AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL075710

Key Points:
- A deep learning architecture is established to estimate ground-level $PM_{2.5}$ by fusing satellite and station observations
- A geo-intelligent model is developed to incorporate geographical correlation into deep learning for performance improvement
- This model shows a superior estimation accuracy (R^2 = 0.88, RMSE = 13.03 μg/m^3) at national scale

Supporting Information:
- Supporting Information S1

Correspondence to:
H. Shen,
shenhf@whu.edu.cn

Citation:
Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating ground-level $PM_{2.5}$ by fusing satellite and station observations: A geo-intelligent deep learning approach. *Geophysical Research Letters*, *44*. https://doi.org/10.1002/2017GL075710

Received 17 SEP 2017
Accepted 19 NOV 2017
Accepted article online 22 NOV 2017

Estimating Ground-Level $PM_{2.5}$ by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach

Tongwen Li[1], Huanfeng Shen[1,2,3], Qiangqiang Yuan[2,4], Xuechen Zhang[1], and Liangpei Zhang[2,5]

[1]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China, [2]The Collaborative Innovation Center for Geospatial Technology, Wuhan, China, [3]The Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, Wuhan, China, [4]School of Geodesy and Geomatics, Wuhan University, Wuhan, China, [5]The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

Abstract Fusing satellite observations and station measurements to estimate ground-level $PM_{2.5}$ is promising for monitoring $PM_{2.5}$ pollution. A geo-intelligent approach, which incorporates geographical correlation into an intelligent deep learning architecture, is developed to estimate $PM_{2.5}$. Specifically, it considers geographical distance and spatiotemporally correlated $PM_{2.5}$ in a deep belief network (denoted as Geoi-DBN). Geoi-DBN can capture the essential features associated with $PM_{2.5}$ from latent factors. It was trained and tested with data from China in 2015. The results show that Geoi-DBN performs significantly better than the traditional neural network. The out-of-sample cross-validation R^2 increases from 0.42 to 0.88, and RMSE decreases from 29.96 to 13.03 μg/m^3. On the basis of the derived $PM_{2.5}$ distribution, it is predicted that over 80% of the Chinese population live in areas with an annual mean $PM_{2.5}$ of greater than 35 μg/m^3. This study provides a new perspective for air pollution monitoring in large geographic regions.

1. Introduction

$PM_{2.5}$ (particulate matter), with an aerodynamic diameter of less than 2.5 μm, is associated with many adverse health effects, such as respiratory problems and cardiovascular disease (Bartell et al., 2013; Brauer et al., 2012; Crouse et al., 2012). Previous studies (Chen et al., 2013) have shown that a 3 year reduction in average life expectancy and a 14% increase in overall mortality would result from a 100 μg/m^3 increase in the concentration of respirable particulate matter. As a result, $PM_{2.5}$ pollution has attracted widespread attention in recent years and has become the focal point of international air pollution research (Engel-Cox et al., 2013).

With the launch of satellites and the continuous improvements in data retrieval technology, estimating ground-level $PM_{2.5}$ using satellite remote sensing has become a promising approach for the monitoring of $PM_{2.5}$ pollution (Engel-Cox, Hoff, et al., 2004; Hoff & Christopher, 2009; Liu, Chen, et al., 2009; Martin, 2008). Three main kinds of methods have been applied to estimate $PM_{2.5}$ using satellite-derived aerosol optical depth (AOD): chemical simulation models (Liu et al., 2004; van Donkelaar et al., 2010), statistical models (Liu, Paciorek, & et al., 2009; Song et al., 2014), and semiempirical models (Lin et al., 2015). Among them, the statistical models are much easier to implement and can obtain a competitive accuracy in $PM_{2.5}$ estimation (Liu, 2014). As a result, many different statistical models have been developed to explore the quantitative relationship between satellite-derived AOD and ground-measured $PM_{2.5}$ (the AOD-$PM_{2.5}$ relationship) (Liu et al., 2007, 2008). For example, the linear regression model establishes a simple linear relationship between AOD and $PM_{2.5}$ (Zhang et al., 2009). Considering more meteorological parameters, the multiple linear regression model was developed by Gupta and Christopher (2009b). To account for the spatial heterogeneity of the AOD-$PM_{2.5}$ relationship, a geographically weighted regression model was introduced (Hu et al., 2013). Moreover, some more complex mixed-effect models (Lee et al., 2011) and generalized additive mixed models (Kloog et al., 2011; Liu et al., 2012) have also been developed to estimate ground-level $PM_{2.5}$. All these statistical models are used to represent the relationship between $PM_{2.5}$ and the latent factors.

However, the levels of $PM_{2.5}$ concentration are related to many factors (Hystad et al., 2011; Paciorek et al., 2008; van Donkelaar et al., 2006; Weber et al., 2010), such as meteorological conditions (e.g., temperature, wind speed, and relative humidity), land use type, population, road networks, and so on. This situation has increased the difficulty of using the traditional statistical models to estimate $PM_{2.5}$. Unlike the traditional methods, nonlinear and nonparametric machine learning algorithms may have the potential to address

this problem (Lary et al., 2014). For example, Gupta and Christopher (2009a) used a back-propagation neural network (BPNN) to estimate surface-level $PM_{2.5}$ in the southeastern United States; an artificial neural network algorithm was trained with Bayesian regularization to estimate $PM_{2.5}$ in eastern China (Wu et al., 2012), and a generalized regression neural network (GRNN) was reported to outperform the traditional models at national scale in China (Li et al., 2017). These neural network models show great advantages in estimating ground-level $PM_{2.5}$.

