Table of Contents

*** Note: Complete details of model set-up are available as part of our Replication Data Set at https://dx.doi.org/10.7910/DVN/OJESO0 ***

S1 WRF Model: Post-processing and Evaluation

We evaluate WRF output against publicly available, 24h-averaged Chinese Meteorological Administration (CMA) observational data. CMA observational data is not used in the NCEP FNL reanalysis WRF initialization fields. CMA provides daily averages of surface pressure, wind speed, temperature, and relative humidity. Access to higher temporal resolution observational data is limited. We convert hourly (d01) and half-hourly (d02, d03) WRF output to daily averages before evaluation. We use a combination of NCAR Command Language v6.1.2 (NCL; http://dx.doi.org/10.5065/D6WD3XH5) and R v2.9.0 (https://www.r-project.org/) to process the observed and simulated output. The standard post-processing toolbox, consisting of the WRF Unified Post Processor and METv4.1 Point-Stat Tool (http://www.dtcenter.org/code/) is not used here because of the low temporal resolution of observational data and file format mismatches. However, we base our evaluation method and procedures on the METv4.1 Point-Stat Tool. Both the METv4.1 and our version of the Point-Stat tool match WRF forecast fields to observation point locations for comparison. For surface observations, no interpolation is performed. Forecasts are instead matched to nearest CMA surface station observation point. Fig. S1 displays a map of the CMA surface network in 2006 and 2008, with approximate WRF domains overlaid with CMA station 54511 (C54511; 39.8N, 116.47E) highlighted in d03. We display sample evaluation results from C54511 in Fig. S3 through Fig. S5, using observed and simulated fields from 2006. In the evaluation, WRF forecast fields are matched to the nearest observation point.

Comparing against publicly available 2006 CMA data from 35 stations across the d02 and d03 domains (Fig S1), the median modeled wind speed was 15% higher than observations, with a median absolute deviation of 16%. We emphasize that a more robust evaluation of WRF windspeed (or other meteorological) biases relative to observations would require access to higher temporal resolution meteorological observations. Currently, we are restricted by data availability to 24-hour averages which blur smaller timescale processes and therefore likely underestimates the WRF surface wind speed bias relative to observations. We do not include d01 comparisons in this analysis, as the distance between nearest station and WRF gridcell center can be on the order of tens of kilometers, decreasing the information and value of the comparison. The graphics associated with the d02 and d03 comparisons are available from https://dx.doi.org/10.7910/DVN/OJESO0 as "006_WRFvCMAplots_2006_d0X.pdf".

S2 STILT Model Set-up and Run Details

The version of WRF-STILT[1] used in this study corresponds to STILT release r701 of the AER-NOAA branch at the STILT svn repository[2], and Release-3-5 of the WRF-STILT interface[3]. Spin-up periods are removed from the WRF meteorological data and the WRF netcdf output files are converted to .arl format (Air Research Laboratory; https://ready.arl.noaa.gov/HYSPLIT_data2arl.php#INFO) prior to being ingested into STILT.

In this study, we transport an ensemble of 500 particles 7-days back in time to model footprints for each measurement hour at the receptor. The receptor (Miyun; 40°29′N, 116°46.45′E, 152 m above sea level (asl)) has the measurement inlet (STILT particle "release" point) located 6m above ground level (agl) (Fig. S2). We employ dynamic regridding, which accounts for resolution changes among the nested WRF domains. Mixing height is derived from WRF PBL heights; we set the surface layer as 50% of the mixed layer height. Footprints are integrated hourly. We set up the STILT runs as "pleasantly parallel" by running each month of a year simultaneously; hours within a month are run serially.

When the receptor release occurs outside of peak daylight hours, stratification of the PBL becomes significant. Therefore, as is common practice in virtually all emissions optimization/assessment studies, we model the 1100 to 1600 (local time) subset. These daylight hours represent a typical window for which STILT reliably models transport (e.g., 4). We examine the unadjusted model performance at all times, averaged seasonally and diurnally, in Sec S7.

S3 Anthropogenic CO_2 inventories

In order to facilitate comparison among the three anthropogenic inventories used in this study, we interpolate the two global inventories (EDGAR, 0.1ºx0.1º; CDIAC, 1ºx1º) to the same 0.25ºx0.25º grid as the regional inventory (ZHAO). We use the NCL Earth System Modeling Framework (ESMF) Conserve regridding method which minimizes deviation of the variable's integral between source and destination grids. We evaluate the impact of regridding in Fig. S6 by comparing annual totals ($MtCO_2$) before and after regridding. The ZHAO inventory remains on its native grid. We show that regridding does not appreciably affect the total emissions reported for mainland China by EDGAR and CDIAC, providing confidence in our representation of the two original inventories.

The ZHAO inventory provides estimates of total annual emissions for 2005 through 2009. In addition, the 2005 and 2009 ZHAO emissions are spatially allocated to a 0.25º x0.25º grid. We average the 2005 and 2009 percent contributions of each grid cell to the total emissions to provide weights for spatially allocating 2006 through 2008 total annual emissions. Fig. S7 evaluates the validity of this assumption by identifying regions where the 2009 gridcell contribution to the total emissions is outside +/- 2% of its 2005 contribution (Fig. S7a) and +/-50% of its 2005 contribution

[1] https://www.bgc-jena.mpg.de/bgc-systems/projects/stilt/pmwiki/pmwiki.php?n=WRFSTILT.WRF-STILT
[2] https://projects.bgc-jena.mpg.de/STILT/svn/branches
[3] available from http://files.aer.com/external/CarbonSoftware

(Fig. S7b). We find the assumption to be valid; the mean change per gridcell from 2009 relative to 2005 is -0.011% with a 2-σ of 15%.

