Supplement for

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

Dongren Liu [a], Baofeng Di [a,b], Yuzhou Luo [c], Xunfei Deng [d], Hanyue Zhang [a], Fumo Yang [a,e], Michael L. Grieneisen [c], Yu Zhan [a,e,f *]

[a] Department of Environmental Science and Engineering, Sichuan University, Chengdu 610065, China
[b] Institute for Disaster Management and Reconstruction, Sichuan University, Chengdu 610200, China
[c] Department of Land, Air, and Water Resources, University of California, Davis CA 95616, United States
[d] Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China
[e] National Engineering Research Center for Flue Gas Desulfurization, Chengdu 610065, China
[f] Sino-German Centre for Water and Health Research, Sichuan University, Chengdu 610065, China
[g] Medical Big Data Center, Sichuan University, Chengdu 610041, China

Number of pages: 22
Number of figures: 10
Number of tables: 9

S.1 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 population density, road density, as well as land use data were extracted from the Gridded Population of the World, the OpenStreetMap, and GlobeLand30 database, respectively (CIESIN, 2016; OSP, 2016; Jun et al., 2014).
- The planetary boundary height (PBLH) was obtained from the Modern-Era Retrospective Analysis for Research and Application (GMAO, 2015).
- The Normalized Difference Vegetation Index (NDVI) were retrieved from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite retrievals (Didan et al., 2015).
- The monthly emission inventories were obtained from the Multi-resolution Emission Inventory for China (MEIC) database (Data under known emission inventories for 2012, 2014, and 2016 were processed with linear interpolation to get the contributions of emissions for the remaining years of 2013 and 2015) (Li et al., 2017).

S.2 Algorithm of random forests (Breiman, 2001)

For *tree* = 1 to 500:

- Randomly draw a sample from the training data with replacement, and the size of each sample is the same as the training data;
- A tree is grown from a single node, and the following steps are repeated until only one observation is present in each terminal node:
  - ✧ Randomly select one third of the predictors as potential splitters;
  - ✧ Find the split that reduces the squared error the most;

Average the predictions of all trees as the model output.

Table S1. The MOPITT-CO (mean ± standard deviation; mg m^{-3}) for the whole nation, North China Plain (NCP), and the Central Tibetan Plateau (CTP) during 2013-2016.

Region	Spring	Summer	Fall	Winter	Annual
Nation	0.25 ± 0.18	0.22 ± 0.17	0.23 ± 0.16	0.29 ± 0.16	0.25 ± 0.17
NCP	0.40 ± 0.24	0.40 ± 0.21	0.42 ± 0.21	0.52 ± 0.20	0.44 ± 0.22
CTP	0.14 ± 0.08	0.18 ± 0.14	0.14 ± 0.08	0.14 ± 0.07	0.15 ± 0.09
Naqu	0.15 ± 0.05	0.15 ± 0.01	0.20 ± 0.21	0.13 ± 0.10	0.17 ± 0.02
Qamdo	0.15 ± 0.02	0.13 ± 0.02	0.12 ± 0.02	0.11 ± 0.02	0.13 ± 0.44

[a] Region: We ranked the multi-year average CO concentrations of all sites in the CTP region, with Naqu (31.4733° N, 92.0505° E), Qamdo (31.1254° N, 97.1809° E), Naqu (31.4846° N, 96.0657° E), and Qamdo (31.1278° N, 97.1804° E) in the top four. Naqu: two monitoring sites in Naqu area; Qamdo: two monitoring sites in Qamdo area.

Table S2. Coverage rates of MOPITT-CO satellite 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. S1 for the coverage maps.

Table S3. The ground-level CO observations (mean ± standard deviation; mg m^{-3}) for the whole nation, North China Plain (NCP), and the Central Tibetan Plateau (CTP) during 2013-2016.

Region	Spring	Summer	Fall	Winter	Annual
Nation	1.97 ± 0.24	0.81 ± 0.17	1.01 ± 0.25	1.39 ± 0.38	1.04 ± 0.29
NCP	1.34 ± 0.24	1.09 ± 0.16	1.48 ± 0.22	2.20 ± 0.39	1.52 ± 0.28
CTP	0.98 ± 0.32	0.73 ± 0.24	0.93 ± 0.27	1.23 ± 0.34	0.96 ± 0.30
Naqu	1.60 ± 0.78	0.92 ± 0.52	1.99 ± 0.98	2.29 ± 1.26	1.54 ± 0.50
Qamdo	1.73 ± 1.19	0.62 ± 0.39	1.38 ± 0.78	1.61 ± 0.99	1.34 ± 0.44

[a] Region: We ranked the multi-year average CO concentrations of all sites in the CTP region, with Naqu (31.4733° N, 92.0505 °E), Qamdo (31.1254° N, 97.1809° E), Naqu (31.4846° N, 96.0657° E), and Qamdo (31.1278° N, 97.1804° E) in the top four. Naqu: two monitoring sites in Naqu area; Qamdo: two monitoring sites in Qamdo area.

Table S4. Correlations between ground-level CO observations and MOPITT-CO during 2013-2016 for China ($\mu \pm \sigma$; mg m^{-3}) (unit conversion: 1.0 ppb = 873.36 mg m^{-3}).

Region	2013	2014	2015	2016	Spring	Summer	Fall	Winter	Annual
Central China	0.62	0.22	0.33	0.24	0.09	0.19	0.21	0.28	0.30
East China	0.44	0.34	0.46	0.37	0.12	0.24	0.34	0.36	0.41
North China	0.35	0.33	0.39	0.39	0.29	0.26	0.34	0.45	0.37
Northeast China	0.23	0.39	0.36	0.33	0.29	0.34	0.34	0.29	0.34
Northwest China	0.51	0.39	0.33	0.28	0.29	0.12	0.27	0.28	0.35
South China	0.65	0.55	0.55	0.57	0.46	0.52	0.59	0.50	0.58
Southwest China	0.18	0.02	0.18	0.23	0.05	0.12	0.23	0.21	0.17
Nation	0.44	0.33	0.36	0.37	0.24	0.33	0.34	0.36	0.44
CTP	-0.18	-0.07	-0.09	0.13	0.14	0.24	0.10	0.01	0.01
NCP	0.36	0.30	0.27	0.36	0.18	0.13	0.25	0.35	0.32

Table S5. Performance of the refined RF-STK model in predicting the ground-level CO concentrations for China during 2013-2016.

