# Evaluation of inverse estimates of North American net ecosystem exchange of CO2 from different observing systems using ACT-America airborne observations

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

[1]The Pennsylvania State University
[2]University of Colorado Boulder
[3]James Madison University
[4]Colorado State University

November 22, 2022

**Abstract**

Quantification of regional terrestrial carbon dioxide (CO2) fluxes is critical to our understanding of the carbon cycle. We evaluate inverse estimates of net ecosystem exchange (NEE) of CO2 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 (i) in situ measurements ('IS'), 29 (ii) the Orbiting Carbon Observatory-2 (OCO-2) v9 column CO 2 (XCO2) retrievals over land ('LNLG'), (iii) OCO-2 v9 XCO 2 retrievals over ocean ('OG'), and (iv) a combination of all these observational constraints ('LNLGOGIS'). We use independent CO2 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 CO2 mole fractions, influence functions from a Lagrangian transport model, 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.

# Evaluation of 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], Daniel Wesloh[1], Tobias Gerken[1,4], Zachary R. Barkley[1], Klaus Keller[5,7], David Baker[6], 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 School of Integrated Sciences, James Madison University, Harrisonburg, VA, USA
[5]Department of Geosciences, The Pennsylvania State University, University Park, PA, USA
[6]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA
[7]Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA

**Key Points:**

- The IS and LNLG inversions are the most reliable products of CarbonTracker in temperate North America, superior to OG or LNLGOGIS inversions.
- Errors in these CarbonTracker regional flux estimates are not strongly dependent on the observational data sources.
- CarbonTracker overestimates seasonal NEE for the Eastern and Central US, as a result, the annual NEE from CarbonTracker may underestimate continental uptake of $CO_2$.

Corresponding author: Yu Yan Cui, `yqc5573@psu.edu`

**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 (i) in situ measurements ("IS"), (ii) the Orbiting Carbon Observatory-2 (OCO-2) v9 column $CO_2$ ($XCO_2$) retrievals over land ("LNLG"), (iii) OCO-2 v9 $XCO_2$ retrievals over ocean ("OG"), and (iv) 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, 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 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.

## 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. Masson-Delmotte et al., 2018; Keller et al., 2008; Ciais et al., 2014). $CO_2$ 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. Pan et al., 2011; Hayes et al., 2012; Liu et al., 2017; Hu et al., 2019; 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 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 that differences in annual budgets are mainly located in the northern and tropical portions of the globe. The OCO-2-based inversion produced results close to the traditional in situ inversion, but the data they used did not allow them to distinguish between the quality of OCO-2-based fluxes and in situ-based fluxes. 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). 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 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. Crowell et al., 2019; Miller & Michalak, 2020). 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 (MIP) (`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.

**Table 1.** Aircraft data from four ACT-America campaigns used in the study

| | Flight months | Flight days | Flight (hours) | ABL data (%) |
|---|---|---|---|---|
| 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 |

# 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 1) 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; 2) the land nadir/land glint (LNLG) retrievals of column-integrated $CO_2$ from OCO-2 v9; 3) OCO-2 ocean glint (OG) v9 retrievals; and 4) 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 (Figure S1).

## 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. 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 atmospheric boundary layer (ABL) conditions. The detailed instrument, deployment and dataset of ACT-America are described in (Davis et al., 2018; Wei et al., in prep for this issue). 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 et al. (2019a, 2019b) which were nudged to the 25-km ECMWF-ERA5 reanalysis data (Hersbach et al., 2020).

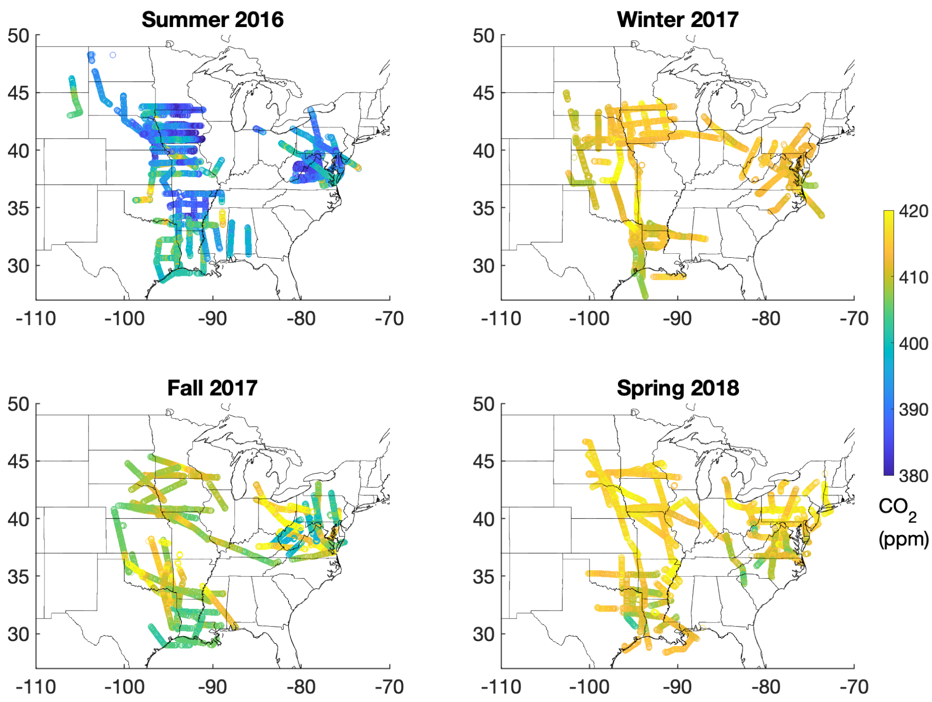



**Figure 1.** Boundary layer $CO_2$ mole fractions (unit: ppmv) sampled during four ACT-America campaigns.

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 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). A validation of the suite of influence functions was conducted. Based on the same flux inputs, background values, and meteorological fields, we compared the FLEXPART-WRF simulated $CO_2$ mole fractions with the WRF-Chem forward simulations along flight tracks and found that they agreed well. The suite of influence functions plays a key role in our evaluation described in Section 2.4.2.

