JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD034406

Special Section:
Carbon Weather: Toward the next generation of regional greenhouse gas inversion systems

Evaluation of CarbonTracker's Inverse Estimates of North American Net Ecosystem Exchange of CO_2 From Different Observing Systems Using ACT-America Airborne Observations

Yu Yan Cui[1], Andrew R. Jacobson[2,3], Sha Feng[1,4], Daniel Wesloh[1], Zachary R. Barkley[1], Li Zhang[1], Tobias Gerken[1,5], Klaus Keller[6,7], David Baker[8], and Kenneth J Davis[1,7]

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA, [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [3]NOAA Earth System Research Laboratory, Global Monitoring Laboratory, Boulder, CO, USA, [4]Now at Atmospheric Sciences & Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA, [5]Now at School of Integrated Sciences, James Madison University, Harrisonburg, VA, USA, [6]Department of Geosciences, The Pennsylvania State University, University Park, PA, USA, [7]Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA, [8]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

Abstract Quantification of regional terrestrial carbon dioxide (CO_2) fluxes is critical to our understanding of the carbon cycle. We evaluate inverse estimates of net ecosystem exchange (NEE) of CO_2 fluxes in temperate North America, and their sensitivity to the observational data used to drive the inversions. Specifically, we consider the state-of-the-science CarbonTracker global inversion system, which assimilates (a) in situ measurements (IS), (b) the Orbiting Carbon Observatory-2 (OCO-2) v9 column CO_2 (XCO_2) retrievals over land (LNLG), (c) OCO-2 v9 XCO_2 retrievals ocean-glint (OG), and (d) a combination of all these observational constraints (LNLGOGIS). We use independent CO_2 observations from the Atmospheric Carbon and Transport (ACT)—America aircraft mission to evaluate the inversions. We diagnose errors in the flux estimates using the differences between modeled and observed biogenic CO_2 mole fractions, influence functions from a Lagrangian transport model, Bayesian inference, and root-mean-square error (RMSE) and bias metrics. The IS fluxes have the smallest RMSE among the four products, followed by LNLG. Both IS and LNLG outperform the OG and LNLGOGIS inversions with regard to RMSE. Regional errors do not differ markedly across the four sets of posterior fluxes. The CarbonTracker inversions appear to overestimate the seasonal cycle of NEE in the Midwest and Western Canada, and overestimate dormant season NEE across the Central and Eastern US. The CarbonTracker inversions may overestimate annual NEE in the Central and Eastern US. The success of the LNLG inversion with respect to independent observations bodes well for satellite-based inversions in regions with more limited in situ observing networks.

Plain Language Summary Biological CO_2 fluxes, an important component of the earth's climate system, remain uncertain, especially at continental and sub-continental spatial domains. Different global CO_2 observing systems imply significantly different net biological fluxes of CO_2. We use independent CO_2 measurements from an extensive multi-seasonal aircraft campaign to evaluate biological CO_2 flux estimates derived from four different observational systems entered into a common data analysis system. The observations include both ground and satellite-based measurements. We found that one of the the satellite-based CO_2 estimates performs nearly as well as the estimates based on ground-based measurements. This suggests that the satellite data may serve to estimate regional variations in biological CO_2 fluxes in portions of the globe with more limited ground-based observing networks. The inversions all appear to overestimate dormant season release of biological CO_2 to the atmosphere, thus may underestimate the net uptake of CO_2 by ecosystems in the Central and Eastern United States.

Key Points:

- In situ measurements and the land nadir/land glint inversions are the most reliable products of CarbonTracker in temperate North America, superior to ocean-glint or LNLGOGIS inversions
- Errors in these CarbonTracker regional flux estimates are not strongly dependent on the observational data sources
- CarbonTracker overestimates seasonal net ecosystem exchange (NEE) for the Eastern and Central US, thus the annual NEE may underestimate continental uptake of CO_2

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Y. Cui,
yqc5573@psu.edu

Citation:

Cui, Y. Y., Jacobson, A. R., Feng, S., Wesloh, D., Barkley, Z. R., Zhang, L., et al. (2021). Evaluation of CarbonTracker's inverse estimates of North American net ecosystem exchange of CO_2 from different observing systems using ACT-America airborne observations. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD034406. https://doi.org/10.1029/2020JD034406

Received 14 DEC 2020
Accepted 30 MAY 2021

Author Contributions:

Conceptualization: Yu Yan Cui, Daniel Wesloh, David Baker, Kenneth J Davis

Data curation: Yu Yan Cui, Andrew R. Jacobson, Sha Feng, Zachary R. Barkley, Li Zhang, Tobias Gerken
Formal analysis: Yu Yan Cui
Investigation: Yu Yan Cui
Methodology: Yu Yan Cui, Daniel Wesloh, Zachary R. Barkley, Tobias Gerken, Klaus Keller, David Baker, Kenneth J Davis
Resources: Andrew R. Jacobson, Sha Feng
Supervision: David Baker, Kenneth J Davis
Validation: Yu Yan Cui, Zachary R. Barkley, Klaus Keller, Kenneth J Davis
Visualization: Yu Yan Cui
Writing – original draft: Yu Yan Cui
Writing – review & editing: Yu Yan Cui

1. Introduction

Accurate quantification of carbon dioxide (CO_2) fluxes from different sources is an important input to the design of climate policies (e.g., Ciais et al., 2014; Keller et al., 2008; Rogelj et al., 2018). CO_2 flux related to terrestrial net ecosystem exchange (NEE) is one of the major components. It is challenging to quantify CO_2 NEE fluxes due to complex biosphere processes, together with the biosphere-atmosphere interactions (e.g., Tian et al., 2016). Both bottom-up and top-down approaches (e.g., Hayes et al., 2012; Hu et al., 2019; Liu et al., 2017; Pan et al., 2011; Thompson et al., 2020) have been used to characterize and quantify CO_2 NEE fluxes using data from a wide range of observation platforms.

The top-down approach is an optimization framework to improve a priori flux estimates, that are informed, for example, by ecosystem carbon-stock inventories or carbon flux models (e.g., Haynes et al., 2019). Atmospheric CO_2 measurements, on which the top-down method relies, can contribute powerful constraints to the bottom-up methods (e.g., Ogle et al., 2015). Different atmospheric CO_2 measurement platforms such as boundary-layer CO_2 mole fractions from ground-based networks (e.g., Andrews et al., 2014; Miles et al., 2012) and column-averaged CO_2 mole fractions (XCO_2) from satellites (e.g., Liu et al., 2020), aim to complement each other. Measurement biases, atmospheric transport errors, or representation errors, however, may cause difficulty in assimilating these measurements within the optimization process.

