Supplementary information

LGHAP: a Long-term Gap-free High-resolution Air Pollutants concentration dataset derived via tensor flow based multimodal data fusion

Kaixu Bai[1,2*], Ke Li[1], Mingliang Ma[3], Kaitao Li[4], Zhengqiang Li[4], Jianping Guo[5*], Ni-Bin Chang[6], Zhuo Tan[1], Di Han[1]

[1]Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai 200241, China

[2]Institute of Eco-Chongming, 20 Cuiniao Rd., Chongming, Shanghai 202162, China

[3]School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan 250101, China

[4]State Environmental Protection Key Laboratory of Satellite Remote Sensing, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

[5]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China

[6]Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL, USA

[*]**Correspondence to:** Kaixu Bai (kxbai@geo.ecnu.edu.cn) and Jianping Guo (jpguocams@gmail.com)

Text S1: Satellite-based AOD products

• MAIAC AOD product

The multiangle implementation of atmospheric correction (MAIAC) is the latest algorithm used for processing MODIS Collection 6 data record, providing not only land surface reflectance but AOD products simultaneously with daily/1-km resolution. Since the first release of MAIAC algorithm in 2011 (Lyapustin et al., 2011), a variety of improvements had been made in cloud/snow detection, aerosol retrieval, and atmospheric correction, and the ultimate goal was to facilitate global scale processing and improve the data accuracy (Lyapustin et al., 2018). Please refer to Lyapustin et al. (2018) for more details related to the main changes. The MAIAC product was named after MCD19, and the AOD product was MCD19A2, which can be accessible via the Land Product Distributed Active Archive Center (LP DAAC). Many validation studies had confirmed that MAIAC AOD had a comparable or even better accuracy than those derived from the Dark Target and Deep Blue algorithms with higher spatial resolution (Goldberg et al., 2019; Lyapustin et al., 2018; Xiao et al., 2017). Consequently, this AOD product was used as the baseline dataset to help generate gap-filled high resolution AOD dataset in China.

• MISR AOD

The Multi-angle Imaging SpectroRadiometer (MISR) is one of five science instruments aboard the polar-orbiting Terra satellite. MISR observes the Earth globally at nine different view zenith angles in four spectral bands (446, 558, 672, and 866 nm) with a cross-track ground spatial resolution of 275 m–1.1 km. Its multi-angular imaging design has enabled to retrieve AOD, surface bidirectional reflectance factors, and hemispherical reflectances over a wide variety of land surface types, providing an extensive dataset of global Earth observations since the early 2000 (Diner et al., 2004; Garay et al., 2020; Witek et al., 2018). The characteristic of multi-angle earth observation is benefit to retrieve aerosol properties more intensively (Guo et al., 2012). Detailed descriptions to the MISR aerosol retrieval methodology over land, e.g., aerosol model selection and radiative transfer theory, can be found in Diner et al. (2004). Compared to the nadir sensor such as MODIS, MISR is more sensitive to aerosol components given the multi-angular imaging design. In this study, the latest version (V23) of

MISR aerosol product with the resolution of 4.4-km (Garay et al., 2020), was used to support the generation of gap-filled AOD imagery.

• VIIRS AOD

The Visible Infrared Imaging Radiometer Suite (VIIRS) is a new instrument aboard the polar-orbiting Suomi National Polar-orbiting Partnership (Suomi-NPP) spacecraft which was launched in 2011. As a new generation of operational moderate resolution-imaging radiometer, VIIRS has 22 imaging and radiometric bands covering wavelengths from 0.41 to 12.5 micrometers, providing more than twenty environmental data records to continue the Earth observation mission initiated by the AVHRR on NOAA and MODIS on Terra and Aqua satellites. Compared to the well-known MODIS, VIIRS has finer spatial resolution with a larger swath. It can provide AOD retrievals at a resolution of 0.75-km (IP product) and 6-km (Environmental Data Record) (Jackson et al., 2013; Levy et al., 2013). A specific aerosol retrieval algorithm was designed to infer AOD from VIIRS observations, with changes not only in theoretical basis and retrieval limitations but also data quality flagging (Jackson et al., 2013). Ground-based validation results indicated a generally good accuracy of VIIRS AOD (R^2=0.73) when compared to AERONET AOD observations (Xiao et al., 2016).

