Supplementary material

Identifying $PM_{2.5}$-Related Health Burden in the Context of the Integrated Development of Urban Agglomeration Using Remote Sensing and GEMM Model

Lili Xu [1,2], Binjie Chen [3], Chenhao Huang [1,2], Mengmeng Zhou [4], Shucheng You [5], Fangming Jiang [1,2], Weirong Chen [1,2] and Jinsong Deng [1,2,*]

[1] College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China; 22114126@zju.edu.cn (L.X.); 3180100429@zju.edu.cn (C.H.); 12314004@zju.edu.cn (F.J.); 22214137@zju.edu.cn (W.C.)
[2] Zhejiang Ecological Civilization Academy, Huzhou 313300, China
[3] Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China; chenbinjie@nbu.edu.cn
[4] School of Business, Changzhou University, Changzhou 213159, China; mm_zhou2022@cczu.edu.cn
[5] Resource Investigation and Monitoring Department, Land Satellite Remote Sensing Application Center, MNR, Beijing 100048, China; yousc@lasac.cn
* Correspondence: jsong_deng@zju.edu.cn; Tel.: +86-571-8898-2623

Table of Contents

Figure S1. The annual $PM_{2.5}$ concentrations in the Yangtze River Delta from 2015 to 2019.

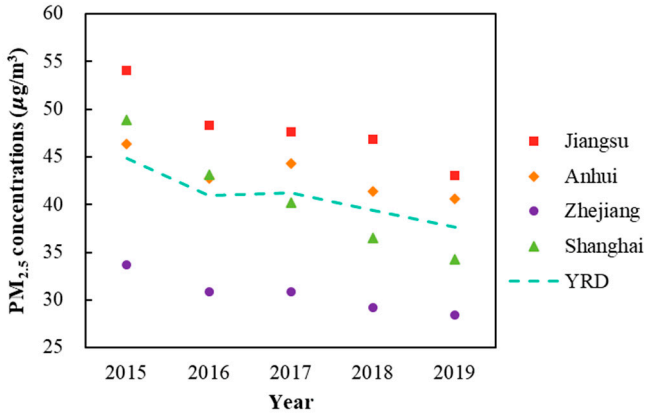


Figure S2. The spatiotemporal variations of annual $PM_{2.5}$ concentrations in the Yangtze River Delta from 2015 to 2019.

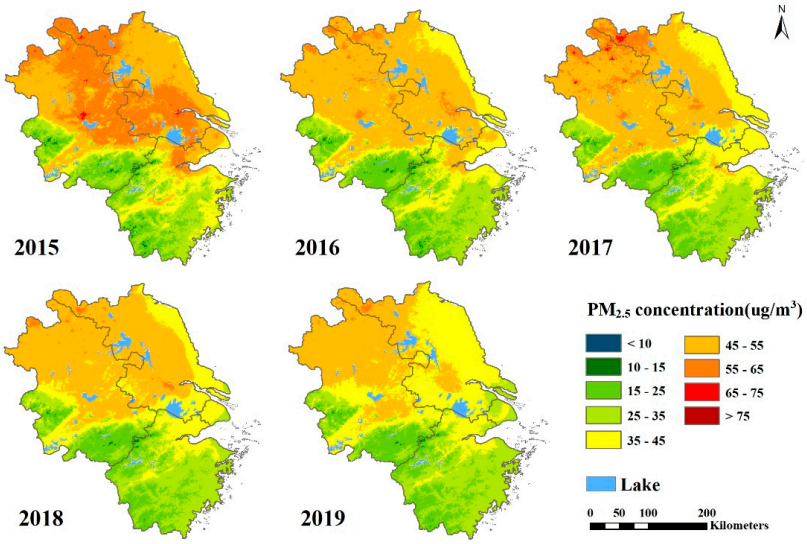


Figure S3. The change in annual PM2.5 concentrations from 2015 to 2019. The purple oval shows the area of most significant change.

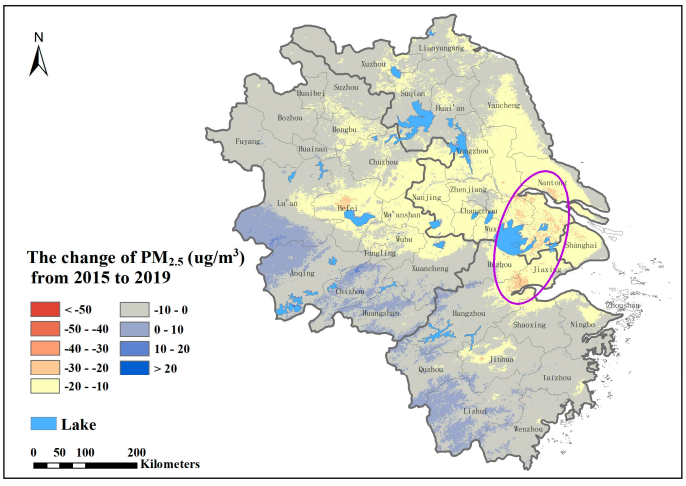


Figure S4. Spatial distribution of $PM_{2.5}$-related premature mortality in the Yangtze River Delta in 2015.

