Joint CO_2 Mole Fraction and Flux Analysis Confirms Missing Processes in CASA Terrestrial Carbon Uptake over North America

Sha Feng[1], Thomas Lauvaux[1,2], Christopher A. Williams[3], Kenneth J. Davis[1,5], Yu Zhou[3], Ian Baker[4], Zachary R. Barkley[1], Daniel Wesloh[1]

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

[2]Laboratoire des Sciences du Climat et de l'Environnement, CEA, CNRS, UVSQ/IPSL, Université Paris-Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette cedex, France

[3]Graduate School of Geography, Clark University, Worcester, MA, USA

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

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

Corresponding author: Sha Feng (sfeng@psu.edu)

Key Points:

- In summer, uncertainty in biogenic fluxes is the largest source of variability in atmospheric CO2.
- Atmospheric CO_2 mole fraction and eddy-flux measurements confirm a sink over North America not represented in the CASA terrestrial model.
- Eddy-flux measurements suggest weak daytime uptake and excessive nighttime respiration in CASA for summer.

Abstract

Terrestrial biosphere models (TBMs) play a key role in the detection and attribution of carbon cycle processes at local to global scales and in projections of the coupled carbon-climate system. TBM evaluation commonly involves direct comparison to eddy-covariance flux measurements. This study uses atmospheric CO_2 mole fraction ($[CO_2]$) measured *in situ* from aircraft and tower, in addition to flux-measurements from summer 2016 to evaluate the CASA TBM. WRF-Chem is used to simulate $[CO_2]$ using biogenic CO_2 fluxes from a CASA parameter-based ensemble and CarbonTracker version 2017 (CT2017) in addition to transport and CO_2 boundary condition ensembles. The resulting "super ensemble" of modeled $[CO_2]$ demonstrates that the biosphere introduces the majority of uncertainty to the simulations. Both aircraft and tower $[CO_2]$ data show that the CASA ensemble net ecosystem exchange (NEE) of CO_2 is biased high (NEE too positive) and identify the maximum light use efficiency E_{max} a key parameter that drives the spread of the CASA ensemble. These findings are verified with flux-measurements. The direct comparison of the CASA flux ensemble with flux-measurements indicates that modeled $[CO_2]$ biases are mainly due to missing sink processes in CASA. Separating the daytime and nighttime flux, we discover that the underestimated net uptake results from missing sink processes that result in overestimation of respiration. NEE biases are smaller in the CT2017 posterior biogenic fluxes, which assimilates observed $[CO_2]$. Flux tower analyses, however, reveal an unrealistic overestimation of nighttime respiration in CT2017.

1 Introduction

The Paris agreement calls for "a balance between anthropogenic emissions by sources and removals by sinks of greenhouse gases in the second half of this century" to limit the increase in global average temperature below 2°C above preindustrial levels (UNFCCC, 2015). Monitoring and tracking this commitment require accurate quantification of terrestrial carbon exchange with the atmosphere. The rapid increase of carbon dioxide (CO_2) has been partially offset by natural biogeochemical processes, including uptake by terrestrial and oceanic ecosystems (Friedlingstein et al., 2019). Understanding terrestrial carbon exchange is crucial for assessing biosphere-atmosphere interactions, for diagnosing terrestrial ecosystem contributions to the global coupled carbon-climate system and, ultimately, for reducing the uncertainty in climate projections (Friedlingstein et al., 2014).

Though the net global flux of CO_2 to the atmosphere is well constrained (Tans and Conway, 2005; Tans et al., 1990), regional-to-continental biogenic CO_2 fluxes are not well characterized in current carbon estimation approaches (Crowell et al., 2019).Two approaches are commonly taken to quantify biogenic CO_2 fluxes at the continental scale. "Top-down" approaches use an optimization process by which atmospheric CO_2 mole fraction ($[CO_2]$) measurements combined with an atmospheric transport model are used to constrain the *a priori* estimation of the spatial and temporal distribution of biologic CO_2 fluxes (Enting et al., 1995). Here we use "[]" to represent atmospheric mole fractions and distinguish it from flux space, and this convention is used hereafter. Determining and reducing the uncertainty associated with transport and with flux priors have been the focus of the atmospheric inversion community (Baker et al., 2006; Gurney et al., 2002).

By contrast, "bottom-up" approaches involve direct simulation of biogenic CO_2 fluxes using inventories (Houghton et al., 1996), upscaled flux measurements (e.g., Baldocchi et al., 2001; Jung et al., 2020), or terrestrial biosphere models (TBMs), without any constraint from atmospheric mole fraction observations. TBMs simulate surface CO_2 fluxes from site level to global scale as they integrate ecological and meteorological drivers (Fung et al., 1987). They simulate the land carbon component in Earth system models used for climate projections and commonly used as the *a priori* estimations for "top-down" approaches. To benefit from both ecosystem measurements and atmospheric mole fractions, joint assimilation systems have also been developed (Kaminski et al., 2002). The optimization procedure relies on an adjoint model of the biogeochemical processes such as the Biosphere Energy-Transfer Hydrology (BETHY) model included in the Carbon Cycle Data Assimilation System (CCDAS) at large scales (Knorr, 2000). Such a system was enhanced to assimilate satellite ecosystem products available over the globe (Kaminski et al., 2012) or eddy-flux and mole fraction measurements over continents (Koffi et al., 2013). These approaches optimize a number of model parameters in part of the underlying TBMs, and have been applied using remote $[CO_2]$ from towers or satellites (e.g., Scholze et al., 2019). On sub-continental scales, atmospheric $[CO_2]$ require fine-resolution models to simulate the complex atmospheric dynamics (e.g., Feng et al., 2019a) and dense tower networks (Andrews et al., 2014) combined with eddy-flux tower networks (e.g. AmeriFlux) covering a wide array of ecosystems and climatic zones.

TBMs have been shown to vary widely in their projections of terrestrial CO_2 sink strengths, not only in magnitude but even in sign (Baker et al., 2006; Gurney et al., 2002). Huntzinger et al. (2011) evaluated flux variability from four TBMs over North America and the potential impact on the inversion results. They found that the diurnal variability in surface fluxes

within the near field of tower [CO_2] observations appear to have a significant impact on the high-frequency variations in the atmospheric data, and, thus, the inversion needs to adjust the temporal (and spatial) variability of the prior fluxes. Feng et al. (2019a) compared the modeled [CO_2] errors attributed to biogenic CO_2 fluxes, fossil fuel emissions, atmospheric transport, and large-scale boundary inflow from daily to annual timescales and discovered that the biogenic CO_2 fluxes dominate the model errors across timescales, implying that [CO2] observations hold promise for evaluating and improving TBMs. Improving TBMs to generate accurate and spatially explicit surface biogenic CO_2 fluxes at high resolution in time and space remains an urgent need for diagnosis, attribution, and projection of terrestrial carbon dynamics, as well as for "top-down" analysis systems that rely on TBMs for a priori surface flux estimation.

Several multiple model intercomparison projects have been conducted to characterize or synthesize current understanding of land–atmosphere carbon exchange and inform the uncertainty or confidence surrounding projections of future exchange and feedbacks with the climate system, such the Multi-Scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP; Huntzinger et al., 2013) Trends in Net Land-Atmosphere Carbon Exchange(TRENDY1; Sitch et al., 2008), the regionally focused Large Scale Biosphere Atmosphere-Data Model Intercomparison Project (LBAMIP2; Gonçalves et al., 2013), and the International Land-Atmosphere Benchmarking Project (ILAMB; Luo et al., 2012). These model intercomparison projects were built upon the protocols that specify standard model inputs, simulations and simulation setup procedures. These model intercomparison exercises have been used to explore the uncertainty in model simulations that arises from internal variability, boundary conditions, and parameter values for structural uncertainty from different model fluctuations (Schwalm et al., 2015). However, due to the complexity of TBMs, it is challenging to trace errors in individual models to misrepresentation of specific ecological processes or inappropriate model parameters through these model intercomparison projects. This merits deep exploration of just one modeling framework but with a perturbed parameter ensemble.

Recently, Zhou et al. (2020) introduced an ensemble of biogenic CO_2 fluxes simulated by the Carnegie-Ames-Stanford-Approach (CASA) biosphere model for North America at the resolutions of 5 km for North America and ~500 m for the U.S. CONUS region from 2003 to 2019 by perturbing three model parameters – maximum light use efficiency E_{max}, optimal temperature of photosynthesis T_{opt}, and temperature response of respiration Q_{10} – on the basis of the CASA biome types. Those parameters were chosen as a result of a sensitivity test of simulated biogenic CO_2 fluxes to a series of CASA parameters. Furthermore, the range of E_{max} values was determined at the ecosystem level by comparison with eddy covariance measurements of net CO_2 flux. Zhou et al. (2020a) illustrated that the pruned L2 ensemble has good agreement with the flux data and outperforms many other TBMs at diurnal and annual scales, even while having too wide a spread in comparison with flux measurements.

Atmospheric [CO_2] measurements provide another opportunity for evaluating modeled biogenic CO_2 fluxes, as evidenced by the fact that continental [CO_2] gradients are mainly attributed to the biosphere (e.g., Feng, et al., 2019a; Feng, et al., 2019b). The major differences between mole fraction and flux measurements fall in the size of the surface influence area, or footprint, that influences a given measurement, and also the upwind memory of the samples. The size of a flux tower footprint is only about 1-km (e.g., McCaughey et al., 2006), and the measurements carry nearly instantaneous information of surface fluxes. Owing to these two factors, the flux measurements do not directly represent regional to continental fluxes, as the local fluxes captured may not be representative of broader scale patterns. One objective of this

study is to explore the coherence between CO_2 flux and mole fractions with respect to modeled biogenic CO_2 flux evaluations. The CO_2 flux measurements selected for this study are from the AmeriFlux network (https://ameriflux.lbl.gov/), which has more than 150 active flux sites sampling a wide range of sites from the Amazonian rainforests to the North Slope of Alaska.

An *in situ* [CO_2] tower typically has a footprint of hundreds of kilometers (e.g., Gloor et al., 2001; Sweeney et al., 2015) and carries the integrated atmospheric [CO_2] signals from day and night. Aircraft measurements have even larger footprints with broad spatial sampling. TBM-modeled CO_2 fluxes can be evaluated against [CO_2] measurements by means of Lagrangian and forward Eulerian transport modeling. Both are subject to transport model errors (e.g., Pillai et al., 2012). In Lagrangian modeling, the modeled biogenic CO_2 mole fractions ([CO_2bio] hereafter) can be directly calculated by convolving biogenic CO_2 fluxes with the influencing areas of the [CO_2] measurements that are simulated by a Eulerian transport model (Uliasz et al., 1994). The difficulty of this approach is to define the observed [CO_2bio] due to mixed signals in the [CO_2] measurements (e.g., Ogle et al., 2015). In Eulerian transport modeling, the CO_2 transport is treated as a passive tracer (Sarrat et al., 2007). The modeled total [CO_2] is the sum of biogenic, fossil fuel, oceanic, fire CO_2 components in conjunction with boundary conditions as described in Feng, et al. (2019a; 2019b). The modeled error therefore can be from any or multiple components in addition to model transport (Feng, et al. (2019a; 2019b). Note that these model error sources are also of concern in the "top-down" estimation. Another objective of this work is to explore to what degree the atmospheric [CO_2] data can be used to evaluate TBMs. Here we use CASA (Potter et al., 1993) for demonstration and adopt the Eulerian transport model WRF-Chem (Grell et al., 2005; Skamarock et al., 2008) to serve this objective.

In this study, we employ both ground-based and aircraft *in situ* [CO_2] data for the model evaluations. The NASA-funded Atmospheric Carbon and Transport (ACT)-America project was designed to improve the CO_2 and methane (CH_4) flux estimates by reducing transport and flux uncertainties. Two aircraft measured atmospheric CO_2, CH_4, and other gas species over Mid-Atlantic, Mid-West, and Southern Gulf regions in fair and frontal weather regimes. Typical flights encompassed 4-6 hours of midday conditions, encompassing 400-800 km in the horizontal and altitudes ranging from 300 m to 9000 m above ground level. Two aircraft flew together, collecting two to four vertical levels (level legs), often stacked one above the other, and typically about 8-12 vertical profiles per flight day. Pal et al. (2020) found that large horizontal and vertical gradients of [CO_2] exist across frontal boundaries based on the data collected from the summer 2016 campaign. The cross-frontal [CO_2] contrasts are greatest in the atmospheric boundary layer (ABL), ranging from 5 to 30 ppm, while the contrasts are about 3-5 ppm in the free troposphere (FT). In the vertical dimension, higher [CO_2] appears in the FT than in the ABL in the cold sector while the opposite pattern appears in the warm sector. Averaged ABL-to-FT [CO_2] differences can be about 12 and -6 ppm in the warm and cold sectors, respectively. These unique flights were designed to be highly sensitive to the seasonal magnitudes of regional-scale carbon fluxes and to provide broad spatial coverage that cannot be obtained with the current long-term observing network. The third objective is to explore to what degree the aircraft [CO_2] measurements and tower measurements yield consistent evaluations of the modeled fluxes. Merging these two [CO_2] observing systems lends more confidence to our conclusions regarding the modeled fluxes and provides a test of the two observing systems.

The three objectives of this study are addressed through comparing the WRF-Chem [CO_2] simulated using the CASA L2 flux ensemble members to the airborne and tower-based

182 [CO_2] data. The WRF-Chem model setup and methods are described in Section 2. All the data
183 used in this work are described in Section 3. The modeled and observed [CO_2], CASA parameter
184 constraints, and causes of the model errors are illustrated in Section 4. Sections 5 and 6 are
185 discussion and conclusions, respectively.