Evaluating current top-down CO_2 flux estimates from the different platforms with independent observations is a promising avenue to improve them. Chevallier et al. (2019) compares six global CO_2 atmospheric inversions from the combinations of three measurements platforms (i.e., Orbiting Carbon Observatory-2 - OCO-2 or Greenhouse Gas Observing Satellite - GOSAT column retrievals, and boundary-layer in situ measurements) using a large number of independent aircraft measurements in the free troposphere. They provide a cross-comparison among different inversion estimates as well as mole fraction-based comparisons between inversions and the aircraft measurements. They found the overall performance of inversions based on in situ data and based on OCO-2 XCO2 observations to be similar, however, they show that the posterior fluxes diverge for the northern and tropical parts of the continents. Seasonal, regional evaluation of the posterior fluxes is needed. The global inversions are temporally and spatially resolved products, and many aircraft field campaigns take place at a regional scale. This opens up the opportunity for further in-depth regional evaluations.

The Atmospheric Carbon and Transport–America (ACT-America) mission, conducted flights east of the Rocky Mountains in the United States (US) during Summer 2016, Winter 2017, Fall 2017, Spring 2018, and Summer 2019 (Davis et al., 2018; Davis et al., 2021). The multi-seasonal aircraft CO_2 sampling of ACT-America provides a unique opportunity for regional evaluation of CO_2 flux estimates. Extensive atmospheric CO_2 measurements from the atmospheric boundary layer (ABL) to the upper free troposphere during four seasons from ACT-America enable researchers to rigorously assess and potentially distinguish the biases and accuracy of different inversion estimates for temperate North America.

OCO-2 gathers XCO_2 measurements globally using nadir and glint observations over land, and glint observations over the oceans (Eldering, O'Dell, et al., 2017; Eldering, Wennberg, et al., 2017). The OCO-2 retrievals are continually being improved (e.g., Miller & Michalak, 2020; O'Dell et al., 2018). Independent observation campaigns can test the ability of the OCO-2 v9-based inversions to estimate regional-scale fluxes with accuracy and precision. Temperate North America has one of the densest in situ-based greenhouse gas monitoring networks in the world. An evaluation of the OCO-2 v9 based flux estimates, along with the evaluation of in situ-based CO_2 flux estimates together can be used to assess the complementary role of the two platforms. Additionally, a multi-platform strategy that combines in situ- and satellite-based platforms to constrain CO_2 NEE is promising but requires independent evaluation.

In this study, we implement a method to evaluate the in situ-based, OCO-2 v9-based, and two-system-combined inversions of CO_2 NEE in temperate North America using airborne observations from the ACT-America mission. Specifically, We evaluate the state-of-the-science CarbonTracker global inversion system's inverse NEE estimate for North America from four different set of observations, created as part of OCO-2 v9 model intercomparison project (https://www.esrl.noaa.gov/gmd/ccgg/OCO2_v9mip/). We evaluate the capability of the four different observing systems to quantify CO2 NEE in temperate North America. The

details of the evaluation framework are described in Section 2. Results and discussion are presented in Section 3. We conclude in Section 4.

2. Materials and Methods

2.1. CarbonTracker CO_2 NEE Flux Products

We evaluate four CO_2 flux products in the study, which are from CarbonTracker global inversion system (Jacobson et al., 2020). Following the protocol of OCO-2 v9 MIP, CarbonTracker performed a series of global CO_2 flux experiments for 2015–2019 driven by a variety of observation platforms, including CO_2 measurements from a) in situ data (IS) compiled in the GLOBALVIEW+ 5.0 (Cooperative Global Atmospheric Data Integration Project, 2019) and NRT v5.1 (CarbonTracker Team, 2019) ObsPack products; b) the land nadir/land glint (LNLG) retrievals of column-integrated CO_2 from OCO-2 v9; c) OCO-2 ocean glint (OG) v9 retrievals; and d) a combination of the in-situ and satellite data (LNLGOGIS). These global flux products are mapped onto 1-degree grid cells at 3-hourly intervals.

CarbonTracker of the OCO-2 v9 MIP use the Open-Data Inventory for Anthropogenic Carbon dioxide (ODIAC) 2018 fossil fuel emission inventory and the Global Fire Emissions Database (GFED) 4.1 s wildfire inventory. The prior biological CO_2 fluxes are from the GFED 4.1 s land net ecosystem exchange inventory post-processed by CarbonTracker 2019. The optimization process in the CarbonTracker system is based on the square root ensemble Kalman filter of (Whitaker & Hamill, 2002). Details of the CarbonTracker inversion system are described in (Peters et al., 2007; Jacobson et al., 2020) (https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2019B_doc.php).

2.2. ACT-America Aircraft Campaign

We use CO_2 measurements from the Summer 2016, Winter 2017, Fall 2017, and Spring 2018 ACT-America campaigns. These are the times for which CO_2 flux products are available from CarbonTracker, as part of the OCO-2 v9 MIP. The maps illustrate the seasonal average of CO_2 NEE fluxes and the spatial coverage of in-situ and OCO-2 LNLG/OG data during the ACT-America campaign periods are shown in Figure S1 and S2. Each ACT-America campaign flew over the same three sub-regions of the United States (US): the Mid-Atlantic, Midwest, and Gulf Coast. For most flight days, two aircraft (a NASA Langley B200 and a NASA Wallops C130) flew together measuring atmospheric CO_2 mole fractions and other atmospheric variables in patterns designed to sample the variability in atmospheric GHGs within mid-latitude weather systems and the associated regional surface fluxes. All flights were conducted during midday hours (15-0 UCT) in order to sample well mixed ABL conditions. The detailed instrument, deployment and data set of ACT-America are described in (Davis et al., 2021; Wei et al., 2021). The calibration of the CO_2 measurements are described by (Baier et al., 2020). About 35% of the flight time was within the ABL, the portion of the atmosphere most sensitive to regional GHG surface fluxes. In this study, we use the ABL measurements excluding the take-off and landing portions, and aggregate these CO_2 measurements across 30-s intervals (Figure 1, Table 1) to construct the receptors in the Lagrangian particle dispersion modeling that described in Section 2.3.

2.3. Influence Functions for ACT Flight Data

Upwind fluxes influence the aircraft samples. We explicitly quantify the source-receptor relationship (i.e., influence function) using a Lagrangian particle dispersion model (FLEXPART-WRF) (Brioude et al., 2013) in a backward mode. The simulations of FLEXPART-WRF are driven by the 27-km WRF-Chem simulated meteorology from the base line simulation described in Feng, Lauvaux, Davis, et al. (2019); Feng, Lauvaux, Keller et al. (2019) which were nudged to the 25-km ECMWF-ERA5 reanalysis data (Hersbach et al., 2020). In the study, we aggregated the set of influence functions to be the 1×1 degree resolution in terms of the flux evaluation.