## 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. A tracer indicating $CO_2$ mole fractions from the boundary condition of the domain (characterized with the CarbonTracker $CO_2$ 4-D simulations) is explicitly simulated in the WRF-Chem configuration (Feng et al., 2019a). We interpolated tracer values at each receptor point to represent 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

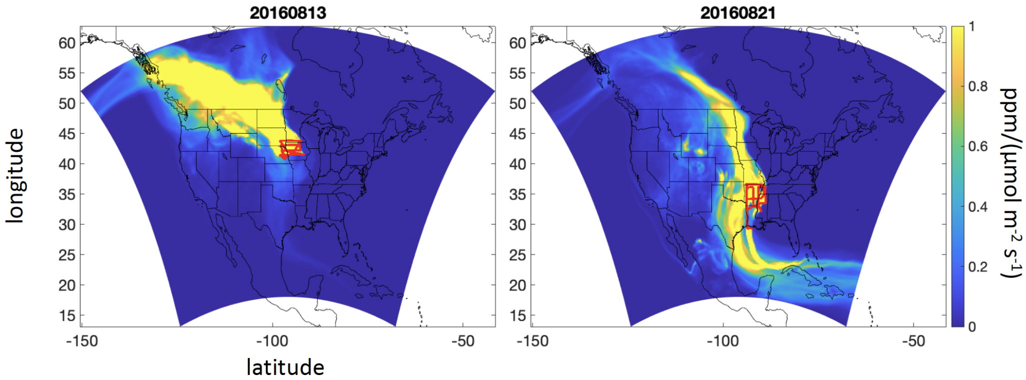



**Figure 2.** Examples of influence functions (unit: ppm/(μmol m-2 s-1)) used to quantify the relationship between the upwind sources and downwind receptors along the flights. Ten-day cumulative influence functions for two flight days are shown, including a fair-weather day, 13 August 2016 (left) and a frontal day, 21 August 2016 (right).

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 MSL with the corresponding ACT-America measurements and find good agreement (Figure S2). We do not explicitly compute uncertainty in the background in this study, but this comparison, and the work of Feng et al. (2019a) suggests that the uncertainty is less than about 1 ppm.

#### *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:

$$y_{ACTbio} = y - y_{bkg} - HE_{ff} - HE_{fire} - HE_{ocn} \tag{1}$$

, 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. $E_{ff}$ is obtained from the ODIAC2018 fossil fuel emission inventory, $E_{fire}$ is from the GFED4.1s wildfire inventory respectively, and $E_{ocn}$ is from the posterior ocean fluxes of the IS, LNLG, OG, or LNLGOGIS experiments, respectively.

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:

$$y_{modelbio} = HE_{bio} \tag{2}$$

### 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 native spatial and temporal resolutions (i.e 3-hourly, 1x1 degree). 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 apply the Bayesian solution to optimize the flux products using the ACT-America data and define a equation to be the differences between the flux products and their optimizations by ACT-America.

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

, where $H$ (dimension: m x n, m: receptors, n: states (spatial clusters associated with the time intervals) is the influence function, $R$ (dimension: m x m) and $B$ (dimension: n x n) represent the covariance of the model-data mismatch and the prior flux errors, respectively. $\varepsilon$ (dimension: n x 1) is a spatially- and temporally- resolved quantity and it represents the errors in the flux product compared with the ACT-America referenced fluxes. $\varepsilon$ is in units of µmol/m$^2$/s and it has positive and negative signs. A lower magnitude of $\varepsilon$ indicates the flux product is closer to the ACT referenced value. Positive values in $\varepsilon$ identify grid clusters where flux products overestimate the NEE of $CO_2$, and vice versa.

We explicitly solve $\varepsilon$ in the function (equation 3). $R$ is assumed to be squared residual between $y_{modelbio}$ and $y_{ACTbio}$. $B$ is given to be 100% relative uncertainty of the flux product 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(Cui et al., 2015). 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 $\varepsilon$. We focus on the grid cells associated with the large values of influence functions for each campaign (details refer to Figure S3), 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 Figure S4). $R$ 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 (20days) to the same time period of day in the prior flux errors. We then calculate the weighted average of $\varepsilon$ (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.

## 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

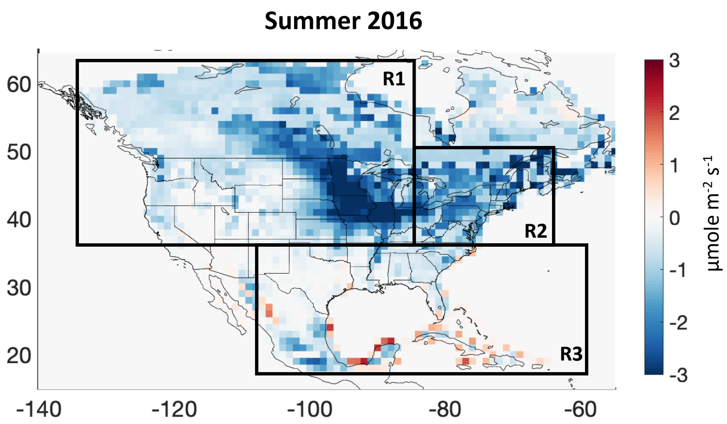



**Figure 3.** Three sub-domains are determined in the study: R1 denotes the Midwestern US and Western Canada areas; R2 denotes the Eastern US area; and R3 denotes the Southern US area. We only focus on the grid cells associated with the high values of the influence functions in the three domains. Details 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.

winter. The OG flux product has the worst performance across the winter, fall, and spring. 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 $\varepsilon$ by campaign in Figure 5, based on equation (3). 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 $\varepsilon$ including the signs for the three sub-domains (Figure 6, and the corresponding spatial maps are shown in Figure S5) 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 South-

