



Carbon dioxide emissions in Northern China based on atmospheric observations from 2005 to 2009

Archana Dayalu^{1,2}, J. William Munger^{1,2}, Yuxuan Wang^{3,4}, Steven C. Wofsy^{1,2}, Yu Zhao⁵, Thomas
5 Nehrkorn⁶, Chris Nielsen², Michael B. McElroy², Rachel Chang⁷

¹Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA

²School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

³Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA

⁴Department of Earth System Sciences, Tsinghua University, Beijing, China

10 ⁵School of the Environment, Nanjing University, Nanjing, China

⁶Atmospheric and Environmental Research, Lexington, MA, USA

⁷Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Canada

Correspondence to: Archana Dayalu (adayalu@seas.harvard.edu)

15 **Abstract.** China has pledged reduction of carbon dioxide emissions per unit GDP by 60-65% relative to
2005 levels, and to peak carbon emissions overall by 2030. However, disagreement among available
inventories makes it difficult for China to track progress toward these goals and evaluate the efficacy of
regional control measures. In this study, we evaluate three anthropogenic CO₂ inventories by tracking
the fidelity of predicted concentrations of CO₂ in the atmosphere to observations, focusing on the key
20 commitment period for the Paris accords (2005) and the Beijing Olympics (2008). One inventory is
China-specific and two are spatial subsets of global inventories. The inventories differ in spatial
resolution, basis in national or subnational statistics, and reliance on global or China-specific emission
factors. We use a unique set of historical atmospheric observations from 2005-2009 to evaluate the
three CO₂ emissions inventories within China's heavily industrialized and populated Northern region
25 accounting for ~33-41% of national emissions. Each anthropogenic inventory is combined with
estimates of biogenic CO₂ within a high-resolution atmospheric transport framework to model the time
series of CO₂ observations. Model-observation mismatch in concentration units is translated to mass
units and used to optimize the original inventories in the measurement influence region, largely
corresponding to Northern China. Except for the peak growing season, where assessment of



anthropogenic emissions is entangled with the strong vegetation signal, we find the China-specific inventory based on subnational data and domestic field-studies agrees significantly better with observations than the global inventories at all timescales. On average, over the study time period, the China-specific inventory has substantially larger (20%) emissions for all China than the global

5 inventories. Our analysis uses observations to support and justify increased development of China-specific inventories in tracking China's progress towards reducing emissions. Here we are restricted to a single measurement site; effectively optimizing inventories at relevant spatial scales requires multiple high temporal resolution observations. We emphasize the need for a denser observational network in future efforts to measure and verify CO₂ emissions for China both regionally and as a whole.

10



1 Introduction

Achievement of emissions targets and reliable projection of climate trends requires accurate and independently validated greenhouse gas emissions inventories, particularly from the world's leading emitters. This paper focuses on independent testing of three CO₂ emission inventories for northern
5 China during the critical period 2005 through 2009. China's contribution to world CO₂ emissions has been steadily growing, becoming the largest in the world in 2006. China has accounted for 60% of the overall growth in global CO₂ emissions over the past 15 years (EIA, 2017). Under the United Nations Framework Convention on Climate Change (UNFCCC) 2015 Paris Climate Agreement, China has committed to reduce its carbon intensity (CO₂ emissions per unit GDP) by 60-65% relative to the
10 baseline year of 2005, and to peak carbon emissions overall by or before 2030.

China's emissions inventories for CO₂ have a large uncertainty, as indicated by differences in data reported at national and provincial levels. In 2012 this discrepancy was approximately half of China's 2020 emission reduction goals (EIA, 2017; NDRC, 2015; Guan et al., 2012; Zhao et al., 2012). Our
15 study addresses the critical need for independent and observational testing of emissions estimates to enable China to successfully achieve its policy targets. Moreover, China is under mounting pressure to address severe regional air pollution events that are often associated with CO₂ emissions sources—vehicles, power plants and other fossil fuel-burning operations. China's 11th Five Year Plan (11th FYP) of 2006-2010 included aggressive measures to retire inefficient coal-fired power plants and improve
20 energy efficiency in other industries starting in 2007 (Zhao et al., 2013; Nielsen & Ho, 2013). A number of pollution control measures that were implemented specifically in preparation for the 2008 Beijing Summer Olympics were also largely in effect by the end of 2007 (Nielsen & Ho, 2013; Wang et al., 2010). Our results thus provide an assessment of the effectiveness of these steps within our study period in our study region, encompassing much of northern China.

25

The efficacy and impact of many of China's emissions reduction measures can be determined only if there exist both an accurate baseline against which to compare future changes in emissions, and a reliable methodology for tracking those emissions. As evaluating all existing inventories is outside the



scope of our analysis, we focus on investigating the performance of three bottom-up anthropogenic inventories that represent the dominant methods currently employed for estimating China's CO₂ emissions. Two of the inventories, the Emissions Database for Global Atmospheric Research (EDGAR) and Carbon Dioxide Information Analysis Center (CDIAC), are spatial subsets from larger global
5 models of CO₂ emissions (PBL, 2013; Andres et al., 2016). They rely on national-level energy statistics and global default values for sectoral emission factors, and they estimate activity levels using generalized proxies (e.g. population). The third inventory (ZHAO) is specific to China, with greater reliance on energy statistics at provincial and individual facility levels as well as emission factors from domestic field studies (Zhao et al., 2012). The ZHAO inventory represents increased efforts in recent
10 years to incorporate more China-specific data into emissions inventories (e.g., <http://www.meicmodel.org/>; Shan et al., 2016).

To our knowledge, none of the China-specific CO₂ inventories have been evaluated with independent atmospheric observations. The official national total for China's 2005 CO₂ emissions from energy
15 related activities, used as the benchmark for the Paris commitment, is approximately 5.4Gton CO₂ (NDRC, 2015). ZHAO, EDGAR, and CDIAC report total 2005 energy-related CO₂ emissions that are higher by 31% (7.1Gton), 9%(5.9Gton), and 7%(5.8Gton) respectively. As the official national total is not available in a spatially allocated format, it cannot be tested by observations and we refer to it only as a benchmark in our analysis. We will show that the China-specific inventory (ZHAO) provides
20 excellent agreement with observations while the others do not. The result provides guidance for assessing, and potentially updating, the Paris agreement base year emissions.

In order to independently evaluate and optimize existing bottom-up estimates of China's CO₂ emissions, we employ a top-down approach using five years (January 2005 through December 2009) of
25 continuous hourly-averaged CO₂ observations measured in Miyun, China, at a site 100km northeast of Beijing (Wang et al., 2010). Modeled concentrations of CO₂ are obtained from convolving hourly CO₂ surface flux estimates with surface influence maps derived from the Stochastic Time-Inverted Lagrangian Transport Model driven with meteorology from the Weather Research and Forecasting



Model version 3.6.1 (WRF-STILT; Lin et al., 2003; Nehrkorn et al., 2010). NOAA CarbonTracker (CT2015) provides modeled estimates of advected upwind background concentrations of CO₂ that are enhanced or depleted by processes in the study region. As atmospheric CO₂ concentrations are significantly modulated by vegetation during the regional growing seasons we additionally prescribe

5 hourly biosphere fluxes of CO₂ using data-driven outputs from the Vegetation, Photosynthesis, and Respiration Model (VPRM) adapted for China (Mahadevan et al., 2012; Dayalu et al., 2017). VPRM provides a functional representation of biosphere fluxes based on data from remote sensing and eddy flux towers. The WRF-STILT-VPRM framework has been successfully adapted for similar emissions optimization studies in North America in regions where biogenic fluxes dominate surface processes

10 (e.g., Sargent et al., 2018; Karion et al. 2016; Matross et al., 2008). The relative magnitudes of biogenic fluxes and anthropogenic emissions in the Northern China region are comparable (Dayalu et al., 2017). In contrast to extensive measurement networks that exist in North America, continuous high-temporal resolution measurements of CO₂ necessary for inventory evaluation applications are sparse and very few datasets are available in China (Wang et al. 2010). Despite being restricted to a single measurement

15 station, our site provides valuable information and constraints on emissions inventories because it receives air at different times from one of the heaviest emitting regions of China, and clean air at other times. Our inventory optimization is confined to the Northern China region, but this region accounts for 33-41% of China's total annual CO₂ emissions from fossil-fuel combustion. Translating model-observation mismatch from concentration units (ppm) to mass units (Mton CO₂), we conduct a basic

20 benchmark optimization of the inventories for the 2005-2009 measurement time period. Our optimizations are resolved at the policy-relevant timescales of season and year.

Section 2 of this paper describes the observation CO₂ record used in this analysis. Sect.3 details the analysis methods, including WRF-STILT model configuration, a discussion of the main features of the

25 inventories, error evaluation, and inventory optimization methods. We present the results in Sect. 4, beginning with an assessment of seasonality impacts. We then compare inventory performance against observations across multiple timescales from hourly to annual. We conclude Sect. 4 with optimization results including a brief examination of regional carbon intensity over the study time period.