We computed a suite of influence functions across 98 flight days, at the same spatial and temporal resolution of the meteorological driver (27 km and hourly) covering the entire domain (Figure 2). Each receptor of the influence function is the 30-s interval along flight tracks, characterized by a box with boundaries between the maximum and minimum latitude/longitude as well as between the maximum and minimum

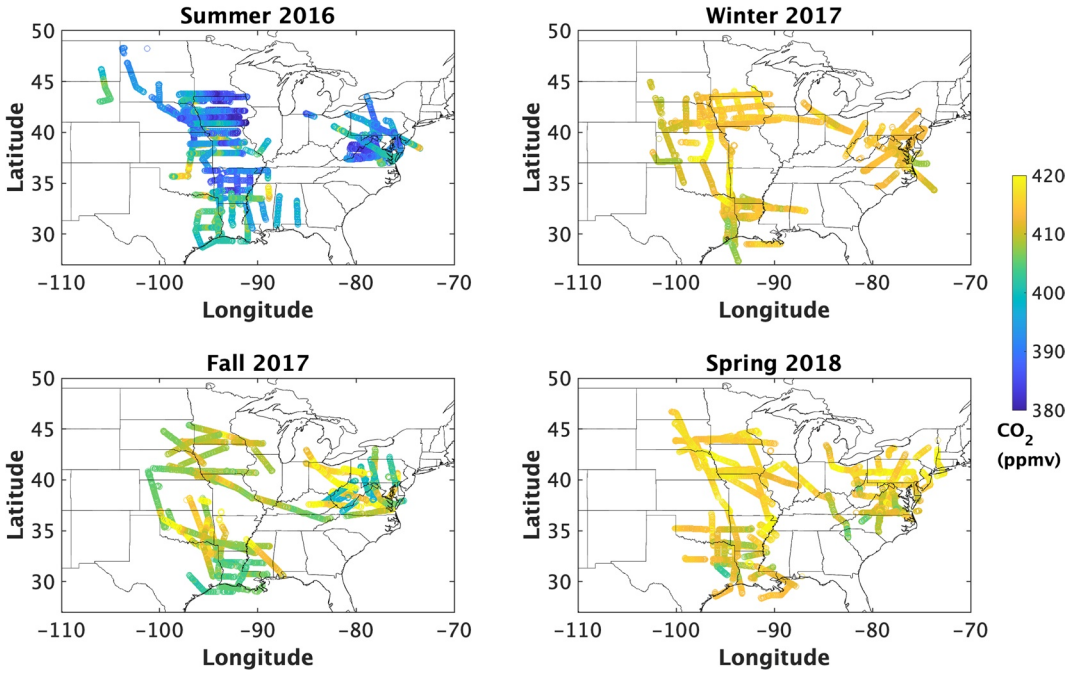

Figure 1. Boundary layer CO_2 mole fractions (unit: ppmv) sampled during four Atmospheric Carbon and Transport-America campaigns.

heights during the 30-s interval. Each receptor box released 5,000 particles and simulated their transport and dispersion backward for 10 days (Cui et al., 2015, 2017, 2019). Some validations of the suite of influence functions were conducted. Based on the same flux inputs, boundary conditions, and meteorological fields, we compared the FLEXPART-WRF simulated CO_2 mole fractions with the WRF-Chem forward simulations along flight tracks to evaluate the ability of the current influence function setup to reproduce corresponding WRF-Chem simulations in the domain. We found that they agreed well. The suite of influence functions plays a key role in our evaluation described in Section 2.4.2. Evaluation of the WRF transport fields has been performed in other ACT studies (e.g., Feng, Lauvaux, Davis, et al., 2019). Additional evaluation using the ACT airborne data is underway.

2.4. Biogenic CO_2 Component

2.4.1. Background Determination

To evaluate the surface fluxes in our domain, we subtract the CO_2 background values from the ACT CO_2 measurements to obtain an estimate of the CO_2 mole fraction enhancements and depletions caused by surface fluxes in the domain. The CO_2 boundary conditions in the WRF-Chem configuration are from CarbonTracker (Feng, Lauvaux, Davis, et al., 2019). We interpolate the boundary values along the flight tracks to determine the background-value elements in y_{bkg}. For the ACT Summer 2016 campaign, we used the 4-D simulations of atmospheric CO_2 mole fractions from the CarbonTracker 2017 product, while for the rest of the campaigns we used values from the CarbonTracker 2019-Near Real Time version 2 product. Upper free tropospheric mole fractions can provide another estimate of continental background conditions (Baier et al., 2020). We compare the simulated background mole fractions along ACT-America flight tracks above 4,000 mean sea level with the corresponding ACT-America measurements and find good agreement (Figure S3). We do not explicitly compute uncertainty in the background in this study, but this comparison, and the work of Feng, Lauvaux, Davis, et al., 2019 suggests that the uncertainty is less than about 1 ppm.

Table 1
Aircraft Data From Four Atmospheric Carbon and Transport-America Campaigns Used in the Study

	Flight months	Flight days	Flight (hours)	Atmospheric boundary layer data fraction (%)
Summer 2016	Jul–Aug	25	248	34
Winter 2017	Feb–Mar	25	218	35
Fall 2017	Oct–Nov	22	245	33
Spring 2018	Apr–May	26	261	32