• AATSR AOD

The Advanced Along-Track Scanning Radiometer (AATSR) is one of the Announcement of Opportunity (AO) instruments onboard the European Space Agency (ESA) satellite Envisat, which was primarily designed for measuring sea surface temperature (Wen et al., 2019). The AATSR was designed to have seven spectral channels at wavelengths of 0.55, 0.67, 0.87, 1.63, 10.7, and 12 μm, which observes at dual view as one is the nadir direction and the other is forward direction with a viewing angle of 55^{o} from nadir view. The nadir spatial resolution is 1 km × 1 km with a swath width of 512 pixels (Che et al., 2016). By taking advantage of the features of dual view, AOD can be retrieved from the AATSR observations both over land and ocean. Previous studies had indicated that AOD from AATSR had a generally good accuracy, with R of 0.88 and RMS of 0.24 when compared with AOD observations from AERONET (de Leeuw et al., 2018). In this study, we used ensemble AOD product provided by Climate Change Initiative (CCI) that aggregating three new versions of AATSR aerosol datasets with spatial resolution of 0.1° (ADV v3.11, ORAC v4.10, and SU v4.32) (de Leeuw et al., 2018; Wen et al., 2019; Xie et al., 2018).

• POLDER AOD

The Polarization and Directionality of the Earths Reflectance (POLDER) is a first satellite-based multi-angle polarization imaging radiometer, which was developed by the French space agency CNES. The third POLDER radiometer (POLDER-3) was deployed on PARASOL satellite, which was operated from March 2005 to October 2013 with an Equator crossing time at 13:30 (local time). POLDER-3 used a 274 pixels × 242 pixels CCD detector array with a pixel size of 5.3 km ×6.2 km at nadir. It measures in 9 spectral channels from blue (0.443 μm) to near-infrared (1.020 μm), providing the polarization measurements at 0.490 μm, 0.670 μm and 0.865 μm at up to 16 different angles with a size of 2100 km ×1600 km (Formenti et al., 2018; Tan et al., 2019). With the Generalized Retrieval of Atmosphere and Surface Properties (GRASP) algorithm (Dubovik et al., 2011), POLDER can provide spectral AOD, fine aerosol optical depth (AODF), coarse aerosol optical depth (AODC), absorbed aerosol optical depth (AAOD), the Ångström exponent (AE), complex refractive index (CRI) and single scattering albedo (SSA) observations (Formenti et al., 2018; Tan et al., 2019). The inter-comparison results indicated that the POLDER AOD had a similar and even higher quality when compared to the MODIS AOD products derived from DT and DB algorithms (Chen et al., 2020; Tan et al., 2019).

Text S2: Data coverage ratio weighted nonlinear correlation coefficient

To assess the relative contribution of each gridded AOD product to the final gap-free AOD dataset, here we developed a new statistical indicator on the basis of the nonlinear correlation coefficient (NCC) that was proposed in Wang et al. (2005). The NCC was derived from the concept of entropy to assess the mutual information between two discrete variables. As shown in Sun et al. (2021), the NCC can be calculated as follows: given two variables x =[$x_1, x_2, \cdots, x_N$] and y =[$y_1, y_2, \cdots, y_N$] with N number of data pairs, we first sorted the data in each variable in an ascending order, which were then divided into *n* equal bins. The NCC between x and y is defined as:

$$NCC(x, y) = H(x) + H(y) - H(x, y) \quad (1)$$

where $H(x)$ is a revised entropy of x and can be calculated via the following equation:

$$H(x) = -\sum_{i=1}^{b} \frac{n_i}{N} \log_b \frac{n_i}{N} \tag{2}$$

In Eq. (2), n_i is the number of elements in the ith bin. Similarly, the joint entropy of (x, y) can be calculated as:

$$H(x, y) = -\sum_{i=1}^{b} \sum_{j=1}^{b} \frac{n_{ij}}{N} \log_b \frac{n_{ij}}{N} \tag{3}$$

Considering NCC only evaluates the mutual information between two variables, which cannot indicate the impact of data coverage ratio of each product, here we proposed to weight NCC by the data coverage ratio of each product to account for the contribution of different data coverage, and the new indicator is thus termed as weighted NCC (WNCC) which can be formulated as:

$$WNCC = w * NCC \tag{5}$$

where w is the valid data coverage ratio of each product. Therefore, both data coverage ratio and mutual information are accounted for to better indicate the contribution of each product to the final fused dataset.

References

Che, Y., Xue, Y., Mei, L., Guang, J., She, L., Guo, J., Hu, Y., Xu, H., He, X., Di, A., Fan, C., 2016. Technical note: Intercomparison of three AATSR Level 2 (L2) AOD products over China. Atmos. Chem. Phys. 16, 9655–9674. https://doi.org/10.5194/acp-16-9655-2016

Chen, C., Dubovik, O., Fuertes, D., Litvinov, P., Lapyonok, T., Lopatin, A., Ducos, F., Derimian, Y., Herman, M., Tanré, D., Remer, L. A., Lyapustin, A., Sayer, A. M., Levy, R. C., Hsu, N. C., Descloitres, J., Li, L., Torres, B., Karol, Y., Herrera, M., Herreras, M., Aspetsberger, M., Wanzenboeck, M., Bindreiter, L., Marth, D., Hangler, A. and Federspiel, C.: Validation of GRASP algorithm product from POLDER/PARASOL data and assessment of multi-angular polarimetry potential for aerosol monitoring, Earth Syst. Sci. Data, 12(4), 3573–3620, doi:10.5194/essd-12-3573-2020, 2020.