Concluding remarks are provided in Sect. 5. Additional methodological details are provided in the accompanying Supplementary Information (SI) and at <https://doi.org/10.7910/DVN/OJES00>.

2 CO₂ observations

This study uses five years (2005-2009) of continuous hourly averaged CO₂ observations (LI-COR
5 Biosciences Li-7000), made at a site in Northern China (Miyun; 40°29'N, 116°46.45'E). The Miyun
receptor is an atmospheric measurement station in a rural site 100 km northeast of the Beijing urban
center. It was established in 2004 by collaborating researchers at the Harvard China Project and
operated by researchers at Tsinghua University. The site is strategically located to capture both clean
continental background air from the west/northwest and polluted air from the Beijing region to the
10 southwest. Miyun is located south of the foothills of the Yan mountains; the region consists of
grasslands, small-scale agriculture, and mixed temperate forest. Further descriptions of the site and
details of the instrumentation of the CO₂ observations are in provided in Wang et al. (2010). Average
annual data coverage in this time period was 83% (range: 78% to 92%).

3 Methods

15 We evaluate performance of the ZHAO, EDGAR, and CDIAC inventories by modelling five years of
hourly CO₂ observations using the Stochastic Time-Inverted Lagrangian Transport Model (STILT; Lin
et al., 2003) run in backward time mode driven by high resolution meteorology from the Weather
Research and Forecasting Model version 3.6.1 (WRF). The WRF-STILT tool models the surfaces that
influenced each measurement hour in the study domain (Figure 1). Hourly vegetation CO₂ fluxes are
20 prescribed by the VPRM adapted for China (Mahadevan et al., 2008, Dayalu et al., 2017). We
categorize seasons by months based on regional growing season patterns, which are heavily dominated
by winter wheat/corn dual-cropping regions in the North China Plain (Dayalu et al. 2017). Winter wheat
emergence in the spring and corn emergence in later summer shift the seasonal patterns such that
regional seasons are more appropriately represented when months of year are grouped as January,



February, March (JFM/Winter); April, May, June (AMJ/Spring); July, August, September (JAS/Summer); and October, November, December (OND/Fall), respectively.

Ultimately, modeled concentrations of CO₂ are obtained from convolving hourly surface flux estimates with surface influence maps derived from the WRF-STILT framework. NOAA CarbonTracker (CT2015) provides estimates of advected upwind background concentrations of CO₂ that are enhanced or depleted by processes in the study region. Our final modeled-measurement data set is the subset consisting of local daytime values (1100h to 1600h) filtered to include only non-missing observations and CT2015 background values satisfying true background criteria as described in the SI, Sect. S6. We optimize inventories based on model-measurement mismatch of this final data subset. Model components are described individually below and in more detail in the SI, Sects S2 through S6.

3.1 WRF-STILT Model Configuration

The WRF-STILT particle transport framework and optimal configuration have been extensively tested in several studies using mid-latitude receptors (e.g., Sargent et al., 2018; McKain et al., 2014; Kort et al., 2013; McKain et al. 2012; Miller et al., 2012). WRF is configured with 41 vertical levels and two-way nesting in three domains, with the outermost domain covering nearly seven administrative regions (Figure 1, Figure 2), defined according to convention in Piao et al. (2009). The domain resolutions from coarsest to finest are 27km (d01), 9km (d02), and 3km (d03). Initial and lateral WRF boundary conditions are provided by NCEP FNL Operational Model Global Tropospheric Analyses at 1°x1° spatial 6-hourly temporal resolution (NCEP, 1999). Nudging of fields is implemented in the outer domain only, and never within the Planetary Boundary Layer (PBL). WRF output is evaluated against publicly accessible 24-hourly averaged observational datasets from the Chinese Meteorological Administration (CMA); finer temporal resolution meteorological data is not publicly available. Additional WRF run details and results from comparison with observations are presented in SI Sect. S2.

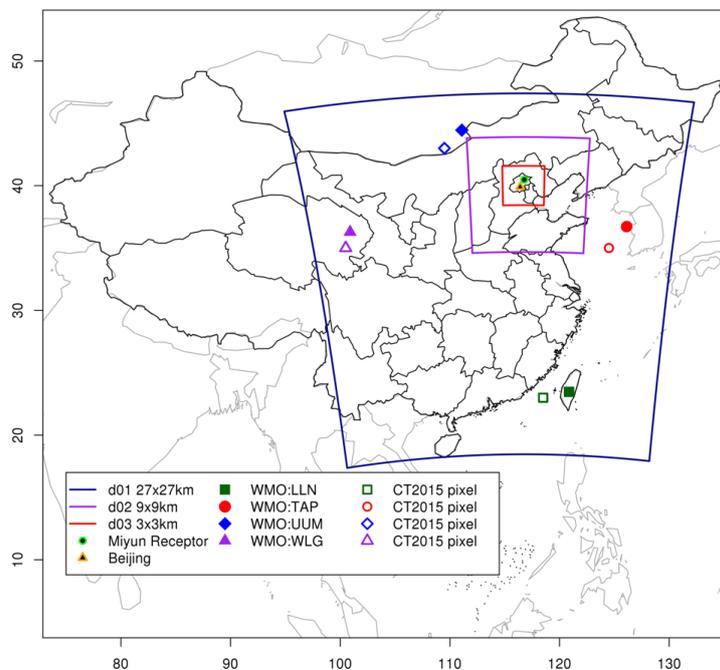


Figure 1. Study domain configuration. Miyun receptor and Beijing center are located within the innermost domain at a resolution of 3x3km. NOAA/ESRL/WMO (WMO) flask sampling sites used to evaluate bias in CT2015 modeled backgrounds are the solid shapes; nearest CT2015 comparison pixel is the corresponding unfilled shape.

The STILT model is configured in backward time mode, with the particle release point set as the Miyun sample inlet height of 158m above sea level (masl), corresponding to 6m above ground level (magl). Each hourly footprint (CO_2 concentration attributed to each unit of flux as $\text{ppm } \mu\text{mol}^{-1}\text{m}^2\text{s}$) is calculated from releasing 500 particles until they reach the outer domain boundaries up to seven days back in time.

- The STILT $0.25^\circ \times 0.25^\circ$ footprint map for each measurement hour enables assessment of regions in the study domain to which the receptor is most sensitive. We calculate STILT surface influence at the 50th ($L_{0.50}$), 75th ($L_{0.75}$), and 90th ($L_{0.90}$) percentile levels (Figure 2). $L_{0.90}$ —the region estimated as containing 90% of surfaces influencing measurement—is selected as the inventory optimization region.



The 1100 to 1600 local time period is used, as STILT performs best with a well-mixed boundary layer at the time of particle release (McKain et al., 2015). Further details are available in SI Sect. S3.

The complete list of WRF-STILT settings and STILT footprint files are available from
5 <http://dx.doi.org/10.7910/DVN/OJES00>.

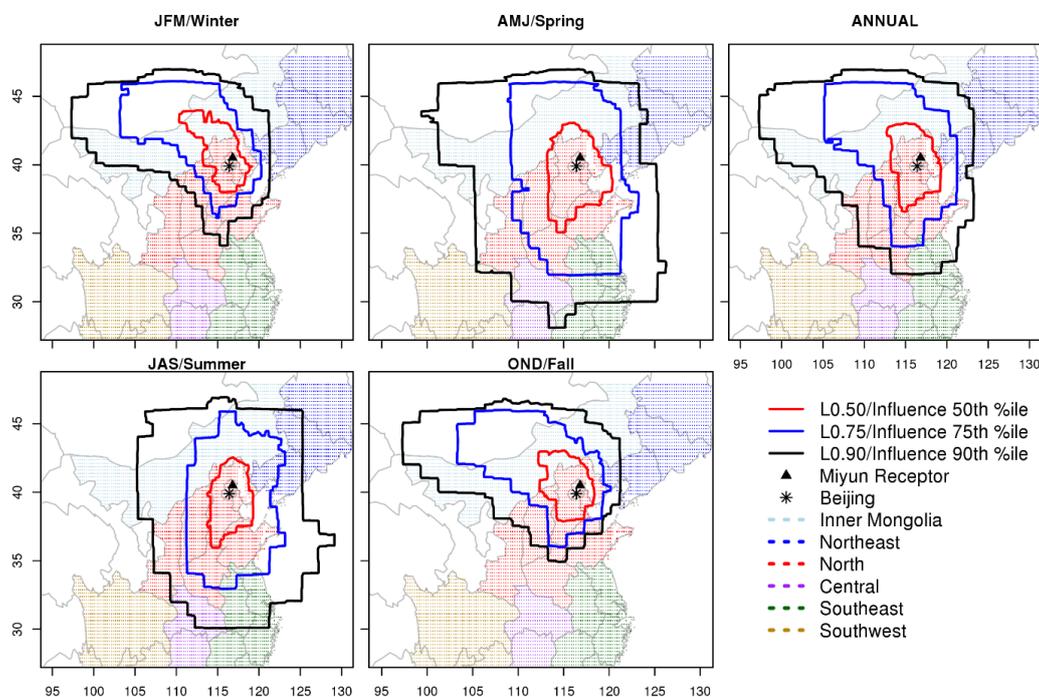


Figure 2. 2005-2009 mean seasonal (a-d) and Annual (e) footprint contours, as percentiles of influence highlighted by administrative region. Red, blue, and black contour lines represent 50th, 75th, and 90th percentile regions respectively. Stippling represents location of 0.25°x0.25° footprint and inventory gridcell centers, colored by relevant administrative regions. Northern China (red stippling) is the administrative region with predominant influence on Miyun observations, followed by Inner Mongolia and Northeast China. Southeast and Central China have minimal representation, and only during the spring and summer seasons.