Total uncorrected emissions for each anthropogenic inventory are calculated on the 0.25ºx0.25º grids and provided in Table S1. We provide emissions summed for each administrative region in the study domain, each STILT influence contour, and all China. Differences among the inventories zoomed to the L_90 influence region, are displayed in Fig. S8. Miyun and Beijing are encompassed by the L_0.25 contour. We display the average gridcell emissions of ZHAO (Fig. S8a) and the differences of EDGAR and CDIAC relative to ZHAO (Fig. S8b and Fig. S8c, respectively). In heavily emitting regions, ZHAO is typically higher than EDGAR and CDIAC. In regions where ZHAO is consistently lower than CDIAC, the differences are lower than the instances where ZHAO is higher. Note that, in the case of CDIAC, the uniformity of the differences includes artefacts from downscaling the gridded CDIAC inventory from 1°x1° to 0.25°x0.25°.

S4 CT2015: Background Concentration Selection and Evaluation of Model Bias

We derive estimates of background CO_2 concentrations from NOAA CarbonTracker (CT2015; https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2015/). CT2015 enables us to estimate concentrations of CO_2 prior to interaction with the surfaces in the study domain. Background value selection is summarized as follows. For each hour, the end x-y-z-time coordinates for each of 500 particles is found and linked to its corresponding CT2015 CO_2 concentrations using a spatiotemporal nearest neighbor approach. Only instances where a particle originated at the edge of the outermost domain and/or an altitude greater than or equal to 3000masl is included in the average background concentration calculation for that hour. If less than 75% of particles for an hour have valid background concentrations, that hour is not used in subsequent analyses. This selection criteria for background CO_2 mole fractions enables realistic modeling of true background conditions that have not interacted with the domain within each hourly measurement's maximum seven-day regional influence period. For the five-year study period, this method of boundary selection retains approximately 85% of hourly modelled values per year and across years.

The CT2015 model for the study domain is heavily trained by observations made approximately weekly via flask sampling at four World Meteorological Organization (WMO) sites in the region (https://www.esrl.noaa.gov/gmd/dv/site/). Mt. Waliguan to the west of the receptor (WLG) represents free tropospheric background air; Ulaan Uul (UUM) in Mongolia represents clean continental air; Tae-ahn Peninsula (TAP) in South Korea represents urban-influenced air from the east; Lulin (LLN) in Taiwan represents urban-influenced air from the southeast. TAP and LLN become more prominent in their representation upwind/background air sites during the spring and summer months when the East Asian Monsoon begins to influence regional air trajectory patterns. WLG and UUM are prominent in their representation of upwind/background air at all times of the year but particularly weight background air during the winter and fall seasons.

We quantify bias in the background model by evaluating observations against the nearest CT2015 model pixel and level. Observations are filtered using highest quality flask sample points only. Fig. S10(top panel) displays the time series of 3-hourly modeled CT2015 values and observed WMO measurements. Deviation of residuals from a normal distribution are displayed in Fig. S10

(bottom panel). The typical 1-σ model bias is 2ppm, but not all of the distributions are normal. For UUM, and therefore, CT2015 parameterization of clean continental background, the model-measurement residuals largely follow a normal distribution centered around 0. The clean continental background generally exhibits well-mixed behavior and is not defined by large excursions in the CO_2 signal. At the high-altitude WLG site representative of the free troposphere, the residuals follow a normal distribution centered around 0 but deviate from normal during instances where significant excursions in the CO_2 signal are present. This is also the case at LLN (distribution centered near 2.5ppm). TAP residuals deviate significantly from normal. In general CT2015 does not capture CO_2 events that are significantly different from global means; CT2015 underestimates uptake processes and overestimates lower or higher than global mean.

As not all deviations from observations can be represented as normal distributions, we place the model-measurement residuals at the four WMO sites in an error pool and select as part of the overall bootstrapping procedure for the modeling framework.

As shown in Fig. S10, LLN shows CO_2 depletion relative to CT2015 suggesting that for this analysis it is not representative as a background site. (CT is not responsive to all sites). The LLN observed CO_2 drawdown compared to modeled CT2015 suggests that LLN sees a lot of surface influence on account of its location in the middle of an island in vegetated surroundings. Moreover, LLN is not an important sector for the influence region of this study; we include it primarily for reference for future studies considering regions of China that would be more sensitive to the sector associated with LLN.

S5 Scaling Results and Methodology

We translate the resulting mole fraction (ppm) mismatch between observed and modeled ΔCO_2 to inventory corrections at annual and seasonal timescales. We scale in the L_0.90 contour (Fig. S9) which represents regions that substantially influence the receptor without disproportionally weighting pixels that contribute very little to the observed signal. Table S2 provides seasonal fluxes for each year before and after scaling. Annual scaling results are in Table 2 of the main text.

At annual scales, the dominant contributor to the CO_2 signal are anthropogenic emissions; correction at annual scales is therefore applied only to the anthropogenic emissions inventories. The other significant contributors include longer term biological and ocean carbon sinks and interannual variability within these components, but for this study region, these components are embedded in the background concentrations. In particular, 13% of the northern China ecosystems and 20% of northeastern China's ecosystems are mixed forests. However, the ecosystems with greatest influence on this single site are croplands with high intra-annual carbon turnover rates. The heavily cropped L_0.90 influence region implies rapid turnaround of vegetation carbon stocks at the annual scale, justifying this assumption. At these timescales, we derive the $\Delta CO_{2,obs}/\Delta CO_{2,mod}$ ratio which represents the factor by which the annual anthropogenic inventory must be scaled in order to match observations. We use a model of the mean method to derive the annual scaling factors,

$$SF = \frac{\overline{\Delta CO_2}_{,obs_{hh}}}{\overline{\Delta CO_2}_{,mod_{hh}}}$$

where *hh* represents each local afternoon hour (1100 to 1600) in the year. SF>1 implies the model underestimates CO_2 concentrations while SF<1 implies the model overestimates CO_2 concentrations. We obtain 95% confidence bounds by bootstrapping uncertainties in the numerator and denominator separately, and obtaining the 0.025 and 0.975 quantiles from the ratio of the means of the two distributions. The annual influence contours are overlayed on the IGBP land use map in Fig. S9, and shows the dominant grassland/cropland influence on the modeled Miyun signal at annual scales. As stated previously, The Miyun CO_2 signal is certainly affected by other biological/oceanic/interannual variability; but these are not demonstrated to be significant parts of the regional (ΔCO_2) signal. These are longer term features embedded in the background concentrations.

