



HAL
open science

Seasonal Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-Scale Boundary Inflow in Atmospheric CO₂ Simulations Over North America

Sha Feng, Thomas Lauvaux, Kenneth Davis, Klaus Keller, Yu Zhou, Christopher Williams, Andrew Schuh, Junjie Liu, Ian Baker

► **To cite this version:**

Sha Feng, Thomas Lauvaux, Kenneth Davis, Klaus Keller, Yu Zhou, et al.. Seasonal Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-Scale Boundary Inflow in Atmospheric CO₂ Simulations Over North America. *Journal of Geophysical Research: Atmospheres*, 2019, 124 (24), pp.14325-14346. 10.1029/2019JD031165 . hal-04237265

HAL Id: hal-04237265

<https://hal.science/hal-04237265>

Submitted on 11 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2019JD031165

Special Section:

Carbon and Weather: Results from the Atmospheric Carbon and Transport -- America Mission

Key Points:

- Our calibrated ensemble system demonstrates that biogenic fluxes dominate the uncertainty in atmospheric boundary layer CO₂
- Transport uncertainty is highly correlated with biogenic flux uncertainty in space and time
- Large-scale CO₂ boundary inflow dominates uncertainty in the atmospheric column-integrated CO₂ uncertainty

Supporting Information:

- Supporting Information S1

Correspondence to:

S. Feng,
sfeng@psu.edu

Citation:

Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., et al. (2019). Seasonal characteristics of model uncertainties from biogenic fluxes, transport, and large-scale boundary inflow in atmospheric CO₂ simulations over North America. *Journal of Geophysical Research: Atmospheres*, 124, 14,325–14,346. <https://doi.org/10.1029/2019JD031165>

Received 12 JUN 2019

Accepted 8 DEC 2019

Accepted article online 11 DEC 2019

Published online 25 DEC 2019

©2019. American Geophysical Union.
All Rights Reserved.

Seasonal Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-Scale Boundary Inflow in Atmospheric CO₂ Simulations Over North America

Sha Feng¹, Thomas Lauvaux^{1,2}, Kenneth J. Davis^{1,3}, Klaus Keller^{3,4}, Yu Zhou⁵, Christopher Williams⁵, Andrew E. Schuh⁶, Junjie Liu⁷, and Ian Baker⁶

¹Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA,

²Laboratoire des Sciences du Climat et de l'Environnement, CEA, CNRS, UVSQ/IPSL, Université Paris-Saclay, Orme des Merisiers, Gif-sur-Yvette, France, ³Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA, ⁴Department of Geoscience, The Pennsylvania State University, University Park, PA, USA,

⁵Department of Biology, Graduate School of Geography, Clark University, Worcester, MA, USA, ⁶Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA, ⁷NASA Jet Propulsion Laboratory, The California Institute of Technology, Pasadena, CA, USA

Abstract Regional estimates of biogenic carbon fluxes over North America from both atmospheric inversions (“top-down” approach) and terrestrial biosphere models (“bottom-up”) remain highly uncertain. We merge these approaches with an ensemble-based, regional modeling system able to diagnose and quantify the causes of uncertainties in top-down atmospheric estimates of the terrestrial sink over North America. Our ensemble approach quantifies and partitions the uncertainty stemming from atmospheric transport, the biosphere, and large-scale CO₂ boundary inflow (boundary conditions). We use meteorological data, CO₂ fluxes, and CO₂ mole fraction measurements to assure the reliability of the ensemble system. Our results show that all uncertainty components have clear seasonal variations. The biogenic flux component dominates modeled boundary layer CO₂ uncertainty, ranging from 2.5 ppm in summer and winter to 1.5 ppm in fall and spring. Spatially, it remains highly uncertain in the U.S. Corn Belt regions. Transport uncertainty reaches a maximum of 2.5 ppm in the summer months and stays at 1.2 ppm for the rest of the year and is highly correlated with the biogenic CO₂ fluxes. Boundary conditions play the smallest role in atmospheric boundary layer CO₂ uncertainty with a magnitude smaller than 1 ppm. However, boundary conditions are the most important uncertainty component in column-averaged CO₂ (XCO₂). The spatiotemporal variations of the uncertainties in modeled XCO₂ are similar to those in atmospheric boundary layer CO₂.

Plain Language Summary Uncertainty in future uptake of carbon dioxide (CO₂) by terrestrial ecosystems drives divergent projections of future climate and uncertainty in prescriptions for climate mitigation. North American ecosystems are currently a significant sink of atmospheric CO₂, but this sink is difficult to measure and thus understand. Terrestrial uptake and emission of CO₂ from a continent such as North America can be inferred from atmospheric CO₂ observations. This approach, however, requires an accurate understanding of all the sources of uncertainty in simulating atmospheric CO₂. These include errors in the simulations of atmospheric transport, fossil fuel emissions of CO₂, and transport of CO₂ from the rest of the Earth's continents and oceans (boundary conditions), in addition to the uncertainty terrestrial CO₂ uptake. Here we quantify the uncertainty in all of these components of an atmospheric CO₂ simulation over North America using an ensemble of simulations. Observations of atmospheric winds and mixing, atmospheric CO₂ mole fractions, and land-atmosphere CO₂ fluxes are used to ensure that the ensemble cover an appropriate range of uncertainty in each element of our simulation. Our results show that over North America, (i) terrestrial biosphere CO₂ fluxes dominate the uncertainty in the atmospheric boundary layer CO₂, (ii) atmospheric transport uncertainty is highly correlated with biosphere flux uncertainty in space and time, and (iii) inflow of CO₂ from other continents and oceans plays the least important role in North American lower atmosphere CO₂ but is the largest source of uncertainty in atmospheric column-integrated CO₂. Our study shows that atmospheric transport and continental boundary conditions are important limits to our ability to infer terrestrial ecosystem CO₂ uptake using, respectively lower atmosphere and atmospheric column CO₂ measurements.

1. Introduction

The evolution of the carbon cycle is one of the main sources of uncertainty in future climate projections (Friedlingstein et al., 2013; IPCC, 2013). Global terrestrial ecosystems (sink) have kept pace with fossil fuel emissions (source), partially offsetting the continuous growth in anthropogenic CO₂ emissions (Le Quére et al., 2009; Liu et al., 2011; Sarmiento et al., 2010). Atmospheric CO₂ measurements indicate that half of the anthropogenic CO₂ emissions are offset by carbon sinks in the land and ocean (Ballantyne et al., 2012). However, allocating and partitioning carbon sinks and sources, in space and time, remain challenging (Schimel et al., 2015) due to our limited understanding of the mechanisms driving the carbon allocation, hence the climate variability (e.g., Friedlingstein et al., 2013).

Two main approaches, “bottom-up” and “top-down,” have been used to study the spatial distribution and temporal variations of CO₂ fluxes to understand and quantify the land carbon sinks and sources (e.g., Baker et al., 2006; Bousquet et al., 2000; Ciais et al., 2005; Crowell et al., 2019; Gurney et al., 2002; Huntzinger et al., 2013; Schwalm et al., 2015; Sitch et al., 2008). Bottom-up approaches, such as carbon flux predictions from terrestrial biosphere models and fossil fuel emission inventories, attempt to quantify carbon exchanges between atmosphere and land surface (e.g., Huntzinger et al., 2013; Schwalm et al., 2015; Sitch et al., 2008). Top-down approaches, referred to as atmospheric inversions, indirectly reconstruct surface CO₂ fluxes from observed spatial and temporal gradients in atmospheric CO₂ mole fractions (e.g., Baker et al., 2006; Bousquet et al., 2000; Ciais et al., 2005; Crowell et al., 2019; Gurney et al., 2002). Atmospheric inversions usually require an atmospheric transport model driven by prescribed surface fluxes (prior, usually from bottom-up approaches) and minimize the mismatch between modeled and observed CO₂ mole fractions to obtain the optimized flux estimates (posterior) in a Bayesian framework (Tarantola, 2005). Hence, the posterior fluxes are subject to the uncertainty from prior fluxes, atmospheric transport model, observations, and assumptions made during the optimization process (e.g., Crowell et al., 2019; Gourdjji et al., 2012; Lauvaux & Davis, 2014; Peylin et al., 2013; Sarmiento et al., 2010). Sarmiento et al. (2010) showed that large disagreement exists in the land and ocean flux estimates between top-down and bottom-up approaches. Thanks to relatively dense observing network, the flux estimates consistently showed carbon uptake over the Northern Hemisphere extratropical land and ocean. The Second State of the Carbon Cycle Report (USGCRP, 2018) shows that atmospheric inversions are in closer agreement between the different systems compared to the spread in terrestrial biosphere models (TBMs; Chapter 2, Figure 2.5) and tend to agree now with the smaller net sink from TBMs (699 Tg C year⁻¹ (±12%) from inversions against 606 Tg C year⁻¹ (±75%) for TBMs). However, large uncertainties remain. Past inverse flux estimates vary from 0 to −1.5 Pg C/year over North America from 11 global inversion systems (Peylin et al., 2013), in spite of having the densest surface CO₂ observing network compared to the rest of the globe. Since these inverse results used different transport models, prior fluxes, and inversion techniques, it is difficult to identify the sources of uncertainty.

Previous efforts focused on evaluating the impact of atmospheric transport in global and regional inversions, such as the Atmospheric Tracer Transport Model Intercomparison (TransCom) Project (e.g., Baker et al., 2006; Gurney et al., 2002; Law et al., 2008; Peylin et al., 2013; Stephens et al., 2007). Errors in horizontal and vertical transport are both major contributors to the overall transport uncertainty. Lin and Gerbig (2005) reported that uncertainties in CO₂ due to advection errors can be up to 60%. Gerbig et al. (2008) and Kretschmer et al. (2012) demonstrated that approximately 30% of the regional biosphere-atmosphere signals have been attributed to errors in the planetary boundary layer (PBL) height over Europe during summer. The sensitivity of simulated CO₂ mole fractions to the model physics, especially PBL schemes and land surface models (LSMs), represents a significant fraction of the observed model-data mismatch (Feng et al., 2016; Kretschmer et al., 2014; Lauvaux & Davis, 2014). Díaz-Isaac et al. (2018a) also examined the sensitivity of the modeled atmospheric CO₂ mole fractions to the model physics, including LSMs, PBL schemes, cumulus schemes, and microphysics, and meteorological reanalysis data. Their results showed that the modeled PBL CO₂ mole fractions are most sensitive to LSMs and PBL parameterizations, though other components of the model such as the meteorological driver data and the cumulus parameterizations are nonnegligible contributing to 10–20% of the model errors. Varying either PBL or LSM parameterizations led to ~4 ppm variability in modeled midday, PBL CO₂ mole fractions. Berner et al. (2011) compared a multiphysics

(model parameterizations) approach, stochastic perturbations, and a combination of both regarding the representations of model uncertainty in a mesoscale ensemble system. The results showed that the combination of a multiphysics scheme and stochastic perturbations provides a more accurate representation of model transport errors.

In contrast, a limited number of studies have been dedicated to investigate the impact of the flux uncertainty on simulated CO₂ mole fractions due to the lack of robust flux uncertainty estimates (Broquet et al., 2013). At continental and global scales, only biogenic flux uncertainties are considered in atmospheric inversions, while fossil fuel CO₂ emissions are assumed to be perfect (e.g., Baker et al., 2006; Basu et al., 2017; Crowell et al., 2019; Gurney et al., 2002; Liu et al., 2017). This assumption is inherited from the relatively low uncertainty in fossil fuel emissions at national scales (Andres et al., 2014). A priori biogenic fluxes are obtained from TBMs simulating ecosystem processes to produce estimates of carbon sources and sinks (e.g., Baker et al., 2010; Mao et al., 2012; Zeng et al., 2005). Due to climatic sensitivity, atmospheric conditions, nutrient and water availability, environmental driver data, soil conditions, and TBMs' estimates vary widely. To understand the land-atmosphere carbon exchange and feedbacks within the climate system, several efforts involving multiple TBM intercomparisons have been investigating the robustness of TBMs, such as the North American Carbon Program Regional and Continental Interim Synthesis activities (Huntzinger et al., 2012), the Trends in Net Land-Atmosphere Carbon Exchange (<http://dgvn.ceh.ac.uk/node/21>), the Large Scale Biosphere Atmosphere-Data Model Intercomparison Project (<http://www.climatemodeling.org/lbamip/>), the International Land-Atmosphere Benchmarking Project (<http://www.ilamb.org/>), and the Multi-Scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP, <https://nacp.ornl.gov/>). These efforts have shown that TBMs tend to show a wide range of annual flux estimates and responses to climate variability with large discrepancies in the spatial distribution (USGCRP, 2018).

In the past decade, regional inversions have reached higher spatial and temporal resolutions (Carouge et al., 2010; Lauvaux et al., 2012; Schuh et al., 2013). Schuh et al. (2013) compared inverse carbon estimates among mesoscale, continental scale, and global scale inversions during an intensive regional measurement campaign (Mid-Continent Intensive, Miles et al., 2012) over the highly productive agricultural regions of the midwestern United States. The inverse flux estimates of the regional total converged (Schuh et al., 2013) but with highly divergent spatial patterns, and the higher-resolution regional inverse flux estimates had better agreement with the gridded agricultural inventory (Ogle et al., 2015).

Regional inversions introduce another source of uncertainty compared to global-scale systems: the large-scale CO₂ boundary inflow acting as lateral boundary conditions. Inaccurate representation of boundary conditions propagates additional errors into the domain, leading to systematic biases in the optimized CO₂ flux estimates (Schuh et al., 2010). Göckede et al. (2010) applied constant offsets to boundary conditions taken from CarbonTracker and found that annually averaged offsets of 0.1 ppm could introduce 3.5 Tg C year⁻¹ (~10% of the net annual uptake) in the state of Oregon. Lauvaux et al. (2012) preprocessed the boundary conditions for each tower in their Corn Belt study and at a time using the surface tower measurement without an adjoint model and computed their associated uncertainty using independent aircraft measurements before the inversion. They found that incorrect boundary conditions could contribute half a parts per million bias to the simulations and a ± 24 Tg C errors in the 7-month flux estimates. Gourdji et al. (2012) compared two plausible sets of CO₂ boundary conditions: one taken from the CarbonTracker posterior CO₂ mole fractions and the other empirically derived from aircraft profiles and marine boundary layer data and found that they both lead to large differences in the flux estimates at annual continental scale. Schuh et al. (2013) implied that the correction of boundary conditions on the flux estimations can represent up to 15% of the annual carbon balance.