3.3 Anthropogenic CO₂ Emissions Inventories

ZHAO, EDGAR, and CDIAC report estimates of total annual emissions of CO₂ at 0.25° x0.25°, 0.1°x0.1°, and 1°x1° original grid resolutions, respectively. We regridded the EDGAR and CDIAC
5 inventories to the 0.25° x0.25° resolution, using NCAR Command Language version 6.2.1 Earth System Modeling Framework conserve regridding algorithm to preserve the integral of emissions (Brown et al., 2012). Assessment of impact of regridding on total emissions for EDGAR and CDIAC is discussed in SI Sect. S4, and Figure S8. We present the main components and defining features of the three anthropogenic CO₂ inventories below.

10

The ZHAO inventory provides estimates of total annual emissions for 2005 through 2009. In addition, spatial location of emissions is given for years 2005 and 2009 on a 0.25° x0.25° grid. Using 2005 and 2009 gridded values, we calculate an average percent contribution of each grid cell to the total emissions. The average contributions are used as weights to spatially allocate 2006, 2007, and 2008
15 total annual emissions. We evaluate and justify this assumption in detail in the SI Sect. S4, and Figure S9. The ZHAO inventory represents one of the first statistically rigorous bottom-up CO₂ inventories for China. It relies on provincial- and facility-level data rather than national level data resulting in national CO₂ emissions estimates that are typically higher than those using national statistics. Satellite observations of criteria air pollutants (e.g., nitrogen dioxide, which serves as a proxy for fossil fuel
20 combustion) show greater agreement with provincial statistics. The increased use of China-specific emission factors and activity levels based on domestic field studies is a shift from other inventories that rely heavily on global averages to estimate processes occurring in China. Despite the increased incorporation of China-specific field data, the largest sources of uncertainty to the ZHAO inventory are industrial emission factors, and activity levels across all sectors. Total uncertainty in the inventory is
25 estimated as -9% to +11%. (Zhao et al., 2012).



The EDGAR emissions database continues to be a major prior in atmospheric studies, and the CO₂ inventory is used to inform key global scientific results considered by the UNFCCC Conference of Parties. The EDGAR global inventory takes total annual estimates of national emissions and downscales emissions to a 0.1° x 0.1° as a function of road/shipping networks, population density, energy/manufacturing point sources, and agricultural land. Estimates for China are available for all five years as gridded inventories. Reported uncertainties for global emissions are ±10% (<http://themasites.pbl.nl/tridion/en/themasites/edgar/documentation/uncertainties/index-2.html>).

However, this applies to global averaged uncertainty; the uncertainty for China is expected to be much higher.

10

We include the CDIAC inventory here due to its historical prevalence as a benchmark inventory for global indicators, including evaluations of carbon intensity provided by the World Bank (World Bank, 2017). The CDIAC inventory allocates estimates of national emissions to a 1° x 1° grid, primarily distributed according to human population density. A thorough assessment of 2σ uncertainties in the CDIAC spatial allocation of emissions shows considerable spread in regional uncertainties (Andres et al., 2016).

Based on multi-year means (2005 to 2009) and 95% confidence intervals derived from two-sample t-tests, we find that within the L_{0.90} optimization region EDGAR and CDIAC report emissions that are significantly lower than ZHAO by typically 20% (-24%, -16%) and 36% (-37%, -34%), respectively. Across China's administrative regions, the highest discrepancy between the global and regional inventories is in Northern China (ZHAO is approximately 30% higher than both EDGAR and CDIAC). In addition, Northern China represents one of the administrative regions with the highest CO₂ emissions density (2.3 to 3.3 kilotonnes of CO₂ per square kilometer, compared to the average of 0.7 ktCO₂ km⁻² averaged across China) and is therefore a particularly rich spatial subset for emissions inventory evaluation. A detailed breakdown of emissions by region of China is provided in the SI Table S2. Spatial differences are displayed in SI Figure S10.



3.4 Vegetation Flux Inventory

We prescribe biological contributions to the CO₂ signal by adapting the VPRM for the study domain which provides 0.25°x0.25° gridded estimates of hourly CO₂ net ecosystem exchange (*NEE*) from 2005 to 2009 (Dayalu et al., 2017). The VPRM is driven by 8-day 500m MODIS surface reflectance values and 10-minute averages of WRF downward shortwave radiation and surface temperature fields. The *NEE* is calibrated using eddy flux measurements representing each ecosystem type classified according to the International Geosphere-Biosphere Programme (IGBP) scheme. Eddy flux data are obtained from FluxNet and ChinaFlux collaborators. The L_{0.90} region is dominated by croplands (Figure S11a), in particular the winter wheat and corn dual-cropping that characterizes the North China Plain (Dayalu et al., 2017). See SI Sect. S5 for a summary of the VPRM adapted for the study region. VPRM output and additional model details are provided at <http://dx.doi.org/10.7910/DVN/RQLGLH>.

3.5 Background Concentrations

Appropriate quantification of background CO₂ concentrations (i.e., the CO₂ concentration at the lateral edges of the model domain and/or prior to interaction with domain surface processes) enables realistic assessment of the study domain's contribution to atmospheric CO₂ at varying timescales. CT2015 estimates of CO₂ concentrations are provided on a 3°x2° grid at upwind background locations. Background values are selected and corrected for large-scale biases using methodology similar to Karion et al. (2016) and is detailed in the SI Sect. S6.

20 3.6 Quantifying Regional Changes to Background CO₂ Concentrations: ΔCO₂

We define hourly ΔCO₂ as a regional change (enhancement or depletion) imparted to concentrations of CO₂ advected from the boundary (CO_{2,CT2015}) such that for each observation hour:

$$\Delta CO_{2,obs} = CO_{2,obs} - CO_{2,CT2015}$$



For each modeled hour, h , where i and j represent the surface gridcell locations:

$$\Delta CO_{2,mod} = \sum_{0h}^{-168h} foot_{ij} \times (ANTH_{ij} + VPRM_{ij})$$

- 5 Note that for the modeled enhancement or depletion, only the VPRM fluxes change hourly; as stated previously, the anthropogenic fluxes are temporally invariant over the course of a year.

Without a sufficiently dense network of high temporal resolution observations, full-scale inverse modeling approach to inventory optimization is inappropriate. Instead, we translate concentration (ppm) mismatch between observed and modeled ΔCO_2 into additive inventory corrections in mass units (Mton CO_2) at annual and seasonal timescales. We optimize emissions within the L_{0.90} contour (Figure 2) which represents regions that substantially influence the receptor, to avoid disproportionately weighting pixels that contribute very little to the observed signal. Details on inventory optimization methodology is provided in the SI Sect. S8.

15 3.6.1 Uncertainty Analysis

The sources of uncertainty in calculations of ΔCO_2 include uncertainty in CT2015 background concentrations, CO_2 observations, STILT footprints, anthropogenic inventories, and the VPRM vegetation inventory. We obtain 95% confidence bounds for ΔCO_2 by following a procedure similar to McKain et al., 2015 that involves bootstrapping daily averages of hourly afternoon values. For monthly and seasonal timescales, we obtain 95% confidence intervals for $\Delta CO_{2,obs}$ by performing a bootstrap on probability distributions of errors in both the CT2015 and observations 1000 times. (See SI Sect. S6 and Figure S17 for details on parameterizing CT2015 uncertainty.) The relevant quantiles are obtained from the resulting distribution, and are reported relative to the mean $\Delta CO_{2,obs}$ of the original data subset. We follow a slightly modified approach for $\Delta CO_{2,mod}$ in that we construct monthly and seasonal residual pools from daily averages of hourly afternoon $CO_{2,mod} - CO_{2,obs}$. The residuals—the deviation of the model from the true observed values—represent the total uncertainty in the model and therefore



aggregates the effects of uncertainty in the footprints, background, and inventories. Monthly and seasonal 95% confidence intervals of $\text{CO}_{2,\text{mod}}-\text{CO}_{2,\text{obs}}$ are then obtained from the distribution of bootstrapping the residual pools 1000 times. We then obtain the mean and 95% confidence interval of $\Delta\text{CO}_{2,\text{mod}}$ by applying the relevant quantiles of the residuals to the mean $\Delta\text{CO}_{2,\text{obs}}$ of the original data subset. Similar to McKain et al., 2015 distributions of seasonal averages obtained from the above method are used to estimate annual averages and 95% confidence intervals.

