Supplementary Materials

Satellite-based mapping of high-resolution ground-level $PM_{2.5}$ with VIIRS IP AOD in China through spatially neural network weighted regression

Yijun Chen[1], Sensen Wu[1, 3*], Yuanyuan Wang[2], Feng Zhang[1, 3], Renyi Liu[1, 3], and Zhenhong Du[1, 3]

[1] School of Earth Sciences, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China.

[2] Ocean Academy, Zhejiang University, Zhoushan, China.

[3] Zhejiang Provincial Key Laboratory of Geographic Information Science, 148 Tianmushan Road, Hangzhou 310028, China.

† Corresponding author: Sensen Wu (wusensengis@zju.edu.cn)

Contents of this file

Number SI pages: 15

Number Figures: 7

Number Tables: 4

Supplementary Methods

Neural network architecture optimization of GNNWR model

A five-layer spatial weighted neural network (SWNN) architecture of the GNNWR model was adopted in this study to estimate $PM_{2.5}$ concentrations, which included one input layer, one output layer, and three hidden layers. The input layer of the SWNN was the distances from the estimated point to all training points so that its length was equal to the number of training points. The output layer of the SWNN was the non-stationary weights of the independent variables so that its length was the count of independent variables.

Hence, the neuron sizes of the three hidden layers in SWNN should be determined. The neuron size of one layer is commonly less than that of its previous layer and greater than that of its subsequent layer. In addition, neuron size is usually a number of power 2 for superior efficiency. Therefore, a simple search strategy was conducted in which the optimal neuron sizes of three hidden layers were selected from the following lists: [512, 128], [128, 64, 32], and [32, 16, 8], respectively. Accordingly, 18 models are developed and compared using two statistical measures (i.e., R^2 and RMSE), and [512, 128, 32] was found to be the best architecture of the three hidden layers for the GNNWR of $PM_{2.5}$ estimations in our experiments.

Training and validation procedures of the GNNWR model

The training and validation procedures of the GNNWR model for $PM_{2.5}$-AOD modeling are shown in Figure S2.

Step 1: The datasets including $PM_{2.5}$ and independent variables were randomly divided into two subsets: training-validation dataset (85%) and testing dataset (15%). The training-validation dataset was split into 10 smaller sets (i.e., 10 folds). Then, the data of 9 folds were used as training dataset for model training and the remaining fold data was used to validate the obtained model.

Step 2: The OLS coefficients denoting the average relationship were pre-calculated using the training dataset. The neural network parameters of GNNWR were initialized, such as the maximal iterations, neuron sizes, learning rate, and batch size.

Step 3: Obtain the first or next mini-batch of data from the training dataset.

Step 4: Compute the loss function (mean square error, MSE) and update the neural network weights and biases by executing a gradient descent algorithm on the mini-batch data.

Step 5: Check whether an epoch was completed. If an epoch was not completed, obtain the next mini-batch of data (Step 3) and repeat Step 4. If an epoch was completed, proceed to the next step.

Step 6: Conduct early stopping strategy and check the number of training epochs. The MSE value of the validation dataset was calculated to judge overfitting. During the training process, the best validation accuracy and its corresponding model parameters were recorded. When the MSE value of the validation dataset did not decrease for a predefined threshold of continuous epochs, the model was considered overfitted, and the training process would be stopped. Then, the last recorded best model would be chosen as the obtained model. In addition, if the maximum number of training epochs was reached, we also stopped the training process.

Otherwise, we disordered the training dataset and updated the epoch number. Then, go back to Step 3.

Step 7: The above training process was repeated for 10 iterations until all folds were validated. The final model was built with the best model parameters achieve by the 10-fold cross-validation with all training-validation samples.

Step 8: The testing dataset was used for the external evaluation of the final model.