de Leeuw, G., Sogacheva, L., Rodriguez, E., Kourtidis, K., Georgoulias, A. K., Alexandri, G., Amiridis, V., Proestakis, E., Marinou, E., Xue, Y. and Van Der A, R.: Two decades of satellite observations of AOD over mainland China using ATSR-2, AATSR and MODIS/Terra: Data set evaluation and large-scale patterns, Atmos. Chem. Phys., 18(3), 1573–1592, doi:10.5194/acp-18-1573-2018, 2018.

Diner, D.J., Martonchik, J. V., Kahn, R.A., Pinty, B., Gobron, N., Nelson, D.L., Holben, B.N., 2005. Using angular and spectral shape similarity constraints to improve MISR aerosol and surface retrievals over land. Remote Sens. Environ. 94, 155–171. https://doi.org/10.1016/j.rse.2004.09.009

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A. and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, Atmos. Meas. Tech., 4(5), 975–1018, doi:10.5194/amt-4-975-2011, 2011.

Formenti, P., Mbemba Kabuiku, L., Chiapello, I., Ducos, F., Dulac, F. and Tanré, D.: Aerosol optical properties derived from POLDER-3/PARASOL (2005–2013) over the western Mediterranean Sea – Part 1: Quality assessment with AERONET and in situ airborne observations, Atmos. Meas. Tech., 11(12), 6761–6784, doi:10.5194/amt-11-6761-2018, 2018.

Garay, M. J., Witek, M. L., Kahn, R. A., Seidel, F. C., Limbacher, J. A., Bull, M. A., Diner, D. J., Hansen, E. G., Kalashnikova, O. V., Lee, H., Nastan, A. M. and Yu, Y.: Introducing the 4.4 km spatial resolution Multi-Angle Imaging SpectroRadiometer (MISR) aerosol product, Atmos. Meas. Tech., 13(2), 593–628, doi:10.5194/amt-13-593-2020, 2020.

Goldberg, D. L., Gupta, P., Wang, K., Jena, C., Zhang, Y., Lu, Z. and Streets, D. G.: Using gap-filled MAIAC AOD and WRF-Chem to estimate daily PM2.5 concentrations at 1 km resolution in the Eastern United States, Atmos. Environ., 199(November 2018), 443–452, doi:10.1016/j.atmosenv.2018.11.049, 2019.

Guo, J., Gu, X., Cheng, T., Xie, D. and Chen, H.: Regional trend analysis of the aerosol optical depth comparing to MODIS and MISR aerosol products, in 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 3654–3657, IEEE., 2012.

Jackson, J. M., Liu, H., Laszlo, I., Kondragunta, S., Remer, L. A., Huang, J. and Huang, H.-C.: Suomi-NPP VIIRS aerosol algorithms and data products, J. Geophys. Res. Atmos., 118(22), 12,673-12,689, doi:10.1002/2013JD020449, 2013.

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

Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I. and Korkin, S.: Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables, J. Geophys. Res., 116(D3), D03210, doi:10.1029/2010JD014985, 2011.

Lyapustin, A., Wang, Y., Korkin, S. and Huang, D.: MODIS Collection 6 MAIAC algorithm, Atmos. Meas. Tech., 11(10), 5741–5765, doi:10.5194/amt-11-5741-2018, 2018.

Sun, Z., Chang, N. Bin, Chen, C. F., Mostafiz, C. and Gao, W.: Ensemble learning via higher order singular value decomposition for integrating data and classifier fusion in water quality monitoring, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 14, 3345–3360, doi:10.1109/JSTARS.2021.3055798, 2021.

Tan, Y., Li, E., Zhang, Z., Lin, X., Chi, Y., Zhou, L., Wu, C. and Wang, Q.: Validation of POLDER-3/GRASP aerosol products using AERONET measurements over China, Atmos. Environ., 215(April), 116893, doi:10.1016/j.atmosenv.2019.116893, 2019.

Wang, Q., Shen, Y., and Zhang, J. Q.: A nonlinear correlation measure for multivariable data set, Phys. D, 3–4, 287–295, doi:10.1016/j.physd.2004.11.001, 2005.

Wen, Y. N., Che, Y. H., Guang, J., Xie, Y. Q., Shi, Z., Zhang, Y. and Li, Z. Q.: VALIDATION of AEROSOL PRODUCTS from ESA/AATSR over CHINA and AOD FUSION BASED on UNCERTAINTIES, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch., 42(3/W9), 181–185, doi:10.5194/isprs-archives-XLII-3-W9-181-2019, 2019.