186 **2 Materials and Methods**

187 2.1 Transport model setup

188 All transport model simulations use WRF-Chem version 3.6.1 (Grell et al., 2005;
189 Skamarock et al., 2008). The modification made to transport greenhouse gases as passive tracers
190 (Lauvaux et al., 2012) allows us to carry ensemble tracers for biogenic CO_2 fluxes (Section 2.2)
191 and [CO_2] boundary conditions (Section 2.4) in one transport run (Section 2.3). In each transport
192 run, WRF-Chem carries 39 CO_2 tracers for the ensembles of boundary conditions and biogenic
193 fluxes, ocean, fossil fuel, and biomass burning fluxes. The modeled total [CO_2] is the sum of a
194 boundary condition, a biogenic flux, and the oceanic flux, fossil fuel emission, and biomass
195 burning CO_2 tracers as described in Feng et al. (2019b, 2019a). The [CO_2] boundary condition
196 tracers are propagated into WRF-Chem hourly with the consideration of the conservation of
197 mass (Butler et al., 2020). Five global CO_2 inversion/reanalysis systems are used for [CO_2]
198 boundary conditions (Section 2.4) and 29 biogenic CO_2 fluxes are used for the biogenic CO_2
199 tracers (Section 2.2). The CO_2 oceanic flux, fossil fuel emission, and biomass burning are taken
200 from CarbonTracker version 2017 (CT2017; Peters et al., 2007).
201 The same model configurations in Feng, et al. (2019a, 2019b) are used except for the
202 meteorological initial and boundary conditions. In this study, we used the ERA5 reanalysis
203 (Hersbach et al., 2020)and benchmark the model transport by nudging the WRF-Chem
204 simulation to ERA5. The wind evaluations show that nudging clearly improves model transport
205 (see Text S1 in Supporting Information). A suite of transport runs is created for uncertainty
206 quantification (Section 2.3). Choices of the model physics schemes are summarized in
207 Table *1*. All WRF-Chem simulations have a horizontal resolution of 27 km ×27 km for
208 the period from July 18 to August 28, 2016 covering the ACT-America summer 2016 aircraft
209 campaign hourly.

210 2.2 Biogenic CO_2 flux ensemble: CASA and CT2017

211 We include 29 biogenic CO_2 fluxes in each transport run as separate tracers: the 27-
212 member CASA L2 net ecosystem exchange (NEE) ensemble, the mean of the CASA NEE
213 ensemble, and the CT2017 posterior biogenic CO_2 flux.

214 The CASA ensemble members (Zhou, et al., 2020b) were generated by perturbing the
215 maximum light use efficiency (E_{max}), optimal temperature of photosynthesis (T_{opt}), and
216 temperature response of respiration (Q_{10}) with the consideration of the biome types in CASA, as
217 described in Zhou et al.(2020a) . These three perturbed parameters were determined to dominate
218 the sensitivity of modeled CO_2 fluxes to the model parameters according to an Extended Fourier
219 Amplitude Sensitivity Testing analysis. The initial range for each parameter was broadly
220 sampled for the L1 ensemble. Parameter ranges were subsequently narrowed to those consistent
221 with AmeriFlux data, resulting in a L2 ensemble. CASA simulates gross primary productivity
222 (GPP), total ecosystem respiration (R_e), and NEE at monthly resolution. The monthly GPP and
223 R_e fluxes were then downscaled to 3-hourly resolution using the Olsen and Randerson (2004)
224 method. Two sets of flux products are included in the official release of the CASA flux

ensemble, one at 500-m resolution covering the U.S. CONUS region and the other at 5-km covering a broader swath of North America. We used the L2 5-km CASA ensemble for this study. Details about the CASA ensemble products can be found in Zhou, et al. (2020a, 2020b)

Unlike CASA that directly simulates biogenic CO_2 fluxes ("bottom-up"), CT2017, a "top-down" flux estimate, optimizes the *a priori* fluxes (CASA; Potter et al., 1993) by assimilating observed $[CO_2]$ (Peters et al., 2007). CT2017 global 3-hourly posterior biogenic fluxes at $1° \times 1°$ are used in this study. Note that, unlike the CASA ensemble used spun up to equilibrium, the CT2017 biogenic flux prior from CASA include an assumed global terrestrial biospheric sink of 2 PgC/yr.

Both CASA and CT2017 biogenic CO_2 fluxes used have a 3-hourly temporal resolution. To downscale to hourly fluxes for the transport model simulations, values in each 3-hourly flux file are repeated hourly over the period they represent

2.3 Transport ensemble

The transport ensemble runs are generated using the combination of multiple physical parameterizations and the stochastic kinetic energy backscattering scheme (SKEBS; Berner et al., 2009; Shutts, 2005). Through this combination, model meteorological initial conditions and physics were perturbed at the same time, introducing transport uncertainty due to the model dynamics and physics. Previous studies demonstrated that model errors can best captured by a combination of by the combination of multi-physics and SKEBS (Berner et al., 2011, 2015) as opposed to a single perturbation scheme. Feng, et al. (2019b, 2019a) for the first time applied this combination to simulations of $[CO_2]$ and demonstrated that a relatively small transport ensemble can represent model $[CO_2]$ transport uncertainty. We use the root-mean-square deviation (RMSD) of the simulated $[CO_2]$ transport ensemble from the ensemble mean to represent transport uncertainty. Similar calculations are used to estimate biogenic flux and boundary condition uncertainty.

Here we varied the land surface models (LSMs) and planetary boundary layer (PBL) schemes in WRF-Chem based on the sensitivity study conducted by Díaz-Isaac et al. (2018). The ensemble members are (1) Mellor-Yamada Nakanishi and Niino Level 2.5 (MYNN 2.5) PBL scheme with Noah LSM, (2) Mellor-Yamada-Janjic PBL scheme with RUC LSM, (3) Yonsei University PBL scheme with five-layer thermal diffusion LSM. The associated surface layer parameterizations are MYNN, Eta, and MM5, respectively. A summary of transport ensemble members can be found in Table 2. Note that these transport ensemble runs are free runs, not nudged to ERA5.

2.4 Boundary condition ensemble

A suite of optimized $[CO_2]$ from five global inversion systems are collected for this study. They are CT2017 (Peters et al., 2007, with updates documented at http://carbontracker.noaa.gov), TM5 as described in Basu et al. (2016), GEOS-Chem developed in the Carbon Monitoring System (CMS; Liu et al., 2014), GEOS-Chem as described in Schuh et al. (2019), and PCTM as described in Barker et al. (2004). These global modeled $[CO_2]$ fields are propagated into WRF-Chem hourly as separate tracers following Butler et al. (2020).

2.5 Footprint analysis

To understand the fluxes influencing the observations used in this study, the Lagrangian Particle Dispersion Model (LPDM; Uliasz et al., 1994) is used to create surface influence functions (footprints) for each flight in the Summer 2016 ACT aircraft campaign (see description in Section 3.1). Particles are released along the time of each aircraft transect within the boundary layer, and are traced backwards in time over a two-week period using meteorology provided by the WRF nudged-transport simulation. The influence by the surface is represented by the number of particles interacting with the surface grid, i.e., below 50m above ground, are summed up, providing a temporal and spatial function that relates the signal observed by the aircraft to the surface fluxes responsible for that signal (Seibert & Frank, 2004).

In addition to aircraft [CO_2] measurements, we also use a subset of the ground-based [CO_2] tower measurements from the NOAA ObsPack GlobalViewPlus package (Cooperative Global Atmospheric Data Integration Project, 2019; see Section 3.2). We create footprints for the towers by releasing particles at 21 UTC backward in time over two-week period using the nudged transport from July 18 to August 28, 2016 across 23 different tower sites in the US. The selection of 21 UTC limits the tower observations to well-mixed ABL conditions, minimizing model transport uncertainty.

2.6 Model evaluation metrics

To demonstrate the importance of the uncertainty in the biosphere, we first compare the contribution of biosphere, transport, and boundary conditions to the modeled [CO_2] uncertainty. We use the RMSD of a given component to illustrate the ensemble spread and associated uncertainty.

We focus on the performance of the individual biogenic CO_2 flux members through the biome-based model biases, which is determined by the CASA biome map and the footprint of the measurements. The uncertainties in modeled [CO_2] associated with the individual biogenic flux members are determined by the spread of the transport and boundary conditions ensembles.

3 Data

3.1 ACT-America aircraft data

Two aircraft, the NASA Langley Beech-craft B200 King Air and NASA Goddard Space Flight Center's C-130H Hercules aircraft, were used to collect high quality *in situ* measurements of greenhouse gases, other gas species, and meteorological fields over Mid-Atlantic (MA), Mid-West (MW), and South Gulf (South) regions of the United States. The flight dates and patterns can be found Table S1. In this study, we used ACT-America L3 Merged *in situ* Atmospheric Trace Gases and Flask Data (Davis et al., 2018). This product provides integrated measurements and metadata flag information, including flight pattern, airmass type, and boundary layer information at five-second intervals. More information about the ACT-America campaign and measurements can be found at https://actamerica.ornl.gov/. The nearest point interpolation is applied to extract modeled [CO_2] along the flight tracks.

3.2 NOAA ObsPack GlobalViewPlus [CO_2] product

We also use tower-based *in situ* [CO_2] data from the NOAA ObsPack GlobalViewPlus product (Cooperative Global Atmospheric Data Integration Project, 2019; see locations in **Figure *1***). Twenty-three tower locations (Table 3) from this data package were selected for investigations. The ObsPack product collects greenhouse gas data from providers around the globe and reformats the data into the ObsPack framework in support of carbon cycle modeling studies (Masarie et al., 2014). The [CO_2] data are first organized into hourly data, then nearest point interpolation is applied to extract modeled [CO_2] at the tower locations.

3.3 AmeriFlux CO_2 flux measurements

We also include CO_2 flux measurements to evaluate our findings regarding biogenic CO_2 flux members. The CO_2 flux data are obtained from the eddy covariance measurements from the AmeriFlux network (https://ameriflux.lbl.gov). Seventy-one flux tower sites from the domain of interest were used. The locations and information of the sites can be found in ***Figure 1*** and Table S2. We obtained a single estimate of NEE for each flux tower location from the non-gap-filled NEE values reported by AmeriFlux, with a preference for eddy-covariance measurements with a storage correction. Tower-measured NEE was averaged to three-hour intervals to match the time resolution of CASA and CT2017. No intervals were excluded if they had any reported data.

4 Results

4.1 Spatiotemporal variability of [CO_2]

We select five of the 25 research flights (**Figure *2***) to illustrate the typical flight patterns for fair and frontal weather regimes. As expected, both models and observations show large [CO_2] gradients in the ABL, leading to large variations in [CO_2bio]. Pal et al. (2020) reported that an elevated [CO_2] band was repeatedly observed along the cold frontal boundary, a feature also captured by the simulations. The 8/4/2016 frontal case shows a narrow, elevated [CO_2] band at the frontal boundary (~25 ppm difference in [CO_2] across the front). We examined this frontal case with a 3-km × 3-km cloud-resolving resolution model and showed that this elevated [CO_2] band has a maximum width of ~200 km and a length of over 800 km extending from northeastern Kansas to northeastern Iowa. **Figure *2*** shows that the frontal boundaries are associated not only with elevated [CO_2] but also with highly variable [CO_2bio]. More than 5 ppm RMSD of [CO_2bio], caused by variability among the CASA ensemble fluxes, appears in the ABL along the frontal boundaries. Enhanced by the baroclinic instability, the cold airmass on the west lifted the warm airmass aloft. ABL [CO_2] penetrates into the free troposphere along the frontal lifting. RMSD in [CO_2bio] greater than 3 ppm, an indicator of strong surface influence, reaches up to 3.5 km above sea level on 7/18 and 1.7 km on 8/4 (**Figure *2***). On the contrary, both [CO_2] and [CO_2bio] have less variability for the fair-weather cases.

We use the ABL [CO_2] observations to evaluate biogenic CO_2 fluxes since, as **Figure *2*** illustrates, these data are the most sensitive to variations in the biological CO_2 fluxes. We limit our work to ABL [CO_2] for the rest of this analysis. **Figure *3*** shows the averaged [CO_2] sampled by aircraft and simulations. Note that the modeled [CO_2] are from the transport-nudged simulation (described in Section 2.1), while the uncertainties are determined by spreads of the ensemble runs associated with different components described in Sections 2.2, 2.3, and 0. All ACT-America aircraft collected afternoon samples aiming at well-mixed ABL conditions. The

most outstanding feature overall is that the biosphere is the most uncertain component in the simulation for the fair and frontal weather regimes except for three cases in the South and one in the Midwest. The footprint analysis (Figure S4) shows that the Southern flights are mainly influenced by Gulf onshore flow and, thus had a limited flux footprint over land. The one Midwestern flight was an OCO-2 underflight in the western Dakotas where biological fluxes are small.

Observations tend to be encompassed by the ensemble model spread except on four flight days: 7/22, 7/25, 7/26, and 8/12, on which the model shows large discrepancy with observations. Our preliminary investigation indicates that the disagreements on 7/25 and 7/26 are mainly caused by unrealistically strong uptake to the west of the Appalachia area in the CT2017 biogenic fluxes; 8/12 is due to errors in the long-range transport.

In the following sections, we focus on investigating the coherence between aircraft and tower [CO_2] data and evaluating the model biases across the CASA ensemble members. CT2017 serves as the reference for this exercise. As an inversion product that is constrained by atmospheric data, we expect that CT2017 should agree well with observed [CO_2] even though the ACT-America aircraft data were not assimilated in CT2017. ***Figure 4*** shows the model skill in the [CO_2] simulations driven with different biogenic CO_2 fluxes. Note that both aircraft and tower [CO_2] data are kept at native temporal resolutions in the comparisons, reflecting the model performance in capturing the spatiotemporal variability of [CO_2] at the synoptic scale (no diel cycle due to only afternoon samples used). In general, both aircraft and tower [CO_2] comparisons show that skill of the CASA members is similar to CT2017 in capturing the variability in ABL [CO_2] as indicated by the similar correlations with observations, but better in capturing the magnitude of this variability. CT2017 overestimates the magnitude of variability in ABL [CO_2]. Tower and aircraft yield similar results.

Aligning the aircraft and tower comparisons together, the CASA-simulated [CO_2] tend to somewhat underestimate the variability of aircraft [CO_2] but overestimate that of tower [CO_2]. Given the different sampling strategies between aircraft and tower measurements, we hypothesize that the models have a tendency to underestimate the spatial variability of atmospheric [CO_2] but overestimate temporal variability at the synoptic scale.

Another feature worth noting is that the CASA-simulated [CO_2] are clustered into two groups, evident in both aircraft and tower [CO_2] comparisons. Given the same transport and boundary conditions were used, the clustering is likely driven by the difference in biogenic fluxes associated with the values of the three parameters used for generating the CASA flux ensemble.

