sets, the operating characteristics of low-cost PM sensors need to be considered. Low-cost PM sensors can capture the high spatial heterogeneity of microenvironments, thus increasing the RMSE of interpolation.[37] For example, in closely spaced locations, low-cost PM sensors sometimes report different $PM_{2.5}$ concentrations because of small-scale pollution events. Thus, the data is discontinuous, and nearby sensors may not report similar concentrations.

Because the data sets have different numbers of data points and different spatial distributions, using LOOCV to calculate the RMSEs to average absolute errors for every data point

**Table 4. Meteorological Conditions on Sampling Days**

| | Chinese New Year Day | Tomb Sweeping Day | Ghost Festival Day | Chinese National Day |
|---|---|---|---|---|
| sampling period | 01/20−02/15 | 04/01−04/06 | 09/04−09/06 | 09/28−10/10 |
| pressure (hPa) | 1007.2 | 994.5 | 990.3 | 999.1 |
| temperature (°C) | 1.9 | 14.9 | 22.1 | 16.2 |
| RH (%) | 55.1 | 66.3 | 85.8 | 84.1 |


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

Table 5. Measurement Site Number for CN-EPA, Low-Cost PM Sensor, Combined $PM_{2.5}$, and AOD Datasets on Sampling Days

| | Chinese New Year Day 01/20−02/15 | Tomb Sweeping Day 04/01−04/06 | Ghost Festival Day 09/04−09/06 | Chinese National Day 09/28−10/10 |
|---|---|---|---|---|
| CN-EPA data set | 18 | 24 | 46 | 47 |
| low-cost PM sensor | 66 | 69 | 97 | 107 |
| AOD data set | 66 | 67 | 54 | 64 |

Table 6. LOOCV of Seven Datasets[a]

| | Chinese New Year Day 01/20 −02/15 | Tomb Sweeping Day 04/01 −04/06 | Ghost Festival Day 09/04 −09/06 | Chinese National Day 09/28 −10/10 |
|---|---|---|---|---|
| | **RMSE** | | | |
| CN-EPA | 12.04 | 12.99 | 10.17 | 7.51 |
| LCS | 59.93 | 10.53 | 11.71 | 10.77 |
| AOD | 10.46 | 12.94 | 15.35 | 17.83 |
| LCS + CN-EPA | 54.23 | 10.57 | 11.72 | 10.00 |
| LCS + AOD | 43.10 | 14.07 | 10.67 | 15.35 |
| AOD + CN-EPA | 16.17 | 15.35 | 11.51 | 17.83 |
| CN-EPA + LCS + AOD | 41.81 | 13.77 | 11.24 | 14.67 |
| | **Mean Absolute Error** | | | |
| CN-EPA | 9.99 | 10.57 | 7.48 | 5.34 |
| LCS | 19.71 | 7.76 | 8.97 | 7.29 |
| AOD | 7.11 | 8.05 | 4.70 | 10.23 |
| LCS + CN-EPA | 18.43 | 8.06 | 8.89 | 6.87 |
| LCS + AOD | 11.90 | 11.17 | 6.14 | 10.29 |
| AOD + CN-EPA | 12.64 | 10.42 | 7.37 | 11.10 |
| CN-EPA + LCS + AOD | 11.83 | 10.18 | 8.45 | 8.76 |

[a]CN-EPA, low-cost PM sensor (LCS), AOD, LCS + CN-EPA, LCS + AOD, AOD + CN-EPA, and CN-EPA + LCS + AOD datasets on sampling days.

might not be directly comparable between the different data sources. Thus, we used a similar method to LOOCV but did not calculate the RMSEs for every data point. The CN-EPA data set was used as a reference in each event to calculate the RMSEs and average absolute errors between measured and predicted $PM_{2.5}$. We used the CN-EPA data points in each event by omitting one data point and calculating the value at the location of the omitted point using the remaining data points. Only the CN-EPA sites were systematically omitted for this calculation, and the AOD and low-cost sensor data were all used for each calculation. In addition to RMSEs, normalized RMSEs were calculated by dividing the RMSE by the mean of the measurement data.

The results are shown in Table 7. Unlike the LOOCV results, the AOD data sets show higher RMSEs, NRMSEs, and absolute errors in all events. This indicates that the use of the AOD derived $PM_{2.5}$ in the ground-level $PM_{2.5}$ concentrations may not be suitable in our study. Additionally, the low-cost PM sensor combining with the CN-EPA data set shows the RMSEs, NRMSEs, and absolute errors were improved in this study.

Comparing with different events, the Chinese New Year Day had the largest difference for the RMSEs and average absolute errors. The RMSEs of low-cost PM sensor, AOD, and CN-EPA combining low-cost PM sensor data sets were 27.93, 35.47, and 30.31 and the average absolute errors were 22.26, 32.79, and

