Table of Contents

Abstract

The Weather Research and Forecasting Model (WRF-ARW) meteorological output, coupled with the Stochastic Time-Inverted Lagrangian Transport model (STILT; a Lagrangian Particle Dispersion Model) has been extensively applied to emissions studies particularly in North American mid-latitudes. In addition, the Vegetation, Photosynthesis, and Respiration Model (VPRM)—a simple biosphere model providing output of hourly vegetation CO_2 fluxes—is being increasingly used in CO_2 optimization studies as part of the WRF-STILT-VPRM modeling framework. This methods paper describes the configuration and processes involved in constructing and evaluating the WRF-STILT-VPRM framework for application to CO_2 emissions optimization studies in China. We present configuration details, optimal settings, and results of evaluation of WRF meteorological fields and VPRM fluxes with observations. We evaluate assumptions associated with selection of CO_2 background concentrations from the NOAA CarbonTracker model (CT2015) and also examine the impact of processing anthropogenic inventories for use in the overall model framework.

S1 Introduction

The Weather Research and Forecasting Model Advanced Research WRF (WRF-ARW) has been coupled with the Stochastic Time-Inverted Lagrangian Transport Model (STILT) is a lagrangian particle dispersion model developed by (1). It has been extensively tested and applied to numerous inventory optimization studies (e.g., 1-5). In addition, the Vegetation, Photosynthesis, and Respiration Model (VPRM) described in (6) has been coupled with WRF-STILT in studies examining CO_2 exchange during the growing season (e.g., 7-8).

In this document, we detail our data processing methods, including configuration and evaluation output from modeling framework (Fig. S1). Sec S2 describes the WRF model and project-specific settings and presents results from evaluation against Chinese Meteorological Administration (CMA) observational data sets. Sec S3 describes the STILT model and settings. Sec S4 describes the construction and processing of the VPRM and its driving data and evaluates hourly model performance in detail. Sec S5 assesses bias in the CT2015 by comparison to observational data. Sec S6 provides supplementary information related to processing the anthropogenic inventories used in this study. Sec S7 presents model output of hourly CO_2 relative to observations over the entire study time period. We conclude with Sec S8 where we present complete seasonal and annual results from inventory optimization.

S2 WRF Model

S2.1 Model Description

The Stochastic Time-Inverted Lagrangian Particle model (STILT) has been optimized for being driven with meteorology from the WRF-ARW modeling system (9). WRF typically conserves mass to a high degree (9).

WRF can be run in ideal or real-data mode. For the purposes of the WRF-STILT framework, the model is configured to be initialized in the real data mode via the optional WRF Pre-processing System (WPS). The second component of the WRF model used in this study is the WRF-ARW solver itself, which consists of the real-data initialization and numerical integration programs. Refer to the official WRFv3.6 User Manual (http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.6_july/UG_July2014.pdf) for further details including general installation and configuration steps.

S2.2 Project-Specific Model Set-up and Run Details

We set up the study region as a two-way nested run with three domains (Table S1). The outermost domain (d01) has the coarsest spatial resolution (27kmx27km). The resolution of domain d02 is 9kmx9km, and domain d03 has the finest resolution (3kmx3km). d01 is the parent domain for d02; d02 is the parent domain for d03.

Within the WRF-STILT framework, we run independent 30-hour WRF simulations (9-10). That is, a month with 31 days involves 31 independent 30-hour simulations. The 30 hours allows for

rejection of the first 6 hours as model equilibration, or spin-up, such that what remains is the requested 24-hour period of valid output. In this study, we start each daily run at 18:00 UTC the previous day and specify a run duration of 30 hours from start. In addition to running the model for short periods of time, error growth during the forecast is further limited by implementing a nudging scheme in the outer domain and never within the PBL. Temperature and Q are not nudged in the PBL as a rule; winds are generally not nudged in the PBL but can be if the temporal resolution of the initialization data is sufficiently fine (10). We customize the WRF Registry.EM_COMMON file prior to recompiling the model as described by http://www.stilt-model.org and (11). The changes made include high resolution output of Temperature and downward shortwave radiation fields for use with the VPRM model. See https://dx.doi.org/10.7910/DVN/OJESO0 and (12) for customization and namelist files.

The combination of short run durations and nudging attempts to minimize WRF output deviation from real meteorological observations/measurements. We run each daily simulation on a single node with 15 cores. See Appendix C in (12) for general steps involved in running WRF for a typical 30h period.

S2.3 Real Data Simulation Mode: WRF Pre-processing System

The purpose of the WRF Pre-processing System (WPS) is to create a *real data-based* initialization file for the WRF-ARW solver in cases where real data simulations are required. This mode is necessary for the WRF-STILT framework.

WPS requires static land data sets (MODIS or USGS) and GRIB1/GRIB2 meteorological data and to run. We use the IGBP-Modified MODIS 20-category Land Use data set for consistency with VPRM configuration (Sec. S4). WRF is initialized with NCEP FNL GRIB1 files: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999 (13). The NCEP FNL fields are on a 1x1 grid, with a 6-hour temporal resolution. All WPS settings are described in its configuration (namelist) file (https://dx.doi.org/10.7910/DVN/OJESO0). Fig. S2 shows the study domain as defined in WPS.

We discard the May 31, 2005 to June 1, 2005 period from analysis due to a discontinuity in the soil levels in the initialization data set. Specifically, the number of expected metgrid soil levels changed from two to four.

S2.4 WPS and WRF Namelist Configuration

https://dx.doi.org/10.7910/DVN/OJESO0 provides the WPS and WRF namelists. The individual variables are thoroughly documented in the WRF-ARWv3.6 User's manual. Deviations from default are based upon optimal WRF-STILT settings as described in (9-11).

S2.5 Post-processing and Evaluation of WRF Output

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/) as shown in the shaded grey area of Fig. S3, is provided for reference but 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. S2 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. S4 through Fig. S6, using observed and simulated fields from 2006. In the evaluation, WRF forecast fields are matched to the nearest observation point.

S3 STILT Model

S3.1 Model Description

STILT (http://www.stilt-model.org), a Lagrangian Particle Dispersion Model (LPDM), is an adjoint used to attribute sensitivies of downwind tracer concentrations made at a receptor to upwind surface influences (sources or sinks) of the tracer. Like all LPDMs, its strength lies in its ability to capture sub-gridscale transport. It does so by interpolating sub-grid scale particle locations using turbulent velocity statistics driven by the meteorology fields. This "stochastic" component of STILT is what enables capture of fine structures and inhomogeneity in transport. Specifically, STILT enables modeling concentrations of a tracer by computing an Influence function, *I*, described in Equation 1 below, reproduced from (1).

$$C(x_r, t_r) = \int_{t_0}^{t} dt \int_V d^3V \, I(\boldsymbol{x_r}, t_r | x, t) \, S(x, t) + \int_V d^3V \, I(\boldsymbol{x_r}, t_r | x, t_0) \, C(x, t_0) \quad (1)$$

where:

C	=	*mixing ratio, in mol/mol, of tracer at specified location and time*
x_r	=	*spatial location of receptor*
t_r	=	*time of observation at receptor located at x_r*
t_0	=	*initial time, less than t_r, provides the boundary condition*
t	=	$t_0 < t < t_r$
d^3V	=	*volume integral representation of* $\int_{x_i}^{x_i + \Delta x} dx \int_{y_j}^{y_j + \Delta y} dy \int_{z_k}^{z_k + \Delta z} dz$
		where horizontal elements are x and y and vertical element is z.
$I(\boldsymbol{x_r}, t_r | x, t)$	=	*Influence Function.* **This is what STILT provides.** *Represents the* **fraction** *of a fluid element (or "particle") that reached location=x_r, time=t_r given that the entire fluid element was at location x at a previous time t. Influence function is expressed in units of inverse volume (vol^{-1}). Note that boldface $\boldsymbol{x_r}$ indicates it is a location vector with x and y elements.*
$S(x,t)$	=	*Surface flux of tracer at location x, time t. Units are (mol/mol) s^{-1}; i.e., mixing ratio per time*

The STILT model computes the influence of upwind surface fluxes on downwind concentrations and ultimately outputs surface footprints. A sample footprint map is shown in Fig. S7. Footprint quantities are expressed as concentration changes at the receptor due to discretized *surface* fluxes and is expressed as $\frac{\mu mol \, mol^{-1}}{\mu mol \, m^{-2} s^{-1}}$ or simply $\frac{ppm}{\mu mol \, m^{-2} s^{-1}}$. Once the particle velocity has been computed, sub-gridscale particle locations are interpolated for a chosen timestep. The sub-gridscale particle locations then produce particle densities by transporting, for statistical robustness, an *ensemble* of particles back in time according to WRF meteorological fields described previously. See (1) for details of the derivation of footprint elements.

STILT examines near-field influences by designating surfaces with which PBL air has come into contact before arrival at the observation location. Given that PBL ventilation time is on the order of 4 days, the near-field surface influence domain affected by PBL processes can extend from regional (10^2 km) to continental (10^3 km) scales, corresponding to spatial scales of the WRF d01 mother domain for this study (1, Table S1). Ultimately, inaccurate representation of transport limits our ability to correctly attribute sources to measured concentrations. STILT assumes well-

mixedness in an atmospheric column of vertical extent of z ≤ h, where h is the mixing height. Footprints have been found to be insensitive to the exact value of h between 10 and 100% of the PBL height such that h is set to PBL height as defined by the WRF output (1). Here we set the mixing height to 50% of the modeled PBL height. The assumption of well-mixedness in the vertical is a consequence of mass conservation and is necessary to prevent the accumulation of particles in low turbulence pockets of the PBL (10).

S3.2 Project-Specific 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). 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 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 unoptimized model performance at all times, averaged seasonally and diurnally, in Sec S7.

S4 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, 0.1º×0.1º; CDIAC, 1º×1º) to the same 0.25º×0.25º grid as the regional inventory (ZHAO) described in detail in (14-16). 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. S8 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

---

[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

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. S9 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. S9a) and +/- 50% of its 2005 contribution (Fig. S9b). 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%.

The original inventories do not embed or provide estimates of intra-annual variability. Previous work (17) has found that temporal variations in CO_2 can be significant, and surface CO_2 can be perturbed from 1.5-8ppm based on time of day and/or day of week (http://cdiac.ornl.gov/ftp/Nassar_Emissions_Scale_Factors). However, in this study we assume anthropogenic CO_2 fluxes are temporally invariant on intra-annual timescales as the effect of applying the weekly and diurnal scaling factors were not statistically significant. No seasonal scaling factors were available.