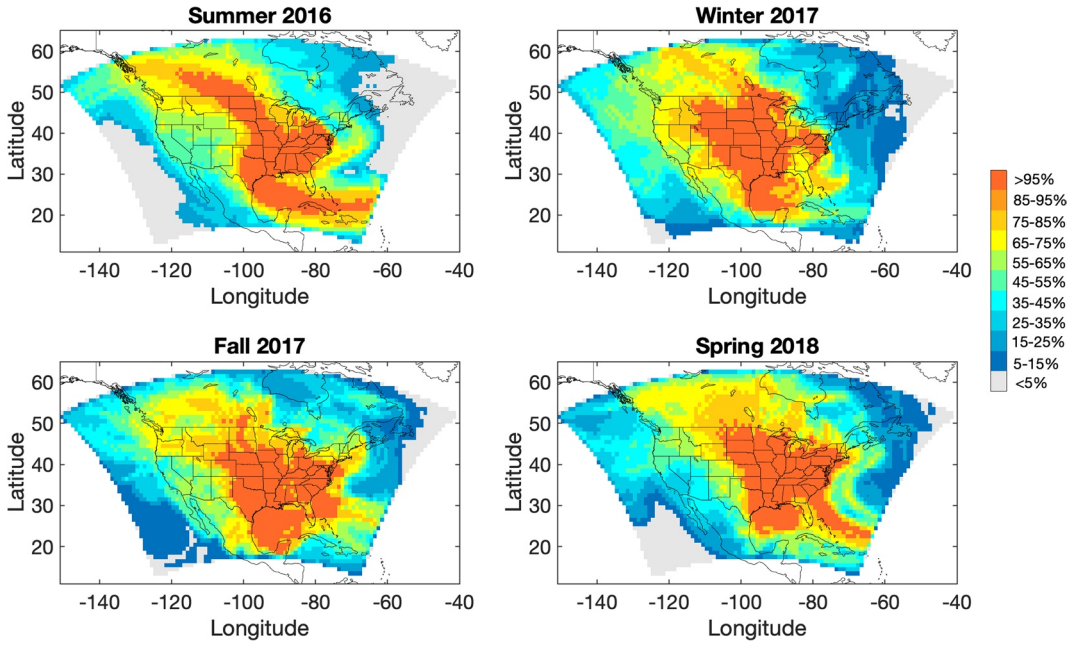

Figure 2. The spatial maps of accumulated 27 × 27 km 10-day influence functions are shown for each campaign, which are colored by the different percentile levels. Influence functions (unit: ppm/(mol m−2 s−1)) are used to quantify the relationship between the upwind sources and downwind receptors along the flights.

2.4.2. ACT Referenced Biogenic CO_2

The atmospheric CO_2 mole fraction continental enhancements and depletions include the influence of different fluxes: biogenic, fossil fuel, fire, and oceanic. To focus on the land biogenic CO_2 component, we remove the influence of the fossil fuel, fire, and oceanic sources on total $CO_2(y)$ by subtracting the component mole fraction enhancements simulated using the influence functions and flux estimates:

$$y_{ACTbio} = y - y_{bkg} - HE_{ff} - HE_{fire} - HE_{ocn}, \quad (1)$$

where H represents the influence functions (see details in 2.3), which are used with the fluxes to produce the atmospheric CO_2 mole fractions along flight tracks. E_{ff}, E_{fire}, E_{ocn} represent CO_2 fluxes from the fossil fuel, fire, and oceanic sources in the domain. E_{ff}, E_{fire}, E_{ocn} are obtained from the CarbonTracker system as part of OCO2 v9 MIP. As described in Section 2.1, E_{ff} is obtained from the ODIAC 2018 fossil fuel emission inventory, E_{fire} is from the GFED4.1 s wildfire inventory respectively. E_{ocn} is from the posterior ocean fluxes of the IS, LNLG, OG, or LNLGOGIS experiments, respectively. ACT-America campaigns were designed to fly over multiple productive ecoregions in Central and Eastern US and usually avoided urban areas, and wildfires in this region are not abundant. We modeled CO_2 enhancements/depletions from the four source components to the ACT-America boundary layer data space. The fire and ocean sources have trivial contributions to ACT-America data, compared with the biological and fossil fuel sources. Fossil fuels have modest influence on ACT-America data. Oda et al. (2018) estimated the annual ,uncertainty estimate of fossil fuel emission from ODIAC 2016 over North American Temperate to be 3.7%. Moreover, we convoluted two fossil fuel emission inventories to the ACT-America boundary layer data space and found the relative errors of mean values to be 2%–11% (Figure S4-S5). The uncertainties from the fossil fuel, fire and ocean fluxes used in Equation 1 are much smaller compared to the uncertainty of NEE of CO_2.

Meanwhile, the modeled biogenic CO_2 enhancements/depletions along the ACT flight tracks are also calculated as well from the four CO_2 NEE flux products (E_{bio}, see Section 2.1) respectively:

$$y_{modelbio} = HE_{bio} \quad (2)$$

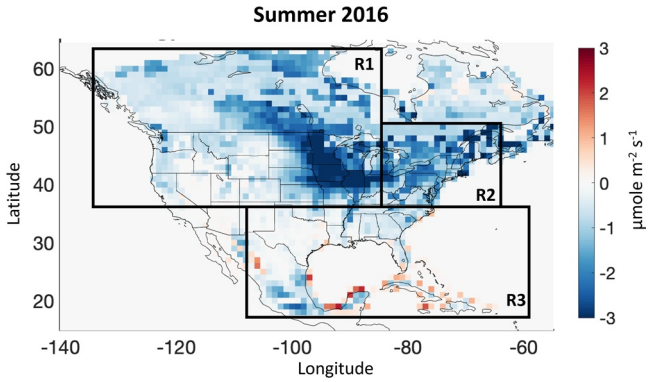

Figure 3. Three sub-domains are determined in the study: R1 denotes the Midwestern US and Western Canada areas; R2 denotes the Eastern US area; and R3 denotes the Southern US area. We only focus on the grid cells associated with the high values of the influence functions in the three domains. Details are described in Support Information. The background is the map of CO_2 NEE fluxes from the IS product, which are averaged values over July and August of 2016.

2.5. Evaluation Framework and Experimental Design

To distinguish and rank the different flux products, we calculate the root-mean-square error (RMSE) between $y_{modelbio}$ and y_{ACTbio}. The value of $y_{modelbio}$ is calculated using the influence functions and the flux products at the 3-hourly 1 × 1 degree spatial and temporal resolutions. The flux product associated with the smaller RMSE value indicates the better performance, and vice versa. The RMSE analysis is applied for all data during each campaign as well as the entire four campaign datasets.

The mole fraction-based analysis above is the net result of upwind biogenic fluxes. It is hence difficult to identify the sub-regional and ecosystem-specific sources of these divergences between the aircraft observations and simulations from the flux products without further diagnosis (Rayner, 2020). Therefore, in the study, we also conduct the flux-based evaluation to further diagnose the errors of flux products at the sub-regional scale. We use the following equations,

$$x = x_0 + BH^T(HBH^T + R)^{-1}(y_{ACTbio} - y_{modelbio}) \tag{3}$$

$$\varepsilon = x_0 - x, \tag{4}$$

and

$$\varepsilon = BH^T(HBH^T + R)^{-1}(y_{modelbio} - y_{ACTbio}), \tag{5}$$

where H (dimension: $m \times n$, m: receptors, n: states (spatial clusters associated with the time intervals) is the influence function, R (dimension: $m \times$ m) and B (dimension: $n \times$ n) represent the covariance of the model-data mismatch and the prior flux errors, respectively. x denotes the optimized flux value (the mode value of the posterior distribution, dimension: n x 1) using ACT-America data, and x_0 denotes the flux products that are evaluated in the study, that is the prior information.

We apply the Bayesian solution to optimize the flux products using the ACT-America data (Equation 3), and use the differences between the flux products and their optimizations by ACT-America (Equations 4 and 5) to evaluate the flux products.

