Supplementary Information

W. Wang, S. Zhao, L. Jiao, M. Taylor, B. Zhang, G. Xu, H. Hou. Estimation of PM2.5 Concentrations in China Using a Spatial Back Propagation Neural Network.

Table of Contents

Glossary of terms and abbreviations

ANN	Artifical Neural Network
AOD	Aerosol Optical Depth
BPNN	Back-Propagation Neural Network
CAAQS	China Ambient Air Quality Standards
CLA	Construction Land Area
CNEMC	China National Environmental Monitoring Centre
DEM	Digital Elevation Model
IQR	Inter-Quartile Range
MLR	Multiple Linear Regression
MODIS	Moderate Resolution Imaging Spectroradiometer
MPE	Mean Prediction Error
NDVI	Normalized Difference Vegetation Index
PCA	Principal Components Analysis
PM	Particular Matter
R^2	Coefficient of Determination
RH	Relative Humidity
RMSE	Root Mean Squared Error
RPE	Relative Prediction Error
SAR	Spatial Autoregression
S-BPNN	Spatial Back-Propagation Neural Network
SEM	Spatial Error Model
SLM	Spatial Lag Model
SLV	Spatial Lag Variable
SSD	Sunshine Duration
TEOM	Tapered Element Oscillating Microbalance
UFA	Universal Function Approximator
WS	Wind Speed

Supplementary Statistics

The statistical indicators used to measure model performance are based on differences and include correlation coefficient (R^2), the root-mean-square error (RMSE, μg/m^3), the mean prediction error (MPE, μg/m^3), and relative prediction error (RPE, %) defined as follows:

$$R^2 = \frac{\sum_{i=1}^{n}(PM_{o,i} - \overline{PM}_o)^2 (PM_{e,i} - \overline{PM}_e)^2}{\sum_{i=1}^{n}(PM_{o,i} - \overline{PM}_o)^2 \sum_{i=1}^{n}(PM_{e,i} - \overline{PM}_e)^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(PM_{o,i} - \overline{PM}_{e,i})}{n}}$$

$$MPE = \frac{\sum_{i=1}^{n} | PM_{o,i} - \overline{PM}_{e,i} |}{n}$$

$$RPE = \frac{RMSE}{\overline{PM_o}}$$

where n is the total number of data records, and PM_o, PM_e are the observed $PM_{2.5}$ concentration and model-estimated $PM_{2.5}$ concentrations, respectively. $\overline{PM_o}, \overline{PM_e}$ are the mean $PM_{2.5}$ concentrations of observation and model-estimated, respectively.

Supplementary Table

Table S1. Data sources, spatial scales and units for each variable in the sample data set.

Category	Variable	Units	Spatial scale	Data source
Measurements	$PM_{2.5}$	μg/m^3	N/A	Official database of the China Environmental Monitoring Center (CEMC): http://106.37.208.233:20035/
Satellite data	AOD (550nm)	N/A	10 km	Atmospheric Archive and Distribution System (LAADS Web): http://ladsweb.nascom.nasa.gov/data/search.html
Synoptic conditions	WS	m/s	N/A	Atmospheric Archive and Distribution System (LAADS Web (http://data.cma.cn/)
	RH	%	N/A	
	Pressure	Pa	N/A	
	Temperature	℃	N/A	
	Precipitation	mm	N/A	
	SSD	h	N/A	
PM2.5 source emission data	Construction land area	km^2	10 km	GIMCP: http://www.dsac.cn/
	Road length	km	10 km	OpenStreetMap: www.openstreetmap.org/
	NDVI	km	1km	Resource and Environment Data Cloud Platform: http://www.resdc.cn/
	Population density	people	1 km	LandScan: https://web.ornl.gov/sci/landscan/
	Pollution sources		N/A	MEP website: http://www.mep.gov.cn/
	DEM base height	m	30 m	Geospatial Data Cloud: http://www.gscloud.cn/

