A novel approach of Landsat 8 imagery to predict PM$_{2.5}$ concentrations in a south-eastern coastal city of China

Lijuan Yang
Ocean College of Minjiang University, Fuzhou, 350118, China.
2611@mju.edu.cn
E-mail: subinarzhong@aliyun.com

Abstract. Satellite remote sensing data with moderate- to high- resolution has been commonly used in deriving spatial coverage of PM$_{2.5}$ concentrations in urban areas. Previous studies focusing on city-scale PM$_{2.5}$ estimation mainly retrieved aerosol optical depth from moderate-to high- resolution remote sensing data. In this study, the spectral response experiment was carried out to explore the sensitivity of spectral wavelengths to PM$_{2.5}$ concentrations in Fuzhou, China. The results showed that the near-infrared reflectance was much more sensitive to PM$_{2.5}$ than other wavelengths. We also found that the difference vegetation index (DVI) presented higher correlation with PM$_{2.5}$ than other indexes. A linear mixed effects (LME) model was then developed to explain the variability of PM$_{2.5}$, and the results showed that the overall $R^2$ of LME model using DVI and meteorological parameters reached 0.80 with RMSE of 7.82 $\mu$g/m$^3$. The results suggested that the proposed LME model using DVI from Landsat 8 OLI could be effectively used for predicting PM$_{2.5}$ in a city scale.

1. Introduction
Many previous and recent epidemiological studies have reported a significant association between exposure to fine particles with aerodynamic diameter less than 2.5 um, i.e., PM$_{2.5}$, and various adverse health effects including asthma$^{[1-2]}$, cardiovascular problems$^{[3-4]}$, lung cancer and mortality$^{[5-6]}$. Typically, the most direct way to obtain the accurate PM$_{2.5}$ concentrations is from the ground PM$_{2.5}$ monitoring sites, which however, cannot be used for retrieving the spatial coverage of PM$_{2.5}$ since the number of these sites is limited and sparse. Satellite with various sensors on board can provide remote sensing imagery with extensive coverage of large area, and the retrieved aerosol optical depth (AOD) representing the atmospheric turbidity has been widely used in assessing the air pollution$^{[7-8]}$.

Currently, several AOD products with various resolution have been demonstrated as good proxy to retrieve the ground-level PM$_{2.5}$ in large- and global- coverage, however, it cannot provide more detailed information in a city scale due to its low spatial resolution$^{[9-10]}$. Thus, more and more researchers have diverted to urban PM$_{2.5}$ estimation by using remote sensing data with moderate- to high- resolution. Previous studies of urban PM$_{2.5}$ estimation mainly focused on retrieving AOD from several moderate- to high- resolution data$^{[11]}$, or established the relationship between retrieved AOD and particulate matter concentrations$^{[12]}$. Several studies also used the remote sensing bands and indices to predict ground-level PM$_{10}$ and PM$_{5.0}$$^{[13-14]}$. One of the limitations of these studies is that they used simple regression models to account for the variability between particulate matter concentrations and other variables (e.g., retrieved AOD, meteorological parameters and land use information), which presented relatively lower predictability in estimating PM concentrations. Therefore, researchers have...
gradually focused on urban PM concentrations estimation by using the advanced statistical models along with the meteorological parameters. For example, researchers used a linear mixed effects (LME) model to explain the variability of the AOD retrieved from GF-1 imageries, meteorological factors and PM$_{2.5}$ in Beijing, Shanghai and Wuhan, respectively, and the $R^2$ of estimated and observed PM$_{2.5}$ reached 0.96$^{[15]}$. 

Previous studies that have linked particulate matter concentrations with ultrahigh resolution AOD, or with remote sensing bands and combined indexes, showed a great capability of remote sensing data with moderate- to high- resolution in monitoring atmospheric aerosols in urban areas. Previous work demonstrated that the aerosol information could be obtained by using visible and near-infrared wavelengths of satellite sensors$^{[16]}$, but the response mechanism of different spectral wavelengths to PM$_{2.5}$ concentration is still unknown. Thus, in this study, we explored the spectral response of different wavelengths of remote sensing images with moderate- to high- resolution to PM$_{2.5}$ concentrations, and discussed the most sensitive spectral wavelengths. Three Landsat 8 imageries with spatial resolution of 30 m covering Fuzhou district, a south-eastern coastal city of China, were used in this study. Unlike the previous studies, we developed a new method that combined the most sensitive spectral wavelengths, instead of AOD, to predict the PM$_{2.5}$ concentrations within a linear mixed effects model in Fuzhou area.