Deep learning, which is considered to be the second generation of neural network, may be a potential way to address this situation (Hinton & Salakhutdinov, 2006). However, to date, deep learning has seldom been applied in the estimation of ground-level $PM_{2.5}$; only a few attempts (Ong et al., 2015) have been made to predict time series $PM_{2.5}$ over monitoring stations. On the other hand, intelligent algorithms are commonly used to describe numerical relationships, but they neglect the geographical correlation of environmental variables. Meanwhile, it has been reported that $PM_{2.5}$ concentrations show significant autocorrelation in time and space (Wu et al., 2015). The nearby $PM_{2.5}$ from neighboring stations and the $PM_{2.5}$ observations from nearby days for the same station are informative for estimating $PM_{2.5}$. It is therefore important to incorporate this geographical correlation relationship into the intelligent algorithms.

Consequently, the objective of this study is to develop a geo-intelligent deep learning model to estimate ground-level $PM_{2.5}$. This model is established and evaluated based on satellite observations, meteorological parameters, and ground-level $PM_{2.5}$ measurements from China, which is suffering from serious $PM_{2.5}$ pollution (Peng et al., 2016). This study will provide a new perspective to investigate the spatiotemporal characteristics of air pollution in a large geographic region.

2. Materials and Methods

2.1. Study Region and Data

The study region is China (see Figure S1 in the supporting information). The study period is from 1 January 2015 to 31 December 2015.

The data include five main parts. (1) Ground-level $PM_{2.5}$: hourly $PM_{2.5}$ for 2015 was obtained from the China National Environmental Monitoring Center (CNEMC) Web site (http://www.cnemc.cn). The number of stations was ~1,500 by the end of 2015. We averaged hourly $PM_{2.5}$ to daily mean $PM_{2.5}$ for the estimation of $PM_{2.5}$. (2) Moderate Resolution Imaging Spectroradiometer (MODIS) AOD: both Terra and Aqua MODIS AOD products were downloaded from the Level 1 and Atmosphere Archive and Distribution System (LAADS, https://ladsweb.modaps.eosdis.nasa.gov/). We adopted Collection 6 10 km AOD products, which are retrieved by combining dark target and deep blue algorithms (Levy et al., 2013). The average of the Terra and Aqua AOD products was employed to estimate daily average $PM_{2.5}$. (3) Meteorological parameters: we extracted relative humidity (RH, %), air temperature at a 2 m height (TMP, K), wind speed at 10 m above ground (WS, m/s), surface pressure (PS, kPa), and planetary boundary layer height (PBL, m) from MERRA-2 meteorological reanalysis data, which were downloaded from the NASA Web site (http://gmao.gsfc.nasa.gov/GMAO_products/). (4) MODIS normalized difference vegetation index (NDVI): MODIS NDVI products (MOD13) were downloaded from the LAADS Web site. (5) Population and road data: these data were obtained from the Socioeconomic Data and Applications Center (SEDAC, http://sedac.ciesin.columbia.edu/data/collection/gpw-v4) and the National Geomatics Center of China (NGCC, http://ngcc.sbsm.gov.cn/), respectively.

Additional information and integration of the input data are provided in Text S1 in the supporting information.

2.2. The Deep Belief Network (DBN) Model for the Estimation of $PM_{2.5}$

The deep belief network (DBN) model, which is one of the most typical deep learning models, was introduced in 2006 (Hinton et al., 2006). A DBN consists of multiple restricted Boltzmann machine (RBM) layers and a back-propagation (BP) layer and can be used for classification or prediction problems. For example, the structure of a DBN with two RBM layers is shown in Figure 1a.

An RBM consists of a visible layer and a hidden layer, where the hidden layer of the prior RBM is the visible layer of the next RBM. Taking the first RBM as an example, from the visible layer ($\mathbf{v}$) to the hidden layer ($\mathbf{h}^1$),

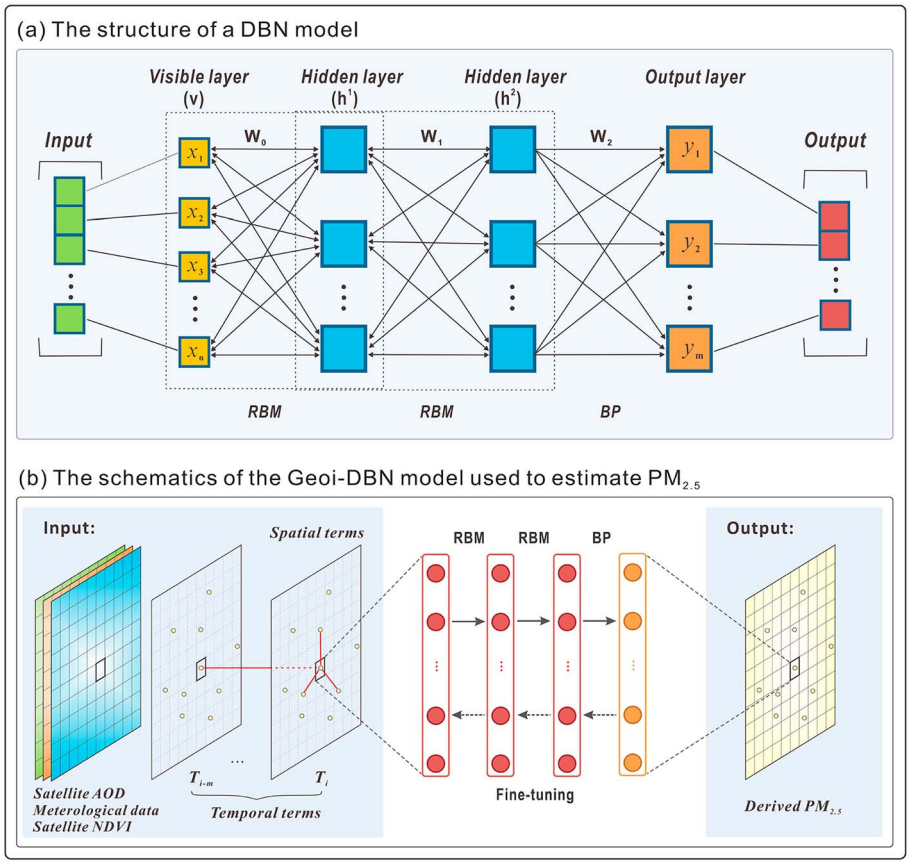

Figure 1. The structure of a DBN and the specific schematics (Geoi-DBN) used to estimate $PM_{2.5}$.