ε (dimension: $n \times 1$) is a spatially and temporally resolved quantity and it represents the errors in the flux product compared with the ACT-America referenced fluxes. ε is in units of μmol/m/s and it has positive and negative signs. A lower magnitude of ε indicates the flux product is closer to the ACT referenced value. Positive values in ε identify grid clusters where flux products overestimate the NEE of CO_2, and vice versa.

R is assumed to be the variance of residuals between $y_{modelbio}$ and y_{ACTbio}. We give a conservative assumption for R. B is given to be 100% relative uncertainty of the flux product (x_0) initially, and we then apply a regularization parameter to B to tune the balance between the contributions of the model-data mismatch and the constraints of the prior estimation based on Equation 3 (Cui et al., 2015, 2017). Given the values of R and tuned B, we explicitly solve ε in Equation 5.

For this study, we focus on the seasonal-level evaluations, thereby we combine all data from each campaign (i.e., each season) as one case, and derive the corresponding spatially and temporally resolved values of ε. We focus on the grid cells associated with the large values of influence functions for each campaign (Figure 2), and aggregate these grid cells in each sub-domain (i.e., R1, R2, and R3 in Figure 3) according to the different ecoregions classified in the CarbonTracker system and obtain total 36, 36, 37, and 33 grid clusters for the four cases, respectively (more details in SI and Figure S6). R is treated as the diagonal matrix in the study. We aggregated the time intervals from the native 3-hourly intervals to the daytime (14–01 UTC) and nighttime (02–13 UTC) scales of each day and used an e-folding temporal correlation scale (20 days) to the same time period of day in the prior flux errors. We then calculate the weighted average of ε (without or within its sign) during each campaign, based on the temporal information constrained by H^TH for each domain (i.e., R1, R2, and R3), to identify the seasonal error levels for the flux products.

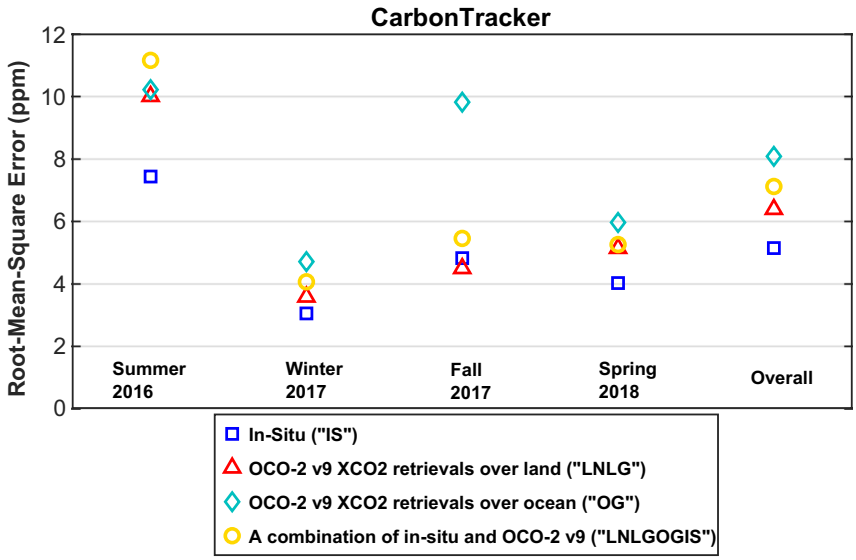

Figure 4. The Root-Mean-Square-Error between $y_{modelbio}$ and y_{ACTbio} from the four flux products are shown, for each Atmospheric Carbon and Transport campaign (Summer 2016, Winter 2017, Fall 2017, and Spring 2018), and combined four campaigns (overall).

3. Results and Discussion

As described in Section 2.5, we use both mole fraction-based and flux-based metrics to evaluate the four sets of NEE inversion products (e.g., IS, LNLG, OG, and LNLGOGIS). First, the mole fraction-based RMSE analysis are shown in Figure 4. We found that the IS flux product has the best performance among the four products during the summer, fall, and spring, and has the second-best performance during the winter time. The performance of the LNLG flux product is second in most seasons and best in the winter. The OG flux product has the worst performance across the winter, fall, and spring and it is consistent with previous studies (e.g., Crowell et al., 2019; O'Dell et al., 2018). The RMSE values integrated over four campaigns show that IS has the best aggregate performance at the annual level, followed by LNLG, OG, and LNLGOGIS. The multi-platform product (LNLGOGIS) performs similarly to the OG flux inversion.

We calculated the averaged absolute values of ε by campaign in Figure 5, based on Equations 3–5, to identify the spatial distribution of errors in the flux products. In general, the four flux products show similar spatial patterns during all four campaigns. The similar spatial patterns indicate that the spatial distributions of errors in the NEE of CO_2 estimates are not strongly dependent on the observational system used. All flux inversions show the largest errors in the Central and Eastern US during the summer time. There are larger errors in the Southern and Eastern US than other areas during the spring. The inversions in winter time show the smallest errors. Although the overall spatial patterns of errors are similar, some differences among the flux products can still be observed at the sub-regional scale. For example, LNLG and LNLGOGIS have similar overall performance with IS in Eastern and Southern US, but much worse than IS in Midwest and Western Canada.

We further calculate the seasonally averaged ε including the signs for the three sub-domains (Figure 6, and the corresponding spatial maps are shown in Figure S7) to identify the seasonal errors for these regions in the flux products. Again, the spatial patterns of the seasonal errors in these CarbonTracker regional flux estimates are not strongly dependent on the observational data sources. During the summertime, we found that all inversions overestimate NEE of CO_2 in the Eastern US (so the magnitude of net photosynthesis is underestimated), but significantly underestimate the flux (net photosynthesis is too large in magnitude) in the Midwest US and western Canada area from the LNLG and LNLGOGIS products. The LNLGOGIS product also underestimates NEE fluxes in the Southern US. The IS fluxes show the overall minimum errors across the three areas. The LNLG fluxes show similar errors with the IS fluxes in the Eastern and Southern US, but larger errors than IS in the Midwest US and Western Canada area in summer. Dormant season