4 Results & Discussion

4.1 Impact of Seasonality on Optimization Region

As shown in Figure 2, we find strong seasonality in footprint extent and influence region, in agreement with previous analysis of Miyun observations by Wang et al. (2010). At annual timescales, the L_{0.90} optimization region is comparable to the WRF d02 extent. Northern China, including Inner Mongolia, dominate the L_{0.90} optimization region both seasonally and annually.

Due to the heavy biosphere influence in the regional growing season, previous work by Wang et al. (2010) used Miyun non-growing season measurements of CO_2 and carbon monoxide (CO) as an anthropogenic tracer to estimate combustion efficiency for China. When compared to bottom-up estimates of national combustion efficiency, observations suggested 25% higher combustion efficiency than bottom-up estimates of national combustion efficiency; however, Wang et al. (2010) note that the regional (Northern China) and seasonal (winter) subsets could contribute to such a discrepancy. The seasonality exhibited in Figure 2 indeed suggests that combustion efficiency estimates derived from non-growing season measurements alone do not represent anthropogenic processes in provinces south of Miyun that are visible in the observations primarily during the growing season. Low emitting regions northwest of Miyun such as Inner Mongolia dominate site influence in the fall and winter; spring and summer correspond to seasons where the higher emitting regions in provinces south heavily influence the Miyun receptor. However, non-growing season CO_2 is influenced by often inefficient district



heating in the northwest. And, while growing season CO₂ is influenced by intense urban activities from Beijing and other cities to the south, vegetation draws down the seasonal CO₂ significantly.

4.2 Unoptimized Models: Performance at multiple timescales

We evaluate unoptimized model performance relative to observations at hourly, seasonal, and annual
5 timescales. While inventory optimizations are performed at the policy relevant scales of seasons and years, examination of the models at shorter timescales provides insight into model bias and error aggregation at longer timescales.

All modeled hourly quantities include the same biological component from VPRM such that the only
10 source of variation among models is the anthropogenic inventory. At all timescales we note that CO_{2,ZHAO+VPRM} and ΔCO_{2,ZHAO+VPRM} consistently agree within uncertainty bounds to observations. With a few exceptions that are discussed in the following sections, CO_{2,EDGAR+VPRM}, CO_{2,CDIAC+VPRM}, ΔCO_{2,EDGAR+VPRM}, and ΔCO_{2,CDIAC+VPRM} systematically underestimate observations.

4.2.1 Hourly

15 We examine the distribution of modeled-measured residuals at hourly timescales for each anthropogenic inventory. While standard deviations are consistent across all models of CO₂ flux (1σ=9ppm; Figure S20) ΔCO_{2,ZHAO+VPRM} exhibits the least bias relative to observations with a non-significant mean residual of 0.25(-0.11,+0.62) ppm. In contrast, ΔCO_{2,EDGAR+VPRM} and ΔCO_{2,CDIAC+VPRM} display significantly greater bias by typically underestimating observations by large amounts:
20 2.1(1.7,2.4) ppm and 3.3(3.0,3.6) ppm, respectively. Here, the 95% confidence intervals are derived from a two-sample t-test. Details of the hourly analysis are provided in *SI Appendix* section S7 and Figures S19-S20, including evaluation of hourly bias by season. The EDGAR and CDIAC underestimation of ΔCO₂ at the hourly scale aggregates at longer timescales of seasons and years as discussed below.



4.2.2 Seasonal

With the exception of the growing season, $\Delta\text{CO}_{2,\text{EDGAR+VPRM}}$ and $\Delta\text{CO}_{2,\text{CDIAC+VPRM}}$ typically underestimate $\Delta\text{CO}_{2,\text{OBS}}$, even within the 95% uncertainty bounds. The VPRM has a sparse calibration

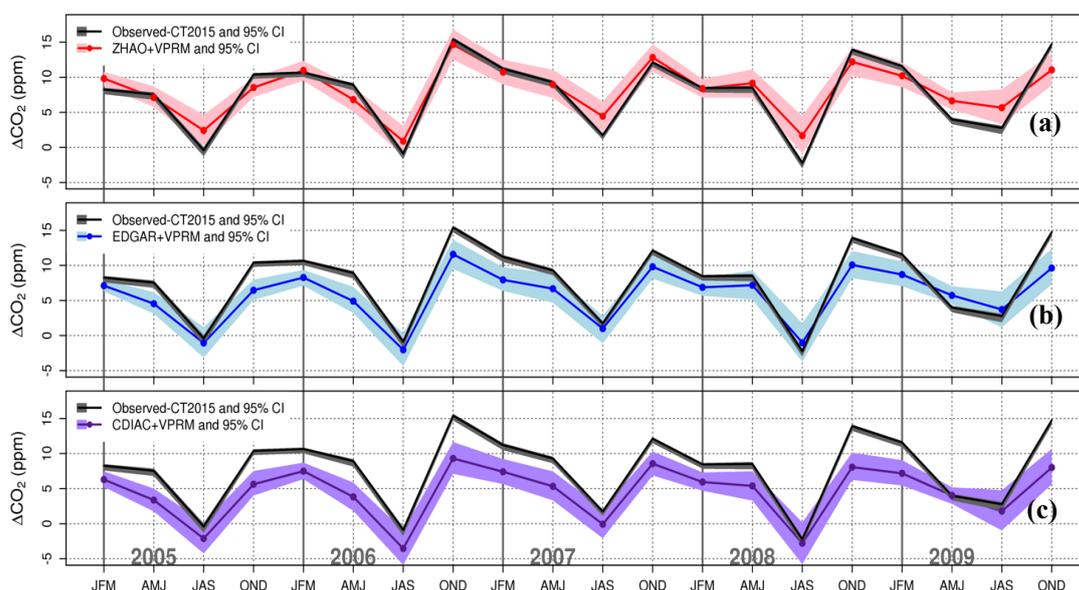


Figure 3. Modeled and Measured Seasonal ΔCO_2 . CT2015 background is subtracted from observations to provide observed ΔCO_2 (black line). 95% confidence bounds are derived from bootstrapping hourly afternoon concentrations for each season.

network, leading to an underestimate of regional CO_2 drawdown during the growing season (Dayalu et al., 2017). Therefore, while $\Delta\text{CO}_{2,\text{ZHAO+VPRM}}$ agrees within 95% confidence bounds with $\Delta\text{CO}_{2,\text{OBS}}$ during the non-growing seasons, $\Delta\text{CO}_{2,\text{ZHAO+VPRM}}$ generally overestimates CO_2 concentrations in the growing season (Figure 3a). $\Delta\text{CO}_{2,\text{EDGAR+VPRM}}$ (Figure 3b) and $\Delta\text{CO}_{2,\text{CDIAC+VPRM}}$ (Figure 3c) display lower CO_2 concentrations and generally result in better agreement with observations during the growing season than at other times of the year; however, based on our analysis at hourly timescales this is an artifact of lower anthropogenic emissions estimates relative to ZHAO that counteracts the VPRM



underestimating drawdown. Even during the growing season, $\Delta\text{CO}_2_{\text{CDIAC+VPRM}}$ agrees with observations typically at its upper confidence limits.

As ZHAO+VPRM demonstrates the least bias relative to observations at hourly and seasonal scales, we model the relative contributions to the monthly signal during the May through September peak regional
5 growing season as defined by Wang et al. (2010). Figure 4 displays the results from partitioning the mean monthly $\Delta\text{CO}_2_{\text{ZHAO+VPRM}}$ signal as a multi-year average into anthropogenic and vegetation contributions. While the WRF-STILT-VPRM framework has been successfully adapted for similar CO_2
inventory optimization studies in North American regions where biogenic fluxes dominate surface processes (Karion et al., 2016; Matross et al., 2006), Figure 4 shows the relative magnitude of biogenic
10 fluxes and anthropogenic emissions in the Northern China region is comparable during peak summer, making it difficult to independently constrain them with observational data. Furthermore, knowledge of the relative contribution of vegetation and anthropogenic processes to the CO_2 signal during the peak growing season is necessary to interpret satellite retrievals of CO_2 over the region (Dayalu et al., 2018).

