ESTIMATING PM2.5 AND PM10 ON ZHUHAI-1 HYPERSPECTRAL IMAGERY

Shengjie Liu[1], *Qian Shi*[2]

[1]Spatial Sciences Institute, University of Southern California, Los Angeles, CA 90089, USA
[2]School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

ABSTRACT

Particulate matter (PM), such as PM2.5 and PM10, was the major pollutant in a severe air pollution episode in 2013 eastern China. Limited by the coverage of stations, fine-scale monitoring at every corner in the city is difficult, if not impossible. Hyperspectral imagery can capture the ground and air information, from which we can estimate the concentrations of PM. In this study, we develop a multitask learning method to estimate the concentrations of PM based on the 10-m hyperspectral data from the newly-launched Zhuhai-1 satellites. We first convert the raw radiance to top-of-atmosphere (TOA) reflectance using the 1985 Wehrli solar irradiance spectrum. Then, we train a multitask network to simultaneously estimate PM2.5 and PM10 concentrations based on the TOA hyperspectral data. Results show that our method leads to estimations of an R-squared of 0.77 for PM2.5 and an R-squared of 0.42 for PM10.

Index Terms— Air quality, PM2.5, PM10, hyperspectral data, multitask learning, Zhuhai-1 satellite

1. INTRODUCTION

Studies have shown that long-term exposure to air pollution increases the risks of lung cancers, heart diseases, cardiopulmonary diseases, and cognitive impairment [1, 2, 3, 4]. With the fast industrialization of mainland China since the 1980s, air quality has degraded to sometimes over 200 $\mu g/m^3$, where the standard by WHO is below 35 $\mu g/m^3$ [5]. Since the severe air pollution in 2013 eastern China, the government has started to build a national network of 1600 stations to monitor air quality, including PM2.5 and PM10. The time frequency is one hour. Although ground stations have filled the gap of lack of data, these stations are unevenly distributed in over 300 cities and mostly installed at locations with prime monitoring conditions, leading to a lack of fine-scale monitoring. Pérez et al. (2010) [6] pointed out that PM2.5 on or near road networks were more severe. With only a few stations in each city, it is not possible to monitor air quality at a fine scale that can distinguish near-road or off-road air pollution [7].

Several studies estimated PM concentrations based on satellite data. Except for the few that estimate PM concentrations directly from TOA reflectance [7, 8], most of the existing studies first estimate aerosol optical depth (AOD), and then estimate PM concentrations based on AOD [9, 10, 11]. Although there are ready-to-use AOD products such as MODIS, VIIRS, Himawari-8 and Sentinel-5, these products are mostly with a spatial resolution of greater than 500 m. It is difficult to use these products to analyze within-city air quality. In recent years, some studies attempted to use medium-resolution satellite data like Landsat-TM/OLI, Sentinel-2 and GF-1 to estimate AOD and PM concentrations [12, 13, 14, 15]. But these satellite data, compared to the most-often-used MODIS, contain fewer spectral bands; and they are not ready-to-use products.

For AOD estimation, the key is to separate land and atmosphere information and accurately estimate the land surface reflectance (down-of-atmosphere, DOA). Current popular algorithms include the Dense Dark Vegetation (DDV), the Deep Blue, and a novel simplified aerosol retrieval algorithm named SARA [16, 17, 18]. The DDV algorithm uses the different reflectance in blue and red channels of vegetation and their linear relationships with infrared to estimate the land surface reflectance of blue and red lights. After that, the contributions from the atmosphere of TOA reflectance can be calculated. But this algorithm only works in areas with dense vegetation. In cities and deserts with high albedo of red and blue lights, the algorithm's prerequisite cannot be met. Additionally, the linear relationship between spectral bands is too rough for ideal estimation [19]. The Deep Blue algorithm assumes that blue light has short wavelength, low DOA reflectance, and high contribution on AOD; therefore, it can be compared with nearby clear DOA reflectance to estimate AOD. But the lack of high-resolution land surface reflectance limits its usage. The SARA algorithm is based on radiative transfer models but the calculation is too complex and requires the asymmetry parameter g and aerosol models. In summary, limitations exist in all three popular AOD algorithms, where the bottleneck is the estimation of land surface reflectance. Additionally, via the estimation of AOD to esti-