Total unoptimized emissions for each anthropogenic inventory are calculated on the 0.25ºx0.25º grids and provided in Table S2. 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. S10. Miyun and Beijing are encompassed by the L_0.25 contour. We display the average gridcell emissions of ZHAO (Fig. S10a) and the differences of EDGAR and CDIAC relative to ZHAO (Fig. S10b and Fig. S10c, 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°.

S5 VPRM Configuration and Evaluation

S5.1 Model Description

The VPRM, a simple biosphere CO_2 flux model, uses satellite and surface observations to quantify net ecosystem exchange (*NEE*) of CO_2 at hourly scales. It has been adapted from the early version described in (6). The adapted model is described in detail in (12). The VPRM *NEE* equation is reproduced below and is computed for each ecosystem category in the domain (Fig. S11a).

$$NEE = -\left(\lambda{\times}T_{scale}{\times}P_{scale}{\times}W_{scale}{\times}\frac{1}{1+\left(\frac{PAR}{PAR_0}\right)}{\times}EVI{\times}PAR\right) + (\alpha{\times}T + \beta)$$

The first parenthesized term represents the photosynthetic Gross Primary Productivity (*GPP*); the second parenthesized term represents ecosystem respiration (R_{eco}). T_{scale}, P_{scale}, and W_{scale} represent temperature sensitivity, phenology, and water availability respectively. *PAR* parameterizes photosynthetically active radiation; *EVI* is the Enhanced Vegetation Index; T is air temperature. λ, α, β, and PAR_0 are parameters adjusted for each ecosystem type during calibration with eddy flux observations.

The *NEE* and gross primary productivity (*GPP)* are provided on a 0.25ºx0.25º grid, matching the anthropogenic inventory described in (16).

S5.2 Data Processing

Fig. S11b displays the entire process involved in modeling hourly Net Ecosystem Exchange of CO_2 (*NEE*; $\mu molCO_2m^{-2}s^{-1}$) across multiple years. As described in (13) we use data derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the Terra Satellite (https://modis.gsfc.nasa.gov) to model *GPP*: surface reflectance (8-day average MOD09A1 surface reflectance bands 2, 6, 1 and 3); IGBP land use categories (MCD12Q1); and ecosystem phenology (MCD12Q2). We do not include MODIS surface reflectance data from the Aqua satellite due to failure of a majority of band 6 detectors after launch. All datasets were downloaded using the Reverb tool in NASA's Earth Observing System Data and Information System (http://reverb.echo.nasa.gov/). We then use the MODIS Mosaic and Reprojection Tool (MRTv4.1) to stich and reproject relevant tiles to a WGS84 datum Geographic Coordinate system on a 500m grid.

We quality control our raw surface reflectance data set by only selecting highest quality data under clear sky conditions. We use the NCL Poisson Grid Filling function to interpolate any missing values that result from our quality filtering (Fig. S12). As described in (13), the quality filtering of the ecosystem timing dates (MCD12Q2) was limited in scope to (i) removing anachronistic dates represented by instances where, for a given pixel, ecosystem times were not in chronological order, and (ii) where a given pixel's date was outside of 1-σ of the mean for the ecosystem class represented by that pixel for its latitude band. The second step of MCD12Q2 quality filtering was not conducted for cropland classes in the 32N-38N latitude band due to bimodality of phenology. We display filtered and filled MCD12Q2 ecosystem timing dates for two sample years (2005 and 2006) in Fig. S13 similarities in patterns across years provides confidence in the robustness of the manual filtering method.

We then calculate the VPRM phenology (Pscale), water stress (Wscale), Enhanced Vegetation Index (*EVI*), and Land Surface Water Index (*LSWI*) parameters from the resulting quality-filtered, filled data set (Fig. S14). Temperature dependence (T, Tscale) and radiation (PAR) are based on hourly-averaged 10-minute outputs of WRF surface temperature and shortwave radiation fields, bilinearly interpolated to the same 500m geographic grid as the MODIS-derived parameters (Fig. S15). Note that in all cases, water pixels are masked.

S5.3 Evaluation of Hourly output at calibration sites

We calibrate hourly VPRM *NEE* output for each major ecosystem type in the domain as described in (13) and examine hourly model bias by evaluating calibrated model output with observed *NEE* (Fig. S16). As discussed in (13), the VPRM's strength lies in its ability to capture hourly processes that are dominated in the short term by temperature and radiation. When integrated over longer timescales (e.g., months and years) non-random errors can aggregate significantly.

The northern mixed forest ecosystem represented by CN-Cha is well captured by the VPRM during non-winter seasons. Respiration is consistently underestimated by the model as evidenced by the nighttime output relative to observations (Fig. S16a). For ecosystems in southern latitudes of the study domain (CN-Din/Evergreen Broadleaf; CN-Qia/Woody Savanna) both season and time of day is captured consistently by the VPRM (Fig. S16b, Fig. S16d). Grassland processes, represented by a single degraded grassland in Inner Mongolia, are best captured during the peak summer but overall underestimate uptake (Fig. S16c). Respiration processes in winter wheat and corn dual croplands (CN-Yuc; Fig. S16e) are consistently and significantly underestimated by the VPRM at all seasons; uptake during the spring winter wheat period is consistently underestimated by VPRM as well.

S6 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. 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.

Background values are selected using methodology similar to (5) and 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.

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. S17(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. S17 (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 an overall bootstrapping procedure for the modeling framework as described in the main text, similar to that described in (4).

S7 Analysis of Modeled and Measured CO_2 at the hourly scale

As discussed in Sec S3.2, STILT does not perform well when the receptor release occurs outside of peak daylight hours, when stratification of the PBL is a significant source of error in particle back-trajectory. Therefore, we model the 1100 to 1600 (local time) afternoon hour subset, which is a typical window for which STILT reliably models transport (e.g., 4). We display diurnal modeled and measured averages by season for the unoptimized models with the local afternoon STILT optimization window highlighted in Fig. S18.

We also provide the complete hourly timeseries of WRF-STILT-VPRM modeled ΔCO_2 using the three anthropogenic inventories, and compare them with hourly observations (Fig. S19). We note that the WRF-STILT transport framework largely captures timing of events; the combination of the WRF vertical mixing parameterizations and vegetation and anthropogenic inventories regulate the magnitude of the regional CO_2 enhancement or depletion relative to the background.

We present hourly local afternoon observations of CO_2 over the entire study time period with modelled CT2015 background conditions (Fig. S20a). We have high confidence in the selection criteria and performance of the CT2015 as evidenced by its ability to represent the baseline of the observations such that local enhancements and depletions are visible as relative to the CT2015 baseline. Local events do not impact the CT2015 modelled background. Therefore, the four WMO stations heavily training CT2015 in the region adequately represent upstream background air advected to Miyun, depending on air mass trajectories.

We evaluate the hourly modelled versus measured ΔCO_2 in the bottom panel, and colour by season. The 1:1 line is displayed for reference. All modelled hourly quantities include the same biological component from VPRM such that the only source of variation among models is the anthropogenic inventory. Fig. S20(b-d) display hourly $\Delta CO_{2,ZHAO+VPRM}$, $\Delta CO_{2,EDGAR+VPRM}$, and $\Delta CO_{2,CDIAC+VPRM}$. For all three anthropogenic inventories evaluated against $\Delta CO_{2,obs}$, the correlation coefficient R^2 is greater than 0.4 when aggregated across all seasons, and for winter and fall seasons. During the spring and summer, $\Delta CO_{2,ZHAO+VPRM}$ has the highest R^2 (0.2 and 0.23, respectively); $\Delta CO_{2,EDGAR+VPRM}$ and $\Delta CO_{2,CDIAC+VPRM}$ have R^2 values of 0.2 or lower. Despite the high R^2 values across seasons and during dormant seasons, $\Delta CO_{2,ZHAO+VPRM}$ is the only model to equally span the 1:1 line; $\Delta CO_{2,EDGAR+VPRM}$ and $\Delta CO_{2,CDIAC+VPRM}$ consistently fall below, suggesting systematic underestimation of measured ΔCO_2 (Fig. S20e-g). The EDGAR and CDIAC underestimation of ΔCO_2 at the hourly scale aggregates more apparently at longer timescales of months, seasons, and years as discussed in the main text.

S8 Optimization Results and Methodology

Complete results from optimization in the L_0.90 influence region are provided in Table S3 and Table S4. Table S3 provides seasonal fluxes for each year before and after optimization. Table S4 provides annual scaling factors, and emissions totals after optimization.

We translate the resulting mole fraction (ppm) mismatch between observed and modeled ΔCO_2 to inventory corrections at annual and seasonal timescales. We optimize in the L_0.90 contour (Fig. S10) which represents regions that substantially influence the receptor without disproportionally weighting pixels that contribute very little to the observed signal.

At annual scales, the dominant contributor to the CO_2 signal are anthropogenic emissions; optimization at annual scales is therefore applied only to the anthropogenic emissions inventories. The heavily cropped L_0.90 influence region implies rapid turnaround of vegetation carbon stocks at the annual scale, justifying this assumption (18). 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.

At the seasonal scale, 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 optimize 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 optimization 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 optimized 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.