Metric[a]	Daily	Monthly	Seasonal	Annual	Spatial
R^2	0.51	0.58	0.65	0.65	0.71
RMSE	0.54	0.40	0.33	0.27	0.22
Slope	0.64	0.66	0.71	0.71	0.75
RPE	50.4%	37.1%	30.3%	24.3%	20.4%
MFB	-0.022	-0.008	-0.013	-0.018	-0.021
MFE	0.35	0.24	0.21	0.17	0.15
MNB	0.701	0.420	0.087	0.011	0.002
MNE	0.98	0.63	0.27	0.18	0.15

[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.

Table S6. Previous studies modeling CO concentrations.

Reference	Model	Study Area	Study Period	Validation	Metric
(Li and Liu, 2011)	CTM	Beijing, China	2000-2010	-	-
(Peng et al., 2007)	MOZART-2	North China	2004	Fitting (Point-by-point comparison)	R^2 = 0.74 (monthly)
(Hu et al., 2016)	CMAQ	China	March to December 2013	Fitting	MNE = 0.59~0.66(monthly)
(Tilmes et al., 2016)	GAM4-CHEM	Global	1995-2011	Correlation	-
(Strode et al., 2016)	GEOSCCM	East. USA; Europe; East. China	2000-2010	-	R = 0.26 (monthly; East. USA) R = 0.39 (monthly; Europe) R = 0.061 (monthly; East. China)
(Yeganeh et al., 2012)	SVM; PLS-SVM	Tehran, Iran	January 2007 to January 2011	Fitting	R^2 = 0.56 (daily; SVM) R^2 = 0.65 (daily; PLS-SVM)
(Hooghiemstra et al., 2012)	4D-Var system	Bahia, Brazil	2007-2009	Bias; correlation; root mean square (RMS)	(Prior vs Posterior) Bias = -8.6; -1.0 RMS = 13.4; 7.9 R = 0.6; 0.8

GEOSCCM: Goddard Earth Observing System (GEOS) chemistry-climate Model (CCM).
4D-Var system: Assimilating National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA/ESRL) Global Monitoring Division (GMD) surface and MOPITT observations jointly by fitting a bias correction scheme.
-: not available.

Table S7. Comparisons of the statistical models in predicting daily [CO] for China during 2013-2016 with the same setting of 10-fold cross-validation.

Metric[a]	RF_0[b]	RF_{fo}[b]	RF_{rw}[b]	RF_{fw}[b]	RF_{fw}-STK[b]	RF_{rw}-STK[b]
R^2	0.56	0.53	0.54	0.53	0.49	0.51
Slope	0.60	0.55	0.57	0.55	0.63	0.64
RMSE	0.50	0.52	0.51	0.52	0.55	0.54
RPE	46.1%	48%	47.1%	48.1%	51.0%	50.4%
MFB	0.0832	-0.013	-0.0076	-0.0128	-0.030	-0.022
MFE	0.31	0.31	0.31	0.31	0.36	0.35
MNB	0.90	0.64	0.66	0.64	0.68	0.70
MNE	1.09	0.90	0.91	0.90	0.97	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. Lower values are better for each metric except R^2 and slope.

[b] RF_0: Random Forest model without adding sample-weighting (with variable selection); $RF_{fo:}$ Random Forest model without adding sample-weighting (with variable selection); RF_{rw}: Random Forest model with adding sample-weighting (with variable selection); RF_{fw}: Random Forest model with adding sample-weighting (without variable selection); RF_{fw}-STK: Random Forest-Spatiotemporal Kriging model with adding sample-weighting (without variable selection); RF_{rw}-STK (i.e., refined RF-STK): Random Forest-Spatiotemporal Kriging model with adding sample-weighting (with variable selection). (Except for RF_0, other models are first trained with the data of log-transformed ground-level CO observations, and then the CO concentrations were predicted as the sums of the STK interpolations and back-transformed RF predictions.)

Table S8. 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
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	90m	-	Spatial
NDVI	-	Normalized Difference Vegetation Index	250m	8 Days	Spatial
POP	people km^{-2}	Population density	30"	-	Spatial
LU10	%	Cultivated land area	30m	-	Spatial
LU20	%	Forest area	30m	-	Spatial
LU30	%	Grassland area	30m	-	Spatial
LU40	%	Shrubland area	30m	-	Spatial
LU50	%	Wetland area	30m	-	Spatial
LU60	%	Waterbody area	30m	-	Spatial
LU80	%	Artificial surface area	30m	-	Spatial
LU90	%	Bareland area	30m	-	Spatial
LU100	%	Permanent frozen land area	30m	-	Spatial
LU255	%	Sea area	30m	-	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 convolution. Spatial: these variables have accompanying variables processed with the spatial convolution.

Table S9. The proportion of annual average sources of various pollutants in the emission inventory for China during 2013-2016 (unit: metric ton).

Emission[a]	Agriculture[b]	Industry[b]	Power[b]	Residential[b]	Transportation[b]
CO	0	63 (39%)	4.5 (3%)	66 (41%)	27 (17%)
OC	0	0.48 (17%)	4.7 (1.7×10^{-7}%)	2.2 (79%)	0.1 (4%)
BC	0	0.49 (32%)	0.002 (10^{-1}%)	0.78 (50%)	0.28 (18%)
VOC	0	18 (62%)	0.08×(1%)	5.4 (19%)	5.2 (18%)
NH_3	9.8 (93%)	0.33 (3%)	0	0.36 (3%)	0.04 (1%)
SO_2	0	12 (60%)	4.8 (23%)	3.2 (15%)	0.3 (2%)

[a] The units of gridded emissions are M (million) mol $Grid^{-1}$ for speciated VOC and T (Ton) $Grid^{-1}$ for all other species.
[b] Meaning of data representation: Total emission of each part on the left and percentage of emissions by sector on the right.