SL's contribution was supported by a USC Dornsife PhD Fellowship of the Spatial Sciences Institute, University of Southern California. QS's contribution was supported by the National Natural Science Foundation of China under Grant 61976234. Contact: S. Liu (liusheng@usc.edu); Q. Shi (shixi5@mail.sysu.edu.cn).

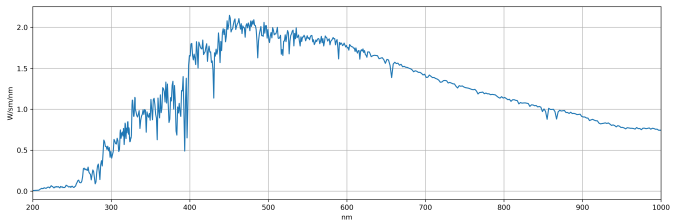


Fig. 1: The 1985 Wehrli Standard Extraterrestrial Solar Irradiance Spectrum

mate PM may lead to further errors.

Since the relationship from TOA to AOD to PM holds, it should be possible to estimate PM directly from TOA. There were a few studies in this direction. For example, Shen et al. (2018) [7] used a restricted Boltzmann machine to estimate PM2.5 based on MODIS imagery with 500 m resolution. Liu et al. (2019) [8] used ensemble learning to directly estimate PM2.5 from Himawari-8 imagery with 2000 m resolution. These studies showed some promising results, but their resolutions were too low for neighborhood-scale analysis. In this study, we directly estimate PM2.5 and PM10 from TOA reflectance at 10 m spatial resolution on the Zhuhai-1 hyperspectral imagery.

2. METHOD AND DATA

2.1. Data

The Zhuhai-1 satellites were launched by Zhuhai Orbita. They provide 32 bands within 443 and 940 nm on a spatial resolution of 10 m. A total of 81 scenes across various Chinese cities were downloaded in 2019. After screening for those with ground stations and valid cloudless data, the number of remaining scenes was 30. A total of 77 data points were extracted for final analysis.

The obtained level-1 hyperspectral data need preprocessing. We first converted them to radiance L_λ (radiometric calibration):

$$L_\lambda = a_\lambda \times \frac{DN_\lambda}{TDIStage_\lambda} + b_\lambda, \tag{1}$$

in which DN_λ is the level-1 raw data, $TDIStage_\lambda$ is a scale factor provided in the header file, and a_λ and b_λ are the multiplicative and additive factors.

The data didn't provide the conversion from radiance to reflectance. We estimated the $ESUN_\lambda$ value based on the 1985 Wehrli standard extraterrestrial solar irradiance spectrum (as shown in Figure 1) and the spectral response functions of all sensors of the Zhuhai-1 satellites (there are a few differences between each sensor). Then we had the TOA reflectance of each band ρ_λ as

$$\rho_\lambda = \frac{\pi L_\lambda d^2}{ESUN_\lambda} cos(\theta_s), \tag{2}$$

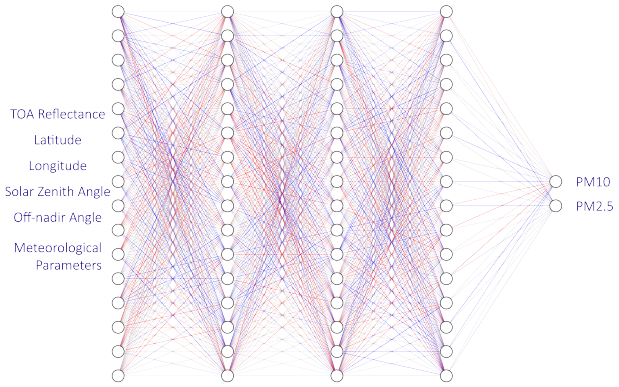


Fig. 2: PMNet

where $ESUN_{\lambda_0}$ was calculated from

$$ESUN_{\lambda_0} = \frac{\int_{\lambda_1}^{\lambda_2} E(\lambda)S(\lambda)d\lambda}{\int_{\lambda_1}^{\lambda_2} S(\lambda)d\lambda}. \tag{3}$$

We provide the ESUNs values of all sensors of the Zhuhai-1 satellites we used in Table 1.

