Supplement of Atmos. Chem. Phys., 19, 12413–12430, 2019
https://doi.org/10.5194/acp-19-12413-2019-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Estimating ground-level CO concentrations across China based on the national monitoring network and MOPITT: potentially overlooked CO hotspots in the Tibetan Plateau

Dongren Liu et al.

Correspondence to: Yu Zhan (yzhan@scu.edu.cn)

S.1 Temporal and spatial convolution

The data of MOPITT retrieved surface CO (MOPITT-CO) are processed with the temporal and spatial convolution to filter noises and fill data gaps. In the first step, the temporal convolution with a 1-dimensioanl Gaussian kernel is employed to process the MOPITT-CO data for each grid cell:

$$M_T(t_0) = \sum_t [M(t) \cdot W_T(t_0 - t)] / \sum_t W_T(t_0 - t) \quad (1)$$

where $M_T(t_0)$ is the output value on day t_0 processed by the temporal convolution, $M(t)$ is the original MOPITT-CO value on day t, and $W_T(t_0 - t)$ is the weighting factor determined by the 1-dimensional Gaussian function:

$$W_T(t_0 - t) = exp[-(t_0 - t)^2 / (2\sigma_T^2)] \quad (2)$$

where the standard deviation (σ_T) is set to 60 according to the sensitivity analysis on the completeness and smoothness of the processed data.

In the second step, the spatial convolution with a 2-dimensional Gaussian kernel is employed to process the output from the previous step day by day:

$$M_{TS}(x_0, y_0) = \sum_{x,y} [M_T(x, y) \cdot W_S(x_0 - x, y_0 - y)] / \sum_{x,y} W_S(x_0 - x, y_0 - y) \quad (3)$$

where $M_{TS}(x_0, y_0)$ is the output value for cell (x_0, y_0) processed by the spatial convolution, $M_T(x, y)$ is the processed MOPITT-CO value from the first step for cell (x, y), and $W_S(x_0 - x, y_0 - y)$ is the weighting factor determined by the 2-dimensional Gaussian function:

$$W_S(x_0 - x, y_0 - y) = exp\{-[(x_0 - x)^2 + (y_0 - y)^2)] / 2\sigma_S^2\} \quad (4)$$

where the standard deviation (σ_S) is set to 0.1 according to the sensitivity analysis on the completeness and smoothness of the processed data.

S.2 Averaging kernel

The averaging kernel (matrix A) adjusts the weights of the "true" state (vector x) and the a priori (vector x_a) in deriving the MOPITT CO retrievals (vector $\hat{x}$) (Deeter et al., 2003; Rodgers, 2000).

$$\hat{x} \approx Ax + (I - A)x_a \quad (5)$$

where I is the identity matrix. Each row of A corresponds to a vertical layer of the CO profile, and the sum of a row shows the overall dependence of the MOPITT CO retrieval at that layer on the a priori information. A small row-sum value indicates strong dependence on the a priori information.

S.3 Algorithm of random forests (Breiman, 2001)

For *tree* = 1 to *N* (e.g., 500 trees in this study):

- Randomly draw a sample from the training data with replacement through bootstrapping;
- A tree is grown from a single node, and the following steps are repeated until the minimum number of observations is present at each terminal node:
  - Randomly select a subset of predictors to be considered at each split;
  - Find the split that reduces the squared error the most;

Average the predictions made by all the decision trees as the output of the random forest.

S.4 Environmental condition data

- The daily weather conditions, including atmospheric pressure, air temperature, precipitation, evaporation, insolation duration, and wind speed, were obtained from 839 meteorological stations (CMA, 2017).
- The elevation data were retrieved from the Shuttle Radar Topography Mission (SRTM) database (Jarvis et al., 2016).
- The data of population density, road density, and land use were extracted from the Gridded Population of the World, the OpenStreetMap, and the GlobeLand30 databases, respectively (CIESIN, 2016; OSP, 2016; Jun et al., 2014).
- The daily planetary boundary height (PBLH) data were obtained from the Modern-Era Retrospective Analysis for Research and Application (GMAO, 2015).
- The Normalized Difference Vegetation Index (NDVI) data were retrieved from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite retrievals (Didan et al., 2015).
- The anthropogenic emission inventories were obtained from the Multi-resolution Emission Inventory for China (MEIC) database (Li et al., 2017). Due to the data availability, the emissions for 2013 and 2015 were linearly interpolated from the available emission data for 2012, 2014, and 2016.

Table S1. List of variable symbols and definitions.

Symbol	Unit	Variable definition	Spatial[a]	Temporal[a]	Convolution[b]
MOPITT	molecule cm^{-2}	MOPITT-retrieved CO surface mixing ratio	0.25°	Day	Temporal and Spatial
DOY	-	Day of year	-	-	-
YEAR	-	Year	-	-	-
EVP	mm	Evaporation	Point	Day	-
PRE	mm	Precipitation	Point	Day	-
PRS	hPa	Atmospheric pressure	Point	Day	-
RHU	%	Relative humidity	Point	Day	-
SSD	hour	Sunshine duration	Point	Day	-
TEM	℃	Temperature	Point	Day	-
WIN	m s^{-1}	Wind speed	Point	Day	-
PBLH	Km	Planetary boundary layer height	0.625°×0.5°	Day	-
ELV	M	Elevation	90 m	-	Spatial
NDVI	-	Normalized Difference Vegetation Index	250 m	8 Days	Spatial
POP	people km^{-2}	Population density	30"	-	Spatial
LU10	%	Cultivated land area	30 m	-	Spatial
LU20	%	Forest area	30 m	-	Spatial
LU30	%	Grassland area	30 m	-	Spatial
LU40	%	Shrubland area	30 m	-	Spatial
LU50	%	Wetland area	30 m	-	Spatial
LU60	%	Waterbody area	30 m	-	Spatial
LU80	%	Artificial surface area	30 m	-	Spatial
LU90	%	Bareland area	30 m	-	Spatial
LU100	%	Permanent frozen land area	30 m	-	Spatial
LU255	%	Sea area	30 m	-	Spatial
ROAD	Km $grid^{-1}$	Road density	Polyline	-	Spatial
eBC	Mg $grid^{-1}$	Emission of black carbon	0.25°	Month	Spatial
eCO	Mg $grid^{-1}$	Emission of CO	0.25°	Month	Spatial
eCO2	Mg $grid^{-1}$	Emission of CO_2	0.25°	Month	Spatial
eNH3	Mg $grid^{-1}$	Emission of NH_3	0.25°	Month	Spatial
eNOx	Mg $grid^{-1}$	Emission of NO_2 and NO	0.25°	Month	Spatial
eOC	Mg $grid^{-1}$	Emission of organic carbon	0.25°	Month	Spatial
ePM25	Mg $grid^{-1}$	Emission of $PM_{2.5}$	0.25°	Month	Spatial
ePMcoar	Mg $grid^{-1}$	Emission of PM-coarse	0.25°	Month	Spatial
eSO2	Mg $grid^{-1}$	Emission of SO_2	0.25°	Month	Spatial
eVOC	Mg $grid^{-1}$	Emission of VOC	0.25°	Month	Spatial

[a] Spatial or temporal resolution of raw data.

[b] Temporal: MOPITT is processed with the temporal and spatial convolution. Spatial: These variables have accompanying variables processed with the spatial convolution.

Table S2. Coverage rates of MOPITT-CO retrievals across China (μ ± σ; %)[a].

Year(s)	Spring	Summer	Fall	Winter	Annual
2013	3.2 ± 2.1	3.0 ± 2.0	4.5 ± 2.6	4.1 ± 2.7	3.7 ± 0.6
2014	3.1 ± 2.1	2.7 ± 2.0	4.3 ± 2.5	4.1 ± 2.4	3.6 ± 0.6
2015	3.0 ± 2.0	3.0 ± 2.1	4.0 ± 2.4	3.8 ± 2.4	3.4 ± 0.6
2016	3.0 ± 2.1	3.0 ± 2.0	4.0 ± 2.7	3.6 ± 2.3	3.4 ± 0.6
2013-2016	3.1 ± 1.6	2.9 ± 1.5	4.2 ± 1.9	3.9 ± 2.0	3.5 ± 0.5

[a] σ stands for the spatial variation. Please refer to Fig. S2 for the coverage maps.

Table S3. Correlations between the daily CO observations from the monitoring network and the MOPITT surface retrievals for China during 2013-2016.