Table 7. LOOCV Results by Using CN-EPA as Reference of Six Datasets[a]

| | Chinese New Year Day 01/20−02/15 | Tomb Sweeping Day 04/01−04/06 | Ghost Festival Day 09/04−09/06 | Chinese National Day 09/28−10/10 |
|---|---|---|---|---|
| | | **RMSE** | | |
| LCS | 27.93 | 12.75 | 16.50 | 9.50 |
| AOD | 35.47 | 34.67 | 20.78 | 15.04 |
| LCS + CN-EPA | 30.31 | 11.12 | 12.03 | 8.56 |
| LCS + AOD | 47.76 | 13.82 | 17.01 | 12.85 |
| AOD + CN-EPA | 21.10 | 16.83 | 12.66 | 13.24 |
| CN-EPA + LCS + AOD | 25.89 | 12.56 | 12.70 | 10.93 |
| | | **Mean Absolute Error** | | |
| LCS | 22.26 | 10.53 | 12.01 | 6.73 |
| AOD | 32.79 | 31.39 | 10.16 | 11.08 |
| LCS + CN-EPA | 17.47 | 9.04 | 9.31 | 6.08 |
| LCS + AOD | 35.66 | 11.86 | 8.73 | 9.55 |
| AOD + CN-EPA | 17.56 | 13.10 | 4.51 | 9.40 |
| CN-EPA + LCS + AOD | 22.51 | 10.06 | 10.57 | 8.26 |
| | | **NRMSE** | | |
| LCS | 0.24 | 0.17 | 0.18 | 0.18 |
| AOD | 0.36 | 0.77 | 0.32 | 0.27 |
| LCS + CN-EPA | 0.24 | 0.15 | 0.22 | 0.17 |
| LCS + AOD | 0.40 | 0.20 | 0.29 | 0.24 |
| AOD + CN-EPA | 0.18 | 0.25 | 0.23 | 0.25 |
| CN-EPA + LCS + AOD | 0.24 | 0.19 | 0.23 | 0.21 |

[a]CN-EPA, low-cost PM sensor (LCS), AOD, LCS + CN-EPA, LCS + AOD, AOD + CN-EPA, and CN-EPA + LCS + AOD datasets on sampling days.



https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

17.47. For the CN-EPA combining low-cost PM sensor AOD data sets, the predictions of the maps were improved by adding the data points covering the whole Xinxiang. Regarding the NRMSEs, the values were from 0.15 to 0.24 and the Chinese New Year Day and Ghost Festival Day had the largest difference for NRMSE.

To analyze the performance of AOD and low-cost PM sensor data sets in urban and rural areas, we designated several CN-EPA monitoring sites in central Xinxiang as urban sites, and sites with no low-cost PM sensors within a 500 m radius as rural sites. Figure 10 shows the locations of urban and rural

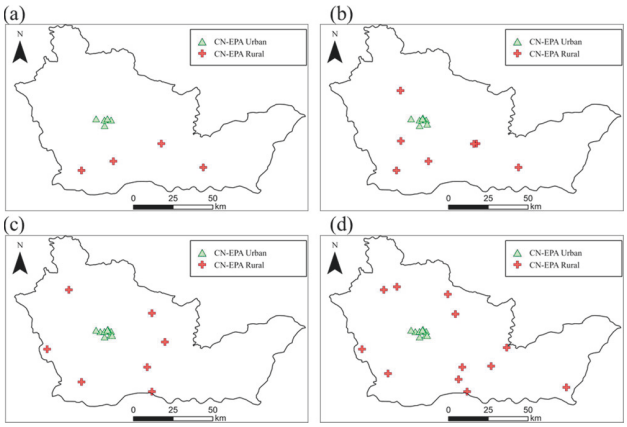



**Figure 10.** CN-EPA site locations in urban and rural areas in four sampling periods: (a) Chinese New Year day, (b) Tomb Sweeping Day, (c) Ghost Festival, and (d) National Day.

sites. Using the same method in previous paragraphs, we compared the predictions from AOD and low-cost PM sensor data sets with the CN-EPA data sets and calculated the RMSEs, NRMSEs, and mean absolute errors. In Table 8, the results show that the AOD data set provides larger RMSEs, NRMSEs, and mean absolute errors in both urban and rural areas. Thus, the AOD data sets may not provide accurate ground-level $PM_{2.5}$ concentrations in urban and rural areas in this study. Regarding the low-cost PM sensor data sets, their performance was similar to urban and rural areas which indicated that the low-cost PM sensor data sets could provide useful information in both urban and rural areas.

**3.5. $PM_{2.5}$ Pollution Map Visualization.** In this section, we plot 2017 pollution maps of $PM_{2.5}$ concentrations in Xinxiang during Chinese New Year, Tomb Sweeping Day, Ghost Festival, and Chinese National Day. During these festivals, fireworks and ghost money burning usually cause severe hazes in the NCP.[2,5,12,37]

From the discussion in Section 3.4, the CN-EPA combined with low-cost PM sensor data set provided more data points and moderate RMSEs, NRMSEs, and mean absolute error. Four data sets, the CN-EPA data set, the calibrated low-cost PM sensor data set, AOD data set, and CN-EPA combined low-cost PM sensor data set, were imported into ArcGIS pro software to obtain $PM_{2.5}$ distribution maps determined by the IDW method.

Figure 11(a-1), (a-2), and (a-3) shows pollution maps from CN-EPA data sets, low-cost PM sensor data sets, and AOD