2.2. Method

Our method is shown in Figure 2. We take inputs of the TOA reflectance, latitude, longitude, solar zenith angle, off-nadir angle and meteorological parameters into a multitask network and output two values: PM2.5 and PM10. In training, we set a learning rate of 0.5 with the AdaDelta optimizer, used a batch size of 4, and adopted the Snapshot Ensemble method [20]. Of the 77 records, 11 were reserved for validation and 66 were used in training.

3. RESULTS AND ANALYSIS

3.1. Accuracy of estimation

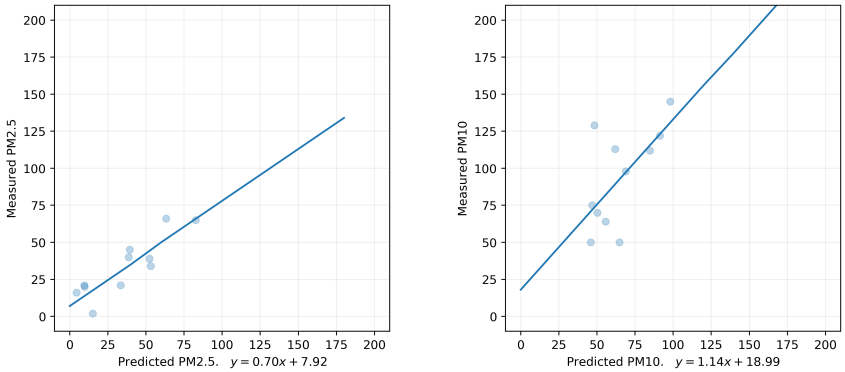
In Figure 3 we show the estimation accuracy in term of R-squared for PM2.5 and PM10. Figure 3a was the fitting on PM2.5 where it achieved an R-squared of 0.77, and Figure 3b was the fitting on PM10 where it achieved an R-squared of 0.42.

3.2. Comparison between hyperspectral and multispectral data

We used the bands 2, 7, 13, and 27 of the Zhuhai-1 hyperspectral data to simulate 4-band multispectral data. Based on the simulated multispectral data, the model achieved an R-squared of 0.22 on PM2.5 estimation (vs. 0.77 of hyperspectral data) and an R-squared of lower than 0.10 on PM10 estimation. This comparison shows the superiority of hyperspectral data in estimating PM.