Region[a]	Spring	Summer	Fall	Winter	2013	2014	2015	2016	2013-2016
Central China	0.08	0.18	0.21	0.27	0.59	0.22	0.32	0.24	0.30
East China	0.12	0.24	0.36	0.37	0.46	0.34	0.48	0.40	0.43
North China	0.30	0.23	0.31	0.42	0.36	0.31	0.39	0.38	0.37
Northeast China	0.28	0.34	0.38	0.30	0.28	0.41	0.39	0.38	0.39
Northwest China	0.31	0.12	0.27	0.31	0.52	0.39	0.35	0.31	0.38
South China	0.49	0.53	0.59	0.52	0.65	0.56	0.55	0.57	0.58
Southwest China	0.10	0.12	0.25	0.20	0.14	-0.02	0.17	0.22	0.17
Nation	0.22	0.31	0.32	0.34	0.42	0.31	0.3	0.34	0.44
Central Tibetan Plateau	0.15	0.25	0.11	0.05	-0.19	-0.09	-0.10	0.14	-0.12
North China Plain	0.18	0.13	0.25	0.35	0.33	0.30	0.27	0.36	0.30

[a] Please refer to Fig. 1 for the locations of these regions.

Table S4. Comparisons of the RF and RF-STK models in predicting daily ground-level CO concentrations across China during 2013-2016 based on the 10-fold cross-validation.

Metric[a]	RF_r[b]	RF[b]	RF_{rw}[b]	RF_w[b]	RF_w-STK[b]	RF_{rw}-STK[bc]	RF_{rw}-STK[b]
R^2	0.56	0.53	0.54	0.53	0.49	0.49	0.51
Slope	0.60	0.55	0.57	0.55	0.63	0.60	0.64
RMSE	0.50	0.52	0.51	0.52	0.55	0.57	0.54
RPE	46.1%	48.0%	47.1%	48.1%	51.0%	53.3%	50.4%
MFB	0.0832	-0.013	-0.0076	-0.0128	-0.030	0.064	-0.022
MFE	0.31	0.31	0.31	0.31	0.36	0.37	0.35
MNB	0.90	0.64	0.66	0.64	0.68	0.74	0.70
MNE	1.09	0.90	0.91	0.90	0.97	1.0	0.98

[a] R^2: coefficient of determination; RMSE: root mean square error (mg m^{-3}); RPE: relative prediction error; MFB: mean fractional bias; MFE: mean fractional error; MNB: mean normalized bias; MNE: mean normalized error.

[b] RF: random forest; STK: spatiotemporal kriging. Subscript r indicates a reduced model through variable selection, and subscript w means that the training samples were inversely weighted by the associated population densities. The CO concentrations were log-transformed to train all the models except for RF_r which was trained with the CO concentrations at native scale.

[c] This RF-STK model was developed with the a priori information rather than the MOPITT retrievals.

Table S5. Previous studies modeling surface CO concentrations.

Reference	Model[a]	Study Area	Study Period	Evaluation Metric[b]
(Hooghiemstra et al., 2012)	4D-Var system	Bahia, Brazil	2007-2009	R=0.6 (daily; prior) R=0.8 (daily; posterior)
(Yeganeh et al., 2012)	SVM; PLS-SVM	Tehran, Iran	2007.01-2011.01	R^2=0.56 (daily; SVM) R^2=0.65 (daily; PLS-SVM)
(Hu et al., 2016)	CMAQ	China	2013.03-2013.12	MNE=0.59~0.66 (daily) MFE=0.86~1.02 (daily)

[a] 4D-Var system: Four-dimensional variational data assimilation system; SVM: support vector machine; PLS-SVM: hybrid model of partial least square and support vector machine; CMAQ: Community Multiscale Air Quality model.

[b] All these studies conducted validation at daily level. Both prior and posterior estimates of the model were evaluated with an independent dataset (Hooghiemstra et al., 2012). Goodness of fit was evaluated in (Yeganeh et al., 2012), and an independent dataset was used for validation in (Hu et al., 2016).

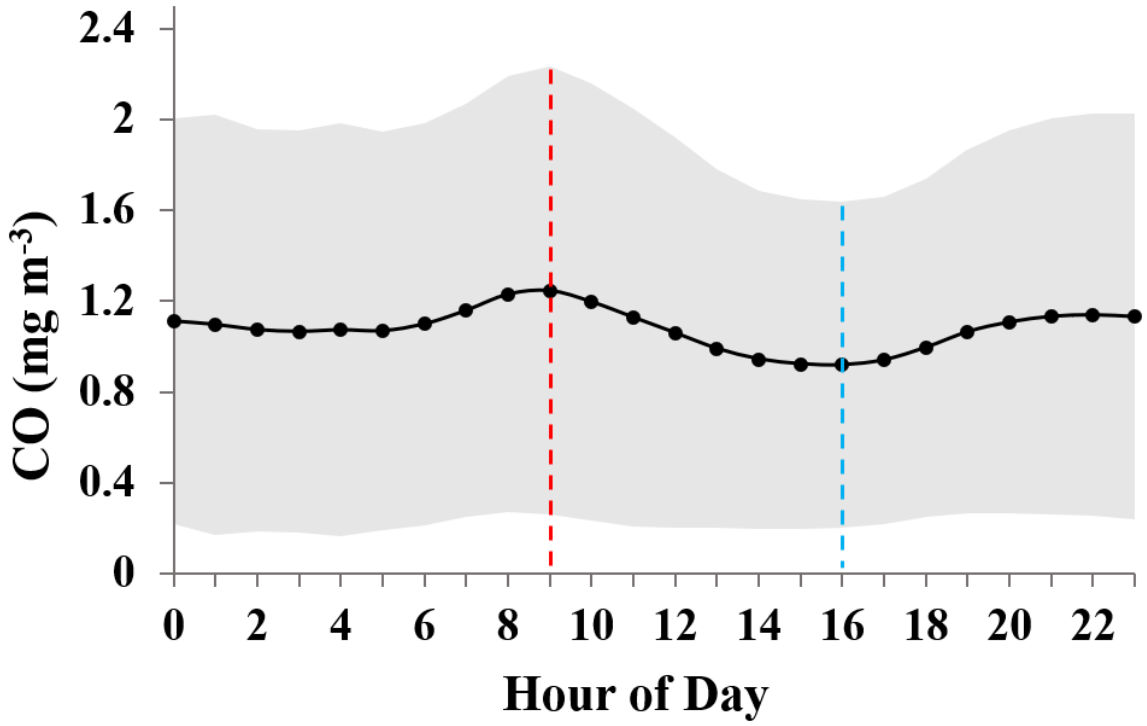

Figure S1: Average diurnal pattern in CO concentrations across 1656 monitoring sites for China during 2013-2016. The peak and the valley appeared at 9am and 4pm (Beijing Standard Time). The shaded area represents the standard deviations.