At the seasonal timescales, however, evaluation of CO_2 processes is complicated by the biogenic flux contribution during the growing season and, to a lesser extent, the effects of ecosystem respiration in the dormant season. At these timescales, we derive additive corrections from converting observation-model mole fraction mismatch to the total CO_2 to be added or subtracted from the inventories. We correct the anthropogenic and vegetation inventories together as it is not possible to distinguish the contributions from our existing observational data set. For each modeled hour we derive a residual-based flux correction, $\Delta\Phi_{hh}$, in $\mu molCO_2 m^{-2} s^{-1}$:

$$\Delta\Phi_{hh} = \frac{\Delta CO_{2,obs_{hh}} - \Delta CO_{2,mod_{hh}}}{\sum_0^{-168h} foot_{hh}}$$

where *hh* represents each local afternoon hour (1100 to 1600) in the season and *h* represents the STILT footprint back-trajectory hour up to 7 days back in time. Given that anthropogenic emissions are positive terms and the biogenic component is a net balance of two opposing terms (uptake and release) of CO_2 during the growing seasons, use of inventory scaling factors for growing season scaling is inappropriate. That is, even a small mole fraction difference between modeled and observed in the growing season can result in meaningless scaling factors when there is a difference in sign involved. While scaling factors are appropriate during dormant seasons, for consistency we apply the same method of additive corrections across all seasons and report the adjusted inventory as fluxes (kg CO_2 m^{-2} $season^{-1}$). The methods are comparable; inventory corrections obtained by both methods during the winter and fall exhibit converging 95% confidence intervals.

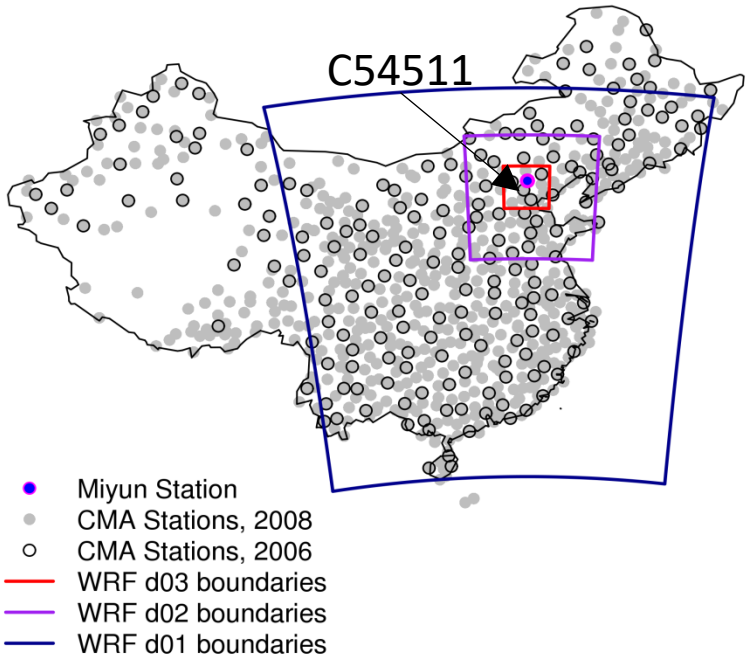

Fig. S1. CMA Station Map (2006, 2008) with WRF domain boundaries. Sample WRF evaluation results are provided for Station 54511 (indicated by arrow on map), near Miyun receptor.

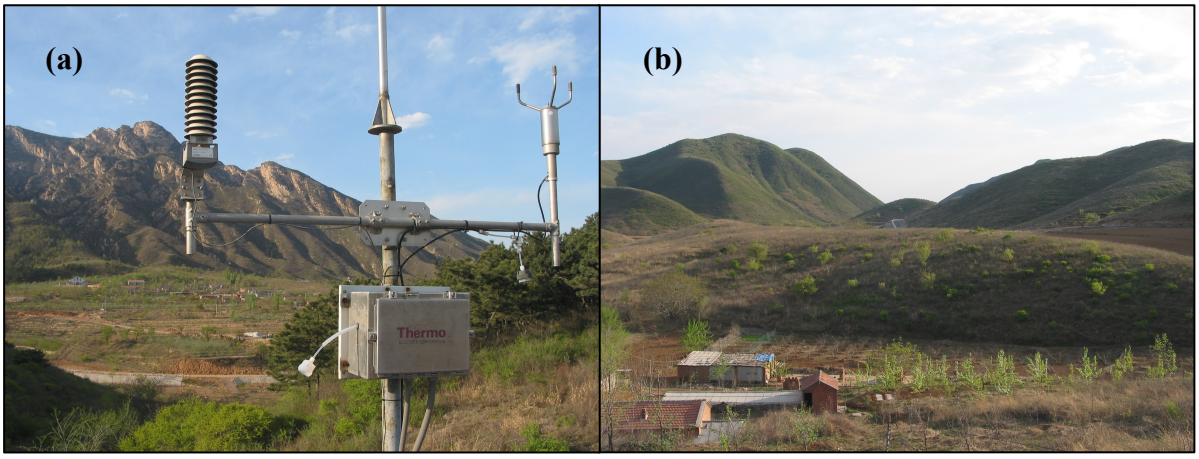

Fig. S2. Miyun Receptor and surroundings, April 2007. (a) Miyun inlet at 6magl/158masl, looking ENE, shows a small rural village in the valley below site, a small patch of short pines, that are generally in downwind direction. Even in spring there is still considerable bare ground. (b) view from Miyun sampling site, looking SW. Foreground shows a farmhouse and various outbuildings that were no longer in active use. Small-scale agricultural fields that were being converted to fruit-tree orchards. Unmanaged lands were grassy/shrub vegetation on hillsides.