Table 1: ESUNs of the Zhuhai-1 Sensors

Band	A-CMOS1	A-CMOS2	A-CMOS3	C-CMOS1	C-CMOS2	C-CMOS3	D-CMOS1	D-CMOS2	D-CMOS3
B01	1994.40	1997.24	1985.72	1999.07	2000.18	1963.43	2001.16	2005.57	1992.02
B02	2026.72	2033.41	2037.89	2035.24	2040.49	2032.90	2042.90	2042.16	2037.44
B03	1905.49	1920.94	1920.05	1916.60	1910.91	1906.35	1911.07	1926.53	1910.55
B04	1792.47	1820.72	1797.49	1808.96	1818.78	1815.91	1805.23	1795.35	1816.00
B05	1899.57	1911.70	1907.54	1889.60	1887.24	1888.81	1900.38	1898.13	1887.45
B06	1857.66	1867.86	1864.57	1871.60	1871.89	1871.87	1872.74	1873.02	1870.85
B07	1812.80	1831.26	1828.39	1839.67	1839.63	1839.48	1836.62	1836.75	1840.57
B08	1816.54	1844.81	1839.62	1840.97	1841.97	1841.46	1839.02	1839.13	1838.81
B09	1778.48	1784.24	1785.88	1781.32	1779.84	1781.34	1784.05	1782.12	1780.22
B10	1741.05	1734.45	1738.05	1730.31	1728.27	1729.15	1729.51	1729.83	1733.55
B11	1675.72	1677.15	1676.03	1677.25	1676.54	1677.29	1676.33	1676.61	1678.76
B12	1636.47	1639.04	1642.66	1637.43	1635.76	1635.53	1635.63	1636.31	1637.76
B13	1500.06	1499.58	1521.23	1528.16	1526.70	1522.82	1517.91	1512.54	1530.66
B14	1533.28	1531.77	1534.99	1531.55	1528.16	1530.81	1530.66	1531.06	1532.62
B15	1467.95	1468.09	1469.24	1470.86	1466.42	1469.42	1467.85	1468.15	1468.62
B16	1410.68	1406.33	1413.09	1411.41	1408.35	1410.05	1410.63	1409.10	1413.15
B17	1363.36	1352.81	1366.45	1362.42	1356.21	1357.92	1359.05	1354.83	1357.53
B18	1329.28	1327.26	1331.48	1329.11	1322.49	1327.41	1326.59	1325.50	1329.97
B19	1278.27	1279.84	1279.40	1275.19	1272.84	1275.32	1275.68	1275.61	1276.80
B20	1243.03	1240.54	1243.85	1240.87	1236.25	1239.80	1239.18	1239.60	1238.70
B21	1195.05	1195.53	1196.70	1194.02	1192.54	1193.20	1193.10	1193.30	1195.02
B22	1171.44	1167.55	1170.68	1165.65	1159.44	1165.17	1165.72	1164.51	1163.43
B23	1123.01	1119.83	1125.35	1123.14	1119.24	1121.50	1119.27	1120.22	1119.76
B24	1085.63	1077.91	1085.03	1085.13	1082.80	1082.75	1081.18	1080.54	1087.94
B25	1050.96	1051.10	1052.10	1048.66	1046.53	1046.77	1049.40	1048.05	1051.21
B26	1004.87	994.54	1000.71	990.48	985.54	988.77	986.60	983.80	995.13
B27	958.19	943.35	961.60	965.67	964.54	965.45	960.54	958.62	966.97
B28	964.80	958.59	961.60	962.79	961.45	962.24	962.41	962.15	968.58
B29	932.15	935.42	932.18	927.27	926.26	923.65	929.68	931.33	926.24
B30	878.39	878.22	882.46	880.63	879.72	876.86	877.39	878.63	875.29
B31	832.16	830.28	833.12	832.46	831.66	829.77	830.93	831.84	829.23
B32	803.54	803.49	818.01	801.55	801.63	799.29	801.11	804.53	806.01



(a) $R^2 = 0.77$. (b) $R^2 = 0.42$.

Fig. 3: Estimation of PM2.5 and PM10. X-axis: estimated values; y-axis: ground-station values.

3.3. Mapping PM2.5 of Nanchang City on 6 Oct 2018

We applied the trained model on a Zhuhai-1 hyperspectral image captured on 6 Oct 2018 over the Nanchang City, a southern Chinese city. At the time of imaging, the off-nadir angle, the solar zenith angle, temperature, relative humidity, wind speed and air pressure were 0.40984, 53.13548, 27.3°C, 29, 3, and 1012 millibars, respectively. The model applied on pixels with band-1 reflectance smaller than 0.2. For missing pixels, we used the random forest where the latitude and longitude as inputs to estimate PM2.5 as outputs. The map is shown in Figure 4.

As shown, road networks were notable with high concentrations of PM2.5. The center left area, which appeared to be a large construction site, had the highest PM2.5 value in this area. Areas with abundant trees, which appeared in red on the right, showed low air pollution on the left. Overall it is possible to distinguish near-road or off-road air pollution.