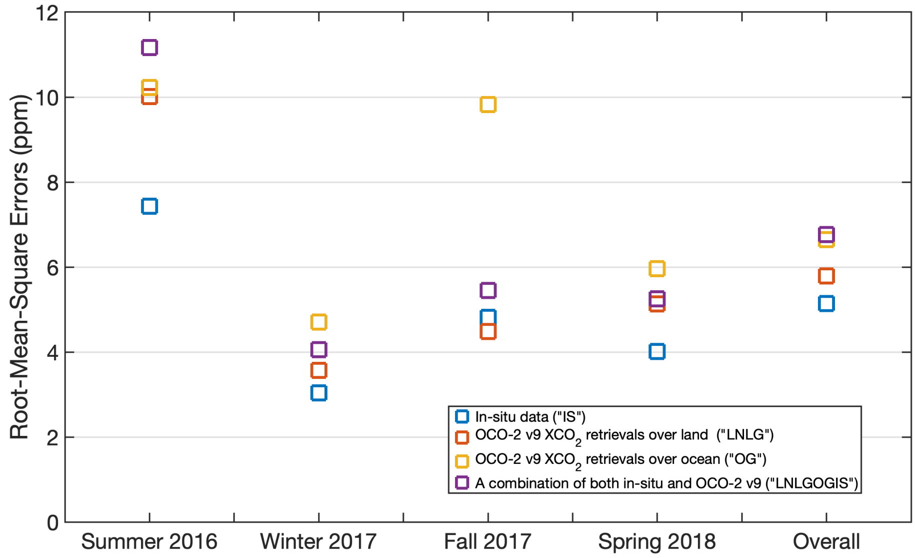



**Figure 4.** The Root-Mean-Square-Error (RMSE) between $y_{modelbio}$ and $y_{ACTbio}$ from the four flux products are shown, for each ACT campaign (Summer 2016, Winter 2017, Fall 2017, and Spring 2018), and combined four campaigns ("overall").

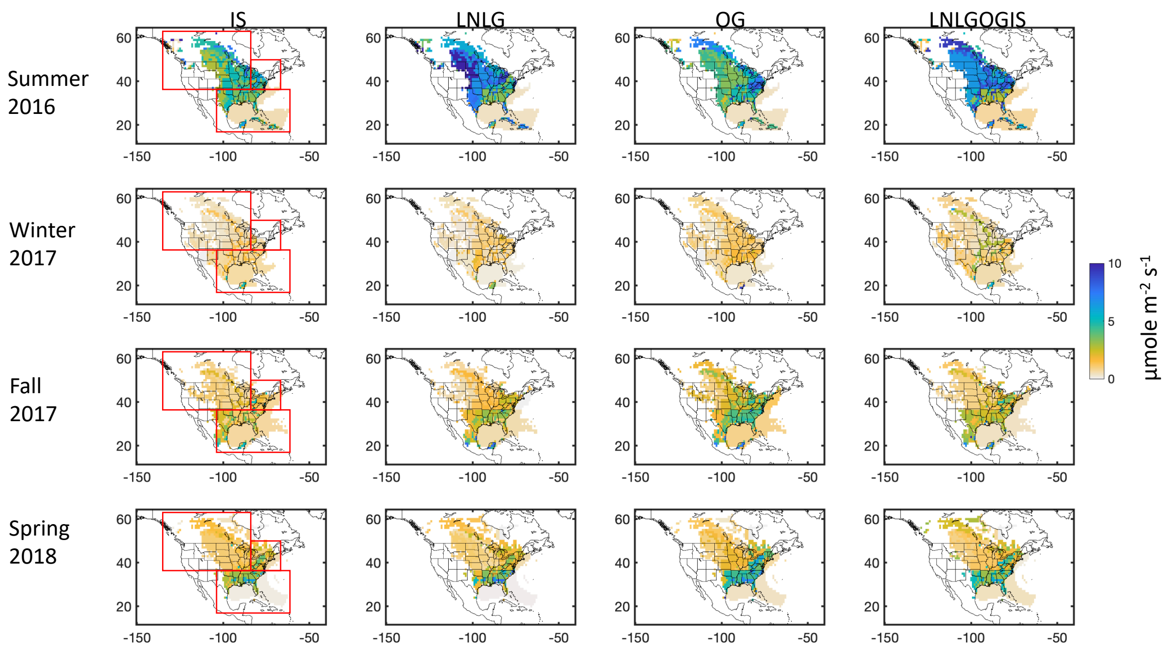



**Figure 5.** Spatial maps of the seasonally averaged $\varepsilon$ values without the positive and negative signs corresponding to the four flux products during each ACT-America campaign, respectively.