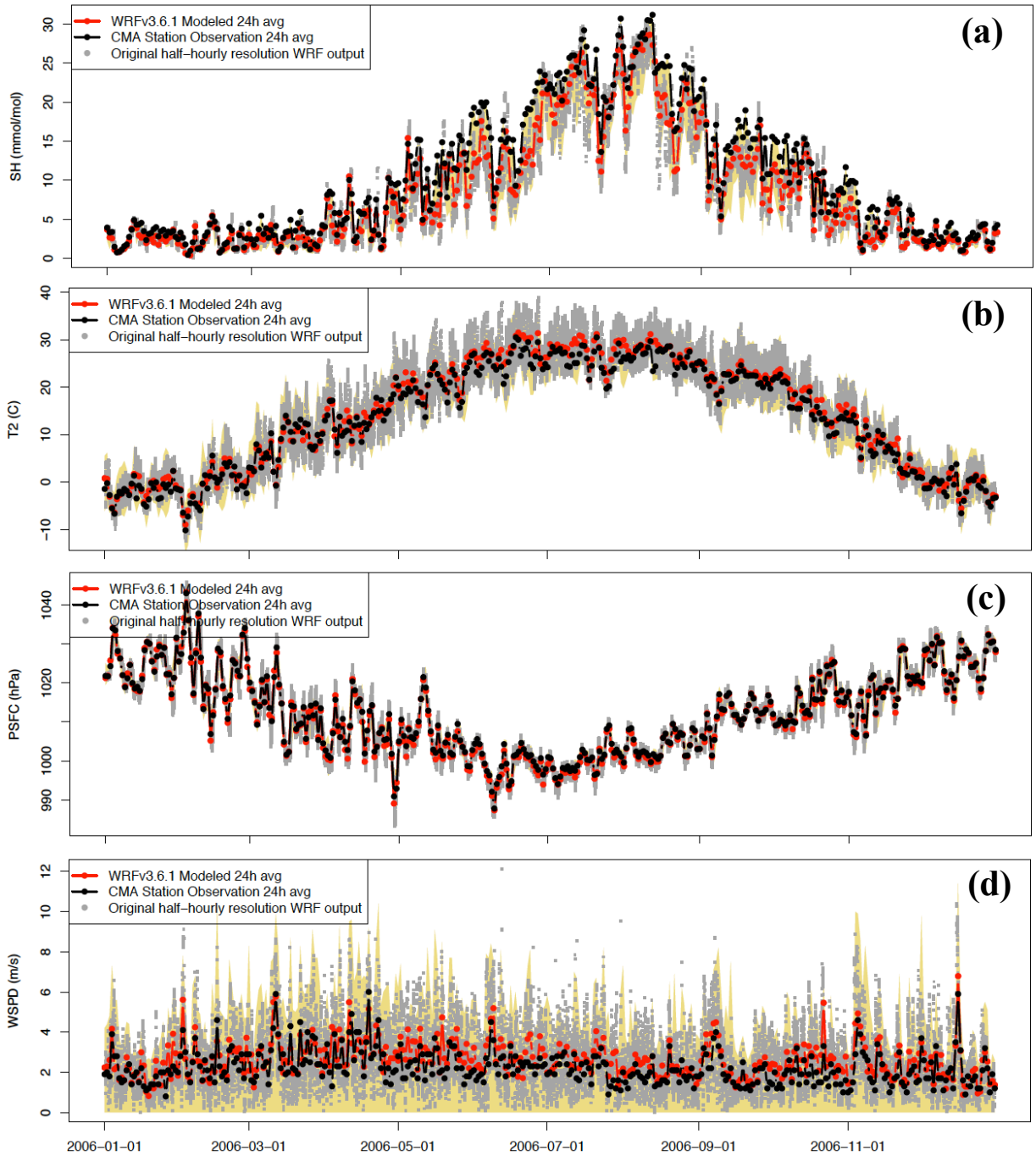

Fig. S3. Evaluation of WRF output against observational data. 2006 Meteorology timeseries for sample WRF gridcell (39.825N, 116.51E) evaluated against nearest CMA Station C54511 (39.800N 116.47E). WRF Meteorology averaged from half-hourly to daily for (a) Specific Humidity; (b) Surface Temperature; (c) Surface Pressure; (d) Surface Wind Speed. Original half-hourly output displayed in grey. Shaded yellow region represents observed daily range; daily minimum for windspeed is not available, but assumed to be 0m/s.