$$h_i^1 = \begin{cases} 1, f\left(\mathbf{W_{0,i}}\mathbf{x} + b_i\right) \geq \mu \\ 0, f\left(\mathbf{W_{0,i}}\mathbf{x} + b_i\right) < \mu \end{cases} \qquad \mu \sim U(0,\ 1) \tag{1}$$

where i refers to the number of the ith neuron and b_i denotes the bias for neuron i. $f(\cdot)$ indicates the transfer function, $f(x) = \frac{1}{1+e^{-x}}$. It is the same for the calculation of the visible layer from the hidden layer. The contrastive divergence algorithm is usually used for training an RBM (Hinton & Salakhutdinov, 2006). The weights are updated in the nth iteration as

$$\mathbf{W_0^{n+1}} = \mathbf{W_0^n} + \varepsilon \cdot \left(\left(\mathbf{h_1^1}\right)^T \mathbf{x} - \left(\mathbf{h_2^1}\right)^T \mathbf{v_1} \right) \tag{2}$$

where ε is the learning rate, $\mathbf{v_1}$ denotes the reconstruction from hidden layer ($\mathbf{h_1^1}$), and $\mathbf{h_1^1}, \mathbf{h_2^1}$ are generated from $\mathbf{x}$, $\mathbf{v_1}$ using equation (1), respectively. The RBMs are pretrained one by one, without supervision, and the trained weights are used to initialize the multilayer neural network. The DBN model then works as a feed-forward neural network (Yue et al., 2017), whereas the error reduction using the BP algorithm is referred to here as "fine-tuning."

Specifically, in our case, the schematics of the geo-intelligent DBN (Geoi-DBN) used to estimate ground-level $PM_{2.5}$ are presented in Figure 1b. The input variables are the satellite-derived AOD, meteorological parameters, NDVI, and spatiotemporally informative terms. Because of the autocorrelation, the nearby n grids of $PM_{2.5}$ measurements and the $PM_{2.5}$ observations from the m prior days for the same grid are informative for estimating $PM_{2.5}$. The nearer observations are more informative than further ones (Tobler, 1970; Yuan et al., 2012). For a specific grid, the spatiotemporally informative terms are represented as

$$S\text{-}PM_{2.5} = \frac{\sum_{i=1}^{n} ws_i PM_{2.5,i}}{\sum_{i=1}^{n} ws_i} \qquad ws_i = \frac{1}{ds_i^2} \tag{3}$$

$$T\text{-}PM_{2.5} = \frac{\sum_{j=1}^{m} wt_j PM_{2.5,j}}{\sum_{j=1}^{m} wt_j} \qquad wt_j = \frac{1}{dt_j^2} \tag{4}$$

$$DIS = \min\left(\frac{1}{ds_i}\right) \quad i = 1, 2, \ldots, n \tag{5}$$

where *ds*, *dt* refer to the spatial and temporal distances, respectively. *m*, *n* are 3 and 10, respectively (Table S1 in the supporting information). The geographical distance (DIS) is used to reflect the heterogeneity of uneven station distribution. Two hidden layers (two RBMs) are then used, and the number of neurons in each hidden layer is 15 (Text S2). The RBM layers are stacked one by one to transfer the input signals to the higher layer. The output layer is a BP layer, which has only one node ($PM_{2.5}$ measurements).

The relationship $PM_{2.5} = f$(AOD, RH, WS, TMP, PBL, PS, NDVI, *S* - $PM_{2.5}$, *T* - $PM_{2.5}$, DIS) is wished to learn from the data records. The process can be divided into three steps:

1. Pretraining. Using the collected data records, the RBMs are trained layer by layer, without supervision. This unsupervised training can extract the essential features associated with $PM_{2.5}$, and they are transferred from the prior RBM to the next RBM layer. Therefore, the higher layer can extract the deeper features related to $PM_{2.5}$.
2. Fine-tuning. Through the prior pretraining step, the initial weights of Geoi-DBN are generated and we can obtain the calculated $PM_{2.5}$. Compared with in situ $PM_{2.5}$ measurements, an estimation error can be obtained, and it is sent back to the Geoi-DBN model to fine-tune the weight coefficients using the BP algorithm.
3. Prediction. This step evaluates the performance of the Geoi-DBN model established on the input data records and predicts the $PM_{2.5}$ values for those locations with no ground stations. Thus, spatially continuous $PM_{2.5}$ data can be reconstructed.

To evaluate the model performance, a sample-based 10-fold cross-validation technique (Rodriguez et al., 2010) was applied to test the model overfitting and predictive power. Furthermore, spatial hold-out cross-validation strategies, that is, site-based cross-validation and leave one province out cross-validation, were used to assess the spatial performance. Details of the model validation are provided in Text S3. We adopted the statistical indicators (Text S4) of the coefficient of determination (R^2), the root-mean-square error (RMSE, $\mu g/m^3$), the mean prediction error (MPE, $\mu g/m^3$), and the relative prediction error (RPE, %) to evaluate the model performance.