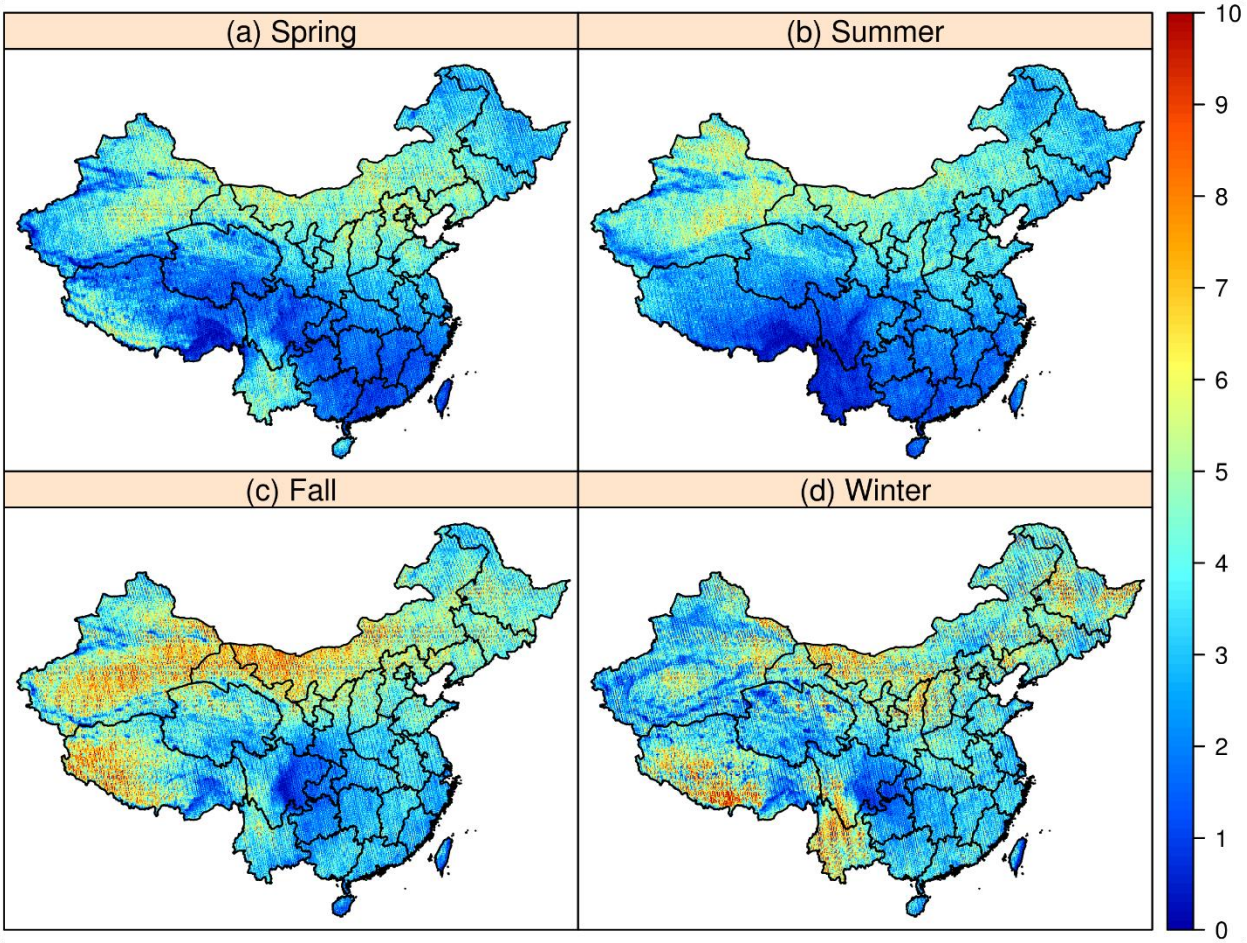

Figure S1. Seasonal Coverage rates of MOPITT-CO satellite retrievals for (a) spring, (b) summer, (c) fall, and (d) winter during 2013-2016 (unit: %).

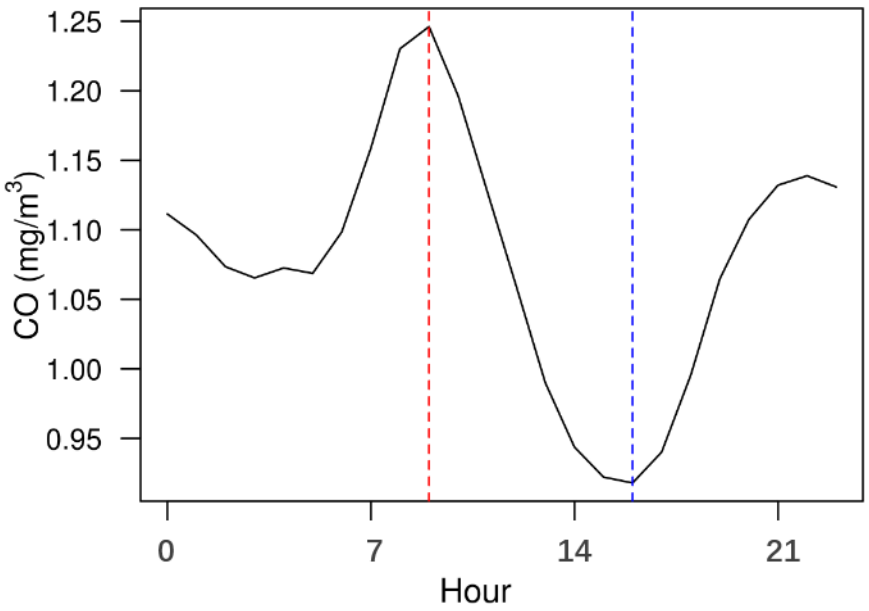

Figure S2. Average diurnal pattern in CO concentrations across 1656 monitoring sites for China during 2013-2016. The peak and the valley appeared at 9am and 4pm, which are indicated by the red line and the blue line, respectively.