Supplementary Figures

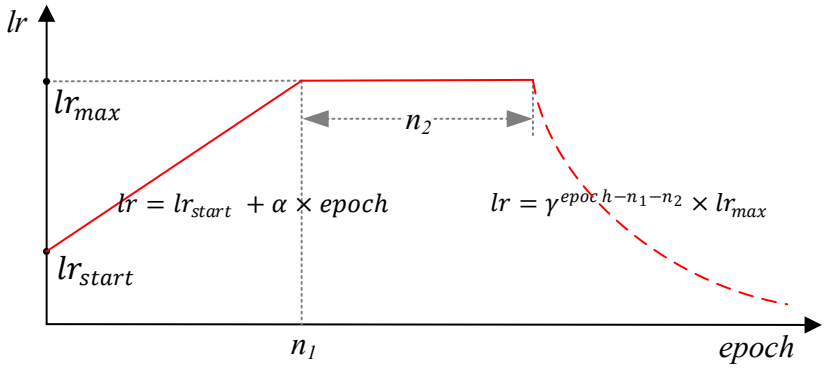


Figure S1. Adaptive learning rate strategy for the GNNWR model.

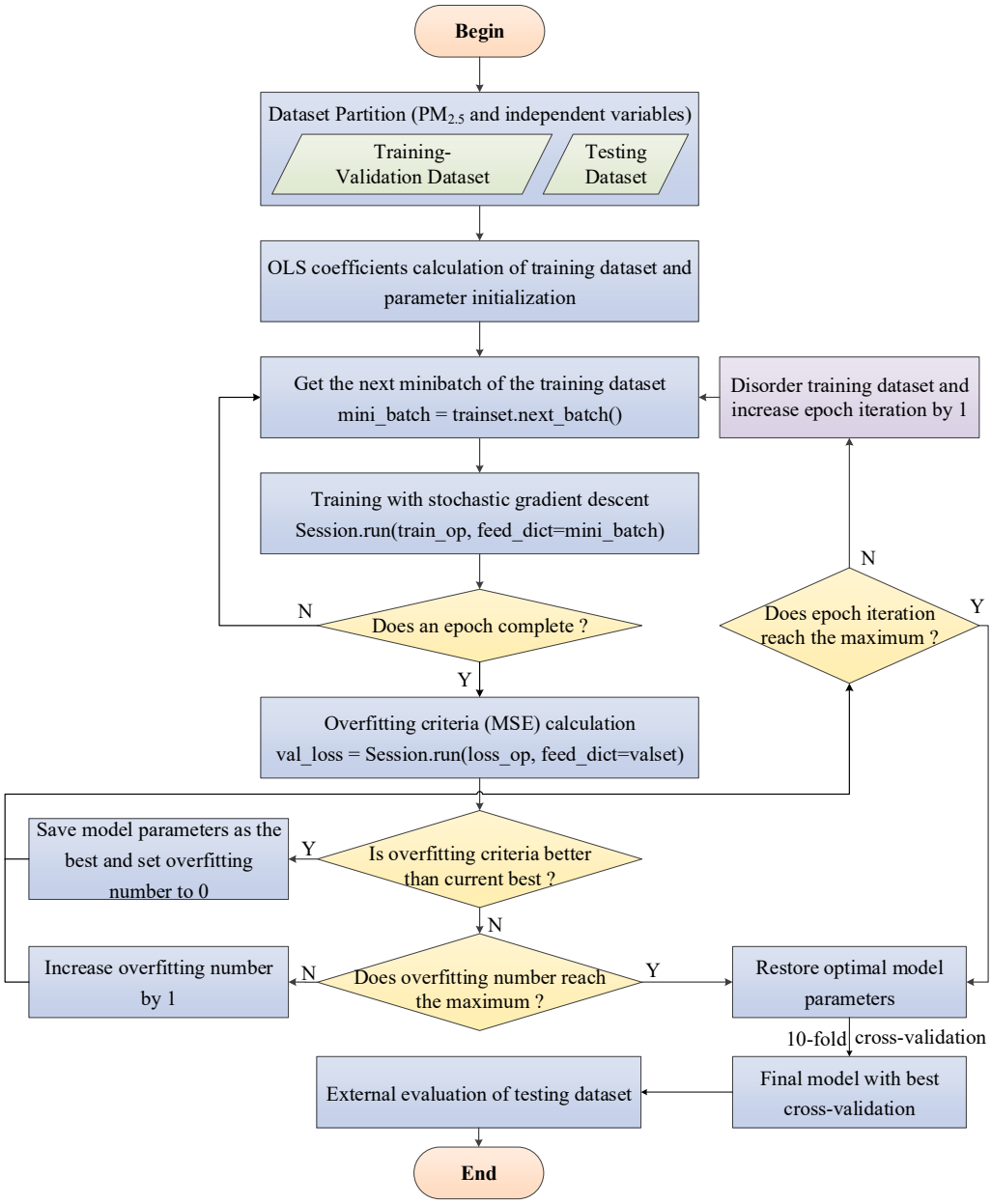