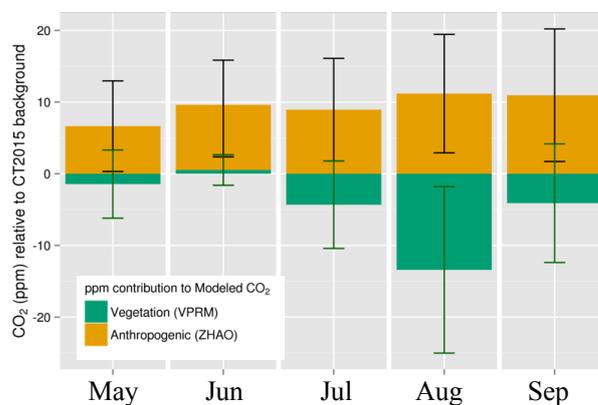


Fig. 4. Modeled mean monthly contribution (ppm) to Miyun CO_2 concentrations from vegetation (VPRM) and anthropogenic (ZHAO) sources. Enhancement and depletion is relative to advected CT2015 background concentrations during the regional growing season (MJJAS), averaged over 2005 to 2009. Vertical lines represent 1- σ of monthly averages (Green: Vegetation; Black: Anthropogenic). Negative values represent depletion from CT2015 background; positive values represent enhancement of CT2015 background.



4.2.3 Annual

Aggregation of uncertainty and anthropogenic inventory biases at shorter timescales becomes most apparent at the annual timescales. For annual budgeting we follow the assumptions of Piao et al. (2009),

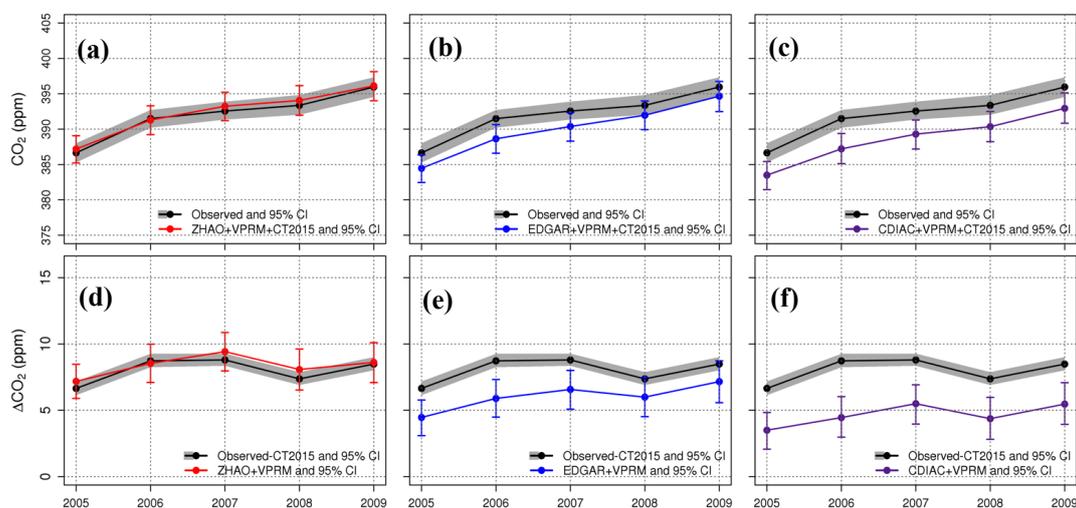


Fig. 5. Mean annual CO_2 and ΔCO_2 over entire study time period. (a-c) CO_2 annual concentration; (d-f) ΔCO_2 (regional enhancement, after removal of advected CT2015 background) with bootstrapped 95% confidence intervals.

5 and Jiang et al. (2016) that agricultural systems are in annual carbon balance because crop biomass has a short residence time. Therefore, while the VPRM is implicitly included in the modeled annual CO_2 and ΔCO_2 , vegetation carbon stocks of the heavily cropped influence region effectively turn over such that only the anthropogenic inventories dominate the modeled CO_2 signal. We evaluate annual CO_2 including CT2015 background (Figure 5a-c) and as regional enhancement relative to background
10 (Figure 5d-f).



For all years, $\text{CO}_{2,\text{ZHAO+VPRM}}$ and $\Delta\text{CO}_{2,\text{ZHAO+VPRM}}$ agree tightly within 95% uncertainty to observations (Figure 5a, Figure 5d). EDGAR+VPRM and CDIAC+VPRM are consistently biased significantly lower than observations.

4.3 Optimization of inventories at seasonal and annual timescales

5 We derive optimized inventories using additive flux corrections at seasonal timescales and multiplicative corrections at annual timescales. Complete seasonal and annual optimization results are provided in the SI Sect. S8, and Tables S3-S4.

The observational record informing the optimization integrates the biological and anthropogenic
10 signals. At the seasonal scale, where biological processes are significant contributors to the signal, we optimize the sum of the anthropogenic and biological fluxes (Figure 6). Optimized non-growing season flux estimates are higher than unoptimized values, partially accounting for the VPRM generally underestimating ecosystem respiration by an additive offset (Dayalu et al., 2017). As the vegetation component is controlled across models, the inter-model variance reflects the relative performance of the
15 anthropogenic estimates. We find that in the non-growing months the original ZHAO+VPRM inventory typically remains within the 95% confidence bounds of the optimized inventory. However, both EDGAR+VPRM and CDIAC+VPRM are consistently significantly lower than their optimized counterparts. This implies that both EDGAR and CDIAC underestimate anthropogenic emissions, and that ZHAO estimates are closer to actual emissions. During the growing seasons, however, the
20 afternoon vegetation signal is significant and the picture is more complex. In the spring, the CO_2 signal at Miyun is significantly affected by the North China Plain winter wheat growing season. The effect of optimization in the spring from 2005 to 2007 is to increase CO_2 emissions with a net positive seasonal flux; however, in 2008 and 2009 we find the net seasonal flux becomes negative such that uptake dominates emissions. The prior models in all cases predict positive flux. During the summer months,
25 ZHAO+VPRM predicts more emissions and/or less uptake relative to EDGAR+VPRM and CDIAC+VPRM. Optimization of summertime fluxes serves to significantly increase ZHAO+VPRM



uptake estimates; the EDGAR+VPRM and CDIAC+VPRM prior estimates are within the 95% confidence bounds of the optimization for reasons discussed previously.

We report annual optimized anthropogenic inventories in the $L_{0.90}$ region in Fig. 7 and Table S4 as
5 $\text{MtCO}_2\text{yr}^{-1}$. As discussed previously, the annual optimizations are applied only to the anthropogenic

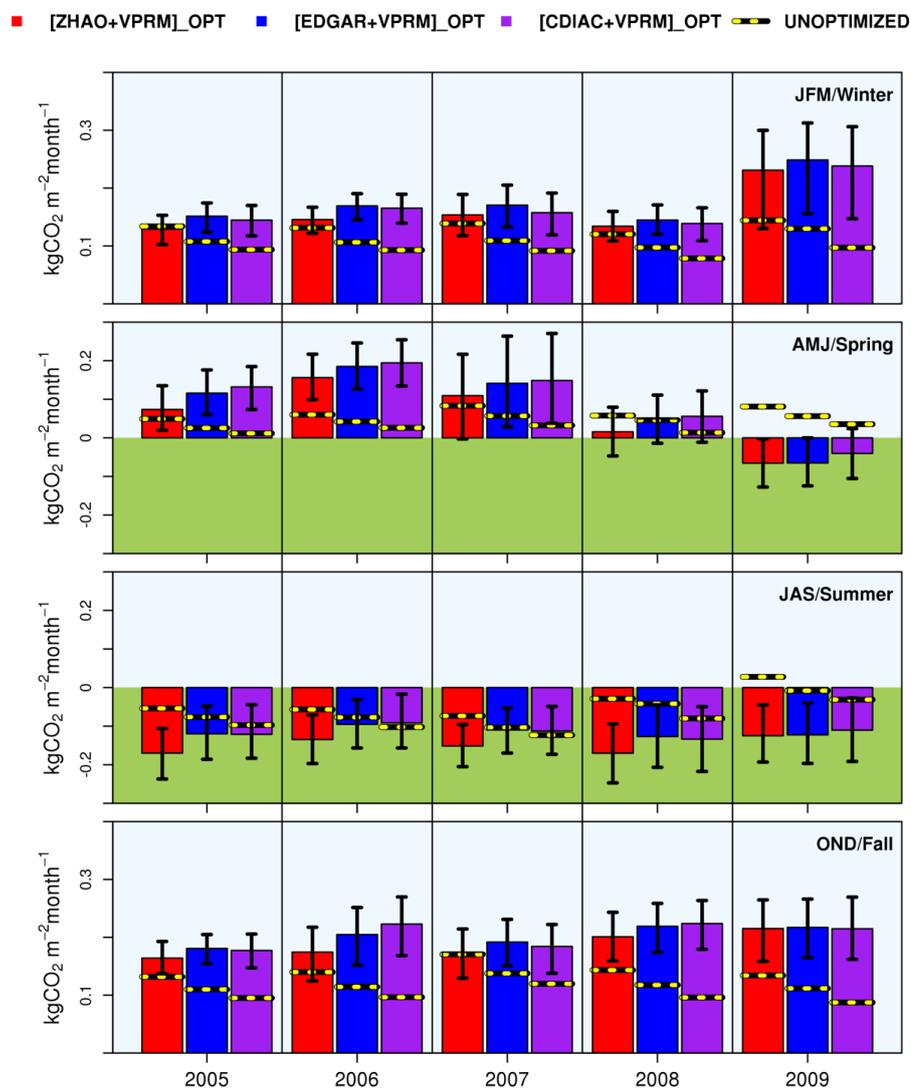


Figure 6. Optimized Seasonal Fluxes in the L_{0.90} region ($\text{kg CO}_2\text{m}^{-2}\text{month}^{-1}$). Anthropogenic and vegetation inventories are optimized together; all models have the same vegetation component (VPRM) and differ only in the anthropogenic inventory source. Shaded green represents negative flux (uptake by biosphere). Optimization based on additive corrections; difference among optimized inventories is due to differing spatial allocations by anthropogenic inventories. Bootstrapped 95% confidence intervals are represented by the black vertical lines.