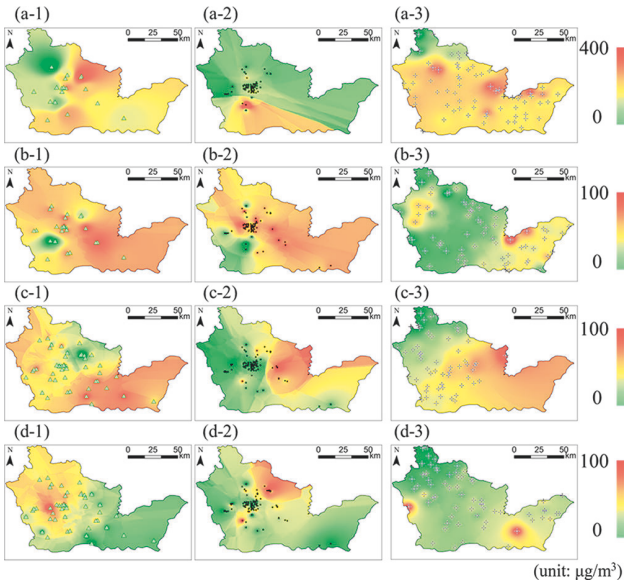



**Figure 11.** Maps of (a-1) CN-EPA, (a-2) low-cost PM sensor, and (a-3) AOD data sets on Chinese New Year Day. Heat-maps of (b-1) CN-EPA, (b-2) low-cost PM sensor, and (b-3) AOD data sets on Tomb Sweeping Day. Heat-maps of (c-1) CN-EPA, (c-2) low-cost PM sensor, and (c-3) AOD data sets on Ghost Festival. Heat-maps of (d-1) CN-EPA, (d-2) low-cost PM sensor, and (d-3) AOD data sets on National Day.

data sets during the Chinese New Year; panels (b-1), (b-2), and (b-3) show pollution maps during Tomb Sweeping Day; panels (c-1), (c-2), and (c-3) show pollution maps during Ghost Festival; panels (d-1), (d-2), and (d-3) show pollution maps during National Day.

Pairwise comparison of Figure 11(a-1) and (a-2), (b-1) and (b-2), (c-1) and (c-2), and (d-1) and (d-2) demonstrates that

**Table 8. Performance of AOD and LCS Datasets in Rural and Urban Areas Compared to CN-EPA Data**

| | Chinese New Year Day 01/20–02/15 | | Tomb Sweeping Day 04/01–04/06 | | Ghost Festival Day 09/04–09/06 | | Chinese National Day 09/28–10/10 | |
|---|---|---|---|---|---|---|---|---|
| | urban | rural | urban | rural | urban | rural | urban | rural |
| | | | | AOD | | | | |
| RMSE | 32.50 | 38.75 | 40.30 | 33.23 | 24.98 | 16.02 | 7.53 | 18.26 |
| mean absolute errors | 29.99 | 37.87 | 39.01 | 30.88 | 21.87 | 14.11 | 5.32 | 15.94 |
| NRMSE | 0.32 | 0.40 | 0.97 | 0.73 | 0.36 | 0.27 | 0.15 | 0.31 |
| | | | | LCS | | | | |
| RMSE | 21.22 | 14.28 | 11.28 | 16.18 | 12.25 | 11.22 | 8.32 | 8.50 |
| mean absolute errors | 16.57 | 11.46 | 9.04 | 12.94 | 9.35 | 8.12 | 4.70 | 6.63 |
| NRMSE | 0.18 | 0.11 | 0.15 | 0.23 | 0.25 | 0.18 | 0.18 | 0.16 |


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

the pollution maps interpolated from CN-EPA data sets are much different compared to maps from the calibrated low-cost PM sensor data sets. The maps interpolated from the low-cost PM sensor data sets indicate more sharp, triangular-like regions, especially in the rural areas of Xinxiang. This difference may be due to the locations of the low-cost PM sensors. Most low-cost PM sensors are located in the densely populated central area of Xinxiang. Relying on data from sensors clustered near the central area of Xinxiang may lead to the failure to predict PM concentrations in more outlying areas. On the other hand, the monitoring stations are homogeneously dispersed in the densely and sparsely populated areas. For this reason, in Figure 11(a-1), (b-1), (c-1), and (d-1) the pollution maps of the CN-EPA data sets are more smooth than Figure 11(a-2), (b-2), (c-2), and (d-2).

During Chinese New Year and Tomb Sweeping Day, the $PM_{2.5}$ concentrations are much higher than during the other two events. The causes may be conventional activities (e.g., fireworks and burning ghost money) and coal combustion for heating.[55,64] The highest $PM_{2.5}$ concentration in Figure 11(a-2) is in a high population density area near downtown Xinxiang. However, as seen in Figure 11(a-1), the CN-EPA monitors did not detect the high PM concentration near the sampling location of the low-cost PM sensor. This difference in detection performance might be caused by microenvironments (e.g., fireworks and biomass burning) that did not affect measurements by CN-EPA monitors. Also, this phenomenon is seen in Figure 11(b-1) and (b-2). From this observation, using the low-cost PM sensor can provide more information on pollution sources that are neglected by CN-EPA measurement sites.

During Ghost Festival, Figure 11(c-1) shows PM concentrations generally higher than Figure 11(c-2), especially in southern Xinxiang where there are fewer low-cost PM sensors. On the other hand, in Figure 11(c-2) low-cost PM sensors measured pollution sources in the middle-eastern area. The pollution source may be due to the tradition of burning ghost money. People in China usually burn ghost money and have many ceremonies to pray for health and safety. These anthropological activities involve burning and may cause small-scale $PM_{2.5}$ pollution in the agricultural area.[5] During National Day, Figure 11(d-1) shows a pollution hot spot in central Xinxiang, and Figure 11(d-2) shows an additional pollution hot spot in north Xinxiang. The low-cost PM sensor data sets provide detailed pollution event records in this area, although they cannot identify causes.

