2022

A Modified vegetation photosynthesis and respiration model (VPRM) for the Eastern USA and Canada, evaluated with comparison to atmospheric observations and other biospheric models

Sharon M. Gourdji
National Institute of Standards and Technology

Anna Karion
National Institute of Standards and Technology

Israel Lopez-Coto
National Institute of Standards and Technology

Subhomoy Ghosh
National Institute of Standards and Technology

Kimberly L. Mueller
National Institute of Standards and Technology

Follow this and additional works at: https://commons.clarku.edu/faculty_geography

See next page for additional authors
Part of the Geography Commons

Authors

Sharon M. Gourdji, Anna Karion, Israel Lopez-Coto, Subhomoy Ghosh, Kimberly L. Mueller, Yu Zhou, Christopher A. Williams, Ian T. Baker, Katharine D. Haynes, and James R. Whetstone

Author Contributions:
Conceptualization: Sharon M. Gourdji, Anna Karion
Data curation: Anna Karion
Formal analysis: Sharon M. Gourdji
Funding acquisition: James R. Whetstone
Investigation: Sharon M. Gourdji, Anna Karion, Israel Lopez-Coto, Subhomoy Ghosh, Kimberly L. Mueller, Yu Zhou, Christopher A. Williams, Ian T. Baker, Katharine D. Haynes

A Modified Vegetation Photosynthesis and Respiration Model (VPRM) for the Eastern USA and Canada, Evaluated With Comparison to Atmospheric Observations and Other Biospheric Models

Sharon M. Gourdji[1] , Anna Karion[1] , Israel Lopez-Coto[1] , Subhomoy Ghosh[1,2] , Kimberly L. Mueller[1] , Yu Zhou[3,4] , Christopher A. Williams[3] , Ian T. Baker[5] , Katharine D. Haynes[5] , and James R. Whetstone[1]

[1]National Institute of Standards & Technology, Gaithersburg, MD, USA, [2]Center for Research Computing, University of Notre Dame, South Bend, IN, USA, [3]Graduate School of Geography, Clark University, Worcester, MA, USA, [4]College of Engineering, Forestry and Natural Sciences, Northern Arizona University, Flagstaff, AZ, USA, [5]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

Abstract Atmospheric CO_2 measurements from a dense surface network can help to evaluate terrestrial biosphere model (TBM) simulations of Net Ecosystem Exchange (NEE) with two key benefits. First, gridded CO_2 flux estimates can be evaluated over regional scales, not possible using flux tower observations at discrete locations for model evaluation. Second, TBM ability to explain atmospheric CO_2 fluctuations due to the biosphere can be directly tested, an important objective for anthropogenic emissions monitoring using atmospheric observations. Here, we customize the Vegetation Photosynthesis and Respiration Model (VPRM) for an eastern North American domain with strong biological activity upwind of urban areas. Parameters are optimized using flux tower observations from a historical database with sites in (and near) the domain. In addition, the respiration model (originally a linear function of temperature) is modified to account for impacts of changing foliage, non-linear temperature, and water stress. Flux estimates from VPRM, the Carnegie-Ames-Stanford Approach (CASA) model and the Simple Biosphere Model v4 (SiB4), are convolved with footprints from atmospheric transport models for evaluation with CO_2 observations at 21 towers in the domain, with roughly half of the towers used here for the first time. Results show that the new respiration model in VPRM helps to correct a growing season sink bias in the atmosphere associated with underestimated summertime respiration using the original model with annual parameters. The new VPRM also better explains fine-scale atmospheric CO_2 variability compared to other TBMs, due to higher resolution diagnostic phenology, the new respiration model, domain-specific parameters, and high-quality input data sets.

Plain Language Summary Photosynthesis and respiration from vegetation and soils contribute to large CO_2 fluctuations in the atmosphere, which mix with CO_2 sources from fossil fuel combustion. Terrestrial biosphere models simulate biological carbon exchange with the atmosphere, which can then be evaluated with atmospheric CO_2 measurements. In this study, we customize a high resolution, data-driven biospheric model, the Vegetation Photosynthesis and Respiration Model (VPRM), for eastern North America, a region with strong biological activity from crops and forests as well as large emission sources. The model equation describing sources to the atmosphere from respiration (i.e., "breathing" from plants and decaying organic matter) is modified to account for increases in foliage and crop biomass during the growing season. Comparisons with other process-based biospheric models and atmospheric CO_2 observations show that the new VPRM model is relatively unbiased and better explains small-scale biospheric CO_2 fluctuations in the atmosphere in this domain compared to other more complex models.

1. Introduction

Carbon dioxide (CO_2) surface fluxes from the terrestrial biosphere produce a large and variable signal in the atmosphere during the growing season due to photosynthesis and ecosystem respiration, especially in biologically productive areas like croplands and forests (Shiga et al., 2014), but also in heavily populated areas containing parks, lawns, gardens, street trees and urban agriculture (Buyantuyev & Wu, 2009; Golubiewski, 2006; Nowak & Crane, 2002; Raciti et al., 2014). Even in the winter (i.e., December to February in the northern hemisphere), CO_2

Methodology: Sharon M. Gourdji, Israel Lopez-Coto
Resources: Anna Karion, Israel Lopez-Coto, Yu Zhou, Christopher A. Williams, Ian T. Baker, Katharine D. Haynes
Software: Israel Lopez-Coto
Supervision: James R. Whetstone
Validation: Sharon M. Gourdji
Visualization: Sharon M. Gourdji
Writing – original draft: Sharon M. Gourdji
Writing – review & editing: Sharon M. Gourdji, Anna Karion, Israel Lopez-Coto, Subhomoy Ghosh, Kimberly L. Mueller, Yu Zhou, Christopher A. Williams, Ian T. Baker, Katharine D. Haynes, James R. Whetstone

sources from ecosystem respiration can have an atmospheric signal as large as that from fossil fuel emissions, especially when integrated over large areas. Therefore, to monitor fossil fuel emissions using atmospheric data, fluctuations in atmospheric CO_2 due to the biosphere must be appropriately modeled and subtracted out from total observed CO_2 mole fractions.

The eastern half of North America, upwind of the heavily populated Northeast Corridor (NEC) of the United States (from Washington D.C. to Boston, MA), is a biologically productive region containing Appalachian deciduous forests, northern mixed forests and southern pine plantations, croplands in the Midwestern Corn Belt and Mississippi river valley, grasslands in Kentucky and Tennessee and coastal and northern wetlands (Figure 1). Each of these ecosystems has strong seasonal signals with varying timing, such that this region provides a valuable case study for testing the ability of terrestrial biosphere models (TBMs) to explain atmospheric CO_2 variability. This region is also particularly data-rich, given a large historical database of flux towers operating in the region with direct observations of Net Ecosystem Exchange (NEE, in ~1 km^2 footprints around each tower) that can be used to parameterize models, as well as an increasingly dense network of surface towers measuring in situ atmospheric CO_2 mole fractions, which can be used for model evaluation.

In this study, we first customize and improve a relatively simple empirical model with a standard light-use efficiency formulation for photosynthesis, that is, the Vegetation Photosynthesis and Respiration Model (VPRM; Mahadevan et al., 2008), for a domain in the eastern United States and Canada (−92°W to −68°W, 33°N to 47°N, Figure 1). Parameters in VPRM are optimized using flux tower observations of NEE collected in and near the domain, thus making it well suited to take advantage of a large amount of data from historical flux towers (i.e., 69 in this study) operating in the region. In the original Mahadevan et al. (2008) formulation, VPRM has a relatively simplistic respiration equation, expressed as a baseline value and linear dependence on temperature. Therefore, we also introduce here a modified VPRM respiration model that includes additional variables: (a) a non-linear temperature response, (b) a vegetation index to better capture seasonality in autotrophic respiration, and (c) a water stress scaling factor and its interactions with temperature to capture drought and soil moisture effects.