3. Results and Discussion

3.1. Evaluation of the Model Performance

To evaluate the deep learning model, we compared the DBN model with BPNN (Gupta & Christopher, 2009a) and GRNN (Li et al., 2017). BPNN, GRNN, and DBN are all data-driven learning models. However, BPNN has a relatively simple structure with three layers; GRNN uses plenty of neurons in the hidden layer to fully approximate functions, and DBN has a more complicated structure and initializes itself by the unsupervised pretraining. As presented in Table 1, when the geographical correlation is not incorporated into the models (original models), Ori-GRNN obtains the best performance (sample-based and site-based cross-validation R^2 of 0.60 and 0.58, respectively), while Ori-BPNN performs the worst. The Ori-DBN model, which has a more complicated structure, is expected to achieve a better performance; however, it obtains a relatively poor result. A possible reason could be that the AOD-$PM_{2.5}$ relationship is not complicated enough for deep learning. From the original models to the geo-intelligent models, the performance is greatly improved. Among these models, the Geoi-DBN model performs the best, whereas the Geoi-GRNN model performs the worst. This differs from the results of the original models. A possible reason is that the spatiotemporally informative

Table 1
The Cross-Validation Performance of the Models

	Sample-based cross-validation				Site-based cross-validation			
Model	R^2	RMSE	MPE	RPE (%)	R^2	RMSE	MPE	RPE (%)
Ori-BPNN	0.42	29.96	21.10	54.74	0.39	29.71	20.80	53.95
Ori-GRNN	0.60	24.22	16.81	44.26	0.58	24.79	17.22	45.01
Ori-DBN	0.54	25.86	18.10	47.24	0.52	26.67	18.52	48.43
Geoi-BPNN	0.84	15.23	10.34	27.75	0.78	17.89	12.09	32.48
Geoi-GRNN	0.82	16.93	12.34	30.83	0.75	19.43	13.88	35.28
Geoi-DBN	0.88	13.03	8.54	23.73	0.82	16.42	10.71	29.82

terms greatly increase the complexity of the AOD-$PM_{2.5}$ relationship. Benefiting from layer-by-layer pretraining, the more complicated relationship between $PM_{2.5}$ and the predictors is better learned in Geoi-DBN than in Geoi-GRNN. Therefore, the Geoi-DBN model achieves the best performance, with out-of-sample cross-validation R^2 and RMSE values of 0.88 and 13.03 μg/m^3, respectively. Furthermore, the spatial hold-out cross-validation results also show that Geoi-DBN has a satisfactory spatial predictive power (Tables 1 and S3). These results demonstrate that the Geoi-DBN model is a promising approach for the estimation of $PM_{2.5}$.

Figure 2 shows the scatterplots for the cross-validation of the Ori-DBN and Geoi-DBN models. For the Ori-DBN model, the sample-based cross-validation R^2 and RMSE values are 0.54 and 25.86 μg/m^3, respectively. The R^2 and RMSE values of the model fitting are 0.59 and 24.53 μg/m^3 (Table S4), respectively. When considering the geographical correlation, the spatiotemporal characteristics of atmospheric $PM_{2.5}$ are better described in the AOD-$PM_{2.5}$ modeling. The model performance is therefore significantly improved. On the other hand, the out-of-sample cross-validation slope for the Geoi-DBN model is 0.88, with an intercept of 6.39 μg/m^3. These findings indicate that the Geoi-DBN model tends to underestimate when the ground-level $PM_{2.5}$ is greater than ~60 μg/m^3. Therefore, the higher $PM_{2.5}$ concentrations may not be sufficiently explained. This issue is further discussed in section 3.4. However, it should be noted that the cross-validation slope of the Geoi-DBN model (0.88) is much greater than that of the Ori-DBN model (0.55). This means that the Geoi-DBN model shows a much lower extent of underestimation than the Ori-DBN model.

The seasonal and spatial performances of the Geoi-DBN model were also evaluated. The out-of-sample R^2 values are 0.84, 0.82, 0.87, and 0.90 for spring, summer, autumn, and winter, respectively. The spatial

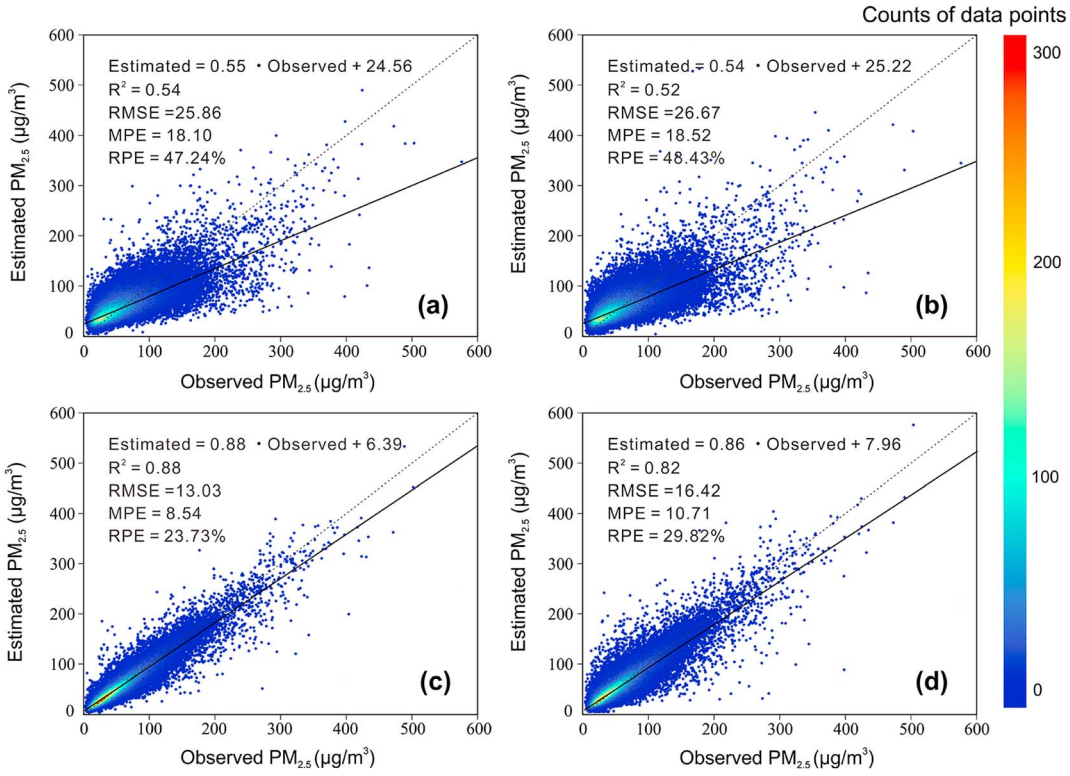

Figure 2. Scatter plots of the cross-validation results: (a and c) sample-based cross-validation results of the Ori-DBN and Geoi-DBN models; (b and d) site-based cross-validation results of the Ori-DBN and Geoi-DBN models.