4.2 Identification of the CASA key parameters

The modeled [CO_2] biases associated with individual CASA flux members and CT2017 with the uncertainty bounds that are determined by the transport and boundary condition ensembles are shown in Figure S6 and S7 for each observation from aircraft and tower comparison, respectively. Applying the footprint analysis to aircraft and tower measurements and the CASA-defined biome map, we obtain the model biases for each biome. We list the top 2 biomes with the most influence on each aircraft and tower measurements in Table 3 and Table 4, respectively. Croplands (CR), deciduous broadleaf forest (DB), and grasslands (GL) are the major biome types sampled by aircraft; CR, DB, GL, and evergreen needleleaf forest (EN) are

mainly sampled by towers (**Figure *5***). The samples primarily influenced by water bodies are removed from the results. Two aspects stand out in the comparisons, modeled [CO_2] from all the flux members are positively biased across all biome types, and one group of the CASA flux members has better agreement with the observations than others. These two aspects are illustrated consistently in the comparisons of both aircraft and tower measurements.

For the first aspect, we find that all flux estimates, including CT2017 and CASA ensemble, overestimate [CO_2]. This is also reflected in **Figure *3***. The model biases seem scaled with the degree of plant productivity, given that we find larger biases associated with CR and DB and smaller biases associated with EN and GL. Zhou et al. (2020) also reported that the monthly-averaged NEE of the CASA ensemble averaged over 13 years had a larger positive bias in CR and DB than EN and GL in summer. When comparing aircraft-based biases to tower-based biases, the WRF-Chem simulated [CO_2] is more positively biased. Assuming that spatial variability in [CO_2] is related to temporal variability, as towers and aircraft basically observe the same weather systems, the larger aircraft-based biases might be caused by different geographic sampling or by the enhanced variability in aircraft data (designed to sample frontal systems) leading to larger biases. As expected, CT2017 has better agreement with the observations overall since its fluxes have been optimized using atmospheric [CO_2] data. However, the fact that WRF-Chem performs better when coupled to optimized CT2017 biogenic CO_2 fluxes confirms that transport model differences remain much smaller than flux differences.

For the second aspect, we discover that distinct groups of CASA members reflect the three parameter values for maximum light use efficiency, E_{max}. The flux members that show the best agreement with the observations mostly have medium E_{max} values while the groups with low and high E_{max} values correspond to larger model biases. Both aircraft and tower measurements identify that E_{max} is the dominant parameter in the CASA ensemble, which is consistent with the sensitivity results of Zhou et al. (2020a).

The performance of the individual flux members are summarized by ranking them as a function of bias (**Figure *6***). Medium E_{max} (E_2) leads to better modeled [CO_2] across different biome types and observation platforms. The groups with low (E_1) and high (E_3) E_{max} tend to have similar biases. A few members with high E_{max} are ranked high (in the top one third), such as P27 and P25 for CR, P27 for DB, P27 for GL, and P27 for EN. For DB and EN, both E_2 and E_3 are assigned the same value (medium E_{max}) in CASA perturbation, explaining why some model members with high E_{max} value are ranked high. However, due to the impact of the long-range transport over upwind biomes (where E_2 not equal E_3), the model [CO_2] bias ranking is different for the DB's and EN's E_{max} groups.

Further investigating the top-performing 11 ensemble members according to Q_{10} values, (**Figure *7***). We find that in general, the flux members with low Q_{10} value (Q_{10}=1.2) are ranked high, followed with medium Q_{10} (Q_{10}=1.4), and then high Q_{10} value (Q_{10}=1.6). In contrast, no T_{opt}-driven grouping is visible (not shown). We conclude that Q_{10} plays a secondary role in CASA-simulated summer NEE.

In summary, the aircraft and tower data deliver consistent results. Given the multiple level-leg sampling and profiling strategies available in the aircraft data in addition to the large vertical gradients in [CO_2bio] that appear in dataset (**Figure *2***), there is potential to impose additional constraints on biogenic fluxes and mixing heights using the vertical gradients of

[CO_2]. Though outside the scope of this study, these vertical gradients will be examined in future work.

4.3 The causes of modeled [CO_2] biases

We unfold the model-tower comparisons and show the modeled [CO_2] biases for each [CO_2] tower we included in the previous analysis in **Figure *8***a. All of the CASA members lead to an overestimate of [CO_2] except at three towers: BAO, ETL, and OSI. The CASA flux members result in modeled [CO_2] that is similar to CT2017 for GL and EN sites with the exception of three sites in the South Gulf region (GCI01, GCI03, and GCI04). Variability in the TBM parameters have more impact on CR and DB where they have the larger sink strength and plant productivity. A similar pattern of biases across biome types can be seen in CT2017, though CT2017 shows slightly better agreement with the observations.

Model transport, boundary condition, and other CO_2 flux components also contribute to modeled [CO_2] errors. Although we illustrated that the contribution is less than that from the biosphere, it can potentially shift the biases uniformly up or down. Therefore, to root out the potential interference, we directly compare the CASA flux members and CT2017 flux member with AmeriFlux eddy-covariance flux measurements. This comparison can confirm if the results from the [CO_2] analysis are consistent with eddy-flux data analysis and lend more insight into the diel cycle of CO_2 fluxes. We group these analyses according to the dominant biomes surrounding each [CO_2] tower.

Due to the uneven distribution of flux and [CO_2] towers, flux towers cannot be always found in the footprint of a given [CO_2] tower. We therefore propagate the flux biases at each [CO_2] tower location using Eq. (1) below. By assuming that a given biome type in the CASA model and CT2017 have a similar behavior (bias) everywhere, we group the flux towers influenced by the same dominant biome together and calculate the overall bias for the given biome F_T. The flux biases at a given [CO_2] tower location, B_i, can be expressed in the following equation.

$$B_i = \frac{1}{24}\sum_{t=1}^{24}\left(\sum_{T=0}^{14} F_T W_T\right) W_t \quad (1)$$

where W_T and W_t are the spatial and temporal weighting functions that propagate the flux biases from the biome level to the [CO_2] site level, respectively. The subscript T denotes the biome: Indexes 1 to 14 are the biome in CASA following Zhou et al. (2020a); index 0 represents water bodies. W_T are the fractional areas of individual biomes relative to the entire influence area of a [CO_2] tower (Figure S2). In the analysis, we find that some towers can be influenced by up to six biome types according to the tower footprint. The temporal weighting function, W_t, is essentially equal across the 24 hours of a day due to the indistinguishable diel cycle after averaging the footprint over the period of interest (Figure S3b and Text S2 in SI).

Figure *8*b shows the derived flux biases associated with each [CO_2] tower over the same time period following Eq. 1. Consistent with the [CO_2] analysis, all of the flux members show positive biases across the [CO_2] towers, indicating the positive biases in modeled [CO_2] are mainly due to the fact that CASA and CT2017 underestimate net uptake. For CASA, this can be attributed to weak plant productivity and/or strong respiration. Zhou et al. (2020) pointed out that there is missing net sink due to the lack of crop harvest and forest recovery in the CASA model,

yielding a net overestimate in annual respiration. Additionally, the relative performance among the CASA E_{max} groups and CT2017 with respect to flux tower data is consistent with the comparison to [CO_2]. The high E_{max} values lead to the largest biases in both [CO_2] and flux space, and CT2017 shows smaller biases for CR and DB. Such coherence across measurement space (mole fraction and flux) and platforms (aircraft, [CO_2] towers, and flux towers) lends a high degree of confidence to these results.

We break the daytime and nighttime flux biases apart to explore what causes the overall weak NEE uptake in CASA. In summer, the daytime fluxes are a combination of GPP and ecosystem respiration; the nighttime fluxes are driven completely by respiration. In the flux observations, the daily, daytime, and nighttime averaged fluxes range from -2 to -1 $\mu\, mol\, m^{-2}s^{-1}$, from -8 to -3 $\mu\, mol\, m^{-2}s^{-1}$, and from 0.5 to 2.5 $\mu\, mol\, m^{-2}s^{-1}$, respectively (Figure S5 in SI). CASA always underestimates the daytime net sink (**Figure *8***c): decreasing E_{max} values leads to larger biases across biomes (sites), and increasing E_{max} values do not offset the weak net uptake in CASA, indicating that tuning the parameters cannot fully counteract the lack of the harvest sink in CASA for summer 2016. CASA appears less biased in simulating the nighttime fluxes but in fact still highly biased compared to the averaged nighttime flux magnitude. On average, CASA underestimates the daily-averaged sink by 1 to 2 $\mu\, mol\, m^{-2}s^{-1}$ for EN, CR, and DB and by 1 $\mu\, mol\, m^{-2}s^{-1}$ or less for GL. The different magnitudes of flux biases across the biomes suggests that the flux bias is scaled with the strength of the seasonal uptake and/or the strength of the annual net carbon exchange (i.e., missing sink).

CR and DB have more distinct E_{max} ranking patterns for the daytime and nighttime flux biases. For CR, low E_{max} leads to larger biases in both daytime and nighttime fluxes due to low daytime net uptake and nighttime respiration. Since E_{max} in CASA directly impacts GPP and indirectly impacts respiration, increasing both GPP and respiration is favorable for capturing CR fluxes in CASA. This is also reflected by the combination of high E_{max} and high Q_{10} values (P25 and P27) being ranked in the top group in **Figure *6***. For DB, however, low E_{max} leads to large, positive flux biases in daytime but the smallest positive bias at night, suggesting that the bigger issue for DB is that the CASA model tends to respire carbon faster than in the actual ecosystem.

In both mole fraction and flux space, CT2017 agrees more closely overall with the observations than the CASA members after averaging (**Figure *8***). However, once we break nighttime and daytime flux apart, CT2017 overestimates the nighttime fluxes for these four biome types and daytime net uptake for DB.

5 Discussion

Multiple CO_2 observation platforms (i.e., aircraft and tower [CO_2] and flux tower data) across measurement space (i.e., concentration vs. flux) identify that E_{max} is the most important parameter driving the spread of the CASA flux ensemble. In concentration space, the modeled [CO_2] biases can be contributed by biogenic flux, transport, boundary conditions and other CO_2 flux components. Although **Figure *3*** demonstrates that, in the sampled cases, the biosphere dominates the model errors, the transport errors are not negligible. The model transport errors can increase or decrease modeled [CO_2] biases in **Figure *5*** as a whole. We choose to focus on the relative performance among these flux members to minimize the impact of the potential transport errors on interpreting the results. However, we acknowledge that the complexity the model transport introduces in the results are not eliminated.

We argue that the transport error is unlikely to change our major conclusion that the biogenic fluxes are biased or change the relative performance across the biogenic CO_2 flux members. This argument is supported in the consistent results from the flux comparison in Section 4.3. This is also reflected by the results that CT2017 has a better agreement with measured [CO_2] (**Figure *5*** and **Figure *6***). While the CT2017 fluxes were optimized with the TM5 transport model (Krol et al., 2005), we consistently found that the associated WRF-simulated [CO_2] bias was the lowest compared to the CASA members. We also note that a slightly different version of CASA (Potter et al., 1993) is used as a prior in CT2017. In our transport model setup, we nudged the WRF-Chem model to the state-of-the-art reanalysis product ERA5 in order to improve the large-scale dynamics. The comparison between the nudged and free run modeled wind fields illustrate the effectiveness and efficiency of this meteorological constraint. Finally, the ensemble of perturbed transport simulations confirms the secondary role of transport errors in [CO_2] model-data residuals. Hence, it seems highly unlikely that transport errors could be the cause of our findings regarding the CASA ensemble.

The CarbonTracker (CT) inversion system works by estimating scaling factors that multiply the prior model (CASA)'s NEE. During the growing season, an increase in the terrestrial sink is equivalent to amplifying the prior model's diel cycle, to perhaps unrealistic levels. These scaling factors are constant for each week and are independent from one week to the next. Thus, the CT fluxes may yield periods of unrealistically large diel cycles, possibly followed by periods with abnormally small daily cycles of NEE. The shape of the day-night flux differences and day-to-day changes cannot be adjusted by CT. To overcome this problem, Schuh et al. (2013) optimized for GPP and respiration separately. Additionally, since CT relies heavily on simulated transport, it is also possible to retrieve exaggerated diurnal cycles in fluxes if that model ventilates the PBL too strongly. While this process could affect fluxes throughout the day, thin and stable nighttime boundary layers are particularly difficult to represent in models of this class.

Section 4.3 provides more insight into the causes of the modeled [CO_2] biases. The biggest assumption we made in Eq. (1) is homogeneity across the entire biome. The optimal method for calculating the flux biases at a [CO_2] tower should be based on the flux towers in a [CO_2] tower's footprint (shown in Figure S2). However, due to the uneven distribution of the flux towers in the domain of interest, this direct calculation is not possible. The method we used to derive the flux biases at each [CO_2] tower takes the influence from multiple biomes into account and results in a bias ranking among the flux members that is consistent with the ranking in mole fraction space, indicating the reliability of this method.

The CASA simulations were spun up equilibrium, and the parameters were adjusted to match the observations. CASA's overestimation of summer NEE likely derives from missing carbon sink processes because of the balanced-biosphere equilibrium starting condition resulting in carbon pools that are too large rather than a problem in the model's parameters (Pietsch & Hasenauer, 2006; Wutzler & Reichstein, 2007; Zhou, Williams, Lauvaux, Davis, et al., 2020). The range of the three perturbed parameters, E_{max}, T_{opt}, and Q_{10}, were determined by comparison to flux measurements over 13 year simulation period. Tuning these parameters cannot solve the fact that carbon stocks are actually dynamic and out of equilibrium. Our results may erroneously indicate that a particular parameter set has the smallest bias when the core problem may be that the carbon pool is out of equilibrium. This is consistent with the finding that CT2017, with its imposed net global land sink, is less biased than the balanced biosphere CASA L2 ensemble.

We therefore do not intend to stress which parameter set is the best. Instead, the highlight of this work is to demonstrate that atmospheric [CO_2] data can be used, with flux tower measurements, to diagnose the performance of TBMs. This study demonstrates the utility of multiple observation platforms for terrestrial biosphere model evaluation.