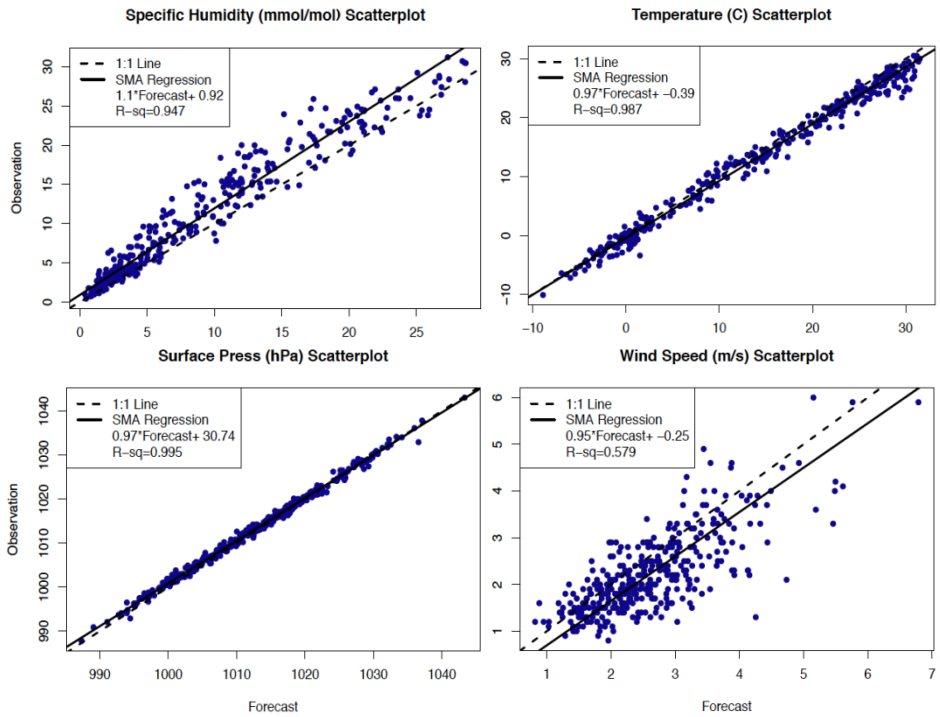

Fig. S4. Observed vs WRF Modeled (Forecast) meteorology for sample WRF gridcell. Gridcell (39.825N, 116.51E) evaluated against nearest CMA Station C54511 (39.800N 116.47E). Time-base of fields is daily average.

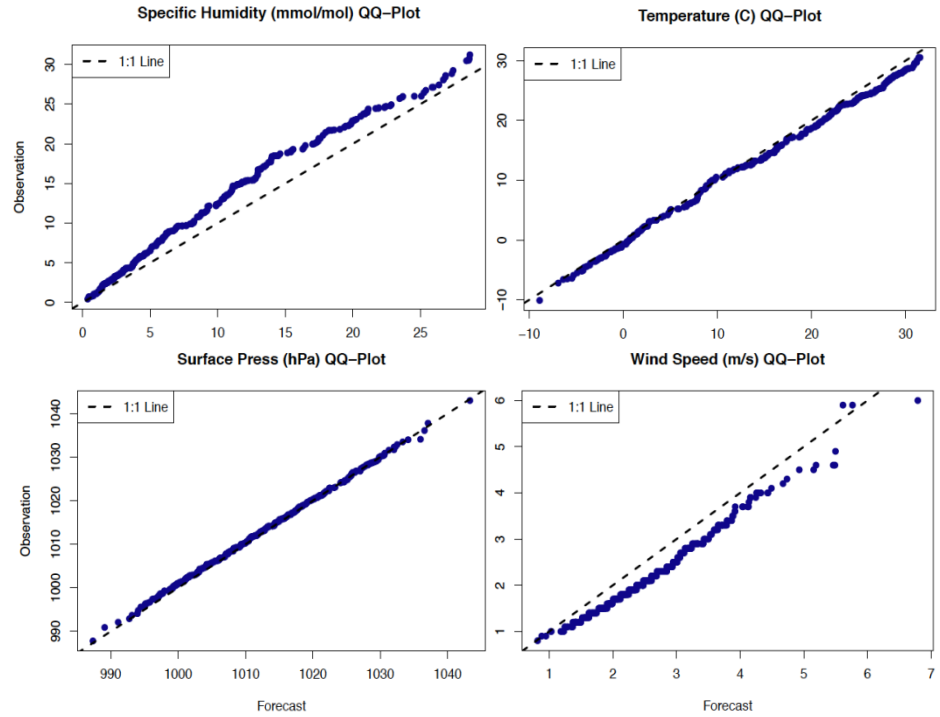

Fig. S5. Q-Q plots of Observed and WRF Modeled (Forecast) meteorology for sample WRF gridcell. Gridcell (39.825N, 116.51E) evaluated against nearest CMA Station C54511 (39.800N 116.47E). Time-base of fields is daily average.

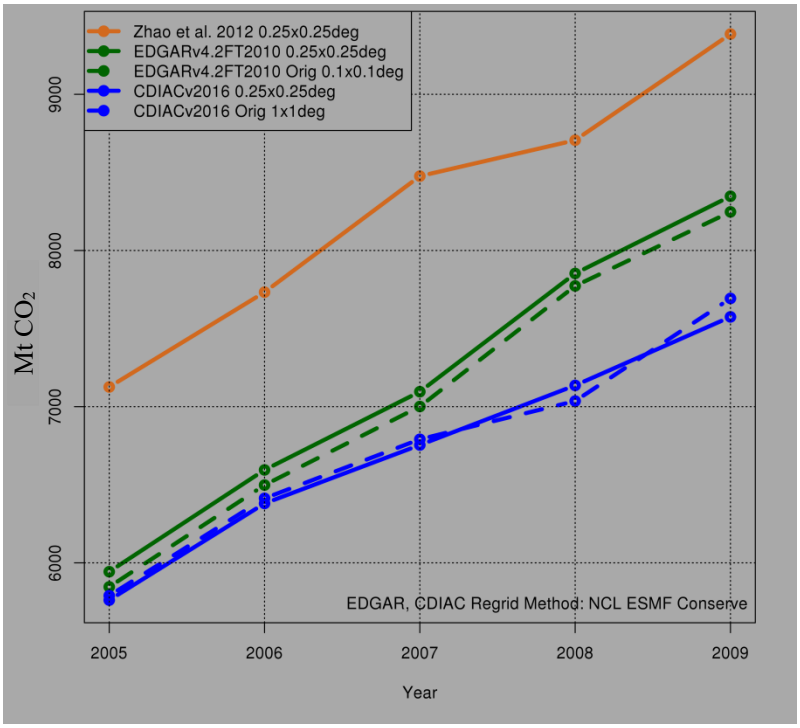

Fig. S6. ZHAO, EDGAR, and CDIAC estimates of total annual CO_2 emissions for Mainland China, 2005 to 2009. EDGAR and CDIAC are regridded to 0.25ºx0.25º grid using the NCL Earth System Modeling Framework Conserve regridding function.