Acknowledgments

References

1. Lin JC, Gerbig C, Wofsy SC, Andrews AE, Daube BC, Davis KJ, Grainger CA (2003) A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *J Geophys Res Atmos*, 108, 4493, doi:10.1029/2002JD003161.
2. Kort EA, Eluszkiewicz J, Stephens BB, Miller JB, Gerbig C, Nehrkorn T, Daube BC, Kaplan JO, Houweling S, Wofsy SC (2008) Emissions of CH4 and N2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, *Geophys Res Lett*, Vol. 35, L18808, doi:10.1029/2008GL034031.
3. Miller SM, Wofsy SC, Michalak AM, Kort EA, Andrews AE, Biraud SC, Dlugokencky EJ, Eluszkiewicz J, Fischer ML, Janssens-Maenhout G., Miller BR, Miller JB, Montzka SA, Nehrkorn T, Sweeney C (2013) Anthropogenic emissions of methane in the United States, *Proc Natl Acad Sci*,110 20018-20022.
4. McKain K, Down A, Raciti SM, Budney J, Hutyra LR, Floerchinger C, Herndon SC, Nehrkorn T, Zahniser MS, and Jackson RB (2015) Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts *Proc Natl Acad Sci* U.S.A., 112 (7) 1941– 1946.
5. Karion A, Sweeney C, Miller JB, Andrews AE, Commane R, Dinardo S, Henderson JM, Lindaas J, Lin JC, Luus KA, Newberger T, Tans P, Wofsy SC, Wolter S, Miller CE (2016) Investigating Alaskan methane and carbon dioxide fluxes using measurements from the CARVE tower, *Atmos Chem Phys*, 16, 5383-5398, doi:10.5194/acp-16-5383-2016.
6. Mahadevan P, Wofsy SC, Matross DM, Xiao X, Dunn AL, Lin JC, Gerbig C, Munger JW, Chow VY, Gottlieb EW (2008) A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM)**,** *Gl Biogeochem Cycles*, 22, GB2005, doi:10.1029/2006GB002735.
7. Luus KA, Commane R, Parazoo NC, Benmergui J, Euskirchen ES, Frankenberg C, Joiner J, Lindaas J, Miller CE, Oechel WC, Zona D, Wofsy SC, Lin JC (2017) Tundra photosynthesis captured by satellite-observed solar-induced chlorophyll fluorescence, *Geophys Res Lett*, 44, 1564–1573, doi:10.1002/2016GL070842.
8. Matross DM, Andrews A, Mahadevan P, Gerbig C, Lin J, Wofsy S, Daube B, Gottleib E, Chow V, Lee J, Zhao C, Bakwin P, Munger J, Hollinger D (2006) Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data. Tellus B, 58: 344–358. doi: 10.1111/j.1600-0889.2006.00206.x.
9. Nehrkorn T, Eluszkiewicz J, Wofsy SC, Lin J, Gerbig C, Longo M, Freitas S (2010) Coupled weather research and forecasting–stochastic time-inverted lagrangian transport (WRF–STILT) model, *Meteorol Atmos Phys*, 107: 51. doi:10.1007/s00703-010-0068-x.
10. Hegarty J, Draxler R, Stein A, Brioude J, Mountain M, Eluszkiewicz J, Nehrkorn T, Ngan F, Andrews A (2013) Evaluation of Lagrangian Particle Dispersion Models with Measurements from Controlled Tracer Releases, *J Appl Meteorol Climatol*, 52, 2623-2637, doi: 10.1175/JAMC-D-13-0125.1.
11. Nehrkorn T (2015) Personal Communication.
12. Dayalu A (2017) Exploring the Wide Net of Human Energy Systems: From Carbon Dioxide Emissions in China to Hydraulic Fracturing Chemicals Usage in the United States, Doctoral dissertation, Harvard University, Graduate School of Arts and Sciences.

13. NCEP National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce (2000) Data from: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999, http://dx.doi.org/10.5065/D6M043C6, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo. (Updated daily.). Accessed† 05 Feb 2014.

14. Andres RJ , Boden TA, Marland G (2016) Data from: "Annual Fossil-Fuel CO_2 Emissions: Mass of Emissions Gridded by One Degree Latitude by One Degree Longitude." Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/ffe.ndp058.2016.

15. European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL) (2013) Data from: "Emission Database for Global Atmospheric Research (EDGAR), release EDGARv4.2 FT2010", http://edgar.jrc.ec.europa.eu.

16. Zhao Y, Nielsen CP, and McElroy M (2012) China's CO_2 emissions estimated from the bottom up: Recent trends, spatial distributions, and quantification of uncertainties, *Atmos Envir*, 59, 214-223.

17. Nassar R, Napier-Linton L, Gurney KR, Andres RJ, Oda T, Vogel FR, Deng F (2013) Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets, J. Geophys. Res. Atmos., 118, 917–933, doi:10.1029/2012JD018196.

18. Shilong P*, et al.* (2009) The carbon balance of terrestrial ecosystems in China. *Nature* 458(7241):1009
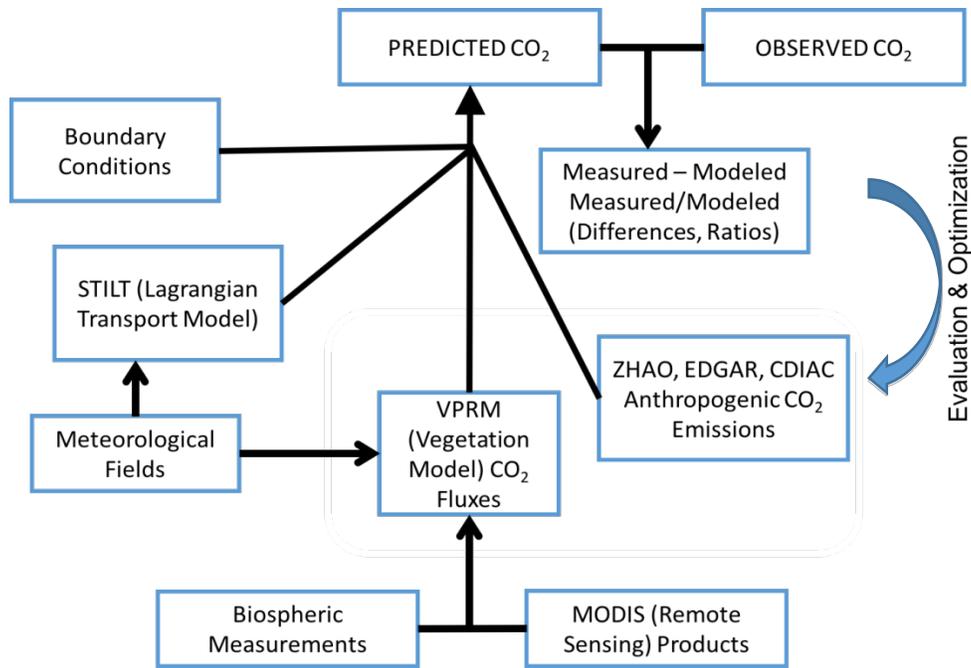
Fig. S1. Overview of CO$_2$ inventory evaluation and optimization
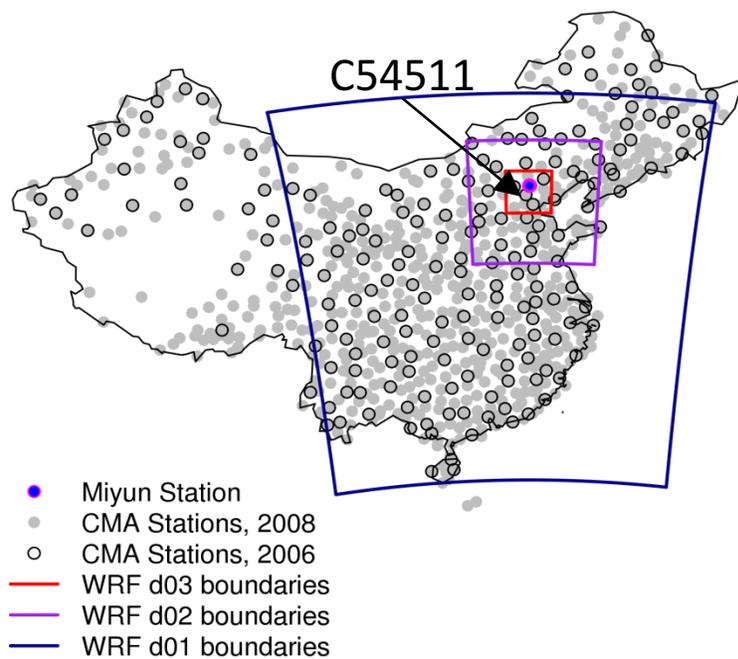