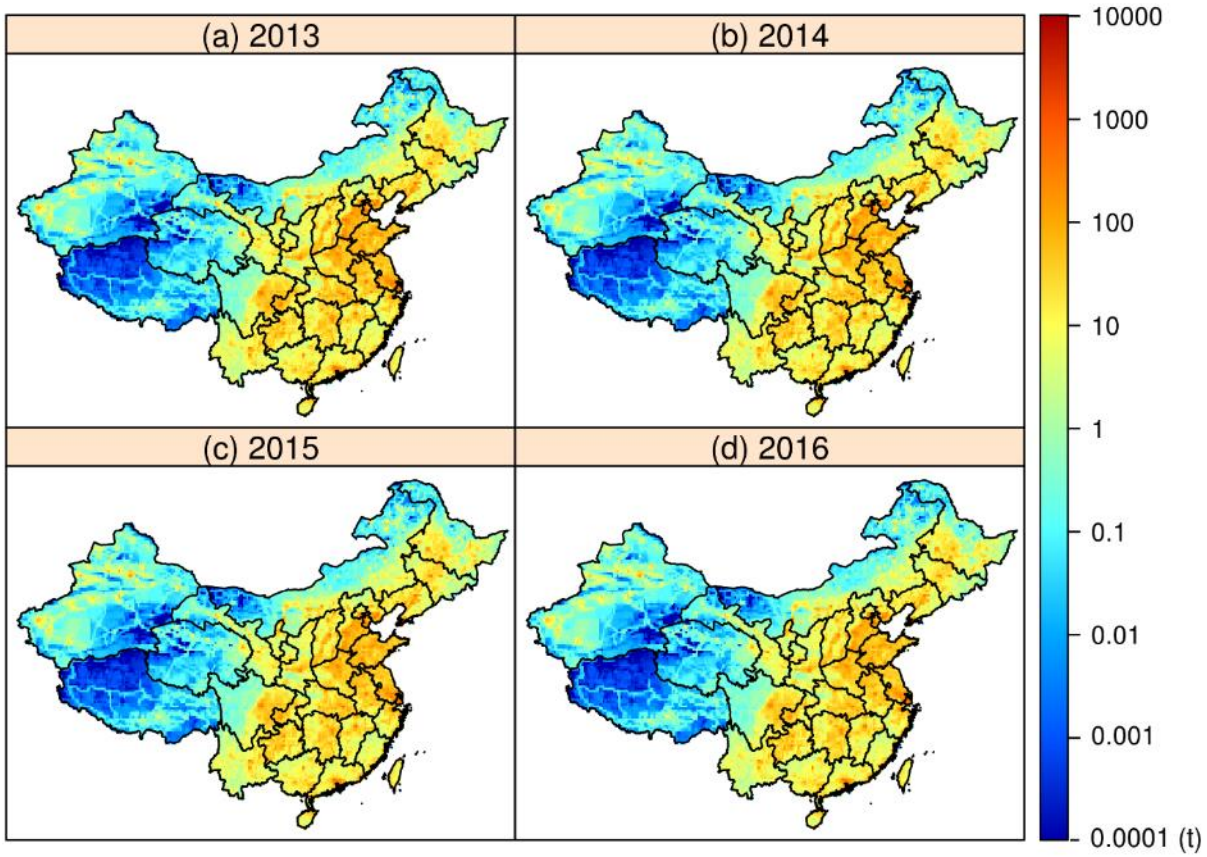

Figure S3. The distribution of CO emission amounts across China from 2013 to 2016 based on the emission inventories of MEIC and MIX (Li et al., 2017).

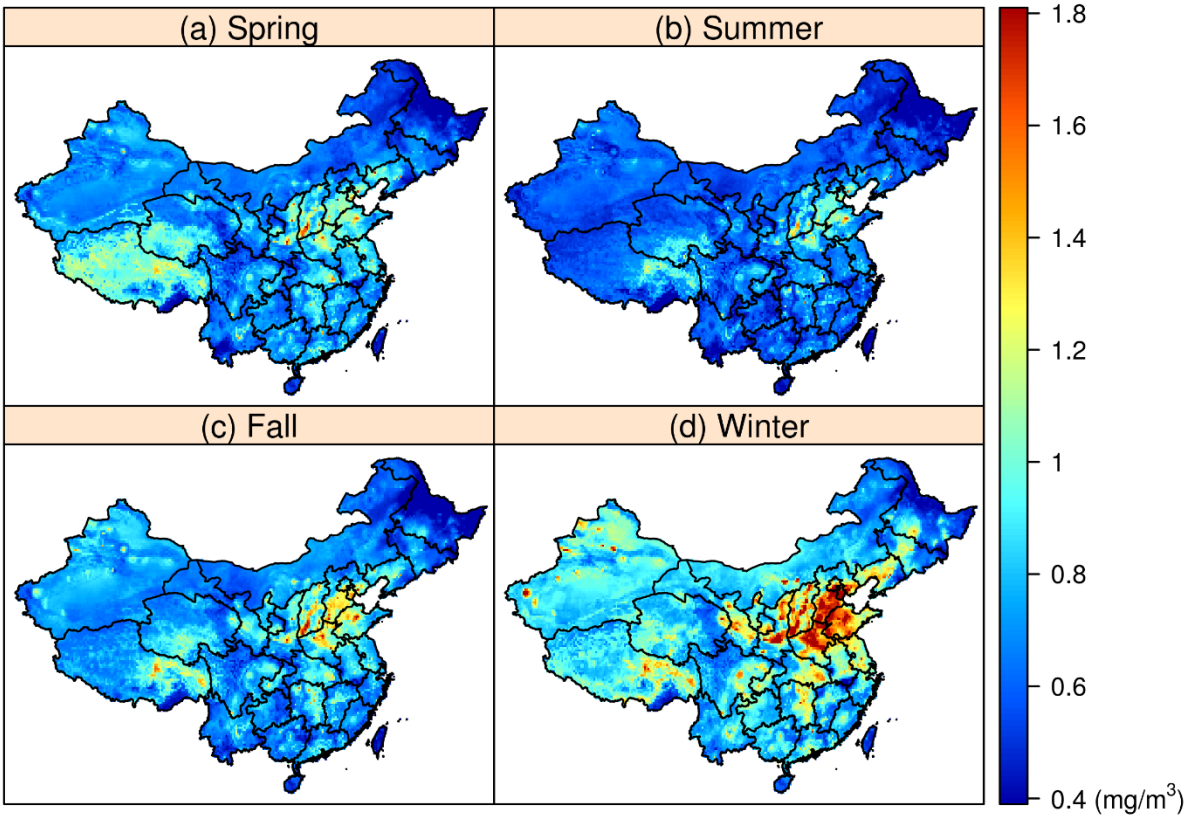

Figure S4. Seasonal averages of the ground-level CO concentrations by the refined RF-STK predicted for (a) spring, (b) summer, (c) fall, and (d) winter during 2013-2016.