In Figure 11(a-3), (b-3), (c-3), and (d-3), the AOD data set provided different hotspots to the CN-EPA and low-cost PM sensor data sets. We can observe some hotspots in the eastern parts of Xinxiang. In the eastern parts of Xinxiang, there are several villages with smaller business areas and less populations than central Xinxiang. The pollution events, which were neglected in the ground monitor data, were observed in the AOD data sets during the Tomb Sweeping Day and the Chinese National Day. The AOD data set also provided lower $PM_{2.5}$ concentrations in north Xinxiang which is a mountainous area with fewer human activities.

Figure 12a–d shows pollution maps interpolating the combination of two data sets: CN-EPA and low-cost PM sensor data sets during the Chinese New Year, Tomb Sweeping Day, Ghost Festival Day, and Chinese National Day. The pollution maps showed higher spatial resolution in four sampling periods. Combining data from monitoring stations and low-cost PM sensors yields a PM distribution map with greater detail of Xinxiang. Increasing the data points for the whole Xinxiang captured pollution sources in the densely populated areas with high resolution in downtown. This sensitivity can help in finding emission sources and estimating exposures. Furthermore, the results can help in regulating human activities to control severe air pollution in

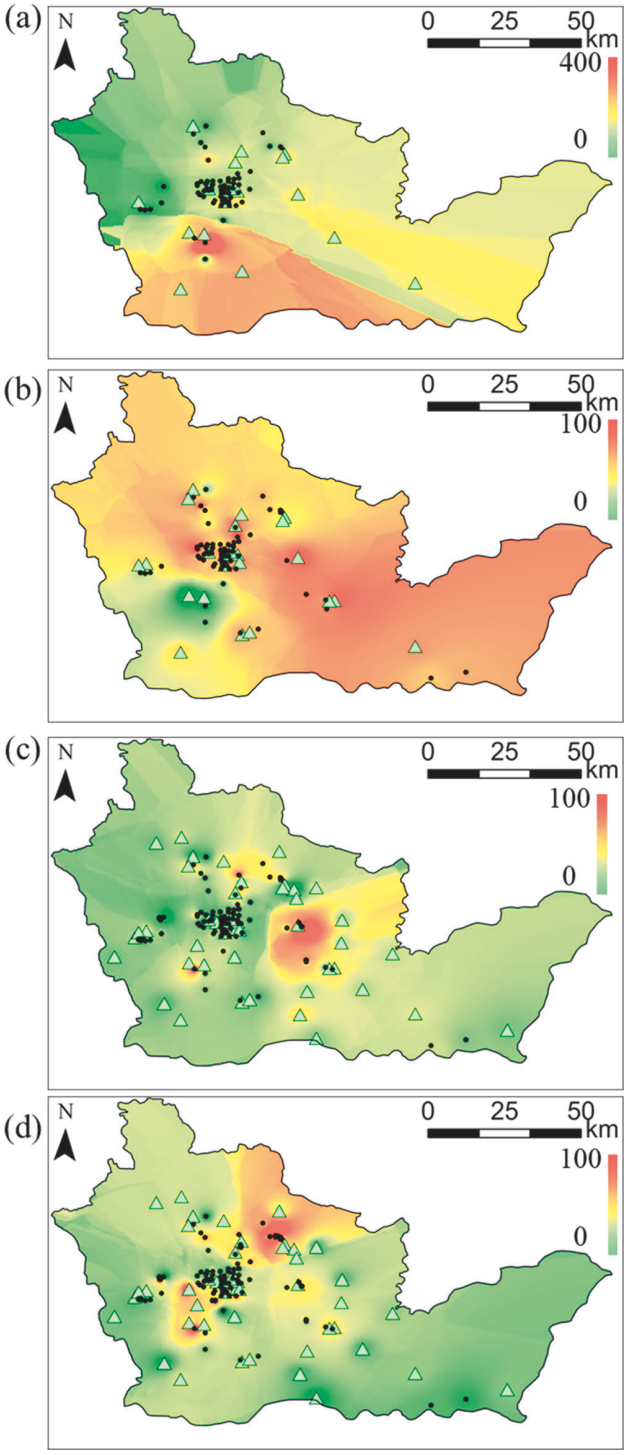



**Figure 12.** Heat-maps of CN-EPA + LCS data sets on Chinese New Year Day, Tomb Sweeping Day, Ghost Festival, and Chinese National Day.


https://doi.org/10.1021/acsearthspacechem.1c00174
ACS Earth Space Chem. 2021, 5, 3022−3035

Xinxiang. To improve the accuracy of prediction in rural areas, we need to consider the spatial distribution and the characteristics of low-cost PM sensor data sets. Increasing the number of data points in rural areas would yield a more homogeneous data set, increasing the accuracy of IDW interpolation.

## IV. CONCLUSIONS

Integrating PM data sets from different measurement methods improves the spatial resolution of PM measurements. This paper illustrated the integration of $PM_{2.5}$ data sets from 48 CN-EPA stations and 144 low-cost PM sensors in Xinxiang, a traditional city in the North China Plain. Additionally, an AOD data set from remote sensing was used to infer ground-level $PM_{2.5}$ predictions.

Inferred $PM_{2.5}$ concentrations from monthly $PM_{2.5}$/AOD ratios provide a better correlation than directly inferring from a whole-year AOD data set. Directly inferred $PM_{2.5}$ concentrations from the MAIAC AOD data set demonstrated weak correlations with an $R^2$ of 0.15. Inferred $PM_{2.5}$ concentrations from monthly $PM_{2.5}$/AOD ratios using MAIAC data provided moderate correlation with $R^2$ values of 0.53. The correlations of AOD and $PM_{2.5}$ varied with season, showing better correlations in spring and winter due to a lower cloud fraction and relative humidity.