Fig. S2. CMA Station Map (2006, 2008) with WRF domain boundaries. Sample WRF evaluation results are provided for Station 54511 (indicated by arrow on map), near Miyun receptor.
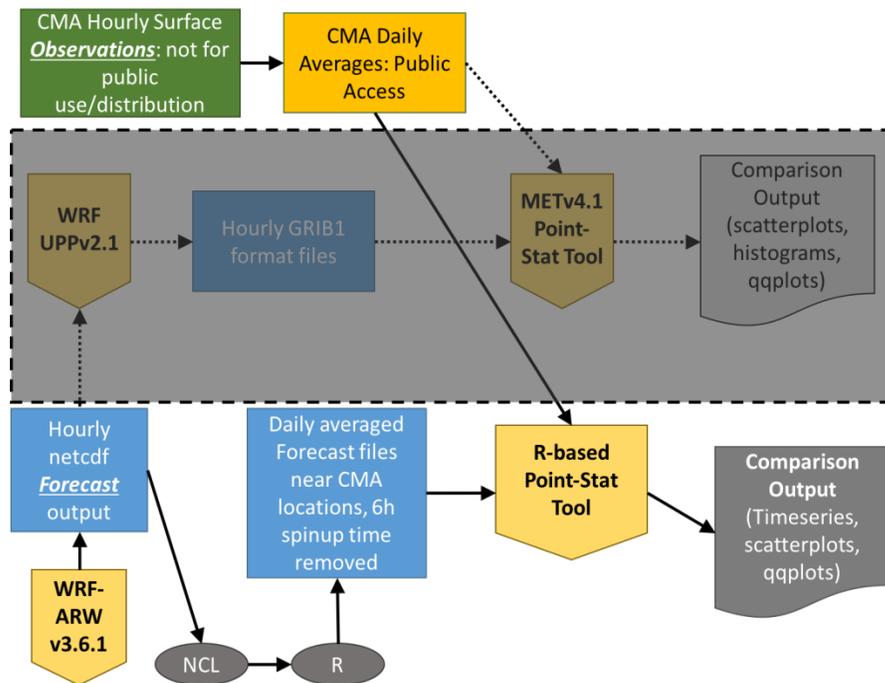
Fig. S3. Evaluation procedure for WRF meteorological output. Grey shaded region refers to standard WRF evaluation toolbox, applicable for future studies where high resolution observational data becomes available.
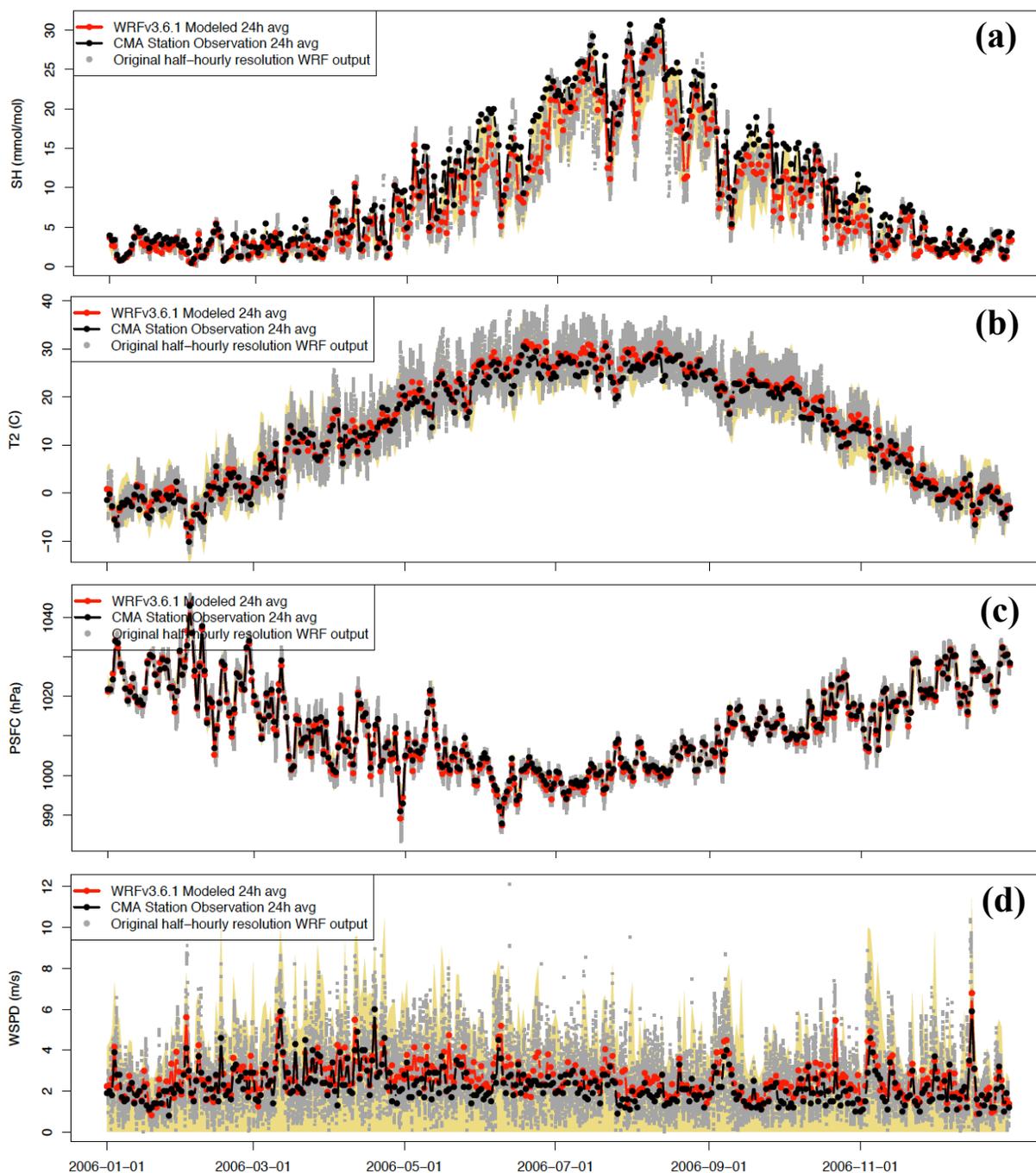
Fig. S4. Evaluation of WRF output against observational data. 2006 Meteorology timeseries for sample WRF gridcell (39.825N, 116.51E) evaluated against nearest CMA Station C54511 (39.800N 116.47E). WRF Meteorology averaged from half-hourly to daily for (a) Specific Humidity; (b) Surface Temperature; (c) 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.
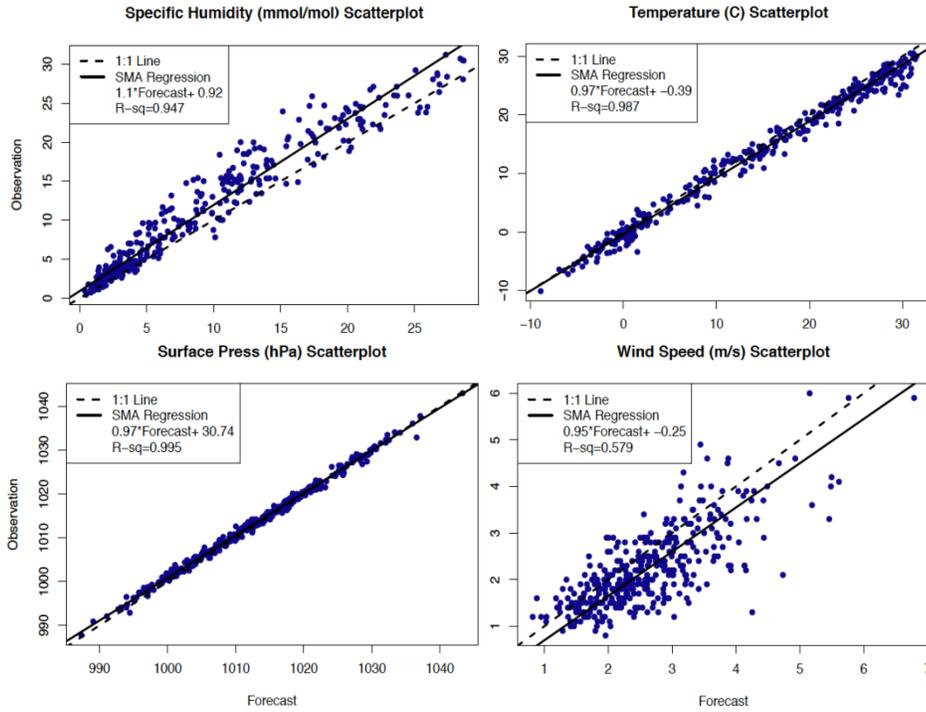
Fig. S5. Observed vs WRF Modeled (Forecast) meteorology for sample WRF gridcell. Gridcell (39.825N, 116.51E) evaluated against nearest CMA Station C54511 (39.800N 116.47E). Time-base of fields is daily average.
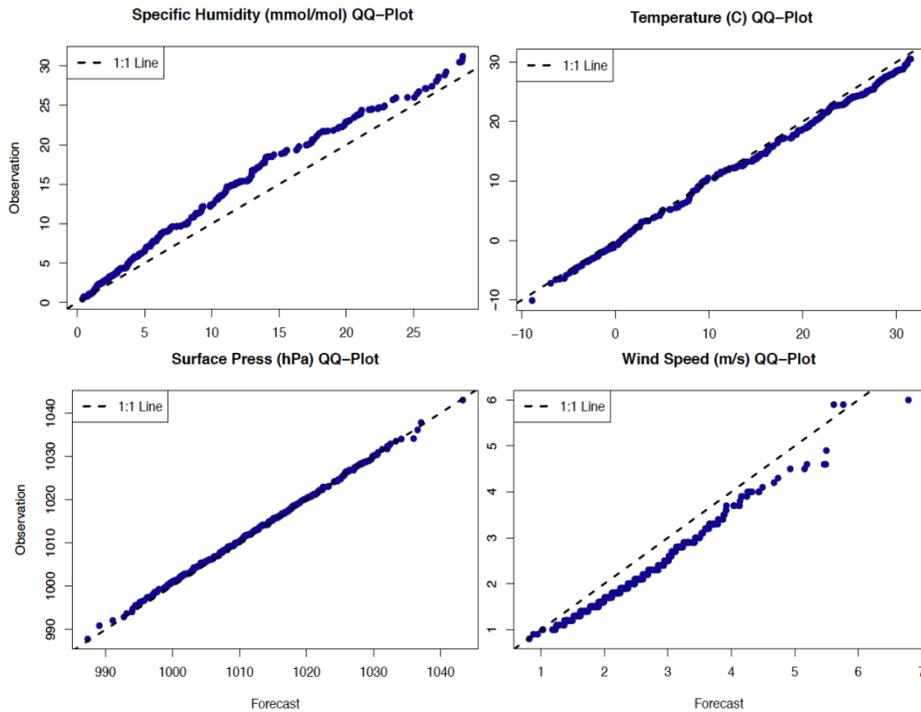

Fig. S6. Q-Q plots of Observed and WRF Modeled (Forecast) meteorology for sample WRF gridcell. Gridcell (39.825N, 116.51E) evaluated against nearest CMA Station C54511 (39.800N 116.47E). Time-base of fields is daily average.
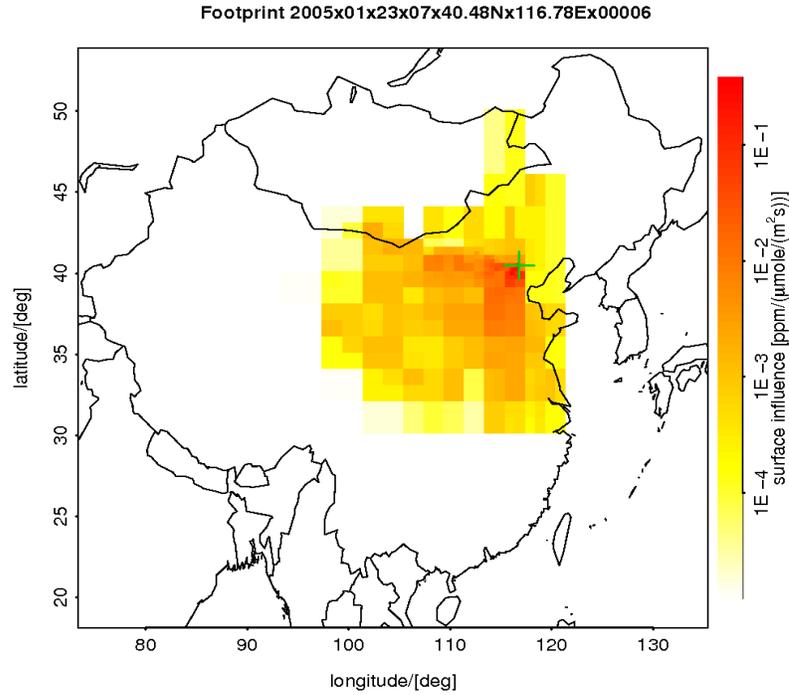
Fig. S7. Sample STILT footprint map. Measurement hour on January 23, 2005 at 0700UTC (1500 Local). Surface influences are provided in ppm $\mu mol^{-1}m^{-2}s^{-1}$. Receptor release point is indicated by the green cross.
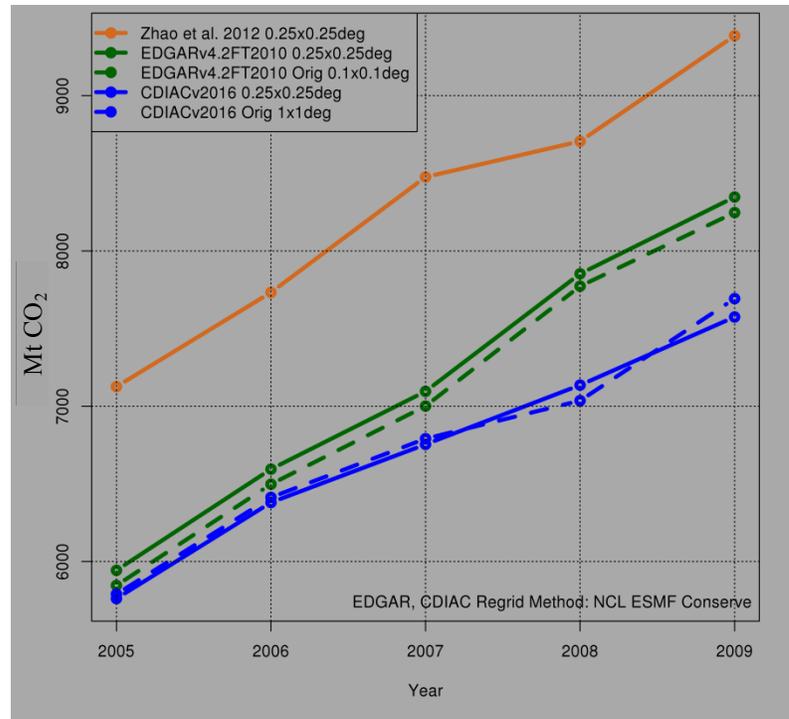