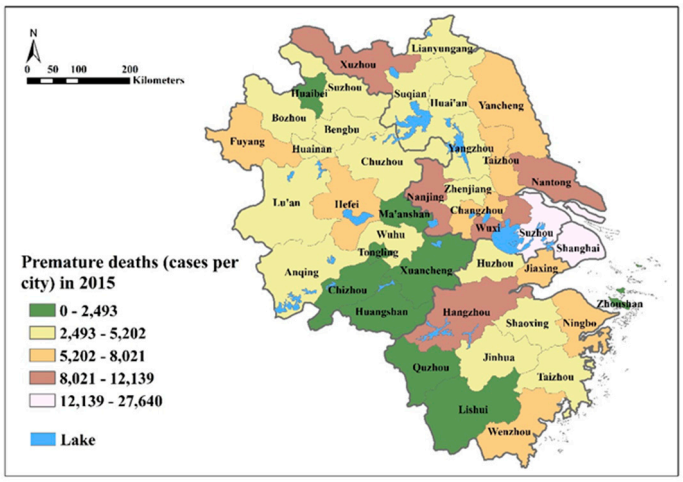


Figure S5. Spatial distribution of $PM_{2.5}$-related premature mortality in the Yangtze River Delta in 2019.

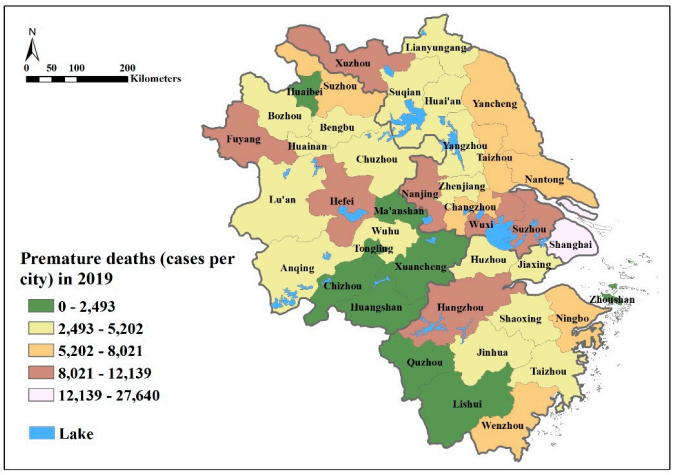
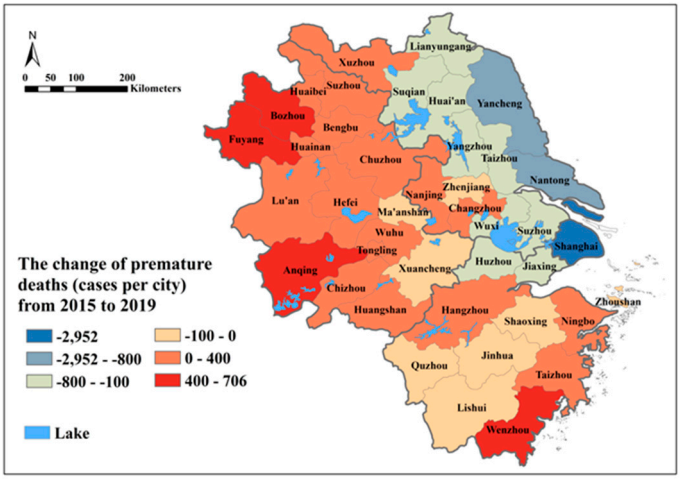


Figure S6. The changes in $PM_{2.5}$-related premature mortality in the Yangtze River Delta from 2015 to 2019.



1. Details of the procedure of estimated $PM_{2.5}$ data by SADNN model

The daily 1-km MAIAC AOD products retrieved from MODIS Terra and Aqua satellites, DEM data, NDVI data, and ERA-5 hourly meteorological data including the 2-meter temperature (T2M), surface pressure (SP), boundary layer height (BLH), relative humidity (RH), 10-meter U wind component (WU10M), and 10-meter V wind component (WV10M), were used to estimate daily $PM_{2.5}$ concentrations across the YRD region from 2015 to 2019. Table S1 summarizes the details of above datasets and their sources of acquisition.