In both global and regional inversions, the magnitudes and structures of the uncertainties from each source component need to be prescribed which predetermines the sensitivity of the posterior fluxes to certain component(s). There have been no studies that directly estimate the uncertainty components in an explicit framework. Uncertainties are prescribed based on expert-level knowledge and statistical metrics to maximize the information content of atmospheric observations (Tarantola, 2005). Additionally, most inverse studies assume that the transport is unbiased and fossil fuel emissions are perfect, while several studies have shown the limits of these assumptions (Chevallier et al., 2010). In this study, we characterize the main uncertainty components explicitly using a regional forward-modeling ensemble framework, including atmospheric

transport, biogenic flux, and boundary condition uncertainties over North America. Uncertainties from fossil fuel emissions have not been fully characterized yet and will remain perfect in our study due to the lack of a rigorous assessment of uncertainties. We aim to build a robust ensemble system to reflect the model uncertainties with a combination of multiphysics scheme and stochastic perturbations, the biogenic flux ensemble with multiple net ecosystem exchange estimates from the MsTMIP TBMs, and a boundary condition ensemble from multiple posterior CO₂ mole fractions produced by global inversions. We validate our ensemble with meteorological, CO₂ mole fraction, and CO₂ flux measurements to validate the representativity of true model and flux uncertainties. The paper is organized as follows. The details of the ensemble framework are described in section 2, in addition to the description of data and evaluation methods. Section 3 describes the results from the evaluation and calibration procedures and temporal and spatial characteristics of the uncertainty components. The representativeness of the ensemble to model uncertainty is discussed in section 4, and the final conclusions are drawn in section 5.

2. Data and Methods

2.1. The Ensemble Modeling Framework

An ensemble-based mesoscale-modeling framework is implemented within our Weather Research and Forecasting model (WRF-CO₂, Lauvaux et al., 2012) derived from the original WRF-Chem model version 3.6.1 passive tracer mode to quantify the uncertainties in modeled atmospheric CO₂ mole fractions from transport, biogenic fluxes, and boundary conditions. The uncertainty is defined here as the root-mean-squared deviation over the members of a given ensemble suite, typically called *ensemble spread*. It is calculated as the standard deviation of modeled CO₂ mole fractions from the associated ensemble suite around the ensemble mean. Initially, the entire ensemble system includes a 10-member transport ensemble suite, an 18-member biogenic flux ensemble suite, and a four-member boundary condition ensemble suite. The chemistry module of the WRF-Chem model was modified to run 22 CO₂ tracers at a time in order to carry 18 biogenic flux members and four CO₂ boundary conditions as passive tracers in each transport run. Therefore, a 720-member ensemble suite (10 transport × 18 biogenic fluxes × 4 boundary conditions) is generated in this work. Further calibrations were applied to the flux ensemble suite to remove the bias and outliers in section 3.1.

The WRF-Chem model was run at 27- by 27-km grids covering most of North America (Figure 1a). The model was driven with the 6 hourly ECMWF Re-Analysis (ERA)-Interim product (Dee et al., 2011) with the resolution of ~80 km and 6-hourly sea surface temperature with a resolution of 12 km. The model configuration was inherited from the WRF-CMS framework developed at Penn State (Butler et al., 2019). It is worth noting that this ensemble framework assures mass conservation along the boundaries of the regional model domain as the globally modeled CO₂ mole fractions propagate into the domain. The details of this framework and mass conservation can be found in Butler et al. (2019).

Hourly simulation was conducted for 5-day periods between 1 December 2009 and 31 December 2010 with a 12-hr meteorological spin-up for each period. The first month of the simulations was considered as the CO₂ simulation spin-up and discarded in the final analysis. Finally, one year long hourly output for the entire year of 2010 was concatenated.

2.1.1. Transport Ensemble Suite

Our approach to build transport ensemble suite was varying meteorological initial and boundary conditions and model physics. We perturbed meteorological initial and boundary conditions based on a stochastic approach, instead of driving WRF-Chem with different reanalysis data. Stochastic perturbations introduce more diversity among the simulation results, producing more skillful ensemble forecasts, compared to physics-based ensembles. In this study, we used the stochastic kinetic energy backscatter schemes (SKEBS, Berner et al., 2009; Shutts, 2005) to create dynamic perturbations and modified the model physics schemes to create physics perturbations. SKEBS uses random stream function perturbations and temperature perturbations with a prescribed kinetic-energy spectrum to create flow-dependent perturbations (i.e. dynamical perturbations). Berner et al. (2009) found that the SKEBS improved probabilistic skill for the ECMWF medium-range forecasts up to 10 days using only a 10-member ensemble. A 10-member transport ensemble suite was therefore created using a combination of a multiphysics scheme and SKEBS. Both were implemented in the WRF-Chem model. A multiphysics scheme consists of choosing combinations of

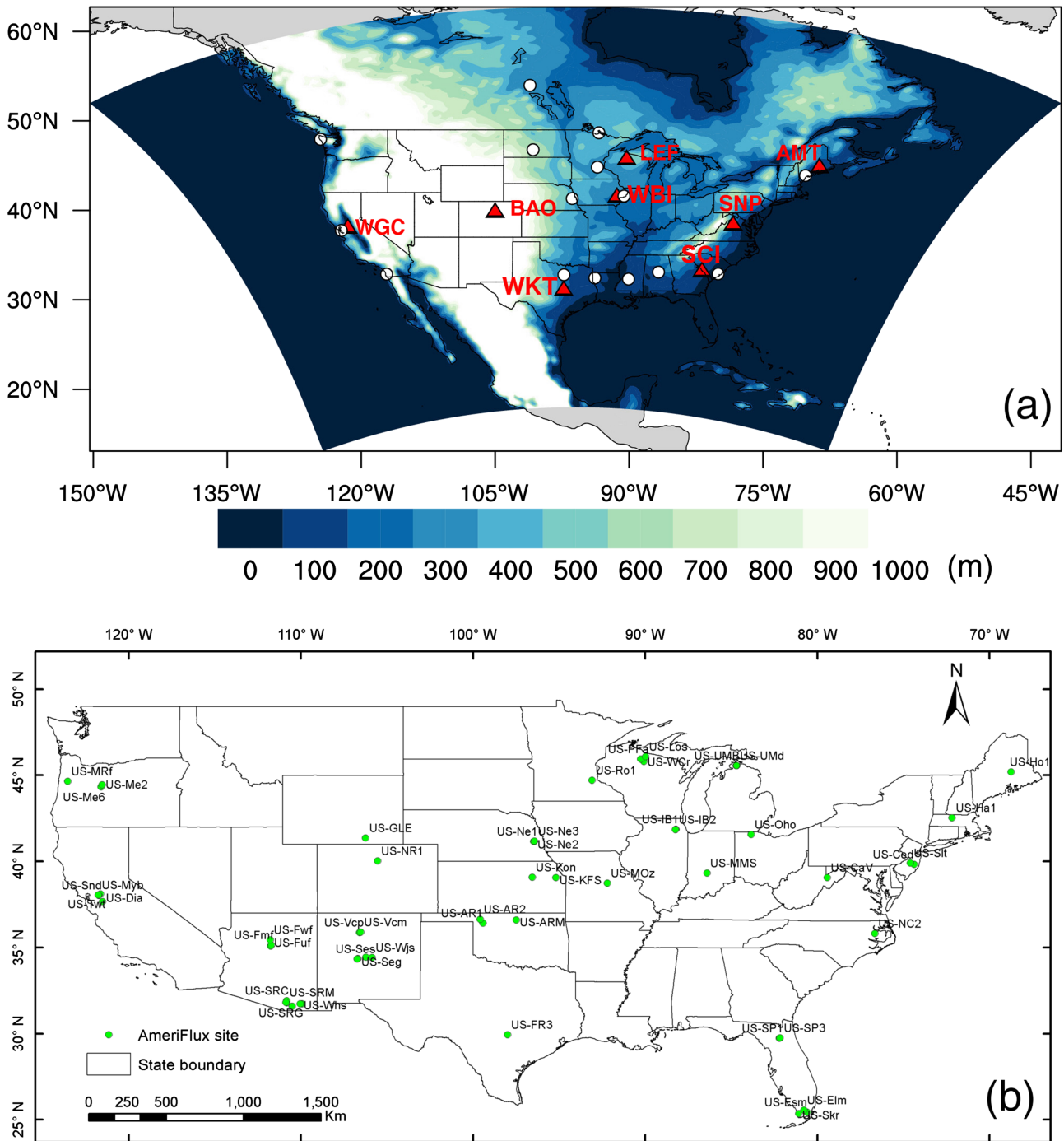


Figure 1. (a) The simulation domain and locations of the observation. Shaded contour is terrain height in meters. Red triangles denote the locations of the NOAA CO₂ towers used in this work, and the names of these towers can be found in Table 1. White dots denote the locations of the NOAA rawinsonde stations. Note that we removed WGC and BAO from the model calibration procedure due to the local contamination. (b) The locations of the AmeriFlux towers. Information describing these towers can be found in Table S1.

different WRF model physics. Combining these two methods likely cover transport uncertainty due to model physics and dynamics. Here we varied the land surface models (LSMs) and planetary boundary layer (PBL) schemes in the WRF-Chem model package based on the findings of Diaz-Isaac et al. (2018b), we chose three combinations: (1) Mellor-Yamada Nakanishi and Niino Level 2.5 (MYNN 2.5) PBL scheme with Noah LSM, (2) Mellor-Yamada-Janjic PBL scheme with RUC LSM, (3) Yonsei University PBL scheme with five-layer thermal diffusion LSM. In addition, three random dynamical perturbations were then applied to each

Table 1
Information of the NOAA CO₂ Towers Used

Code	Location	Intake height (m above ground level)	Surface type
AMT	Argyle, Maine	107	Forest
BAO	Erie, Colorado	300	Grassland/suburban
LEF	Park Falls, Wisconsin	396	Forest/wetland
SCT	Beech Island, South Carolina	305	Agriculture, residential, and industrial
SNP	Shenandoah National Park, Virginia	17	Mountain forest
WBI	West Branch, Iowa	379	Agriculture
WGC	Walnut Grove, California	483	Agriculture, residential, and industrial in the valley
WKT	Moody, Texas	457	Grassland pasture

physics scheme using a SKEBS. With inclusion of a baseline setup (MYNN2.5 and Noah LSM without a SKEBS), these 10 members constitute the transport ensemble suite.

2.1.2. Biogenic Flux Ensemble Suite

The biogenic fluxes ensemble suite comprises 3-hourly biogenic fluxes (net ecosystem exchange, NEE) from 15 TBMs in North American Carbon Program MsTMIP (Fisher et al., 2016; Huntzinger et al., 2013; Wei et al., 2014) including the mean fluxes of the MsTMIP models as an individual tracer, hourly biogenic fluxes from the Colorado State University (CSU) SiB3 model (Baker et al., 2008), and 3-hourly posterior biogenic fluxes from CarbonTracker (Peters et al., 2007) version 2016 (CT2016), for a total of 18 flux members. The list of the biogenic flux ensemble suite is shown in Table 2. The MsTMIP and CSU_SiB3 estimated fluxes fall in the bottom-up estimates, and the CarbonTracker posterior fluxes top-down estimates. All of the biogenic flux members were regridded from the native resolution to 27×27 km to feed in the WRF-Chem transport simulations.

MsTMIP includes a large suite of TBMs with the same simulation protocol in terms of model resolution, simulation period, and initial conditions. The goal of MsTMIP was to quantify the contribution of model structural differences to intermodel variability. These models vary in complexity and in their parameteriza-

tion of canopy conductance (energy and water fluxes), photosynthesis and respiration (carbon fluxes), allocation of carbon between soil and above (carbon pools), and the vegetation dynamics and disturbances. More details can be found in Huntzinger et al. (2013). Here we used the 3-hourly downscaling product of the MsTMIP model suite made available from Fisher et al. (2016) using the downscaling approach described in Olsen and Randerson (2004). The spatial resolution of the MsTMIP suite is $0.5^\circ \times 0.5^\circ$. We kept constant values in the 3-hr window when we generated hourly NEE from the MsTMIP suite.

SiB3 model fluxes (CSU_SiB3) were obtained by forcing the model with Modern-Era Retrospective Analysis (MERRA, Rienecker et al., 2011), and MERRA precipitation was scaled to Global Precipitation Climatology Project (Adler et al., 2003) amounts following the method described in Baker et al. (2010). Phenology was obtained from the MODIS LAI/fPAR product (Baker et al., 2008). CSU_SiB3 provides hourly NEE at $0.5^\circ \times 0.5^\circ$.

CarbonTracker is a global inversion system that assimilates atmospheric CO₂ mole fraction data from surface, tower, and aircraft measurements using a global Transport Model 5 (TM5) coupled to an ensemble Kalman filter (EnKF) optimization framework. Prior fluxes includes the CO₂ fluxes from fossil fuel burning, fires, terrestrial biosphere exchange, and exchanges with the oceans, of which fossil fuel and

Table 2
The Members of the Biogenic Flux Ensemble Suite

Model	Reference
MsTMIP	
BIOME_BGC	Thornton et al. (2002)
CLM	Mao et al. (2012)
CLM4VIC	Lei et al. (2014)
CLASS_CTEM ^a	Huang et al. (2011)
DLEM ^a	Tian et al. (2012)
GTEC	Ricciuto Daniel et al. (2011)
ISAM	Jain Atul and Yang (2005)
LPJ	Sitch et al. (2003)
ORCHIDEE	Krinner et al. (2005)
SIB3	Baker et al. (2008)
SIBCASA	Schaefer et al. (2008)
TEM6	Hayes et al. (2011)
TRIPLEX ^a	Peng et al. (2002)
VEGAS	Zeng et al. (2005)
VISIT	Ito (2010)
MEAN (of MsTMIP)	
CTBIO	Peters et al. (2007)
CSU_SIB3	Baker et al. (2008)

^aThe model has been removed from the final analysis.