The rankings in Section 4.2 show that increasing E_{max} does not necessarily lead to better agreement with the measurements. Increasing E_{max} increases GPP, but the carbon seems to respire away quickly in CASA and does not reside long enough, perhaps because harvest is not represented in this simulation. Key missing processes include the effects of management and land use such as (a) agricultural sinks from management that removes crops and crop residues thus decreasing the size of the carbon pool that might be respired, (b) pastureland sinks from cattle and other grazers that consume plant biomass and store it in their body mass, (c) forest carbon storage as trees and stands mature with sequestration in wood. Additional candidates include stimulation of ecosystem carbon sinks by growth enhancement factors such as rising CO2 concentration, nitrogen deposition, fertilizer additions, and the like. Including these missing processes in the model framework would require a re-evaluation of all model parameters, and we might find that a different E_{max} parameter set was best relative to what was identified in this study and might lead to a better match with this suite of observation. In addition, the daytime and nighttime flux biases of CASA suggest that the day-night cycling of respiration is not being done properly. The modeled respiration is primarily related to temperature. However, previous studies showed that the correlation plots of respiration and temperature are scattered. The high-biased points tend to occur in daytime, and the low-biased points tend to occur at nighttime because the carbon fixation processes are more active during daytime when more labile carbon is ready to be respired. There are different day-night biases in the model structure.

Simplistic temporal downscaling of CASA's monthly carbon fluxes likely contributes to biases in the diel cycle. The 3-hourly CASA ensemble flux products were downscaled from the native monthly resolution using the Olsen and Randerson (2004) method, in which GPP is downscaled with downward shortwave radiation, Re is downscaled with air temperature and Q_{10}, and NEE is the sum of GPP and Re. This makes it difficult to use the diel cycle analyses in section 4.3 to make conclusions about GPP and Re in CASA. The biases in the diel cycle are nonetheless informative, since this temporal downscaling is widely used.

The ability to diagnose errors in GPP and respiration will require models that resolve the diel cycle and possibly additional data dedicated to the photosynthetic process. Carbonyl sulfide (COS) has been proposed as an independent proxy for GPP as it diffuses into leaves in a fashion very similar to CO_2, but in contrast to the latter, is generally not emitted by respiration. Campbell et al. (2017) presented a global, measurement-based estimate of GPP growth during the twentieth century that is based on long-term atmospheric COS records, derived from ice-core, firn and ambient air samples and found that the observation-based COS record is most consistent with simulations of climate and the carbon cycle. Recently, Spielmann et al. (2019) used concurrent ecosystem-scale flux measurements of CO_2 and COS at four European biomes for a joint constraint on CO_2 flux partitioning. Their results demonstrated the importance of using multiple approaches for constraining present-day GPP due to a systematic underestimation under low light conditions with the classical approaches relying merely on CO_2 fluxes. Other studies have used the $\delta^{13}C$ of nocturnal whole-ecosystem respiration as a proxy from which to derive carbon isotope discrimination associated with photosynthesis (Alstad et al., 2007; Bowling et al., 2002;

Flanagan et al., 1996). Joint atmospheric gas constraints may improve our diagnoses of the causes of biases in TBM flux estimates. Studies using ACT-America airborne biogenic tracers, including CO_2, COS, and CO, are submitted to this collection.

6 Conclusions

We evaluate the modeled [CO_2] biases associated with CASA TMB biogenic CO_2 flux ensemble members and the CT2017 posterior biogenic flux using aircraft and tower *in situ* [CO_2] jointly with eddy covariance flux data in summer 2016. Aircraft and tower *in situ* [CO_2] were influenced by grasslands (GL), evergreen needleleaf forest (EN), deciduous broadleaf forest (DB) and croplands (CR). While the mole fraction-based analyses revealed a systematic underestimation of carbon uptake by the balanced-biosphere CASA runs, the flux-based analyses identified a combined effect from an overestimated respiration at night and an under-estimated uptake during the day. The joint observational analysis yields strong confidence that these results span large spatial domains and multiple ecosystems due to the availability of long aircraft transects and a wide network of ground-based measurements. The results from analyzing both mole fraction and flux model-data residuals were consistent. The systematic errors in CASA that span all parameter values suggest that missing processes cannot be properly simulated by adjusting the existing parameters. Analyses indicate that modeled [CO_2] biases are related to biome productivity; the models tend to be biased more for high productivity biomes (DB and CR) and biased less for GL and EN. In particular, the summer harvest sink absent from CASA seems to be responsible for large biases in the Midwest. Lastly, CT2017, an inversion product that is constrained by atmospheric [CO_2] data and has an imposed net biogenic carbon sink, shows better agreement with [CO_2] mole fraction data compared to CASA flux ensemble members. However, the flux-based analyses revealed that the diurnal variations of CT were unrealistically large. We suggest that the scaling of the net daily fluxes in large-scale inversions must be further decomposed into day and night sub-components to reproduce the diel cycle of photosynthetic and ecosystem respiration processes.

Acknowledgments, Samples, and Data

This work was funded by the NASA ACT-America project. ACT - America project is a NASA Earth Venture Suborbital-2 project funded by NASAs Earth Science Division (Grant NNX15AG76G to Penn State and NNX16AN17G to Clark University). T. Lauvaux was also supported by the French research program Make Our Planet Great Again (project CIUDAD). Many thanks to Andrew R. Jacobson at NOAA Global Monitoring Laboratory (GML) for the feedback on the diel cycle of the CT2017 posterior biogenic CO2 fluxes.

The CarbonTracker CT2017 product can be obtained from NOAA GML, Boulder, Colorado, USA from the website at http://carbontracker.noaa.gov. The L2 CASA ensemble suite used in this work can be found at https://doi.org/10.3334/ORNLDAAC/1675. NCEP Real-Time 12-km SST are archived at ftp://polar.ncep.noaa.gov/pub/history/sst/ophi. The ensemble output will be made available from datacommons@psu at https: //doi.org/10.26208/z864-qk7. The Obspack CO_2 Global View Plus data package can be obtained from https://www.esrl.noaa.gov/gmd/ccgg/obspack/. The citations for individual [CO_2] towers are listed in Table S3. AmeriFlux data are obtained from https://ameriflux.lbl.gov/, with citations given in Table S2.

References

Alstad, K. P., Lai, C.-T., Flanagan, L. B., & Ehleringer, J. R. (2007). Environmental controls on the carbon isotope composition of ecosystem-respired CO2 in contrasting forest ecosystems in Canada and the USA. *Tree Physiology*, *27*(10), 1361–1374. https://doi.org/10.1093/treephys/27.10.1361

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., et al. (2014). CO2, CO, and CH4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts. *Atmospheric Measurement Techniques*, *7*(2), 647–687. https://doi.org/10.5194/amt-7-647-2014

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., et al. (2006). TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988–2003. *Global Biogeochemical Cycles*, *20*(1). https://doi.org/10.1029/2004GB002439

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Barker, D. M., Huang, W., Guo, Y.-R., Bourgeois, A. J., & Xiao, Q. N. (2004). A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results. *Monthly Weather Review*, *132*(4), 897–914. https://doi.org/10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2

Basu, S., Miller, J. B., & Lehman, S. (2016). Separation of biospheric and fossil fuel fluxes of CO_2 by atmospheric inversion of CO_2 and $^{14}CO_2$ measurements: Observation System Simulations. *Atmospheric Chemistry and Physics*, *16*(9), 5665–5683. https://doi.org/10.5194/acp-16-5665-2016

Berner, J., Shutts, G. J., Leutbecher, M., & Palmer, T. N. (2009). A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System. *Journal of the Atmospheric Sciences*, *66*(3), 603–626. https://doi.org/10.1175/2008JAS2677.1

Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A., & Snyder, C. (2011). Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations. *Monthly Weather Review*, *139*(6), 1972–1995. https://doi.org/10.1175/2010MWR3595.1

Berner, J., Fossell, K. R., Ha, S.-Y., Hacker, J. P., & Snyder, C. (2015). Increasing the Skill of Probabilistic Forecasts: Understanding Performance Improvements from Model-Error Representations. *Monthly Weather Review*, *143*(4), 1295–1320. https://doi.org/10.1175/MWR-D-14-00091.1

Bowling, D. R., McDowell, N. G., Bond, B. J., Law, B. E., & Ehleringer, J. R. (2002). 13C content of ecosystem respiration is linked to precipitation and vapor pressure deficit. *Oecologia*, *131*(1), 113–124. https://doi.org/10.1007/s00442-001-0851-y

Butler, M. P., Lauvaux, T., Feng, S., Liu, J., Bowman, K. W., & Davis, K. J. (2020). Atmospheric Simulations of Total Column CO2 Mole Fractions from Global to Mesoscale within the Carbon Monitoring System Flux Inversion Framework. *Atmosphere*, *11*(8), 787. https://doi.org/10.3390/atmos11080787

694 Campbell, J. E., Berry, J. A., Seibt, U., Smith, S. J., Montzka, S. A., Launois, T., et al. (2017).
695 Large historical growth in global terrestrial gross primary production. *Nature*, *544*(7648),
696 84–87. https://doi.org/10.1038/nature22030
697 Chen, F., & Dudhia, J. (2001). Coupling an Advanced Land Surface–Hydrology Model with the
698 Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and
699 Sensitivity. *Monthly Weather Review*, *129*(4), 569–585. https://doi.org/10.1175/1520-
700 0493(2001)129<0569:CAALSH>2.0.CO;2
701 Cooperative Global Atmospheric Data Integration Project. (n.d.). Multi-laboratory compilation
702 of atmospheric carbon dioxide data for the period 1957-2018;
703 obspack_co2_1_GLOBALVIEWplus_v5.0_2019_08_12; NOAA Earth System Research
704 Laboratory, Global Monitoring Division http://dx.doi.org/10.25925/20190812.
705 Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., et al. (2019). The
706 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network.
707 *Atmospheric Chemistry and Physics*, *19*(15), 9797–9831. https://doi.org/10.5194/acp-19-
708 9797-2019
709 Davis, K. J., Obland, M. D., Lin, B., Lauvaux, T., O'dell, C., Meadows, B., et al. (2018). ACT-
710 America: L3 Merged In Situ Atmospheric Trace Gases and Flask Data, Eastern USA.
711 *ORNL DAAC*. https://doi.org/10.3334/ORNLDAAC/1593
712 Díaz-Isaac, L. I., Lauvaux, T., & Davis, K. J. (2018). Impact of physical parameterizations and
713 initial conditions on simulated atmospheric transport and CO_2 mole fractions in the US
714 Midwest. *Atmospheric Chemistry and Physics*, *18*(20), 14813–14835.
715 https://doi.org/10.5194/acp-18-14813-2018
716 Enting, I. G., Trudinger, C. M., & Francey, R. J. (1995). A synthesis inversion of the
717 concentration and δ13 C of atmospheric CO2. *Tellus B*, *47*(1–2), 35–52.
718 https://doi.org/10.1034/j.1600-0889.47.issue1.5.x
719 Feng, S., Lauvaux, T., Keller, K., Davis, K. J., Rayner, P., Oda, T., & Gurney, K. R. (2019). A
720 Road Map for Improving the Treatment of Uncertainties in High-Resolution Regional
721 Carbon Flux Inverse Estimates. *Geophysical Research Letters*, *46*(22), 13461–13469.
722 https://doi.org/10.1029/2019GL082987
723 Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., et al. (2019). Seasonal
724 Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-
725 Scale Boundary Inflow in Atmospheric CO2 Simulations Over North America. *Journal*
726 *of Geophysical Research: Atmospheres*, *124*(24), 14325–14346.
727 https://doi.org/10.1029/2019JD031165
728 Flanagan, L. B., Brooks, J. R., Varney, G. T., Berry, S. C., & Ehleringer, J. R. (1996). Carbon
729 isotope discrimination during photosynthesis and the isotope ratio of respired CO2 in
730 boreal forest ecosystems. *Global Biogeochemical Cycles*, *10*(4), 629–640.
731 https://doi.org/10.1029/96GB02345
732 Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., &
733 Knutti, R. (2014). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle
734 Feedbacks. *Journal of Climate*, *27*(2), 511–526. https://doi.org/10.1175/JCLI-D-12-
735 00579.1
736 Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., et al.
737 (2019). Global Carbon Budget 2019. *Earth System Science Data*, *11*(4), 1783–1838.
738 https://doi.org/10.5194/essd-11-1783-2019

739 Fung, I. Y., Tucker, C. J., & Prentice, K. C. (1987). Application of Advanced Very High
740 Resolution Radiometer vegetation index to study atmosphere-biosphere exchange of
741 CO2. *Journal of Geophysical Research: Atmospheres*, *92*(D3), 2999–3015.
742 https://doi.org/10.1029/JD092iD03p02999
743 Gloor, M., Bakwin, P., Hurst, D., Lock, L., Draxler, R., & Tans, P. (2001). What is the
744 concentration footprint of a tall tower? *Journal of Geophysical Research: Atmospheres*,
745 *106*(D16), 17831–17840. https://doi.org/10.1029/2001JD900021
746 Gonçalves, L. G. G. D., Borak, J. S., Costa, M. H., Saleska, S. R., Baker, I., Restrepo-Coupe, N.,
747 et al. (2013). Overview of the large-scale biosphere-atmosphere experiment in amazonia
748 data model intercomparison project (LBA-DMIP). *Agricultural and Forest Meteorology*,
749 *182–183*, 111–127. https://doi.org/10.1016/j.agrformet.2013.04.030
750 Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder,
751 B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric*
752 *Environment*, *39*(37), 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027
753 Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., et al. (2002).
754 Towards robust regional estimates of CO 2 sources and sinks using atmospheric transport
755 models. *Nature*, *415*(6872), 626–630. https://doi.org/10.1038/415626a
756 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020).
757 The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*,
758 *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803
759 Houghton, J. T., Meria Filho, L. G., Lim, B., Treanton, K., Mamaty, I., Bonduki, Y., et al.
760 (1996). *Intergovernmental Panel on Climate Change. Greenhouse Gas Inventory*
761 *Reporting Instructions* (IPCC WGI Technical Support Unit No. Volumes 1, 2 and 3
762 (IPCC, OECD and IEA)). Bracknell, U.K.: Meteorological Office.
763 Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., et al.
764 (2013). The North American Carbon Program Multi-Scale Synthesis and Terrestrial
765 Model Intercomparison Project – Part 1: Overview and experimental design.
766 *Geoscientific Model Development*, *6*(6), 2121–2133. https://doi.org/10.5194/gmd-6-2121-
767 2013
768 Huntzinger, Deborah N., Gourdji, S. M., Mueller, K. L., & Michalak, A. M. (2011). The utility
769 of continuous atmospheric measurements for identifying biospheric CO2 flux variability.
770 *Journal of Geophysical Research: Atmospheres*, *116*(D6).
771 https://doi.org/10.1029/2010JD015048
772 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D.
773 (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER
774 radiative transfer models. *Journal of Geophysical Research: Atmospheres*, *113*(D13).
775 https://doi.org/10.1029/2008JD009944
776 Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020).
777 Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of
778 the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-
779 17-1343-2020
780 Kain, J. S. (2004). The Kain–Fritsch Convective Parameterization: An Update. *Journal of*
781 *Applied Meteorology*, *43*(1), 170–181. https://doi.org/10.1175/1520-
782 0450(2004)043<0170:TKCPAU>2.0.CO;2