Figure S2. Training and validation procedures of the GNNWR model for $PM_{2.5}$-AOD modeling.

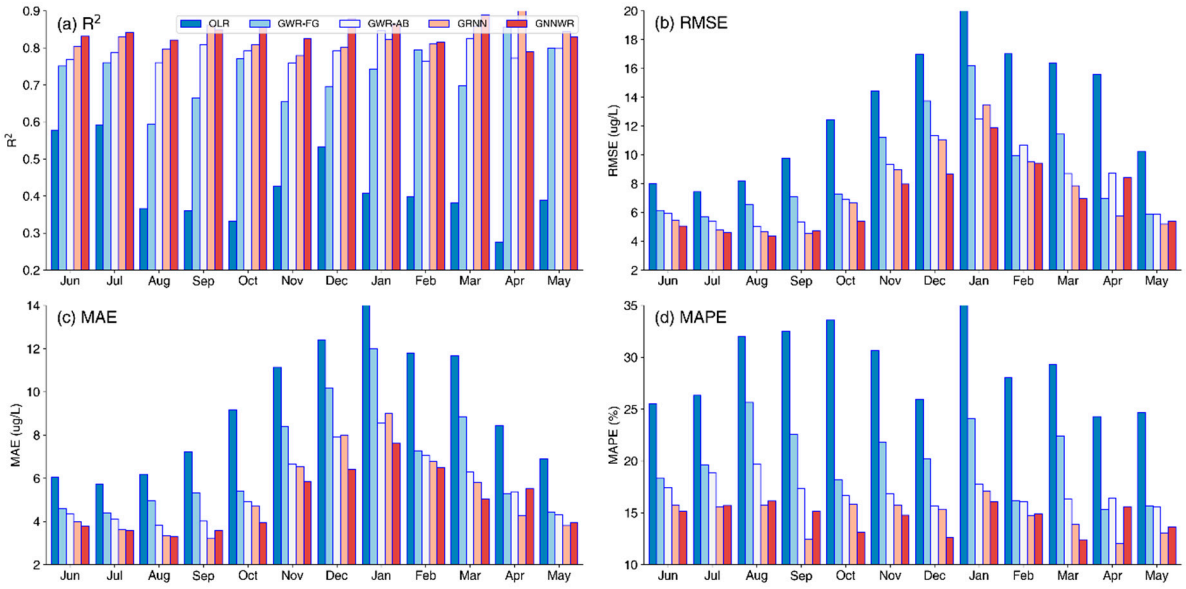


Figure S3. Training performance of $PM_{2.5}$ estimations in each month of the GNNWR model.