The procedure included three main steps. First, due to large AOD data gaps in the MAIAC AOD products, the data gaps were imputed to obtain full-coverage AOD data. We first fused the Terra and Aqua AOD products by fitting the Terra and Aqua AOD products through linear regression on a monthly scale. After combining the Terra and Aqua AOD products, an RF model was developed to impute the remaining MAIAC AOD data gaps. The AOD gap-filling RF model, referred to as RF-AOD model, is expressed as

$$AOD_{st} = f(AOD_{CAMS_{st}}, DEM_s, NDVI_{st}, T2M_{st}, SP_{st}, BLH_{st}, RH_{st}, WU10M_{st}, WV10M_{st}, LON_s, LAT_s, DOY_t) \quad (1)$$

where f () represents the RF algorithm, s denotes the location of a grid and t is the day. The dependent variable is the AOD to be imputed. The independent variables include CAMS AOD, DEM, NDVI, T2M, SP, BLH, RH, WU10M, WV10M, longitude (LON), latitude (LAT) and day of the year (DOY). Validated against the AERONET AOD data, the imputed AOD data had a high correlation coefficient (R) value of 0.89 and a low RMSE value of 0.24.

Second, an interpretable self-adaptive deep neural network (SADNN) model, incorporating AOD, meteorological and other auxiliary predictors, was developed to estimate daily spatially-continuous $PM_{2.5}$ concentrations. The SADNN model was developed based on the standard DNN by incorporating the attention module after the input layer (Figure S7). After the attention module, several hidden layers were connected before the output layer. The number of hidden layers and the number of hidden neurons in each hidden layer were determined through cross-validation (CV). Based on the trade-off between model accuracy and efficiency, we set the number of hidden layers to 6, and the number of neurons in each hidden layer to 512 in the SADNN model. The SADNN model incorporating the predictors is expressed as

$$PM_{2.5_{st}} = f(AOD_{imputed_{st}}, DEM_s, NDVI_{st}, T2M_{st}, SP_{st}, BLH_{st}, RH_{st}, WU10M_{st}, WV10M_{st}, LON_s, LAT_s, DOY_t) \quad (2)$$

where f () represents the SADNN algorithm, s denotes the location of a grid and t is the day. The dependent variable is the estimated $PM_{2.5}$ concentrations. The independent variables include the imputed AOD, DEM, NDVI, T2M, SP, BLH, RH, WU10M, WV10M, LON, LAT and DOY.

A total of 2,998,459 pieces of valid data for the training and validation of the SADNN model was obtained. We executed five-fold sample-based and site-based CV approaches to validate the robustness of the model. The dataset was randomly split into five subsets, with each subset containing approximately 20% of the samples for sample-based CV and 20% of the ground sites for site-based CV. Then one of the subsets was treated as the testing data, and the model was fitted with the remaining datasets. This process was repeated five times to ensure that all of the data were tested. For the sample (site)-based CV, the model had a high R^2 value of 0.86 (0.84) and low RMSE and MAE values of 13.07 (14.30) $\mu g/m^3$ and 8.23 (8.82) $\mu g/m^3$, respectively. For the SADNN model performance for each $PM_{2.5}$ monitoring site, approximately 80%

of the monitoring stations reported R^2 values higher than 0.70 and RMSE values lower than 15 $\mu g/m^3$, indicating excellent performance in estimating $PM_{2.5}$ concentrations.

Finally, the validated SADNN model, incorporating the imputed AOD, meteorological and other auxiliary data, was used to estimate daily spatially-continuous $PM_{2.5}$ concentrations across the YRD region from 2015 to 2019.

Figure S7. Structure of the SADNN model and the attention module.