4. CONCLUSION

In this study, we co-estimated PM2.5 and PM10 from hyperspectral data at 10 m spatial resolution. Results show that an R-squared of 0.77 for PM2.5 and an R-squared of 0.42 for PM10 can be achieved even with a limited set of samples. Near-road air pollution and construction-related air pollution were successfully distinguished in a case study of Nanchang City, China. Future studies may benefit from the detailed es-

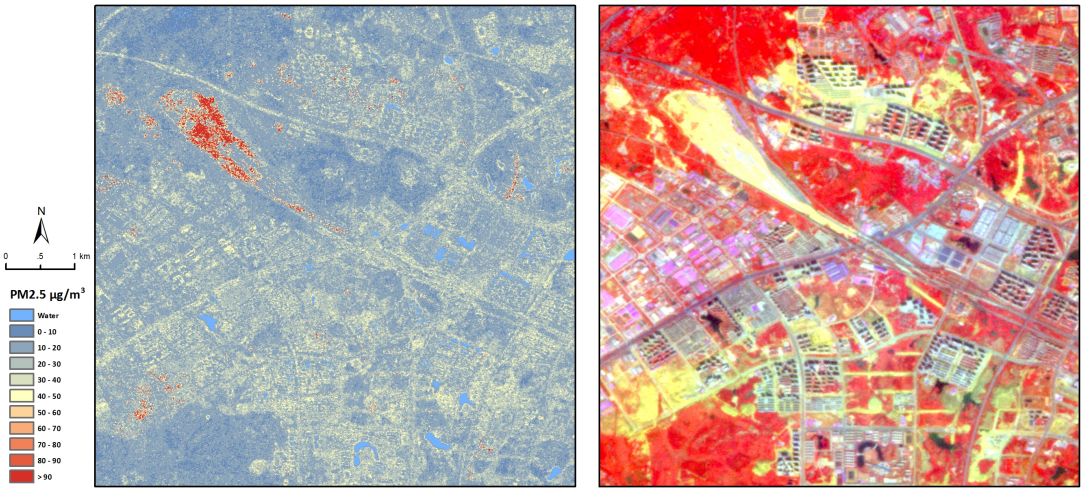


Fig. 4: Subset of the estimated PM2.5 of Nanchang City on 6 Oct 2018. **Left:** predicted PM2.5. **Right:** False Color Map.

timation to investigate people's health and their exposure to air pollution.

5. REFERENCES

[1] Douglas W Dockery, C Arden Pope, Xiping Xu, John D Spengler, James H Ware, Martha E Fay, Benjamin G Ferris Jr, and Frank E Speizer, "An association between air pollution and mortality in six us cities," *N. Engl. J. Med.*, vol. 329, no. 24, pp. 1753–1759, 1993.

[2] C Arden Pope III, Richard T Burnett, Michael J Thun, Eugenia E Calle, Daniel Krewski, Kazuhiko Ito, and George D Thurston, "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution," *Jama*, vol. 287, no. 9, pp. 1132–1141, 2002.

[3] William F McDonnell, Naomi Nishino-Ishikawa, Floyd F Petersen, Lie Hong Chen, and David E Abbey, "Relationships of mortality with the fine and coarse fractions of long-term ambient pm 10 concentrations in nonsmokers," *J. Expo. Sci. Environ. Epidemiol.*, vol. 10, no. 5, pp. 427, 2000.

[4] Petkus et al., "Associations between air pollution exposure and subtypes of cognitive performance in older women," *Alzheimer's & Dementia*, vol. 17, pp. e056228, 2021.

[5] World Health Organization, *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*, World Health Organization, 2006.

[6] Pérez et al., "Variability of particle number, black carbon, and pm10, pm2. 5, and pm1 levels and speciation: influence of road traffic emissions on urban air quality," *Aerosol. Sci. Technol.*, vol. 44, no. 7, pp. 487–499, 2010.