The low-cost PM sensor data set is strongly correlated with the CN-EPA data set, especially in spring. A good correlation between the low-cost PM sensor data set and the CN-EPA data set with an $R^2$ of 0.82 was obtained. Seasonally, summer showed the lowest $R^2$, 0.69, and demonstrated more fluctuations due to meteorological factors (wind speed, temperature, relative humidity, and air mass) and specific human activity. The higher RMSEs were found in fall and winter which may be due to contribution of localized sources such as coal stoves used for heating.

The cross-validation of seven data sets was performed to evaluate the predictions of data sets. The AOD data set showed the lowest values of RMSEs from the LOOCV method, but after combining the AOD data sets with CN-EPA data sets both the RMSEs and mean absolute errors increased. This indicated that the AOD data set might not reflect the ground-level $PM_{2.5}$ concentrations in this study. On the other hand, the RMSEs and mean absolute errors of low-cost PM sensor data sets improved after combining with CN-EPA data sets. The performance of low-cost PM sensor and AOD data sets in rural and urban areas showed that AOD data sets had higher RMSEs, NRMSEs, and mean absolute errors than low-cost PM sensor data sets. To reduce the effect of the data number of different data sets, the RMSEs, the NRMSEs, and the mean absolute errors were calculated by comparing the predictions and the measurements at the location of CN-EPA data sets. The AOD data set showed higher RMSEs, NRMSEs, and mean absolute errors than other data set combinations. Thus, we decided to use the CN-EPA and low-cost PM sensor data sets to visualize the pollution maps.

Finally, the CN-EPA combined with low-cost PM sensor data sets were better than the CN-EPA data set in identifying local hotspots in downtown and rural Xinxiang during four events that generate serious haze: Chinese New Year, Tomb Sweeping Day, Ghost Festival, and National Day. Pollution maps were created via IDW interpolation, and the cross-validations were used to evaluate their accuracy. The CN-EPA data set showed lower RMSEs than the low-cost PM sensor data set and the combined $PM_{2.5}$ data sets. The results demonstrated the Chinese New Year Day had a higher difference to the other three events.

## AUTHOR INFORMATION


### Corresponding Author

**Pratim Biswas** − *Aerosol and Air Quality Research Laboratory, Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States; College of Engineering, University of Miami, Coral Gables, Florida 33146, United States;* orcid.org/0000-0003-1104-3738; Email: pbiswas@miami.edu

### Authors

**Chun-Ying Chao** − *Aerosol and Air Quality Research Laboratory, Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* Present Address: Department of Civil and Environmental Engineering, Rice University, Houston, Texas 77005, United States

**Huang Zhang** − *Aerosol and Air Quality Research Laboratory, Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* Present Address: Institute of Nuclear and New Energy Technology, Key Laboratory of Advanced Reactor Engineering and Safety, Tsinghua University, Beijing 100084, China

**Melanie Hammer** − *Atmospheric Composition Analysis Group, Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* orcid.org/0000-0001-8443-7979

**Yu Zhan** − *Department of Environmental Science and Engineering, Sichuan University, Chengdu 610065, China;* orcid.org/0000-0002-8473-2799

**David Kenney** − *Aerosol and Air Quality Research Laboratory, Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

**Randall V. Martin** − *Atmospheric Composition Analysis Group, Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsearthspacechem.1c00174


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The work was partially supported by the National Science Foundation (Award Number 2020673) as part of CAMS-Net. Additional support from MAGEEP and the Lopata Endowment is gratefully acknowledged.

## REFERENCES

(1) An, Z.; Huang, R. J.; Zhang, R.; Tie, X.; Li, G.; Cao, J.; Zhou, W.; Shi, Z.; Han, Y.; Gu, Z.; Ji, Y. Severe haze in northern China: A synergy of anthropogenic emissions and atmospheric processes. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (18), 8657−8666.