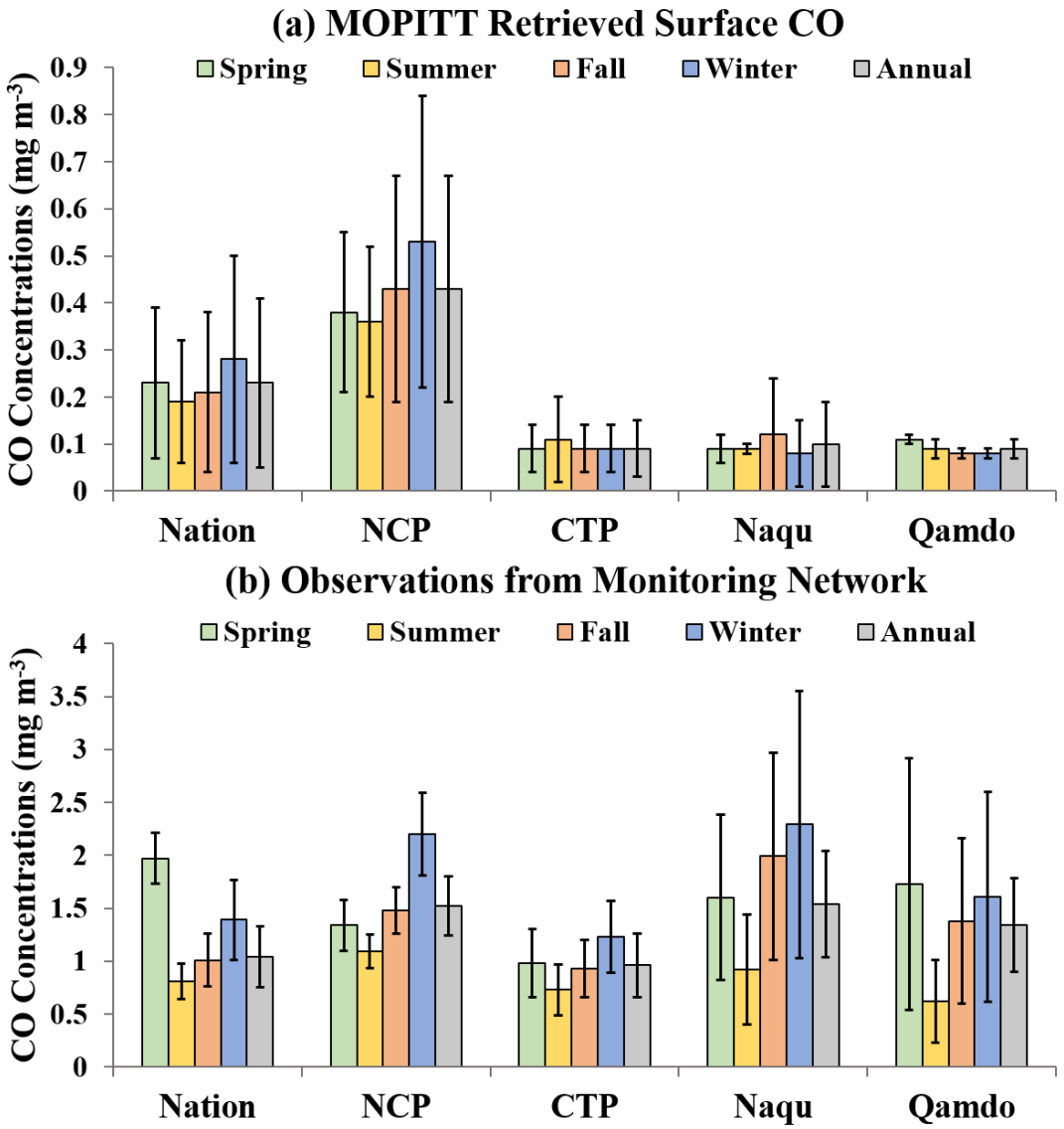

Figure S2: Seasonal averages of the ground-level CO concentrations (mg m^{-3}) for the whole China, the North China Plain (NCP), the Central Tibetan Plateau (CTP), Naqu, and Qamdo during 2013-2016 based on (a) the MOPITT retrieved surface CO and (b) the observations from the monitoring network. The error bars represent the standard deviations. Naqu and Qamdo are two main cities in CTP (Fig. 8).

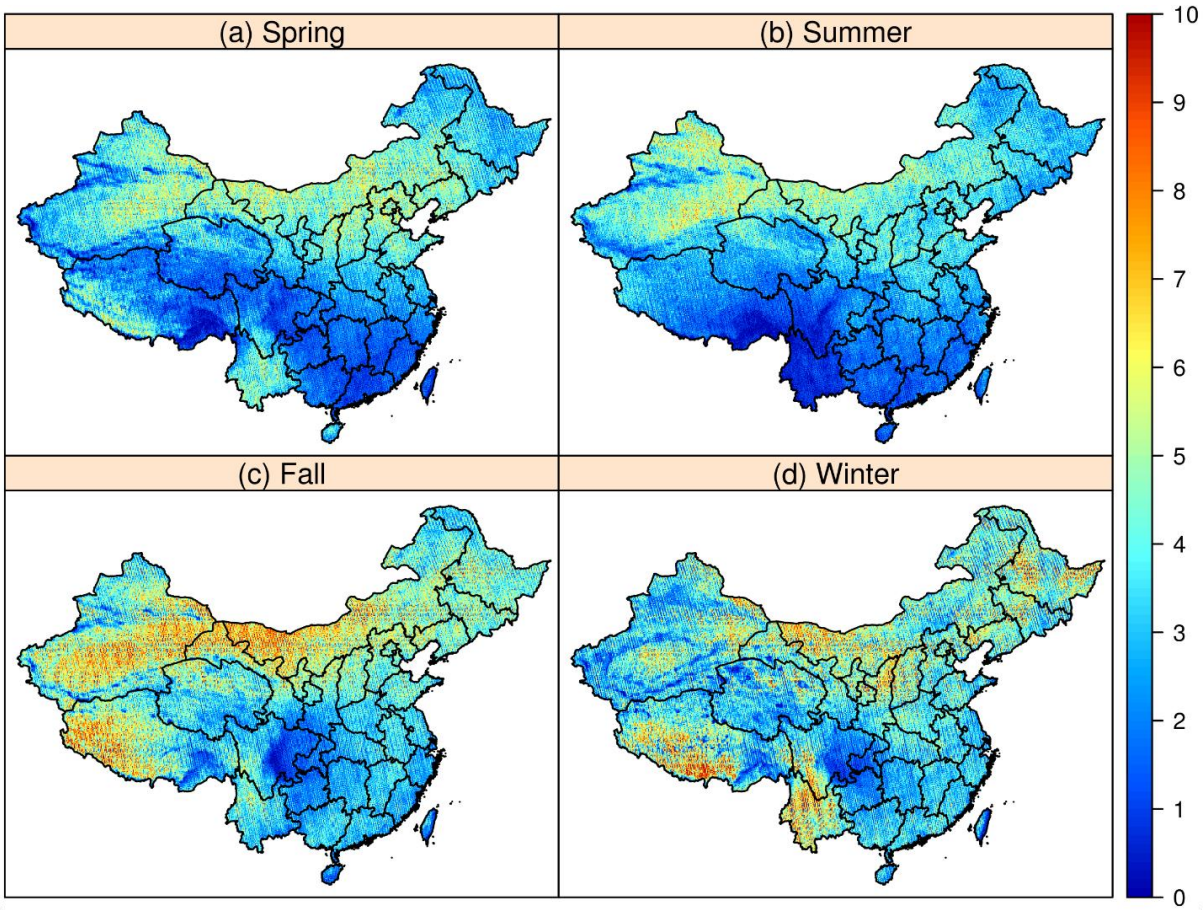

Figure S3: Coverage rates (%) of MOPITT-CO retrievals for (a) spring, (b) summer, (c) fall, and (d) winter during 2013-2016 across China. The coverage rate at each grid cell was calculated as the percentage of days with MOPITT-CO retrievals in each season.

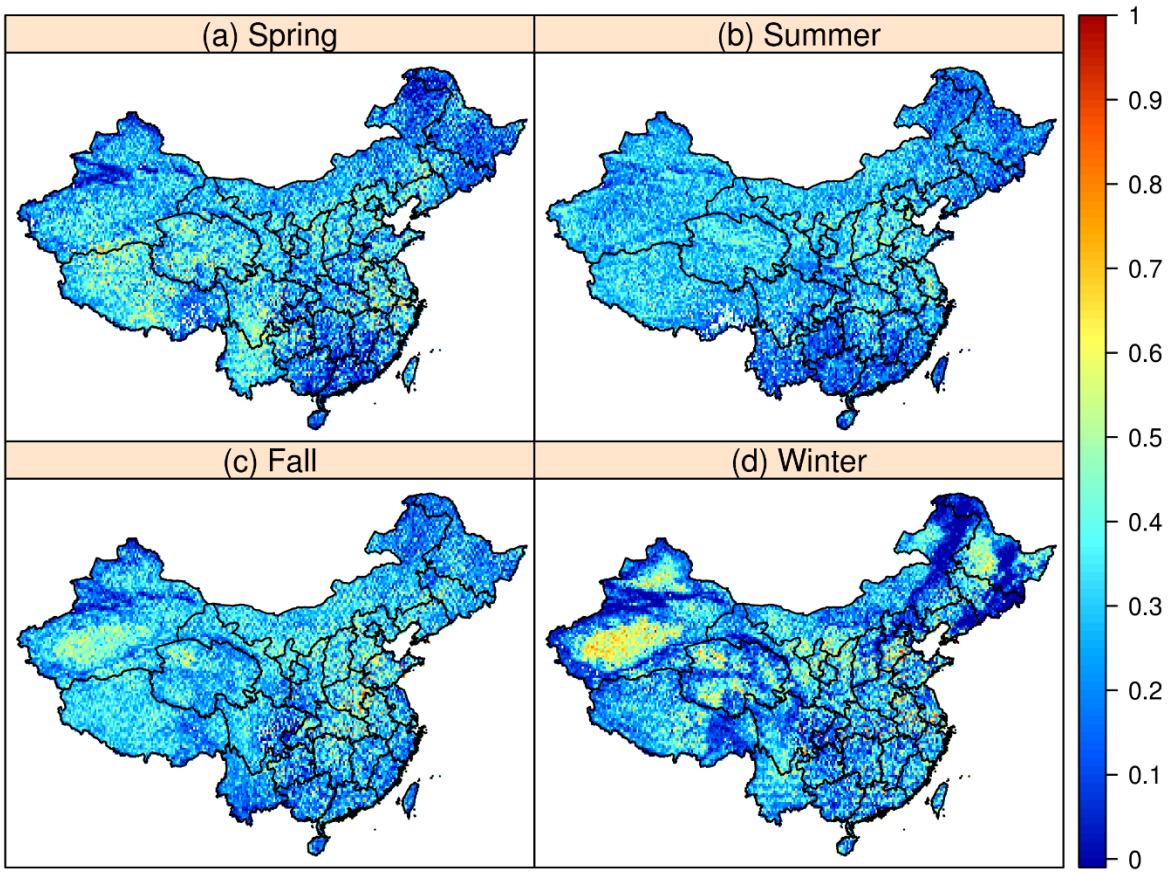

Figure S4: Seasonal means of the averaging-kernel row-sum values associated with the MOPITT retrieved surface CO for (a) spring, (b) summer, (c) fall, and (d) winter during 2013-2016 across China. Small row-sum values indicate strong dependence of the MOPITT retrievals on the a priori information. Please refer to "S.2 Averaging kernel" for more explanation.