Witek, M. L., Garay, M. J., Diner, D. J., Bull, M. A. and Seidel, F. C.: New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water, Atmos. Meas. Tech., 11(1), 429–439, doi:10.5194/amt-11-429-2018, 2018.

Xiao, Q., Zhang, H., Choi, M., Li, S., Kondragunta, S., Kim, J., Holben, B., Levy, R.C., Liu, Y., 2016. Evaluation of VIIRS, GOCI, and MODIS Collection 6 AOD retrievals against ground sunphotometer observations over East Asia. Atmos. Chem. Phys. 16, 1255–1269. https://doi.org/10.5194/acp-16-1255-2016

Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A. and Liu, Y.: Full-coverage high-resolution daily PM2.5 estimation using MAIAC AOD in the Yangtze River Delta of China, Remote Sens. Environ., 199(March), 437–446, doi:10.1016/j.rse.2017.07.023, 2017.

Xie, Y., Xue, Y., Che, Y., Guang, J., Mei, L., Voorhis, D., Fan, C., She, L. and Xu, H.: Ensemble of ESA/AATSR Aerosol Optical Depth Products Based on the Likelihood Estimate Method with Uncertainties, IEEE Trans. Geosci. Remote Sens., 56(2), 997–1007, doi:10.1109/TGRS.2017.2757910, 2018.

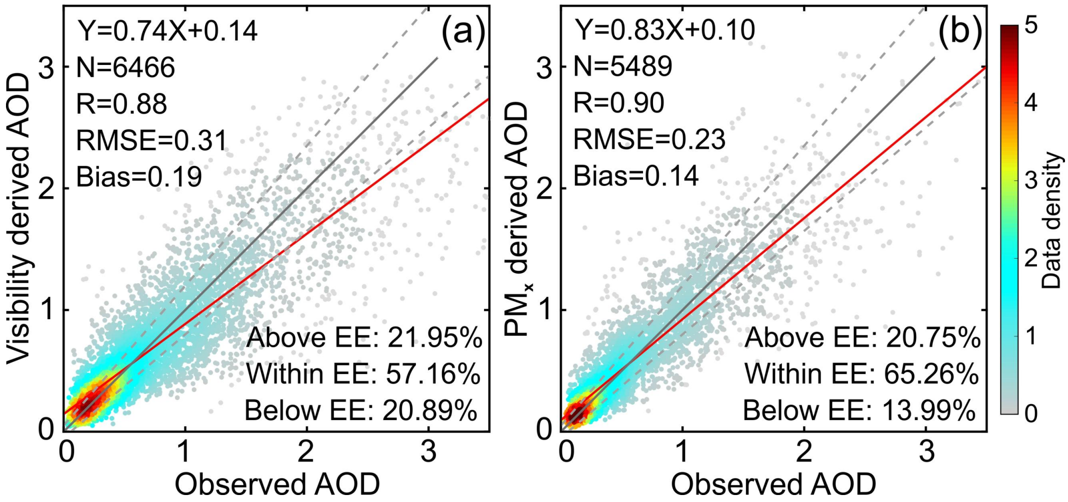

Figure S1. Scatter plots between ground AOD observations and AOD inferred from (a) atmospheric visibility and (b) air pollutants concentration.