Table 3
Configurations of Global Inversion Systems (Boundary Conditions)

Label	Transport model	Meteorological driver	Horizontal resolution	Temporal resolution	CO ₂ fluxes				Optimization technique
					Biosphere	Fossil fuel	Ocean	Fire	
CT2016	TM5	ERA- Interim	2° × 3°	3 hourly	CT-flux, optimized from CASA	CT/Miller and ODIAC	CT-flux, optimized from Takahashi	GFED4	ENKF
Basu	TM5	ERA- Interim	2° × 3°	3 hourly	CASA	EDGAR	Takahashi	GFED2	4D-Var
CMS	Geos-Chem	Geos-5	4° × 5°	3 hourly	CMS flux from CASA-GFED3	CDIAC	Darwin project	GFED3	L-BFGS
Schuh	Geo-Chem	ERA-Interim	2.5° × 2°	3 hourly	CASA	EDGAR	Takahashi	GFED2	Lag ENKF (5 weeks window)

fire emissions are fixed based on bottom-up estimates of their distribution and magnitude, whereas biogenic and oceanic fluxes are adjusted to match the atmospheric CO₂ data (Peters et al., 2007). Three-hourly posterior biogenic CO₂ fluxes at 1° × 1° were used and marked as CTBIO in the biogenic flux ensemble suite. We kept constant values for each hour in the 3-hr window when we generated hourly biogenic fluxes from CTBIO. A summary of the CarbonTracker system version 2016 (CT2016) can be found in the first row of Table 3.

2.1.3. Boundary Condition Ensemble Suite

We collected optimized CO₂ mole fractions from four global inversion systems to construct our boundary condition ensemble suite. The CO₂ mole fractions from global inversion systems used here correspond to optimized surface fluxes with offline transport generated by different general circulation models. The boundary condition ensemble suite includes CT2016 (Peters et al., 2007), a variational inversion using the TM5 (the same transport model as used in CT2016) as described in Basu et al. (2016), inverse fluxes from the Carbon Monitoring System (CMS) using the Goddard Earth Observing System (GEOS)-Chem model (Liu et al., 2014), and an EnKF optimized mole fractions using GEOS-Chem (Schuh et al., 2015). The two state-of-the-art global transport models used here (i.e., TM5 and GEOS-Chem) provide a wide range of boundary conditions representing transport uncertainties in addition to inverse flux differences at the global scale (Schuh et al., 2019). A summary of the four inversion systems can be found in Table 3.

2.1.4. The Construction of the Modeled Total Atmospheric CO₂ Mole Fractions and the Uncertainties

The modeled total CO₂ mole fractions comprise the sum of the tracers for biosphere, boundary conditions, fossil fuel and fire emissions, and oceanic fluxes for a given transport run. Fossil fuel emissions, fire, and oceanic fluxes were taken from CT2016 for all ensemble members. CT2016 obtains fossil fuel CO₂ emissions by aggregating the annual global total fossil fuel CO₂ emissions (Marland et al., 2008) spatially based on the Electronic Data Gathering, Analysis and Retrieval (EDGAR) inventories (<https://themasites.pbl.nl/tridion/en/themasites/edgar/index.html>) with an imposed seasonal cycle from Blasing et al. (2005), monthly fire CO₂ emissions from the Global Fire Emissions Database version v4.1s (GFED4, Giglio et al., 2006; van der Werf et al., 2006), and oceanic CO₂ fluxes from Takahashi et al. (2002).

We constructed the modeled total atmospheric CO₂ mole fractions for a given ensemble suite by varying the associated CO₂ tracer variables and having other tracers fixed. For example, the tracers of the biogenic flux ensemble suite have the same boundary condition tracer and transport but vary biogenic flux tracers; the transport suite has the same boundary condition and flux tracers but varies transport simulations; the boundary suite has the same transport and fluxes but

Table 4
Annual Bias and Mean Absolute Error (MAE) in Modeled CO₂ Mole Fractions Among the Members of the Biogenic Flux Ensemble Suite to the NOAA CO₂ Tower Measurements (See Details in Table 1) on the Monthly Timescale for the Year of 2010

Model	Bias (ppm)	MAE (ppm)
ISAM	0.89	3.62
CSU_SIB3	0.79	2.00
SIB3	0.66	4.17
ORCHIDEE	0.60	2.83
BIOME_BGC	0.30	3.25
VEGAS	0.19	1.80
CLM4VIC	-0.09	2.33
CTBIO	-0.10	1.35
DLEM	-0.14	2.18
TEM6	-0.20	1.66
MEAN	-0.23	1.85
GTEC	-0.30	6.65
LPJ	-0.39	3.74
SIBCASA	-0.41	3.19
VISIT	-0.62	1.80
TRIPLEX	-1.69	4.25
CLASS_CTEM	-1.87	4.71

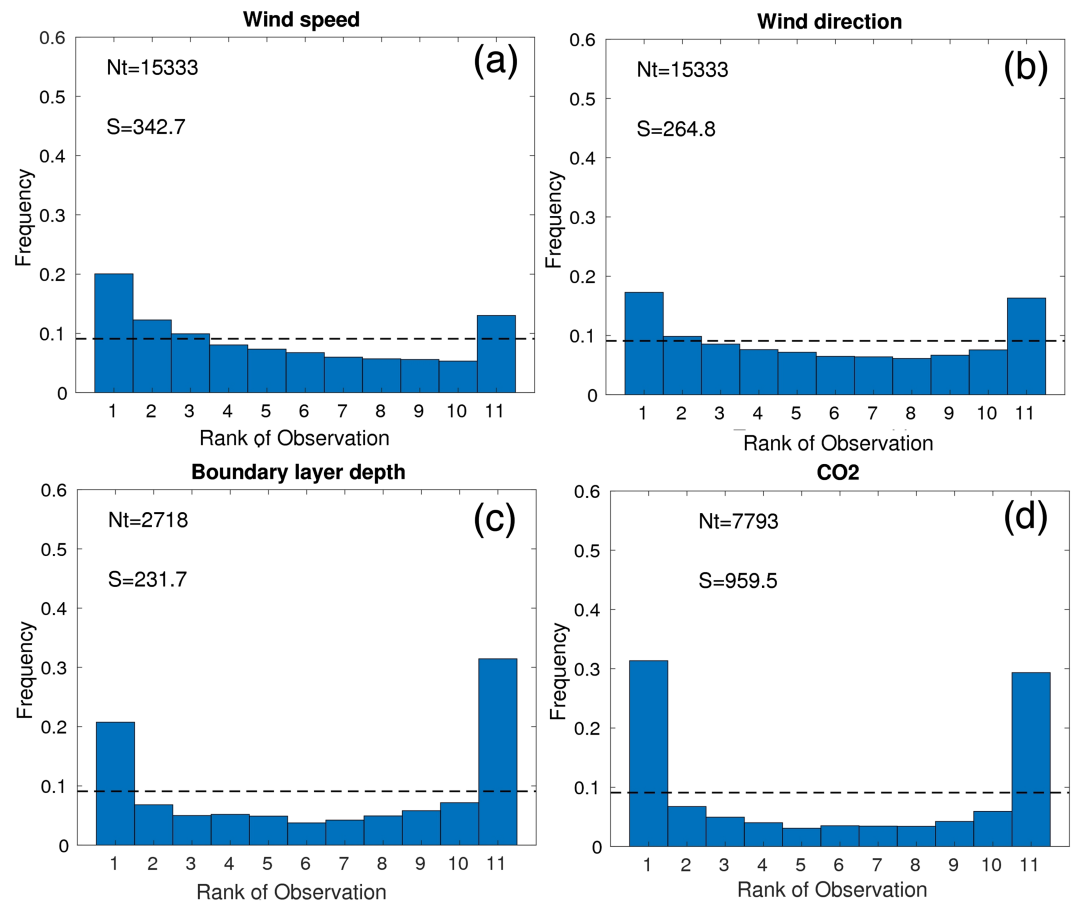


Figure 2. Rank histograms of the transport ensemble suite for 0 UTC (a) 925 hPa wind speed, (b) 925 hPa wind direction, (c) PBL depth, and (d) midday average PBL CO₂ mole fractions. Dashed lines denote the ideal values of the rank histograms. N_t denotes the total number of the observations used. S denotes the rank histogram score of the ensemble. The dashed line is the ideal frequency of a flat distribution.

varies boundary condition tracers. We selected the CT2016 CO₂ mole fractions, the CT2016 posterior biogenic fluxes, MYNN+Noah LSM without SKEBS as the reference of boundary conditions, biogenic fluxes, and transport, respectively. The selection of CT2016 as the reference is owing to the reason that it is commonly applied by the community. Additionally, the evaluation of biospheric flux members in section 3.1.2 illustrates that CT2016 is fairly reliable.

The root-mean-square difference of the ensemble members from the ensemble mean is used to quantify the uncertainty in modeled CO₂ mole fractions from the given component. This definition stands if the ensemble is unbiased, which is not necessarily true for biospheric flux ensemble suite. We therefore will calibrate biospheric flux members in detailed manners and discard biased outliers from the final analysis. Details can be founded in section 3.1.2.

Note that we can only implement surface fluxes and boundary conditions as independent tracers in each transport run. Transport is coupled with individual tracers and cannot be separated from other components; transport always scales with a given tracers. Because of this reason, the contribution of the transport component to total atmospheric CO₂ mole fractions is invalid and not presentable, unlike the contribution from biosphere and boundary condition shown in Figures 8 and 9.

2.2. Data

The observations we used to evaluate the robustness of the ensemble spread include the PBL winds and depths, atmospheric CO₂ mole fractions, and CO₂ surface fluxes. We obtained meteorological wind data

from the National Oceanic and Atmospheric Administration (NOAA) rawinsonde stations (<https://ruc.noaa.gov/raobs/>; see locations in Figure 1) and calculated the PBL depths from the same data set using the virtual potential temperature gradient approach. We obtained the in situ CO₂ data (Andrews et al., 2014) from the NOAA ObsPack GlobalViewPlus product (Cooperative Global Atmospheric Data Integration, 2018) (see locations in Figure 1 and information in Table 1). Only the measurements of eight towers from this data package were selected for investigations. ObsPack data products collect greenhouse gas data from providers over the globe and reformat the data into the ObsPack framework to simulate and support carbon cycle modeling studies (Masarie et al., 2014). Another set of data we used is CO₂ flux measurements from the AmeriFlux network (<https://ameriflux.lbl.gov>). We included 65 sites in our analysis. The locations and information of the sites can be found in Figure 1b and Table 2, respectively.

2.3. Evaluation and Calibration Methods

The rank histogram, also known as Talagrand diagram (e.g., Anderson, 1996; Hamill, 1997; Hamill & Colucci, 1997; Talagrand et al., 1997), has been widely used to evaluate the reliability of an ensemble, namely, the representativeness of the ensemble to the model uncertainties. Details about the rank histogram and its applications can be found in Hamill et al. (2001). Briefly, the rank histogram is the probability density function of the observation data points among the corresponding modeled values created by sorting the observation and models increasingly for a given data point. It takes space and time into account. Ideally, the rank histogram of a robust ensemble should represent the equiprobability of ranked observations among models and hence be characterized by a flat distribution with an associated score equal to 1 by definition. However, most ensembles tend to be underdispersive and underestimate the true uncertainty of the atmospheric evolution (Buizza et al., 2005), namely, the ensemble spread is too narrow. In this case, a considerable number of observations fall into the outer bins, which leads to “U-shaped” rank histograms with scores greater than 1. Díaz-Isaac et al. (2018a) used rank histograms to define the optimal selection of members among 45 to produce the most accurate representation of transport model errors. Here we use a simpler approach by starting with a small-size ensemble and gradually adding members to the ensemble until the flatness of the histogram is acceptable. Typically, underdispersive ensembles show large scores, as described in Garaud and Mallet (2011). We rely on previous studies to define the acceptance level in terms of flatness of the histogram (cf. section 3). Note that the rank histogram analyses are applied to modeled PBL wind speed, direction, and height, in addition to CO₂ mole fractions.

3. Results

3.1. Evaluation and Calibration of the Ensemble System

3.1.1. Evaluation of the Transport Ensemble Suite

A reliable ensemble system implies that the ensemble spread represents the actual model uncertainty. To evaluate the reliability of our ensemble, we performed ten different transport simulations to produce a reasonably low flatness score. Figures 2a–2c shows the rank histograms of the transport ensemble suite for PBL wind speed, wind direction, and depth. The evaluation of the transport ensemble suite is solely based on meteorological data. The bins on both ends of the rank histograms represent the frequency of observations falling outside the envelope of the ensemble. Statistically, observations should fall outside the ensemble range at the same frequency as other ranks (or bins) if the ensemble is representative of the model uncertainty. For this given transport ensemble we built, the two outer bins are 2 times greater than the rest of the bins for wind speed and wind direction and a factor of 3 for PBL depth. It also appears that this transport model ensemble tends to overestimate wind speed and direction and underestimate PBL depths. The rank histograms have an “L” shape. We acknowledge that the transport ensemble suite is slightly biased and underdispersive. However, the flatness scores remain low (100–340, compared to 148 with the 101-member ensemble in Garaud & Mallet, 2011) and acceptable to investigate further the model uncertainties based on the spread of the ensemble system. In practice, a small ensemble tends to be underdispersive. Existing operational systems attempt to increase the dispersion by increasing the number of ensemble members. Previous studies from a large number of ensemble members show that the two outer bins are usually a factor of 3 greater than the ideal distribution even for a large ensemble (e. g., 50 members in Berner et al., 2009; 101 members in Garaud & Mallet, 2011; 45 member in Díaz-Isaac et al., 2018b). Lauvaux et al. (2019) showed that small size ensembles are sufficient to produce reliable statistics except spurious structures in space. We conclude that the transport ensemble suite is fairly representative and skillful despite our small

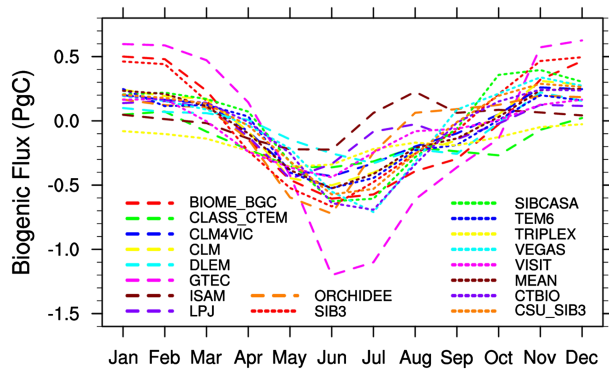


Figure 3. Monthly means of the biogenic flux ensemble members over North America for the year 2010.

number of members, but clearly not sufficient to represent the full spatiotemporal structures of model errors. Here we focus on the model uncertainty which can be represented by a smaller number of calibrated members.