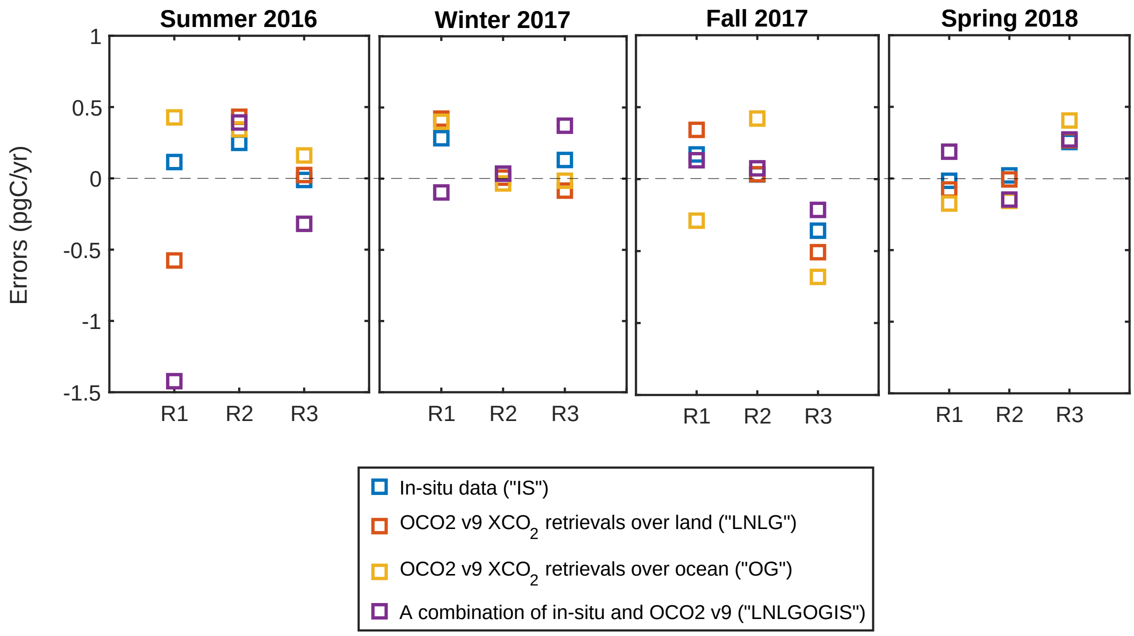



**Figure 6.** The integrated regional errors ($\varepsilon$ values) refer to the daily flux estimation from the four flux products shown for each ACT-America campaign, respectively.

ern US, but larger errors than IS in the Midwest US and Western Canada area in summer. Dormant season 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 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 $\varepsilon$ 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 $\varepsilon$ by daytime and nighttime for each case are calculated as well (Figure S6 and S7), respectively. 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 $\varepsilon$ (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 $\varepsilon$ (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 $\varepsilon$ (negative values during the daytime, and positive values during the nighttime) are seen again in the Southern US. In the spring, opposing patterns of $\varepsilon$ (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 in situ (IS) inversion and the inversion using the land-nadir, land-glint (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 ocean-glint (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. 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 dataset 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.

## 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 $CO_2$in 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. The authors declare, to their knowledge, no conflicts of interest with the submission of this manuscript. 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://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1593`). The WRF-Chem simulations used in the manuscript can be found online (`https://doi.org/10.26208/49kd-b637`), and FLEXPART-WRF simulations can be found online (`ftp://evs2ftp.ornl.gov/Inversion_Models/Influence_Function/`).


## References


Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., ... Tans, P. P. (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.

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.

Brioude, J., Arnold, D., Stohl, A., Cassiani, M., Morton, D., Seibert, P., ... Wotawa, G. (2013). The lagrangian particle dispersion model flexpart-wrf version 3.1. *Geoscientific Model Development*, *6*(6), 1889–1904.

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.

Ciais, P., Dolman, A. J., Bombelli, A., Duren, R., Peregon, A., Rayner, P. J., ... Zehner, C. (2014). Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system. *Biogeosciences*, *11*(13), 3547–3602.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., ... Jones, D. B. A. (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.

Cui, Y. Y., Brioude, J., Angevine, W. M., Peischl, J., McKeen, S. A., Kim, S.-W., ... Trainer, M. (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.

Cui, Y. Y., Brioude, J., McKeen, S. A., Angevine, W. M., Kim, S.-W., Frost, G. J., ... Trainer, M. (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.

Cui, Y. Y., Henze, D. K., Brioude, J., Angevine, W. M., Liu, Z., Bousserez, N., ... Trainer, M. (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.

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. *ORNL DAACm, Oak Ridge, Tennessee, USA.*, *https://doi.org/10.3334/ORNLDAAC/1593*.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., ... Yoshimizu, J. (2017). The orbiting carbon observatory-2: first 18 months of science data products. *Atmospheric Measurement Techniques*, *10*(2), 549–563.

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).

Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., ... Baker, I. (2019a). 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*(24), 14325-14346.

Feng, S., Lauvaux, T., Keller, K., Davis, K. J., Rayner, P., Oda, T., & Gurney, K. R. (2019b). A road map for improving the treatment of uncertainties in high-resolution regional carbon flux inverse estimates. *Geophysical Research Letters*, *46*(22), 13461-13469.

Hayes, D. J., Turner, D. P., Stinson, G., McGuire, A. D., Wei, Y., West, T. O., ... Cook, R. B. (2012). Reconciling estimates of the contemporary north american 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.

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.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049.

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).

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, 2020. https://doi.org/10.25925/39m3-6069*.

Keller, K., McInerney, D., & Bradford, D. F. (2008). Carbon dioxide sequestration: how much and when? climatic change. *Climatic Change*, *88*(8), 267.

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*, *2020*, 1–53.

Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., ... Eldering, A. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 el niño. *Science*, *358*(6360). doi: 10.1126/science.aam5690

Masson-Delmotte, V., Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., Shukla, P., ... Waterfield, T. (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. *In press*.

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 co2mole fractions detected with a tower-based network in the u.s. upper midwest. *Journal of Geophysical Research: Biogeosciences*, *117*(G1).

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.

Ogle, S. M., Davis, K., Lauvaux, T., Schuh, A., Cooley, D., West, T. O., ... Denning, A. (2015). An approach for verifying biogenic greenhouse gas emissions inventories with atmospheric CO2 concentration data. *Environmental Research Letters*, *10*(3), 034012.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., ... Hayes, D. (2011). A large and persistent carbon sink in the world's forests. *Science*, *333*(6045), 988–993.

Rayner, P. (2020). Data assimilation using an ensemble of models: a hierarchical approach. *Atmospheric Chemistry and Physics*, *20*(6), 3725–3737.

Thompson, R. L., Broquet, G., Gerbig, C., Koch, T., Lang, M., Monteil, G., ... Rödenbeck, C. (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.

Tian, H., Lu, C., Ciais, P., Michalak, A., Canadell, J., Saikawa, E., ... Wofsy, S. (2016). The terrestrial biosphere as a net source of greenhouse gases to the atmosphere. *Nature*, *531*, 225–228.

Wei, Y., Shrestha, R., Pal, S., Gerken, T., McNelis, J., Singh, D., ... Davis., K. (in prep for this issue). The atmospheric carbon and transport (act) – america datasets: Description, management, and delivery. *Earth and Space Science*.