VPRM is first run historically using site-specific meteorological data at the flux tower locations, evaluating flux estimates with observations reserved from the parameter optimization. In this manner, we distinguish between various model setups, choosing the best ones for further evaluation. For three chosen setups, VPRM is then run in a gridded implementation at a 0.02°, hourly resolution from November 2016 to October 2017. Gridded flux estimates are evaluated by comparison with output from two other TBMs, as well as with comparison to near-surface atmospheric CO_2 mole fraction observations at 21 surface towers in the domain (Figure 2).

Unlike flux tower observations, atmospheric CO_2 measurements can help evaluate regional-scale gridded CO_2 flux estimates integrated in time and space across the landscape. In fact, despite decades of research into physiological leaf and canopy scale carbon cycling, regional-scale estimates of CO_2 flux have remained highly uncertain in part due to the lack of validation data to date at these scales (e.g., Huntzinger et al., 2012). The dense atmospheric CO_2 observing network currently available for this domain can thus help to close this gap in process-based understanding of landscape-level CO_2 fluxes, as called for in the 2021 North American Carbon Program Science Implementation Plan (Sections 3.2–3.4 in Williams et al., 2021), that is, the need for "studies that bridge understanding of surface CO_2 fluxes across scales, due to the varying importance of ecosystem processes operating at different scales."

Two TBMs used for intercomparison here are commonly used in North America, especially as priors in atmospheric inversion studies (e.g., L. Hu et al., 2019). These are the Carnegie-Ames Stanford Approach (CASA) model (Potter et al., 1993; Randerson et al., 1996; Zhou, Williams, Lauvaux, Davis, et al., 2020) and the Simple Biosphere model, version 4 (SiB4; Haynes, Baker, Denning, Stöckli, et al., 2019; Haynes, Baker, Denning, Wolf, et al., 2019; Sellers et al., 1986, 1996). VPRM, CASA, and SiB4 represent a range of biospheric modeling approaches that vary from the most empirical (VPRM) to the most mechanistic (SiB4), allowing for an investigation of the relative impact of model complexity versus other factors in explaining atmospheric CO_2 variability. Thus, this comparison also addresses another component of the NACP Science Implementation Plan, that is, the need for improved statistical methods and evaluation data sets for assessing increased model complexity (including the assessment of "trade-offs between increasing model complexity and measurable improvement of model reliability," Williams et al., 2021).
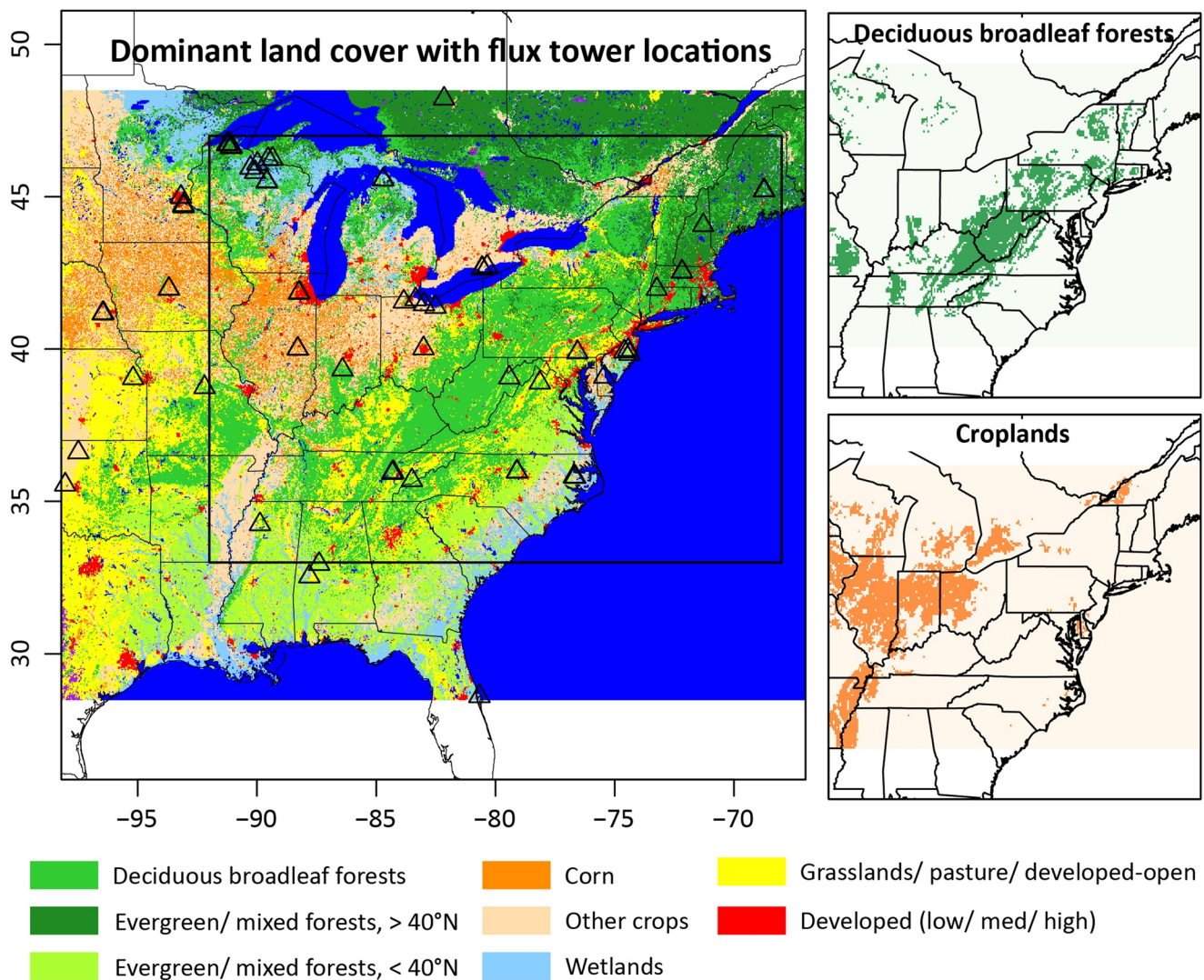
Figure 1. Map of dominant land cover at 0.02° resolution for eight plant functional types (PFTs), with a rectangle around the flux simulation domain (left panel). (The shrubland PFT is not shown here because it is not the dominant land cover for any pixel at this spatial resolution.) Also shown are the locations of the 69 flux towers included in the parameter optimization (Table S1 in Supporting Information S1). Deciduous broadleaf forest and cropland pixels used for spatial aggregation of gridded fluxes are shown in the panel on the right at 0.1°, where these pixels contain more than 50% coverage in the underlying land cover maps for VPRM and CASA, and 25% coverage for SiB4 at 0.5°.

CO_2 fluxes from each of the TBMs are multiplied by footprints from an atmospheric transport and dispersion model to simulate atmospheric CO_2 enhancements due to the biosphere (using mean enhancements from two transport models to help reduce transport model uncertainty). Simulated and observed CO_2 enhancements are then compared across towers and throughout the year to assess the TBMs' ability to reproduce spatial gradients across towers, seasonality, and finer-scale (diurnal and synoptic) temporal variability in the atmosphere.