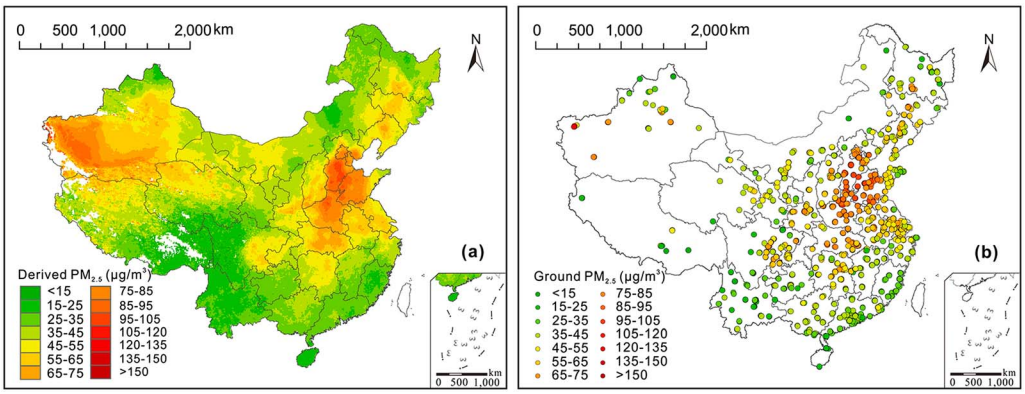

Figure 3. Annual mean distribution of (a) derived $PM_{2.5}$ and (b) ground-measured $PM_{2.5}$ in China.

distribution of R^2 (Figure S3) shows that higher values of R^2 are found over the areas (e.g., East China) with denser monitoring stations.

3.2. Mapping of $PM_{2.5}$ Concentrations

The annual and seasonal mean distributions of $PM_{2.5}$ in China are mapped in Figures 3a and S4, using our previous mapping strategy (Li et al., 2017). The annual distribution has a similar spatial distribution to the ground-measured $PM_{2.5}$, which is presented in Figure 3b. We also compared our results with previous satellite-derived $PM_{2.5}$ data (van Donkelaar et al., 2016). The results are shown in Figure S5. In our results, the levels of $PM_{2.5}$ are higher in the northern regions than in the southern regions. Meanwhile, a heavily polluted region is identified in the North China Plain. As reported in a previous study (Chen et al., 2008), the climate of this region is characterized by stagnant weather, with weak wind and a relatively low boundary layer height, which results in the atmospheric conditions for the accumulation, formation, and processing of aerosols. This is one of the main reasons for the serious $PM_{2.5}$ pollution in this area. Additionally, the $PM_{2.5}$ concentrations are generally higher in the inland regions (e.g., Hunan, Hubei, and Hunan) and lower in the coastal regions (e.g., Guangdong and Fujian). The regions with the least $PM_{2.5}$ pollution are in Hainan and Yunnan provinces, which benefit from the low levels of anthropogenic emissions and favorable meteorological conditions for atmospheric dispersion. Last but not least, a very high level of $PM_{2.5}$ pollution is found in the northwest, especially the Xinjiang Autonomous Region. A possible reason is that the dust particles in this desert region make a significant contribution to the accumulation of $PM_{2.5}$ (Fang et al., 2016).

3.3. Exposure Analysis Over China

The distribution of population in China in 2015 is presented in Figure S6. The World Health Organization (WHO) interim target (IT)-1 and IT-3 for annual mean $PM_{2.5}$ are 35 and 15 $\mu g/m^3$, respectively (WHO, 2006). As shown in Figure 4, the population-weighted estimated annual mean $PM_{2.5}$ is 53.34 $\mu g/m^3$, which greatly exceeds the WHO IT-1 standard. Almost all regions of China (except for the northwest) show population-weighted averages that are greater than the spatial averages. These findings indicate that more people are living in more polluted regions. Figure 4 also shows that over 80% of the Chinese population live in areas that exceed the WHO IT-1 standard. Spatially, South China has the highest percentage of population living in areas meeting the WHO IT-1 standard, whereas Central and North China have the lowest.

3.4. Discussion

$PM_{2.5}$ assessment by fusing satellite and station observations involves lots of different factors, which inherently results in big data. In this situation, a deep learning model may better estimate $PM_{2.5}$. Therefore, we made efforts to incorporate urban big data (i.e., population data and road network) into the deep learning model. The results show that these predictors have almost no positive (or even passive) effect on model performance (Table S5). The possible reason could be that these predictors are not real time and cannot reflect the temporal variation of the AOD-$PM_{2.5}$ relationship. Moreover, they have too coarse spatial resolutions to provide additional information. It is possible that the model performance would be greatly improved if fine-scale and real-time data (e.g., daily street-level traffic flow) could be obtained.