3.1.2. Calibration of the Biogenic Flux Ensemble Suite

Initially, our biogenic flux ensemble suite consists of 18 members including TBMs and inverse fluxes. Figures S1 and S2 show the monthly mean of all of biogenic flux members in January and July. In winter (January, Figure S1 in the supporting information), most of biogenic flux members show positive CO₂ fluxes over North America with a few exceptions. CLASS_GTEM and TRIPLEX show strong CO₂ uptake in the Midwest and South, respectively. In summer (July, Figure S2), most biogenic flux members show strong CO₂ uptake over North America except ISAM, LPJ, ORCHIDEE, and TRIPLEX. The spatial patterns are highly variable. We notice

that the summertime strong CO₂ uptake over the U.S. Midwest, corresponding to the large agricultural area of the Corn Belt, is only visible in GTEC, SiB3CASA, TEM5, VEGAS, MEAN, CTBIO, and CSU_SiB3. Some of the members, such as CLASS_CTEM, ISAM, ORCHIDEE, LPJ, and TRIPLEX show positive CO₂ fluxes over the Corn Belt in summer. Previous studies have reported the largest uptake in North America from fast-growing crops such as corn and soybean during summertime (Lokupitiya et al., 2016; Ogle et al., 2015; West et al., 2011). When summing these highly varying biogenic flux maps over North America, most MsTMIP models show similar monthly variations including amplitudes and phases, except GTEC with a larger seasonal amplitude, ISAM with a short period of net carbon uptake, and Biome-BGC and SiB3 with strong winter respiration (Figure 3).

We propagated these surface biogenic fluxes into modeled atmospheric CO₂ mole fractions and evaluated biogenic flux members in the flux and mole fraction spaces. We used Taylor diagrams to present the comparisons of the monthly means of the biogenic flux members with flux measurements (Figure 4a) and the associated modeled atmospheric CO₂ mole fractions with the NOAA CO₂ towers (Figure 4b). The skill of MsTMIP members has similar rankings in both flux and mole fraction spaces. For example, in both spaces, GTEC remains outside the range of the ensemble, and CTBIO and VEGAS always outperform other

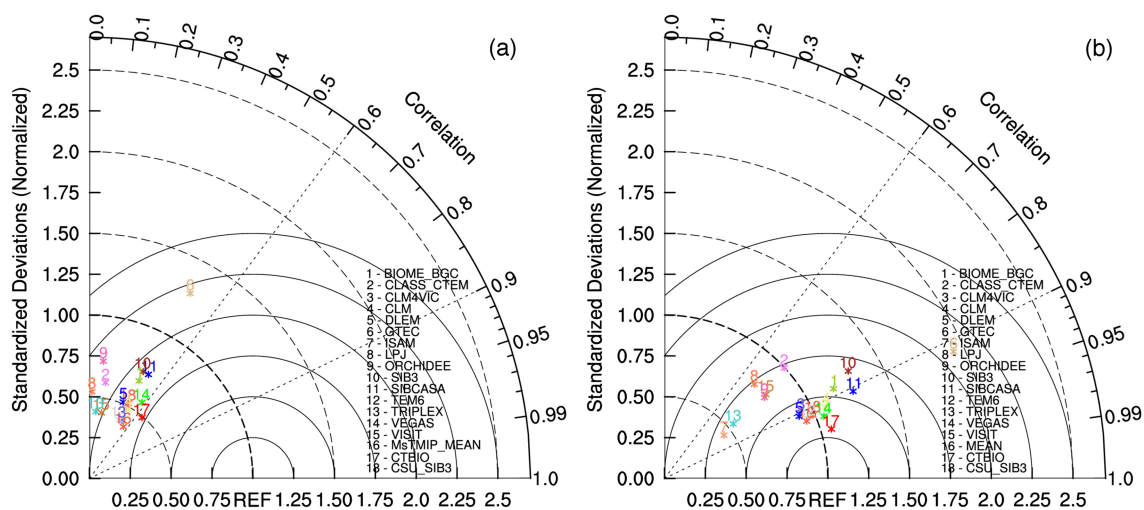


Figure 4. Taylor diagrams of the biogenic flux ensemble suite in the space of (a) fluxes and (b) mole fractions. The standard deviations of models are normalized by the standard deviations of observations (“REF”). The observations used in (a) are from the AmeriFlux flux tower measurements. The locations and information can be found in Figure 1b and Table S1. Modeled fluxes are determined by the value of the nearest grid cell to a flux tower. All hour data are used to calculate monthly averaged CO₂ fluxes with the requirement of more than 30% of data availability at valid AmeriFlux sites. The observations used in (b) are from the NOAA CO₂ tall tower measurements. The locations and information can be found in Figure 1a and Table 1. Modeled CO₂ mole fractions are determined by the value of the nearest grid cell to a concentration tower. Daytime hours are used to calculate monthly averaged CO₂ mole fractions.

simulations for the closest distance to the observations (“REF” in Figure 4). GTEC seems an outlier in seasonal cycle (Figure 3). The consistent rankings in both spaces suggest that biogenic fluxes drive the variations of atmospheric CO₂ mole fractions at the monthly timescale despite atmospheric transport and boundary conditions, implying larger uncertainty from biogenic fluxes than from transport and/or boundary conditions.

While the flux and mole fraction spaces exhibit qualitatively consistent features across the biogenic flux members, they differ quantitatively. First, the biogenic flux members in mole fraction space display higher correlations than in flux space. The correlation coefficients range from ~0.55 to ~0.95 in mole fraction space (Figure 4b), while from ~0 to ~0.6 in flux space (Figure 4a). Second, the variability over time in mole fraction space is more consistent between model and observation than in flux space. The normalized standard deviations in mole fraction space range from 0.75 to 1.25 (except that GTEC is 1.85, ISAM and TRIPLEX ~0.5; Figure 4b) while ranging from 0.25 to 0.75 (except that GTEC is 1.25) in flux space (Figure 4a). As a result, the mole fraction space comparisons have smaller root-mean-square errors than the flux space comparison. The differences between the two spaces are likely related to three reasons. The first reason is due to the much larger footprint sizes of the mole fraction tower measurements as compared to flux tower measurements. The typical footprint size of a CO₂ mole fraction tower is about 10⁶–10⁸ km² (e.g., Gloor et al., 2001; Sweeney et al., 2015), while a flux tower is about 1 km² (e.g., Costa-e-Silva et al., 2015; McCaughey et al., 2006; Nagler et al., 2005). Flux towers measure surface fluxes locally. The atmosphere, as an integrator, mixes the CO₂ signals across a wider spatial domain. Second, the hemispheric background plays the major role in the seasonal variations of atmospheric CO₂ mole fractions, and the transport model captures this correlation very well setting a good baseline of the atmospheric CO₂ mole fractions. Third may cause by different model resolutions in two spaces: CTBIO is at 1°×1° and the rest TBMs are at 0.5°×0.5° in flux space; all modeled CO₂ mole fractions are at 27 × 27 km.

We also examined the annual and monthly averaged model bias across all biogenic flux ensemble members (Table 4). Note that all of the modeled atmospheric total CO₂ mole fractions have the same transport and boundary conditions with various biogenic flux members. Given –0.1 ppm of annual bias and 1.31 ppm of monthly averaged mean absolute error (MAE) in CTBIO, biogenic fluxes predominate the biases in the annual CO₂ mole fraction shown in Table 4. The model annual biases appear smaller than the monthly biases overall, usually 1 order of magnitude smaller expect for TRIPLEX and CLASS_CTEM. TRIPLEX and CLASS_CTEM also produce much larger negative biases over the continent, while most TBMs have annual biases within 1 ppm. It appears that TRIPLEX and CLASS_CTEM are outliers in net annual fluxes. Consistent with Figures 3 and 4, GTEC holds largest bias on the monthly timescale, while most of the biogenic members show monthly averaged MAE below 4 ppm. To remain statistically coherent to produce quasi-normal statistics, we remove GTEC, TRIPLEX, and CLASS_CTEM from the biogenic flux members for the final uncertainty assessment hereafter. The remaining 15 biogenic flux members compose the calibrated biogenic flux ensemble suite and will be used to calculate the simulated CO₂ uncertainty from biogenic fluxes hereafter. Figure 5a illustrates the rank histogram of the calibrated biogenic flux ensemble suite for the atmospheric CO₂ mole fractions. The nearly flat histogram implies that our ensemble for fluxes is dispersive enough to represent flux errors in CO₂ mole fractions. The flatness also suggests that biogenic fluxes dominate the model uncertainty, since the transport errors would lead to the ensemble that is under-dispersive (Figure 2d). With the inclusion of transport ensemble suite (Figure 5b), the rank histogram is even flatter, and the score reduce from 20.2 to 1.8, suggesting that a properly dispersive ensemble has been constructed.

3.1.3. Inclusion of the Boundary Condition Uncertainty and Evaluation

Due to the limited number of ensemble members we collected from four global chemical transport models, we used a different approach to add the uncertainty from CO₂ boundary conditions into the total modeled CO₂ uncertainty. Only the uncertainties at monthly timescale from boundary conditions are included in the evaluation. We defined it as the root-mean-square deviations of the monthly averaged CO₂ mole fractions across boundary conditions from their ensemble means. Hence, the synoptic scale variability is not incorporated but the seasonal variability of large-scale CO₂ meridional gradients is. This approach is commonly used to introduce measurement errors into rank histograms because observations are all unique and hence no ensemble is available (Saetra et al., 2004). The inclusion of the boundary condition uncertainty reduces the score from 1.8 to 1.3 (Figure 5b versus Figure 5c). Our final ensemble that embraces the model

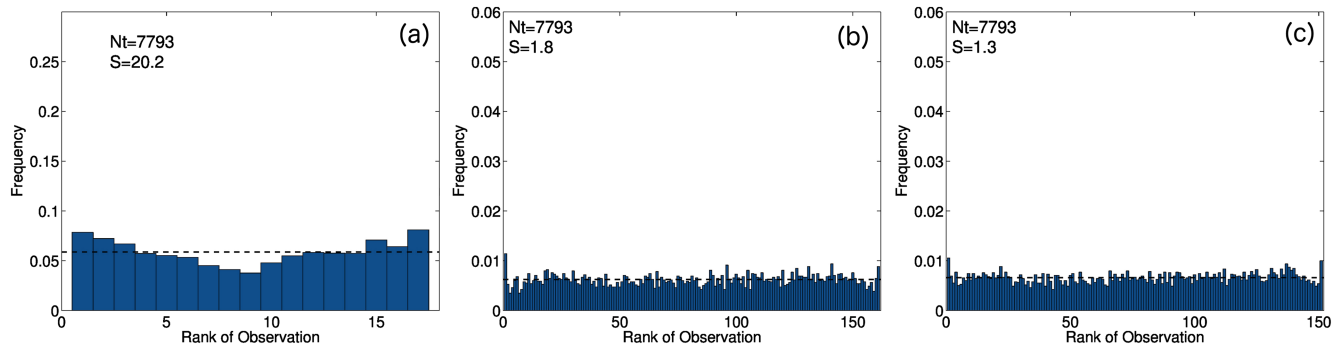


Figure 5. Rank histograms of (a) calibrated biogenic flux ensemble suite, (b) the combination of the calibrated biogenic flux and transport ensemble suites, and (c) the full ensemble with the inclusion of the uncertainties of transport, biogenic fluxes, and boundary conditions for the atmospheric CO₂ mole fractions. Dashed lines denote the ideal values of the rank histograms. N_t denotes the total number of the observations used from six NOAA CO₂ tower sites. The locations and information can be found in Figure 1a and Table 1. Data points represent individual hourly mean values, and the histograms are aggregated over all sites and daytime hours of the entire year. S denotes the rank histogram score of the ensemble.

uncertainties from transport, biogenic fluxes, and boundary conditions has excellent flatness scores, and further analysis will use the calibrated ensembles. See Figure 6 for a summary of the evaluation and calibration processes. We acknowledge here that an ideal ensemble should include additional boundary conditions and possibly transport members, but we conclude that our current ensemble shows sufficient dispersion to approach the true model uncertainties.

3.2. Seasonal Variations of Model Uncertainties

3.2.1. Comparisons with In Situ CO₂ Measurements

Figure 7 illustrates the monthly means of the observed and modeled atmospheric CO₂ mole fractions at the CO₂ tower locations we selected. Note that modeled CO₂ mole fractions were calculated from the ensemble means. Due to the contamination from local sources (in the Central Valley, California, and the Front Range, Colorado), we removed two towers from the final comparison and listed them in Figure S3. The overall features of the atmospheric CO₂ mole fractions from the observations follow the typical seasonal cycle, low in summer and high in winter. The amplitudes of the seasonal variations change from site to site. It appears that the northern sites (AMT, LEF, SNP, and WBI) are drawn down to much lower mole fractions (~375 ppm) in summertime than the southern sites (SCT and WKT, ~387 ppm), while the winter signals remain at the similar level of ~400 ppm across sites.

The comparisons of the spread across the ensemble suites suggest that the total model uncertainty is largely attributed to biogenic flux uncertainty, consistent with findings in section 3.1. The boundary condition and transport uncertainties have much smaller values for each month at the tower locations.

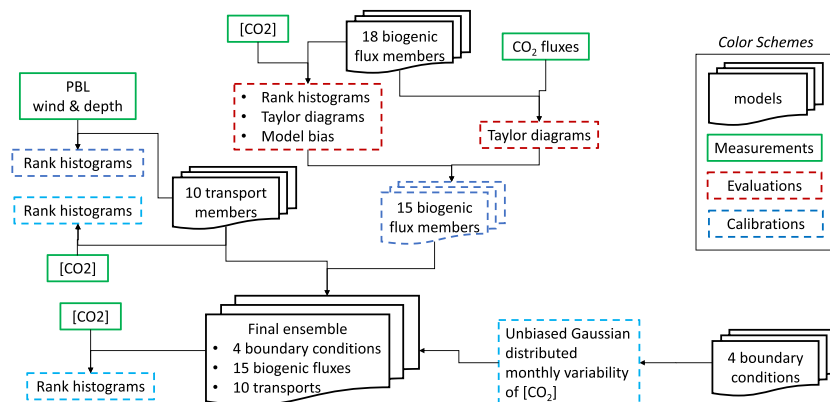


Figure 6. Flow chart of the summary of the ensemble evaluation and calibration processes.