# Supporting Information for "Evaluation of inverse estimates of North American net ecosystem exchange of $CO_2$ from different observing systems using ACT-America airborne observations"

Y. Y. Cui[1], A. R. Jacobson[2,3], S. Feng[1], D. Wesloh[1], T. Gerken[1,4], Z. R. Barkley[1], K. Keller[5,7], D. Baker[6], and K. 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 School of Integrated Sciences, James Madison University, Harrisonburg, VA, USA

[5]Department of Geosciences, The Pennsylvania State University, University Park, PA, USA

[6]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

[7]Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA

This supplement includes six figures as the supporting analysis. Detailed descriptions are contained in Text S1.

## Contents of this file

1. Text S1
2. Figures S1 to S7

## Introduction

---

**Text S1.** The averaged $CO_2$ NEE flux maps corresponding to each campaign time period are shown in Figure S1, to illustrate the estimation of $CO_2$ NEE flux from the different products. In the study, we evaluated the four flux products for the ecoregions in the three domains with the day/night intervals during the campaign. We only evaluated the spatially and temporally resolved fluxes which were constrained by the aircraft campaign.

Figure S2 shows the comparisons between background values of $CO_2$ mole fractions along flight tracks ($>$ 4,000 m MSL) and the ACT-America measured $CO_2$. The difference ($>$ 4,000 m MSL) between the background values and the ACT $CO_2$ measurements up to 1.1, 0.45, 0.65, and 0.37 ppm, for the four campaigns, respectively. The differences are much smaller than the ambient $CO_2$ mole fractions.

In this study, the suitable regions restricted by $H^T H$ are analyzed. For the computational efficiency, we only focused on the grid cells associated with the large values of influence functions. Specifically, the grid cells associated with the values of the accumulated influence functions (each campaign) that are greater than the $65^{th}$ percentile of the accumulated influence functions are considered. The corresponding maps are shown in Figure S3.

Figure S4 shows the eco-regions are considered for each campaign in the study. The eco-regions are defined in CarbonTracker system (`https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2019B_doc.php#tth_sEc9`). ACT-America covers 17 different ecosystems. We aggregate the grid cells by ecosystems in each of the three domain (shown in Figure 3) to obtain the grid clusters used in our analysis.

The spatial map of $\varepsilon$ values, at the seasonal level, refer to the daily, daytime and nighttime flux estimation from the four flux products are shown in Figure S5, Figure S6 and Figure S7, respectively.

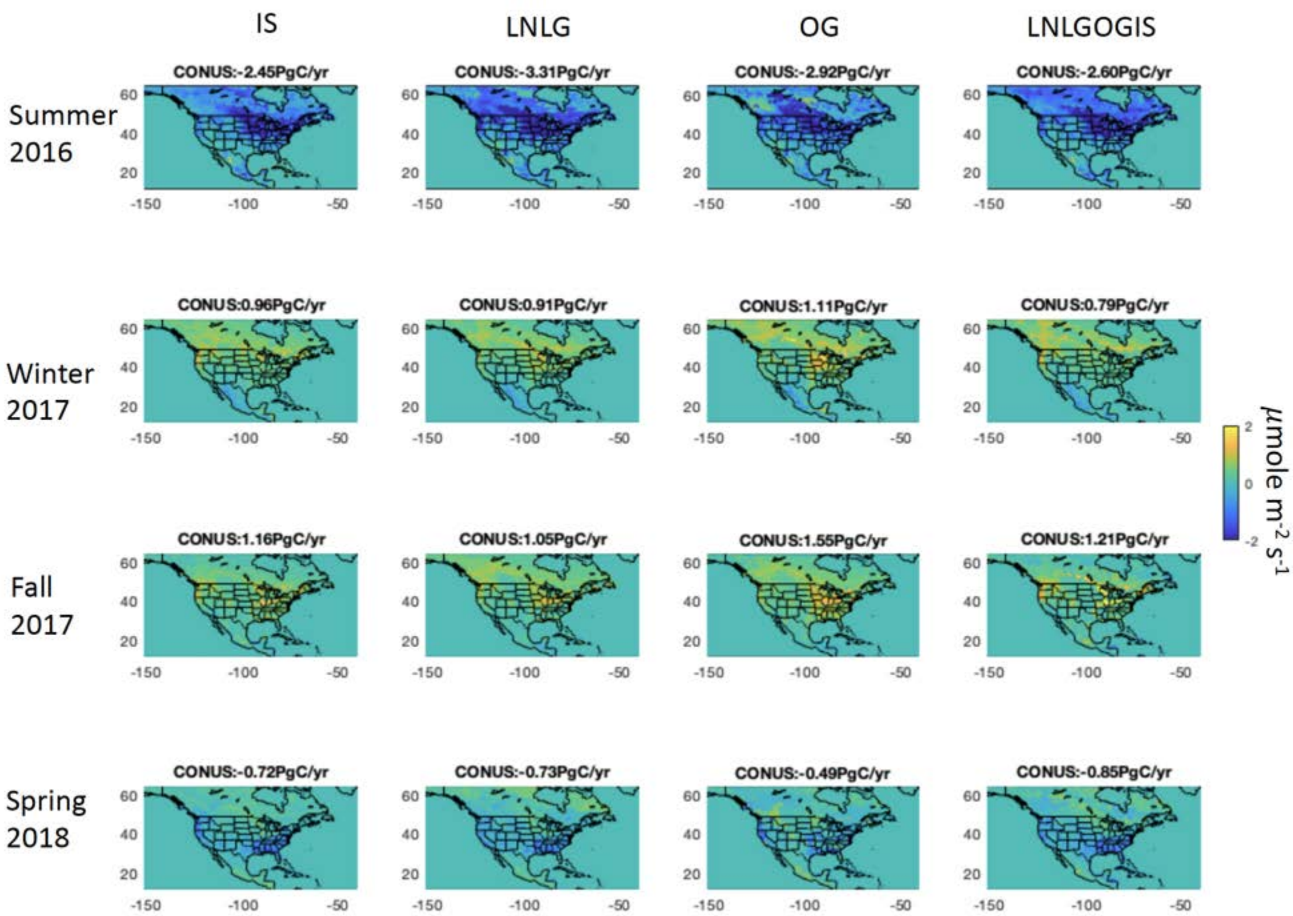



**Figure S1.** Average $CO_2$ NEE from four flux products corresponding to the ACT-America campaign periods, respectively. ACT-Summer2016: July-August; ACT-Winter2017: January-March; ACT-Fall2017: September-November; ACT-Spring2018: April-May.