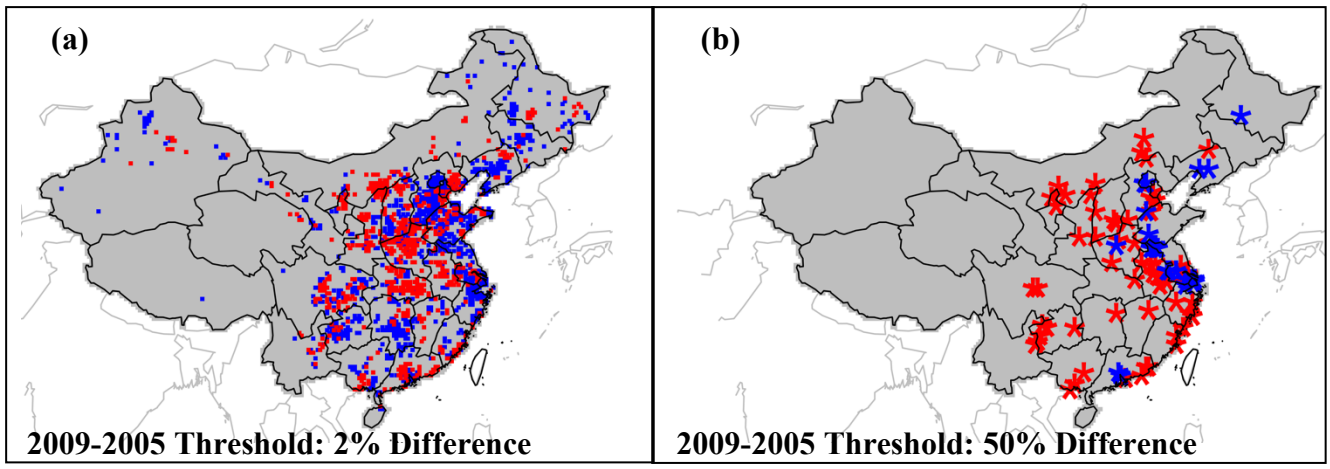

Fig. S7 Spatial Allocation of ZHAO inventories (2006-2008). Mean percent difference of gridcell contribution to total emissions is -0.011% ±15% (2-σ). We highlight instances where 2009 gridcell contribution to total annual emissions differs from its 2005 value by (a) more than 2% and (b) more than 50%. Blue represents a relative DECREASE in 2009 relative to 2005; red represents a relative INCREASE; grey represents values WITHIN the specified threshold.

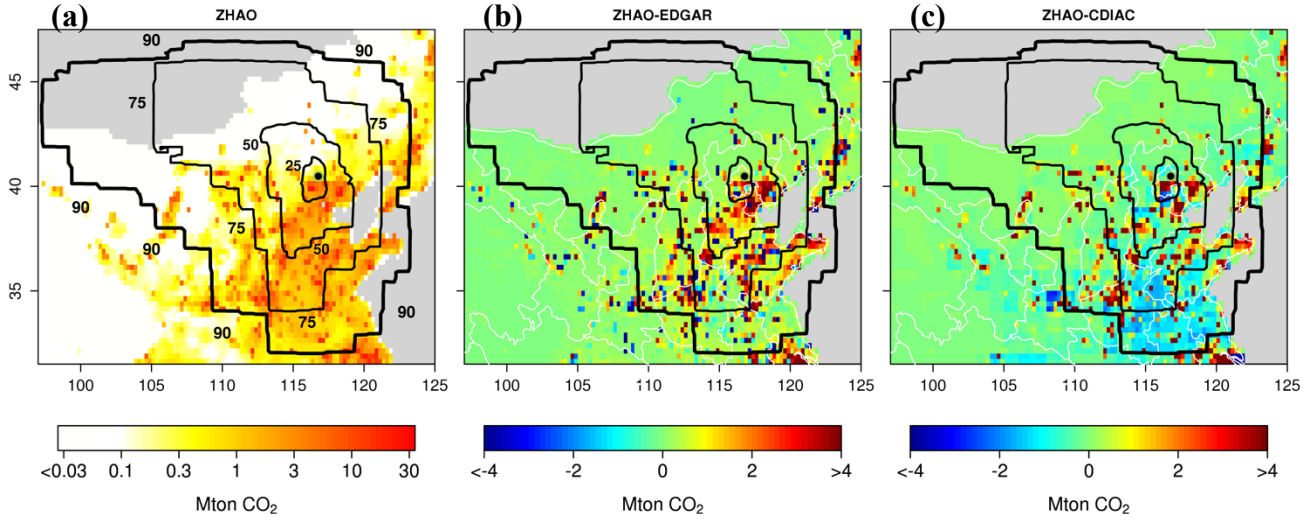

Fig. S8 Mean annual anthropogenic emissions (Mt CO2 yr-1, 2005-2009) zoomed to approximate d02 extent. Black contour lines represent the 25th, 50th, and 75th, and 90th percentiles of multi-year mean annual STILT footprint influences. (3a) displays emissions estimated by ZHAO; black and green circle represents Miyun receptor. (3b) displays EDGAR inventory difference relative to ZHAO; (3c) displays CDIAC inventory difference relative to ZHAO. ZHAO is consistently higher than EDGAR and CDIAC in the Beijing area. Both EDGAR and CDIAC are regridded from their original grids to the ZHAO grid via ESMF Conserve regridding technique.