(2) Bai, L.; He, Z.; Ni, S.; Chen, W.; Li, N.; Sun, S. Investigation of PM2.5 absorbed with heavy metal elements, source apportionment and their health impacts in residential houses in the North-east region of China. *Sustainable Cities and Society* **2019**, *51*, 101690.
(3) Liu, L.; Liu, Y.; Wen, W.; Liang, L.; Ma, X.; Jiao, J.; Guo, K. Source Identification of Trace Elements in PM2.5 at a Rural Site in the North China Plain. *Atmosphere* **2020**, *11* (2), 179.
(4) Yang, K.; Li, Q.; Yuan, M.; Guo, M.; Wang, Y.; Li, S.; Tian, C.; Tang, J.; Sun, J.; Li, J.; Zhang, G. Temporal variations and potential sources of organophosphate esters in PM2.5 in Xinxiang, North China. *Chemosphere* **2019**, *215*, 500−506.
(5) Ye, C.; Chen, R.; Chen, M. The impacts of Chinese Nian culture on air pollution. *J. Cleaner Prod.* **2016**, *112*, 1740−1745.
(6) World Health Organization. https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health (accessed 10/15/2020).
(7) Cao, J.; Chow, J. C.; Lee, F. S. C.; Watson, J. G. Evolution of PM2.5 Measurements and Standards in the U.S. and Future Perspectives for China. *Aerosol Air Qual. Res.* **2013**, *13* (4), 1197−1211.
(8) Lin, Y.; Zou, J.; Yang, W.; Li, C. Q. A Review of Recent Advances in Research on PM2.5 in China. *Int. J. Environ. Res. Public Health* **2018**, *15* (3), 438.
(9) Bai, Z.; Han, J.; Azzi, M. Insights into measurements of ambient air PM 2.5 in China. *Trends Environ. Anal. Chem.* **2017**, *13*, 1−9.
(10) Chen, X.; Zhang, L. W.; Huang, J. J.; Song, F. J.; Zhang, L. P.; Qian, Z. M.; Trevathan, E.; Mao, H. J.; Han, B.; Vaughn, M.; Chen, K. X.; Liu, Y. M.; Chen, J.; Zhao, B. X.; Jiang, G. H.; Gu, Q.; Bai, Z. P.; Dong, G. H.; Tang, N. J. Long-term exposure to urban air pollution and lung cancer mortality: A 12-year cohort study in Northern China. *Sci. Total Environ.* **2016**, *571*, 855−61.
(11) Cheng, Y.; Zheng, G.; Wei, C.; Mu, Q.; Zheng, B.; Wang, Z.; Gao, M.; Zhang, Q.; He, K.; Carmichael, G.; Poschl, U.; Su, H. Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China. *Science Advances* **2016**, *2* (12), No. e1601530.
(12) Feng, J.; Sun, P.; Hu, X.; Zhao, W.; Wu, M.; Fu, J. The chemical composition and sources of PM2.5 during the 2009 Chinese New Year's holiday in Shanghai. *Atmos. Res.* **2012**, *118*, 435−444.
(13) Huang, T.; Yu, Y.; Wei, Y.; Wang, H.; Huang, W.; Chen, X. Spatial-seasonal characteristics and critical impact factors of PM2.5 concentration in the Beijing-Tianjin-Hebei urban agglomeration. *PLoS One* **2018**, *13* (9), No. e0201364.
(14) Jiao, W.; Hagler, G.; Williams, R.; Sharpe, R.; Brown, R.; Garver, D.; Judge, R.; Caudill, M.; Rickard, J.; Davis, M.; Weinstock, L.; Zimmer-Dauphinee, S.; Buckley, K. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-costsensor performance in a suburban environment in the southeastern UnitedStates. *Atmos. Meas. Tech.* **2016**, *9* (11), 5281−5292.
(15) Li, J.; Mattewal, S. K.; Patel, S.; Biswas, P. Evaluation of Nine Low-cost-sensor-based Particulate Matter Monitors. *Aerosol Air Qual. Res.* **2020**, *20* (2), 254−270.
(16) Xu, L.; Chen, F.; Zhong, X.; Zhang, L. e.; Ye, R.; Cai, W.; Rao, Q.; Li, J. Spatial disequilibrium of fine particulate matter and corresponding health burden in China. *J. Cleaner Prod.* **2019**, *238*, 117840.
(17) Carvalho, H. The air we breathe: differentials in global air quality monitoring. *Lancet Respir. Med.* **2016**, *4* (8), 603−605.
(18) Apte, J. S.; Messier, K. P.; Gani, S.; Brauer, M.; Kirchstetter, T. W.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Vermeulen, R. C.; Hamburg, S. P. High-resolution air pollution mapping with Google street view cars: exploiting big data. *Environ. Sci. Technol.* **2017**, *51* (12), 6999−7008.