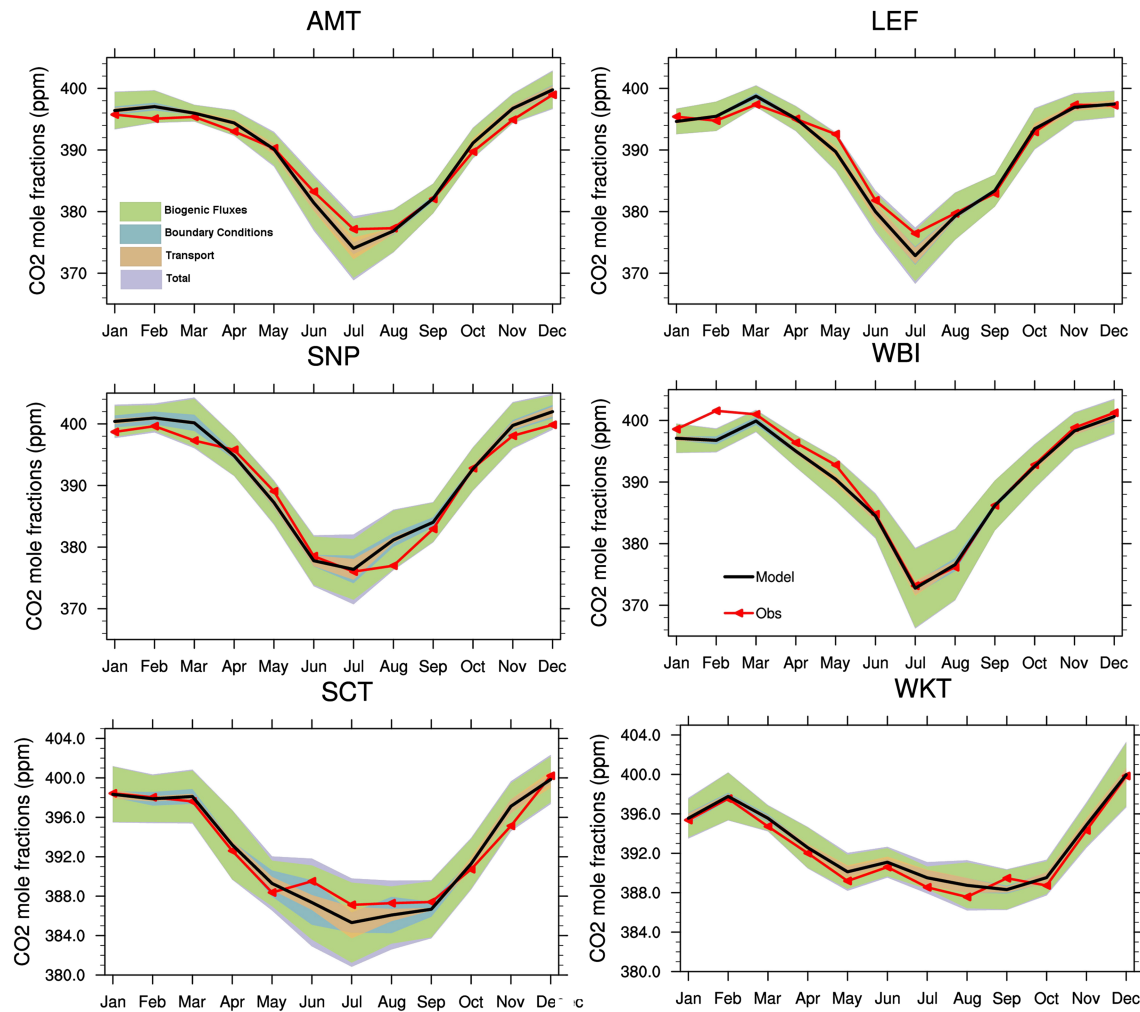


Figure 7. Monthly mean of the modeled and observed midday CO₂ mole fraction at the NOAA CO₂ towers. The locations and information can be found in Figure 1a and Table 1. Red lines denote the observations, and black lines the modeled means for all of the calibrated ensemble members. Shaded areas denote the ensemble spreads of the calibrated full, biogenic flux, transport, and boundary condition ensembles.

3.2.2. Spatiotemporal Variability of Model Uncertainties

Figure 8 shows the spatial distributions of the ensemble means for each component in January and July. In such a setup, the mean contribution of transport to the atmospheric CO₂ is invalid, but the contribution of transport to uncertainty in atmospheric CO₂ can be represented by the transport ensemble spread. For simplicity, the modeled CO₂ mole fractions at the fifth level, ~550 m above ground level, and 20 UTC are used to represent the well-mixed midday atmospheric boundary layer conditions. The convention of using the CO₂ mole fraction data in well-mixed midday atmospheric boundary layer conditions follows the common practice in atmospheric inversions.

The biospheric uptake causes a drawdown in summer of more than 12 ppm in the PBL across agricultural areas, like the U.S. Corn Belt in July (Figure 8d), and a net increase of more than 6 ppm of CO₂ during winter (Figure 8c). Boundary conditions are the result of large-scale atmospheric circulation (Figures 8a and 8b). Seasonal patterns depend on the locations of the jet stream over North America combined with large-scale CO₂ meridional gradients. Interestingly, the spatial patterns of modeled column-averaged CO₂ (XCO₂) are very similar to the PBL CO₂ but display a smaller magnitude and gradient (Figure 9), suggesting the free troposphere and stratosphere generally have little horizontal structure and mainly serve to dilute the patterns present in the PBL.

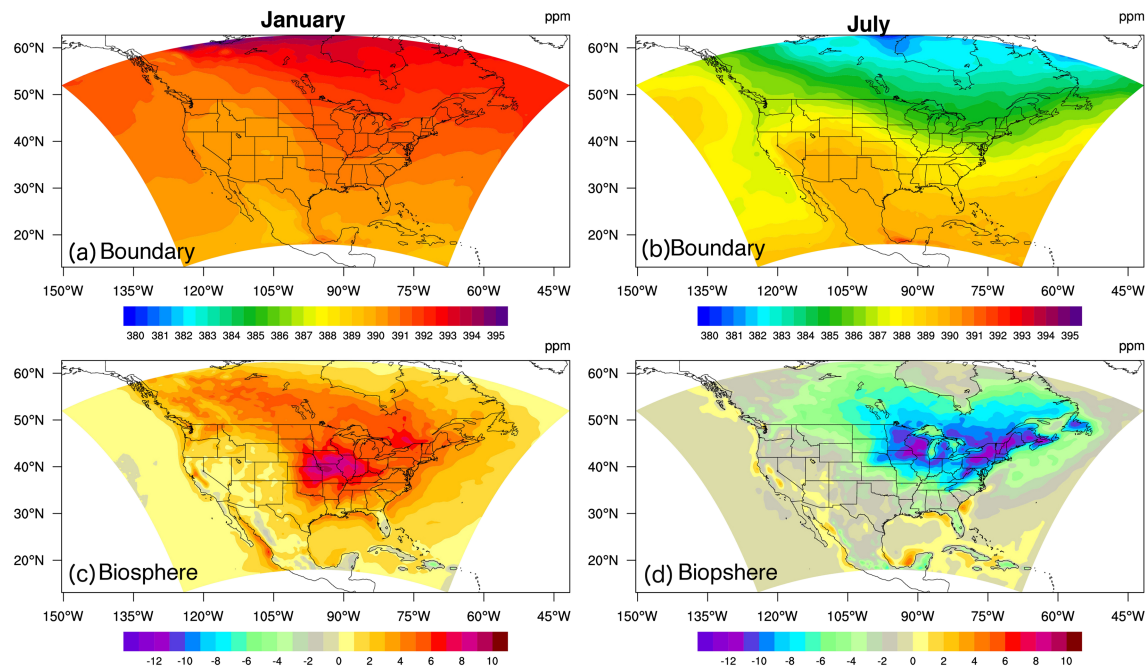


Figure 8. Monthly mean of the modeled atmospheric boundary layer CO₂ mole fractions for (a) and (b) boundary condition components and (c) and (d) the biogenic flux components. (a) and (c) The monthly means at fifth model level for January 2010 at 20 UTC and (b) and (d) for July.

Figure 10 displays the spatial distributions of the ensemble spreads for each component in January and July. Spatial gradients of the transport uncertainties are remarkably similar to the biogenic flux ensemble means (Figure 10e versus Figure 8c for January; Figure 10f versus Figure 8d for July), as the atmospheric CO₂ mole fractions scale directly with surface fluxes. The pattern correlations of the absolute transport uncertainties and biogenic flux ensemble means are 0.65 for winter and 0.81 for summer, respectively. Areas with larger surface flux magnitudes will generate large variations once propagated to the atmosphere. Transport uncertainty averaged over North America as a whole is 1.3 ppm in January (winter) and 2.8 in July (summer) (see

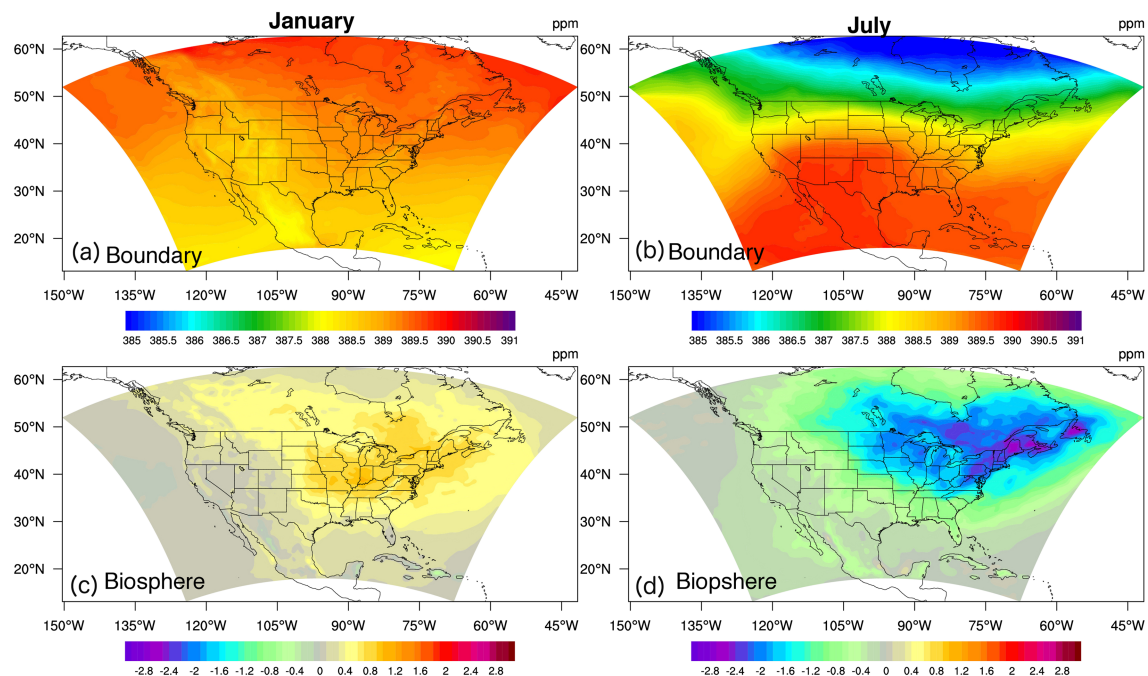


Figure 9. Same as Figure 8 but for column-integrated CO₂ (XCO₂).

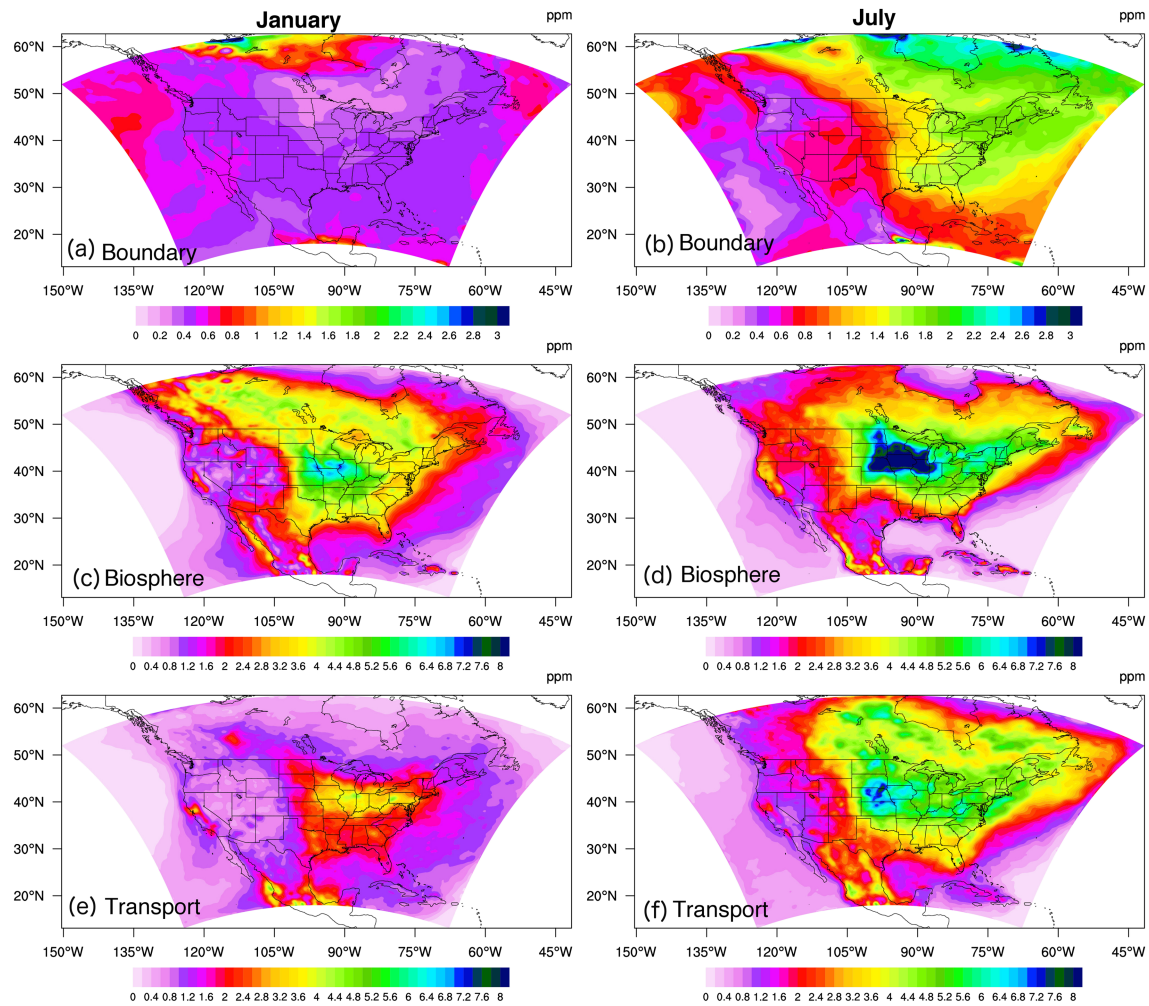


Figure 10. Root-mean-square difference from the ensemble monthly mean of the modeled surface atmospheric boundary layer CO₂ mole fractions for the (a and b) boundary conditions, (c and d) biogenic flux, and (e and f) transport components. (a, c, and e) The fifth model level for January 2010 at 20 UTC and (b, d, and f) for July. Note that the color scales for (a and b) boundary conditions differ from the others.