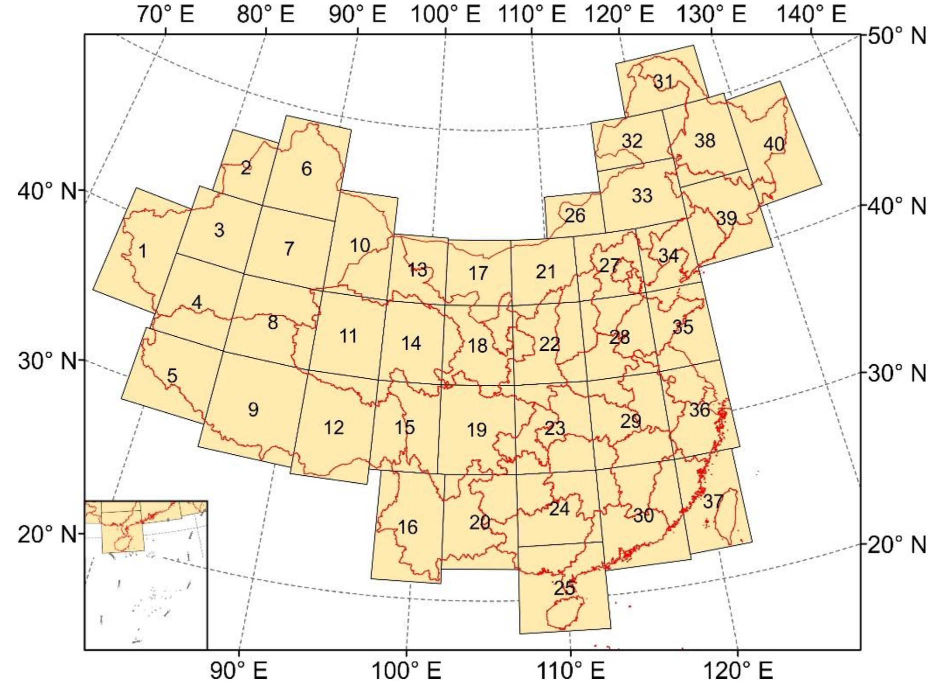

Figure S2. Spatial distribution of 40 subregions that was divided to facilitate AOD gap filling in the mainland China.

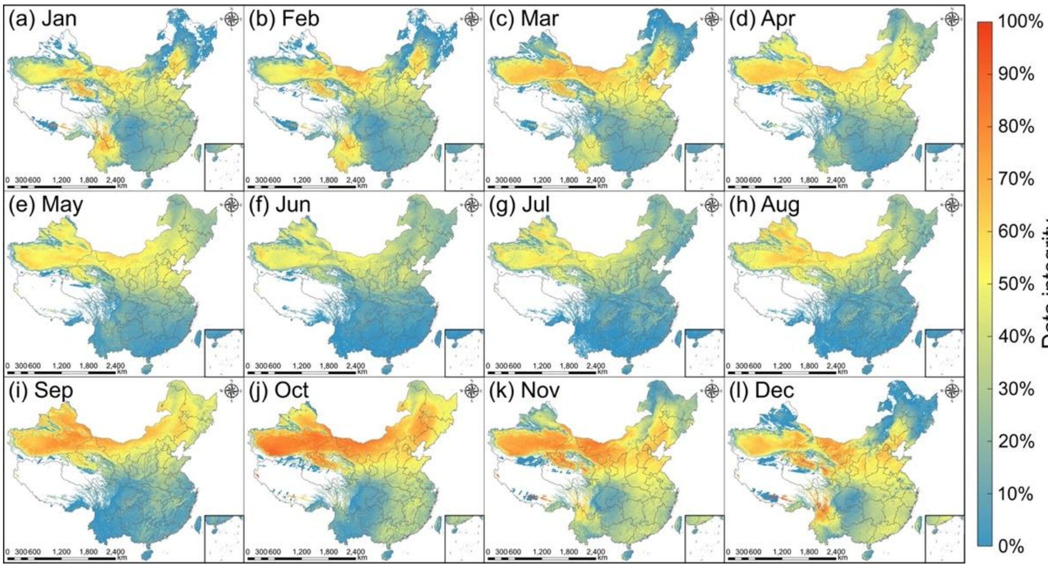

Figure S3. Monthly mean data integrity of AOD from Terra/MODIS in China during the period of 2000 to 2020.

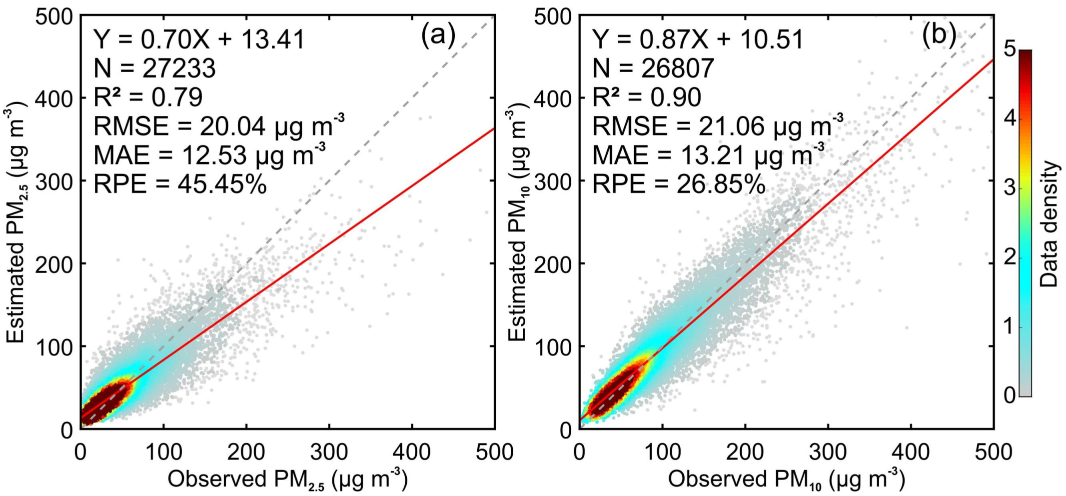

Figure S4. Sample-based cross validation accuracy of random forest models for (a) $PM_{2.5}$ and (b) PM_{10} mapping.