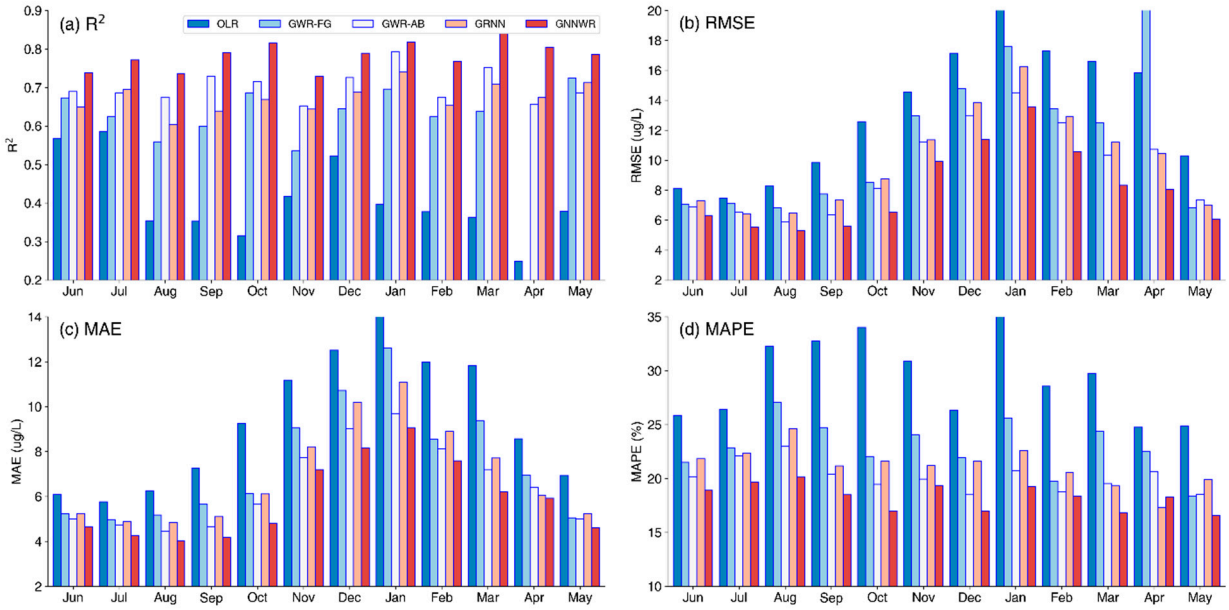


Figure S4. Cross-validation performance of $PM_{2.5}$ estimations in each month of the GNNWR model.