783 Kaminski, T., Knorr, W., Rayner, P. J., & Heimann, M. (2002). Assimilating atmospheric data
784 into a terrestrial biosphere model: A case study of the seasonal cycle. *Global*
785 *Biogeochemical Cycles*, *16*(4), 14-1-14–16. https://doi.org/10.1029/2001GB001463
786 Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., & Mathieu, P.-P.
787 (2012). Consistent assimilation of MERIS FAPAR and atmospheric CO_2 into a terrestrial
788 vegetation model and interactive mission benefit analysis. *Biogeosciences*, *9*(8), 3173–
789 3184. https://doi.org/10.5194/bg-9-3173-2012
790 Knorr, W. (2000). Annual and interannual CO2 exchanges of the terrestrial biosphere: process-
791 based simulations and uncertainties. *Global Ecology and Biogeography*, *9*(3), 225–252.
792 https://doi.org/10.1046/j.1365-2699.2000.00159.x
793 Koffi, E. N., Rayner, P. J., Scholze, M., Chevallier, F., & Kaminski, T. (2013). Quantifying the
794 constraint of biospheric process parameters by CO_2 concentration and flux measurement
795 networks through a carbon cycle data assimilation system. *Atmospheric Chemistry and*
796 *Physics*, *13*(21), 10555–10572. https://doi.org/10.5194/acp-13-10555-2013
797 Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., et al.
798 (2005). The two-way nested global chemistry-transport zoom model TM5: algorithm and
799 applications. *Atmospheric Chemistry and Physics*, *5*(2), 417–432.
800 https://doi.org/10.5194/acp-5-417-2005
801 Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., et al. (2012).
802 Constraining the CO2 budget of the corn belt: exploring uncertainties from the
803 assumptions in a mesoscale inverse system. *Atmospheric Chemistry and Physics*, *12*(1),
804 337–354. https://doi.org/10.5194/acp-12-337-2012
805 Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., et al. (2014). Carbon
806 monitoring system flux estimation and attribution: impact of ACOS-GOSAT XCO2
807 sampling on the inference of terrestrial biospheric sources and sinks. *Tellus B: Chemical*
808 *and Physical Meteorology*, *66*(1), 22486. https://doi.org/10.3402/tellusb.v66.22486
809 Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., et al.
810 (2012). A framework for benchmarking land models. *Biogeosciences*, *9*(10), 3857–3874.
811 https://doi.org/10.5194/bg-9-3857-2012
812 Masarie, K. A., Peters, W., Jacobson, A. R., & Tans, P. P. (2014). ObsPack: a framework for the
813 preparation, delivery, and attribution of atmospheric greenhouse gas measurements.
814 *Earth System Science Data*, *6*(2), 375–384. https://doi.org/10.5194/essd-6-375-2014
815 McCaughey, J. H., Pejam, M. R., Arain, M. A., & Cameron, D. A. (2006). Carbon dioxide and
816 energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada.
817 *Agricultural and Forest Meteorology*, *140*(1), 79–96.
818 https://doi.org/10.1016/j.agrformet.2006.08.010
819 Nakanishi, M., & Niino, H. (2006). An Improved Mellor–Yamada Level-3 Model: Its Numerical
820 Stability and Application to a Regional Prediction of Advection Fog. *Boundary-Layer*
821 *Meteorology*, *119*(2), 397–407. https://doi.org/10.1007/s10546-005-9030-8
822 Ogle, S. M., Davis, K., Lauvaux, T., Schuh, A., Cooley, D., West, T. O., et al. (2015). An
823 approach for verifying biogenic greenhouse gas emissions inventories with atmospheric
824 CO 2 concentration data. *Environmental Research Letters*, *10*(3), 034012.
825 https://doi.org/10.1088/1748-9326/10/3/034012
826 Olsen, S. C., & Randerson, J. T. (2004). Differences between surface and column atmospheric
827 CO2 and implications for carbon cycle research. *Journal of Geophysical Research:*
828 *Atmospheres*, *109*(D2). https://doi.org/10.1029/2003JD003968

Pal, S., Davis, K. J., Lauvaux, T., Browell, E. V., Gaudet, B. J., Stauffer, D. R., et al. (2020). Observations of Greenhouse Gas Changes Across Summer Frontal Boundaries in the Eastern United States. *Journal of Geophysical Research: Atmospheres*, *125*(5).

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., et al. (2005). An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations. *Journal of Geophysical Research: Atmospheres*, *110*(D24). https://doi.org/10.1029/2005JD006157

Peters, Wouter, 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

Pietsch, S. A., & Hasenauer, H. (2006). Evaluating the self-initialization procedure for large-scale ecosystem models. *Global Change Biology*, *12*(9), 1658–1669. https://doi.org/10.1111/j.1365-2486.2006.01211.x

Pillai, D., Gerbig, C., Kretschmer, R., Beck, V., Karstens, U., Neininger, B., & Heimann, M. (2012). Comparing Lagrangian and Eulerian models for CO_2 transport – a step towards Bayesian inverse modeling using WRF/STILT-VPRM. *Atmospheric Chemistry and Physics*, *12*(19), 8979–8991. https://doi.org/10.5194/acp-12-8979-2012

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., & Klooster, S. A. (1993). Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles*, *7*(4), 811–841. https://doi.org/10.1029/93GB02725

Sarrat, C., Noilhan, J., Dolman, A. J., Gerbig, C., Ahmadov, R., Tolk, L. F., et al. (2007). Atmospheric CO_2 modeling at the regional scale: an intercomparison of 5 meso-scale atmospheric models. *Biogeosciences*, *4*(6), 1115–1126. https://doi.org/10.5194/bg-4-1115-2007

Scholze, M., Kaminski, T., Knorr, W., Voßbeck, M., Wu, M., Ferrazzoli, P., et al. (2019). Mean European Carbon Sink Over 2010–2015 Estimated by Simultaneous Assimilation of Atmospheric CO2, Soil Moisture, and Vegetation Optical Depth. *Geophysical Research Letters*, *46*(23), 13796–13803. https://doi.org/10.1029/2019GL085725

Schuh, A. E., Lauvaux, T., West, T. O., Denning, A. S., Davis, K. J., Miles, N., et al. (2013). Evaluating atmospheric CO2 inversions at multiple scales over a highly inventoried agricultural landscape. *Global Change Biology*, *19*(5), 1424–1439. https://doi.org/10.1111/gcb.12141

Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., et al. (2019). Quantifying the Impact of Atmospheric Transport Uncertainty on CO2 Surface Flux Estimates. *Global Biogeochemical Cycles*, *33*(4), 484–500. https://doi.org/10.1029/2018GB006086

Schwalm, C. R., Huntzinger, D. N., Fisher, J. B., Michalak, A. M., Bowman, K., Ciais, P., et al. (2015). Toward "optimal" integration of terrestrial biosphere models. *Geophysical Research Letters*, *42*(11), 4418–4428. https://doi.org/10.1002/2015GL064002

Seibert, P., & Frank, A. (2004). Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode. *Atmospheric Chemistry and Physics*, *4*(1), 51–63. https://doi.org/10.5194/acp-4-51-2004

873 Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems.
874 *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 3079–3102.
875 https://doi.org/10.1256/qj.04.106
876 Sitch, S., Huntingford, C., Gedney, N., Levy, P. E., Lomas, M., Piao, S. L., et al. (2008).
877 Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon
878 cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Global*
879 *Change Biology*, *14*(9), 2015–2039. https://doi.org/10.1111/j.1365-2486.2008.01626.x
880 Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Barker, D., Duda, G., et al. (2008). A Description
881 of the Advanced Research WRF Version 3. https://doi.org/10.5065/D68S4MVH
882 Spielmann, F. M., Wohlfahrt, G., Hammerle, A., Kitz, F., Migliavacca, M., Alberti, G., et al.
883 (2019). Gross Primary Productivity of Four European Ecosystems Constrained by Joint
884 CO2 and COS Flux Measurements. *Geophysical Research Letters*, *46*(10), 5284–5293.
885 https://doi.org/10.1029/2019GL082006
886 Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., et al. (2015).
887 Seasonal climatology of CO2 across North America from aircraft measurements in the
888 NOAA/ESRL Global Greenhouse Gas Reference Network. *Journal of Geophysical*
889 *Research: Atmospheres*, *120*(10), 5155–5190. https://doi.org/10.1002/2014JD022591
890 Tans, P. P., & Conway, T. J. (2005). *Monthly Atmospheric CO2 Mixing Ratios from the NOAA*
891 *CMDL Carbon Cycle Cooperative Global Air Sampling Network, 1968-2002* (No.
892 cdiac:NOAA/CMDL). Environmental System Science Data Infrastructure for a Virtual
893 Ecosystem. https://doi.org/10.15485/1464220
894 Tans, Pieter P., Fung, I. Y., & Takahashi, T. (1990). Observational Contrains on the Global
895 Atmospheric Co2 Budget. *Science*, *247*(4949), 1431–1438.
896 https://doi.org/10.1126/science.247.4949.1431
897 Thompson, G., Rasmussen, R. M., & Manning, K. (2004). Explicit Forecasts of Winter
898 Precipitation Using an Improved Bulk Microphysics Scheme. Part I: Description and
899 Sensitivity Analysis. *Monthly Weather Review*, *132*(2), 519–542.
900 https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2
901 Uliasz, M., Bartochowska, M., Madany, A., Piwkowski, H., Parfiniewicz, J., & Rozkrut, M.
902 (1994). Application of the Mesoscale Dispersion Modeling System to Investigation of Air
903 Pollution Transport in Southern Poland. In S.-E. Gryning & M. M. Millán (Eds.), *Air*
904 *Pollution Modeling and Its Application X* (pp. 27–34). Boston, MA: Springer US.
905 https://doi.org/10.1007/978-1-4615-1817-4_4
906 United Nations Framework Convention on Climate Change (UNFCCC). (2015). *Decision*
907 *1/CP.21: Adoption of the Paris Agreement* (No. FCCC/CP/2015/L.9/Rev.1. 12 Dec).
908 Paris: UNFCCC. Retrieved from
909 https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf
910 Wutzler, T., & Reichstein, M. (2007). Soils apart from equilibrium – consequences for
911 soil carbon balance modelling. *Biogeosciences*, *4*(1), 125–136.
912 https://doi.org/10.5194/bg-4-125-2007
913 Zhou, Y., Williams, C. A., Lauvaux, T., Davis, K. J., Feng, S., Baker, I., et al. (2020). A
914 Multiyear Gridded Data Ensemble of Surface Biogenic Carbon Fluxes for North
915 America: Evaluation and Analysis of Results. *Journal of Geophysical Research:*
916 *Biogeosciences*, *125*(2), e2019JG005314. https://doi.org/10.1029/2019JG005314

917 Zhou, Y., Williams, C. A., Lauvaux, T., Feng, S., Baker, I. T., Wei, Y., et al. (2020). ACT-
918 America: Gridded Ensembles of Surface Biogenic Carbon Fluxes, 2003-2019.
919 https://doi.org/10.3334/ORNLDAAC/1675
920
921

Figure and Table captions

Figure *1*. CASA biome type, and the locations of the [CO_2] towers (red circled crosses) and AmeriFlux flux towers (black triangles). The colored area is the model domain.

Figure *2*. Simulated [CO_2] (left column) and the RMSD of the biogenic [CO_2] ([CO_2bio]) from the mean of modeled [CO_2bio] (right column) for five flights in the ACT Summer 2016 aircraft campaign: (a) 7/18 19 UTC (front in MA); (b) 7/21 19 UTC (fair Wx in MA): (c) 8/4 18 UTC (front in MW); (d) 8/9 19 UTC (fair Wx in MW); (e) 8/22 22 UTC (fair Wx in South). The aircraft sampled [CO_2] are overlaid with the simulations (left column). The black dots on the RMSD panels are the flight paths. Potential temperature contours are overlaid to indicate the frontal locations.

Figure *3*. Daily afternoon averaged [CO_2] from the ACT aircraft and simulations. The error bars of the simulated [CO_2] are the spread of the full, biogenic flux, transport, and boundary condition ensembles. Note that the spread of boundary condition is very small and not visible in the figure. Only the ABL samples are included in the analysis. The point-down arrows denote the frontal cases during the campaign. Gray dashed lines denote the transit.

Figure 4. Taylor diagrams of simulated [CO_2] compared with the observations. The standard deviations of models are normalized by the standard deviations of observations ("REF"). The model [CO_2] are associated with the (29) biogenic flux members from CASA and CT2017. The color scheme indicates CT2017, CASA ensemble mean, CASA low E_{max} (E_1), CASA medium E_{max} (E_2), and CASA high E_{max} (E_3) groups. "P#" in the legend denotes the index of CASA ensemble member in the original product. The observations used in (a) are from the ABL legs of the ACT-America aircraft measurements. The flight dates and other information are listed in Table S1. The observations used in (b) are from the [CO_2] tower measurements at afternoon hours (19-22 UTC). The locations and information can be found in Figure 1 and Table 3. Both aircraft and tower [CO_2] data are kept at native temporal resolutions. Modeled [CO_2] are determined by the value of the nearest grid cell to sample locations. The period of interest is from July 18 to August 28, 2016 covering the ACT-America summer 2016 aircraft campaign.

Figure *5*. The model [CO_2] are associated with the (29) biogenic flux members from CASA and CT2017. The color scheme indicates CT2017, CASA ensemble mean, CASA low E_{max} (E_1), CASA medium E_{max} (E_2), and CASA high E_{max} (E_3) groups. "P#" on the x-axis are the indexes of CASA ensemble members in the original product.

Figure *6*. The final ranking of the modeled [CO_2] bias among the biogenic flux members shown in Figure 5. The model skills are ranked decreasingly, meaning the model biases increase from top to bottom. The name convention of the CASA ensemble members follows P# /$T_{opt}E_{max}Q_{10}$. # is the index of the CASA members in the original product. At the position of T_{opt}, "0" means the default value; "2" means the default value plus 2 degrees C; "-2" means the default value minus 2 degrees C. For E_{max}, "1", "2", and "3" denotes the low, medium, and high E_{max} in Table 4 of Zhou et al. (2020). "2","4","6" at the Q_{10} position indicate that Q_{10} = "1.2", "1.4", and "1.6", respectively.