inventory, as the signal at the annual timescale is effectively dominated by anthropogenic emissions; net ecosystem fluxes are expected to be relatively minor at the L_{0.90} extent in comparison. For all years, the emissions estimated by the original ZHAO inventory lie within the 95% confidence bounds of the optimized ZHAO inventory. However, for EDGAR and CDIAC, the original inventories consistently underestimate observations. Averaged over the five-year study period, EDGAR and CDIAC are typically lower than observations in the L_{0.90} region by 30% and 70% respectively (Fig. 5). Averaged across the five years, this translates to EDGAR and CDIAC being scaled relative to their unoptimized values in the L_{0.90} region by 1.3 and 1.7, respectively (Fig. 7; Table S4). We note that the relatively simple methodology involved in downscaling the coarse 1° x 1° CDIAC emissions allocation contributes to its particularly poor performance relative to observations. Specifically, when high emissions from a localized region are distributed across an entire 1° x 1° grid, their influence is diluted, highlighting that

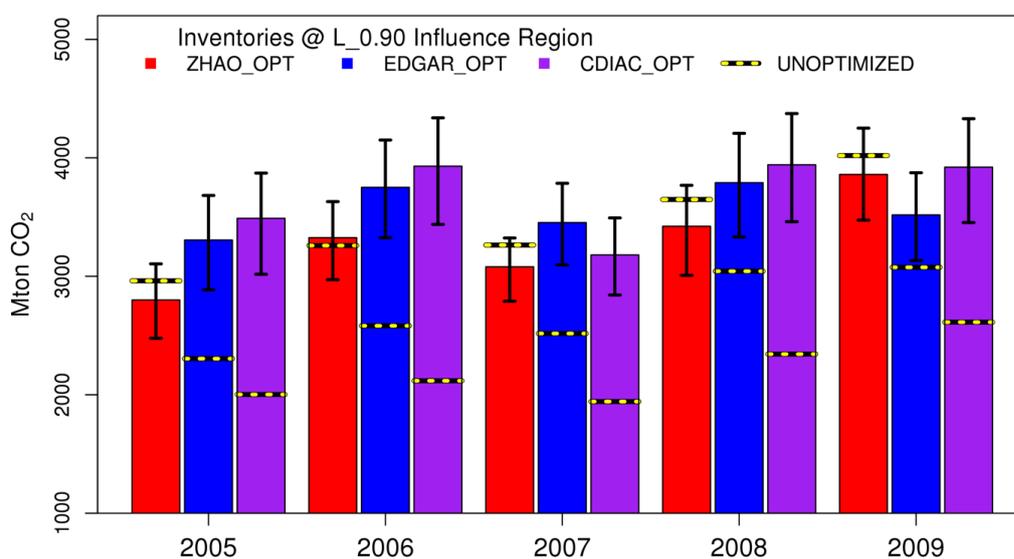


Fig. 7. Annually optimized emissions for 90th percentile of influence region. Optimization is based on multiplicative scaling factors. Difference among optimized inventory means is due to differing spatial allocations in original anthropogenic inventories. Bootstrapped 95% confidence intervals are represented by the black vertical lines. *Note the y-axis origin begins at 1000 Mton CO₂ for visual clarity.



CDIAC, or any other similar coarse-resolution inventory, is not well suited for this type of regional study making them difficult to test with observations.

In all optimizations, the differences in means among inventories is a result of differing proxies for spatial allocation among the inventories. In other words, the optimal annual mean and distribution for each inventory is based on the inventory's existing method for spatial allocation. Determining optimal spatial allocation is outside the scope of this study as doing so requires data from a network of high-resolution observations, which does not exist.

4.4 Regional Patterns in Emissions from 2005 to 2009

We examine the statistical significance of the inter-annual observed concentration and enhancement differences using a two sample t-test (Table 1). The observed concentrations including advected global background (Figure 5, top row) display an overall increasing trend of 1.87 (1.8, 1.9) ppm CO₂ yr⁻¹ between 2005 and 2009, in agreement with flask samples obtained from nearby WMO sites between 2007 and 2010 (Liu et al., 2014). The inter-annual increases are statistically significant (Table 1). However, when we remove the modeled background to more closely examine regional patterns that would otherwise be drowned out by the global signal, we find that the regional (Δ CO₂) trend does not parallel the increasing global trend (Figure 5, bottom row; Table 1). Regionally, the observed enhancements increase from 2005 to 2006 and plateau in 2007 before decreasing in 2008. Enhancements increase again in 2009.

20

25

30



Table 1. Inter-annual observed CO₂ and ΔCO₂ differences. Differences are of observations between consecutive years. 95% confidence intervals are derived from a two-sample t-test. Italicized entries denote instances where the inter-annual difference is not statistically significant (confidence interval includes zero).

5

Time Interval (y₂-y₁)	CO_{2,OBS} (ppm) Mean Difference (95% CI)	ΔCO_{2,OBS} (ppm) Mean Difference (95% CI)
2006-2005	4.86 (4.5, 5.2)	2.08 (1.9, 2.3)
2007-2006	1.08 (0.69, 1.5)	<i>0.0693 (-0.15, 0.29)</i>
2008-2007	0.772 (0.37, 1.2)	-1.43 (-1.6, -1.2)
2009-2008	2.60 (2.2, 3.0)	1.12 (0.88, 1.4)
2009-2005	9.31 (8.9, 9.7)	1.84 (1.6, 2.0)

In Figure 8a we estimate Gross Regional Product (GRP) for eight of China's 34 provincial-level administrative units, specifically those encompassed significantly by the L_{0.90} influence contour:

10 Beijing, Tianjin, Henan, Shanxi, Shandong, Hebei, Inner Mongolia, and Liaoning. We suggest that industrial energy efficiency improvements beginning in 2007 under the 11th FYP, preparations and staging of the 2008 Beijing Summer Olympics, and the global financial crisis in late 2008 followed by a large Chinese fiscal stimulus in 2009 are likely contributors to the observed interannual variation in regional CO₂ emissions (Figure 5d-e) while also compatible with a doubling of GRP from 2005 to 2009
15 (Figure 8a). In addition, earlier work by Wang et al. (2010) extends Miyun observations of CO₂ growth rate to all of China and estimates a lower growth rate than previously suggested. However, Figure S9 suggests local reductions in regions influencing Miyun, possibly in preparation for the Beijing Olympics, are partially offset by increases elsewhere. A larger network of sites would be needed to quantify this further in order to evaluate the CO₂ growth rate for other regions in China and for China as
20 a whole.

As policy targets are often measured as relative changes over multiple years, an important component of emissions inventories is their ability to accurately capture multi-year changes. Observations indicate enhancements above background CO₂ increased by 28% (22%, 34%) between 2005 and 2009.



ZHAO+VPRM estimates a 20% increase over the same time period while EDGAR+VPRM and CDIAC+VPRM estimate 61% and 56% increases respectively.

4.4 Implications for Assessing National Carbon Emission Targets

China has pledged a 60-65% reduction in carbon intensity by 2030 and has additionally set a benchmark
5 of 40-45% reduction in carbon intensity by 2020, where both targets are relative to the baseline year
2005 (NDRC, 2015; Guan et al., 2014). However, Guan et al. (2014) found that provincial trends in
carbon intensity can vary significantly from national trends. Using the GRP values shown in Figure 8a,
we calculate a Northern China regional carbon intensity (Figure 8b). The eight provinces are those that
are encompassed significantly by the L_{0.90} influence contour: Beijing, Henan, Shanxi, Tianjin,
10 Shandong, Hebei, Inner Mongolia, and Liaoning. We also estimate an L_{0.90} regional carbon intensity
based on the official national energy-related CO₂ emissions in NDRC (2015); we scale the national total
by 39% (35%,42%) which is the mean (range) contribution of the L_{0.90} region to the national
emissions in 2005, averaged across the three unoptimized gridded emissions inventories. We emphasize
that carbon intensity values are inherently uncertain due to complexities in GRP and Gross Domestic
15 Product (GDP) calculations such as double-counting due to inter-provincial trade or spatial mismatch
between emissions and economic data. Nevertheless, the analysis provides valuable insight into trends
rather than precise values.