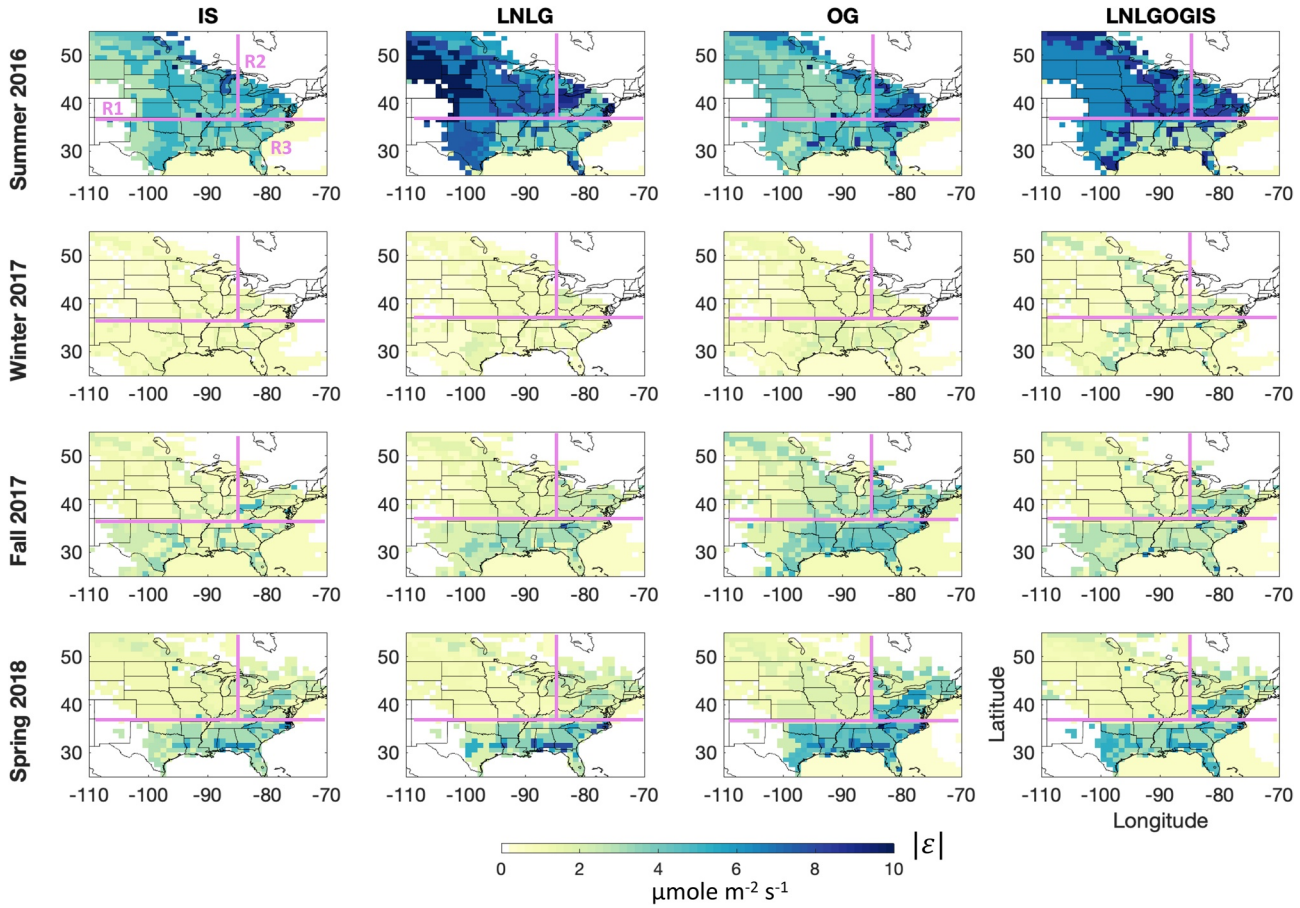

Figure 5. The spatial maps of seasonal ε absolute values without the positive and negative signs are shown, corresponding to the four flux products during each Atmospheric Carbon and Transport-America campaign, respectively. The three sub-regions are shown with the light pink lines.

NEE is generally overestimated in the inversions. The LNLG fluxes show a larger overestimate of NEE in Midwest and Western Canada during the wintertime compared with IS, but show a smaller overestimate of NEE in the Eastern and Southern US areas. During the fall, all inversions overestimate NEE of CO_2 in the Eastern US and underestimate NEE of CO_2 in the Southern US. The IS fluxes show fewer errors than the LNLG fluxes in the Midwest US and Western Canada and Southern US, but LNLG also shows a similar

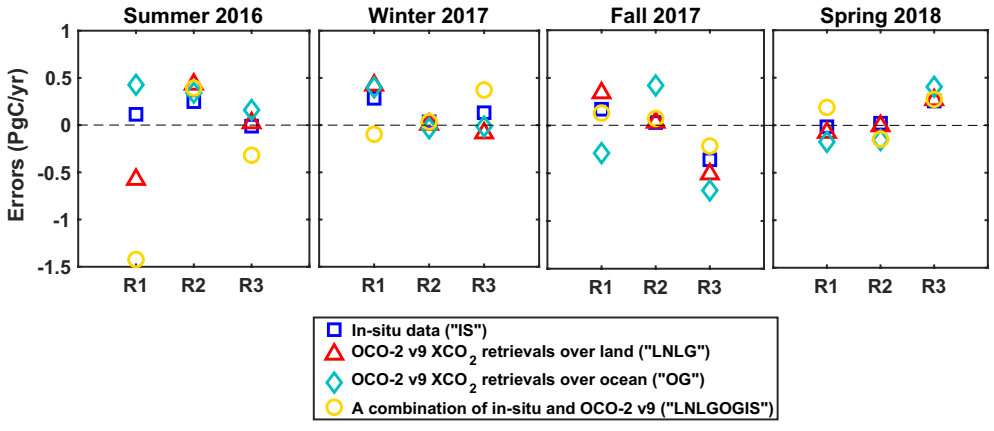

Figure 6. The integrated regional errors (ε values) refer to the daily flux estimation from the four flux products shown for each Atmospheric Carbon and Transport-America campaign, respectively.

overestimate of NEE in the Eastern US during the fall. The OG fluxes show the largest errors across the three domains. All inversions overestimate NEE of CO_2 in the Southern US during spring. The LNLG flux biases are similar in pattern and magnitude to the IS fluxes for the three domains.

Extrapolating these results across seasons suggests that the inversions generally amplified the seasonal cycle of NEE in Midwest and Western Canada by underestimating summer NEE or overestimating dormant season NEE, especially for the LNLG products. When we consider ε results across the four campaigns we found that the annual NEE of CO_2 fluxes have the positive errors in in Midwest and Western Canada and Eastern US from the IS and LNLG fluxes, but the LNLG fluxes show negative errors in the Southern US. The IS fluxes have the best seasonal performance and LNLG has the best annual performance across the three areas (i.e., the Central and Eastern temperate North America).