Fig. S8. ZHAO, EDGAR, and CDIAC estimates of total annual CO_2 emissions for Mainland China, 2005 to 2009. EDGAR and CDIAC are regridded to 0.25°x0.25° grid using the NCL Earth System Modeling Framework Conserve regridding function.
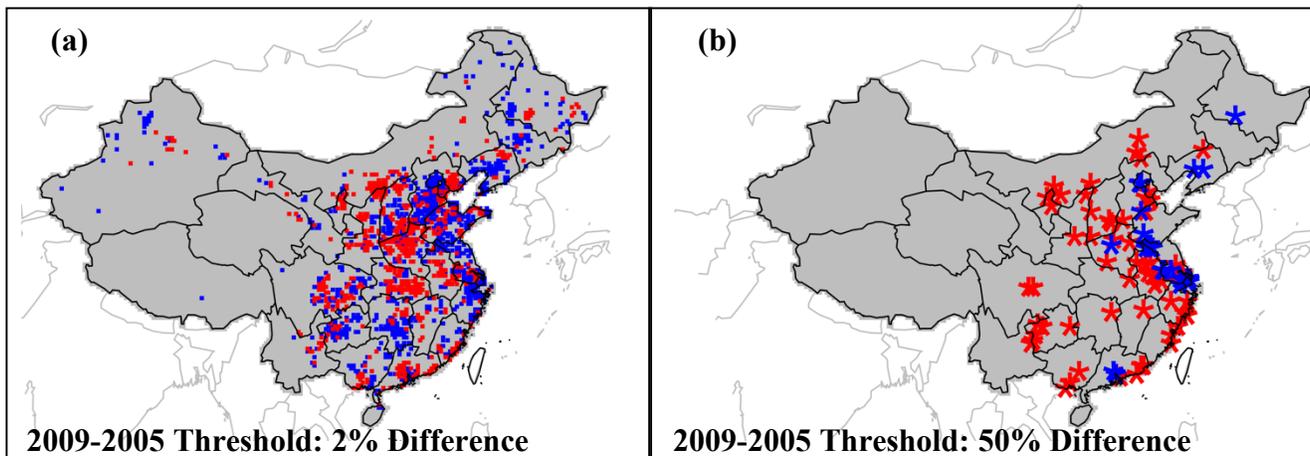
22

Fig. S9 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.
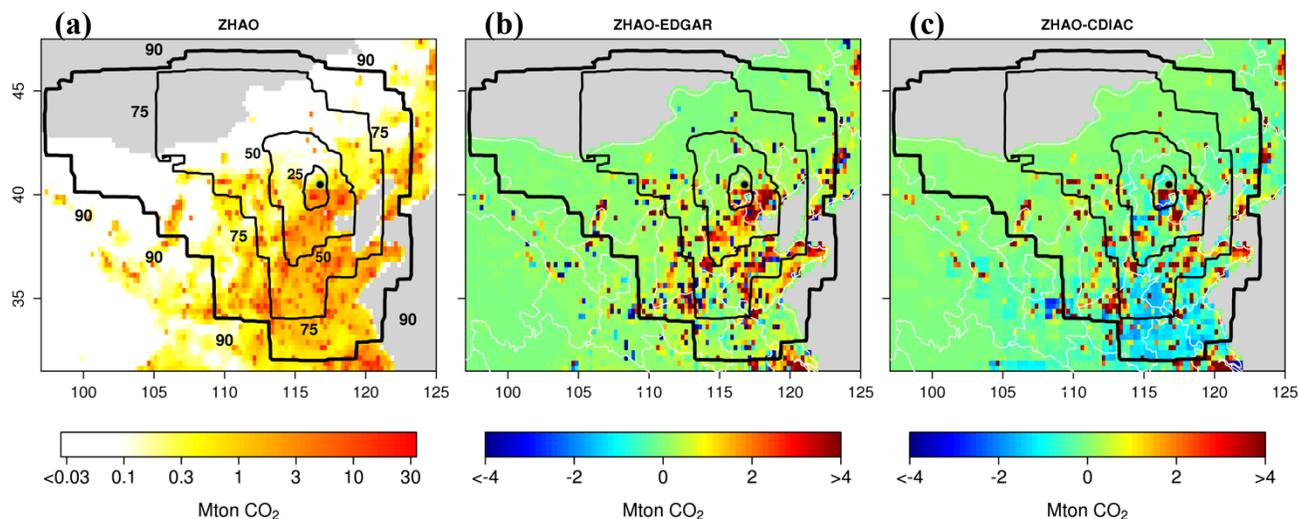