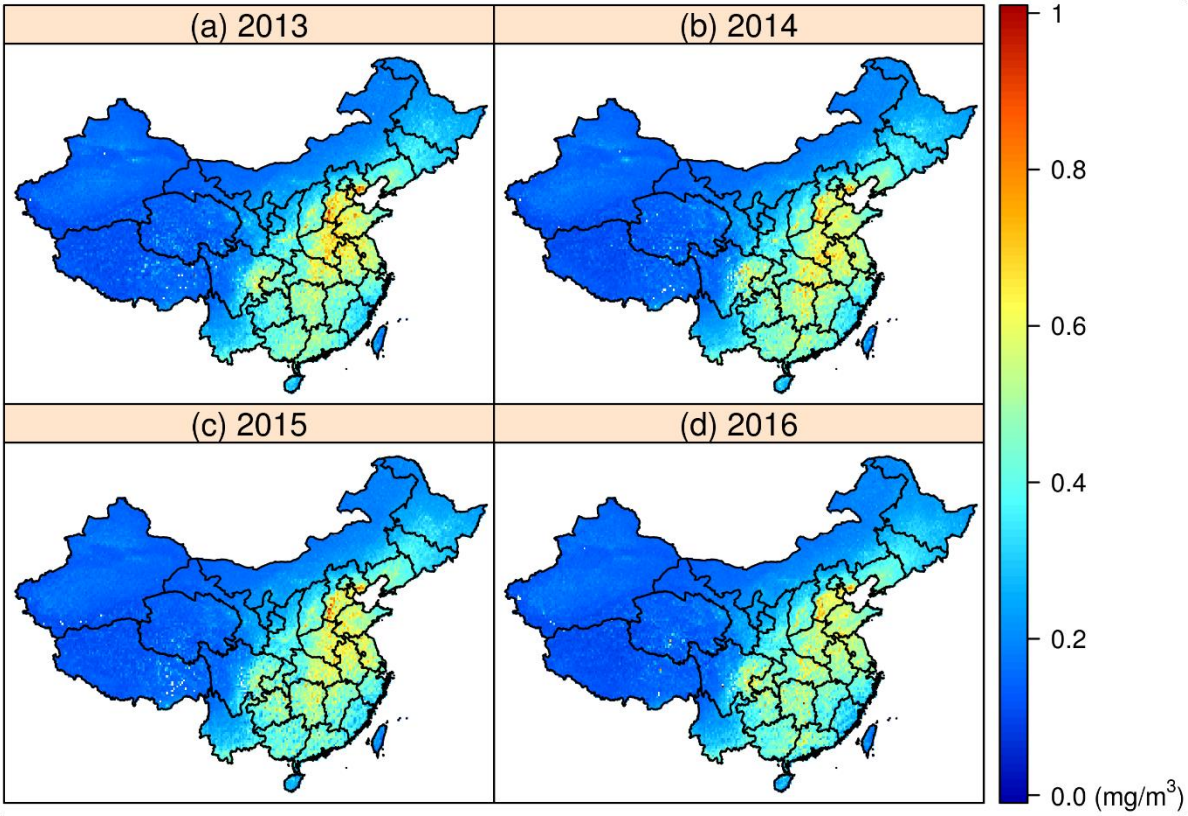

Figure S5. Annual averages of converted surface CO retrieved from the MOPITT Level-2 product for (a) 2013, (b) 2014, (c) 2015, and (d) 2016. (unit conversion: 1.0 ppb = 873.36 mg m^{-3}).