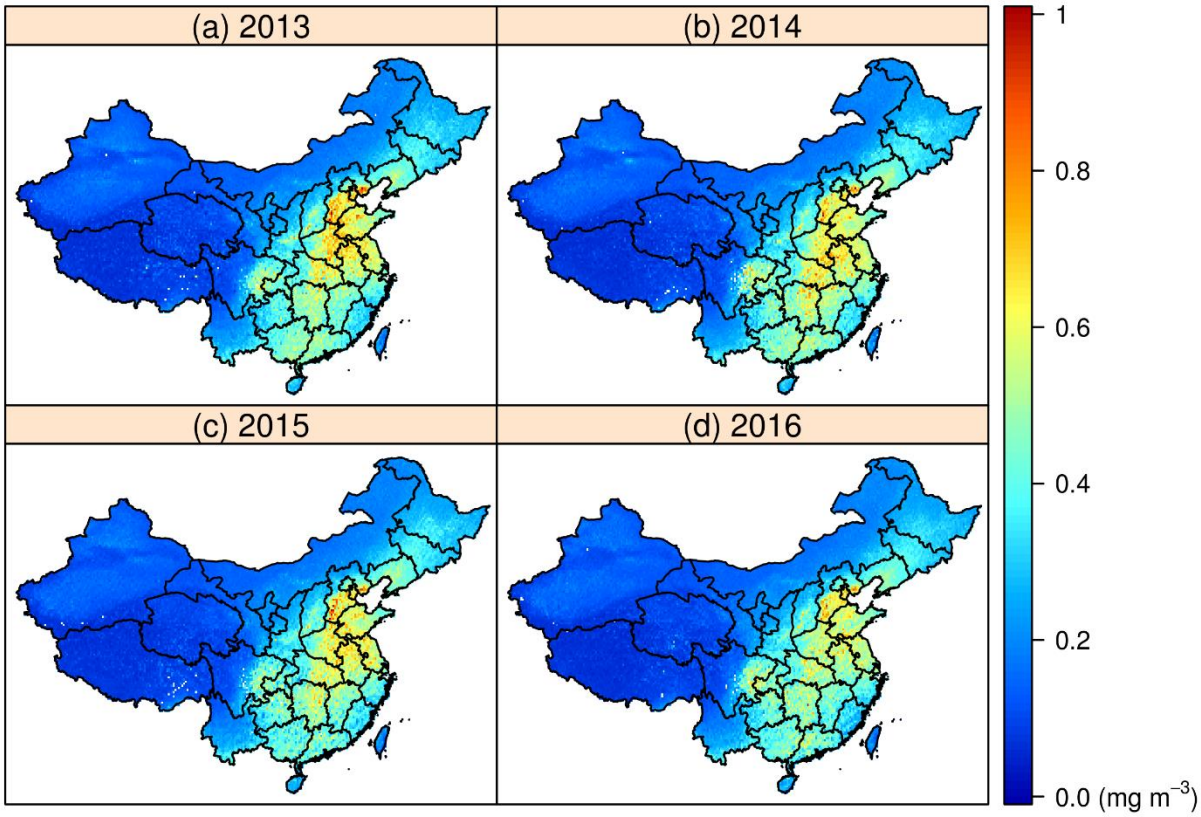

Figure S5: Annual averages of the MOPITT retrieved surface CO concentrations for (a) 2013, (b) 2014, (c) 2015, and (d) 2016.

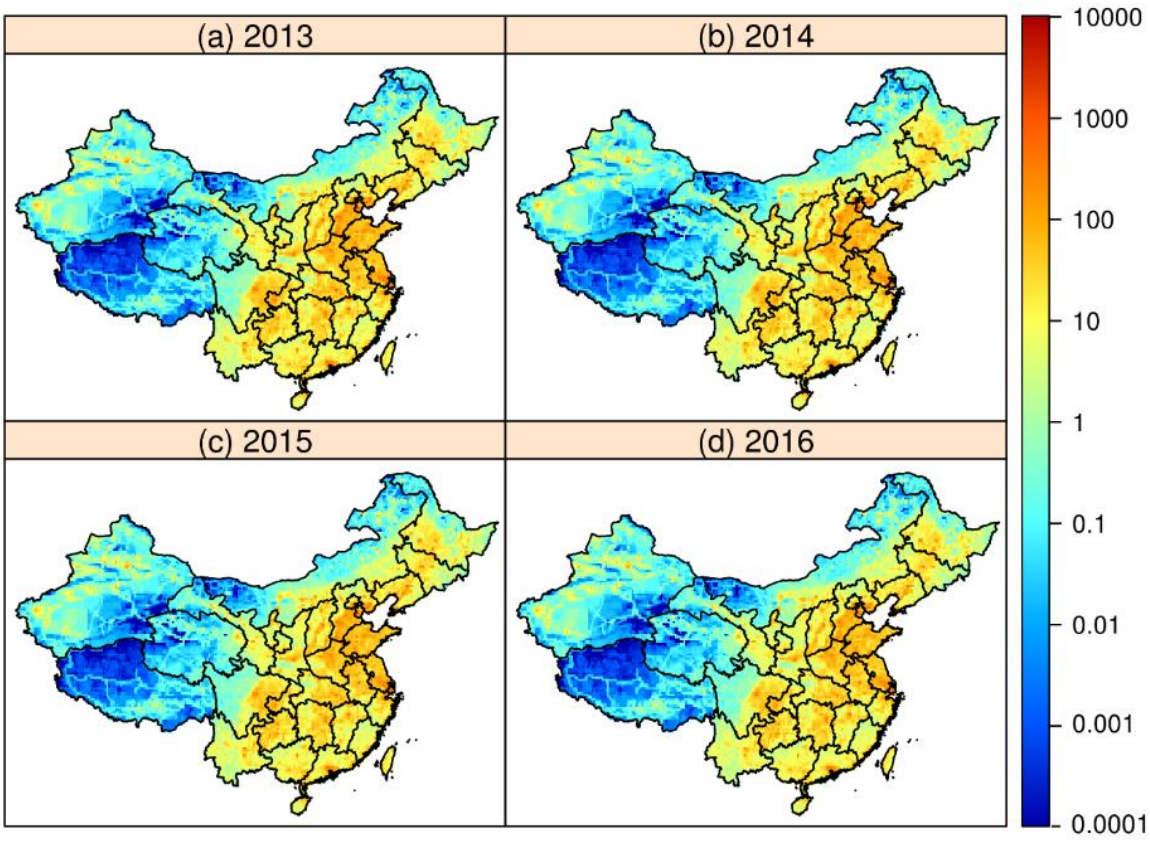

Figure S6: Annual total CO emissions (t) in (a) 2013, (b) 2014, (c) 2015, and (d) 2016 from anthropogenic sources across China. Due to the data availability, the CO emissions for 2013 and 2015 were linearly interpolated from the available data for 2012, 2014, and 2016 (Li et al., 2017).