Fig. S10 Mean annual Anthropogenic emissions (Mt CO$_2$ yr^{-1}, 2005-2009) zoomed to approximate d02 extent. Black contour lines represent the 25th, 50th, and 75th, and 90th percentiles of multi-year mean annual STILT footprint influences. (3a) displays emissions estimated by ZHAO; black and green circle 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.
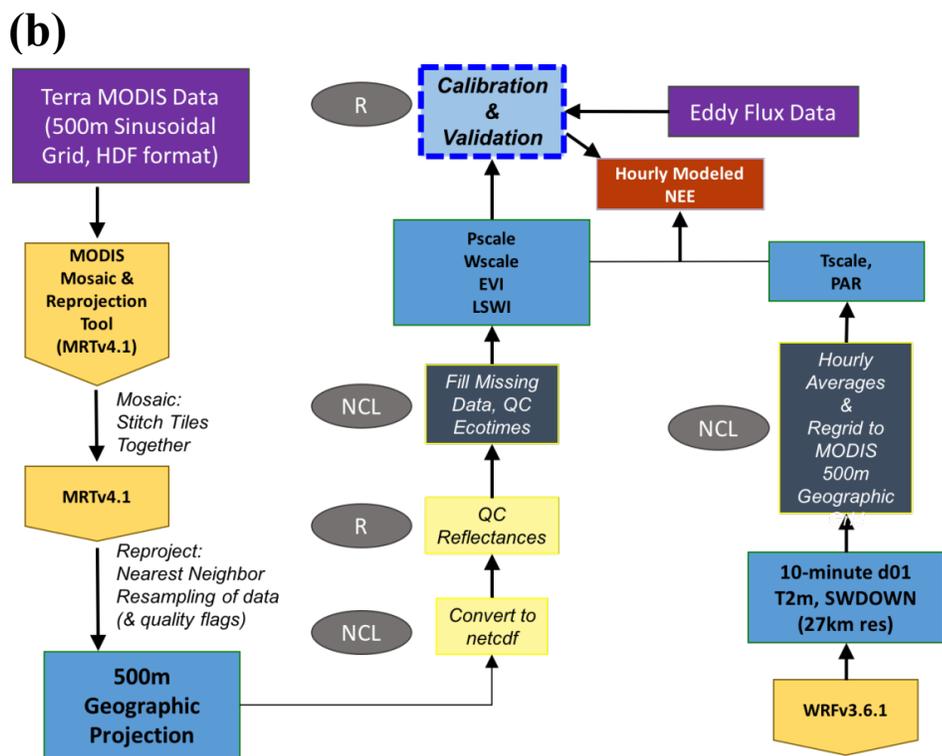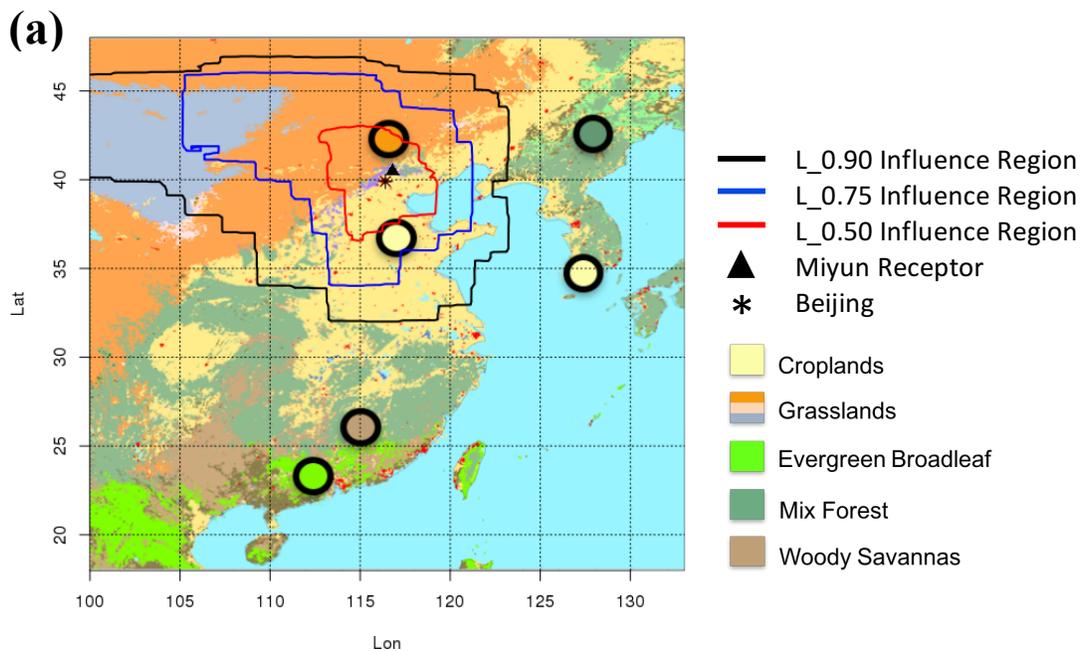
Fig. S11. VPRM Processing (a) for each pixel associated with an IGBP ecosystem category in domain and (b) schematic of VPRM data processing, quality control, and model calibration leading to hourly NEE output at each pixel. Note western edge of domain is slightly truncated in (a).
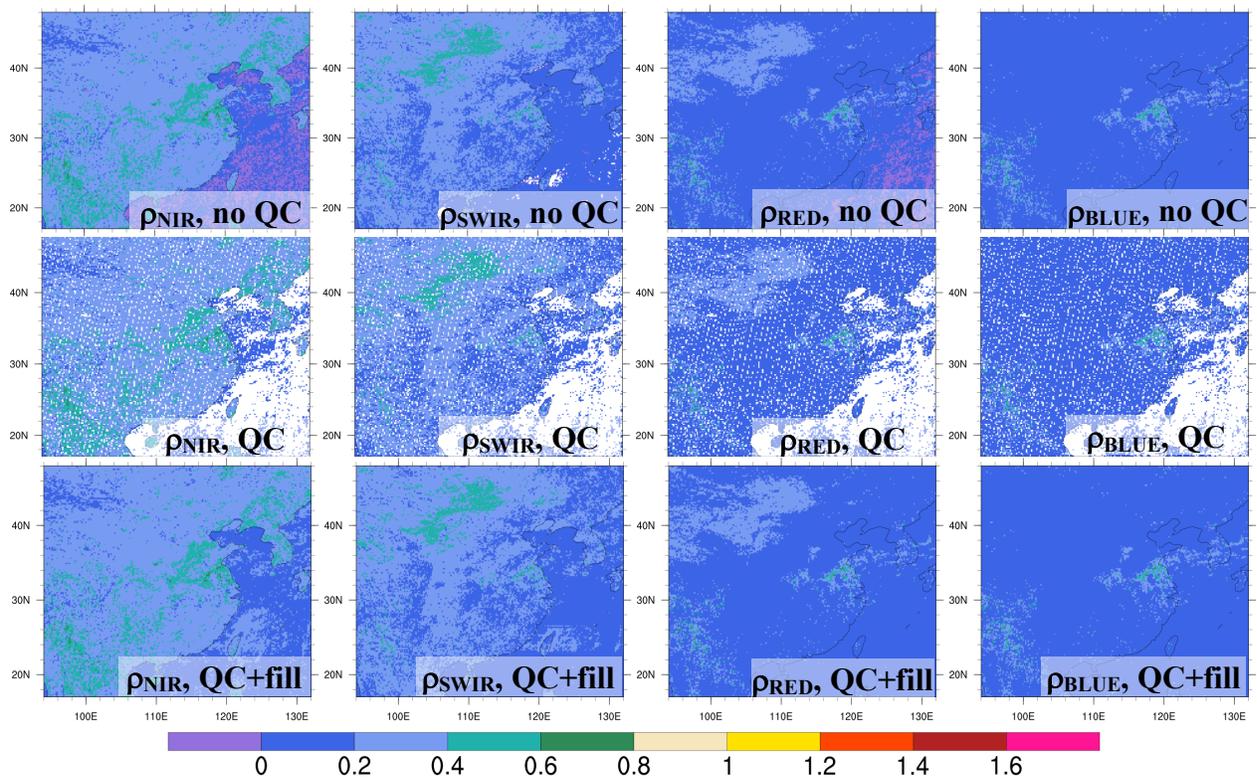
Fig. S12. MOD09A1 Surface reflectances (ρ) in the near infra-red (NIR); shortwave infrared (SWIR); red (RED); and blue (BLUE) for a sample day during peak growing season in July 2006 (day 201). We display original, unfiltered data (top row); QC filtered subset (middle row); and poisson-filled filtered subset (bottom row).
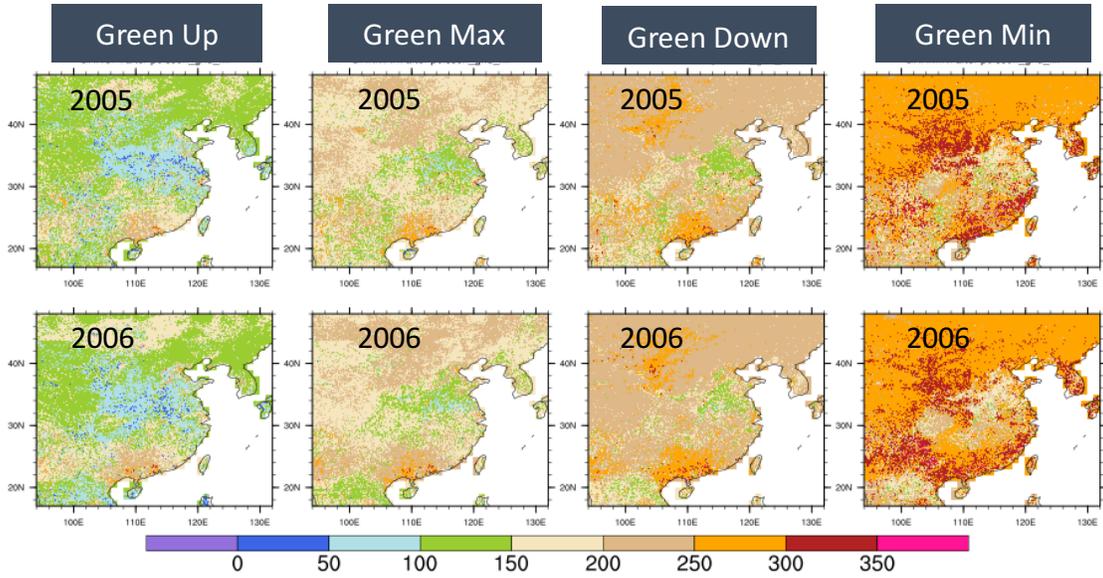
Fig. S13. Quality filtered MCD12Q2 Ecosystem Timing Dates. Reported as day of year for 2005 (top row) and 2006 (bottom row)
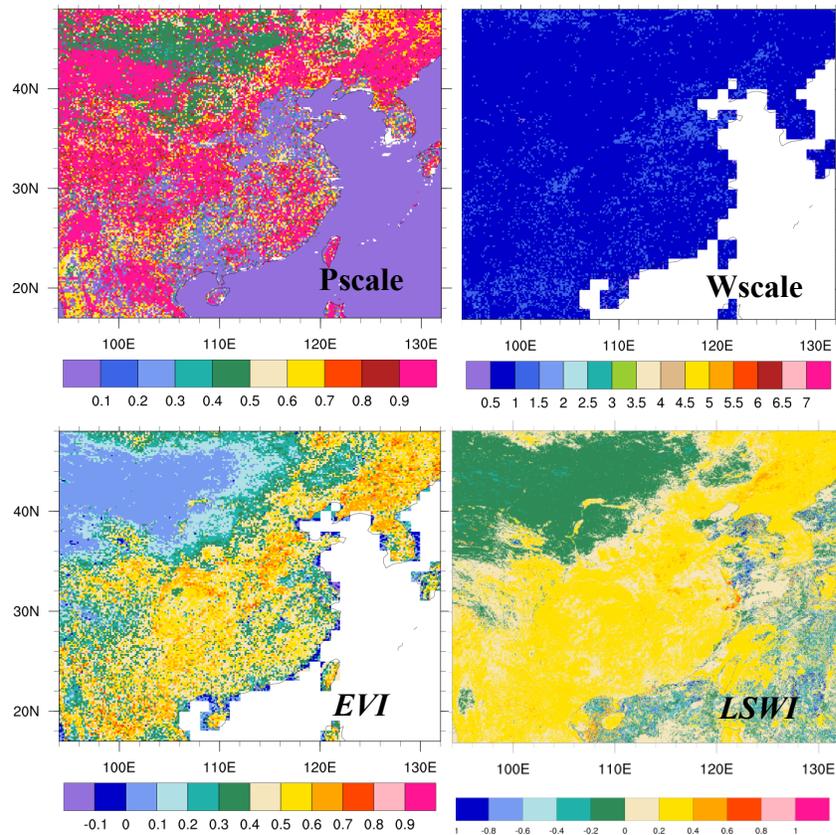
Fig. S14. Pscale, Wscale, EVI, LSWI for sample peak growing season day. Day is in July 2006 (day 201).Timing is based on ecosystem event onset.
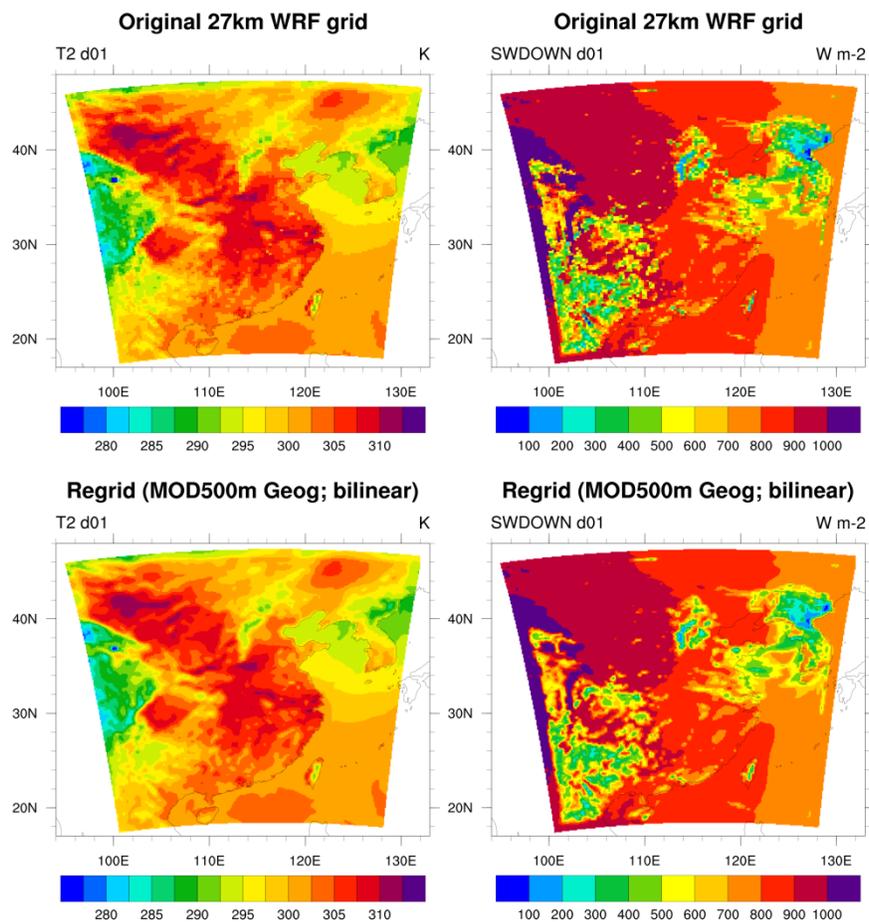
Fig. S15. WRF Surface Temperature (T2) and Shortwave radiation (SWDOWN) in d01 for a sample day in July 2006. Top row: fields on original 27km WRF grid; bottom row: fields regridded using bilinear interpolation (NCL; Earth System Modeling Framework) to 500m geographic grid.
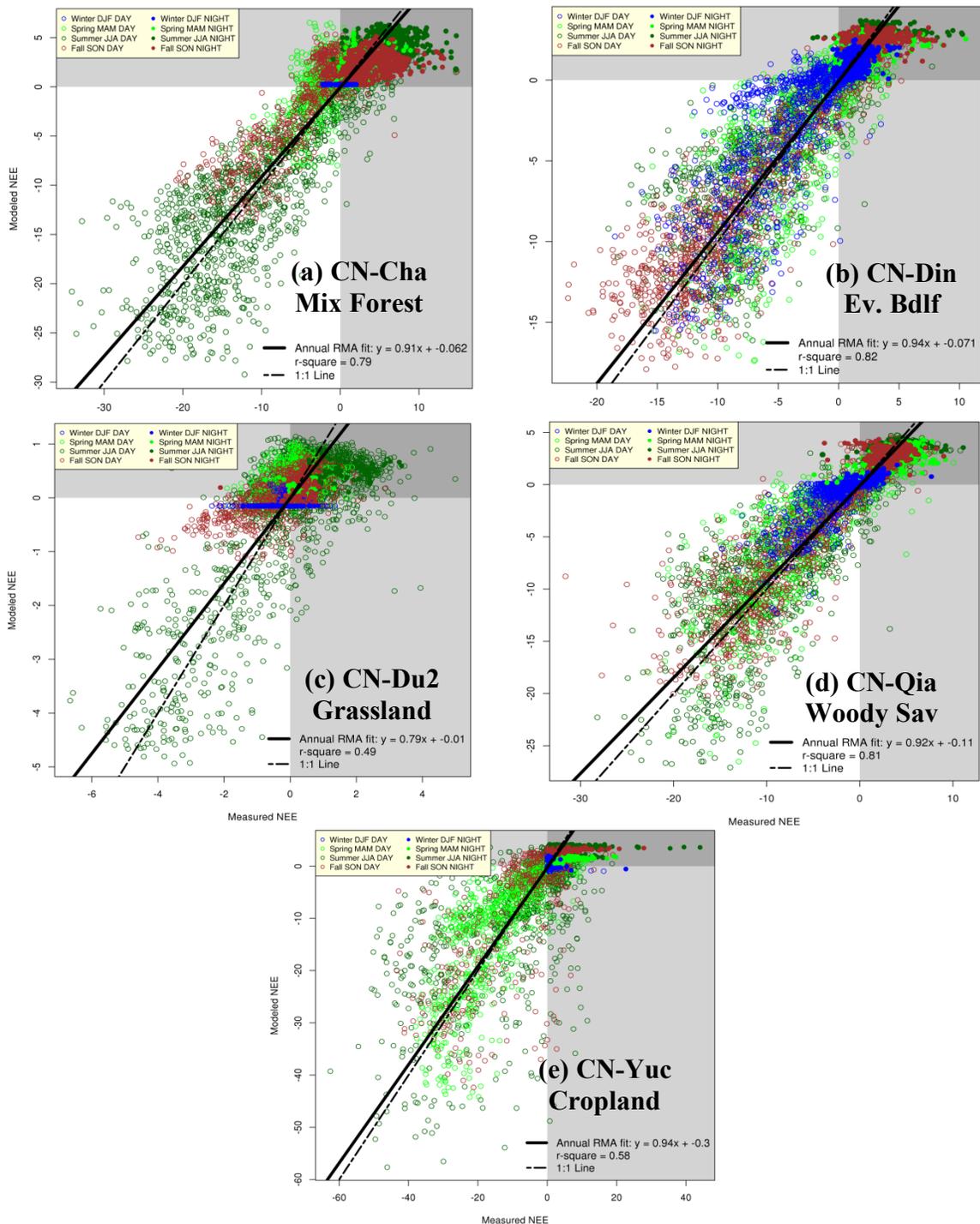