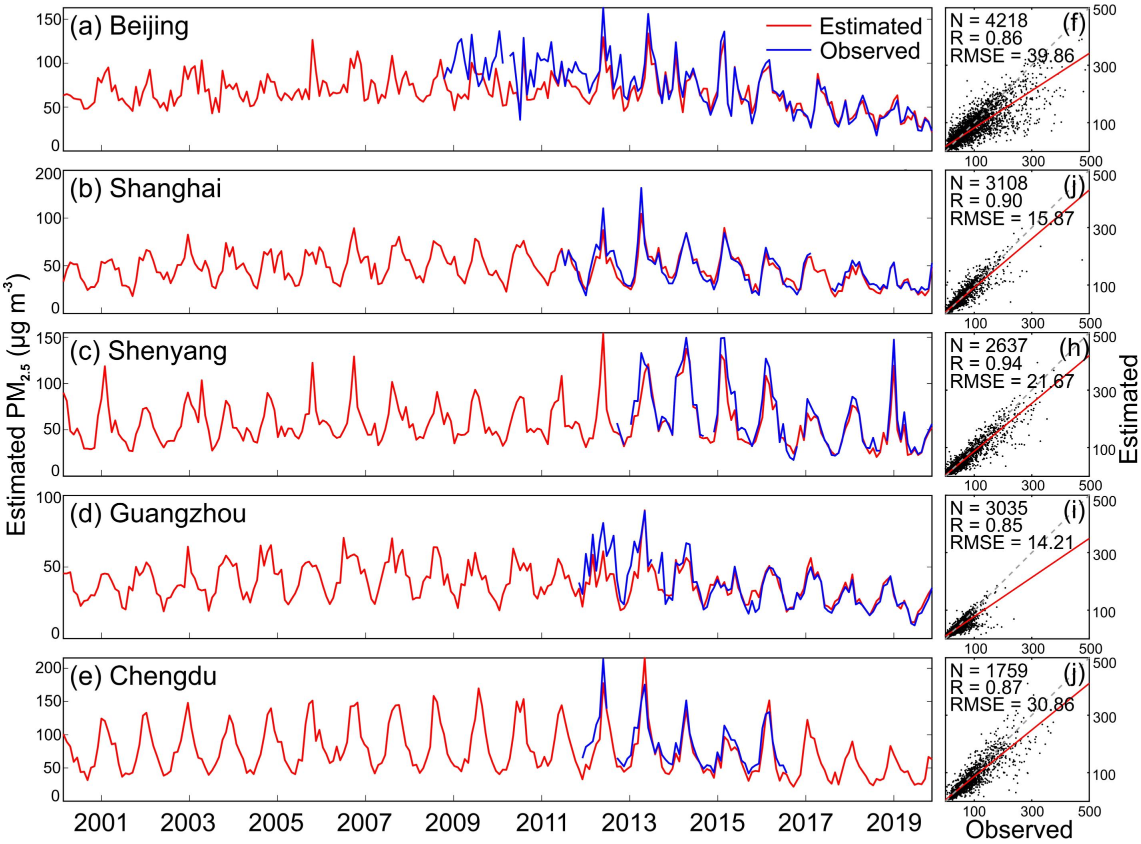

Figure S5. Comparison of SCHAP $PM_{2.5}$ time series with measurements from United States Embassy in China. (a–e) Temporal variations of monthly $PM_{2.5}$ concentration and (f–j) the associated daily scatter plots.