By improving a single relatively data-driven biospheric model (i.e., VPRM) and evaluating flux estimates from multiple TBMs using a large atmospheric CO_2 data set from multiple towers in this domain, this study aims to provide a prototype framework for continuous model evaluation and improvement. From such a feedback loop, we begin initial steps toward outlining the key features of a TBM in eastern North America that will operate well across spatial and temporal scales to explain fine-scale atmospheric CO_2 variability due to the biosphere for the ultimate purpose of emissions monitoring.

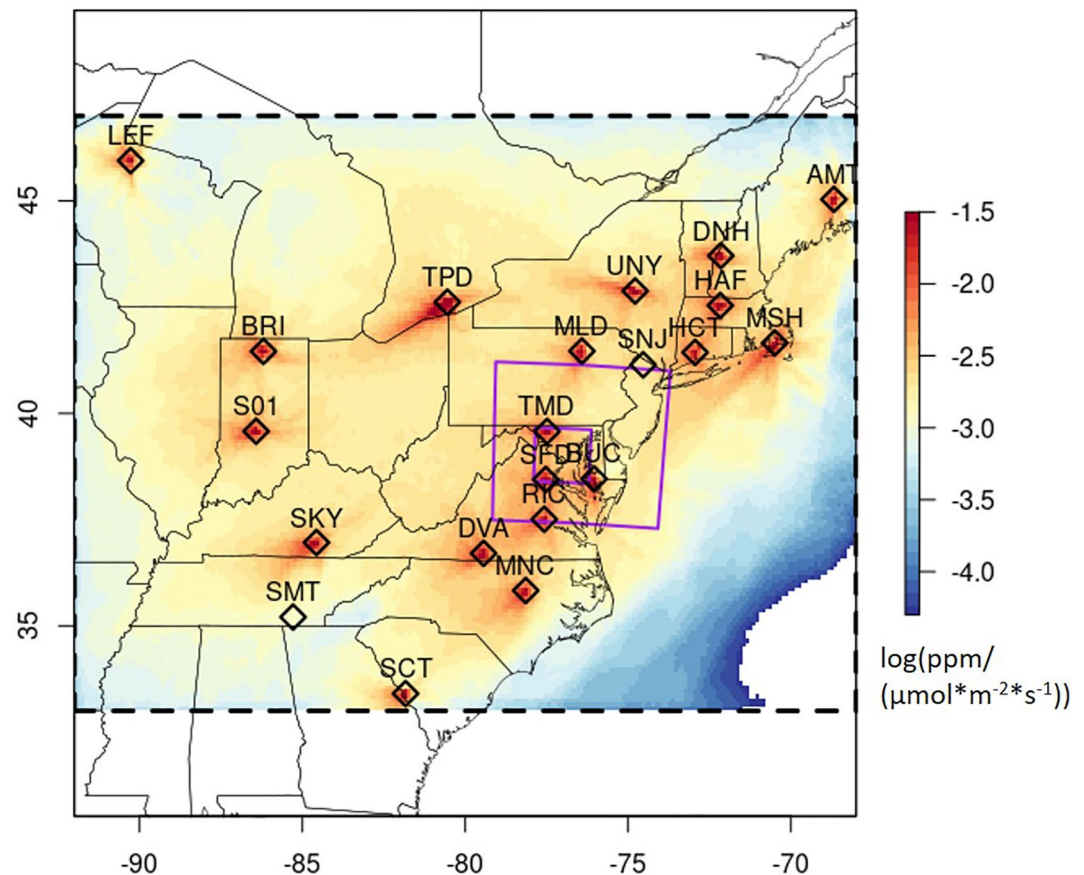**Figure 2.** Map of CO_2 observational towers and mean July 2017 afternoon integrated footprints (summed across towers and all hours back from observation time and then averaged across observation times and WRF-STILT and NAMS-STILT footprints). The inner nests for the WRF simulation are shown in purple. SNJ and SMT had no CO_2 observations in July 2017; therefore, their footprints are not included in the map for this month.

2. Methods

In this section, we first describe the original VPRM model, and then its specific implementation and modifications for this study. We also briefly describe the CASA and SiB4 models used to assess uncertainty across relatively independent TBMs. Finally, we describe the flux tower observations, atmospheric CO_2, and ancillary data sets used for model evaluation and outline overall analyses for assessing model performance.

2.1. Vegetation Photosynthesis and Respiration Model

TBMs typically estimate carbon sources to the atmosphere from ecosystem respiration (R_e, or the sum of autotrophic, R_a, and heterotrophic, R_h, respiration) and carbon uptake from photosynthesis (also known as Gross Ecosystem Exchange, or GEE), with NEE defined as their sum (i.e., R_e + GEE). Most TBMs (including VPRM) also differentiate flux dynamics across land cover categories, referred to as plant functional types (PFTs), and then merge PFT-specific fluxes into a set of total flux estimates across the landscape. Where TBMs differ, they do so in their physiological representations of GEE and R_e, spatial and temporal resolutions, input data sets, PFT classifications, and phenology schemes.

The VPRM implemented here is a diagnostic light-use efficiency model that calculates GEE and R_e using remote-sensing and meteorological model inputs, along with parameters optimized using flux tower observations for each of the PFTs shown in Figure 1. Thus, it is a relatively empirical, data-driven model compared to more process-based TBMs developed for the purpose of simulating long-term carbon balance or coupling with global climate models (Fisher et al., 2014). The original model from Mahadevan et al. (2008) was deliberately simplistic,

with the assumption that parameters and/or fluxes would be optimized in atmospheric inverse models. In particular, the respiration equation is a simple linear function of temperature, which does not account for seasonal or spatial variability in biomass and litter inputs to soil carbon pools.

2.1.1. VPRM Model Equations

The VPRM photosynthesis algorithm has a standard light-use efficiency formulation, with GEE defined as:

$$GEE = \lambda * T_{scale} * P_{scale} * W_{scale1} \frac{1}{\left(1 + \frac{PAR}{PAR_0}\right)} * PAR * EVI \tag{1}$$

where λ is a potential light-use efficiency factor, T_{scale}, P_{scale}, and W_{scale1} are temperature, phenology, and water stress scaling factors, as defined in Mahadevan et al. (2008), PAR is Photosynthetically Active Radiation, PAR_0 is the half-saturation constant of PAR, and EVI is the remotely sensed Enhanced Vegetation Index. The term $\frac{1}{\left(1 + \frac{PAR}{PAR_0}\right)}$ is a scaling factor describing the efficiency of photosynthesis at high light levels (which relates to canopy structure). Overall, PAR*EVI represents the amount of absorbed radiation, with GEE modeled as potential uptake (i.e., λ*PAR*EVI) downregulated by each of the scaling factors (which vary from 0 to 1). Parameters optimized with flux tower data for each PFT include λ and PAR_0, as well as the optimal temperature (T_{opt}) in the T_{scale} equation:

$$T_{scale} = \frac{(T - T_{min})(T - T_{max})}{[(T - T_{min})(T - T_{max}) - (T - T_{opt})^2]} \tag{2}$$

T_{min} and T_{max} are set equal to 0 and 45°C respectively in this study (i.e., temperatures below or above which photosynthesis shuts down).

W_{scale1} is defined for all PFTs as:

$$W_{scale1} = \frac{1 + LSWI}{1 + LSWI_{max}} \tag{3}$$

where LSWI is the remotely sensed Land Surface Water Index (Chandrasekar et al., 2010) calculated using bands 2 and 6 from MODIS reflectance data, and $LSWI_{max}$ is the site-specific maximum daily LSWI from May to October (derived from a multi-year mean; X. Xiao et al., 2004).