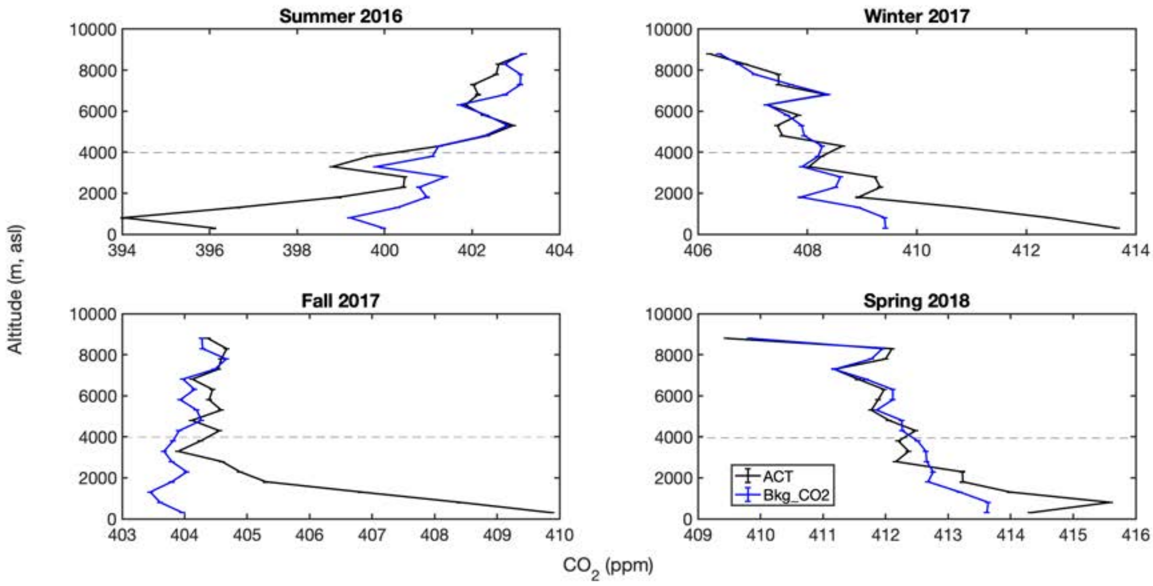



**Figure S2.** Vertically (every 500m interval) averaged $CO_2$ mole fractions are shown from the ACT-America measurements ("ACT") as well as the background values simulated by WRF-Chem (see Feng et al., 2019a, "Bkg_$CO_2$"), for each campaign. All flight data are included except take-off and landing portions.

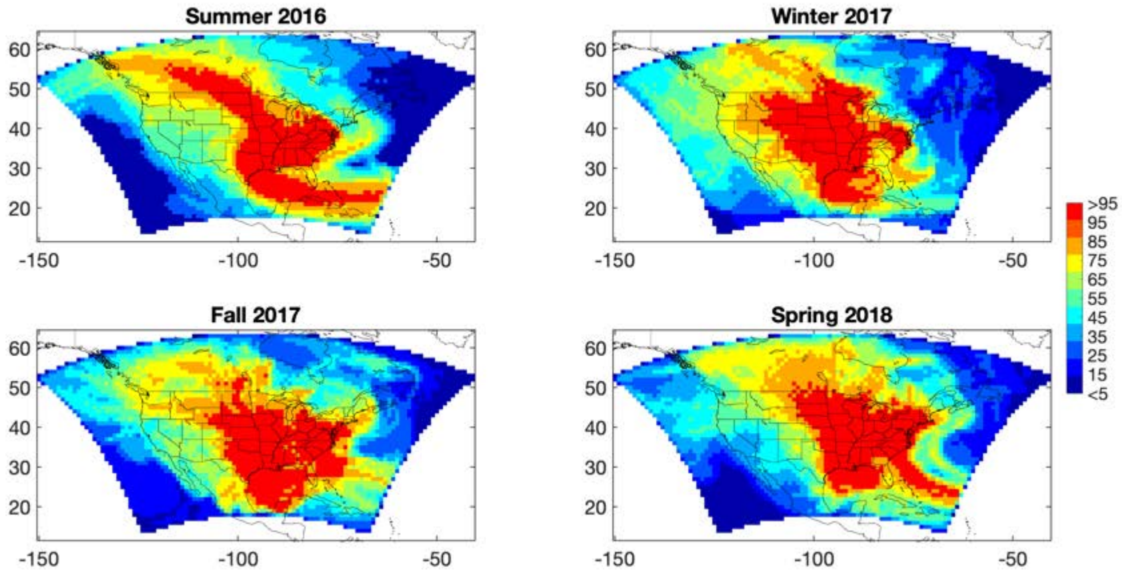
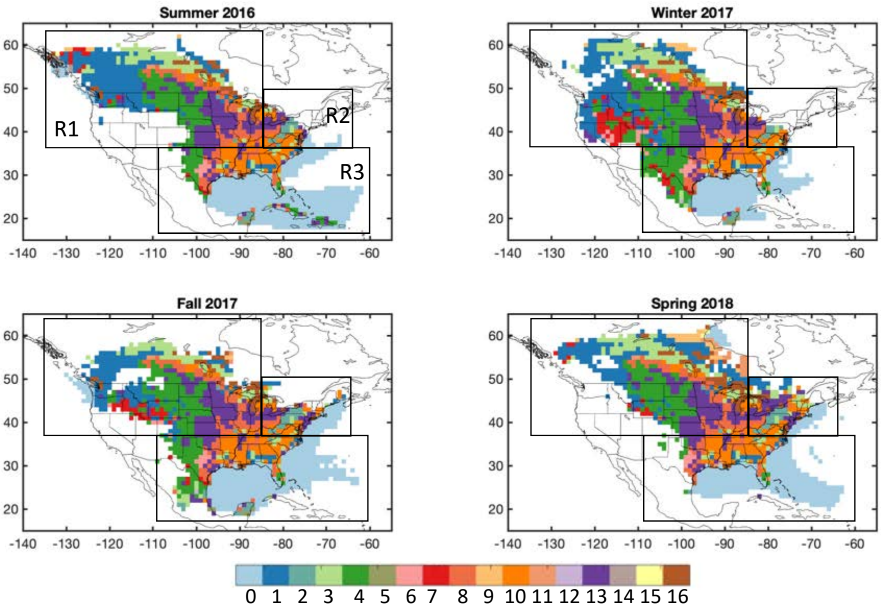