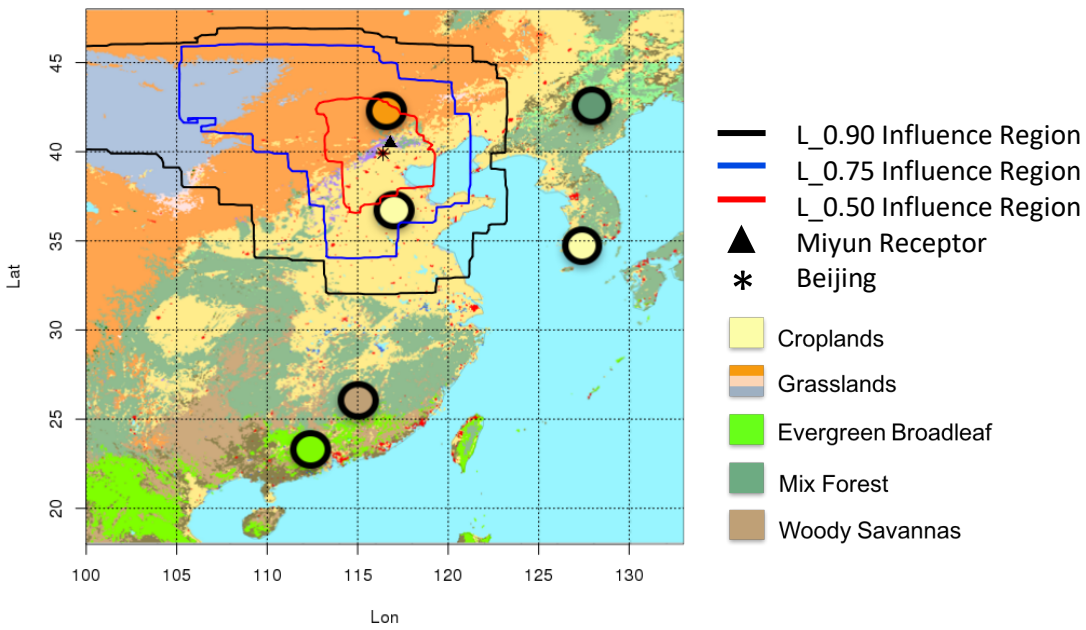

Fig. S9. IGBP land use categories in domain overlaid with STILT influence contours. Note western edge of domain is slightly truncated.

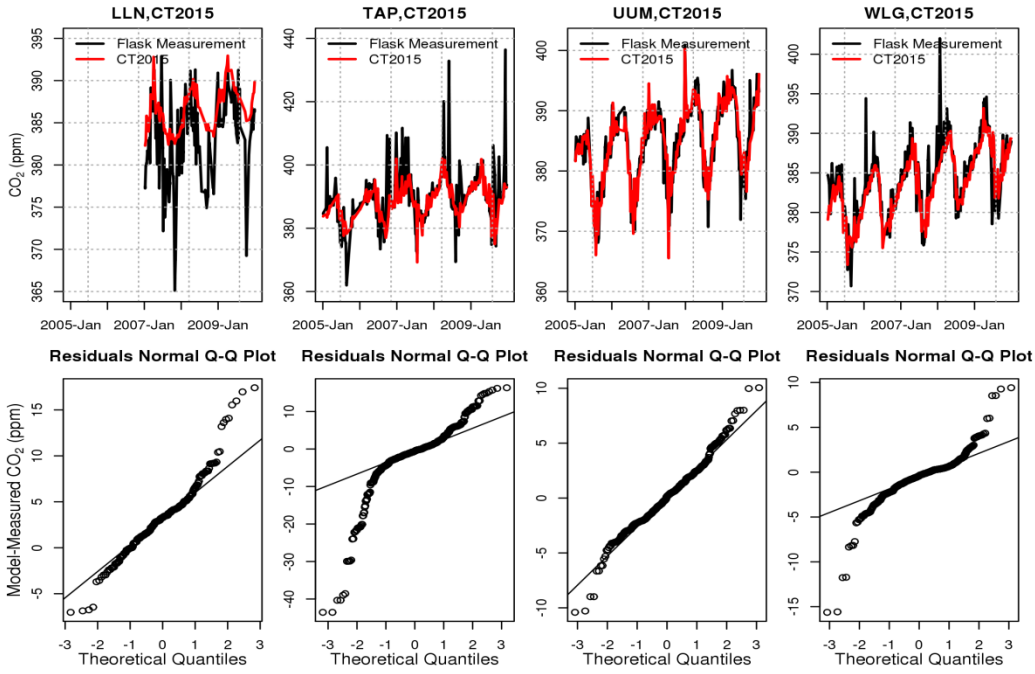

Fig. S10. Evaluation of CT2015 model bias. ~Weekly flask samples from WMO sites (LLN, TAP, UUM, WLG) used to train CT2015 compared with nearest CT2015 pixel. Top row: timeseries. Bottom row: QQ plots of model-measurement residuals.

Table S1. Comparison of unadjusted annual anthropogenic CO_2 emissions ($TgCO_2$) by region. EDGAR and CDIAC are reported as percent differences relative to ZHAO. *: Based on sums AFTER spatial allocation of ZHAO inventories but are <0.1% different from original inventory totals.