The original VPRM model formulation estimates R_e as a baseline value plus a linear function of temperature:

$$R_e = \beta + \alpha * T_{air} \tag{4}$$

where T_{air} is the surface air temperature (in °C), β is the baseline respiration (in $\mu mol*m^{-2}*s^{-1}$), α is the temperature sensitivity of respiration (in $\mu mol*m^{-2}*s^{-1}/°C$), and α and β are optimized parameters. In Mahadevan et al. (2008), T_{air} below a threshold (T_{low}) is set equal to T_{low} to account for continued R_e in winter, when soils remain warmer than air temperatures. Here, we just set predicted negative R_e values to 0.

To help improve the respiration model beyond this simple linear formulation, we developed a new R_e equation that incorporates additional predictor variables. To define this new respiration model (Equation 5), various multivariate regressions were run against observed nighttime mean NEE at the flux towers to determine a single equation that is consistent with physiologic understanding of autotrophic and heterotrophic respiration and improves model fit across PFTs. Candidate models were evaluated by comparing adjusted R^2s across different sets of potential covariates, which, like other model selection algorithms, penalizes the addition of spurious predictor variables. Biases in monthly mean seasonal and diurnal cycles were also assessed to distinguish between candidate models.

Following this procedure, the equation for R_e is updated as follows:

$$R_e = \beta + \alpha_1 * T' + \alpha_2 * T'^2 + \gamma * EVI + \theta_1 * W_{scale2} + \theta_2 * W_{scale2} * T' + \theta_3 * W_{scale2} * T'^2 \tag{5}$$

where α_1, α_2, γ, θ_1, θ_2, and θ_3 are optimized parameters, and T' is a modified air temperature variable intended to capture soil temperatures that remain warmer than air temperatures in winter:

$$T_{air} < T_{crit} : \quad T' = T_{crit} - T_{mult} * (T_{crit} - T_{air}) \tag{6}$$

$$T_{air} \geq T_{crit} : \quad T' = T_{air} \tag{7}$$

where T_{crit} is a low-temperature threshold (in °C) and T_{mult} is a scalar from 0 to 1 which is multiplied by air temperatures below T_{crit}. This modification of low air temperatures for R_e is like the fixed T_{low} threshold in Mahadevan et al. (2008), but here T' can still slope downward at low temperatures.

In the new respiration model, a slightly different water stress scaling factor was used than in the GEE equation, with W_{scale2} defined as a normalized LSWI:

$$W_{scale2} = \frac{LSWI - LSWI_{min}}{LSWI_{max} - LSWI_{min}} \tag{8}$$

where $LSWI_{max}$ is as defined previously and $LSWI_{min}$ is the site-specific minimum LSWI across a full year (from a multi-year mean). W_{scale2} gave a slightly better model fit than W_{scale1} for grasslands, soybean/other crops, and shrubland, and equivalent fits for other PFTs; hence this definition of W_{scale} was chosen for the respiration equation.

In the final R_e equation, the squared temperature term introduces a non-linear temperature response, while EVI introduces realistic seasonality and spatial patterns. The W_{scale2} parameter and its interactions with temperature help to account for water stress, especially at high temperatures when soils tend to dry out. Literature supports the use of these additional factors to help explain R_e fluxes, given that autotrophic respiration has large seasonal increases associated with canopy development (Jassal et al., 2007), and current photosynthetic uptake is known to account for a large portion of above and below-ground respiration during the growing season (Amthor, 2000; Högberg et al., 2001). Furthermore, soil moisture limits both autotrophic and heterotrophic respiration during drought periods (Flexas et al., 2006; Meir et al., 2008; Molchanov, 2009), and respiration response to temperature is typically described with a Q10, non-linear relationship (e.g., N. Meyer et al., 2018). These additional factors have also previously been suggested as needed improvements to the VPRM R_e equation in X. Li et al. (2020).

2.1.2. VPRM Parameter Optimization and PFT Classification

VPRM parameters are optimized using hourly NEE observations from flux towers in and near the domain, with parameters optimized separately for each PFT using only towers associated with each land cover classification. Flux data is included from 69 towers operational at any time from 2001 to the present (Figure 1; Table S1 in Supporting Information S1), which allows for the inclusion of many more locations than relying only on towers currently collecting data. In fact, 46 of the 69 flux towers used in this study were not included in either Mahadevan et al. (2008) or Hilton et al. (2013, 2014), although some areas of the domain (e.g., urban/suburban lawns, parks and gardens, and Appalachian deciduous forests and northern mixed forests in Canada) remain under-represented, potentially leading to spatial biases in optimized parameters. (Additional uncertainty is associated with the PFT classification and the extent to which variation across the landscape can be explained by PFTs in general, as discussed in more detail below.) In addition, u-star filtering of flux tower data to remove observations with insufficient turbulence results in disproportionate filtering out of nighttime relative to daytime data. To address this concern, we perform bootstrapping on data from under-sampled hours to create a more even distribution across the diurnal cycle (Chernick, 2007). More details on the flux tower data processing are included in the supplemental material.

All parameters with the original GEE and R_e models (i.e., λ, PAR_0, β, and α in Equations 1 and 4) are optimized simultaneously minimizing least squares between modeled fluxes and flux tower observations (using the DEoptim package in R). It should be noted that this optimization procedure ensures zero bias on an average basis across all hours and months in the observational data set but does not guarantee that optimized parameters will capture the peaks of either the diurnal or seasonal cycles. Therefore, we first optimize a set of annual (i.e., time-invariant) parameters, and then four sets of seasonally varying parameters (for December to February, March to May, June to August and September to November; Table S2 in Supporting Information S1), which should help to better account for flux seasonality.

For the modified R_e model (Equation 5), all R_e parameters (α₁, α₂, γ, θ₁, θ₂, and θ₃) are estimated as coefficients from regressions using nighttime NEE observations, while T_{crit} and T_{mult} are optimized by maximizing R^2 values in multiple regressions across a range of realistic values. "Observed" daytime hourly GEE is then derived as observed daytime NEE minus predicted daytime respiration, and GEE parameters are separately optimized by

again minimizing least squares between model and observations (using the DEoptim package in R, Table S3 in Supporting Information S1). The optimization of GEE and R_e parameters separately here should help to better capture diurnal cycle amplitude by ensuring zero bias for daytime and nighttime fluxes respectively, compared to optimizing all parameters simultaneously, which only ensures zero bias on a 24-h basis. Only one set of time-invariant parameters are optimized with the new R_e model, given that EVI and LSWI already help to account for seasonal changes.

PFTs are used here to classify flux towers and then spatially scale optimized parameters across the landscape, as in Mahadevan et al. (2008). PFTs are known to be imperfect, if useful, ecological classification tools (e.g., Atkin et al., 2015; Bonan et al., 2002; Wullschleger et al., 2014; J. Xiao et al., 2011). Therefore, the original PFT classification from the AmeriFlux database was re-examined for this study to identify how tower-specific optimized parameters cluster across and between PFTs using the original VPRM model (Figure S1 in Supporting Information S1, with similar results for VPRM$_{new}$). Based on an initial analysis, evergreen needleleaf and mixed forests were merged and then separated between regions north and south of 40°N, where fast-growing young pine plantations in the south behave differently from older-growth forests in boreal areas. This is consistent with previous work (Hilton et al., 2013; J. Xiao et al., 2011), showing that stand age and disturbance history may be as equally important as climate and PFTs for understanding NEE variations at large regional scales. Also, as in previous work, corn was separated from other crops given differences in C4 versus C3 photosynthesis. While there remains some overlap between site-specific optimized parameters across PFTs, the PFT classification used in this study still has ecological relevance, for example, crops clearly behave differently from other unmanaged ecosystems in terms of GEE parameters, with clear differences between corn and other crops in terms of PAR$_0$ and optimal temperature for photosynthesis. One topic for future research may be an improved classification scheme for grasslands, which have a relatively larger spread in optimized parameters across sites compared to other PFTs, likely due to a mix of C3 and C4 grasses as well as differences in management and land use (e.g., mowing and grazing) across sites.