2. Materials and Methods

2.1. Study area
Fuzhou is an important port city that located in the coastal zonal area of Fujian province, between the Yangtze River Delta (YRD) to the north and the Pearl River Delta (PRD) to the south. The study area now includes six urban districts and six surrounding suburban districts (Fig. 1). The air quality has declined due to the urban expansion and intensive human activities in Fuzhou, for instance, the daily PM$_{2.5}$ concentration in Fuzhou exceeded the fourth grade standard of China (115 $\mu$g/m$^3$) in 14% of the days during the entire winter in 2015. Moreover, the air pollution in the study area has also been exacerbated by the pollutants originating from YRD region in cold season.

2.2. Data

2.2.1. Landsat 8 satellite data.
The Landsat 8 satellite with two on-board instruments, i.e., Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS), was launched on February 11, 2013. It provides digital earth images with nine spectral bands (band1-7, 9: 30 m, band8: 15 m) and two thermal wavelengths (100 m). Three cloud-free Landsat 8 imageries covering Fuzhou area from 2014 to 2016 were collected in this study. The atmospheric correction on the data of Landsat 8 imageries was carried out before the application.

2.2.2. PM$_{2.5}$ observation data.
The corresponding hourly PM$_{2.5}$ concentrations measured at 45 monitoring sites for three days were collected (Fig. 1), and the averaged PM$_{2.5}$ concentrations from 10:00 to 12:00 were used as the dependent variable. Before data modeling, all the abnormal PM$_{2.5}$ data (i.e., PM$_{2.5}$$< 0$) were discarded to reduce the estimation error.

2.2.3. Meteorological data.
There are only one meteorological station located in the center of study area with available data, thus, the meteorological data from Goddard Earth Observing System Data Forward Processing (GEOS-5 FP) were used in this study. The spatial resolution of GEOS-5 FP data is $0.25^\circ$ (latitude) $\times 0.3125^\circ$ (longitude) in China, and two meteorological variables including relative humidity (RH, %) and wind speed (WS, m/s) were obtained in this study. For data integration, we interpolated these meteorological data to a 30 m grid using the inverse distance weighting method.
2.3. model development.

For comparison, we defined PM$_{2.5}$ concentrations with $\leq$30, 30-50, and >50 $\mu$g/m$^3$ as low, moderate, and high level, respectively. Three land-cover types including vegetation, bare soil and built-up area were selected for spectral response analysis. By comparing the spectral response curve of three land-cover types in different PM$_{2.5}$ levels, the most sensitive spectral wavelengths, or the combined wavelengths, which presented the highest correlation coefficient with PM$_{2.5}$ concentrations were finally determined.

A simple regression model incorporating the most sensitive spectral wavelength/combined wavelengths and meteorological variables was first developed in this study. A linear mixed effects (LME) model that contains two effects, i.e., fixed effect and random effect, was further applied to predict the PM$_{2.5}$ concentrations. Here in the LME model, we defined the fixed effect as the average impact of all the independent variables on PM$_{2.5}$ for the entire period, while the random effect explains the variation between PM$_{2.5}$ and the independent variables. The LME model was performed in R language:

$$PM_{2.5ij} = (a + b_j) + (c + d_j) \cdot SSW_{ij} + (x + y_j) \cdot WS_{ij} + (k + z_j) \cdot RH_{ij} + \epsilon_{ij}$$ (1)

where $PM_{2.5ij}$, $SSW_{ij}$, $WS_{ij}$ and $RH_{ij}$ are the four variables in LME model, which represent the hourly average PM$_{2.5}$, the value of the sensitive spectral wavelength, the wind speed and relative humidity on day $j$ at site $i$, respectively. $a$ and $b_j$ represent the fixed and random intercept for the entire period; $c$, $x$ and $k$ are the fixed slopes for the sensitive spectral wavelength, WS and RH, respectively; $d_j$, $y_j$ and $z_j$ are the random slopes for the sensitive spectral wavelength, WS and RH, respectively; and $\epsilon_{ij}$ represents the random error.

3. Results and Discussion

3.1. DataSpectral response analysis

Three days representing three different PM$_{2.5}$ levels, i.e., September 27, 2015 (28 $\mu$g/m$^3$), June 25, 2016 (40 $\mu$g/m$^3$) and April 17, 2014 (87 $\mu$g/m$^3$), were selected for the spectral characteristics analysis. The results showed that three types all presented higher near-infrared reflectance (i.e., NIR or b5) in clear days with low PM$_{2.5}$ concentrations than those with high PM$_{2.5}$, namely, the NIR reflectance.
decreased with the increase of PM$_{2.5}$ concentrations. The mid-infrared reflectance varied much and presented unstable relationship in different PM$_{2.5}$ levels for three land-cover types. We achieved a statistical correlation coefficient ($r$) value of 0.50, 0.61 and 0.56 (P<0.01), respectively. These $r$ values were much higher than those obtained from other days with less available PM$_{2.5}$ data ($r$=0.30–0.50, P<0.05). Eight NIR-related indexes including ratio vegetation index (RVI), difference vegetation index (DVI), normalized difference vegetation index (NDVI), soil adjusted vegetation index (SAVI), atmospherically resistant vegetation index (ARVI), modified normalized difference vegetation index (MNDVI), normalized difference moisture index (NDMI) and normalized difference built-up index (NDBI) were also employed to validate their relationship with PM$_{2.5}$ concentrations. Our results indicated that six vegetation indexes were generally stronger relative to PM$_{2.5}$ than NDMI and NDBI, while DVI presented the highest $r$ value of ~0.50 in three PM$_{2.5}$ levels.