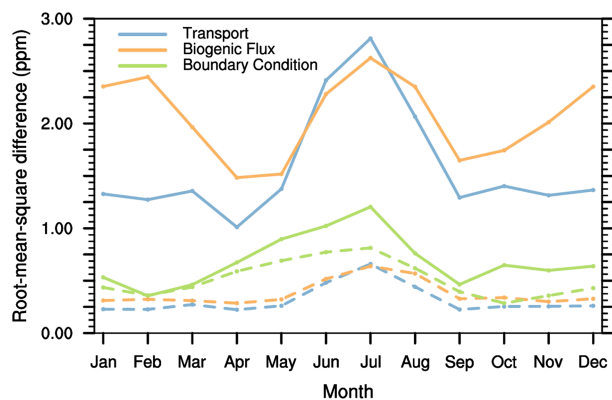


Figure 11. Monthly aggregated root-mean-square difference (uncertainty) of the modeled PBL CO₂ (solid lines) and XCO₂ (dashed lines) from transport, biogenic fluxes, and boundary conditions over North America.

Figure 11). In section 4, we will discuss on the implications of confounding error structures to disaggregate flux and transport errors in atmospheric inversions.

In winter, the biogenic flux uncertainty is ~4 ppm over large regions due to relatively small biogenic CO₂ signals from soil and plant respiration (Figure 10c). In summer, the uncertainty increases significantly up to 8 ppm in the U.S. Corn Belt (Figure 10d) but remains relatively smaller in Canada's boreal forest area, with ~3.5 ppm in winter and ~4 ppm in summer (Figures 10c and 10d). The large gradients in the biogenic flux uncertainty exist across the western and eastern United States over the Great Plains for both seasons. In general, areas with large biogenic flux uncertainties (Figures 10c and 10d) correspond to strong biogenic CO₂ mole fractions (Figures 8c and 8d), confirming that uncertainties scale with the mean flux magnitude, as often assumed in regional inversions (Lauvaux et al., 2012). On average, the biogenic flux uncertainty is 2.4 ppm in winter and 2.6 ppm in summer (Figure 11), with low uncertainties in spring and fall

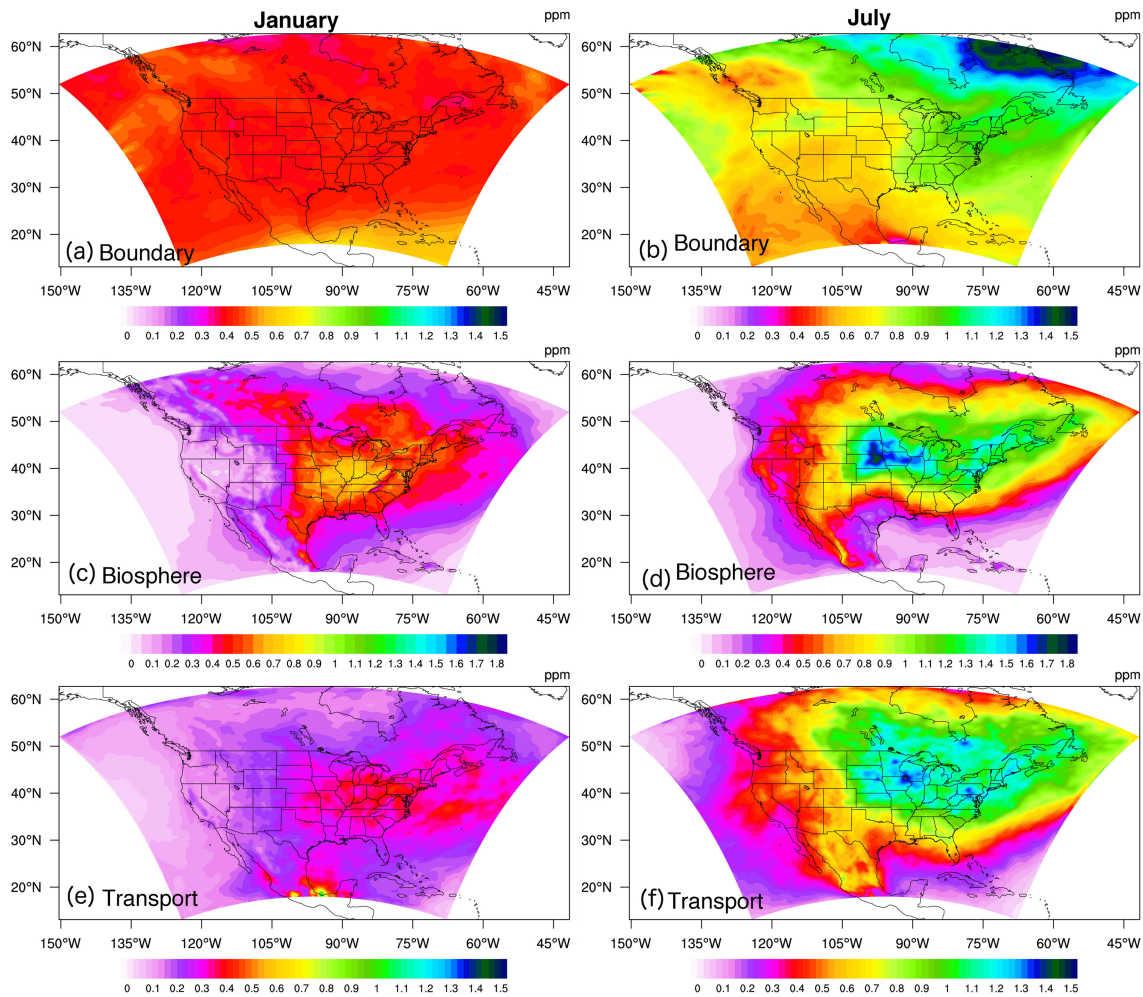


Figure 12. Same as Figure 10 but for column-integrated CO_2 (XCO_2).

(1.5 ppm and 1.6 ppm, respectively). The magnitudes of biogenic flux uncertainties are very similar in mole fraction space, while they are not in flux space (0.16 Pg C for winter and 0.26 Pg C for summer). The shallower PBL heights in winter may amplify the uncertainty in biogenic CO_2 mole fractions.

Boundary condition uncertainty is overall the smallest in PBL for both seasons (Figures 10a and 10b). In winter, due to the reduced biogenic CO_2 fluxes in the Northern Hemisphere, the difference between global modeled CO_2 mole fractions is small (less than 1 ppm). The four different boundary CO_2 signals propagating into the model domain create 0.5 ppm uncertainty over North America. In summer, the uncertainty increases to ~ 0.8 ppm in the western United States and ~ 2 ppm elsewhere (1.2 ppm on average). The largest spatial gradients in the boundary condition uncertainty appear over the Great Plains. It is worth mentioning that the boundary condition uncertainty is as important as other components near the western boundary of the model domain and even more important near the northern boundary, indicating the importance of boundary condition errors where the inflow variability is prevalent.

Figure 12 shows the monthly averaged uncertainty maps for the XCO_2 counterpart. The uncertainties in the PBL CO_2 are propagated into the whole column and create similar spatial patterns but with much smaller magnitudes, with the exception of boundary condition uncertainty (Figures 12a and 12b). Both transport and flux uncertainties decrease by 75% in the column, suggesting the uncertainties from these two sources are mainly in the PBL or lower troposphere.

The uncertainties from boundary conditions are more persistent throughout the whole column (Figure 11). Both the PBL and column-averaged uncertainties are approximately 0.5 ppm in winter. In summer,

uncertainty in XCO_2 due to boundary conditions is approximately 1–2 ppm (~1.6 ppm) on the east and 0.7 ppm (~0.8 ppm) on the west of the domain in terms of column-averaged (PBL) CO_2 mole fractions. The large-scale CO_2 boundary inflow has more relative impact on uncertainty in XCO_2 than on PBL CO_2 , even more than the transport and flux components. This finding is consistent with Chen et al. (2019) who used an EnKF approach over a month to quantify the same components. They showed that the uncertainties from boundary conditions are persistent throughout the entire atmospheric column, while those from transport and biogenic fluxes rapidly decrease with altitude.

The domain-averaged uncertainties in both modeled PBL CO_2 and XCO_2 vary seasonally (Figure 11). In the PBL, biogenic CO_2 flux uncertainty is high in summer (due to strong negative NEE, Figure 5) and winter (strong positive NEE) months and low in spring and fall (due to low NEE); transport uncertainty stays high in summer months and low throughout the rest of the year; boundary condition uncertainty has seasonal variations that is similar to transport uncertainty but much smaller in magnitude.

4. Discussions

An ensemble provides an estimate of the probability distribution of model variables, given an estimate of the probability distribution of the drivers and the model physics. The initial selection of the ensemble members is arbitrary, and ensemble systems tend to underestimate the true uncertainty (Berner et al., 2015; Buizza et al., 2005; Hagedorn et al., 2008). A reliable ensemble system should ensure that its spread encompasses the variability of the observations. We used rank histograms as a tool to determine the reliability of our ensemble and to diagnose errors analyzing its mean and spread. Three major sources of errors in regional inversions were considered: the atmospheric transport model, biogenic fluxes, and boundary conditions. The final calibrated ensemble system appeared reliable enough with respect to the representativeness of the model uncertainties. Ideally, the estimated uncertainties presented here should also be studied in the same physical space used in inversions, namely, flux space for biogenic fluxes. Starting from our calibrated MsTMIP flux ensemble, error covariances can be extracted, regularized, and filtered from spurious error correlations, inevitable in small-size ensembles (e.g., Ménétrier et al., 2015). Only with these additional steps can our uncertainties be used in future regional inversions.

Our selection of the transport members was based on empirical evidence from Díaz-Isaac et al. (2018a) that a calibrated small-sized ensemble could recover the model errors built from a large-sized ensemble (such as a 45-member ensemble). This 10-member transport ensemble appeared to be biased in Figure 2. It tends to overestimate the PBL wind speed and direction and underestimate the PBL depths. We focused on the representativeness of the ensemble spread in modeled CO_2 . Each individual simulation is affected by added perturbations, to the exception of the reference simulation. Hence, no specific quantification of transport errors was performed for each member. Future studies should introduce an evaluation of the simulations, similar to the biogenic flux ensemble, to identify the most appropriate transport by using unperturbed simulations. A rigorous investigation of the selected transport members in the future could help improve model physics impacting atmospheric CO_2 simulations and help construct better transport members along with the consideration of computational cost.

We found that the CO_2 boundary conditions play an important role in modeled CO_2 and, in particular, in XCO_2 . Here we only include four global models to describe modeled CO_2 uncertainty from boundary conditions. This small ensemble likely under samples the true model uncertainty, a limitation to performing a direct evaluation of each global model solution. During the evaluations (rank histograms), we only considered the variability across the global models. Biases in global models are likely a more urgent issue to address for improving the accuracy of the inverse flux estimates at regional scale (Schuh et al., 2013), which is out of the scope of our work. Uncertainties from boundary conditions depend on the locations of the domain boundaries, whether the oceanic or terrestrial influence dominates, and depend on the size of the simulation domain as diffusion will remove spatial gradients as a function of travel time. The boundary uncertainties also depend on the amplitude of the discrepancies among global models. Our simulation domain was defined in part to reduce errors by extending our simulation domain to both oceans, except over Canada. While larger domains will reduce spatial and temporal variations, large biases from global models will remain in domain-limited simulations.

The biogenic CO₂ flux members we selected mainly from the M_sTMIP project. These TBMs represent a variety of model processes but have uniform meteorological and land surface drivers (Fisher et al., 2016; Huntzinger et al., 2013). Still, these TBMs show tremendous spatiotemporal variations across models (Figures 5, S1, and S2). To construct an unbiased biogenic flux ensemble spread, we did a few more text to find out the outliers on the monthly (seasonal) and annual timescales. The initial ensemble was filtered to remove the unrealistic members and hence provide more reliable estimates of the biogenic flux uncertainty. The flatness of the rank histogram confirmed the reliability of our ensemble, but additional uncertainty from model parameters and driver data has not been included here, which could further increase the actual spread of the ensemble. Ecosystem measurements, such as soil carbon stocks, soil respiration, and carbon storage, could provide an independent assessment of future TBM ensembles before being coupled to an atmospheric system.

Lastly, fossil fuel CO₂ emissions (FFCO₂), oceanic fluxes, and biomass burning can cause model uncertainty as well. For the domain of interest (cf. Figure 1a), oceanic fluxes are negligible. For example, CO₂ uptake is $-6.775 \text{ Tg C year}^{-1}$ along North Eastern Pacific, $-0.135 \text{ Tg C year}^{-1}$ along Californian Current along the coast, $-2.1 \text{ Tg C year}^{-1}$ along Gulf of Mexico, and $-2.723 \text{ Tg C year}^{-1}$ along Florida Upwelling (Laruelle et al., 2014), while the terrestrial sink is about $-766 \pm 250 \text{ Tg C year}^{-1}$ over North America (USGCRP, 2018). However, the averaged biomass burning are about $83.8 \pm 36.4 \text{ Tg year}^{-1}$ and $191.3 \pm 60.0 \text{ Tg year}^{-1}$ for temperate and boreal North America during 2002–2011, respectively (Shi et al., 2015) and fossil fuel emissions represent $1,775 \pm 105 \text{ Tg C year}^{-1}$ over North America (USGCRP, 2018). The uncertainties associated with these two components are not negligible compared to the net terrestrial sink and should be considered in future studies.