An urban PFT was also defined for purposes of modifying respiration fluxes in areas with large impervious surface coverage. However, given that there are no specifically "urban" flux towers in the database, parameters for the urban PFT are assumed to be the same as for deciduous broadleaf forests (as in Hardiman et al., 2017), presumably the native vegetation of most cities in this domain. As in Hardiman et al. (2017), a correction was made in the urban PFT to reduce heterotrophic respiration (assumed as $0.5 * R_e$) by the fraction of impervious surfaces, although the autotrophic respiration correction was not applied given the difficulty in identifying reference pixels outside of every city in the domain. However, for the GEE and modified respiration model (Equation 5), fluxes are also suppressed in urban areas due to lower EVI values.

The urban PFT is defined from the underlying land cover maps to include developed land with low, medium and high intensity in the National Land Cover Database (NLCD; Jin et al., 2019; Yang et al., 2018) in the USA, and "urban/developed" in the Canadian land cover product. The "developed-open" category in the NLCD (i.e., cemeteries, gardens, lawns, and parks) is instead classified as grasslands, with the assumption (in the absence of other information) that these areas behave like grassland sites in the flux tower database. This hypothesis deserves to be tested further in future work, particularly given the large variation in optimized grassland parameters across sites discussed above (Figure S1 in Supporting Information S1).

2.1.3. VPRM Implementation and Evaluation

Using the model equations in Section 2.1.1 and optimized parameters, VPRM is run hourly at two different spatial scales for model evaluation: (a) site-specific runs at flux towers from 2001 to the present (for those years containing observational data for comparison) using tower-specific meteorological and high-resolution MODIS data and (b) gridded runs across the full domain at a 0.02° spatial resolution for a single year from November 1, 2016 to October 31, 2017. Each site-specific run is defined for the flux tower PFT, whereas for the gridded runs, fluxes are derived as weighted averages across PFTs using gridded fractional land cover maps to define the weights (Figure 1). Input data sets for the gridded runs include hourly air temperature and radiation from the High Resolution Rapid Refresh model (HRRR; Benjamin et al., 2016), and EVI and LSWI from the MODIS Terra and Aqua satellites, which are interpolated between satellite overpass dates in overlapping 16-day and 8-day composite periods. VPRM inputs are described in more detail in the Supporting Information.

Table 1
Comparison of Features Across Multiple VPRM Setups Evaluated in This Study

	VPRM$_{ann.orig}$	VPRM$_{ann}$	VPRM$_{ann.ND}$	VPRM$_{seas}$	VPRM$_{new}$
Respiration model	Original model: linear function of temperature	Original model: linear function of temperature	Original model: linear function of temperature	Original model: linear function of temperature	Expanded model including EVI, non-linear temperature, and interactions with water stress
Parameter seasonality	No	No	No	Yes (winter, spring, summer, fall)	No
Optimization technique	GPP and R_e parameters optimized simultaneously	GPP and R_e parameters optimized simultaneously	R_e parameters optimized with night-time NEE data; GPP parameters optimized with daytime NEE - predicted R_e	GPP and R_e parameters optimized simultaneously	R_e parameters optimized with night-time NEE data; GPP parameters optimized with daytime NEE - predicted R_e
Flux tower data set gap-filled and resampled?	No	Yes	Yes	Yes	Yes

Note. Model setups shaded in light gray were implemented for the gridded runs.

Site-specific VPRM is run for five different cases (described in Table 1), including four cases using the original R_e model (VPRM$_{ann.orig}$, VPRM$_{ann}$, VPRM$_{ann.ND}$, and VPRM$_{seas}$) and one with the modified R_e model (VPRM$_{new}$). With the original R_e model, we test the impact of bootstrapping nighttime data in the flux tower database using the original (VPRM$_{ann.orig}$) versus resampled database (VPRM$_{ann}$), annual (VPRM$_{ann}$) versus seasonal parameters (VPRM$_{seas}$), and optimizing GEE and R_e parameters simultaneously (VPRM$_{ann}$) versus separately (VPRM$_{ann.ND}$). We additionally isolate the impact of the new respiration model by comparing the old (VPRM$_{ann.ND}$) and new (VPRM$_{new}$) respiration equations using a consistent night/day optimization method. For model evaluation, a fraction of observations in the historical flux tower database is reserved from the parameter optimization. Given the relative sparsity of tower locations, we did not reserve entire sites to be used for model evaluation, but instead randomly sampled data across existing sites in the database (as explained in more detail in the Supporting Information). We acknowledge that this practice will likely over-estimate model performance in simulating spatial variability (Joiner & Yoshida, 2020; Reitz et al., 2021), but it should still be sufficient to distinguish between the quality of different model setups.

For simplicity, only VPRM$_{ann}$, VPRM$_{seas}$, and VPRM$_{new}$ were chosen for further analysis in the gridded runs, as these setups are the most realistic and consistent with previous VPRM publications (e.g., Hilton et al., 2013, 2014; Mahadevan et al., 2008), while demonstrating advances made in this study. Because there is no direct set of observations for validating gridded flux estimates across a landscape, here we rely on atmospheric CO_2 observations and TBM intercomparison to elucidate model performance at regional scales. The availability of simulations from two transport models (WRF-STILT and NAMS-STILT, as further described in Section 2.3) for the atmospheric CO_2 comparisons limited the study to this single year; however, weather patterns in 2016–2017 were within the range of 20 years of interannual variability in most parts of the domain (Figure S2 in Supporting Information S1).

2.2. Other Biospheric Models

We also compare gridded VPRM flux estimates with model output from CASA and SiB4, two process-based TBMs which are commonly used to understand the carbon cycle and as priors in atmospheric inversion studies over the North American continent (e.g., CarbonTracker from NOAA-ESRL, https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/). The CASA and SiB4 flux estimates used here for comparison with VPRM were not run specifically for this study; therefore, driver data and model resolutions were not standardized. However, the spread across models of plausible flux estimates should still provide a measure of VPRM model uncertainty. The principal components and differences between the VPRM, CASA, and SiB4 implementations compared here are described briefly below and also summarized in Table S4 in Supporting Information S1.

The CASA model was first developed in the 1990s to take advantage of remote-sensing data from NASA satellites (Potter et al., 1993; Randerson et al., 1996) and like VPRM, it is a diagnostic light-use efficiency model, incorporating remotely sensed data, meteorological inputs, and light-use efficiency factors to estimate GEE. It uses a diagnostic phenology scheme based on remotely sensed monthly fPAR (fraction of photosynthetically active radiation, Figure S3 in Supporting Information S1). The CASA implementation included here is the ensemble mean from the Level 2 pruned ensemble generated for the ACT-America project (Zhou, Williams, Lauvaux, Davis, et al., 2020; Zhou, Williams, Lauvaux, Feng, et al., 2020), with ensemble members varying across PFT-specific parameter ranges and maximum light-use efficiency parameters calibrated with North American flux tower data. Zhou, Williams, Lauvaux, Davis, et al. (2020) previously demonstrated that this CASA ensemble mean agrees well with flux tower observations compared to other TBMs.