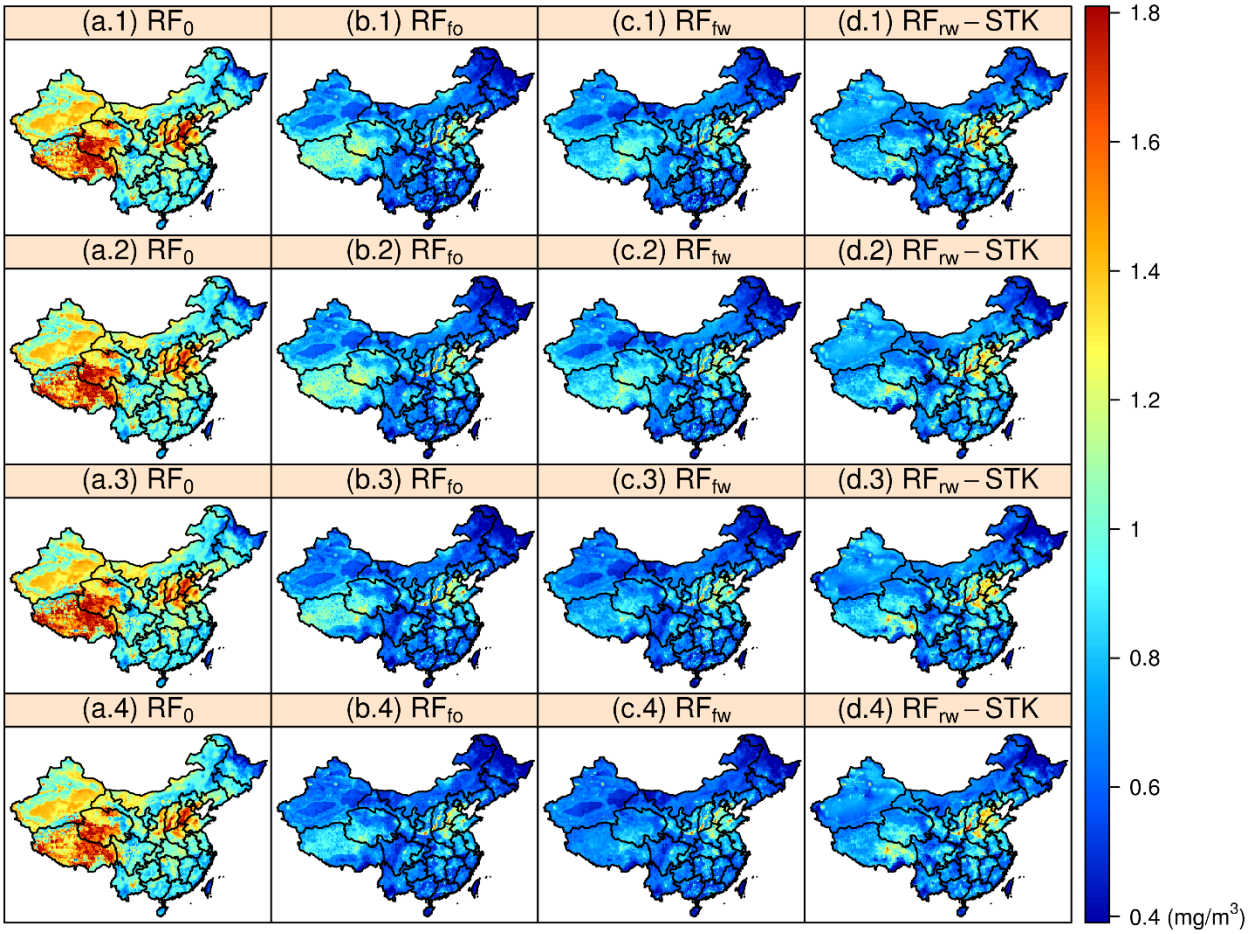

Figure S6. Annual average ground-level CO concentrations from 2013 to 2016 predicted by RF_0: Random Forest model without adding sample-weighting (with variable selection); $RF_{fo:}$ Random Forest model without adding sample-weighting (with variable selection); RF_{fw}: Random Forest model with adding sample-weighting (without variable selection); RF_{rw}-STK (i.e., refined RF-STK): Random Forest-Spatiotemporal Kriging model with adding sample-weighting (with variable selection). (Except for RF_0, other models are first trained with the data of log-transformed CO observations, and then the CO concentrations were predicted as the sums of the STK interpolations and back-transformed RF predictions.)

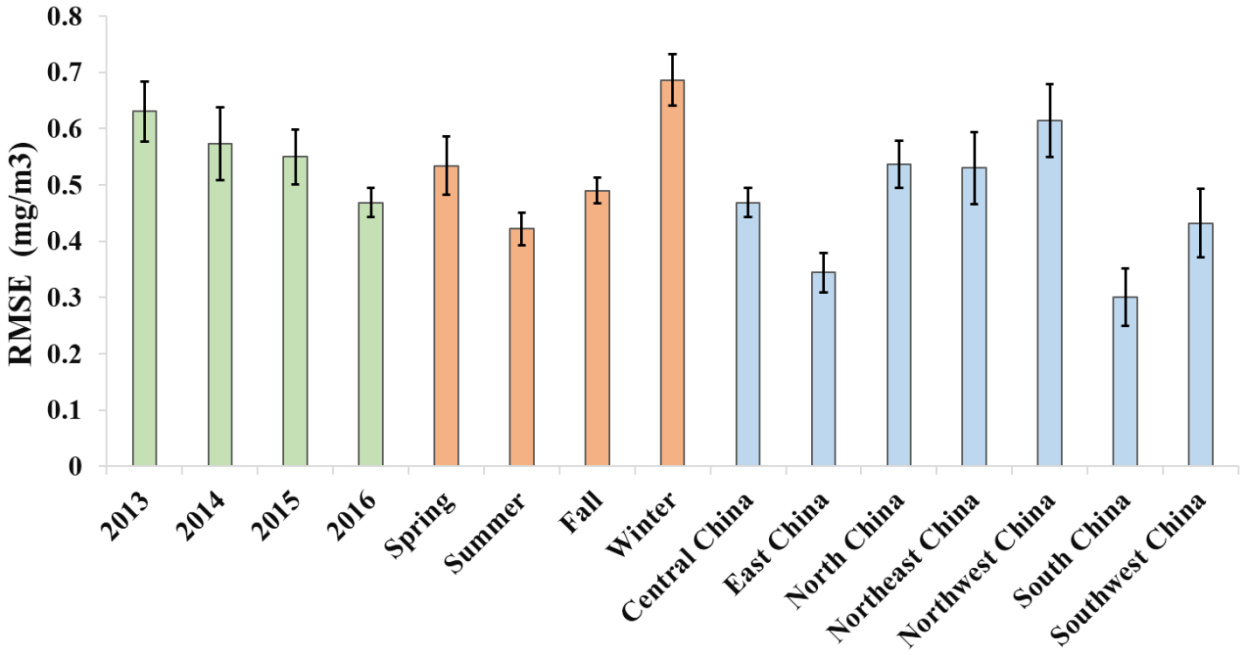

Figure S7. Performance of the refined RF-STK model in predicting daily [CO] by regions, years, and seasons. The mean and standard deviation of the root mean square error (RMSE) over all the 10-fold cross-validations are presented. The annual average numbers of monitoring sites in Central, East, North, 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.