Supplementary Figures

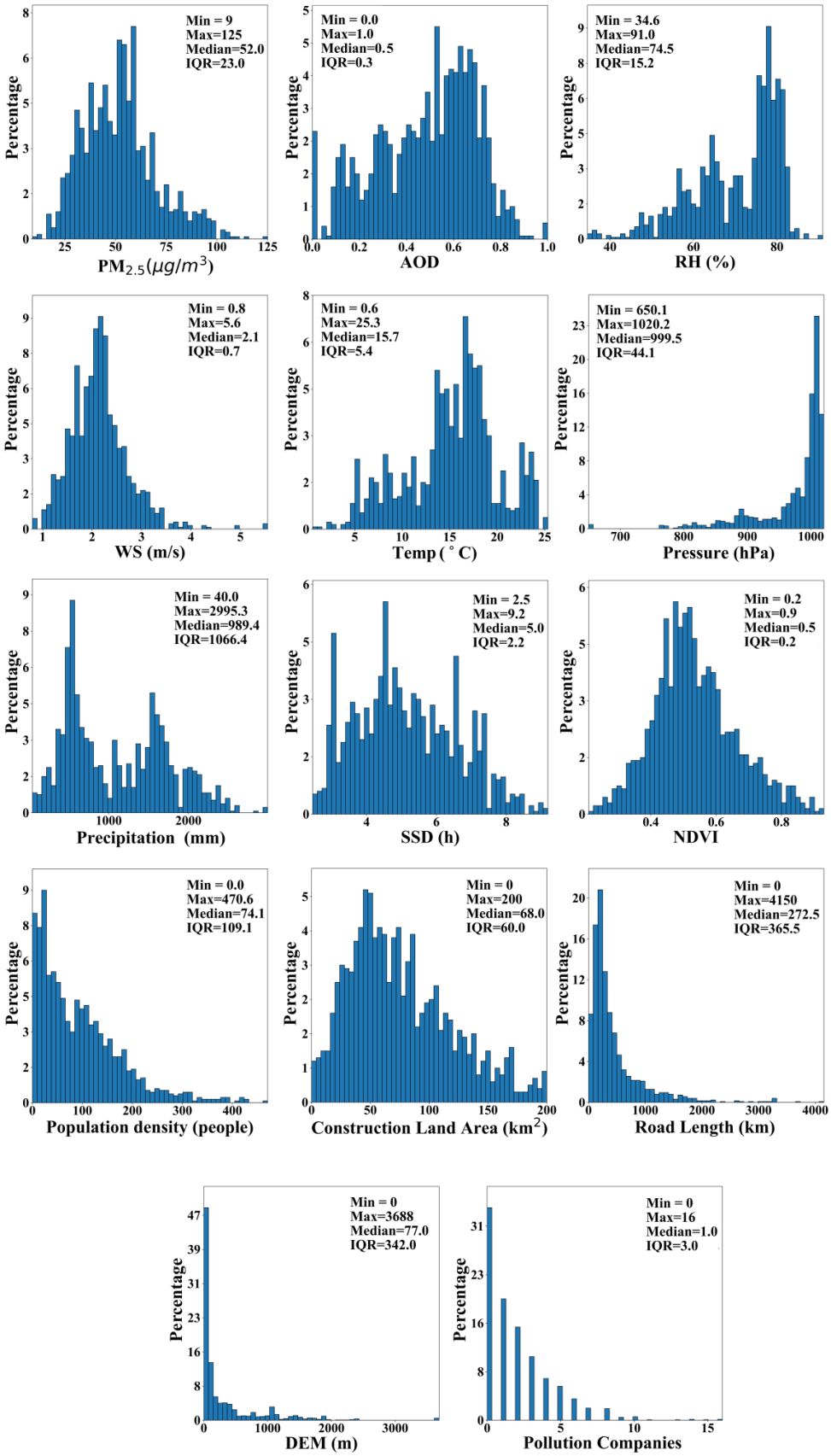

Figure S1. Histogram and median statistics of the candidate explanatory variables used in construction of the S-BPNN model. The sample data is drawn from N = 1280 monitoring sites across China.