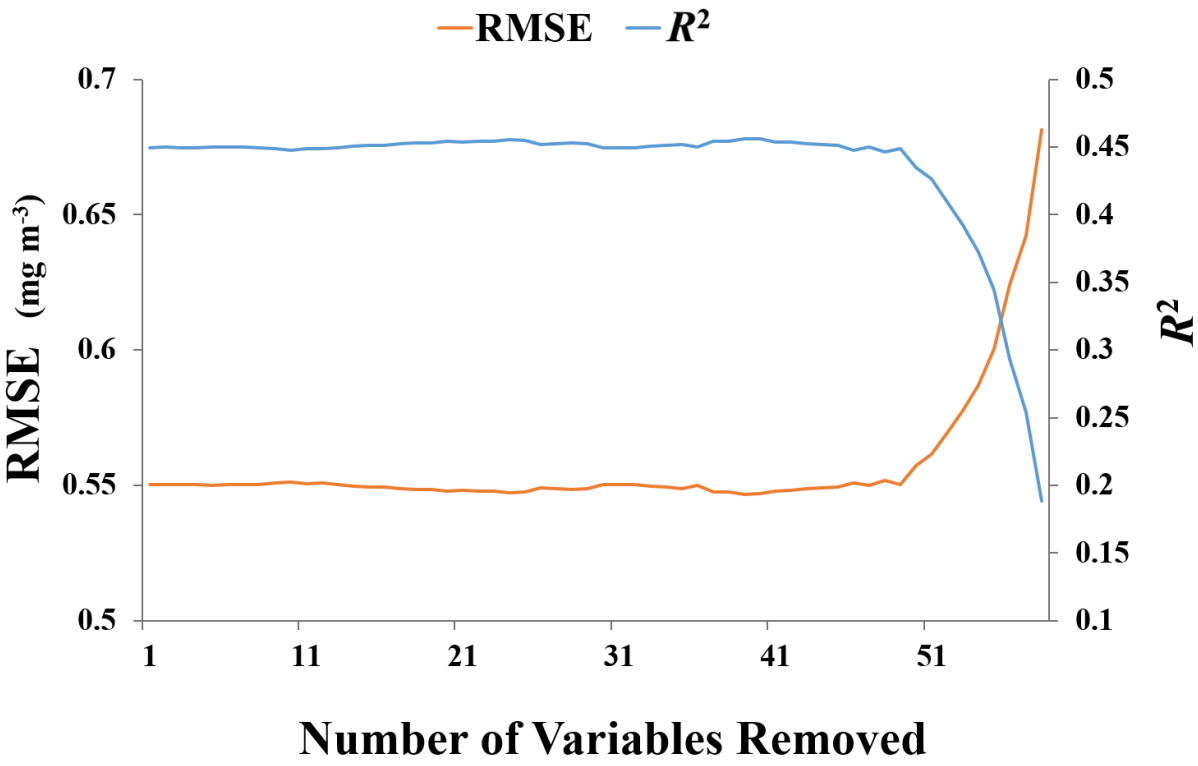

Figure S7: Evolution of the cross-validation RMSE (mg m^{-3}) and R^2 for the random forest submodels through the stepwise backward variable selection process.

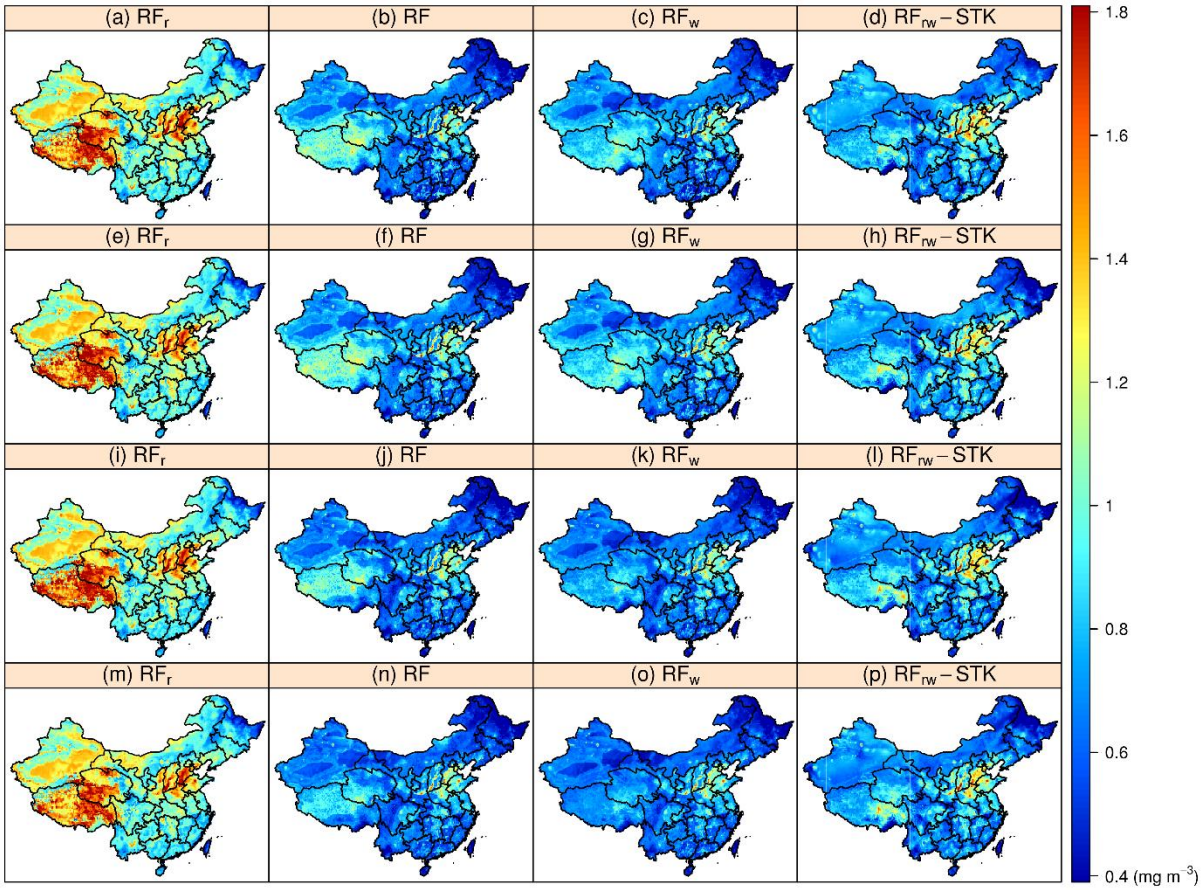

Figure S8: Annual average ground-level CO concentrations from 2013 to 2016 predicted by the RF_r, RF, RF_w and RF_{rw}-STK models with the MOPITT retrievals. RF: random forest; STK: spatiotemporal kriging. Subscript r indicates a reduced model through variable selection, and subscript w means that the training samples were inversely weighted by the associated population densities. The CO concentrations were log-transformed to train all the models except for RF_r which was trained with the CO concentrations at native scale. The predictions for 2013-2016 by each model are presented from top to bottom rows.

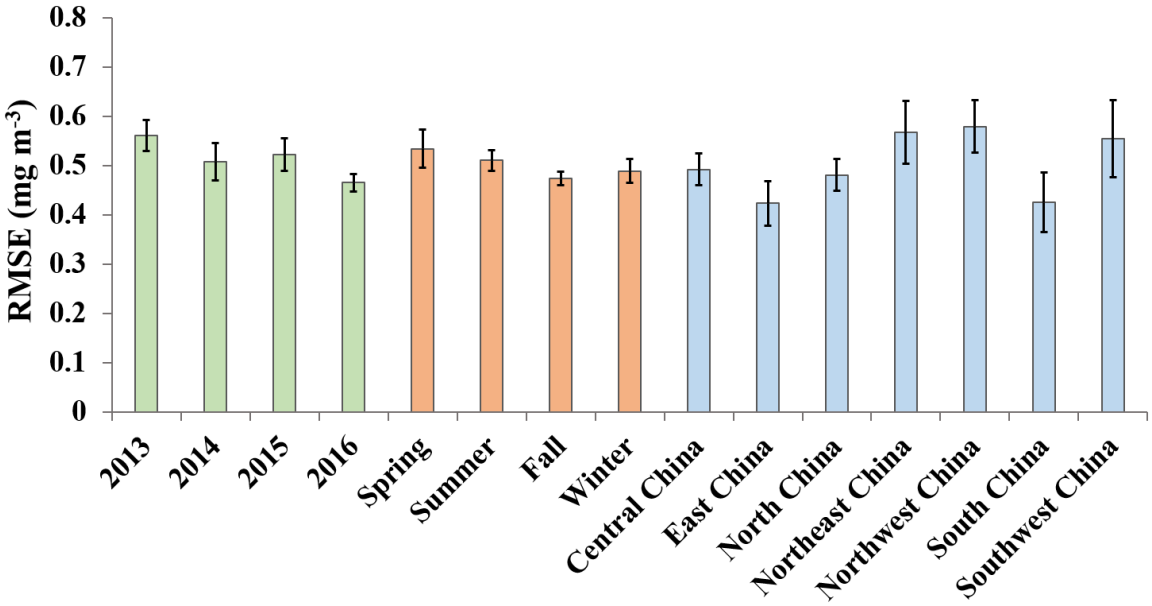

Figure S9: Performance of the RF-STK model in predicting daily CO concentrations by regions, years, and seasons. The mean and standard deviation of the root mean square error (RMSE, mg m^{-3}) over all the 10-fold cross-validations are presented. The numbers of monitoring sites in Central, East, North, Northeast, Northwest, South, and Southwest China are 267, 255, 307, 171, 159, 278, and 219, respectively. The numbers of monitoring sites in 2013, 2014, 2015, and 2016 are 743, 1041, 1542, and 1603, respectively.