(19) Li, H. Z.; Gu, P.; Ye, Q.; Zimmerman, N.; Robinson, E. S.; Subramanian, R.; Apte, J. S.; Robinson, A. L.; Presto, A. A. Spatially dense air pollutant sampling: Implications of spatial variability on the representativeness of stationary air pollutant monitors. *Atmospheric Environment: X* **2019**, *2*, 100012.
(20) van Donkelaar, A.; Martin, R. V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P. J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ. Health Perspect.* **2010**, *118* (6), 847−55.
(21) Gupta, P.; Doraiswamy, P.; Levy, R.; Pikelnaya, O.; Maibach, J.; Feenstra, B.; Polidori, A.; Kiros, F.; Mills, K. C. Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations. *Geohealth* **2018**, *2* (6), 172−181.
(22) Garay, M. J.; Kalashnikova, O. V.; Bull, M. A. Development and assessment of a higher-spatial-resolution (4.4 km) MISR aerosol optical depth product using AERONET-DRAGON data. *Atmos. Chem. Phys.* **2017**, *17* (8), 5095−5106.
(23) Hsu, N. C.; Jeong, M.-J.; Bettenhausen, C.; Sayer, A. M.; Hansell, R.; Seftor, C. S.; Huang, J.; Tsay, S.-C. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres* **2013**, *118* (16), 9296−9315.
(24) Levy, R. C.; Mattoo, S.; Munchak, L. A.; Remer, L. A.; Sayer, A. M.; Patadia, F.; Hsu, N. C. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* **2013**, *6* (11), 2989−3034.
(25) Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. MODIS Collection 6 MAIAC algorithm. *Atmos. Meas. Tech.* **2018**, *11* (10), 5741−5765.
(26) Eck, T. F.; Holben, B.; Reid, J.; Dubovik, O.; Smirnov, A.; O'neill, N.; Slutsker, I.; Kinne, S. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *Journal of Geophysical Research: Atmospheres* **1999**, *104* (D24), 31333−31349.
(27) Gupta, P.; Khan, M. N.; da Silva, A.; Patadia, F. MODIS aerosol optical depth observations over urban areas in Pakistan: quantity and quality of the data for air quality monitoring. *Atmos. Pollut. Res.* **2013**, *4* (1), 43−52.
(28) Li, T.; Shen, H.; Zeng, C.; Yuan, Q.; Zhang, L. Point-surface fusion of station measurements and satellite observations for mapping PM2.5 distribution in China: Methods and assessment. *Atmos. Environ.* **2017**, *152*, 477−489.
(29) van der Werf, G. R.; Randerson, J. T.; Giglio, L.; Collatz, G. J.; Mu, M.; Kasibhatla, P. S.; Morton, D. C.; DeFries, R. S.; Jin, Y.; van Leeuwen, T. T. Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997−2009). *Atmos. Chem. Phys.* **2010**, *10* (23), 11707−11735.
(30) Hu, X.; Waller, L. A.; Lyapustin, A.; Wang, Y.; Al-Hamdan, M. Z.; Crosson, W. L.; Estes, M. G.; Estes, S. M.; Quattrochi, D. A.; Puttaswamy, S. J.; Liu, Y. Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment* **2014**, *140*, 220−232.
(31) Yang, Q.; Yuan, Q.; Yue, L.; Li, T.; Shen, H.; Zhang, L. The relationships between PM2.5 and aerosol optical depth (AOD) in mainland China: About and behind the spatio-temporal variations. *Environ. Pollut.* **2019**, *248*, 526−535.
(32) Hammer, M. S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A. M.; Hsu, N. C.; Levy, R. C.; Garay, M. J.; Kalashnikova, O. V.; Kahn, R. A.; Brauer, M.; Apte, J. S.; Henze, D. K.; Zhang, L.; Zhang, Q.; Ford, B.; Pierce, J. R.; Martin, R. V. Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998−2018). *Environ. Sci. Technol.* **2020**, *54* (13), 7879−7890.
(33) van Donkelaar, A.; Martin, R. V.; Brauer, M.; Hsu, N. C.; Kahn, R. A.; Levy, R. C.; Lyapustin, A.; Sayer, A. M.; Winker, D. M. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. *Environ. Sci. Technol.* **2016**, *50* (7), 3762−72.
(34) Johnson, K. K.; Bergin, M. H.; Russell, A. G.; Hagler, G. S. W. Field Test of Several Low-Cost Particulate Matter Sensors in High and Low Concentration Urban Environments. *Aerosol Air Qual. Res.* **2018**, *18* (3), 565−578.
(35) Lin, Y. C.; Chi, W. J.; Lin, Y. Q. The improvement of spatial-temporal resolution of PM2.5 estimation based on micro-air quality