The Simple Biosphere Model (SiB), despite its name, is a complex process-based model that simulates both the carbon and energy cycles and can be used to predict future carbon cycle dynamics (Sellers et al., 1986, 1996). GEE is estimated in SiB using the mechanistic Farquhar et al. (1980) enzyme-kinetic photosynthesis algorithm. This study uses output from SiB4 (Haynes, Baker, Denning, Stöckli, et al., 2019; Haynes, Baker, Denning, Wolf, et al., 2019), which has a prognostic (predictive) phenology scheme specifically developed to account for grassland and crop calendars. The SiB4 phenology scheme has been shown to perform well for distinct crop types (corn, soybeans, and winter wheat; Lokupitiya et al., 2009) and grasslands (Haynes, Baker, Denning, Stöckli, et al., 2019; Haynes, Baker, Denning, Wolf, et al., 2019) at discrete flux tower locations; however, it may fail to represent actual phenology across a landscape due to unmodeled management effects (e.g., fertilization, irrigation, equipment availability for planting and harvest), and disturbance events like storms or insect outbreaks.

CASA and SiB4 both have process-based respiration models, which track flows across carbon pools (albeit with a varying number of live and dead pools, Table S4 in Supporting Information S1). Respiration fluxes are then similarly determined from each pool as a function of photosynthetic supply, pool-specific turnover and decay-rate constants, and environmental stress factors.

CASA and SiB4 have varying spatial and temporal resolutions for model structure as well as in their specific implementations here. First, CASA has a native monthly temporal resolution, whereas SiB4 outputs hourly fluxes (like VPRM), with a sub-hourly calculation of photosynthesis, and daily updates to an internally calculated Leaf Area Index. An algorithm was introduced by Fisher et al. (2016) to downscale monthly CASA fluxes to 3-hourly using temperature and radiation data (further interpolated to hourly for this study); however even with this downscaling, phenology is still tracked monthly in CASA.

In terms of spatial resolution, the CASA implementation used here was run at 500 m for the coterminous USA and at 5 km for North America (with 500 m fluxes in the USA merged with 5 km fluxes in Canada for this study). Only the dominant land cover in each pixel is represented in CASA. In contrast, SiB4 fluxes are estimated here at the relatively coarse spatial resolution of 0.5° but using a weighted land-cover approach (like VPRM) to represent sub-pixel variability. This weighting approach has been shown to make land surface model performance less sensitive to spatial resolution (D. Li et al., 2013). The dominant land cover approach in CASA is partially compensated for by its high spatial resolution in the USA but could be potentially problematic in representing fragmented land covers in Canada at the 5 km resolution.

Gridded flux estimates from all three TBMs (VPRM, CASA, and SiB4) are aggregated/disaggregated to 0.1° resolution for further analysis, a computationally tractable scale that also allows for comparison of spatial patterns. However, to test the relative impact of varying spatial resolution across the TBMs, a sensitivity test is performed aggregating VPRM and CASA flux estimates to 0.5° from May to October (to match the resolution of SiB4) and simulating atmospheric CO_2 with the original (0.1°) versus coarsened (0.5°) fluxes.

2.3. Atmospheric CO_2 Evaluation of Gridded Fluxes

Gridded flux estimates are evaluated by convolving them with footprints from an atmospheric transport model and then comparing simulated with observed atmospheric CO_2 mole fractions sampled continuously at multiple towers throughout the domain. The footprints (or sensitivity to fluxes in space and time) of atmospheric observations vary by tower location, inlet height, and weather patterns, although the towers generally see the influence of high-resolution surface fluxes (in space and time) near the tower, and a more diffuse integrated signal coming from farther away. Thus, atmospheric CO_2 data can help to evaluate biospheric CO_2 flux estimates at coarser
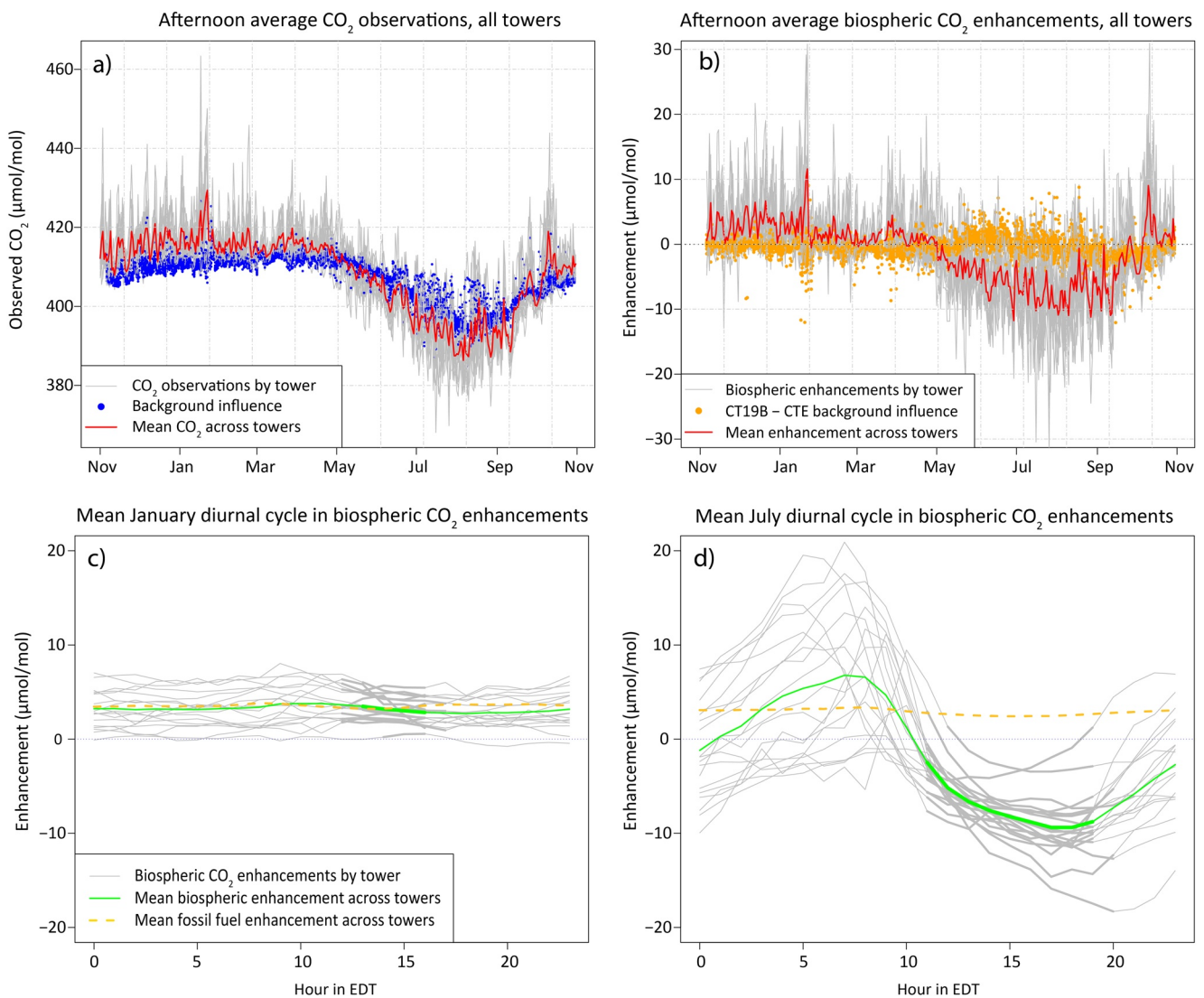
Figure 3. Atmospheric CO_2 observations across towers compared in four different ways: (a) total afternoon average CO_2 mole fractions across full year for each tower (gray), mean across towers (red) and modeled "optimal" background contribution (blue), (b) afternoon average biospheric enhancements across full year for each tower (gray), mean across towers (red) and difference in CT19 B versus CTE background conditions (orange), (c) mean hourly diurnal cycle of biospheric enhancements in January for each tower (gray), mean biospheric enhancement across towers (green) and mean FF contribution across towers (gold), and (d) the same as (c) but in July. For (c) and (d), afternoon hours have a thicker line width, with "afternoon" defined as described in the supplemental material. Convolutions and background conditions are as described in the text for all sub-plots unless otherwise indicated.

spatial and temporal scales over more parts of the domain compared to the limited spatial footprints (∼1 km^2) sampled by the flux towers.