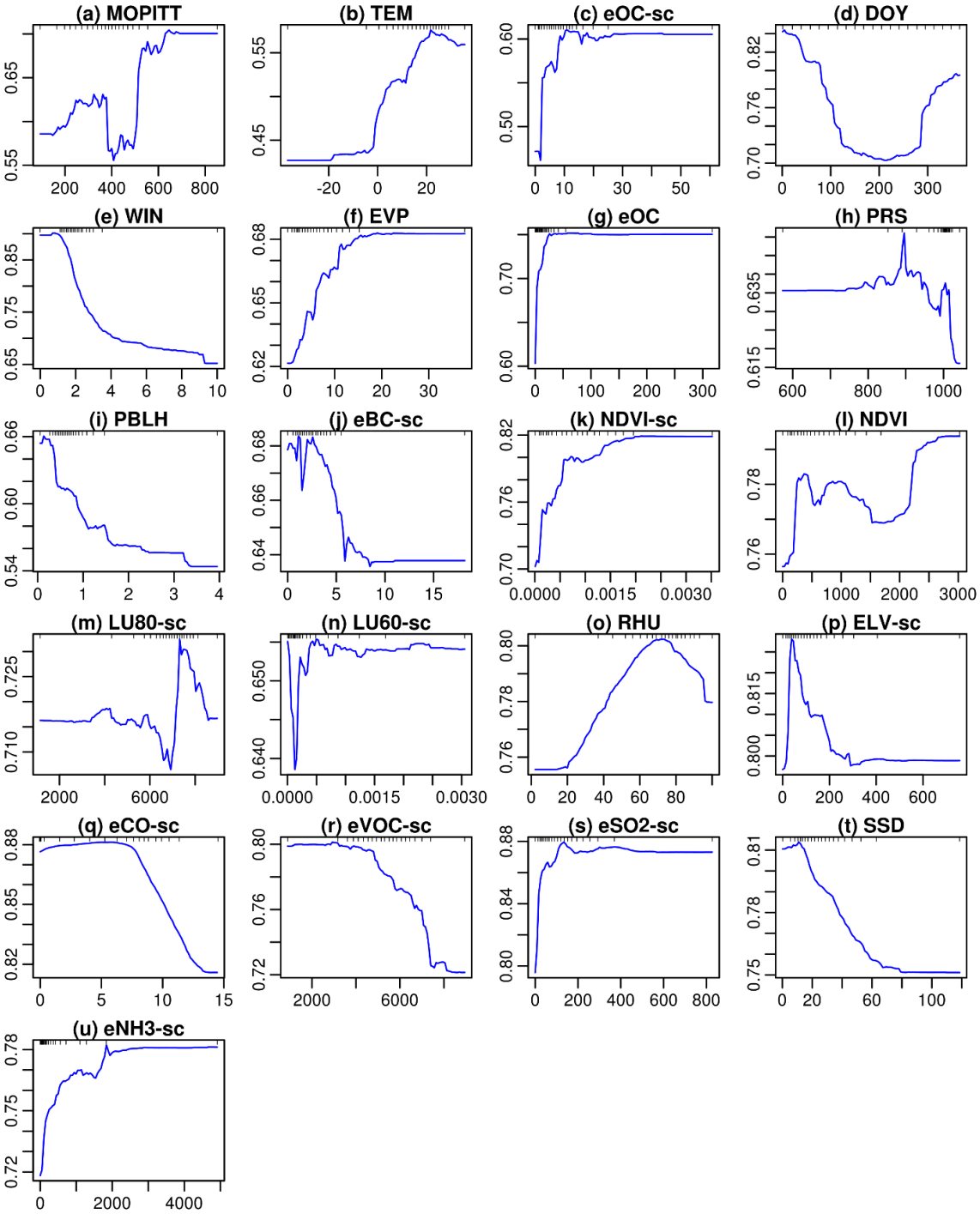

Figure S10: Partial dependence plots of the random forest submodel for delineating the relationship between each predictor variable and the ground-level CO concentrations. Partial dependence (Y axis) is the effect of a predictor variable (X axis) on the CO concentrations when the values of all the other predictor variables are fixed at their averages (Friedman, 2001). The subplots are arranged in the order of variable importance. Please refer to Table S1 for the descriptions and units of the predictor variables. The rug plot indicates the data density. Note that the partial dependence estimations are of high uncertainty given low data densities.

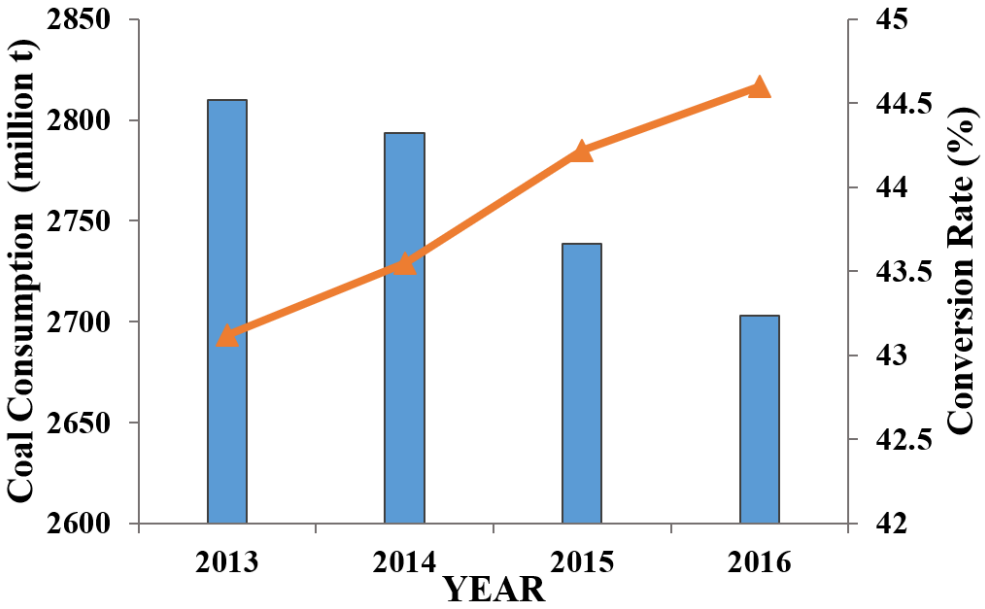

Figure S11: Coal consumption amounts and energy conversion rates in the sector of power generation and heating for China during 2013-2016 (CSY, 2018).