Fig. S16 Hourly Modeled vs. Measured NEE at VPRM calibration sites. Point color and fill represent season and time of day (open circles = day time). Light grey shading represents regions where either Modeled or Measured *NEE* are >0 (net release); dark grey represents regions where both Modeled and Measured *NEE* are >0.
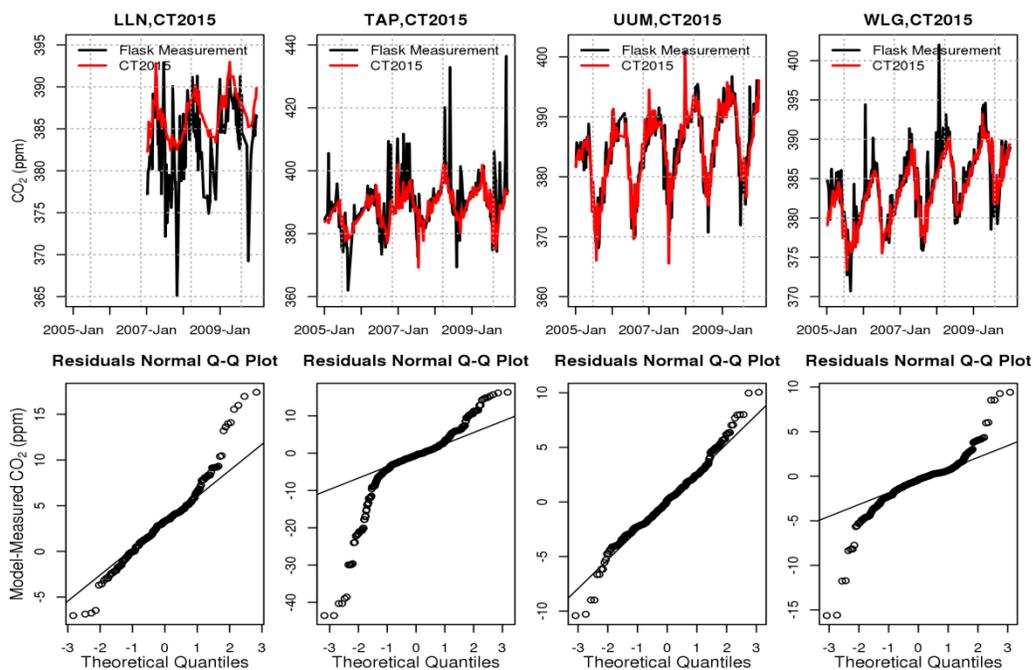
Fig. S17. Evaluation of CT2015 model bias. ~Weekly flask samples from WMO sites (LLN, TAP, UUM, WLG) used to train CT2015 compared with nearest CT2015 pixel. Top row: time series of measurements; bottom row: Normal Q-Q plot of model-measured residuals.
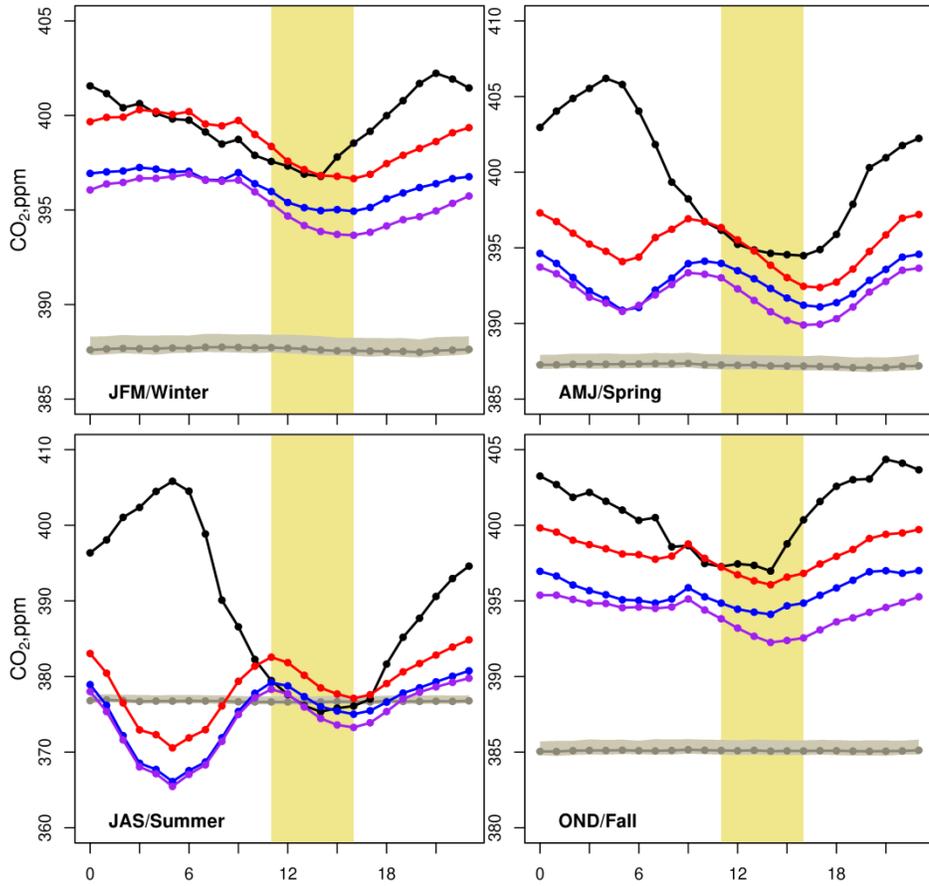
Fig. S18. Diurnal UNOPTIMIZED modeled and measured CO$_2$, averaged by season from 2005 to 2009. Local time is on x-axis. Optimization window (1100 to 1600) is highlighted in yellow. Comparison of CO$_2$ observations (black) with CO$_2$ modeled by ZHAO+VPRM (red), EDGAR+VPRM (blue), and CDIAC+VPRM (purple). CT2015 background is in grey.
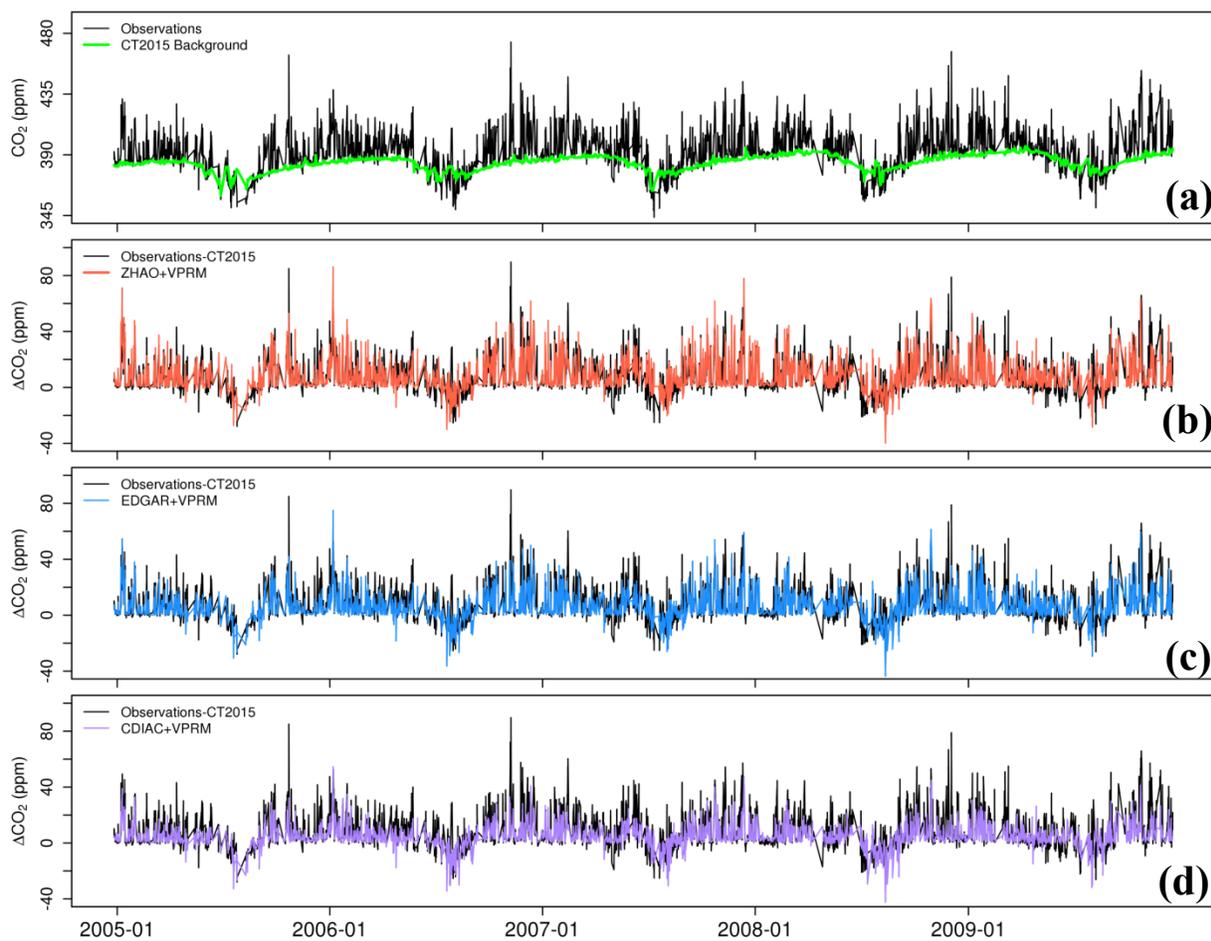
Fig. S19. Hourly Timeseries of CO_2 and ΔCO_2 (1100 to 1600, local) over study time period.
(a) Observed CO_2 (ppm) with modeled background highlighted; observed ΔCO_2 plotted against
(b) $\Delta CO_{2,\text{ZHAO+VPRM}}$; (c) $\Delta CO_{2,\text{EDGAR+VPRM}}$; (d) $\Delta CO_{2,\text{ CDIAC+VPRM}}$.
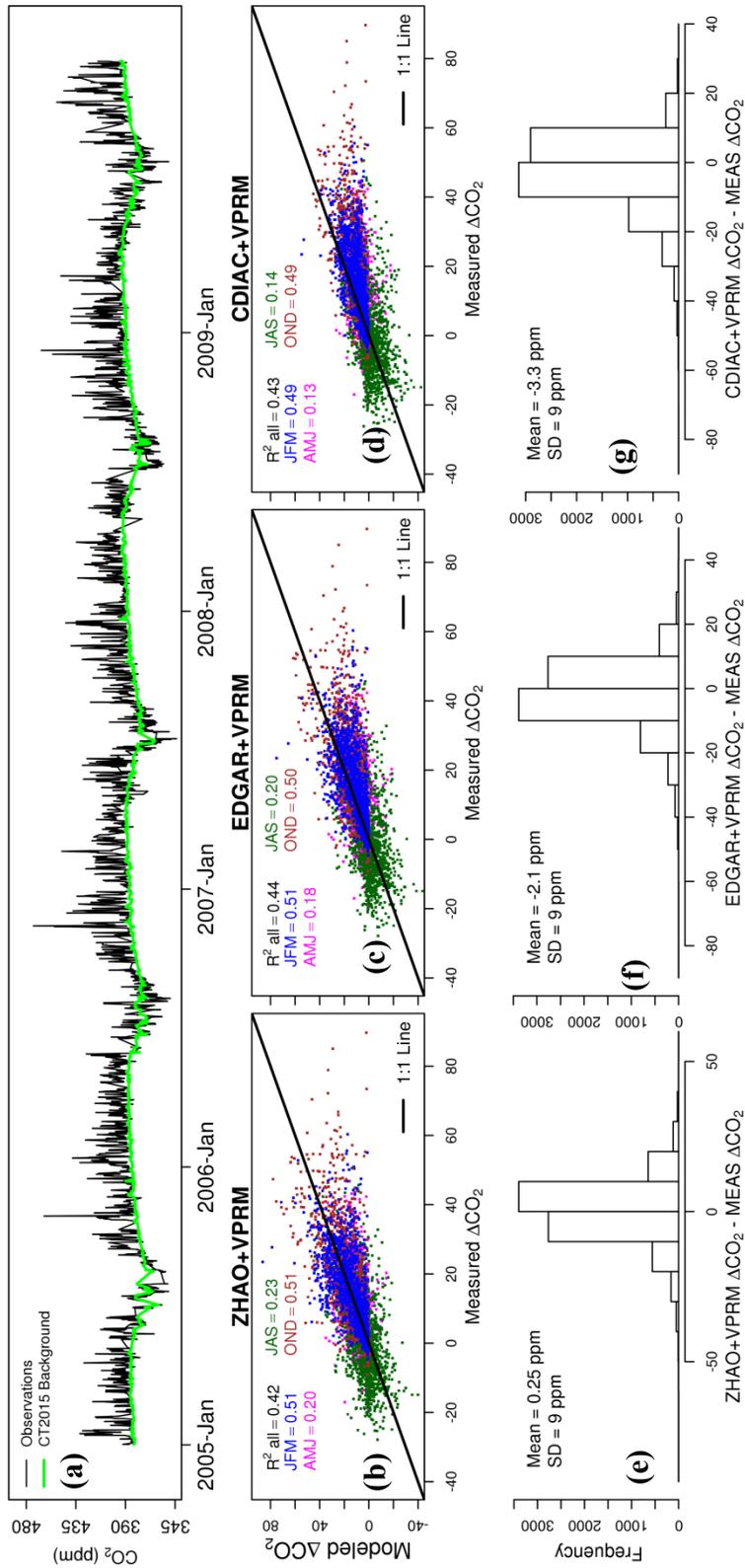
Fig. S20 Hourly (1100 to 1600 Local Time) Modeled and Measured CO₂ and ΔCO₂. Measured CO₂ and modelled CT2015 background concentrations are displayed in (a). Modelled versus measured ΔCO₂ for each anthropogenic inventory is shown in b-d, colored by season. Histograms of modelled-measured residuals are shown in e-g. The VPRM vegetation component is included in all modelled ΔCO₂ values.