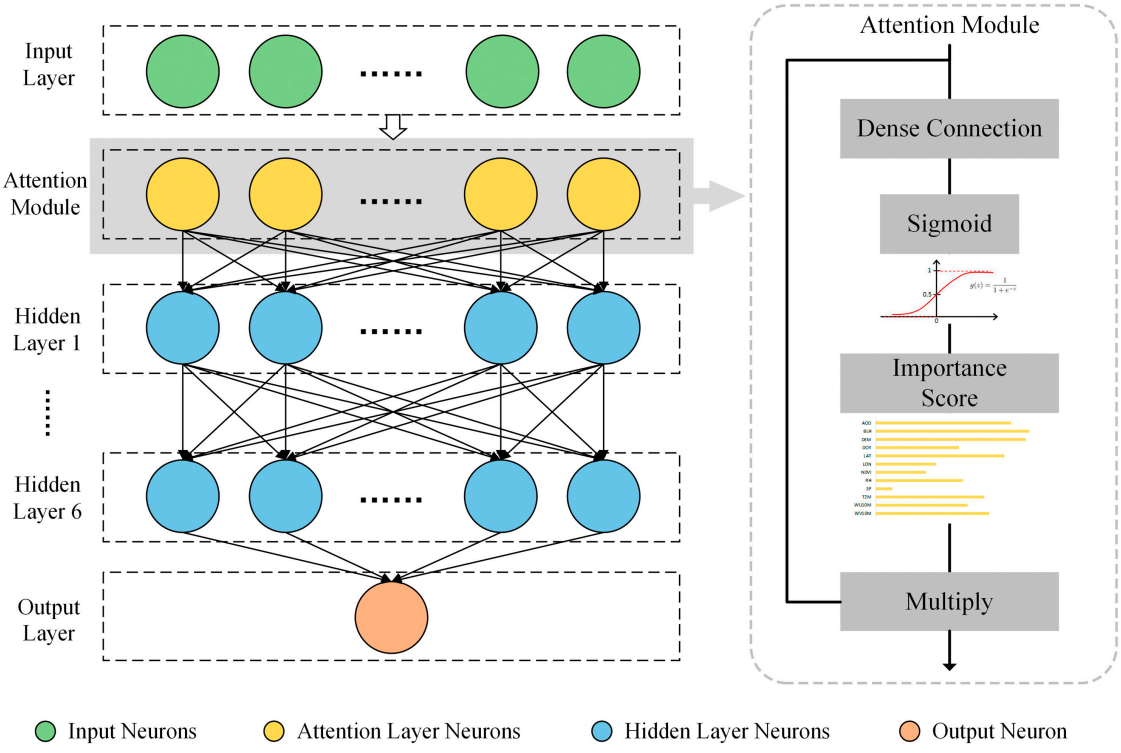


Table S1. Detail description of the datasets used in SADNN model.

Product	Variable	Unit	Spatial Resolution	Temporal Resolution	Source
$PM_{2.5}$	$PM_{2.5}$	ug/m^3	-	Hourly	CNEMC, http://www.cnemc.cn/
AERONET	AERONET AOD	-	-	Daily	AERONET, https://aeronet.gsfc.nasa.gov/
MCD19A2	MAIAC AOD	-	1km	Daily	NASA, https://ladsweb.modaps.eosdis.nasa.gov/
CAMS	CAMS AOD	-	0.125°	3-Hour	ECMWF, https://ads.atmosphere.copernicus.eu/
ERA-5	2-meter Temperature	K	0.125°	Hourly	ECMWF, https://cds.climate.copernicus.eu/
	Surface Pressure	hPa	0.125°	Hourly	
	10-meter U-Wind Component	m/s	0.125°	Hourly	
	10-meter V-Wind Component	m/s	0.125°	Hourly	
	Boundary Layer Height	m	0.125°	Hourly	
	Relative Humidity	%	0.125°	Hourly	
ASTER GDEM V1	DEM	m	30m	-	GSCloud, http://www.gscloud.cn
MOD13A3	NDVI	-	1km	Monthly	NASA, https://ladsweb.modaps.eosdis.nasa.gov/

2. Spatial-temporal variations of $PM_{2.5}$ concentrations

Figure S1 shows the spatial-temporal changes in $PM_{2.5}$ concentrations at the provincial level in the YRD region from 2015 to 2019. In 2019, annual $PM_{2.5}$ concentrations in the YRD region were 37.64 μg/m^3, 34.3 μg/m^3, 43.04 μg/m^3, 40.56 μg/m^3 and 28.43 μg/m^3 in Shanghai Municipality, Jiangsu Province, Zhejiang Province, and Anhui Province, respectively. From 2015 to 2019, an overall decrease in $PM_{2.5}$ concentrations could be found, indicating that the JPCAP policy had a significant effect. Overall, annual $PM_{2.5}$ concentrations in the YRD region decreased by 16.13%, from 44.88 μg/m^3 in 2015 to 37.64 μg/m^3 in 2019. Among provinces, annual $PM_{2.5}$ concentrations have decreased the most with a reduction of 14.55 μg/m^3 in Shanghai Municipality, followed by 11.00 μg/m^3, 5.75μg/m^3, and 5.26 μg/m^3 in Jiangsu Province, Zhejiang Province, and Anhui Province, respectively. The five-year percentage decline of $PM_{2.5}$ concentrations was 29.79%, 20.36%, 15.61%, and 12.42% for Shanghai Municipality, Jiangsu Province, Zhejiang Province, and Anhui Province, respectively.