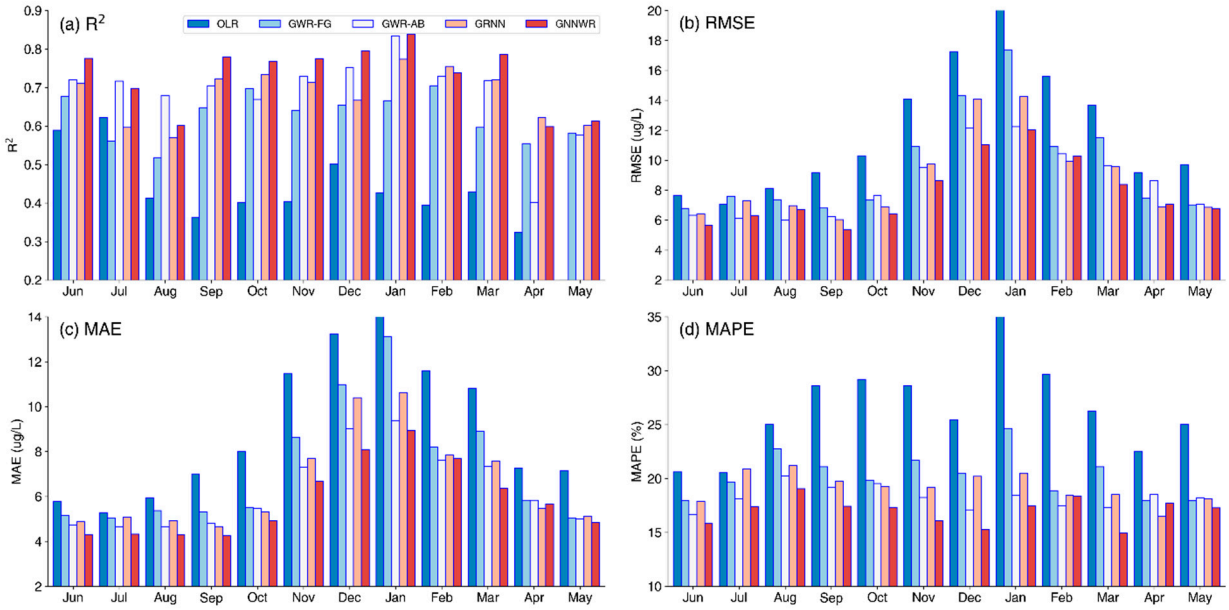


Figure S5. Prediction performance of $PM_{2.5}$ estimations in each month of the GNNWR model.

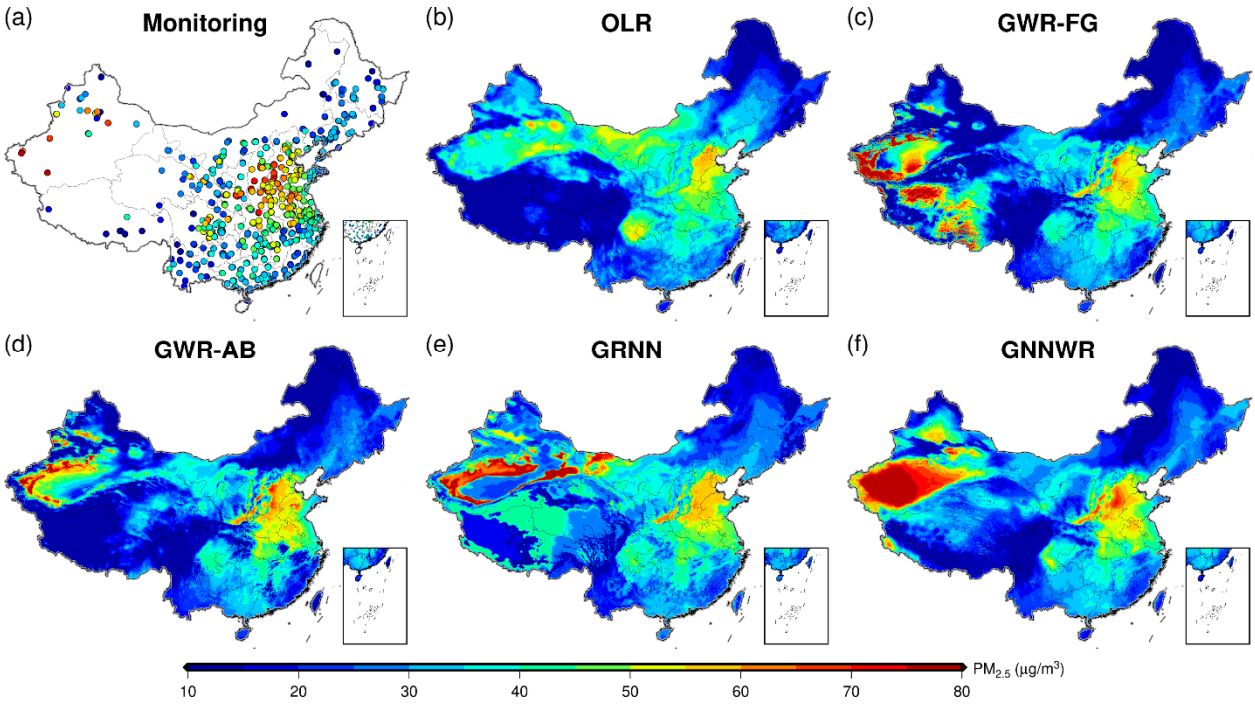


Figure S6. Spatial distributions of annual mean ground-measured (a) $PM_{2.5}$ concentrations and satellite-estimated (b-f) $PM_{2.5}$ concentrations of different models in China.