Over the study time period, the GRP of the L_{0.90} region more than doubled (Figure 8a), evidently
20 correlated to a significant increase in emissions. Coinciding with the 2008 Beijing Summer Olympics,
the region's contribution to China's GDP grew from approximately 13.5% in 2007 to nearly 16% in
2008, representing a 20% increase, before plateauing into 2009 (Figure 8a). As noted in Guan et al.
(2014), reductions in carbon emissions intensity can come about via two main pathways: the first,
within industries, through increased energy efficiency combined with expanded production capacity; the
25 second, across the economy, through structural shifts from energy-intensive industrial sectors to service
sectors. The doubling of GRP suggests enlarged production capacity as a driver for regional carbon
intensity reductions. From 2005 to 2009, carbon intensity for the L_{0.90} region decreased by 47%



(28%,65%), based on a one-sample t-test of pooled emissions intensity changes across optimized inventories. Analysis presented by organizations such as the World Bank (World Bank, 2017) suggests China's carbon intensity at the national level decreased by 20% in 2009 relative to 2005. However, we note that the carbon emissions data source for the World Bank carbon intensity calculations is CDIAC.

- 5 We have shown that at least for the L_0.90/Northern China region, CDIAC emissions data for the base year 2005 was 70% (50%, 90%) too low.

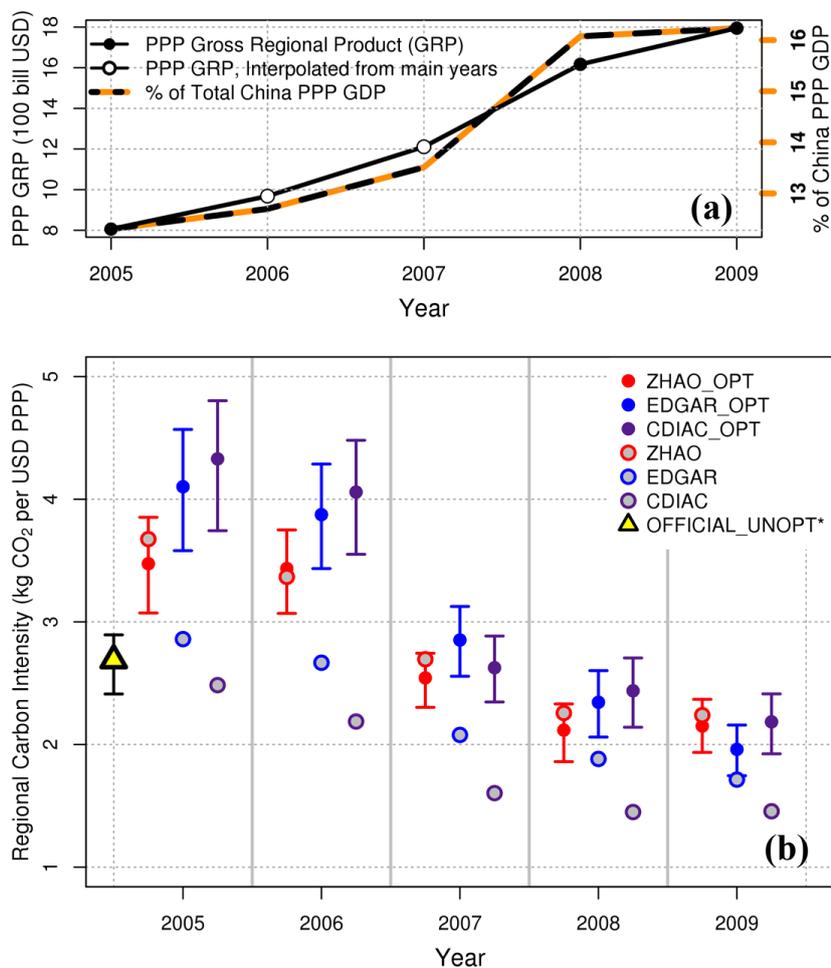


Fig. 8. Estimates of Regional Carbon Intensity ($\text{kg CO}_2 \text{ USD}_{\text{PPP}}^{-1}$). (a) PPP GRP by year and as a % of China's national GDP. No PPP GRP values were available for 2006 and 2007; PPP GRP for these years was instead calculated by linearly interpolating Nominal GRP/PPP GRP for 2005, 2008, and 2009. (b) Regional Carbon intensity using optimized (solid) and unoptimized (grey) CO₂ estimates. Uncertainty bars are bootstrapped 95% confidence intervals. GRP, GDP data from IMF, World Bank, China Statistical Yearbook. Provinces used in GRP calculation are those significantly encompassed by L_{0.90} contour: Beijing, Henan, Shanxi, Tianjin, Shandong, Hebei, Inner Mongolia, and Liaoning. *Estimated by scaling the official national emissions total by the average contribution (39%) of L_{0.90} region to total emissions in 2005. Uncertainty bars represent the % contribution range estimated by ZHAO, EDGAR, and CDIAC in 2005 (35%, 42%).



5 Conclusions

Continuous hourly CO₂ observations, significantly influenced by the heavily CO₂-emitting Northern China region, are used in a top-down evaluation and optimization of three bottom-up CO₂ flux inventories. We focus on the policy-relevant time interval from 2005 to 2009, noting that 2005 is China's baseline year for carbon commitments. The three inventories are distinct in their anthropogenic component, with a common biogenic flux component provided by the VPRM, a simple satellite data-driven biosphere model. The ZHAO anthropogenic emissions inventory incorporates a regional approach to China's CO₂ emissions estimation, using activity data at the provincial and facility-levels as well as domestic emission factors. The EDGAR and CDIAC emissions inventories incorporate a greater reliance on global averages and China's national statistics and international default emission factors, and depend more heavily on proxies (e.g., population) to allocate the emissions geographically. The three anthropogenic inventories are selected such that the typical range of inventory methods available for China are well-represented.

We find strong seasonality in L_{0.90} footprint extent and influence region, with the northwest dominating non-growing season and a more uniform influence in the growing season. The Northern China administrative region, excluding Inner Mongolia, dominates the L_{0.90} influence region (Figure 2). Within the L_{0.90} inventory optimization region, EDGAR and CDIAC are lower than ZHAO by 20% and 36% respectively. Across administrative regions, the highest discrepancy between the global and regional inventories is in Northern China, where the ZHAO inventory estimates emissions that are on average 30% higher than both EDGAR and CDIAC (SI, Table S2).

We find the ZHAO+VPRM inventory agrees very closely with observations, much better than the nationally referenced inventories at all timescales. During the peak growing season, the regional enhancement to background CO₂ concentrations is approximately zero, due to an agriculturally dominated vegetation signal that is equal and opposite in sign to the anthropogenic signal, so this conclusion applies strictly to the other three seasons. At annual timescales, the anthropogenic signal dominates and we find that EDGAR and CDIAC underestimate emissions in the Northern China region



- by an average of 30% and 70% respectively. In contrast, the ZHAO inventory gives *a priori* results very close to observations and is not significantly affected by the optimization. Note that the EDGAR and CDIAC inventories can differ from -10% to -20% relative to ZHAO in their national emissions totals (Table S2). Since the ZHAO methodology gives comparatively accurate and higher results for the
- 5 influence region dominated by Northern China, we hypothesize that the proxies used in the global inventories have biases that likely result also in overestimation in other regions of China. However, observational data from strategically located stations in and around the eastern half of China are required to explore this hypothesis.
- 10 Our results, backed by a robust high-resolution time series of CO₂ observations, show that assessments of China's CO₂ emissions require regional inventories with a methodology such as that employed in ZHAO, where China-specific field and facility-level data are used with increased reliance on provincial energy statistics. We show that global inventories can significantly underestimate annual CO₂ emissions in China and that the regional partitioning can be far off. In situ CO₂ observations interpreted within a
- 15 high-resolution model framework such as described in this study provide a powerful constraint to test and correct spatially explicit inventories. The single station available for the 2005-2009 period was strategically located to provide information on one of the highest CO₂ emitting regions of China. Within that limitation, the observations provide strong evidence supporting the use of China-specific methods, such as those employed in ZHAO, for China's CO₂ emissions inventory derivation. A denser network of
- 20 CO₂ measurement stations in China is required as a basis for effective monitoring, reporting, and verification of regional and national inventories.



Code and Data Availability

Code and data are available at <https://doi.org/10.7910/DVN/OJESO0>. The supplement includes observational and modeled CO₂ time series, WRF and STILT parameter files, and STILT footprint files.