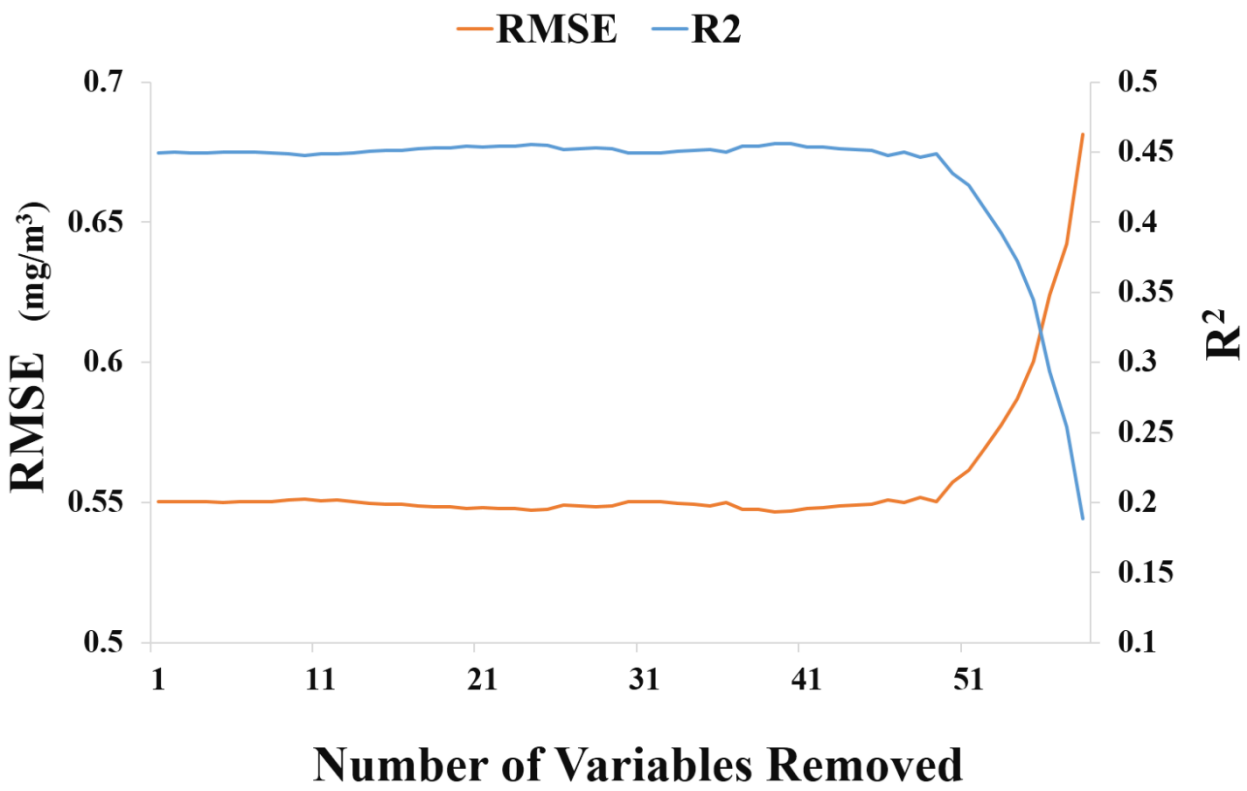

Figure S8. Evolution of cross-validation RMSE (mg m^{-3}) and R^2 for the random forest submodels through the variable selection process. Refer to Table S7 for the detailed descriptions of the variables.

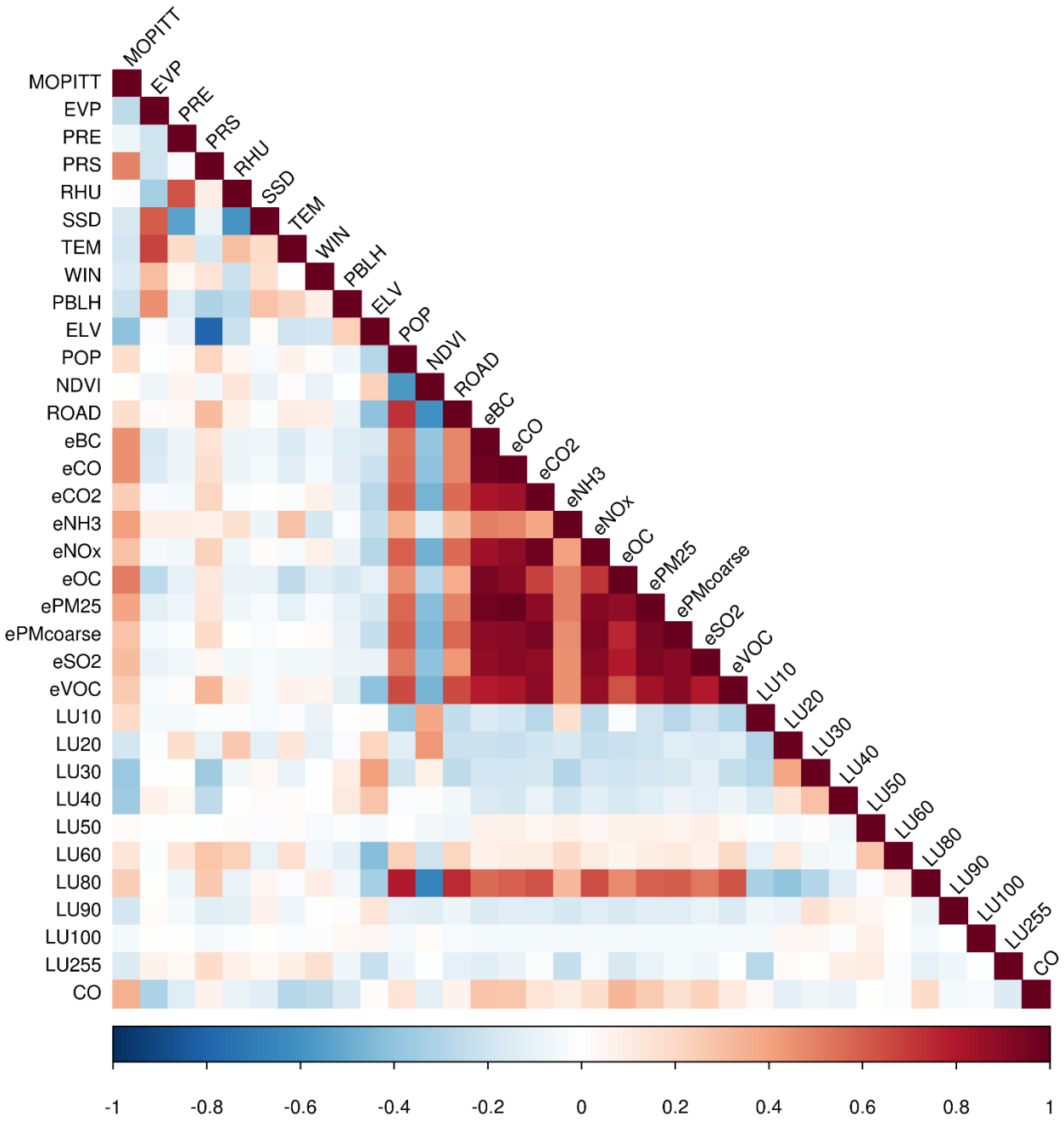

Figure S9. Correlations among the geographic factors and [CO]. The Spearman's rank correlation coefficients are used due to the prevalence of nonlinearity. Please refer to Table S7 for the detailed descriptions of the variable acronyms.