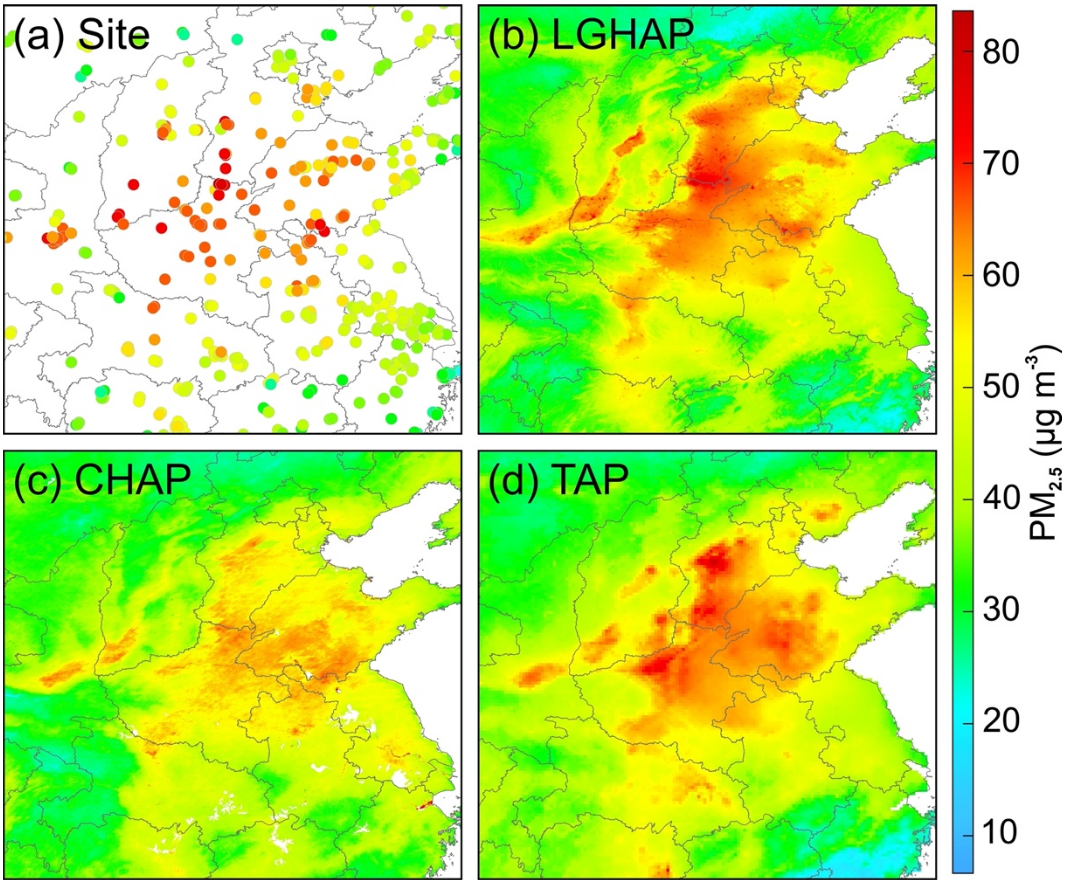

Figure S6. Comparison of spatial distribution of annual mean $PM_{2.5}$ concentration in 2019 in eastern China.

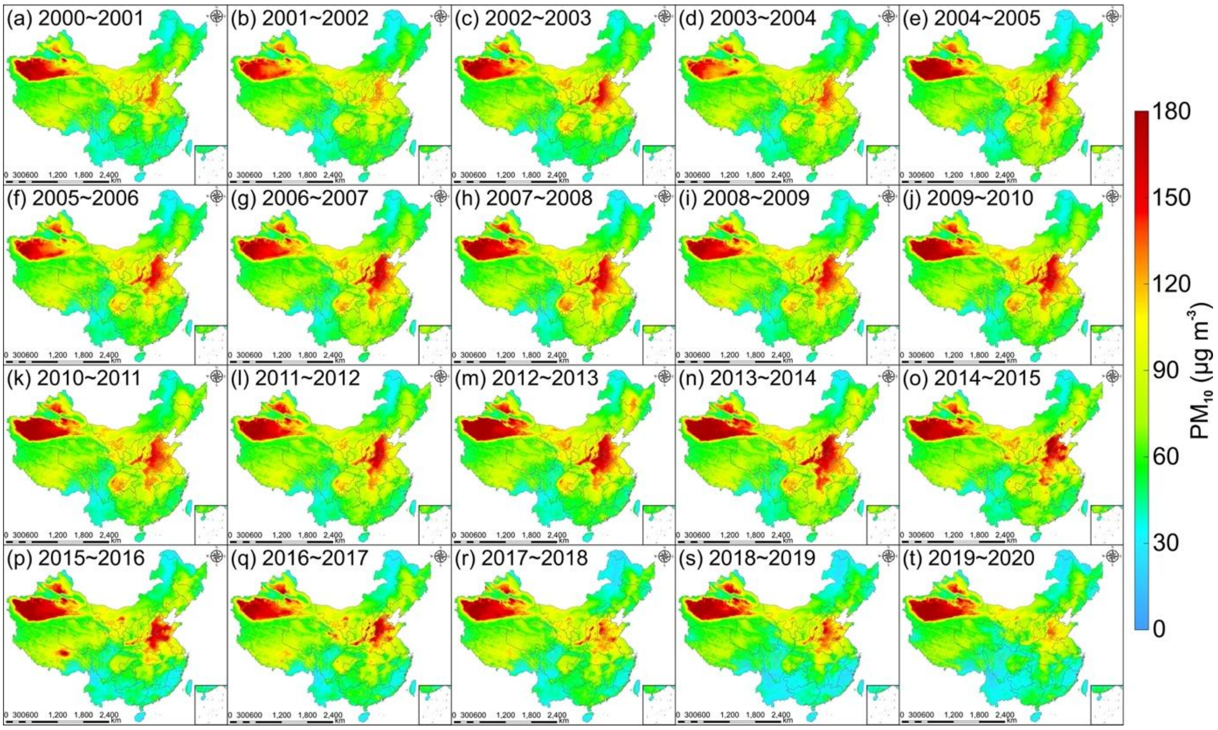

Figure S7. Spatial distribution of wintertime (September to February) averaged PM_{10} concentration from SCHAP during 2000 to 2020 in China.