Figure *7*. Top 11 biogenic flux members in Figure 6 applied different color schemes in order to reveal the secondary parameter, Q_{10}. Note that the flux members associated in the medium E_{max} are with color background; others are with white background.

964 **Figure *8*.** CO_2 mole fraction and flux biases at the [CO_2] towers. (a) Afternoon averaged (19-22
965 UTC) [CO_2] biases. (b) daily, (c) daytime (15-2 UTC), and (c) nighttime (3-14 UTC) averages
966 of the flux biases derived from Eq. (1). The shaded areas are the minimum and maximum biases
967 among different E_{max} groups in the CASA ensemble. The associated solid lines are the mean of
968 the given E_{max} group. Additionally, CT2017 and the CASA ensemble mean are denoted in green
969 and purple dotted lines. The horizontal gray line is unbiased. The vertical gray lines divide the
970 towers into various biomes determined by the footprint analysis. The dominate biomes, GL -
971 grassland, EN - evergreen needleleaf forest, Wa - water, CR - croplands, DB - deciduous
972 broadleaf forest, MF - mixed forest for [CO_2] towers are listed in Table 3. A mixed type is listed
973 if the difference between the fractions of the top 2 biome types is within 10 %.

974

975 **Table 1.** WRF-Chem model physics parameterization choices

976 **Table 2.** Perturbations of the transport ensemble

977 **Table 3.** Selected [CO2] towers and top 2 biome types

978 **Table 4.** Top 2 biome types in the ACT-America summer 2016 flight samples
979

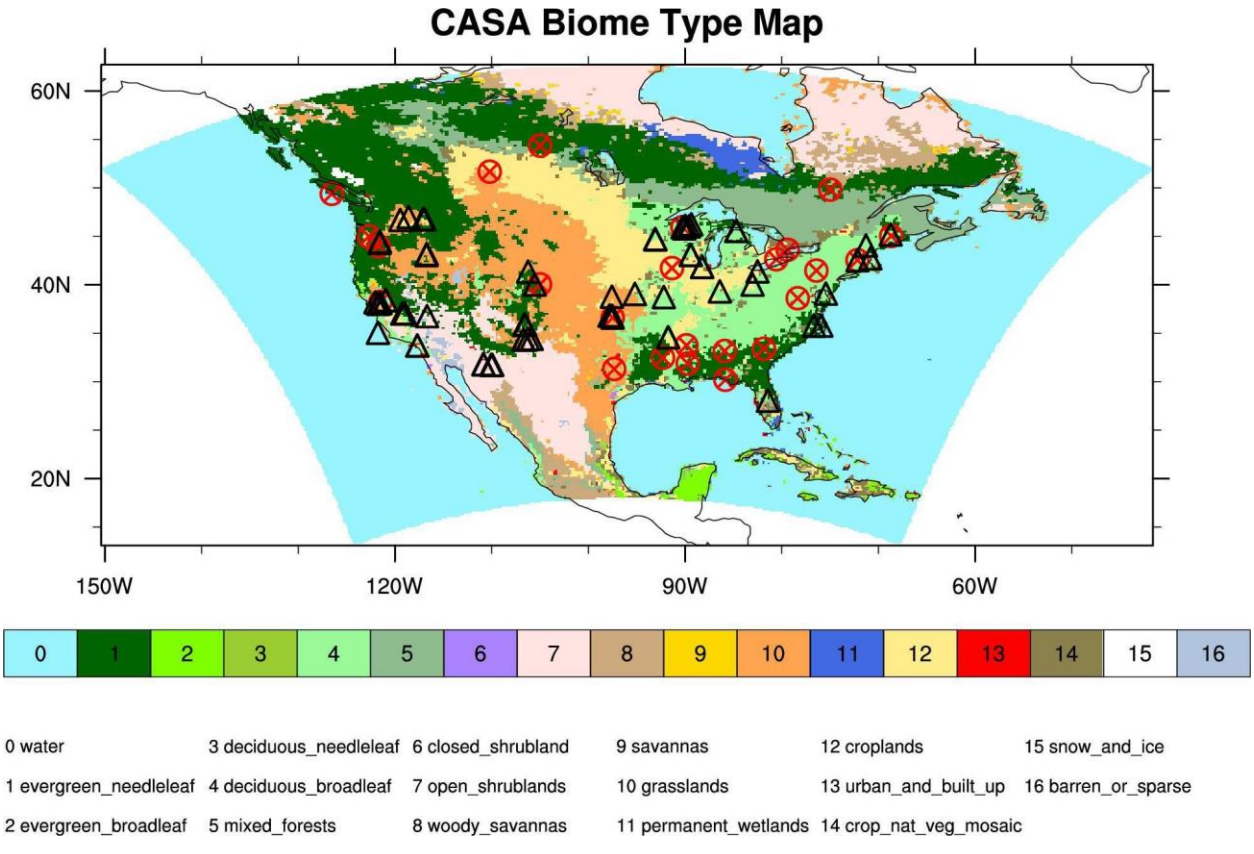

Figure 1. CASA biome type, and the locations of the [CO_2] towers (red circled crosses) and AmeriFlux flux towers (black triangles). The colored area is the model domain.

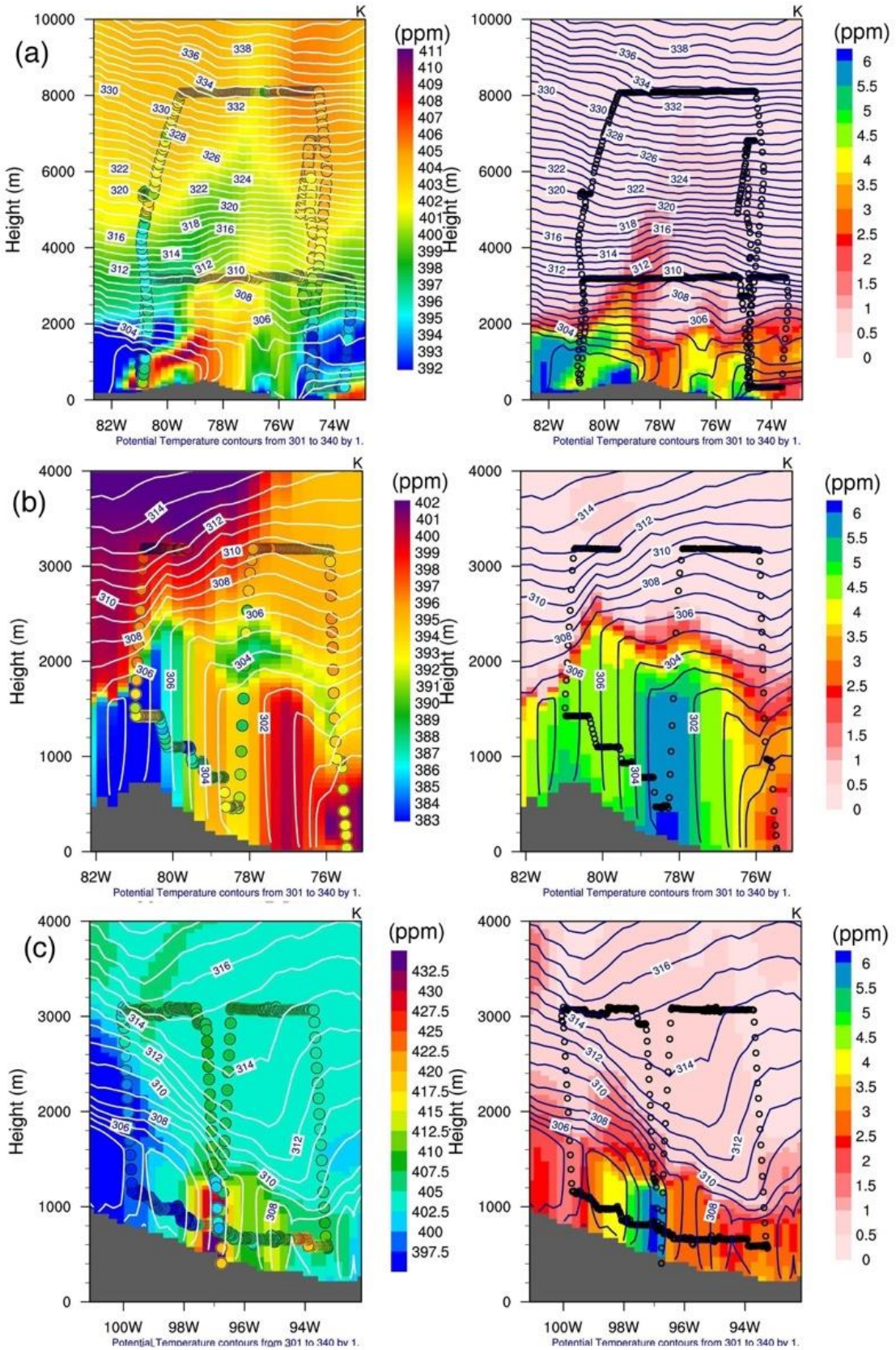

Figure 2. (to be continued)

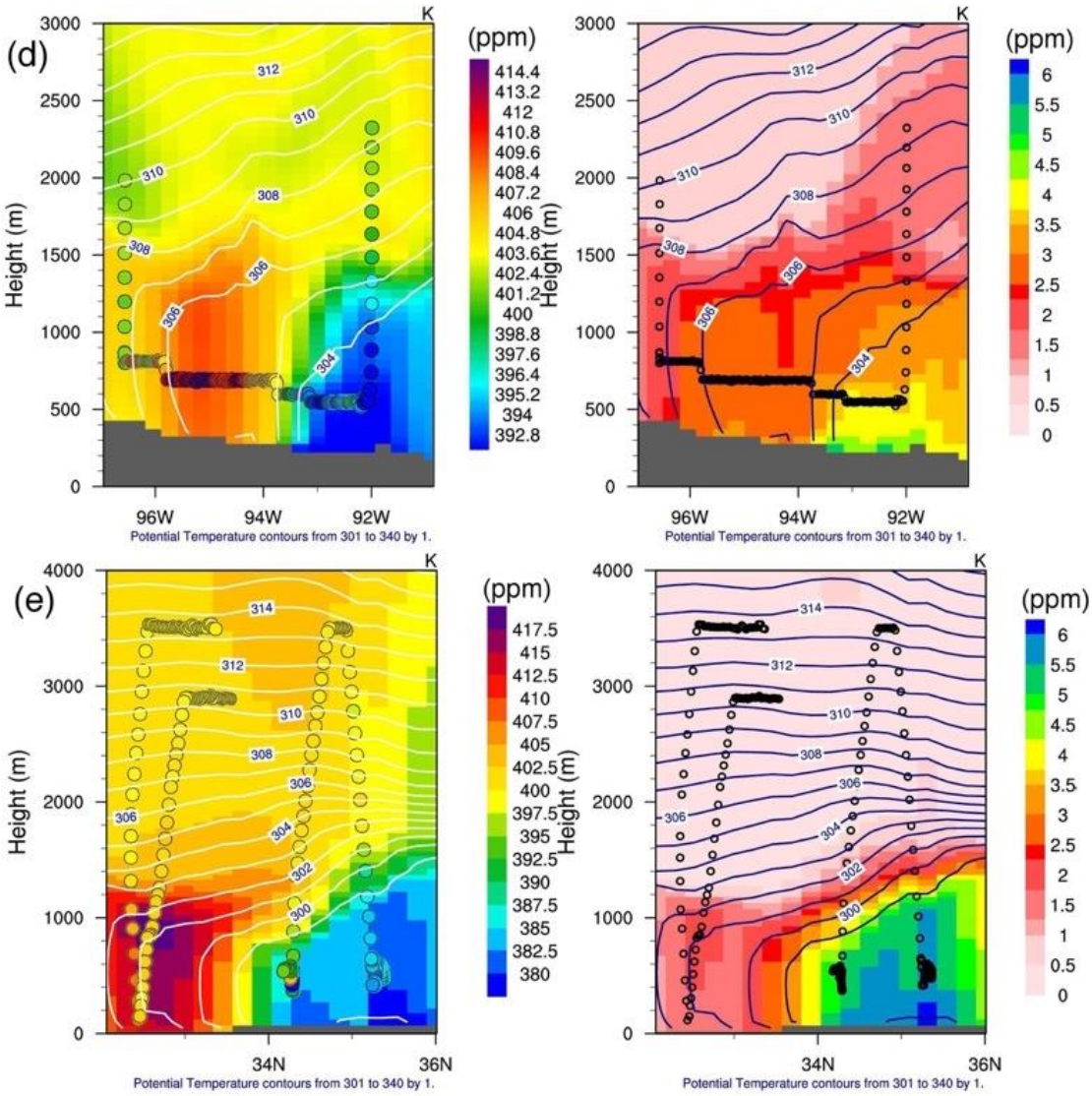

Figure 2. Simulated [CO_2] (left column) and the RMSD of the biogenic [CO_2] ([CO_2bio]) from the mean of modeled [CO_2bio] (right column) for five flights in the ACT Summer 2016 aircraft campaign: (a) 7/18 19 UTC (front in MA); (b) 7/21 19 UTC (fair Wx in MA): (c) 8/4 18 UTC (front in MW); (d) 8/9 19 UTC (fair Wx in MW); (e) 8/22 22 UTC (fair Wx in South). The aircraft sampled [CO_2] are overlaid with the simulations (left column). The black dots on the RMSD panels are the flight paths. Potential temperature contours are overlaid to indicate the frontal locations.