In practice, FFCO₂ is commonly assumed perfectly known and are not optimized. The design of our ensemble system followed this convention. However, the uncertainty of the total FFCO₂ is about 8% globally and 4% for the United States (Andres et al., 2014). Most of global gridded FFCO₂ products, such as Open-source Data Inventory for Anthropogenic CO₂ (ODIAC; Oda et al., 2018), Carbon Dioxide Information Analysis Center (CDIAC; Andres et al., 1996), EDGAR (<http://edgar.jrc.ec.europa.eu/>), and Fossil Fuel Data Assimilation System (Asefi-Najafabady et al., 2014), have the same constraints to the annual total estimates uncertainty within 8% (Oda et al., 2018). They were disaggregated to the grid-level based on different spatial proxies with various assumptions. Disaggregating the total FFCO₂ emissions from national to grid level is expected to potentially introduce significant uncertainty, which can be propagated into the atmospheric signals through atmospheric transport model simulations. However, the direct assessment of the errors associated with disaggregation cannot be easily done because of the lack of evaluation data (Oda et al., 2018). A companion study including the modeled CO₂ uncertainty from fossil fuel emissions at various timescales demonstrates that uncertainty in fossil fuel emissions is a key factor in the uncertainty surrounding biospheric flux estimates and can considerably impair our ability to quantify regional fluxes (Feng et al., 2019). In this study, following the common practice of atmospheric inversion, we only focus on the main three uncertainty components (i.e., biogenic fluxes, transport, and boundary conditions).

We mixed CO₂ data across towers and seasons and evaluation on the daytime PBL signals across the entire year of 2010. We expect that we would find biases and other problems seasonally, or at individual towers. These finer-resolution spatial and temporal statistics need to be considered in future work. Observations that are not representative of the goal of the study, that is, BAO and WGC towers (Figure S3), can lead to false conclusions about the characteristics of the ensemble. The National Aeronautics and Space Administration (NASA) Atmospheric Carbon and Transport-America aircraft campaigns have sampled greenhouse gases and meteorological data across four seasons and three regions in the United States in various weather regimes and along OCO-2 overpass paths in order to provide more comprehensive and regionally specific data constraints for future model development.

5. Conclusions

We constructed a calibrated ensemble in order to quantify of the uncertainty components in modeled atmospheric CO₂ mole fractions, both in the PBL and for column-integrated mole fractions. We considered the model uncertainties caused by transport, biogenic fluxes, and boundary conditions. Initially, the ensemble

Acknowledgments

Primary funding for this research was provided by NASA's Earth Sciences Division as part of the Atmospheric Carbon and Transport (ACT)-America Earth Venture Suborbital mission (grant NNX15AG76G to Penn State). We thank M. P. Butler at the Pennsylvania State University for generating the codes that incorporate the global modeled CO₂ mole fractions into the regional model with the conservation of mass (archived at https://github.com/psu-inversion/WRF_boundary_coupling), S. Basu at NOAA ESRL GMD, Boulder, Colorado, United States, for providing TM5 modeled CO₂ mole fractions, J. Berner at the National Center for Atmospheric Research for technical support to implement the stochastic kinetic energy backscatter scheme, and S. Biraud for providing the list of AmeriFlux DOIs. The 3-hourly output from Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP; <http://nacp.ornl.gov/MsTMIP.shtml>) can be found on the Modeling and Synthesis Thematic Data Center at Oak Ridge National Laboratory (ORNL; <http://nacp.ornl.gov>). CarbonTracker CT2016 results were provided by NOAA ESRL, Boulder, Colorado, United States, from the website at <http://carbontracker.noaa.gov>. A set of GEOS-Chem simulated CO₂ mole fractions are provided by NASA Carbon Monitoring System (<https://carbon.nasa.gov/>). The CO₂ mole fraction data used in this work were prepared by the in situ tower and aircraft programs at National Oceanic and Atmospheric Administration (NOAA) Global Greenhouse Gas Reference Network. The data set was archived in `obspack_co2_1_GLOBALVIEWplus_v3.1_2017-10-18` (<https://doi.org/10.15138/G3T055>). The AmeriFlux Network data can be downloaded from <http://ameriflux.lbl.gov/>. The details of the sites used can be found in Table S1. Funding for AmeriFlux data resources was provided by the U.S. Department of Energy's Office of Science. The meteorological data used are from the NOAA rawinsonde stations (<https://ruc.noaa.gov/raobs/>). Computing resources were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. The WRF-Chem model output used for this study is available at datacommons.psu.edu (doi:10.26208/7a4p-q224).

system consisted of 10 transport ensemble members, 18 biogenic flux members, and 4 boundary condition members. We applied rank histograms to evaluate the reliability of the ensemble suites and calibrated the ensemble system. The aggregated ensemble produced a sound rank histogram for boundary layer CO₂ mole fractions, with evaluation of individual components to the extent possible. This work represents significant progress toward ensemble representation of atmospheric CO₂ simulations and the quantification of all the major sources of uncertainty in regional modeling.

In addition to using rank histograms to evaluate the biogenic CO₂ flux ensemble suite as a whole, we also assessed the biogenic CO₂ flux members individually using Taylor diagrams and a few additional statistical metrics. In the model rankings, biogenic fluxes from the global inversion system CarbonTracker outperformed all of the TBMs even in flux space, indicating the added value of atmospheric inversions. Spatially, some of the TBMs in the MsTMIP suite fail to capture the strong uptake over the U.S. Corn Belt in summer (Figure S2). In addition, some of the models overestimate the seasonal amplitude and mischaracterized the seasonality compared to other TBMs (Figure 3). Further model-data comparisons of both fluxes and mole fractions resulted in similar rankings of biogenic CO₂ members, implying that atmospheric data and eddy covariance flux measurements provide consistent and complementary information on surface fluxes.

The seasonal characteristics of modeled CO₂ uncertainties from transport, biogenic CO₂ fluxes, and CO₂ boundary conditions were studied using the calibrated ensemble members. All displayed seasonal variations in both PBL CO₂ and XCO₂. In PBL CO₂, the biogenic flux uncertainty played the most important role and reached a maximum of ~2.5 ppm in the winter and summer months and a minimum of ~1.5 ppm in the transition months. Transport uncertainty was as large as biogenic flux uncertainty in summer but stayed low throughout the rest of the year, ranging from 1.0 to 2.8 ppm. Both are highly uncertain in the U.S. Corn Belt area where vegetation fluxes are large. The spatial correlation between biogenic and transport model uncertainty, both driven by the actual magnitude of biogenic CO₂ fluxes, suggests that future inverse systems should include more rigorous estimates of transport errors to limit confounding error structures.

Boundary condition uncertainty plays the least role in the PBL CO₂ in general, ranging from 0.4 ppm to 1.2 ppm. It can reach over 2 ppm along the northern boundary of the domain. However, the influence of the boundary conditions remains persistent throughout the entire atmospheric column, while transport and biogenic fluxes only dominate PBL CO₂. CO₂ boundary conditions therefore become the most important source of uncertainties in modeled XCO₂.

The inverse results with well-calibrated full ensemble will provide us with direct impact of individual uncertainty source component on the optimized CO₂ fluxes. However, a careful handling of the building error covariances with the consideration of the temporal and spatial correlations is required and deserves another publication.

Author Contributions

S. Feng and T. Lauvaux designed the experiment. S. Feng conducted the experiment and the initial analysis and wrote the first draft of the paper. K. Keller and K. Davis contributed to the design of the model framework. Y. Zhou and C. Williams conducted the comparison of net ecosystem exchange between the terrestrial models, CarbonTracker, and flux tower measurements. A. Schuh and J. Liu provided GEOS-Chem modeled CO₂ mole fractions, and I. Baker provided SIB3_CSU 3 hourly biogenic CO₂ fluxes for the atmospheric CO₂ simulations. All authors edited and approved the final manuscript.

References

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., et al. (2003). The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–Present). *Journal of Hydrometeorology*, 4(6), 1147–1167. [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2)
- Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7), 1518–1530.
- Andres, R. J., Boden, T. A., & Higdon, D. (2014). A new evaluation of the uncertainty associated with CDIA estimates of fossil fuel carbon dioxide emission. *Tellus Series B: Chemical and Physical Meteorology*, 66(1), 23616.
- Andres, R. J., Marland, G., Fung, I., & Matthews, E. (1996). A 1° × 1° distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990. *Global Biogeochemical Cycles*, 10(3), 419–429.

- Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., et al. (2014). CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: Instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts. *Atmospheric Measurement Techniques*, 7(2), 647–687. <https://doi.org/10.5194/amt-7-647-2014>
- Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., et al. (2014). A multiyear, global gridded fossil fuel CO₂ emission data product: Evaluation and analysis of results. *Journal of Geophysical Research: Atmospheres*, 119(17), 20,213–20,231.
- Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., et al. (2006). TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO₂ fluxes, 1988–2003. *Global Biogeochemical Cycles*, 20, n/a. <https://doi.org/10.1029/2004GB002439>
- Baker, I., Denning, S., & Stöckli, R. (2010). North American gross primary productivity: Regional characterization and interannual variability. *Tellus B*, 62(5), 533–549.
- Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., & da Rocha, H. R. (2008). Seasonal drought stress in the Amazon: Reconciling models and observations. *Journal of Geophysical Research – Biogeosciences*, 113(G1).
- Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P., & White, J. W. C. (2012). Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature*, 488, 70.
- Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., & Miller, J. B. (2017). The impact of transport model differences on CO₂ surface flux estimates from OCO-2 retrievals of column average CO₂. *Atmospheric Chemistry and Physics Discussions*, 2017, 1–32.
- Basu, S., Miller, J. B., & Lehman, S. (2016). Separation of biospheric and fossil fuel fluxes of CO₂ by atmospheric inversion of CO₂ and ¹⁴CO₂ measurements: Observation system simulations. *Atmospheric Chemistry and Physics*, 16(9), 5665–5683.
- Berner, J., K. R. Fossell, S. Y. Ha, J. P. Hacker, and C. Snyder (2015). Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Monthly Weather Review*, 143(4), 1295–1320.
- Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A., & Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics Representations. *Monthly Weather Review*, 139(6), 1972–1995.
- Berner, J., Shutts, G. J., Leutbecher, M., & Palmer, T. N. (2009). A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *Journal of the Atmospheric Sciences*, 66(3), 603–626.
- Blasing, T. J., Broniak, C. T., & Marland, G. (2005). The annual cycle of fossil-fuel carbon dioxide emissions in the United States. *Tellus B*, 57(2), 107–115.
- Bousquet, P., Peylin, P., Ciais, P., Le Quééré, C., Friedlingstein, P., & Tans, P. P. (2000). Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science*, 290(5495), 1342–1347. <https://doi.org/10.1126/science.290.5495.1342>
- Broquet, G., Chevallier, F., Bréon, F. M., Kadygrov, N., Alemanno, M., Apadula, F., et al. (2013). Regional inversion of CO₂ ecosystem fluxes from atmospheric measurements: Reliability of the uncertainty estimates. *Atmospheric Chemistry and Physics*, 13(17), 9039–9056. <https://doi.org/10.5194/acp-13-9039-2013>
- Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y., & Wei, M. (2005). A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133(5), 1076–1097.
- Butler, M. P., Lauvaux, T., Feng, S., Liu, J., Bowman, K. W., & Davis, K. J. (2019). Mass-conserving coupling of total column CO₂ (XCO₂) from global to mesoscale models: Case study with CMS-Flux inversion system and WRF-Chem (v3.6.1). *Geoscientific Model Development Discussion*, 2019, 1–35.
- Carouge, C., Rayner, P. J., Peylin, P., Bousquet, P., Chevallier, F., & Ciais, P. (2010). What can we learn from European continuous atmospheric CO₂ measurements to quantify regional fluxes—Part 2: Sensitivity of flux accuracy to inverse setup. *Atmospheric Chemistry and Physics*, 10(6), 3119–3129.
- Chen, H. W., Zhang, F., Lauvaux, T., Davis, K. J., Feng, S., Butler, M. P., & Alley, R. B. (2019). Characterization of regional-scale CO₂ transport uncertainties in an ensemble with flow-dependent transport errors. *Geophysical Research Letters*, 46, 4049–4058. <https://doi.org/10.1029/2018GL081341>
- Chevallier, F., Feng, L., Bösch, H., Palmer, P. I., & Rayner, P. J. (2010). On the impact of transport model errors for the estimation of CO₂ surface fluxes from GOSAT observations. *Geophysical Research Letters*, 37(21)L21803 <https://doi.org/10.1029/2010GL044652>
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., et al. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437(7058), 529–533. <https://doi.org/10.1038/nature03972>
- Cooperative Global Atmospheric Data Integration, P. (2018). Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957–2017; obspack_co2_1_GLOBALVIEWplus_v4.1_2018-10-29, edited, NOAA Earth System Research Laboratory, Global Monitoring Division.
- Costa-e-Silva, F., Correia, A. C., Piayda, A., Dubbert, M., Rebmann, C., Cuntz, M., et al. (2015). Effects of an extremely dry winter on net ecosystem carbon exchange and tree phenology at a cork oak woodland. *Agricultural and Forest Meteorology*, 204, 48–57.
- Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., et al. (2019). The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network. *Atmospheric Chemistry and Physics Discussions*, 2019, 1–79.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>
- Diaz-Isaac, L. I., Lauvaux, T., & Davis, K. J. (2018a). Impact of physical parameterizations and initial conditions on simulated atmospheric transport and CO₂ mole fractions in the US Midwest. *Atmospheric Chemistry and Physics*, 18(20), 14,813–14,835.
- Diaz-Isaac, L. I., Lauvaux, T., & Davis, K. J. (2018b). Impact of physical parameterizations and initial conditions on simulated atmospheric transport and CO₂ mole fractions in the US Midwest. *Atmospheric Chemistry and Physics Discussions*, 2018, 1–40.
- Feng, S., Lauvaux, T., Keller, K., Davis, K. J., Rayner, P., Oda, T., & Gurney, K. R. (2019). A road map for improving the treatment of uncertainties in high-resolution regional carbon flux inverse estimates. *Geophysical Research Letters*, 46. <https://doi.org/10.1029/2019GL082987>
- Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A., et al. (2016). Los Angeles megacity: A high-resolution land-atmosphere modelling system for urban CO₂ emissions. *Atmospheric Chemistry and Physics*, 16(14), 9019–9045. <https://doi.org/10.5194/acp-16-9019-2016>
- Fisher, J. B., Sikka, M., Huntzinger, D. N., Schwalm, C., & Liu, J. (2016). Technical note: 3-hourly temporal downscaling of monthly global terrestrial biosphere model net ecosystem exchange. *Biogeosciences*, 13(14), 4271–4277.
- Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2013). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27(2), 511–526.
- Garaud, D., & Mallet, V. (2011). Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: Application to air quality. *Journal of Geophysical Research-Atmospheres*, 116(D19).