**Figure S3.** The values of the accumulated influence functions colored by percentile levels, for each campaign.




0. ocean; 1. Conifer Forest; 2. Broadleaf Forest; 3. Mixed Forest; 4. Crass/Shrub; 5. Tropical Forest; 6. Shrub/Woods; 7. Semitundra; 8. Fields/Woods/Savanna; 9. Northern Taiga; 10. Forest/Field; 11. Wetland; 12. Shrub/Tree/Suc; 13. Crops; 14. Conifer Snowy/Coastal; 15. Wooded Tundra, 16. Water

**Figure S4.** The maps of ecoregions are considered in the study for each campaign.

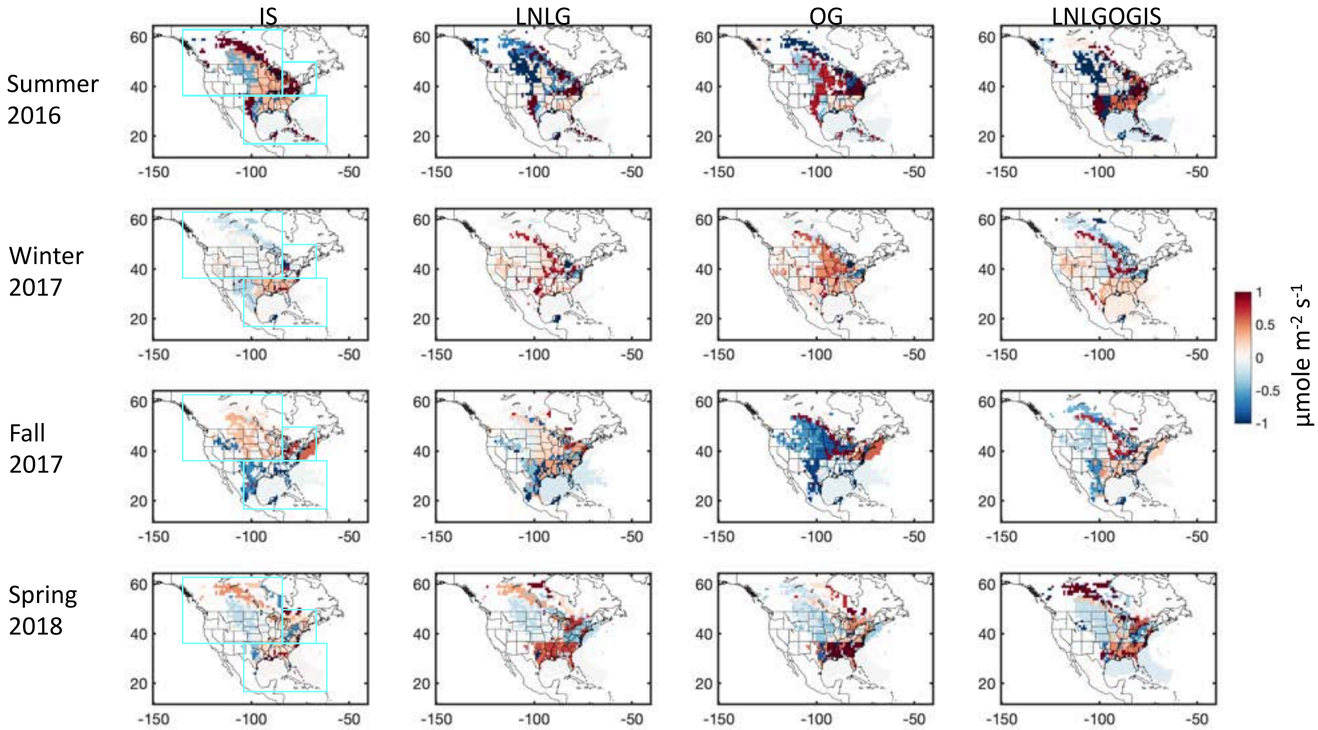



**Figure S5.** The spatial map of $\varepsilon$ values refer to the daily mean flux estimation from the four flux products shown for each ACT-America campaign, respectively.