The seasonally averaged ε by daytime and nighttime for each case are calculated as well (Figure S8 and S9), respectively. We note that these day-night analysis might be influenced by biases in atmospheric transport, including simulation of the nocturnal atmospheric boundary layer. The spatial patterns of the errors during the daytime and nighttime largely match those found for the daily NEE error estimates in Figure 6. During the summertime, opposing patterns of ε (negative values during the daytime, and positive values during the nighttime) in Midwest and Western Canada suggest that both nighttime respiration and net daytime photosynthesis are overestimated in the area. Both positive biases during daytime and nighttime in the Eastern US suggest overestimated biogenic respiration in this region. During the wintertime, positive biases seen in day and night from IS and LNLG in Midwest and Western Canada indicate that respiration is overestimated in the region. The magnitudes of errors in day and night from all flux products are small in the Eastern US. Opposing patterns of ε (negative values during the daytime, and positive values during the nighttime) are seen in the Southern US. Consequently, the overall daily errors in these areas are small in Figure 6. In the fall, opposing patterns of ε (negative values during the daytime, and positive values during the nighttime) are seen again in the Southern US. In the spring, opposing patterns of ε (negative values during the daytime, and positive values during the nighttime) in the three domains suggest that both nighttime respiration and net daytime photosynthesis are overestimated in these areas.

4. Conclusions

We implement a framework to evaluate the NEE of CO_2 flux estimations across the Central and Eastern United States and some of Western Canada. We use this approach on the posterior fluxes from the CarbonTracker global flux inversion system, which, for the OCO2 v9 MIP, was run with four different atmospheric CO_2 data sources.

This study suggests that, in terms of regional variability in NEE of CO_2, the IS inversion and the inversion using the LNLG observations from OCO-2 v9 are likely to be the most reliable products of the CarbonTracker system, superior to inversions based on the OCO-2 v9 OG or all data platforms (LNLGOGIS) data sets. We found, using a error diagnosis metric, that IS generally outperforms the inversions based on OCO-2 v9 observations, but the differences between the IS inversion and the LNLG inversion are relatively small. The OG and LNLGOGIS inversions are clearly inferior to the IS and LNLG inversions with respect to this error metric analysis, and warrant further investigations. This strong performance of the LNLG inversion as compared to the IS inversion is encouraging when considering inverse flux estimates in regions of the world where the in situ observing network is sparse.

The spatially resolved errors for the regional fluxes in CarbonTracker are not strongly dependent on the observational data source. Our results suggest that CarbonTracker overestimates seasonal NEE for the Central and Eastern US, and that, as a result, the annual NEE from CarbonTracker may underestimate continental uptake of CO_2 (annual mean NEE too positive). Summer NEE is positively biased in the Eastern US and negatively biased in Midwest and Western Canada, yielding relatively little total seasonal bias across the continent in summer. In the dormant seasons, the CarbonTracker inversions appear generally to overestimate NEE. It is possible that the FLEXPART-WRF transport model used in our evaluation system may be biased. The differences between the two systems (FLEXPAT-WRF and TM5-ERA-interim) would also cause the uncertainty to the flux evaluation. Conclusive assessment of the magnitude of the errors in seasonal NEE from CarbonTracker will depend on a more rigorous assessment of the transport models, which is

currently being conducted. Nevertheless, we demonstrate that this continental-scale, multi-season airborne data set provides sufficient data to distinguish among inverse flux estimates and posterior identify flux biases, resulting in better understanding of the true NEE from North America.

We propose to extend this evaluation framework to other flux products from both top-down or bottom-up methods, such as other members of the OCO-2 v9 MIP and any available continental-scale biogenic CO_2 flux estimates. We hypothesize that these studies will yield insights that are applicable across the globe, especially in midlatitude ecosystems.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

FLEXPART-WRF model can be found online (https://www.flexpart.eu/wiki/FpLimitedareaWrf). All ACT-America in situ data used in the manuscript can be found online (https://doi.org/10.3334/ORNLDAAC/1556). The WRF-Chem simulations used in the manuscript can be found online (https://www.datacommons.psu.edu/commonswizard/MetadataDisplay.aspx?Dataset=6250), and FLEXPART-WRF simulations can be found online (ftp://evs2ftp.ornl.gov/Inversion_Models/Influence_Function/).

Acknowledgments
The Atmospheric Carbon and Transport (ACT) - America project is a NASA Earth Venture Suborbital 2 project funded by NASA's Earth Science Division (Grant NNX15AG76G to Penn State). We also would like to acknowledge the support for the OCO-2 v9 flux model inter-comparison project provided through NASA Grant #80 NSSC 18K0909. We gratefully thank Joshua P. DiGangi, Bianca Baier who led the measurements of CO_2in the ACT-America campaign, Yaxing Wei and Gao Chen who led data archival efforts, and Thomas Lauvaux who participated the design of the WRF-Chem modeling configuration.

References

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., et al. (2014). CO_2, CO, and CH_4 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

Baier, B. C., Sweeney, C., Choi, Y., Davis, K. J., DiGangi, J. P., Feng, S., & Weibring, P. (2020). Multispecies assessment of factors influencing regional and enhancements during the winter 2017 act-america campaign. *Journal of Geophysical Research: Atmospheres*, *125*(2), e2019JD031339. https://doi.org/10.1029/2019jd031339

Brioude, J., Arnold, D., Stohl, A., Cassiani, M., Morton, D., Seibert, P., et al. (2013). The lagrangian particle dispersion model flexpart-wrf version 3.1. *Geoscientific Model Development*, *6*(6), 1889–1904. https://doi.org/10.5194/gmd-6-1889-2013

Chevallier, F., Remaud, M., O'Dell, C. W., Baker, D., Peylin, P., & Cozic, A. (2019). Objective evaluation of surface- and satellite-driven carbon dioxide atmospheric inversions. *Atmospheric Chemistry and Physics*, *19*(22), 14233–14251. https://doi.org/10.5194/acp-19-14233-2019

Ciais, P., Dolman, A. J., Bombelli, A., Duren, R., Peregon, A., Rayner, P. J., et al. (2014). Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system. *Biogeosciences*, *11*(13), 3547–3602. https://doi.org/10.5194/bg-11-3547-2014

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*, *19*(15), 9797–9831. https://doi.org/10.5194/acp-19-9797-2019

Cui, Y. Y., Brioude, J., Angevine, W. M., Peischl, J., McKeen, S. A., Kim, S.-W., et al. (2017). Top-down estimate of methane emissions in California using a mesoscale inverse modeling technique: The San Joaquin valley. *Journal of Geophysical Research: Atmospheres*, *122*(6), 3686–3699. https://doi.org/10.1002/2016jd026398