5

Author Contributions

A.D., J.W.M, and S.C.W designed the research. A.D. performed the research with guidance from all co-
10 authors. Y.W. and J.W.M monitored, maintained, and provided access to the Miyun hourly
observational data set. Y.Z. provided the China-specific anthropogenic inventory. WRF-STILT
simulations were performed by A.D. with assistance from T.N. A.D. constructed the vegetation CO₂
inventory. A.D. wrote the paper with contributions from all co-authors.

15

Competing Interests

The authors declare no competing interests.

20

Acknowledgments

We acknowledge the Harvard-China Project and the Harvard Global Institute for funding this study. We
thank Zhiming Kuang for providing computational resources. We also thank Jenna Samra, Maryann
Sargent, and Victoria Liublinska for helpful discussion.

25



References

- Andres, R.J., Boden, T.A., and Marland, G.: Annual Fossil-Fuel CO₂ 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, 2016a.
- Andres, R. J., Boden, T. A., and Higdon, D. M.: Gridded uncertainty in fossil fuel carbon dioxide emission maps, a CDIAC example, Atmos. Chem. Phys., 16, 14979-14995, doi:10.5194/acp-16-14979-2016, 2016b.
- Brown D, Brownrigg R, Haley M, & Huang W (2012) *The NCAR Command Language (NCL) v6.0. 0*. UCAR/NCAR Computational and Information Systems Laboratory, Boulder, CO. Available at <http://dx.doi.org/10.5065/D6WD3XH5>.
- Dayalu, A., Munger, W., Wofsy, S. C., Wang, Y., Nehrkorn, T., Zhao, Y., McElroy, M. B., Nielsen, C., and Luus, K.: VPRM-CHINA: Using the Vegetation, Photosynthesis, and Respiration Model to partition contributions to CO₂ measurements in Northern China during the 2005–2009 growing seasons, Biogeosciences Discuss., <https://doi.org/10.5194/bg-2017-504>, in review, 2017.
- European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL): Emission Database for Global Atmospheric Research (EDGAR), release EDGARv4.2 FT2010, <http://edgar.jrc.ec.europa.eu>, 2013.
- Guan, D., Liu, Z., Geng, Y., Lindner, S., and Hubacek, K.: The gigatonne gap in China's carbon dioxide inventories, Nat. Clim. Chg, 2,672–675, doi:10.1038/nclimate1560, 2012.
- Guan, D., Klasen, S., Hubacek, K., Feng, K., Liu, Z., He, K., Geng, Y., and Zhang Q.: Determinants of stagnating carbon intensity in China, Nat. Clim. Chg, 4, 1017-1023, doi:10.1038/nclimate2388, 2014.
- Hegarty, J., Draxler, R., Stein, A., Brioude, J., Mountain, M., Eluszkiewicz, J., Nehrkorn, T., Ngan, F., and Andrews, A.: 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, 2013.
- Karion, A., Sweeney, C., Miller, J. B., Andrews, A. E., Commane, R., Dinardo, S., Henderson, J. M., Lindaas, J., Lin, J. C., Luus, K. A., Newberger, T., Tans, P., Wofsy, S. C., Wolter, S., and Miller, C. E.: 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, 2016.



- Kort, E. A., Angevine, W.M., Duren, R., and Miller, C.E.: Surface observations for monitoring urban fossil fuel CO₂ emissions: Minimum site location requirements for the Los Angeles megacity, *J. Geophys. Res. Atmos.*, 118, 1577–1584, doi:10.1002/jgrd.50135, 2013.
- 5 Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *Journal of Geophysical Research-Atmospheres*, 108, 4493, doi:10.1029/2002JD003161, 2003.
- 10 Liu, Z., Guan, D., Wei, W., Davis, S.J., Ciais, P., Bai, J., Peng, S., Zhang, Q., Hubacek, K., Marland, G., Andres, R.J., Crawford-Brown, D., Lin, J., Zhao, H., Hong, C., Boden, T.A., Feng, K., Peters, G.P., Xi, F., Liu, J., Li, Y., Zhao, Y., Zeng, N., and He, K.: Reduced carbon emission estimates from fossil fuel combustion and cement production in China. *Nature* 524, 335–338, 2015.
- 15 Mahadevan, P., Wofsy, S.C., Matross, D.M., Xiao, X., Dunn, A.L., Lin, J.C., Gerbig, C., Munger, J.W., Chow, V.Y. and Gottlieb, E.W.: A satellite-based biosphere parameterization for net ecosystem CO₂ exchange: Vegetation Photosynthesis and Respiration Model (VPRM), *Global Biogeochem. Cycles*, 22, GB2005, doi:10.1029/2006GB002735, 2008.
- 20 Matross, D. M., Andrews, A., Pathmathevan, M., Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Lee, J. T., Zhao, C. L., Bakwin, P. S., Munger, J. W., and Hollinger, D. Y.: Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data, *Tellus Series B-Chemical and Physical Meteorology*, 58, 344-358, 2006.
- 25 McKain, K., Down, A., Raciti, S. M., Budney, J., Hutyra, L. R., Floerchinger, C., Herndon, S. C., Nehrkorn, T., Zahniser, M. S., and Jackson, R. B.: 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, 2015.
- 30 McKain, K., Wofsy, S.C., Nehrkorn, T., Eluszkiewicz, Ehleringer, J.R., and Stephens, B.B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region, *Proc. Nat. Acad. Sci.*, 109(22), 8423–8428, 2012.
- 35 Miller, S.M., Kort, E.A., Hirsch, A.I., Dlugokencky, E.J., Andrews, A.E., Xu, X., Tian, H., Nehrkorn, T. Eluszkiewicz, J., Michalak, A.M., and Wofsy, S.C.: Regional sources of nitrous oxide over the United States: Seasonal variation and spatial distribution, *J. Geophys. Res.*, 117, D06310, doi:10.1029/2011JD016951, 2012.
- 40 Nassar, R., Napier-Linton, L., Gurney, K.R., Andres, R.J., Oda, T., Vogel, F.R., and Deng, F.: Improving the temporal and spatial distribution of CO₂ emissions from global fossil fuel emission data sets, *J. Geophys. Res. Atmos.*, 118, 917–933, doi:10.1029/2012JD018196, 2013.



- Nielsen, C. and Ho, M.: Clearer Skies Over China: Reconciling Air Quality, Climate, and Economic Goals, MIT Press, ISBN-13: 9780262019880, DOI:10.7551/mitpress/9780262019880.001.0001, 2013.
- 5 NCEP National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce: 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, 2000.
- 10 NDRC National Development Reform Commission: Enhanced Actions on Climate Change: China's Intended Nationally Determined Contributions, 2015.
- Nehrkorn, T., Eluszkiewicz, J., Wofsy, S.C., Lin, J., Gerbig, C., Longo, M., and Freitas, S.: Coupled
15 weather research and forecasting–stochastic time-inverted lagrangian transport (WRF–STILT) model, Meteorol. Atmos. Phys., 107: 51. doi:10.1007/s00703-010-0068-x, 2010.
- Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S., and Wang, T.: The carbon balance of
20 terrestrial ecosystems in China, Nature, 458(7241), 1009–1013, 2009.
- Sargent, M., Barrera, Y., Nehrkorn, T., Hutyra, L., Gately, C., Jones, T., McKain, K., Sweeney, C.,
Hegarty, J., Hardiman, B., Wofsy, S.: Anthropogenic and biogenic CO₂ fluxes in the Boston urban
region, Proc. Natl. Academy Sci (*in press*), 2018.
- 25 Shan, Y., Liu, J., Liu, Z., Xu, X., Shao, S., Wang, P., Guan, D.: New provincial CO₂ emission
inventories in China based on apparent energy consumption data and updated emission factors, Applied
Energy, v184, 742-750, <https://doi.org/10.1016/j.apenergy.2016.03.073>, 2016.
- 30 Wang, Y., Munger, J., Xu, S., McElroy, M., Hao, J., Nielsen, C., and Ma, H.: CO₂ and its correlation
with CO at a rural site near Beijing: Implications for combustion efficiency in China, Atmos. Chem. and
Phys., 10: 8881-8897, 2010.
- World Bank (2017). CO₂ emissions (kg per PPP \$ of GDP). Available at
35 <https://data.worldbank.org/indicator/EN.ATM.CO2E.PP.GD?locations=CN>. Accessed May 12, 2017.
- Zhao, Y., Nielsen, C.P., and McElroy, M.: China's CO₂ emissions estimated from the bottom up:
Recent trends, spatial distributions, and quantification of uncertainties, Atmos. Envir., 59, 214-223,
2012.
- 40 Zhao, Y., Zhang, J., and Nielsen, C. P.: The effects of recent control policies on trends in emissions of
anthropogenic atmospheric pollutants and CO₂ in China, Atmos. Chem. Phys., 13, 487-508,
doi:10.5194/acp-13-487-2013, 2013.