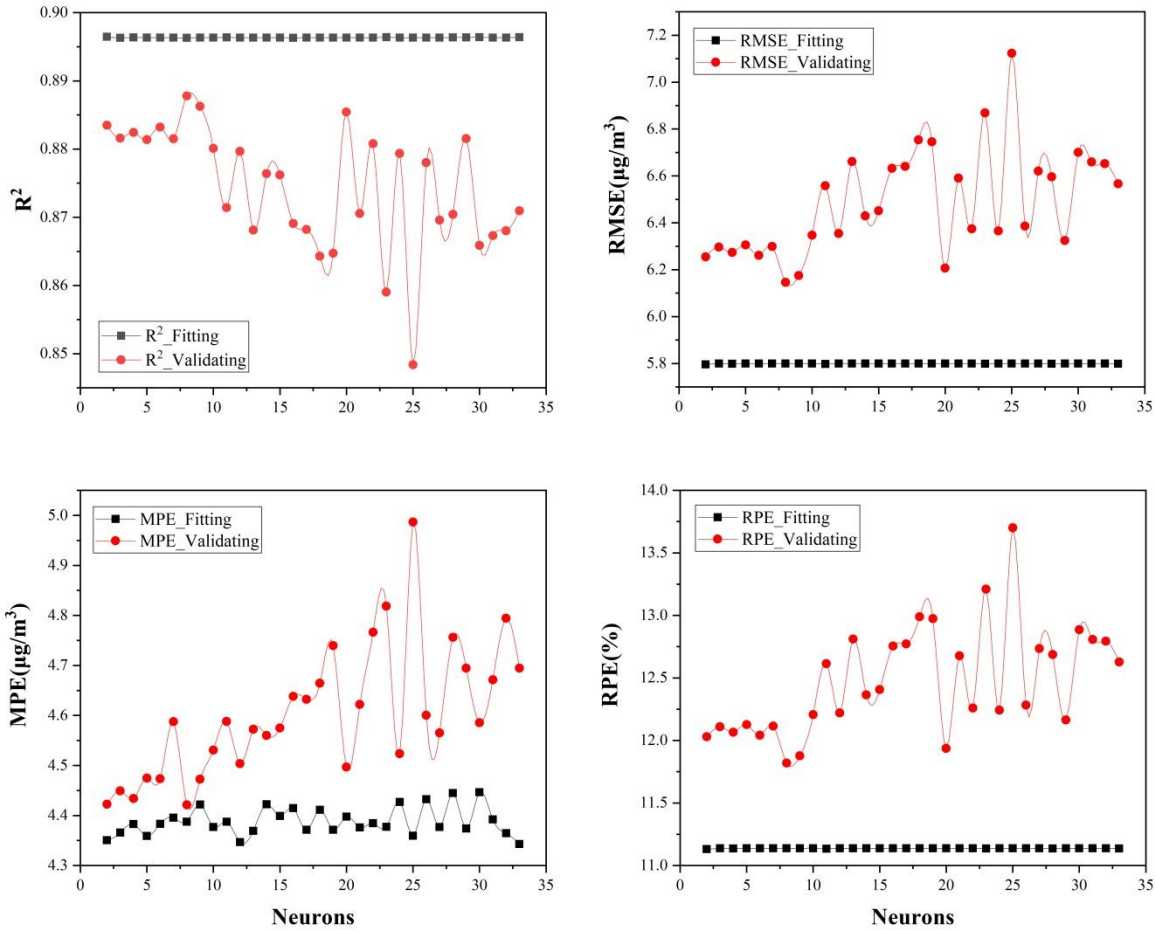

Figure S2. S-BPNN model performance as a function of the number of neurons in the hidden layer.