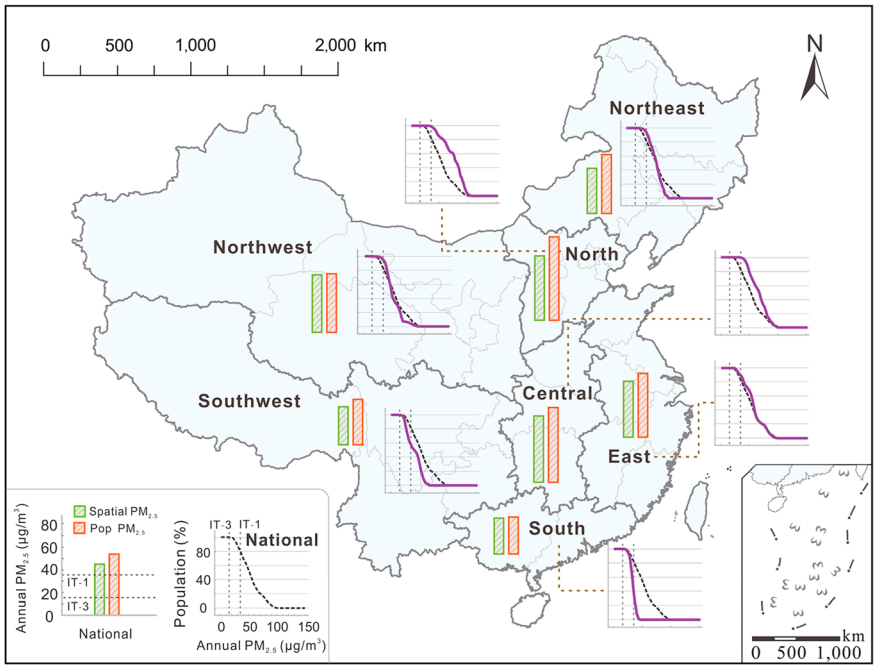

Figure 4. Exposure to $PM_{2.5}$ over China in 2015. Spatial $PM_{2.5}$: spatial mean $PM_{2.5}$. Pop $PM_{2.5}$: population-weighted mean $PM_{2.5}$. The curves represent the percentage of population exposed to $PM_{2.5}$, and the black dashed curve denotes national exposure to $PM_{2.5}$.

The Geoi-DBN cross-validation slope of observed $PM_{2.5}$ versus estimation is 0.88, indicating some evidence of bias. However, it should be noted that the national-scale estimates of $PM_{2.5}$ in China (Fang et al., 2016; Lin et al., 2015; Ma et al., 2016; You et al., 2016) are mostly underestimated (slopes of 0.79–0.83) when ground-level $PM_{2.5}$ is greater than 60 $\mu g/m^3$. This underestimation may be down to several reasons, including the possibility of mixed types and layers of aerosols in the atmosphere (Guo et al., 2016) and the hygroscopicity of urban aerosols (Gupta & Christopher, 2009a). Another possible reason is that we used point-based monitoring data and a spatially averaged modeling framework. The sampling distribution of monitors may not give a great estimation of the spatially averaged concentration for that grid. Therefore, this underestimation is probably a systematic error related to the complicated aerosols in China and the modeling framework.

We summarize the previous studies using machine learning for $PM_{2.5}$ in China in Table S6. Compared with these studies, this study has three main advantages. First, we estimate the spatial $PM_{2.5}$ concentrations using machine learning by incorporating satellite data, while previous study predicted $PM_{2.5}$ over stations (Li et al., 2016). Differing from the traditional machine learning methods (Yao & Lu, 2014), a more advanced deep learning architecture is then utilized for the satellite-based $PM_{2.5}$ estimation. Lastly, while previous studies have used machine learning to simulate the AOD-$PM_{2.5}$ relationship (Wu et al., 2012), this study further considers the geographical correlation to greatly improve model performance.

In this study, a deep learning architecture has been established to estimate ground-level $PM_{2.5}$, achieving a satisfactory performance. However, it should be noted that we applied only one type of deep learning model (i.e., DBN) to model the AOD-$PM_{2.5}$ relationship. Would any other deep learning model work better with this problem? Deep learning has more hidden layers to better represent complex nonlinear relationships, so whether or not we can estimate $PM_{2.5}$ using original satellite reflectance rather than satellite-derived AOD to avoid intermediate error deserves further study. Deep learning is a promising approach for AOD-$PM_{2.5}$ modeling, but there is still room for improvement.

4. Conclusions

Despite the potential application of satellite-based AOD for air quality studies (Engel-Cox, Holloman, et al., 2004; Wang & Christopher, 2003), the estimation of $PM_{2.5}$ involves a large number of factors (Zheng et al.,

2017). We therefore developed a geo-intelligent deep learning model to better represent the AOD-$PM_{2.5}$ relationship. This study introduced the layer-by-layer pretraining technique to the satellite remote sensing assessment of $PM_{2.5}$. In addition, the geographical correlation was adopted to significantly improve the estimation accuracy. The deep learning-based AOD-$PM_{2.5}$ modeling of China accurately estimated $PM_{2.5}$, with out-of-sample cross-validation R^2 and RMSE values of 0.88 and 13.03 $\mu g/m^3$, respectively. It is predicted that over 80% of the Chinese population live in areas with an annual mean $PM_{2.5}$ greater than the WHO IT-1 standard in 2015. Overall, we can say that the proposed approach is promising for air pollution monitoring in large geographical regions.

Acknowledgments
This work was funded by the National Key R&D Program of China (2016YFC0200900) and the National Natural Science Foundation of China (41422108). We are grateful to the CNEMC, the Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC), the NASA Data Center, the NGCC, and the SEDAC for providing the foundational data. The data used are listed in section 2.1 and the supporting information.