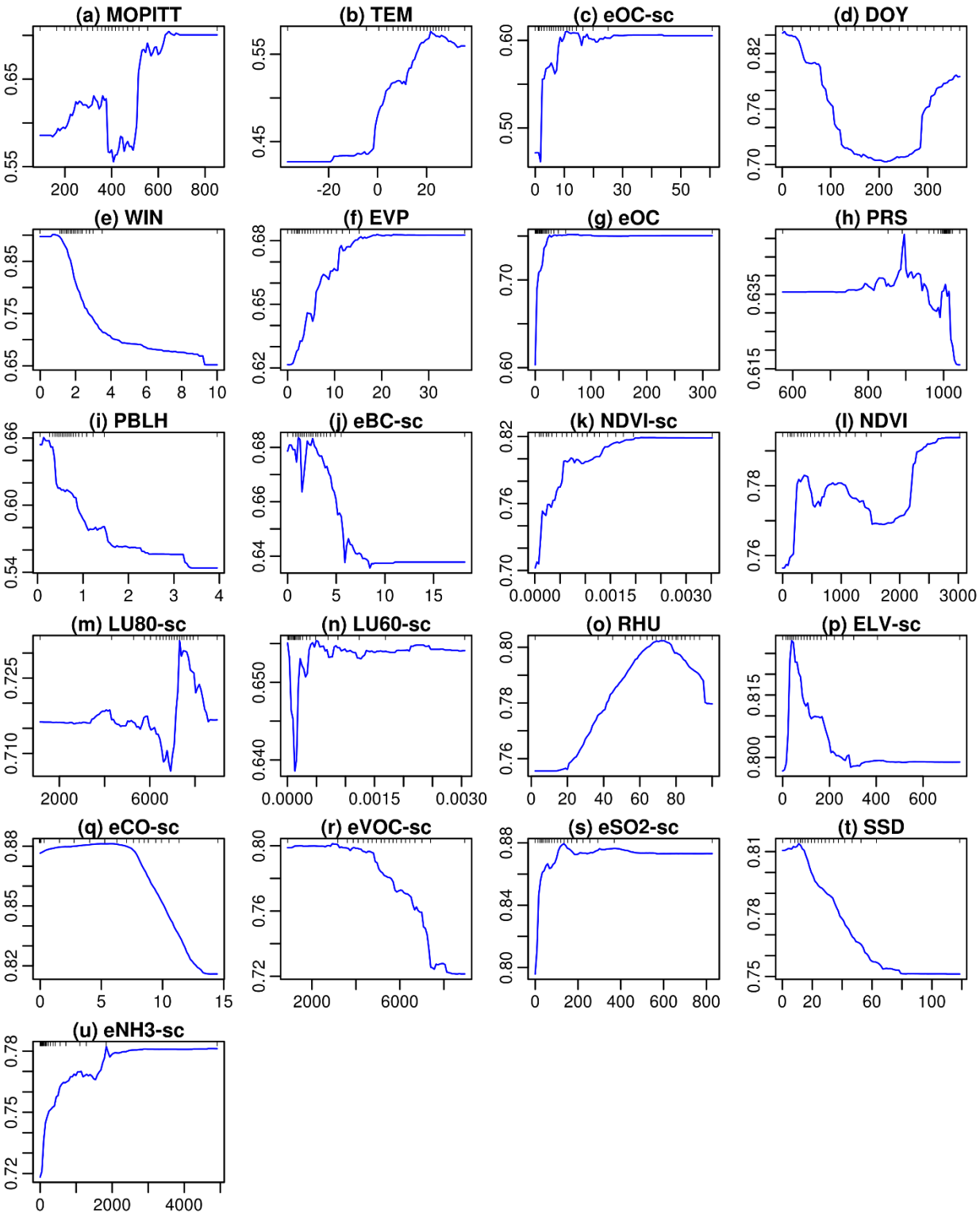

Figure S10. Partial dependence plots for the random forest submodel, indicating the effects of the predictor variables on the [CO] predictions. The rug plot indicates the data density. Note that the partial dependence estimation tends to be unreliable at the two ends of horizontal axis due to their low data densities. Refer to Table S7 for the descriptions of the predictor variables.

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.
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.
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, n/a-n/a, 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, L., and Liu, Y.: Space-borne and ground observations of the characteristics of CO pollution in Beijing, 2000–2010, Atmos. Environ., 45, 2367-2372, 10.1016/j.atmosenv.2011.02.026, 2011.
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.
Peng, L., Zhao, C., Lin, Y., Zheng, X., Tie, X., and Chan, L. Y.: Analysis of carbon monoxide budget in North China, Chemosphere., 66, 1383-1389, 10.1016/j.chemosphere.2006.09.055, 2007.
Strode, S. A., Worden, H. M., Damon, M., Douglass, A. R., Duncan, B. N., Emmons, L. K., Lamarque, J.-F., Manyin, M., Oman, L. D., Rodriguez, J. M., Strahan, S. E., and Tilmes, S.: Interpreting Space-Based Trends in Carbon Monoxide, Atmos. Chem. Phys. Discuss., 1-21, 10.5194/acp-2016-87, 2016.
Tilmes, S., Lamarque, J. F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., Smith, A. K., Neely, R. R., Conley, A., Vitt, F., Val Martin, M., Tanimoto, H., Simpson, I., Blake, D. R., and Blake, N.: Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-ClimateModel Initiative (CCMI), Geosci. Model Dev. Discuss., 1-50, 10.5194/gmd-2015-237, 2016.
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.