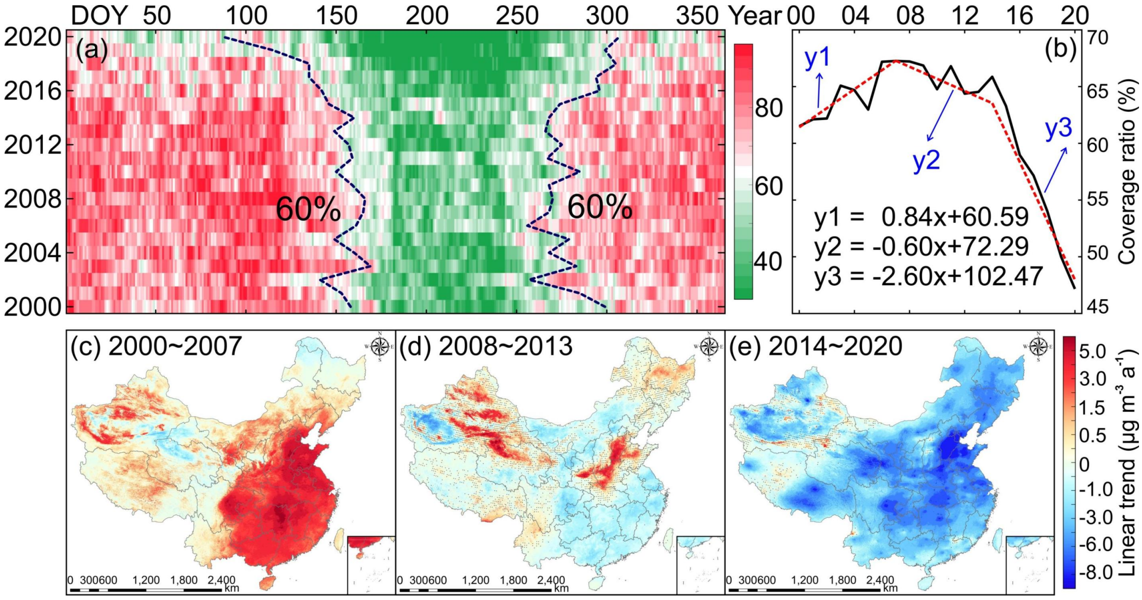

Figure S8. Temporal variations of the proportion of land areas covered with PM_{10} concentration exceeding 50 μg m^{-3} and PM_{10} trends during three different periods. (a) Temporal variations of the land coverage ratio with daily PM_{10} concentration exceeding 50 μg m^{-3} from 2000 to 2000. (b) same as (a) but for annual mean PM_{10} concentration. (c–e) PM_{10} trends during periods of 2000–2007, 2008–2013, and 2014–2020. The dotted regions imply trend estimations are statistically insignificant at the 95% confidence interval.

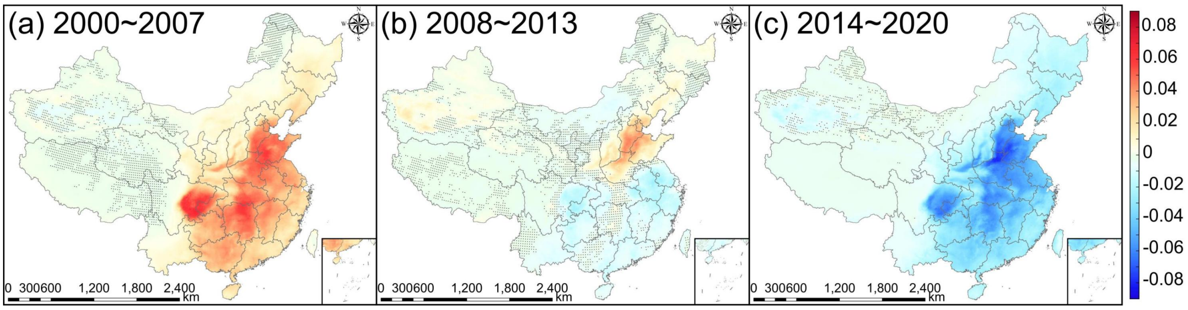

Figure S9. AOD trends during periods of (a) 2000–2007, (b) 2008–2013, and (c) 2014–2020. The dotted regions imply trend estimations are statistically insignificant at the 95% confidence interval.