[7] Huanfeng Shen, Tongwen Li, Qiangqiang Yuan, and Liangpei Zhang, "Estimating regional ground-level pm2. 5 directly from satellite top-of-atmosphere reflectance using deep belief networks," *J. Geophys. Res. Atmos.*, vol. 123, no. 24, pp. 13–875, 2018.

[8] Jianjun Liu, Fuzhong Weng, and Zhanqing Li, "Satellite-based pm2. 5 estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm," *Atmos. Environ.*, vol. 208, pp. 113–122, 2019.

[9] Jianping Guo, Feng Xia, Yong Zhang, Huan Liu, Jing Li, Mengyun Lou, Jing He, Yan Yan, Fu Wang, Min Min, et al., "Impact of diurnal variability and meteorological factors on the pm2. 5-aod relationship: Implications for pm2. 5 remote sensing," *Environ. Pollut.*, vol. 221, pp. 94–104, 2017.

[10] Siwei Li, Everette Joseph, and Qilong Min, "Remote sensing of ground-level pm2. 5 combining aod and backscattering profile," *Remote Sens. Environ.*, vol. 183, pp. 120–128, 2016.

[11] Ying Zhang and Zhengqiang Li, "Remote sensing of atmospheric fine particulate matter (pm2. 5) mass concentration near the ground from satellite observation," *Remote Sens. Environ.*, vol. 160, pp. 252–262, 2015.

[12] Zhang et al., "Estimation of ultrahigh resolution pm2. 5 concentrations in urban areas using 160 m gaofen-1 aod retrievals," *Remote Sens. Environ.*, vol. 216, pp. 91–104, 2018.

[13] Kun Sun, Xiaoling Chen, Zhongmin Zhu, and Tianhao Zhang, "High resolution aerosol optical depth retrieval using gaofen-1 wfv camera data," *Remote Sens.*, vol. 9, no. 1, pp. 89, 2017.

[14] Lin Sun, Jing Wei, Muhammad Bilal, Xinpeng Tian, Chen Jia, Yamin Guo, and Xueting Mi, "Aerosol optical depth retrieval over bright areas using landsat 8 oli images," *Remote Sens.*, vol. 8, no. 1, pp. 23, 2015.

[15] Xinpeng Tian, Qiang Liu, Zhenwei Song, Baocheng Dou, and Xiuhong Li, "Aerosol optical depth retrieval from landsat 8 oli images over urban areas supported by modis brdf/albedo data," *IEEE Geosci. Remote. Sens. Lett.*, vol. 15, no. 7, pp. 976–980, 2018.

[16] PM Teillet and G Fedosejevs, "On the dark target approach to atmospheric correction of remotely sensed data," *Can. J. Remote Sens.*, vol. 21, no. 4, pp. 374–387, 1995.

[17] NC Hsu, M-J Jeong, C Bettenhausen, AM Sayer, R Hansell, CS Seftor, J Huang, and S-C Tsay, "Enhanced deep blue aerosol retrieval algorithm: The second generation," *J. Geophys. Res. Atmos.*, vol. 118, no. 16, pp. 9296–9315, 2013.

[18] Muhammad Bilal, Janet E Nichol, Max P Bleiweiss, and David Dubois, "A simplified high resolution modis aerosol retrieval algorithm (sara) for use over mixed surfaces," *Remote Sens. Environ.*, vol. 136, pp. 135–145, 2013.

[19] SS Gillingham, Neil Flood, TK Gill, and RM Mitchell, "Limitations of the dense dark vegetation method for aerosol retrieval under australian conditions," *Remote Sens. Lett.*, vol. 3, no. 1, pp. 67–76, 2012.

[20] Gao Huang, Yixuan Li, Geoff Pleiss, Zhuang Liu, John E Hopcroft, and Kilian Q Weinberger, "Snapshot ensembles: Train 1, get m for free," *ICLR 2017*, 2017.