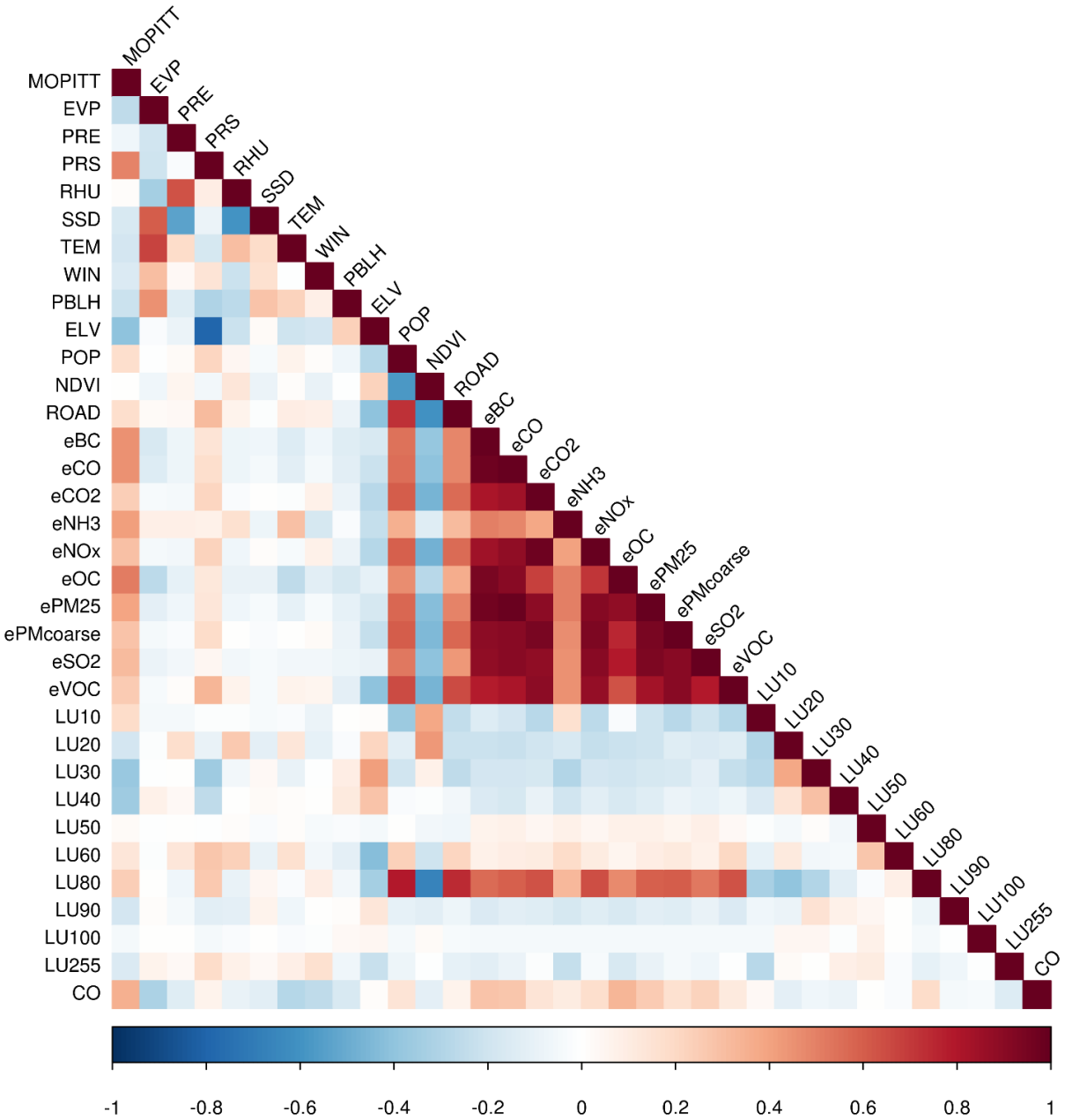

Figure S12: Correlations among the predictor variables and the ground-level CO concentrations, which were measured by the Spearman's rank correlation coefficients. Please refer to Table S1 for the detailed descriptions of the variables.

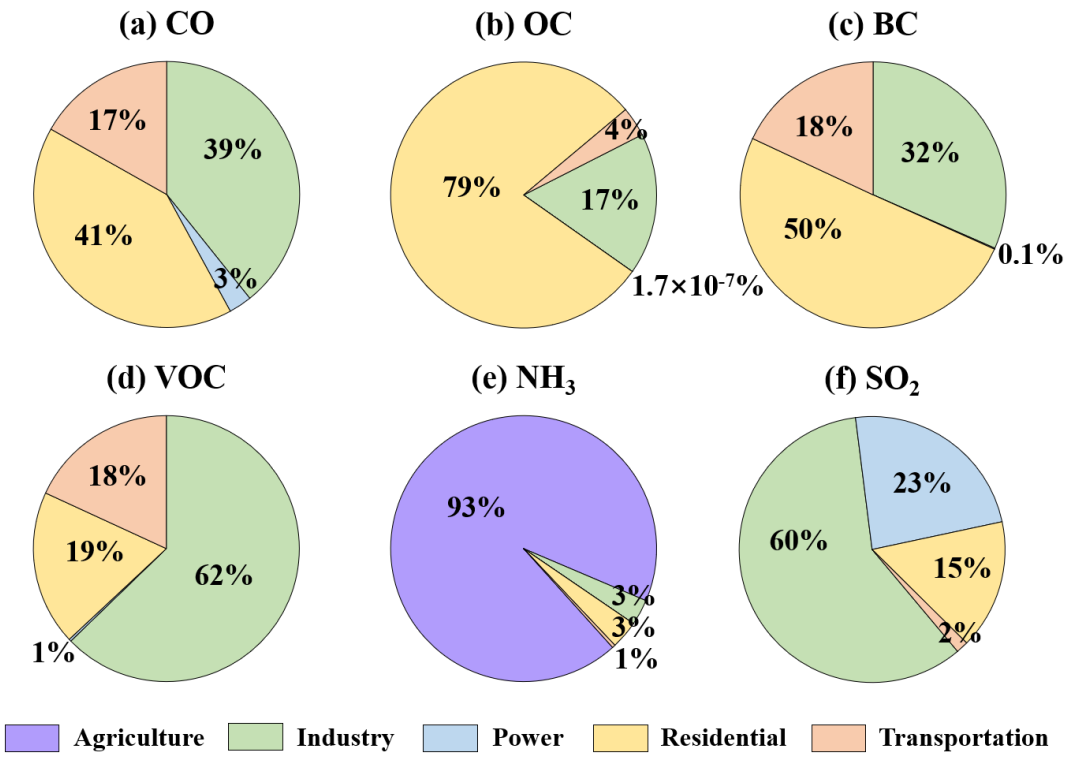

Figure S13: Anthropogenic emission sources of (a) CO, (b) organic carbon (OC), (c) black carbon (BC), (d) volatile organic compound (VOC), (e) NH_3, and (f) SO_2 for China during 2013-2016 (Li et al., 2017).

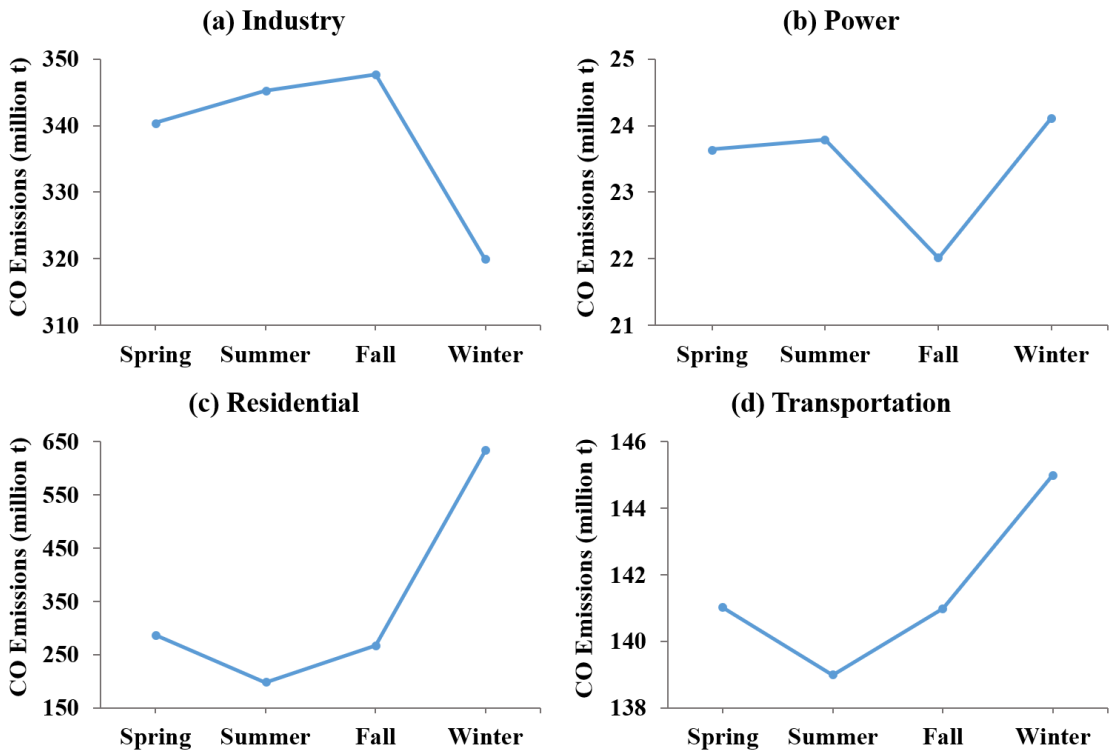

Figure S14: Total CO emissions (million t) from (a) industry, (b) power, (c) residential, and (d) transportation sectors in each season over China during 2013-2016 (Li et al., 2017).