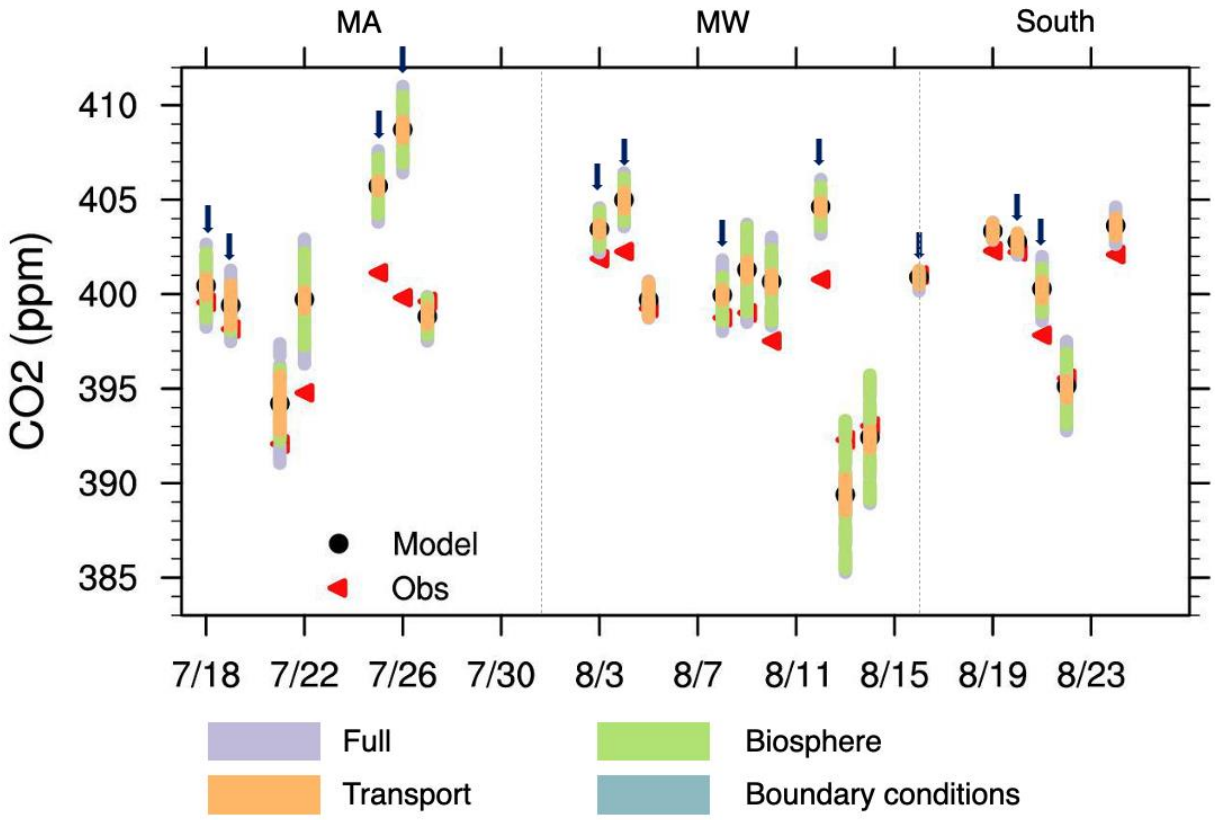

Figure 3. Daily afternoon averaged [CO_2] from the ACT aircraft and simulations. The error bars of the simulated [CO_2] are the spread of the full, biogenic flux, transport, and boundary condition ensembles. Note that the spread of boundary condition is very small and not visible in the figure. Only the ABL samples are included in the analysis. The point-down arrows denote the frontal cases during the campaign. Gray dashed lines denote the transit.

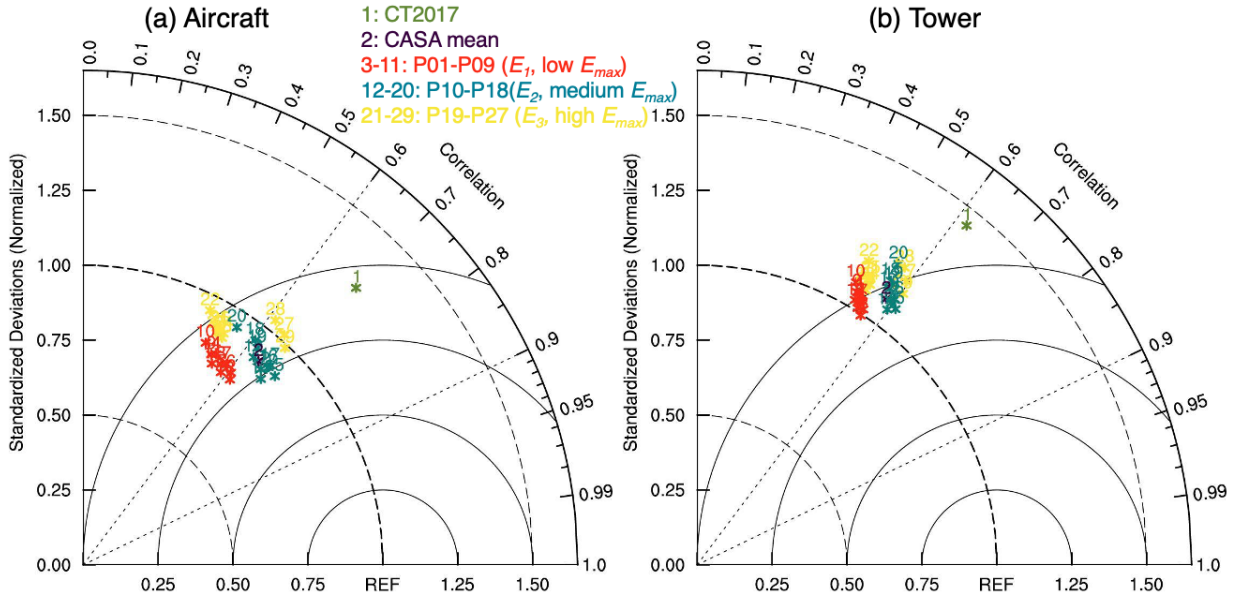

Figure 4. Taylor diagrams of simulated [CO_2] compared with the observations. The standard deviations of models are normalized by the standard deviations of observations ("REF"). The model [CO_2] are associated with the (29) biogenic flux members from CASA and CT2017. The color scheme indicates CT2017, CASA ensemble mean, CASA low E_{max} (E_1), CASA medium E_{max} (E_2), and CASA high E_{max} (E_3) groups. "P#" in the legend denotes the index of CASA ensemble member in the original product. The observations used in (a) are from the ABL legs of the ACT-America aircraft measurements. The flight dates and other information are listed in Table S1. The observations used in (b) are from the [CO_2] tower measurements at afternoon hours (19-22 UTC). The locations and information can be found in Figure 1 and Table 3. Both aircraft and tower [CO_2] data are kept at native temporal resolutions. Modeled [CO_2] are determined by the value of the nearest grid cell to sample locations. The period of interest is from July 18 to August 28, 2016 covering the ACT-America summer 2016 aircraft campaign.

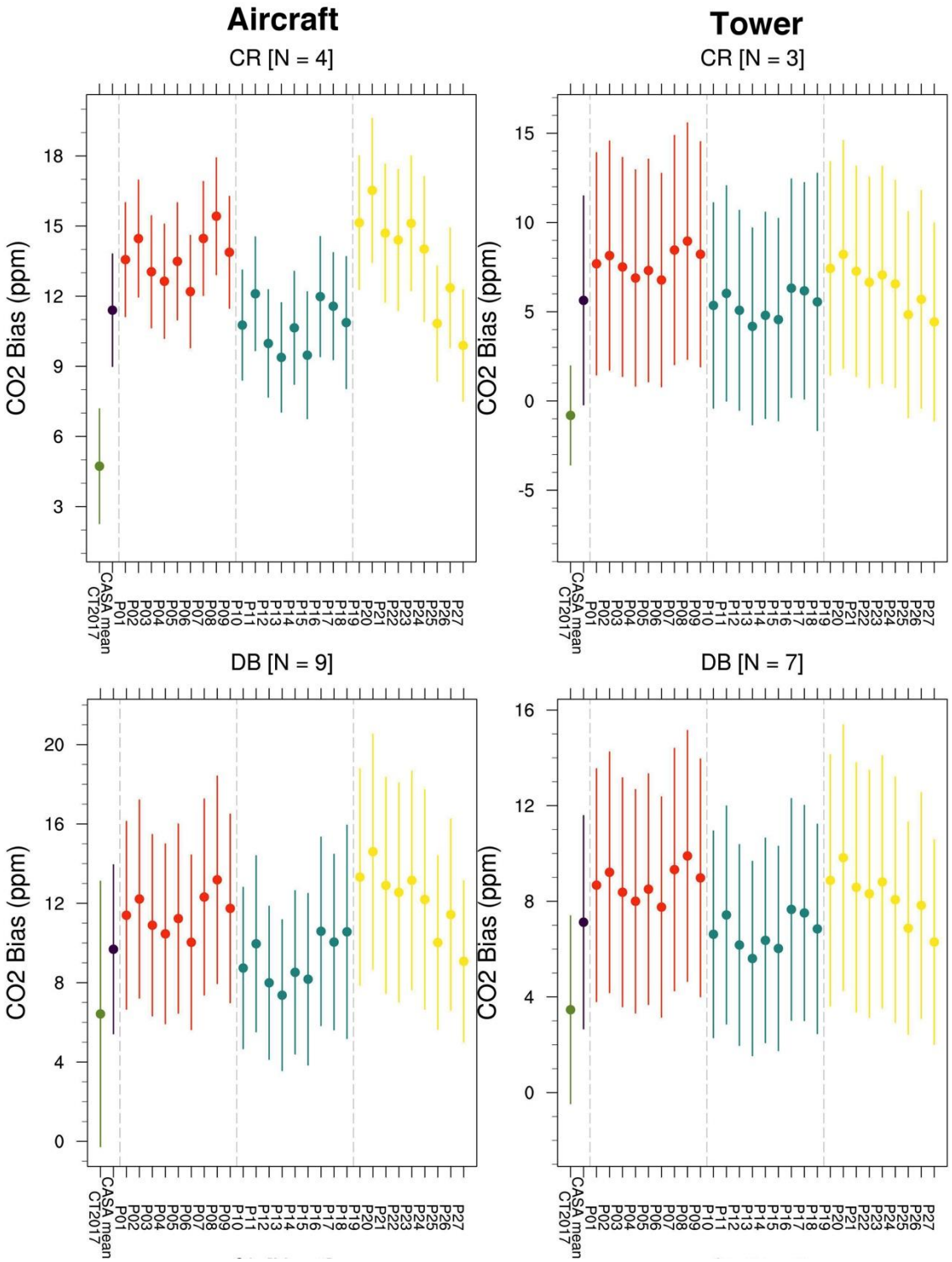

Figure 5. (to be continued)

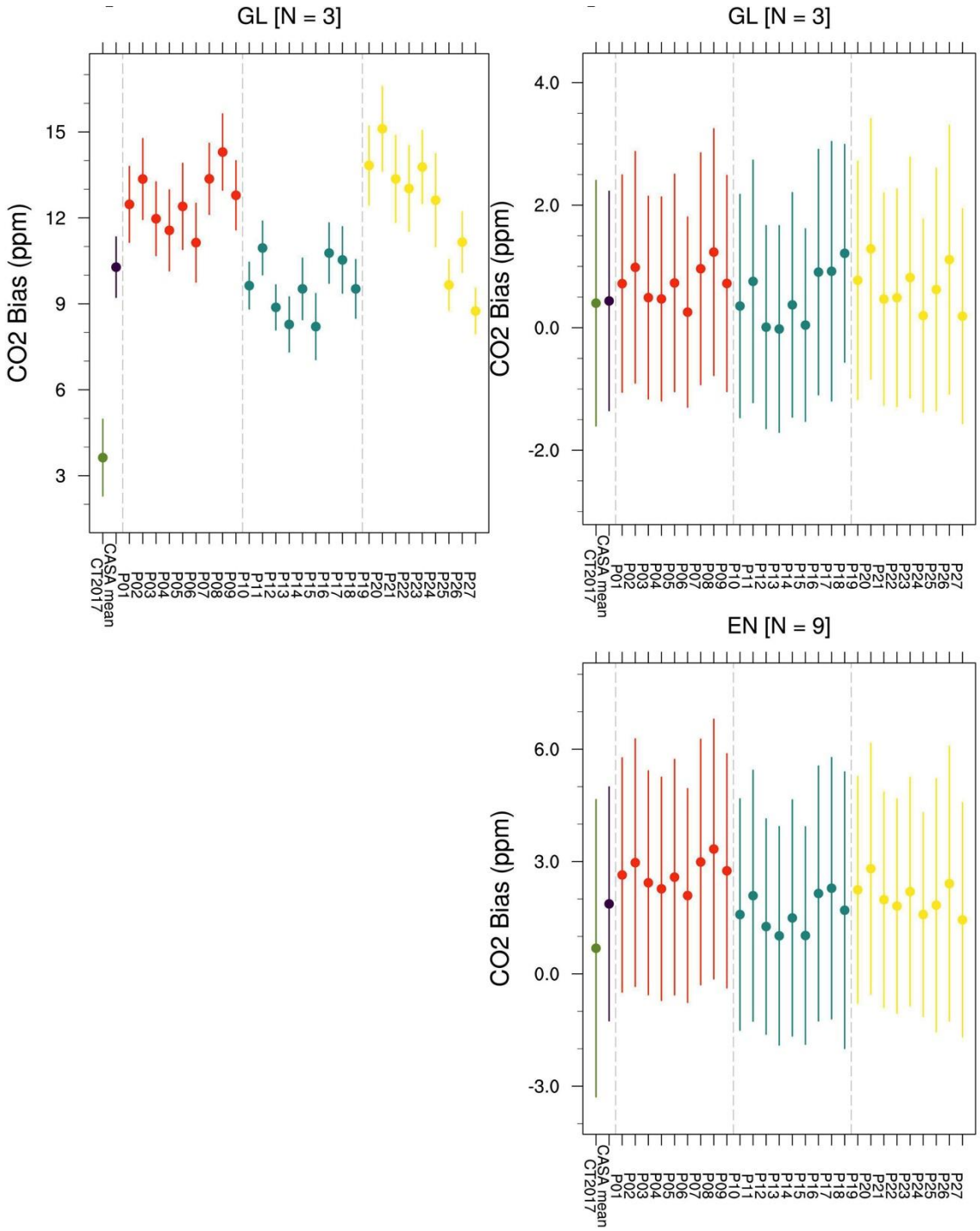

Figure 5. The model [CO_2] are associated with the (29) biogenic flux members from CASA and CT2017. The color scheme indicates CT2017, CASA ensemble mean, CASA low E_{max} (E_1), CASA medium E_{max} (E_2), and CASA high E_{max} (E_3) groups. "P#" on the x-axis are the indexes of CASA ensemble members in the original product.