- Gerbig, C., Körner, S., & Lin, J. C. (2008). Vertical mixing in atmospheric tracer transport models: Error characterization and propagation. *Atmospheric Chemistry and Physics*, 8(3), 591–602.
- Giglio, L., van der Werf, G. R., Randerson, J. T., Collatz, G. J., & Kasibhatla, P. (2006). Global estimation of burned area using MODIS active fire observations. *Atmospheric Chemistry and Physics*, 6(4), 957–974.
- Gloor, M., Bakwin, P., Hurst, D., Lock, L., Draxler, R., & Tans, P. (2001). What is the concentration footprint of a tall tower? *Journal of Geophysical Research-Atmospheres*, 106(D16), 17,831–17,840.
- Göckede, M., Turner David, P., Michalak Anna, M., Vickers, D., & Law Beverly, E. (2010). Sensitivity of a subregional scale atmospheric inverse CO₂ modeling framework to boundary conditions. *Journal of Geophysical Research-Atmospheres*, 115(D24).
- Gourdji, S. M., Mueller K. L., Yadav V., Huntzinger D. N., Andrews A. E., Trudeau M., et al. (2012). North American CO₂ exchange: Iinter-comparison of modeled estimates with results from a fine-scale atmospheric inversion. *Biogeosciences*, 9(1), 457–475. <https://doi.org/10.5194/bg-9-457-2012>.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., et al. (2002). Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models. *Nature*, 415(6872), 626–630. <https://doi.org/10.1038/415626a>
- Hagedorn, R., Hamill, T. M., & Whitaker, J. S. (2008). Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. *Part I: Two-meter temperatures*, *Monthly Weather Review*, 136(7), 2608–2619.
- Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. *Weather and Forecasting*, 12(4), 736–741.
- Hamill, T. M., & Colucci, S. J. (1997). Verification of Eta-RSM short-range ensemble forecasts. *Monthly Weather Review*, 125(6), 1312–1327.
- Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11), 2776–2790.
- Huntzinger, D. N., Schwalm C., Michalak A. M., Schaefer K., King A. W., Wei Y., et al. (2013). The North American Carbon Program Multi-Scale Synthesis and Terrestrial Model Intercomparison Project —Part 1: Overview and experimental design. *Geoscientific Model Development*, 6(6), 2121–2133. <https://doi.org/10.5194/gmd-6-2121-2013>
- IPCC (2013). In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), *Climate change 2013. The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 1535). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- Kretschmer, R., Gerbig, C., Karstens, U., Biavati, G., Vermeulen, A., Vogel, F., et al. (2014). Impact of optimized mixing heights on simulated regional atmospheric transport of CO₂. *Atmospheric Chemistry and Physics*, 14(14), 7149–7172.
- Kretschmer, R., Gerbig, C., Karstens, U., & Koch, F. T. (2012). Error characterization of CO₂ vertical mixing in the atmospheric transport model WRF-VPRM. *Atmospheric Chemistry and Physics*, 12, 2441–2458.
- Laruelle, G. G., Lauerwald, R., Pfeil, B., & Regnier, P. (2014). Regionalized global budget of the CO₂ exchange at the air-water interface in continental shelf seas. *Global Biogeochemical Cycles*, 28(11), 1199–1214.
- Lauvaux, T., & Davis, K. J. (2014). Planetary boundary layer errors in mesoscale inversions of column-integrated CO₂ measurements. *Journal of Geophysical Research: Atmospheres*, 119(2), 490–508.
- Lauvaux, T., Diaz-Isaac, L. I., Bocquet, M., & Bousserez, N. (2019). Diagnosing spatial error structures in CO₂ mole fractions and XCO₂ column mole fractions from atmospheric transport. *Atmospheric Chemistry and Physics Discussions*, 2019, 1–26.
- Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., et al. (2012). Constraining the CO₂ budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system. *Atmospheric Chemistry and Physics*, 12(1), 337–354. <https://doi.org/10.5194/acp-12-337-2012>
- Law, R. M., Peters, W., Rödenbeck, C., Aulagnier, C., Baker, I., Bergmann, D. J., et al. (2008). TransCom model simulations of hourly atmospheric CO₂: Experimental overview and diurnal cycle results for 2002. *Global Biogeochemical Cycles*, 22, n/a. <https://doi.org/10.1029/2007GB003050>
- Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, 2(12), 831–836. <https://doi.org/10.1038/ngeo689>
- Lin, J. C., & Gerbig, C. (2005). Accounting for the effect of transport errors on tracer inversions. *Geophysical Research Letters*, 32, L01802. <https://doi.org/10.1029/2004GL021127>
- Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., et al. (2014). Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT XCO₂ sampling on the inference of terrestrial biospheric sources and sinks. *Tellus Series B: Chemical and Physical Meteorology*, 66(1), 22486. <https://doi.org/10.3402/tellusb.v66.22486>
- Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., et al. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. *Science*, 358(6360), eaam5690. <https://doi.org/10.1126/science.aam5690>
- Liu, J., Fung, I., Kalnay, E., & Kang, J.-S. (2011). CO₂ transport uncertainties from the uncertainties in meteorological fields. *Geophysical Research Letters*, 38(12), L12808. <https://doi.org/10.1029/2011GL047213>
- Lokupitiya, E., Denning, A. S., Schaefer, K., Ricciuto, D., Anderson, R., Arain, M. A., et al. (2016). Carbon and energy fluxes in cropland ecosystems: A model-data comparison. *Biogeochemistry*, 129(1–2), 53–76. <https://doi.org/10.1007/s10533-016-0219-3>
- Mao, J., Thornton, P. E., Shi, X., Zhao, M., & Post, W. M. (2012). Remote Sensing Evaluation of CLM4 GPP for the Period 2000–09. *Journal of Climate*, 25(15), 5327–5342.
- Marland, G., Boden, T. A., & Andres, R. J. (2008). *Global, regional, and national fossil fuel CO₂ emissions. In trends: A compendium of data on global change*. Oak Ridge, TN: edited by O. R. N. L. Carbon Dioxide Information Analysis Center, U.S. Department of Energy.
- Masarie, K. A., Peters, W., Jacobson, A. R., & Tans, P. P. (2014). ObsPack: A framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements. *Earth System Science Data*, 6(2), 375–384.
- McCaughey, J. H., Pejam, M. R., Arain, M. A., & Cameron, D. A. (2006). Carbon dioxide and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada. *Agricultural and Forest Meteorology*, 140(1), 79–96.
- Ménétrier, B., T. Montmerle, Y. Michel, and L. Berre (2015). Linear filtering of sample covariances for ensemble-based data assimilation. *Part I: Optimality criteria and application to variance filtering and covariance localization*, *Monthly Weather Review*, 143(5), 1622–1643.
- Miles, N. L., Richardson, S. J., Davis, K. J., Lauvaux, T., Andrews, A. E., West, T. O., et al. (2012). Large amplitude spatial and temporal gradients in atmospheric boundary layer CO₂ mole fractions detected with a tower-based network in the U.S. upper Midwest. *Journal of Geophysical Research*, 117, G01019. <https://doi.org/10.1029/2011JG001781>
- Nagler, P. L., Scott, R. L., Westenburg, C., Cleverly, J. R., Glenn, E. P., & Huete, A. R. (2005). Evapotranspiration on western U.S. rivers estimated using the *Enhanced Vegetation Index from MODIS and data from eddy covariance and Bowen ratio flux towers*, *Remote Sensing of Environment*, 97(3), 337–351.

- Oda, T., Maksyutov, S., & Andres, R. J. (2018). The Open-source Data Inventory for Anthropogenic CO₂, version 2016 (ODIAC2016): A global monthly fossil fuel CO₂ gridded emissions data product for tracer transport simulations and surface flux inversions. *Earth System Science Data*, 10(1), 87–107. <https://doi.org/10.5194/essd-10-87-2018>
- Ogle, S. M., Davis, K., Lauvaux, T., Schuh, A., Cooley, D., West, T. O., et al. (2015). An approach for verifying biogenic greenhouse gas emissions inventories with atmospheric CO₂ concentration data. *Environmental Research Letters*, 10(3), 034012. <https://doi.org/10.1088/1748-9326/10/3/034012>
- Olsen, S. C., & Randerson, J. T. (2004). Differences between surface and column atmospheric CO₂ and implications for carbon cycle research. *Journal of Geophysical Research-Atmospheres*, 109(D2).
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., et al. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proceedings of the National Academy of Sciences*, 104(48), 18,925–18,930. <https://doi.org/10.1073/pnas.0708986104>
- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., et al. (2013). Global atmospheric carbon budget: Results from an ensemble of atmospheric CO₂ inversions. *Biogeosciences*, 10(10), 6699–6720. <https://doi.org/10.5194/bg-10-6699-2013>
- Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, 24(14), 3624–3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>
- Saetra, Ø., Hersbach, H., Bidlot, J., & Richardson, D. S. (2004). Effects of Observation Errors on the Statistics for Ensemble Spread and Reliability. *Monthly Weather Review*, 132, 1487–1501. [https://doi.org/10.1175/1520-0493\(2004\)132<1487:EOOEOOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1487:EOOEOOT>2.0.CO;2)
- Sarmiento, J. L., Gloor, M., Gruber, N., Beaulieu, C., Jacobson, A. R., Mikaloff Fletcher, S. E., et al. (2010). Trends and regional distributions of land and ocean carbon sinks. *Biogeosciences*, 7(8), 2351–2367.
- Schimel, D., Stephens, B. B., & Fisher, J. B. (2015). Effect of increasing CO₂ on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences*, 112(2), 436–441.
- Schuh, A. E., Baker, D. F., & Jacobson, A. (2015). *GEOS-Chem-CarbonTracker*, in *ESRL/Global Monitoring Annual Conference*. Boulder, CO.
- Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., et al. (2010). A regional high-resolution carbon flux inversion of North America for 2004. *Biogeosciences*, 7(5), 1625–1644.
- Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., et al. (2019). Quantifying the impact of atmospheric transport uncertainty on CO₂ surface flux estimates. *Global Biogeochemical Cycles*, 33(4), 484–500.
- Schuh, A. E., Lauvaux, T., West, T. O., Denning, A. S., Davis, K. J., Miles, N., et al. (2013). Evaluating atmospheric CO₂ inversions at multiple scales over a highly inventoried agricultural landscape. *Global Change Biology*, 19(5), 1424–1439. <https://doi.org/10.1111/gcb.12141>
- Schwalm, C. R., Huntzinger, D. N., Fisher, J. B., Michalak, A. M., Bowman, K., Ciais, P., et al. (2015). Toward “optimal” integration of terrestrial biosphere models. *Geophysical Research Letters*, 42, 4418–4428. <https://doi.org/10.1002/2015GL064002>
- Shi, Y., Matsunaga, T., Saito, M., Yamaguchi, Y., & Chen, X. (2015). Comparison of global inventories of CO₂ emissions from biomass burning during 2002–2011 derived from multiple satellite products. *Environmental Pollution*, 206, 479–487. <https://doi.org/10.1016/j.envpol.2015.08.009>
- Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131(612), 3079–3102.
- Sitch, S., Huntingford, C., Gedney, N., Levy, P. E., Lomas, M., Piao, S. L., et al. (2008). Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Global Change Biology*, 14(9), 2015–2039. <https://doi.org/10.1111/j.1365-2486.2008.01626.x>
- Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., et al. (2007). Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂. *Science*, 316(5832), 1732–1735. <https://doi.org/10.1126/science.1137004>
- Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., et al. (2015). Seasonal climatology of CO₂ across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network. *Journal of Geophysical Research: Atmospheres*, 120(10), 5155–5190. <https://doi.org/10.1002/2014JD022591>
- Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., et al. (2002). Global sea-air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(9–10), 1601–1622. [https://doi.org/10.1016/S0967-0645\(02\)00003-6](https://doi.org/10.1016/S0967-0645(02)00003-6)
- Talagrand, O., Vautard, R., & Strauss, B. (1997). Evaluation of probabilistic prediction systems, paper presented at Workshop on Predictability. *ECMWF, Shinfield Park, Reading*, 20–22 October 1997.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics. Philadelphia: PA.
- USGCRP (U.S. Global Change Research Program) (2018). In N. Cavallaro, G. Shrestha, R. Birdsey, M. A. Mayes, R. G. Najjar, S. C. Reed, P. Romero-Lankao, & Z. Zhu (Eds.), *Second State of the Carbon Cycle Report (SOCCR2): A sustained assessment report* (p. 878). Washington, DC: U.S. Global Change Research Program. <https://doi.org/10.7930/SOCCR2.2018>
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., & Arellano, A. F. Jr. (2006). Interannual variability in global biomass burning emissions from 1997 to 2004. *Atmospheric Chemistry and Physics*, 6(11), 3423–3441.
- Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., et al. (2014). *NACP MSTMIP: Global and North American Driver Data for Multi-Model Intercomparison*. Data set. Oak Ridge, TN: Ridge National Laboratory Distributed Active Archive Center. <https://doi.org/10.3334/ORNLDAAC/1220>
- West, T. O., Bandaru, V., Brandt, C. C., Schuh, A. E., & Ogle, S. M. (2011). Regional uptake and release of crop carbon in the United States. *Biogeosciences*, 8(8), 2037–2046.
- Zeng, N., Qian, H., Roedenbeck, C., & Heimann, M. (2005). Impact of 1998–2002 midlatitude drought and warming on terrestrial ecosystem and the global carbon cycle. *Geophysical Research Letters*, 32, L22709. <https://doi.org/10.1029/2005GL024607>