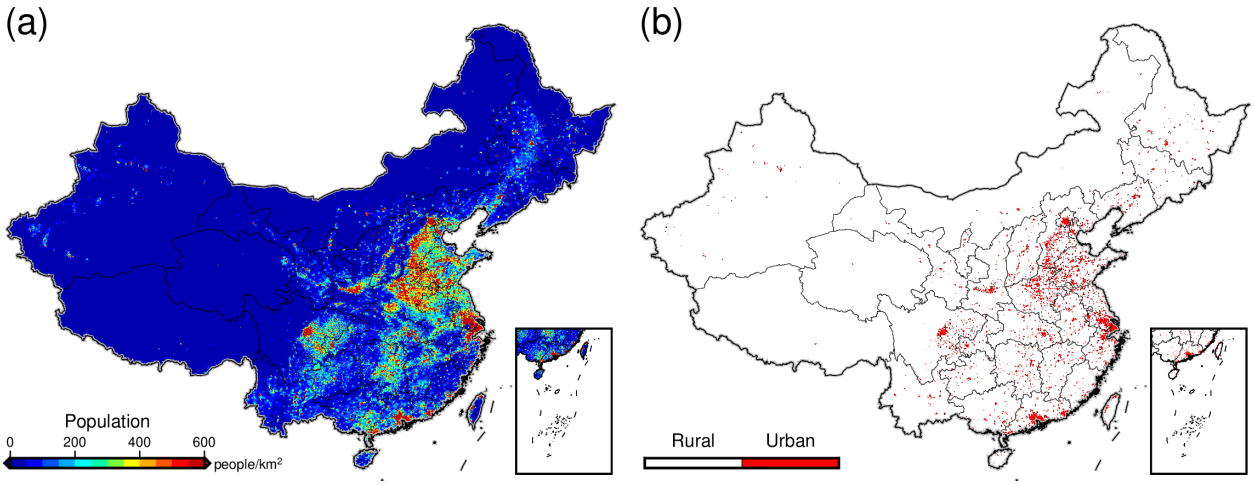


Figure S7. Spatial distributions of population count in 2017 and the urban and rural areas estimated from the population counts in China.

Supplementary Tables

Table S1. Data sources and descriptions of the datasets

Data Type	Dataset	Time	Space	Description	Source
Monitoring	$PM_{2.5}$	hour	point	/	CEMC (http://106.37.208.233:20035/)
Remote Sensing	AOD	day	750-m	aerosol optical depth	VIIRS (http://www.class.noaa.gov)
	NDVI	16-days	500-m	normalized difference vegetation index	MODIS (http://www.class.noaa.gov)
Meteorology	UW	hour	0.1°	10-m u-component wind	ERA5 (https://cds.climate.copernicus.eu)
	VW	hour	0.1°	10-m v-component wind	
	TEMP	hour	0.1°	2-m temperature	
	TP	hour	0.1°	total precipitation	
	BLH	hour	0.25°	boundary layer height	
	SP	hour	0.1°	surface pressure	
	E	hour	0.25°	evaporation	
	RH	hour	0.1°	relative humidity	
	WS	hour	0.1°	10-m wind speed	
	WD	hour	0.1°	10-m wind direction	
Geography	DEM	/	90-m	digital elevation model	SRTM (https://www2.jpl.nasa.gov/srtm)
Society	POP	5-years	0.5'	population count	SEDAC (http://sedac.ciesin.columbia.edu)

Table S2. Exploratory analysis of $PM_{2.5}$ and the explanatory variables

$PM_{2.5}$	AOD	WS	TEMP	TP	BLH	RH	DEM	NDVI
Pearson's r	0.441	-0.117	-0.407	-0.367	-0.199	-0.308	-0.171	-0.357
Sig.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
VIF	1.420	1.436	1.989	2.072	1.364	2.075	1.480	1.529