3.2. Results of model fitting and validation

In view of the high correlation of the difference vegetation index and PM$_{2.5}$ concentrations, the difference vegetation index, instead of NIR, was employed as the main variable to predict the PM$_{2.5}$ concentrations in Fuzhou. In addition to DVI, two meteorological parameters including wind speed and relative humidity were also added as ancillary variables in our models. We aimed to develop an advanced statistical model, i.e., the linear mixed effects (LME) model, to explain the variability of PM$_{2.5}$ concentrations and other independent variables, which has previously been demonstrated to present high predictability in PM$_{2.5}$ estimation by using AOD with coarse resolution. We achieved an overall $R^2$ of 0.80 between the estimated PM$_{2.5}$ and observed values in LME model (Fig. 2).

The RMSE of the LME model was 7.82 $\mu$g/m$^3$. The slopes of fixed effect for DVI, WS and RH were -46.15, -3.39 and -9.76, respectively, indicating that DVI, wind speed and relative humidity were all generally negatively related to PM$_{2.5}$. The WS was generally negatively correlated with PM$_{2.5}$ concentrations, indicating that higher wind speed would facilitate the dilution of particulate matter concentrations. The RH, representing the mass proportion of water vapor in the atmosphere, can also reduce the pollutant concentrations to some extent, such as the heavy rainfall with RH great than 80%. Furthermore, the results of the 10-fold cross validation showed that the average CV $R^2$ values for LME model was 0.78, with CV RMSE of 8.01 $\mu$g/m$^3$, indicating the significant predictability of our LME model.

3.3. Predictions of PM$_{2.5}$ concentrations

We achieved the spatial coverage of ground-level PM$_{2.5}$ concentrations by using DVI and LME model for three days in Fuzhou. Figure 3 illustrates the map of the PM$_{2.5}$ concentrations on September 27,
2015 and June 25, 2016, which also represents three different PM$_{2.5}$ levels. The derived-mean PM$_{2.5}$ concentrations for the entire study area were 28 μg/m$^3$, and 40 μg/m$^3$ on these two dates, respectively, exhibiting higher mean PM$_{2.5}$ concentrations in the center and southeast of the study area, and lower values in west and north. For comparison, the average PM$_{2.5}$ concentrations for each district was then derived, and finally, we obtained much higher PM$_{2.5}$ concentrations in Taijiang, Gulou and Cangshan than other surrounding districts. These three urban districts are the most frequent areas of human activities in the socio-economic, which are characterized by high population density and road density. To sum up, the large population and road density, as well as the weak urban ventilation, are the key reasons for high PM$_{2.5}$ values in three urban districts.

Fig. 3. Spatial distribution of derived PM$_{2.5}$ concentrations for the Fuzhou area (left: 2015/09/27; right: 2016/06/25)

4. Conclusion
In this study, a spectral response experiment was first launched to explore the sensitivity of spectral wavelengths to PM$_{2.5}$ concentrations, by which the most sensitive remote sensing wavelength (i.e., NIR) and combined index (i.e., different vegetation index) were determined. The linear mixed effects model that incorporated the DVI and two meteorological parameters was further developed, and our results revealed that the LME model could explain 80% of the variability between PM$_{2.5}$ and other independent variables. Moreover, the retrieved PM$_{2.5}$ concentrations was also consistent with the urban pattern, showing high PM$_{2.5}$ concentrations in urban districts with large population density and relatively low PM$_{2.5}$ values in suburban districts. In summary, this is the first study that uses the spectrum response experiment aiming moderate- to high- resolution remote sensing data, along with a linear mixed effects model to assess the fine particles pollution in a small scaled region. The results of this thesis would provide a brand new methodology in estimating PM$_{2.5}$ concentrations with more details in city-scale.

Acknowledgement
The authors would like to thank GSFC/NASA for the use of Landsat 8 data. This work was supported by the National Key Research and Development Project (No: 2016YFA0600302).