sensors by using data fusion technique. *Environ. Int.* **2020**, *134*, 105305.
(36) Sayahi, T.; Butterfield, A.; Kelly, K. E. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* **2019**, *245*, 932−940.
(37) Li, J.; Zhang, H.; Chao, C.-Y.; Chien, C.-H.; Wu, C.-Y.; Luo, C. H.; Chen, L.-J.; Biswas, P. Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM2.5. *Atmos. Environ.* **2020**, *223*, 117293.
(38) Lung, S. C.; Wang, W. V.; Wen, T. J.; Liu, C. H.; Hu, S. C. A versatile low-cost sensing device for assessing PM2.5 spatiotemporal variation and quantifying source contribution. *Sci. Total Environ.* **2020**, *716*, 137145.
(39) Tagle, M.; Rojas, F.; Reyes, F.; Vasquez, Y.; Hallgren, F.; Linden, J.; Kolev, D.; Watne, A. K.; Oyola, P. Field performance of a low-cost sensor in the monitoring of particulate matter in Santiago, Chile. *Environ. Monit. Assess.* **2020**, *192* (3), 171.
(40) Shi, X.; Zhao, C.; Jiang, J. H.; Wang, C.; Yang, X.; Yung, Y. L. Spatial representativeness of PM2. 5 concentrations obtained using observations from network stations. *J. Geophys. Res.: Atmos.* **2018**, *123* (6), 3145−3158.
(41) Clements, A. L.; Griswold, W. G.; Rs, A.; Johnston, J. E.; Herting, M. M.; Thorson, J.; Collier-Oxandale, A.; Hannigan, M. Low-cost air quality monitoring tools: from research to practice (a workshop summary). *Sensors* **2017**, *17* (11), 2478.
(42) Kong, L.; Xin, J.; Zhang, W.; Wang, Y. The empirical correlations between PM2.5, PM10 and AOD in the Beijing metropolitan region and the PM2.5, PM10 distributions retrieved by MODIS. *Environ. Pollut.* **2016**, *216*, 350−360.
(43) Sorek-Hamer, M.; Strawa, A. W.; Chatfield, R. B.; Esswein, R.; Cohen, A.; Broday, D. M. Improved retrieval of PM2.5 from satellite data products using non-linear methods. *Environ. Pollut.* **2013**, *182*, 417−23.
(44) Wu, X.; Wang, Y.; He, S.; Wu, Z. $PM_{2.5}/PM_{10}$ ratio prediction based on a long short-term memory neural network in Wuhan, China. *Geosci. Model Dev.* **2020**, *13* (3), 1499−1511.
(45) Chung, A.; Chang, D. P.; Kleeman, M. J.; Perry, K. D.; Cahill, T. A.; Dutcher, D.; McDougall, E. M.; Stroud, K. Comparison of real-time instruments used to monitor airborne particulate matter. *J. Air Waste Manage. Assoc.* **2001**, *51* (1), 109−20.
(46) Barkjohn, K. K.; Gantt, B.; Clements, A. L. Development and application of a United States-wide correction for PM 2.5 data collected with the PurpleAir sensor. *Atmos. Meas. Tech.* **2021**, *14* (6), 4617−4637.
(47) Wallace, L.; Bi, J.; Ott, W. R.; Sarnat, J.; Liu, Y. Calibration of low-cost PurpleAir outdoor monitors using an improved method of calculating PM2. 5. *Atmos. Environ.* **2021**, *256*, 118432.
(48) Zheng, T.; Bergin, M.; Wang, G.; Carlson, D. Local PM2. 5 Hotspot Detector at 300 m Resolution: A Random Forest−Convolutional Neural Network Joint Model Jointly Trained on Satellite Images and Meteorology. *Remote Sensing* **2021**, *13* (7), 1356.
(49) Gupta, P.; Remer, L. A.; Levy, R. C.; Mattoo, S. Validation of MODIS 3 km land aerosol optical depth from NASA's EOS Terra and Aqua missions. *Atmos. Meas. Tech.* **2018**, *11* (5), 3145−3159.
(50) Cherian, R.; Quaas, J. Trends in AOD, clouds, and cloud radiative effects in satellite data and CMIP5 and CMIP6 model simulations over aerosol source regions. *Geophys. Res. Lett.* **2020**, *47* (9), No. e2020GL087132.
(51) Li, L.; Franklin, M.; Girguis, M.; Lurmann, F.; Wu, J.; Pavlovic, N.; Breton, C.; Gilliland, F.; Habre, R. Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling. *Remote sensing of environment* **2020**, *237*, 111584.
(52) Chen, L.; Shi, M.; Gao, S.; Li, S.; Mao, J.; Zhang, H.; Sun, Y.; Bai, Z.; Wang, Z. Assessment of population exposure to PM2.5 for mortality in China and its public health benefit based on BenMAP. *Environ. Pollut.* **2017**, *221*, 311−317.
(53) Wang, L.; Zhang, N.; Liu, Z.; Sun, Y.; Ji, D.; Wang, Y. The Influence of Climate Factors, Meteorological Conditions, and Boundary-Layer Structure on Severe Haze Pollution in the Beijing-Tianjin-Hebei Region during January 2013. *Advances in Meteorology* **2014**, *2014*, 685971.
(54) Zhang, Q.; Ma, Q.; Zhao, B.; Liu, X.; Wang, Y.; Jia, B.; Zhang, X. Winter haze over North China Plain from 2009 to 2016: Influence of emission and meteorology. *Environ. Pollut.* **2018**, *242*, 1308−1318.
(55) Wang, Y.; Wang, Q.; Ye, J.; Li, L.; Zhou, J.; Ran, W.; Zhang, R.; Wu, Y.; Cao, J. Chemical composition and sources of submicron aerosols in winter at a regional site in Beijing-Tianjin-Hebei region: Implications for the Joint Action Plan. *Sci. Total Environ.* **2020**, *719*, 137547.
(56) Chafe, Z.; Brauer, M.; Héroux, M.-E.; Klimont, Z.; Lanki, T.; Salonen, R. O.; Smith, K. R. Residential heating with wood and coal: health impacts and policy options in Europe and North America. *UNECE*; 2015; https://unece.org/environment-policy/publications/residential-heating-wood-and-coal-health-impacts-and-policy-options (accessed 09/15/2020).
(57) Martin, R. V. Satellite remote sensing of surface air quality. *Atmos. Environ.* **2008**, *42* (34), 7823−7843.
(58) Green, M.; Kondragunta, S.; Ciren, P.; Xu, C. Comparison of GOES and MODIS aerosol optical depth (AOD) to aerosol robotic network (AERONET) AOD and IMPROVE PM2.5 mass at Bondville, Illinois. *J. Air Waste Manage. Assoc.* **2009**, *59* (9), 1082−91.
(59) Takemura, T.; Okamoto, H.; Maruyama, Y.; Numaguti, A.; Higurashi, A.; Nakajima, T. Global three-dimensional simulation of aerosol optical thickness distribution of various origins. *Journal of Geophysical Research: Atmospheres* **2000**, *105* (D14), 17853−17873.
(60) Guo, J.; Xia, F.; Zhang, Y.; Liu, H.; Li, J.; Lou, M.; He, J.; Yan, Y.; Wang, F.; Min, M.; Zhai, P. Impact of diurnal variability and meteorological factors on the PM2.5 - AOD relationship: Implications for PM2.5 remote sensing. *Environ. Pollut.* **2017**, *221*, 94−104.
(61) Wang, S.-H.; Hung, W.-T.; Chang, S.-C.; Yen, M.-C. Transport characteristics of Chinese haze over Northern Taiwan in winter, 2005−2014. *Atmos. Environ.* **2016**, *126*, 76−86.
(62) Zhang, Y.; Gu, K. AQPDBJUT Dataset: Picture-Based PM Monitoring in the Campus of BJUT. *arXiv preprint arXiv:2003.08609* **2020**, https://arxiv.org/abs/2003.08609 (accessed 08/17/2021).
(63) Wang, Y.; Li, J.; Jing, H.; Zhang, Q.; Jiang, J.; Biswas, P. Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement. *Aerosol Sci. Technol.* **2015**, *49* (11), 1063−1077.
(64) Zhang, J. J.; Smith, K. R. Household air pollution from coal and biomass fuels in China: measurements, health impacts, and interventions. *Environ. Health Perspect.* **2007**, *115* (6), 848−55.


https://doi.org/10.1021/acsearthspacechem.1c00174
*ACS Earth Space Chem.* 2021, 5, 3022−3035