References

Bartell, S. M., Longhurst, J., Tjoa, T., Sioutas, C., & Delfino, R. J. (2013). Particulate air pollution, ambulatory heart rate variability, and cardiac arrhythmia in retirement community residents with coronary artery disease. *Environmental Health Perspectives*, *121*(10), 1135–1141. https://doi.org/10.1289/ehp.1205914

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., … Thurston, G. D. (2012). Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environmental Science & Technology*, *46*(2), 652–660. https://doi.org/10.1021/es2025752

Chen, Y., Ebenstein, A., Greenstone, M., & Li, H. (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(32), 12,936–12,941. https://doi.org/10.1073/pnas.1300018110

Chen, Z., Cheng, S., Li, J., Guo, X., Wang, W., & Chen, D. (2008). Relationship between atmospheric pollution processes and synoptic pressure patterns in northern China. *Atmospheric Environment*, *42*(24), 6078–6087. https://doi.org/10.1016/j.atmosenv.2008.03.043

Crouse, D. L., Peters, P. A., van Donkelaar, A., Goldberg, M. S., Villeneuve, P. J., Brion, O., … Burnett, R. T. (2012). Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. *Environmental Health Perspectives*, *120*(5), 708–714. https://doi.org/10.1289/ehp.1104049

Engel-Cox, J. A., Hoff, R. M., & Haymet, A. D. J. (2004). Recommendations on the use of satellite remote-sensing data for urban air quality. *Journal of the Air & Waste Management Association*, *54*(11), 1360–1371. https://doi.org/10.1080/10473289.2004.10471005

Engel-Cox, J. A., Holloman, C. H., Coutant, B. W., & Hoff, R. M. (2004). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment*, *38*(16), 2495–2509. https://doi.org/10.1016/j.atmosenv.2004.01.039

Engel-Cox, J. A., Kim Oanh, N. T., van Donkelaar, A., Martin, R. V., & Zell, E. (2013). Toward the next generation of air quality monitoring: Particulate matter. *Atmospheric Environment*, *80*, 584–590. https://doi.org/10.1016/j.atmosenv.2013.08.016

Fang, X., Zou, B., Liu, X., Sternberg, T., & Zhai, L. (2016). Satellite-based ground $PM_{2.5}$ estimation using timely structure adaptive modeling. *Remote Sensing of Environment*, *186*, 152–163. https://doi.org/10.1016/j.rse.2016.08.027

Guo, J., Miao, Y., Zhang, Y., Liu, H., Li, Z., Zhang, W., … Zhai, P. (2016). The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data. *Atmospheric Chemistry and Physics*, *16*(20), 13,309–13,319. https://doi.org/10.5194/acp-16-13309-2016

Gupta, P., & Christopher, S. A. (2009a). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *Journal of Geophysical Research*, *114*, D20205. https://doi.org/10.1029/2008JD011497

Gupta, P., & Christopher, S. A. (2009b). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research*, *114*, D14205. https://doi.org/10.1029/2008JD011496

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hoff, R. M., & Christopher, S. A. (2009). Remote sensing of particulate pollution from space: Have we reached the promised land? *Journal of the Air & Waste Management Association*, *59*(6), 645–675.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., … Liu, Y. (2013). Estimating ground-level $PM_{2.5}$ concentrations in the southeastern U.S. using geographically weighted regression. *Environmental Research*, *121*, 1–10. https://doi.org/10.1016/j.envres.2012.11.003

Hystad, P., Setton, E., Cervantes, A., Poplawski, K., Deschenes, S., Brauer, M., … Demers, P. (2011). Creating national air pollution models for population exposure assessment in Canada. *Environmental Health Perspectives*, *119*(8), 1123–1129. https://doi.org/10.1289/ehp.1002976

Kloog, I., Koutrakis, P., Coull, B. A., Lee, H. J., & Schwartz, J. (2011). Assessing temporally and spatially resolved $PM_{2.5}$ exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment*, *45*(35), 6267–6275. https://doi.org/10.1016/j.atmosenv.2011.08.066

Lary, D. J., Faruque, F. S., Malakar, N., Moore, A., Roscoe, B., Adams, Z. L., & Eggelston, Y. (2014). Estimating the global abundance of ground level presence of particulate matter ($PM_{2.5}$). *Geospatial Health*, *8*(3), S611–S630. https://doi.org/10.4081/gh.2014.292

Lee, H., Liu, Y., Coull, B., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. *Atmospheric Chemistry and Physics*, *11*(15), 7991–8002. https://doi.org/10.5194/acp-11-7991-2011

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., & Hsu, N. C. (2013). The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, *6*(11), 2989–3034. https://doi.org/10.5194/amt-6-2989-2013

Li, T., Shen, H., Zeng, C., Yuan, Q., & Zhang, L. (2017). Point-surface fusion of station measurements and satellite observations for mapping $PM_{2.5}$ distribution in China: Methods and assessment. *Atmospheric Environment*, *152*, 477–489. https://doi.org/10.1016/j.atmosenv.2017.01.004

Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, *23*(22), 22,408–22,417. https://doi.org/10.1007/s11356-016-7812-9

Lin, C., Li, Y., Yuan, Z., Lau, A. K. H., Li, C., & Fung, J. C. H. (2015). Using satellite remote sensing data to estimate the high-resolution distribution of ground-level $PM_{2.5}$. *Remote Sensing of Environment*, *156*, 117–128. https://doi.org/10.1016/j.rse.2014.09.015

Liu, Y. (2014). Monitoring $PM_{2.5}$ from space for health: Past, present, and future directions. *Environmental Manager*, *6*, 6–10.

Liu, Y., Chen, D., Kahn, R. A., & He, K. (2009). Review of the applications of Multiangle Imaging Spectroradiometer to air quality research. *Science in China Series D: Earth Sciences*, *52*(1), 132–144. https://doi.org/10.1007/s11430-008-0149-6

Liu, Y., Franklin, M., Kahn, R., & Koutrakis, P. (2007). Using aerosol optical thickness to predict ground-level $PM_{2.5}$ concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, *107*(1-2), 33–44. https://doi.org/10.1016/j.rse.2006.05.022

Liu, Y., He, K., Li, S., Wang, Z., Christiani, D. C., & Koutrakis, P. (2012). A statistical model to evaluate the effectiveness of $PM_{2.5}$ emissions control during the Beijing 2008 Olympic Games. *Environment International*, *44*(Supplement C), 100–105.