Cui, Y. Y., Brioude, J., McKeen, S. A., Angevine, W. M., Kim, S.-W., Frost, G. J., et al. (2015). Top-down estimate of methane emissions in California using a mesoscale inverse modeling technique: The South Coast Air Basin. *Journal of Geophysical Research: Atmospheres*, *120*(13), 6698–6711. https://doi.org/10.1002/2014jd023002

Cui, Y. Y., Henze, D. K., Brioude, J., Angevine, W. M., Liu, Z., Bousserez, N., et al. (2019). Inversion estimates of lognormally distributed methane emission rates from the haynesville-bossier oil and gas production region using airborne measurements. *Journal of Geophysical Research: Atmospheres*, *124*(6), 3520–3531. https://doi.org/10.1029/2018jd029489

Davis, K., Browell, E., Feng, S., Lauvaux, T., Obland, M., Pal, S., & Williams, C. A. (2021). The atmospheric carbon and transport (act)—America mission. *Bulletin of the American Meteorological Society*. https://www.essoar.org/doi/pdf/10.1002/essoar.10505721.1

Davis, K., Obland, M., Lin, B., Lauvaux, T., O'Dell, C., Meadows, B., & Pauly, R. (2018). *Act-America: L3 merged in situ atmospheric trace gases and flask data, eastern USA*. Oak Ridge, Tennessee, USA: ORNL DAACm. https://doi.org/10.3334/ORNLDAAC/1593

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., et al. (2017). The orbiting carbon observatory-2: First 18 months of science data products. *Atmospheric Measurement Techniques*, *10*(2), 549–563. https://doi.org/10.5194/amt-10-549-2017

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., & Weir, B. (2017). The orbiting carbon observatory-2 early science investigations of regional carbon dioxide fluxes. *Science*, *358*(6360). https://doi.org/10.1126/science.aam5745

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 CO_2 simulations over North America. *Journal of Geophysical Research: Atmospheres*, *124*(24), 14325–14346. https://doi.org/10.1029/2019jd031165

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*(22), 13461–13469. https://doi.org/10.1029/2019gl082987

Hayes, D. J., Turner, D. P., Stinson, G., McGuire, A. D., Wei, Y., West, T. O., et al. (2012). Reconciling estimates of the contemporary NAmerican carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory-based data. *Global Change Biology*, *18*(4), 1282–1299. https://doi.org/10.1111/j.1365-2486.2011.02627.x

Haynes, K. D., Baker, I. T., Denning, A. S., Stöckli, R., Schaefer, K., Lokupitiya, E. Y., & Haynes, J. M. (2019). Representing grasslands using dynamic prognostic phenology based on biological growth stages: 1. Implementation in the simple biosphere model (SiB4). *Journal of Advances in Modeling Earth Systems*, *11*(12), 4423–4439. https://doi.org/10.1029/2018ms001540

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hu, L., Andrews, A. E., Thoning, K. W., Sweeney, C., Miller, J. B., Michalak, A. M., & van der Velde, I. R. (2019). Enhanced North American carbon uptake associated with el niño. *Science Advances*, *5*(6). https://doi.org/10.1126/sciadv.aaw0076

Jacobson, A. R., Schuldt, K. N., Miller, J. B., Oda, T., Tans, P., Andrews, A., & Zimnoch, M. (2020). *Carbontracker CT2019. Model published by NOAA Earth system research laboratory*. Global Monitoring Division. https://doi.org/10.25925/39m3-6069

Keller, K., McInerney, D., & Bradford, D. F. (2008). Carbon dioxide sequestration: How much and when? *Climatic Change*, *88*(8), 267. https://doi.org/10.1007/s10584-008-9417-x

Liu, J., Baskaran, L., Bowman, K., Schimel, D., Bloom, A. A., Parazoo, N. C., & Wofsy, S. (2020). Carbon monitoring system flux net biosphere exchange 2020 (CMS-Flux NBE 2020). *Earth System Science Data Discussions*, 1–53.

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

Miles, N. L., Richardson, S. J., Davis, K. J., Lauvaux, T., Andrews, A. E., West, T. O., & Crosson, E. R. (2012). Large amplitude spatial and temporal gradients in atmospheric boundary layer CO_2 mole fractions detected with a tower-based network in the U.S. upper midwest. *Journal of Geophysical Research: Biogeosciences*, *117*(G1). https://doi.org/10.1029/2011jg001781

Miller, S. M., & Michalak, A. M. (2020). The impact of improved satellite retrievals on estimates of biospheric carbon balance. *Atmospheric Chemistry and Physics*, *20*(1), 323–331. https://doi.org/10.5194/acp-20-323-2020

Oda, T., Maksyutov, S., & Andres, R. J. (2018). The open-source data inventory for anthropogenic co_2, version 2016 (ODIAC2016): A global monthly fossil fuel co_2 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

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., & Velazco, V. A. (2018). Improved retrievals of carbon dioxide from orbiting carbon observatory-2 with the version 8 ACOS algorithm. *Atmospheric Measurement Techniques*, *11*(12), 6539–6576. Retrieved from https://amt.copernicus.org/articles/11/6539/2018/. https://doi.org/10.5194/amt-11-6539-2018

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 CO_2 concentration data. *Environmental Research Letters*, *10*(3), 034012. https://doi.org/10.1088/1748-9326/10/3/034012

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., et al. (2011). A large and persistent carbon sink in the world's forests. *Science*, *333*(6045), 988–993. https://doi.org/10.1126/science.1201609

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), 18925–18930. https://doi.org/10.1073/pnas.0708986104

Rayner, P. (2020). Data assimilation using an ensemble of models: A hierarchical approach. *Atmospheric Chemistry and Physics*, *20*(6), 3725–3737. https://doi.org/10.5194/acp-20-3725-2020

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., et al. (2018). *IPCC, 2018: Global warming of 1.5°C. an ipcc special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (pp. 93–174). Geneva, Switzerland: World Meteorological Organization.

Thompson, R. L., Broquet, G., Gerbig, C., Koch, T., Lang, M., Monteil, G., et al. (2020). Changes in net ecosystem exchange over Europe during the 2018 drought based on atmospheric observations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1810), 20190512. https://doi.org/10.1098/rstb.2019.0512

Tian, H., Lu, C., Ciais, P., Michalak, A., Canadell, J., Saikawa, E., et al. (2016). The terrestrial biosphere as a net source of greenhouse gases to the atmosphere. *Nature*, *531*, 225–228. https://doi.org/10.1038/nature16946

Wei, Y., Shrestha, R., Pal, S., Gerken, T., McNelis, J., Singh, D., & Davis, K. (2021). The atmospheric carbon and transport (act)—America datasets: Description, management, and delivery. *Earth and Space Science*. https://www.essoar.org/doi/pdf/10.1002/essoar.10505692.1

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:edawpo>2.0.co;2