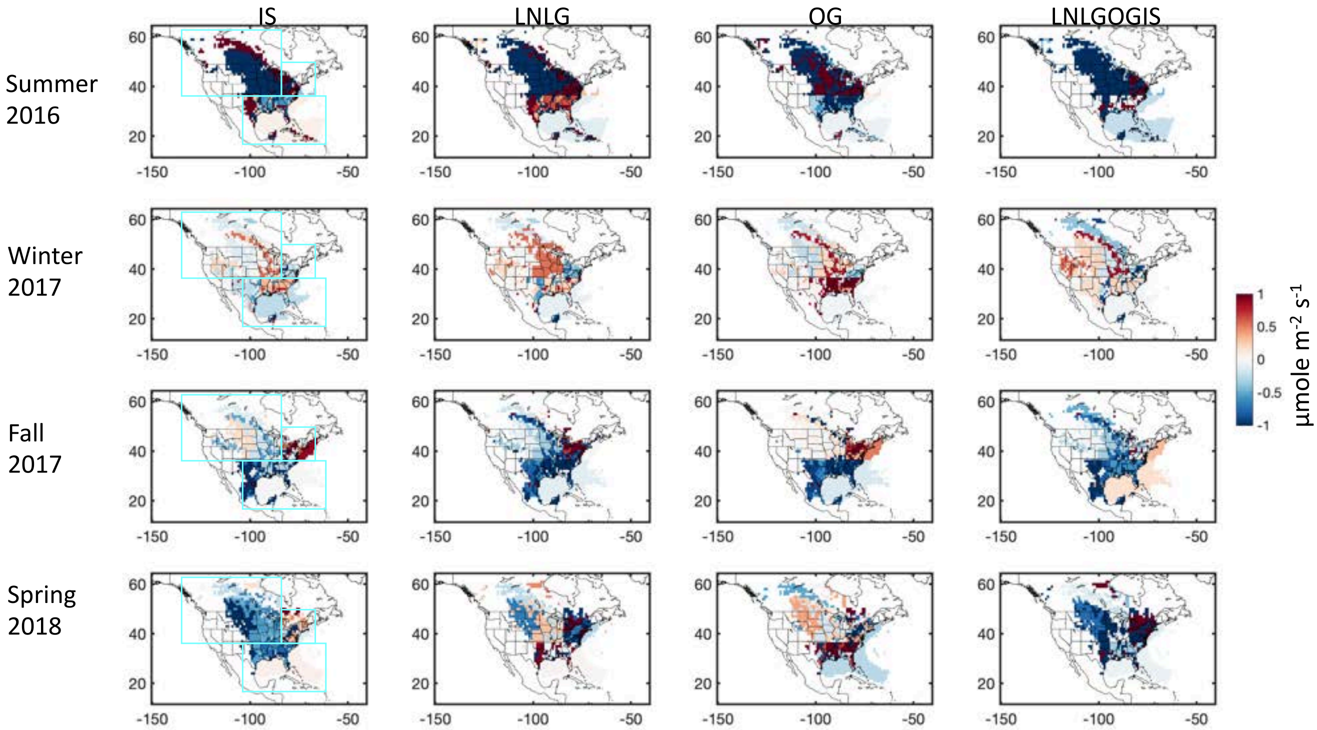



**Figure S6.** The spatial map of $\varepsilon$ values refer to the daytime flux estimation from the four flux products shown for each ACT-America campaign, respectively.

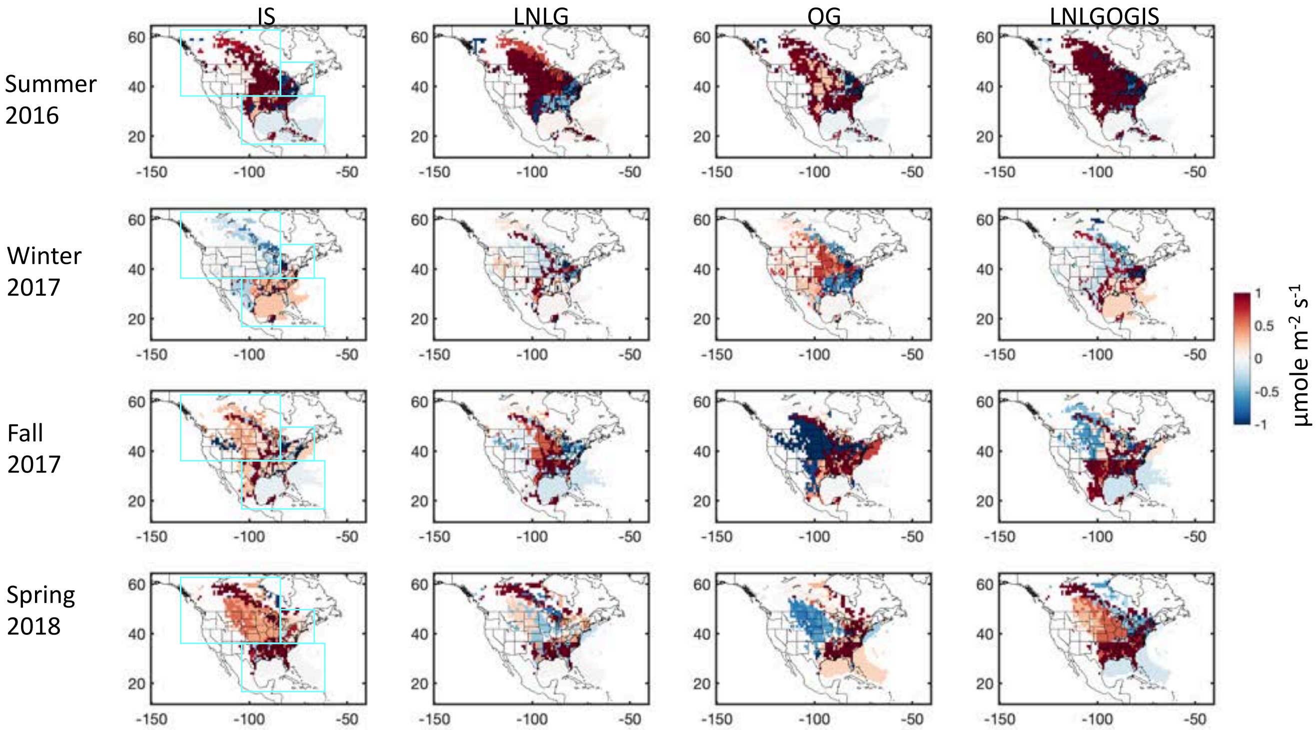



**Figure S7.** The spatial map of $\varepsilon$ values refer to the nighttime flux estimation from the four flux products shown for each ACT-America campaign, respectively.