Figure S2 further shows the detailed spatiotemporal variations of $PM_{2.5}$ concentrations in the YRD region from 2015 to 2019. The spatial patterns of $PM_{2.5}$ concentrations demonstrated distinct geographical features. The worst polluted areas mainly occurred in the northern YRD region. At the same time, better air quality could be found in the eastern coastal cities and the southwestern mountainous areas characterized by favorable climatic conditions and high forest coverage. In addition, the changing pattern also varied by different years. From 2015 to 2016, Anhui Province and Jiangsu Province showed a substantial reduction in annual $PM_{2.5}$ concentrations, improving large areas of polluted regions. However, in 2017, there was a rebound of $PM_{2.5}$ concentrations in the north. From 2017 to 2019, the east coast regions with less $PM_{2.5}$ pollution radiated inland. The less polluted areas expanded apparently, resulting in three pollution levels distributed in roughly three equal parts: relatively economically undeveloped regions of the north, the economically developed eastern seaboard, and the mountainous area in the southwest.

Generally, the air quality in the YRD region has improved significantly from 2015 to 2019, as illustrated in Figure S3. The most significant drops (> 30.0 μg/m^3) could be found in the intersection of Shanghai Municipality, Jiangsu province, and Zhejiang province, that was circled with a purple oval. However, some areas experienced no recovery or even worse conditions. For example, Xuzhou, Wuxi in Jiangsu Province, Suzhou, Bozhou, and Hefei in Anhui Province maintained severe air quality. Southwestern mountainous areas in Zhejiang Province and Anhui Province, initially the least polluted regions, had increased $PM_{2.5}$ concentrations from 0 to 10 μg/m^3. Thus, there were noticeable differences in effectiveness among areas under the context of joint air pollution control.

4. GEMM NCD+LRI model

The parametrization in the GEMM NCD+LRI model builds up the relationship between long term $PM_{2.5}$ exposure and health outcomes, thereby relative risk (RR) of NCD+LRI on concentration(C) was first calculated:

$$RR(C) = e^{\frac{\theta \times \ln\left(\frac{z}{\alpha}+1\right)}{1+e^{\left(-\frac{z-\mu}{v}\right)}}}, where\ z = \max(0, C - 2.4),$$

where θ, α, μ, and v designs the pattern of the $PM_{2.5}$-mortality relationships, which are defined in [1]. Further, the fraction of mortality attributed to $PM_{2.5}$ exposure (AF) could be expressed as:

$$AF(C) = \frac{RR(C) - 1}{RR(C)}$$

Next, premature mortality attributed to long-term exposure to $PM_{2.5}$ for the total population group (aged ≥25 years) in grid j were calculated:

$$M_j(C_j) = P_j \times B \times AF(C_j)$$

where P_j represented the total population amount in grid j; B represented the baseline mortality incidence rates of NCD+LRI for a population group (aged ≥25 years), which were selected from the GBD data in the website of Global Health Data Exchange (GHDx) (http://ghdx.healthdata.org/gbd-results-tool); the baseline mortality incidence rates are varied year by year; and $AF(C_j)$ was the fraction of NCD+LRI attributed to long term $PM_{2.5}$ exposure at the level C_j for a population group.

Table S2. Baseline mortality incidence rates of NCD+LRI for 25 plus population group

Years	Baseline mortality	Source
2015	0.008818496	Global Health Data Exchange (GHDx) (http://ghdx.healthdata.org/gbd-results-tool)
2016	0.008917280	
2017	0.008994136	
2018	0.009103251	
2019	0.009308302	

Table S3. Provincial-level statistics of air pollution related exposure factors in this study

Provinces	Respiratory rate (m^3/day)		Time spent outdoors (hour/day)	
	Mean	Std	Mean	Std
Shanghai	16.18	0.95	3.77	0.61
Jiangsu	15.97	0.46	4.78	1.78
Zhejiang	15.53	0.27	4.54	0.54
Anhui	15.65	0.36	4.42	0.51
National average	15.80		4.67	

Table S4. Parameters for GEMM NCD+LRI model in this study

Age Range (years)	θ	With Chinese standard error θ	α	μ	v
25 plus	0.1430	0.01807	1.6	15.5	36.8

References:

[1] Burnett, R.; et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc. Natl. Acad. Sci. U S A **2018**, 115(38), 9592-9597. https://doi.org/10.1073/pnas.1803222115.