References
[1] Shakya, K. M., M. Rupakheti, A. Shahi, R. Maskey, B. Pradhan, A. Panday, S. P. Puppala, M. Lawrence, and R. E. Peltier. 2017. Near-road sampling of PM2.5, BC, and fine-particle chemical components in Kathmandu Valley, Nepal, Atmospheric Chemistry and Physics, 17: 6503-16.
[2] Yao, F., M. L. Si, W. F. Li, and J. S. Wu. 2018. A multidimensional comparison between MODIS and VIIRS AOD in estimating ground-level PM2.5 concentrations over a heavily polluted region in China. Science of the Total Environment, 819-828.

[3] Kloog, I., B. A. Coull, A. Zanobetti, P. Koutrakis, and J. D. Schwartz. 2012. Acute and Chronic Effects of Particles on Hospital Admissions in New-England, PLoS One, 7.

[4] Mantovani, K. C. C., L. F. C. Nascimento, D. S. Moreira, L. C. P. F. D. Vieira, and N. P. Vargas. 2016. Air pollutants and hospital admissions due to cardiovascular diseases in Sao Jose do Rio Preto, Brazil, Ciencia&SaudeColetiva, 21: 509-15.

[5] Kloog, I., P. Koutrakis, B. A. Coull, H. J. Lee, and J. Schwartz. 2011. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements, Atmospheric Environment, 45: 6267-75.

[6] Abba, E. J., S. Unnikrishnan, R. Kumar, B. Yeole, and Z. Chowdhury. 2012. Fine aerosol and PAH carcinogenicity estimation in outdoor environment of Mumbai City, India, International Journal of Environmental Health Research, 22: 134-49.

[7] Zhong, G. S., X. F. Wang, M. Guo, H. Tani, A. R. Chittenden, S. Yin, Z. Y. Sun, and S. Matsumura. 2017. A Dark Target Algorithm for the GOSAT TANSO-CAI Sensor in Aerosol Optical Depth Retrieval over Land, Remote Sensing, 9.

[8] He, Q. Q., and B. Huang. 2018. Satellite-based mapping of daily high-resolution ground PM2.5 in China via space-time regression modeling, Remote Sensing of Environment, 206: 72-83.

[9] Chen, X. F., Z. Q. Li, S. S. Zhao, L. K. Yang, Y. Ma, L. Liu, D. H. Li, L. L. Qie, and J. Xing. 2018. Using the Gaofen-4 geostationary satellite to retrieve aerosols with high spatiotemporal resolution, Journal of Applied Remote Sensing, 12.

[10] Mei, L. L., V. Rozanov, M. Vountas, J. P. Burrows, and A. Richter. 2018. XBAER-derived aerosol optical thickness from OLCI/Sentinel-3 observation, Atmospheric Chemistry and Physics, 18: 2511-23.

[11] Li, Z. B., D. P. Roy, H. K. K. Zhang, E. F. Vermote, and H. Y. Huang. 2019. Evaluation of Landsat-8 and Sentinel-2A Aerosol Optical Depth Retrievals across Chinese Cities and Implications for Medium Spatial Resolution Urban Aerosol Monitoring, Remote Sensing, 11.

[12] Luo, N., M. S. Wong, W. J. Zhao, X. Yan, and F. Xiao. 2015. Improved aerosol retrieval algorithm using Landsat images and its application for PM10 monitoring over urban areas, Atmospheric Research, 153: 264-75.

[13] Amanollahi, J., C. Tzanis, A. M. Abdullah, M. F. Ramli, and S. Pirasteh. 2013. Development of the models to estimate particulate matter from thermal infrared band of Landsat Enhanced Thematic Mapper, International Journal of Environmental Science and Technology, 10: 1245-54.

[14] Ozelkan, E., M. Karaman, S. Mostamandy, Z. D. U. Avci, and H. Toros. 2015. Derivation of Pm10 Levels Using Obra on Landsat 5 Tm Images: A Case Study in Izmir, Turkey, Fresenius Environmental Bulletin, 24: 1585-96.

[15] Zhang, T. H., Z. M. Zhu, W. Gong, Z. R. Zhu, K. Sun, L. C. Wang, Y. S. Huang, F. Y. Mao, H. F. Shen, Z. W. Li, and K. Xu. 2018. Estimation of ultrahigh resolution PM2.5 concentrations in urban areas using 160 m Gaofen-1 AOD retrievals, Remote Sensing of Environment, 216: 91-104.

[16] Kahn, R., P. Banerjee, D. McDonald, and D. J. Diner. 1998. Sensitivity of multiangle imaging to aerosol optical depth and to pure-particle size distribution and composition over ocean, Journal of Geophysical Research-Atmospheres, 103: 32195-213.