		STILT L_0.25	**STILT L_0.50**	**STILT L_0.75**	**STILT L_0.90**	**IM**	**NE**	**N**	**C**	**SE**	**S**	**SW**	All China
2005	ZHAO	135.1	697.0	1796	3015	252.1	682.8	2244	502.4	1486	519.6	759.5	7126
	EDGAR	*-31%*	*-35%*	*-28%*	*-23%*	*+1.9%*	*+1.2%*	*-32%*	*+1.2%*	*-12%*	*-19%*	*-25%*	*-17%*
	CDIAC	*-49%*	*-44%*	*-42%*	*-36%*	*-48%*	*-32%*	*-32%*	*+13%*	*-23%*	*-1.7%*	*+17%*	*-19%*
2006	ZHAO	124.8	734.4	1922	3273	311.7	690.6	2440	558.9	1590	567.6	822.9	7726*
	EDGAR	*-17%*	*-32%*	*-26%*	*-21%*	*-8.2%*	*+13%*	*-31%*	*+2.1%*	*-7.9%*	*-19%*	*-23%*	*-15%*
	CDIAC	*-39%*	*-41%*	*-40%*	*-34%*	*-54%*	*-26%*	*-31%*	*+13%*	*-21%*	*-0.74%*	*+20%*	*-17%*
2007	ZHAO	136.8	805.0	2107	3588	341.6	757.0	2675	612.6	1743	622.1	902.0	8469*
	EDGAR	*-18%*	*-33%*	*-27%*	*-22%*	*-9.8%*	*+12%*	*-32%*	*+0.76%*	*-9.3%*	*-21%*	*-25%*	*-16%*
	CDIAC	*-41%*	*-43%*	*-42%*	*-37%*	*-55%*	*-29%*	*-33%*	*+9.2%*	*-23%*	*-4.2%*	*+15%*	*-20%*
2008	ZHAO	140.5	826.8	2164	3685	350.9	777.5	2747	629.2	1790	639.0	926.4	8699*
	EDGAR	*-12%*	*-27%*	*-21%*	*-16%*	*-3.8%*	*+18%*	*-26%*	*+7.5%*	*-1.9%*	*-14%*	*-20%*	*-9.7%*
	CDIAC	*-39%*	*-41%*	*-40%*	*-35%*	*-54%*	*-27%*	*-31%*	*+12%*	*-21%*	*-1.4%*	*+19%*	*-18%*
2009	ZHAO	125.1	864.7	2301	3974	424.6	777.2	2967	694.8	1903	693.4	997.2	9370
	EDGAR	*+5.4%*	*-26%*	*-20%*	*-17%*	*-16%*	*+25%*	*-27%*	*+3.5%*	*-1.8%*	*-15%*	*-21%*	*-11%*
	CDIAC	*-26%*	*-40%*	*-40%*	*-36%*	*-60%*	*-22%*	*-32%*	*+8.0%*	*-21%*	*-3.5%*	*+17%*	*-19%*

Table S2. Seasonal Flux Corrections and 95% CI (kg CO_2 m^{-2} $month^{-1}$) for L_0.90 region. Original fluxes are in regular font; corrected fluxes and 95% CI are in bold

		JFM/Winter	AMJ/Spring	JAS/Summer	OND/Fall
2005	**ZHAO**	0.133	0.0492	-0.0540	0.132
		0.129 (0.103, 0.105)	**0.0735 (0.0195, 0.135)**	**-0.170 (-0.237,-0.106)**	**0.164 (0.137, 0.193)**
	EDGAR	0.108	0.0256	-0.076	0.110
		0.151 (0.124, 0.174)	**0.116 (0.0597, 0.176)**	**-0.120 (-0.186, -0.0478)**	**0.181 (0.154, 0.204)**
	CDIAC	0.0937	0.0117	-0.0972	0.0951
		0.144 (0.117, 0.170)	**0.132 (0.0734, 0.185)**	**-0.121 (-0.183, -0.0445)**	**0.177 (0.147, 0.206)**
2006	**ZHAO**	0.131	0.0601	-0.0568	0.140
		0.146 (0.122, 0.167)	**0.156 (0.0990,0.217)**	**-0.135 (-0.197,-0.0708)**	**0.174 (0.124, 0.217)**
	EDGAR	0.106	0.0421	-0.0771	0.114
		0.169 (0.145, 0.190)	**0.185 (0.126, 0.246)**	**-0.0951 (-0.157, -0.0310)**	**0.204 (0.152, 0.251)**
	CDIAC	0.0929	0.0260	-0.102	0.0965
		0.165 (0.139, 0.189)	**0.194 (0.134, 0.254)**	**-0.0912 (-0.157, -0.0171)**	**0.223 (0.168, 0.270)**
2007	**ZHAO**	0.139	0.0831	-0.0735	-0.171
		0.154 (0.118, 0.189)	**0.109 (-0.00290, 0.217)**	**-0.151 (-0.205, -0.0958)**	**0.174 (0.129, 0.214)**
	EDGAR	0.109	0.0569	-0.103	0.138
		0.171 (0.133,0.205)	**0.141 (0.0282, 0.264)**	**-0.110 (-0.170, -0.0528)**	**0.192 (0.151, 0.231)**
	CDIAC	0.0917	0.0323	-0.123	0.119
		0.157 (0.119, 0.191)	**0.149 (0.0381, 0.271)**	**-0.113 (-0.173, -0.490)**	**0.184 (0.138, 0.222)**
2008	**ZHAO**	0.120	0.0577	-0.0290	0.143
		0.134 (0.109, 0.160)	**0.0157 (-0.0470,0.0794)**	**-0.170 (-0.247, -0.0940)**	**0.201 (0.159, 0.243)**
	EDGAR	0.0973	0.0459	-0.419	0.118
		0.145 (0.120, 0.171)	**0.0492 (-0.0140,0.111)**	**-0.127 (-0.207,-0.0447)**	**0.219 (0.174, 0.259)**
	CDIAC	0.0785	0.0135	-0.0800	0.0960
		0.139 (0.109, 0.166)	**0.0559 (-0.0114, 0.122)**	**-0.134 (-0.217, -0.0494)**	**0.224 (0.179, 0.264)**
2009	**ZHAO**	0.144	0.0809	0.0277	0.134
		0.231 (0.130, 0.300)	**-0.0655 (-0.127, -0.00290)**	**-0.125 (-0.193, -0.0449)**	**0.215 (0.158, 0.265)**
	EDGAR	0.130	0.0563	-0.00797	0.112
		0.249 (0.156, 0.313)	**-0.0653 (-0.124, 0.00)**	**-0.122 (-0.197, -0.0399)**	**0.217 (0.165, 0.266)**
	CDIAC	0.0970	0.0355	-0.0312	0.0874
		0.238 (0.147, 0.306)	**-0.0404 (-0.105, 0.0239)**	**-0.110 (-0.192, -0.0267)**	**0.215 (0.162, 0.270)**