Liu, Y., Paciorek, C., & Koutrakis, P. (2008). Estimating daily $PM_{2.5}$ exposure in Massachusetts with satellite aerosol remote sensing data, meteorological, and land use information. *Epidemiology*, *19*(6).

Liu, Y., Paciorek, C. J., & Koutrakis, P. (2009). Estimating regional spatial and temporal variability of PM(2.5) concentrations using satellite data, meteorology, and land use information. *Environmental Health Perspectives*, *117*(6), 886–892. https://doi.org/10.1289/ehp.0800123

Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V., & Sarnat, J. A. (2004). Mapping annual mean ground-level $PM_{2.5}$ concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. *Journal of Geophysical Research*, *109*, D22206. https://doi.org/10.1029/2004JD005025

Ma, Z., Hu, X., Sayer, A. M., Levy, R., Zhang, Q., Xue, Y., … Liu, Y. (2016). Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004-2013. *Environmental Health Perspectives*, *124*(2), 184–192. https://doi.org/10.1289/ehp.1409481

Martin, R. V. (2008). Satellite remote sensing of surface air quality. *Atmospheric Environment*, *42*(34), 7823–7843. https://doi.org/10.1016/j.atmosenv.2008.07.018

Ong, B. T., Sugiura, K., & Zettsu, K. (2015). Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting $PM_{2.5}$. *Neural Computing and Applications*, 1–14.

Paciorek, C. J., Liu, Y., Moreno-Macias, H., & Kondragunta, S. (2008). Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level $PM_{2.5}$. *Environmental Science & Technology*, *42*(15), 5800–5806. https://doi.org/10.1021/es703181j

Peng, J., Chen, S., Lü, H., Liu, Y., & Wu, J. (2016). Spatiotemporal patterns of remotely sensed $PM_{2.5}$ concentration in China from 1999 to 2011. *Remote Sensing of Environment*, *174*, 109–121. https://doi.org/10.1016/j.rse.2015.12.008

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575. https://doi.org/10.1109/TPAMI.2009.187

Song, W., Jia, H., Huang, J., & Zhang, Y. (2014). A satellite-based geographically weighted regression model for regional $PM_{2.5}$ estimation over the Pearl River Delta region in China. *Remote Sensing of Environment*, *154*, 1–7. https://doi.org/10.1016/j.rse.2014.08.008

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, *46*(sup1), 234–240.

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., … Winker, D. M. (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental Science & Technology*, *50*(7), 3762–3772. https://doi.org/10.1021/acs.est.5b05833

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, *118*(6), 847–855. https://doi.org/10.1289/ehp.0901623

van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research*, *111*, D21201. https://doi.org/10.1029/2005JD006996

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and $PM_{2.5}$ mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21), 2095. https://doi.org/10.1029/2003GL018174

Weber, S. A., Engel-Cox, J. A., Hoff, R. M., Prados, A. I., & Zhang, H. (2010). An improved method for estimating surface fine particle concentrations using seasonally adjusted satellite aerosol optical depth. *Journal of the Air & Waste Management Association*, *60*(5), 574–585. https://doi.org/10.3155/1047-3289.60.5.574

WHO (2006). *Air quality guidelines: Global update 2005: Particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization. Retrieved from https://books.glgoo.com/books?id=7VbxUdIJE8wC&lr=&hl=zh-EN&source=gbs_navlinks_s

Wu, J., Li, J., Peng, J., Li, W., Xu, G., & Dong, C. (2015). Applying land use regression model to estimate spatial variation of $PM_{2.5}$ in Beijing, China. *Environmental Science and Pollution Research*, *22*(9), 7045–7061. https://doi.org/10.1007/s11356-014-3893-5

Wu, Y., Guo, J., Zhang, X., Tian, X., Zhang, J., Wang, Y., … Li, X. (2012). Synergy of satellite and ground based observations in estimation of particulate matter in eastern China. *Science of the Total Environment*, *433*, 20–30. https://doi.org/10.1016/j.scitotenv.2012.06.033

Yao, L., & Lu, N. (2014). Spatiotemporal distribution and short-term trends of particulate matter concentration over China, 2006–2010. *Environmental Science and Pollution Research*, *21*(16), 9665–9675. https://doi.org/10.1007/s11356-014-2996-3

You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., & Wang, W. (2016). National-scale estimates of ground-level $PM_{2.5}$ concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. *Remote Sensing*, *8*(3), 184. https://doi.org/10.3390/rs8030184

Yuan, Q., Zhang, L., & Shen, H. (2012). Hyperspectral image denoising employing a spectral-spatial adaptive total variation model. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(10), 3660–3677. https://doi.org/10.1109/TGRS.2012.2185054

Yue, L., Shen, H., Zhang, L., Zheng, X., Zhang, F., & Yuan, Q. (2017). High-quality seamless DEM generation blending SRTM-1, ASTER GDEM v2 and ICESat/GLAS observations. *ISPRS Journal of Photogrammetry and Remote Sensing*, *123*, 20–34. https://doi.org/10.1016/j.isprsjprs.2016.11.002

Zhang, H., Hoff, R. M., & Engel-Cox, J. A. (2009). The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and $PM_{2.5}$ over the United States: A geographical comparison by U.S. Environmental Protection Agency regions. *Journal of the Air & Waste Management Association*, *59*(11), 1358–1369. https://doi.org/10.3155/1047-3289.59.11.1358

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., … Qiu, Y. (2017). Analysis of influential factors for the relationship between $PM_{2.5}$ and AOD in Beijing. *Atmospheric Chemistry and Physics Discussions*, *2017*, 1–57.