Table S1. Resolution and Extent of study WRF domains.

Domain (resolution)	Latitude Range	Longitude Range
d01 (27km x 27km)	17.37N, 47.40N	94.9E, 132.3E
d02 (9km x 9km)	34.58N, 43.90N	111.5E, 122.8E
d03 (3km x 3km)	38.42N, 41.59N	114.8E, 118.6E

Table S2. Comparison of unoptimized 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 totals in (16).

|		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 S3. Seasonal Flux Optimization Results and 95% CI (kg CO$_2$ m^{-2} month^{-1}) for L_0.90 region. Optimizations based on additive corrections. Unoptimized fluxes are in regular font; optimized 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)**

Table S4. Annual scaling factors (95% CI) and optimized emissions for L_0.90 region.

		Scaling Factor (95% CI)	Optimized $MtCO_2$ (95% CI)
2005	**ZHAO**	0.95 (0.84, 1.0)	2800 (2476, 3105)
	EDGAR	1.4 (1.3, 1.6)	3306 (2886, 3683)
	CDIAC	1.7 (1.5, 1.9)	3489 (3017, 3871)
2006	**ZHAO**	1.0 (0.91, 1.1)	3326 (2972, 3631)
	EDGAR	1.5 (1.3, 1.6)	3751 (3325, 4150)
	CDIAC	1.9 (1.6, 2.0)	3930 (3438, 4338)
2007	**ZHAO**	0.94 (0.85, 1.0)	3080 (2789, 3324)
	EDGAR	1.4 (1.2, 1.5)	3454 (3096, 3785)
	CDIAC	1.6 (1.5, 1.8)	3180 (2842, 3493)
2008	**ZHAO**	0.94 (0.82, 1.0)	3422 (3008, 3768)
	EDGAR	1.2 (1.1, 1.4)	3790 (3332, 4207)
	CDIAC	1.7 (1.5, 1.9)	3941 (3461, 4374)
2009	**ZHAO**	0.96 (0.86, 1.1)	3860 (3474, 4251)
	EDGAR	1.1 (1.0, 1.3)	3518 (3133, 3874)
	CDIAC	1.5 (1.3, 1.7)	3921 (3454, 4330)