Ranking	CR		DB*		GL		EN*
	Aircraft (N=4)	Tower (M=3)	(N=9)	Tower (M=7)	Aircraft (N=3)	Tower (M=3)	Tower (M=9)
1	CT2017	CT2017	CT2017	CT2017	CT2017	P12/224	CT2017
2	P13/022	P13/022	P13/022	P13/022	P15/222	P13/022	P13/022
3	P15/222	P27/236	P12/224	P15/222	P13/022	P15/222	P15/222
4	P27/236	P15/222	P15/222	P12/224	P27/236	P27/236	P12/224
5	P12/224	P14/-222	P14/-222	P27/236	P12/224	P24/232	P27/236
6	P14/-222	P25/036	P10/024	P14/-222	P18/226	P06/212	P14/-222
7	P10/024	P12/224	P27/236	P10/024	P14/-222	P10/024	P10/024
8	P25/036	P10/024	CASA mean	P18/226	P10/024	P14/-222	P24/232
9	P18/226	P18/226	P11/-224	P25/036	P25/036	CT2017	P18/226
10	CASA mean	CASA mean	P25/036	CASA mean	CASA mean	CASA mean	P22/032
11	P17/-226	P26/-236	P06/212	P11/-224	P17/-226	P21/234	P25/036
12	P16/026	P11/-224	P17/-226	P17/-226	P16/026	P04/012	CASA mean
13	P11/-224	P17/-226	P04/012	P16/026	P11/-224	P22/032	P21/234
14	P06/212	P16/026	P18/226	P06/212	P06/212	P03/214	P11/-224
15	P26/-236	P24/232	P16/026	P26/-236	P26/-236	P25/036	P06/212
16	P04/012	P22/032	P03/214	P04/012	P04/012	P01/014	P16/026
17	P03/214	P06/212	P05/-212	P24/232	P03/214	P09/216	P23/-232
18	P05/-212	P04/012	P01/014	P22/032	P05/-212	P05/-212	P19/034
19	P01/014	P23/-232	P26/-236	P03/214	P01/014	P11/-224	P04/012
20	P09/216	P21/234	P09/216	P05/-212	P24/232	P19/034	P17/-226
21	P24/232	P05/-212	P24/232	P21/234	P09/216	P23/-232	P26/-236
22	P22/032	P19/034	P02/-214	P01/014	P22/032	P16/026	P03/214
23	P02/-214	P03/214	P07/016	P23/-232	P02/-214	P17/-226	P05/-212
24	P07/016	P01/014	P22/032	P19/034	P21/234	P07/016	P01/014
25	P21/234	P02/-214	P21/234	P09/216	P07/016	P02/-214	P09/216
26	P23/-232	P20/-234	P23/-232	P02/-214	P23/-232	P26/-236	P20/-234
27	P19/034	P09/216	P08/-216	P07/016	P19/034	P18/226	P02/-214
28	P08/-216	P07/016	P19/034	P20/-234	P08/-216	P08/-216	P07/016
29	P20/-234	P08/-216	P20/-234	P08/-216	P20/-234	P20/-234	P08/-216

E1: Low Emax
E2: Medium Emax
E3: High Emax

*E2 = E3 for DB and EN

Figure 6. The final ranking of the modeled [CO_2] bias among the biogenic flux members shown in Figure 5. The model skills are ranked decreasingly, meaning the model biases increase from top to bottom. The name convention of the CASA ensemble members follows P# /$T_{opt}E_{max}Q_{10}$. # is the index of the CASA members in the original product. At the position of T_{opt}, "0" means the default value; "2" means the default value plus 2 degrees C; "-2" means the default value minus 2 degrees C. For E_{max}, "1", “2”, and “3” denotes the low, medium, and high E_{max} in Table 4 of Zhou et al. (2020). "2","4","6" at the Q_{10} position indicate that Q_{10} = "1.2", "1.4", and "1.6", respectively.

Ranking	CR		DB*		GL		EN*
	Aircraft (N=4)	Tower (M=3)	Aircraft (N=9)	Tower (M=7)	Aircraft (N=3)	Tower (M=3)	Tower (M=9)
1	CT2017	CT2017	CT2017	CT2017	CT2017	P12/224	CT2017
2	P13/022	P13/022	P13/022	P13/022	P15/222	P13/022	P13/022
3	P15/222	P27/236	P12/224	P15/222	P13/022	P15/222	P15/222
4	P27/236	P15/222	P15/222	P12/224	P27/236	P27/236	P12/224
5	P12/224	P14/-222	P14/-222	P27/236	P12/224	P24/232	P27/236
6	P14/-222	P25/036	P10/024	P14/-222	P18/226	P06/212	P14/-222
7	P10/024	P12/224	P27/236	P10/024	P14/-222	P10/024	P10/024
8	P25/036	P10/024	CASA mean	P18/226	P10/024	P14/-222	P24/232
9	P18/226	P18/226	P11/-224	P25/036	P25/036	CT2017	P18/226
10	CASA mean	CASA mean	P25/036	CASA mean	CASA mean	CASA mean	P22/032
11	P17/-226	P26/-236	P06/212	P11/-224	P17/-226	P21/234	P25/036

*E2 = E3 for DB and EN

Emax = E2* && Q10 = 1.2 | Emax = E2* && Q10 = 1.4 | Emax = E2* && Q10 = 1.6 | Others

Figure 7. Top 11 biogenic flux members in Figure 6 applied different color schemes in order to reveal the secondary parameter, Q_{10}. Note that the flux members associated in the medium E_{max} are with color background; others are with white background.

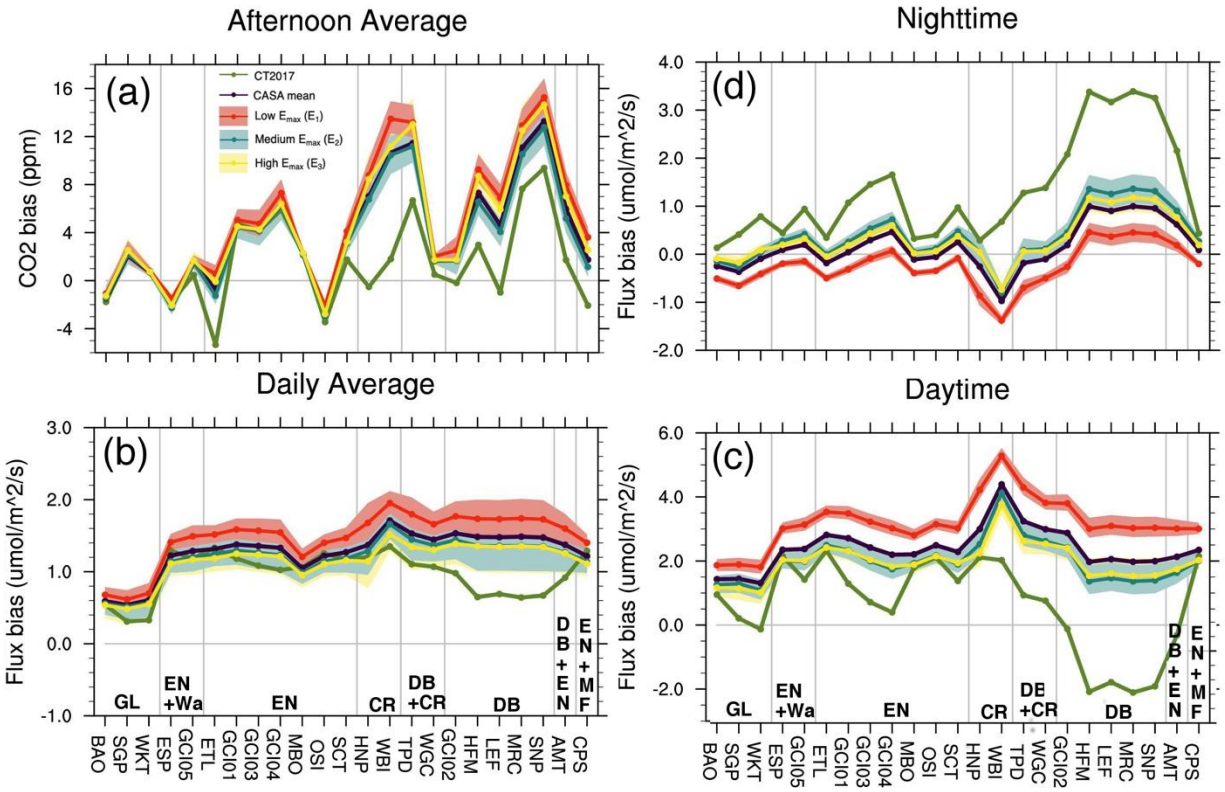

Figure 8. CO_2 mole fraction and flux biases at the [CO_2] towers. (a) Afternoon averaged (19-22 UTC) [CO_2] biases. (b) daily, (c) daytime (15-2 UTC), and (c) nighttime (3-14 UTC) averages of the flux biases derived from Eq. (1). The shaded areas are the minimum and maximum biases among different E_{max} groups in the CASA ensemble. The associated solid lines are the mean of the given E_{max} group. Additionally, CT2017 and the CASA ensemble mean are denoted in green and purple dotted lines. The horizontal gray line is unbiased. The vertical gray lines divide the towers into various biomes determined by the footprint analysis. The dominate biomes, GL - grassland, EN - evergreen needleleaf forest, Wa - water, CR - croplands, DB - deciduous broadleaf forest, MF - mixed forest for [CO_2] towers are listed in Table 3. A mixed type is listed if the difference between the fractions of the top 2 biome types is within 10 %.

Table 1. WRF-Chem model physics parameterization choices

Parameterization	Option used	Reference
Microphysics	Thompson	Thompson et al. (2004)
Longwave Radiation	RRTMG longwave scheme	Iacono et al. (2008)
Shortwave Radiation	RRTMG shortwave scheme	Iacono et al. (2008)
PBL Scheme	MYNN2	Nakanishi & Niino (2006)
Land Surface	Unified Noah land-surface model	Chen & Dudhia (2001)
Cumulus	Kain-Fritsch (new Eta) scheme	Kain (2004)

Table 2. Perturbations of the transport ensemble

Transport run	PBL scheme	Surface layer	LSM	SKEBS
1	MYNN	MYNN	Noah	N
2	MYNN	MYNN	Noah	Y
3	MYNN	MYNN	Noah	Y
4	MYNN	MYNN	Noah	Y
5	MYJ	Eta	RUC	N
6	MYJ	Eta	RUC	Y
7	MYJ	Eta	RUC	Y
8	YSU	MM5	Thermal	N
9	YSU	MM5	Thermal	Y
10	YSU	MM5	Thermal	Y

1068
1069 Table 3. Selected [CO_2] towers and top 2 biome types

Site	Lat	Lon	Intake elevation (m ASL)	1st Biome type (Fraction %)	2nd Biome type (Fraction %)
AMT	45.03	-68.68	160	DB (42.1)	EN (32.2)
BAO	40.05	-105	1884	GL (62.8)	EN (36.4)
CPS	49.82	-74.98	389	EN (46.3)	MF (32.2)
ESP	49.38	-126.54	47	EN (49.6)	Wa (47.1)
ETL	54.35	-104.99	597	EN (67.8)	CR (14.0)
GCI01	32.47	-92.28	165	EN (61.2)	DB (22.3)
GCI02	33.75	-89.85	205	DB (56.2)	CR (27.3)
GCI03	31.89	-89.73	232	EN (57.9)	DB (33.9)
GCI04	33.18	-85.89	428	EN (58.7)	DB (40.5)
GCI05	30.2	-85.83	105	Wa (41.3)	EN (41.3)
HFM	42.54	-72.17	369	DB (87.6)	Wa (10.0)
HNP	43.61	-79.39	97	CM (43.8)	DB (17.4)
LEF	45.95	-90.27	868	DB (81.8)	Wa (7.4)
MBO	43.98	-121.69	2742	EN (73.6)	GL (20.6)
MRC	41.47	-76.42	652	DB (98.3)	CM (1.7)
OSI	45	-122.7	620	EN (74.4)	Wa (12.4)
SCT	33.41	-81.83	420	EN (78.5)	DB (17.4)
SGP	36.61	-97.49	374	GL (77.7)	DB (12.4)
SNP	38.62	-78.35	1025	DB (93.4)	EN (3.3)
TPD	42.64	-80.56	266	DB (32.2)	CR (27.3)
WBI	41.72	-91.35	621	CR (76.9)	DB (17.4)
WGC	38.27	-121.49	483	DB (26.4)	CR (24.0)
WKT	31.31	-97.33	708	GL (67.8)	DB (22.3)

CM - Cropland natural vegetation mosaic; CR – Croplands; DB - Deciduous broadleaf forest; EN - Evergreen needleleaf forest; GL – Grasslands; MF - Mixed forest; Wa - Water.

1070
1071

1072
1073
1074 Table 4. Top 2 biome types in the ACT-America summer 2016 flight samples

Date	1st Biome type (Fraction %)	2nd Biome type (Fraction %)
2016-07-18*	DB (60.1)	CR (17.4)
2016-07-19	DB (40.7)	EN (20.7)
2016-07-21*	DB (54.6)	Wa (15.7)
2016-07-22	DB (60.7)	EN (16.8)
2016-07-25	DB (69.0)	EN (16.4)
2016-07-26	DB (61.0)	CR (17.6)
2016-07-27	DB (30.1)	EN (19.1)
2016-08-03	GL (38.0)	OS (24.5)
2016-08-04*	GL (40.3)	Wa (25.5)
2016-08-05	GL (40.9)	Wa (17.5)
2016-08-08	DB (34.6)	CR (24.1)
2016-08-09*	DB (44.0)	GL (21.6)
2016-08-10	DB (35.7)	CR (20.1)
2016-08-12	Wa (25.1)	CR (20.1)
2016-08-13	CR (72.6)	GL (12.0)
2016-08-14	CR (62.7)	EN (9.0)
2016-08-16	Wa (71.3)	DB (9.4)
2016-08-19	Wa (93.3)	DB (3.3)
2016-08-20	Wa (66.2)	CR (14.1)
2016-08-21	CR (42.4)	Wa (22.0)
2016-08-22*	CR (41.3)	DB (28.4)
2016-08-24	Wa (83.9)	EN (5.2)

CR – Croplands; DB - Deciduous broadleaf forest; EN - Evergreen needleleaf forest; GL – Grasslands; OS - Open Shrublands; Wa - Water.

1075
1076
1077