2.3.1. Atmospheric CO_2 Tower Network

Afternoon hourly average atmospheric CO_2 observations from 21 surface towers are used to evaluate gridded NEE fluxes (Figures 2 and 3; Table S5; Section S2 in Supporting Information S1). The towers, maintained by a variety of data providers (Karion et al., 2020; Miles et al., 2018; Mitchell et al., 2019; NOAA ESRL, 2019; Richardson et al., 2017), were chosen among potential towers as those primarily sited in non-urban areas, excluding locations with challenging topography (for modeling transport) or large fossil fuel emission influence. Although many of these towers have been previously used in atmospheric inversion studies (especially those denoted as NOAA and EC in Table S5 in Supporting Information S1, e.g., L. Hu et al., 2019), roughly half of them (operated by Earth Networks, denoted as EN-NIST in Table S5 in Supporting Information S1) have not been used before in

a comprehensive analysis of upwind CO_2 fluxes. Thus, the combination of CO_2 mole fraction data from new and existing towers provides a unique opportunity to evaluate regional-scale CO_2 fluxes in this domain.

To simulate atmospheric CO_2 enhancements due to the biosphere, modeled NEE fluxes are multiplied by footprints from an atmospheric transport model (Lin et al., 2003), also known as a convolution where flux units of $\mu mol * m^{-2} * s^{-1}$ are converted to $\frac{\mu mol}{mol}$, or the mole fraction enhancement of CO_2 at the observation location and time occurring due to upwind modeled fluxes inside the domain. However, total atmospheric CO_2 measurements are also influenced by fossil fuel and fire emissions and atmospheric CO_2 in air masses advecting into the domain (i.e., background conditions), all of which must be subtracted from observed CO_2 mole fractions for comparison with simulated biospheric enhancements. Thus, convolved NEE fluxes (or simulated biospheric CO_2 enhancements) are compared with "observed" enhancements, defined as total observed CO_2 mole fractions—simulated fossil fuel influence—background conditions, with enhancements having positive or negative signs due to the influence of biospheric sources and sinks. (Fire emissions were also considered, but ultimately not included due to their marginal influence on CO_2 mole fractions in this domain and time period.) The term "observed enhancements" is henceforth used throughout the paper (in comparison to the simulated enhancements from NEE convolutions), although these pseudo-observations are also subject to errors in the transport model, fossil fuel inventories, and background conditions, as discussed below.

Both observed total CO_2 mole fractions (Figure 3a) and biospheric enhancements (Figure 3b) show a strong seasonal cycle and spatial variability across towers, reflecting the influence of seasonality and spatial patterns in the underlying CO_2 fluxes, but also atmospheric mixing. With changes in wind direction and synoptic weather conditions, what towers "see" in the domain at any given moment may be sparse and variable, but mean integrated footprints summed across towers (shown for July in Figure 2) show that most of the domain is observed on average over time by this network of towers, with slightly less sensitivity along the edges of the domain, and stronger sensitivities in summer compared to winter (when wind speeds are higher). Some towers have a stronger influence from croplands, wetlands, forests, or urban areas, although almost all towers have some sensitivity to deciduous broadleaf forests, crops, and grassland/pasture (including developed-open space), showing the broad influence of these land covers throughout the domain (Figure 1 and Table S5 in Supporting Information S1). The tower network is also sensitive to both day and nighttime fluxes on average, with afternoon observations typically most sensitive to nearby fluxes in the morning of the same day and previous night (Figure S4 in Supporting Information S1).

2.3.2. Atmospheric Transport Models

Footprints corresponding to each hourly atmospheric observation were generated from two different transport models: WRF-STILT and NAMS-STILT. WRF (Weather Research and Forecasting model; Skamarock et al., 2008)) was run with a 1-, 3-, and 9-km nest, with the finer scale nests centered around the Washington DC/Baltimore area (Figure 2). NAMS (North American Mesoscale System; NCEI et al., 2020) is a meteorological product with a spatial resolution of 12 km made publicly available by NOAA/ARL ((ftp://arlftp.arlhq.noaa.gov/nams). The Stochastic Time-Inverted Lagrangian Transport model (STILT, Lin et al., 2003) was used in combination with winds from the two meteorological products (WRF and NAMS) to generate footprints on a 0.1° grid across the domain by releasing virtual particles from each observation point and tracing them back in time and space to their source locations and times. Further details of the custom WRF runs and footprint generation is included in Section 3 of the supplemental material.

WRF-STILT with its custom setup and higher spatial resolution may help to better model transport in the Appalachian mountain range that crosses most of the domain (Pillai et al., 2011) and within the two inner nests from Washington D.C. to Philadelphia, as compared to NAMS-STILT. However, without a more in-depth study evaluating the two transport models against observations, it is difficult to know which set of footprints is more accurate in different parts of the domain throughout the year. Therefore, convolutions using WRF-STILT and NAMS-STILT footprints are averaged for this study, as averaging across process-based models is known to help reduce the influence of systematic and random errors (Elder, 2018). Results from atmospheric CO_2 comparisons are also highlighted only if they are robust across each individual transport model (with corresponding analyses using WRF-STILT or NAMS-STILT footprints alone included in the Supporting Information).

2.3.3. Background Conditions

To determine background conditions (Karion et al., 2021; Mueller et al., 2018), STILT virtual particles corresponding to each atmospheric CO_2 observation are traced backwards to their exit locations from the domain, and then a four-dimensional CO_2 mole fraction field is sampled and averaged at exit locations and time periods to determine the background contribution. Modeled four-dimensional CO_2 mole fractions at exit points are taken from two optimized data assimilation products: CarbonTracker CT2019 B (CT19 B; Jacobson et al., 2020) and CarbonTracker Europe 2018 (CTE18; Peters et al., 2010). Background conditions at the towers differ between the two products (CT19 B and CTE18) throughout the year, with mean monthly differences ranging from 0.5 µmol/mol (or parts per million, ppm) in November and December to 1.4 µmol/mol in August and 1.8 µmol/mol in September (Figure 3b), in part due to differences in the underlying biospheric models used as priors in each optimization system. In fact, the difference between background conditions is larger than the "observed" biospheric enhancements for about 26% of observations in winter months (November to April) and 12% of observations from May to October, with even higher percentages in early spring (e.g., 47% in March and 32% in April) as ecosystems transition from CO_2 sources to sinks.

Given that there are limited datasets to evaluate which background product is most accurate at different times of the year, we used the following procedure to choose an "optimal" product in each month. Monthly mean biases across towers are compared between simulated and observed CO_2 enhancements using each potential background condition (CT19 B, CTE18, and their mean) and for each TBM. If one background product has consistently smaller biases across towers for all the TBMs, then we assume that this product is the "better" model in that month for our purposes. This analysis resulted in selecting the mean background in November, February, March, April, August, and September, and the CTE18 product in the remaining months.