Table S3. Settings for the architectures and hyper-parameters for the GNNWR model

SWNN	Input	Hiddens		Output
	1024	[512, 128, 32]		9
Hyper-parameters	Learning Rate	Maximum Epoch	Batch Size	Dropout
	0.1-0.35-0.01	20000	64	0.8

Table S4. Annual mean and population weighted $PM_{2.5}$ ($PPM_{2.5}$) concentrations at province level in China. The $PPM_{2.5}$ includes the whole, urban and rural $PPM_{2.5}$. Rate is the variation rate from the mean $PM_{2.5}$ to the whole $PPM_{2.5}$. U/R is the ratio of urban $PPM_{2.5}$ and rural $PPM_{2.5}$.

Region	Population (people/km^2)	$PM_{2.5}$ (μg/m^3)	Population weighted $PM_{2.5}$ (μg/m^3)			Rate	U/R
			whole	urban	rural		
Xinjiang	10	48.88	63.27	61.18	64.29	29.45%	0.95
Hebei	263	41.34	54.13	57.90	51.47	30.93%	1.12
Shanxi	164	47.18	54.05	57.27	51.78	14.54%	1.11
Henan	409	50.57	53.76	56.14	52.02	6.32%	1.08
Beijing	1005	44.05	51.47	52.08	46.63	16.82%	1.12
Anhui	305	45.87	49.57	50.50	49.10	8.07%	1.03
Shandong	421	46.99	47.97	48.57	47.50	2.09%	1.02
Tianjin	856	48.91	46.87	46.37	48.57	-4.17%	0.95
Shaanxi	130	31.26	45.68	54.68	39.30	46.15%	1.39
Hubei	218	40.00	45.53	48.10	43.70	13.82%	1.10
Jiangsu	568	45.36	44.91	44.56	45.48	-0.99%	0.98
Sichuan	120	20.09	40.08	44.76	36.91	99.55%	1.21
Hunan	238	36.59	38.95	41.51	37.79	6.43%	1.10
Shanghai	3224	38.09	38.43	38.52	33.33	0.90%	1.16
Chongqing	245	33.09	36.76	38.42	35.51	11.07%	1.08
Qinghai	6	26.22	35.15	42.94	32.47	34.06%	1.32
Jiangxi	214	33.08	35.11	37.23	34.01	6.15%	1.09
Zhejiang	417	30.17	34.11	35.10	31.67	13.06%	1.11
Ningxia	92	28.69	33.51	38.01	30.29	16.81%	1.25
Guangxi	153	31.95	33.31	35.37	32.68	4.26%	1.08
Guizhou	147	31.37	31.74	32.15	31.62	1.18%	1.02
Gansu	41	26.86	30.49	37.44	28.33	13.53%	1.32
Guangdong	496	29.52	30.13	30.57	29.01	2.08%	1.05
Liaoning	191	27.16	29.04	30.09	27.93	6.94%	1.08
Jilin	90	25.16	28.21	29.38	27.52	12.13%	1.07
Heilongjiang	50	22.25	27.52	28.95	26.76	23.70%	1.08
Macao	19263	25.78	27.04	27.04	0.00	4.88%	/
Inner Mongolia	14	20.64	25.36	31.59	22.81	22.89%	1.38
Fujian	203	23.75	24.84	25.83	23.57	4.61%	1.10
Taiwan	449	22.54	24.09	24.13	23.88	6.89%	1.01
Yunnan	98	21.06	23.88	25.42	23.60	13.42%	1.08
Hong Kong	2314	25.33	18.10	18.17	14.33	-28.52%	1.27
Hainan	185	18.81	17.56	17.64	17.53	-6.66%	1.01
Tibet	3	19.95	15.04	16.39	14.97	-24.61%	1.09