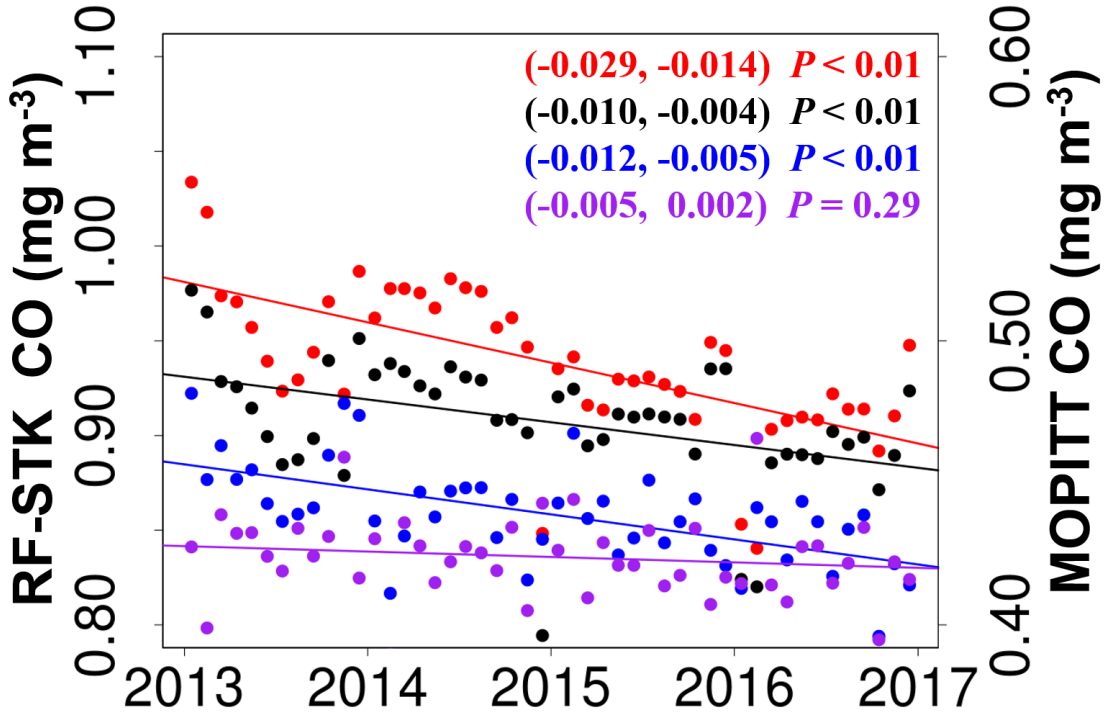

Figure S15: Temporal trends of the population-weighted average ground-level CO concentrations (mg m^{-3}) for China during 2013-2016 based on the actual MOPITT retrieved surface CO (blue solid line), the MOPITT a priori surface CO (purple solid line), the predictions made by the RF-STK model using the actual MOPITT retrieved surface CO (red solid line), and the predictions made by the RF-STK model using the MOPITT a priori surface CO (black solid line). The points in different colors represent the deseasonalized monthly averages for deriving the corresponding trend lines. The 95% confidence intervals of the trends are in parentheses followed by the *P* values.

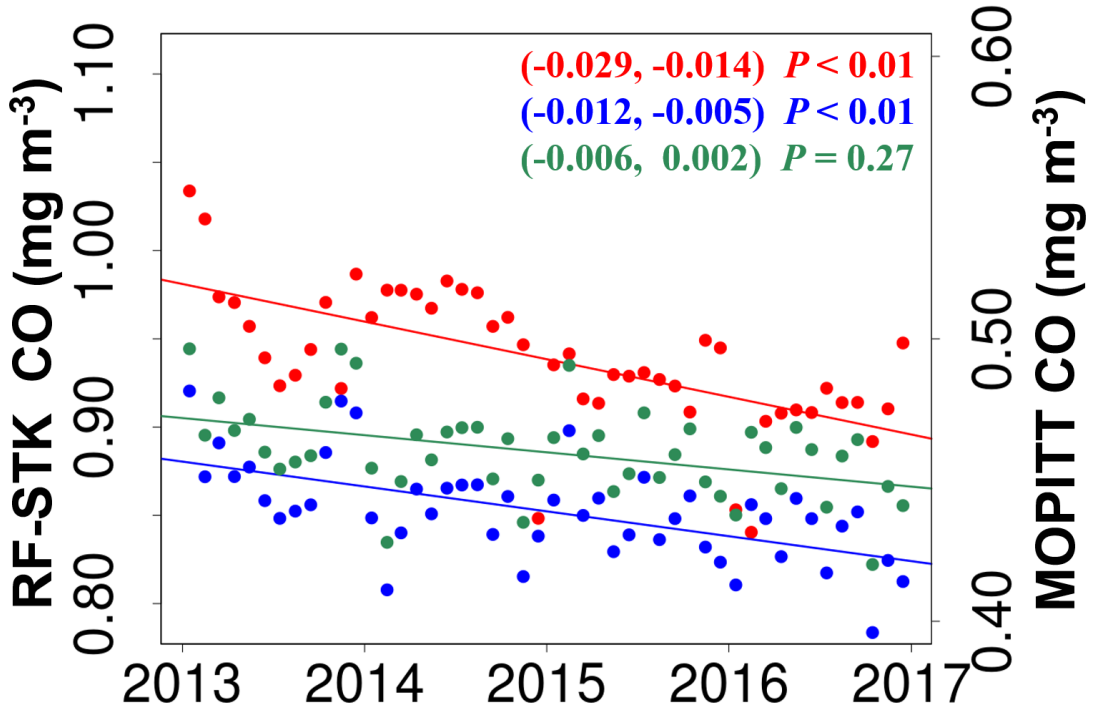

Figure S16: Temporal trends of the population-weighted average ground-level CO concentrations (mg m^{-3}) for China during 2013-2016 based on the actual (blue solid line) and the bias-adjusted (green solid line) MOPITT retrieved surface CO, as well as the predictions made by the RF-STK model using the actual MOPITT retrieved surface CO (red solid line). The bias correction was carried out according to the mean bias drift of -0.69% per year reported in the previous study (Deeter et al., 2017). The points in different colors represent the deseasonalized monthly averages for deriving the corresponding trend lines. The 95% confidence intervals of the trends (mg m^{-3} per year) are in parentheses followed by the *P* values.

References

Breiman, L.: Random Forests, Mach. Learn., 45, 5-32, 2001.

CIESIN: Gridded Population of the World, Version 4 (GPWv4): Population count, NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, 2016.

CMA China meteorology data http://data.cma.gov.cn/, access: 18 Feb 2017, 2017.

China Statistical Yearbook: http://www.stats.gov.cn/tjsj/ndsj/, access: 1 June 2018, 2018.

Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J. F., Ho, S. P., Yudin, V., Attié, J. L., Packman, D., Chen, J., Mao, D., and Drummond, J. R.: Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, J. Geophys. Res-Atmos., 108, 10.1029/2002jd003186, 2003.

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Martínez-Alonso, S., Worden, H. M., and Sweeney, C.: A climate-scale satellite record for carbon monoxide: the MOPITT Version 7 product, Atmos. Meas. Tech., 10, 2533-2555, 10.5194/amt-10-2533-2017, 2017.

Didan, K., Munoz, A. B., Solano, R., and Huete, A.: MODIS vegetation index user's guide (MOD13 series). Version 3.00 (Collection 6), Vegetation Index & Phenology Lab, The University of Arizona, 2015.

Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, Ann. Stat., 29, 1189-1232, 2001.

GMAO: MERRA-2 tavg1_2d_flx_Nx: 2d,1-hourly, time-averaged, single-level, assimilation, surface flux diagnostics V5.12.4, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, 2015.

Hooghiemstra, P. B., Krol, M. C., van Leeuwen, T. T., van der Werf, G. R., Novelli, P. C., Deeter, M. N., Aben, I., and Röckmann, T.: Interannual variability of carbon monoxide emission estimates over South America from 2006 to 2010, J. Geophys. Res-Atmos., 117, 10.1029/2012jd017758, 2012.

Hu, J., Chen, J., Ying, Q., and Zhang, H.: One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system, Atmos. Chem. Phys., 16, 10333-10350, 10.5194/acp-16-10333-2016, 2016.

Hole-filled seamless SRTM data V4.1. International Centre for Tropical Agriculture (CIAT): http://srtm.csi.cgiar.org, access: 26 Sep 2016, 2016.

Jun, C., Ban, Y., and Li, S.: China: Open access to Earth land-cover map, Nature., 514, 434, 2014.

Li, M., Zhang, Q., Kurokawa, J.-i., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17, 935-963, 10.5194/acp-17-935-2017, 2017.

OpenStreetMap contributors Planet dump.: http://planet.openstreetmap.org, access: 7 Sep 2016, 2016.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, Theory and Practice, World Scientific, 2000.

Yeganeh, B., Motlagh, M. S. P., Rashidi, Y., and Kamalan, H.: Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model, Atmos. Environ., 55, 357-365, 10.1016/j.atmosenv.2012.02.092, 2012.