2.3.4. Fossil Fuel Emissions

To remove the influence of fossil fuel emissions on atmospheric CO_2 observations, we use fossil fuel flux estimates from the Vulcan 3.0 product (Gurney et al., 2020) in the United States merged with the FFDAS product (Asefi-Najafabady et al., 2014) in Canada. Both products are defined hourly, with FFDAS at 0.1° and Vulcan 3.0 at 1 km spatial resolution but merged to have a consistent 0.1° resolution. Unfortunately, neither product was available specifically for our year of interest (2016/ 2017); therefore, the merged fossil fuel product for 2015 is adjusted to match the days of week in 2016/ 2017, given that fossil fuel emissions are known to behave differently on weekdays versus weekends (Gurney et al., 2020).

Uncertainty associated with fossil fuel emission estimates is generally considered to be lower than that from biospheric flux estimates (Lauvaux et al., 2021), especially during the growing season; however, emission magnitudes and their fine spatiotemporal patterns are not perfectly known, especially if the emission product was developed for a year other than the one of interest, as in this study. In fact, in winter, fossil fuel and biospheric enhancements for towers in the domain are similar in magnitude and have the same positive sign (Figure 3c); thus, small errors in emissions estimates could have a larger relative impact on atmospheric CO_2 comparisons in winter compared to summer months (Figure 3d). Regardless, we consider the fossil fuel emission product used here sufficient for the purposes of evaluating biospheric model performance, which is supported by minimal differences between convolutions using different emission products (e.g., FFDAS in the USA rather than Vulcan 3.0) compared to the spread associated with varying transport, biospheric model and background conditions (as also seen in other studies over eastern North America, e.g., Martin et al., 2019).

2.4. Model-Data and Model-Model Intercomparisons

Figure 4 shows the flow of analyses used to evaluate VPRM flux estimates, including analyses with flux tower observations (green), intercomparison with CASA and SiB4 flux estimates (orange), and atmospheric CO_2 observations (blue). Site-specific flux estimates from the five VPRM cases are compared to reserved nighttime NEE observations at flux tower locations (to evaluate the respiration component) and then 24-h NEE (to evaluate full seasonal and diurnal cycles). Then, gridded flux estimates from the three selected VPRM cases (VPRM$_{ann}$, VPRM$_{seas}$, and VPRM$_{new}$), CASA and SiB4 are compared in terms of their gridded spatial patterns, as well as seasonal and diurnal cycles aggregated to two spatial groupings: deciduous broadleaf forest and cropland pixels (Figure 1, right panel). Together, these two land cover types make up about half of the land area in the domain and are the predominant land covers upwind of the NEC, thus disproportionately influencing atmospheric CO_2
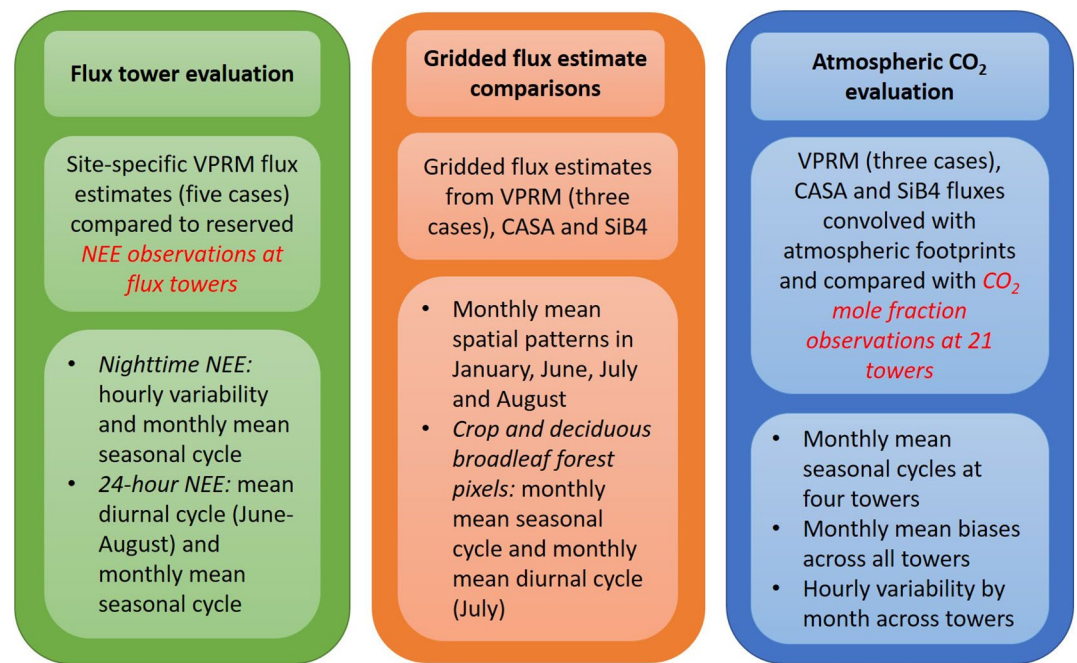
Figure 4. Diagram of planned analyses for VPRM flux evaluation. Each color box delineates a particular type of evaluation (in terms of data sets and scale of analysis), with the second row of sub-boxes describing the model runs and observations, and the third row describing each set of analyses.

in urbanized areas of the East Coast. Moreover, the Corn Belt is known to be one of the most biologically productive ecosystems on Earth during the height of the growing season (Gray et al., 2014; Hilton et al., 2017; Zeng et al., 2014), and therefore has a strong influence on atmospheric CO_2 across the domain, especially in June, July, and August.

Finally, NEE convolutions from the three VPRM's, CASA, and SiB4 are compared to observed atmospheric CO_2 enhancements (as described in Section 2.3) to test the TBMs' ability to capture CO_2 spatial gradients, seasonal cycles, and hourly variability in the atmosphere. The seasonal cycles of monthly mean atmospheric enhancements are first compared at four specific towers, with two predominantly influenced by croplands (S01 in Indiana and TPD in Ontario) and two by forests (DNH in New Hampshire and DVA in Virginia; Figure 2; Table S5 in Supporting Information S1). Then, monthly mean biases are examined across all 21 towers to identify general errors in the TBM flux estimates across the domain (in terms of phenology, flux magnitudes, and spatial patterns) throughout the year. Next, we compare hourly CO_2 variability within each month across towers, which tests the TBMs' ability to match the diurnal cycle, synoptic variability, and spatial gradients seen in the observations by time of year.

For the analyses using flux tower and atmospheric CO_2 data, simulations are compared to observations at both hourly and monthly timescales. At the monthly timescale, mean biases are assessed, whereas at the hourly timescale, the coefficient of determination (R^2) and the Nash-Sutcliffe coefficient (or NSC, Moriasi et al., 2007) are used to evaluate model performance, with the NSC, also known as a model efficiency factor, defined as:

$$NSC = 1 - \frac{\sum_{i=1}^{n}(pred_i - obs_i)^2}{\sum_{i=1}^{n}(obs_i - \overline{obs})^2}.$$

The NSC is calculated using the same equation as a coefficient of determination (R^2), but instead of using a fitted regression model (guaranteed to have zero mean bias), the NSC metric uses actual model simulations, or in this case, NEE convolutions with atmospheric footprints ($pred_i$). Therefore, the NSC can vary from -∞ to 1, with values < 0 indicating that the model performs worse than the observational mean for prediction (due to biases in the model), and values > 0 implying a better model fit than the observational mean. Thus, this metric assesses the
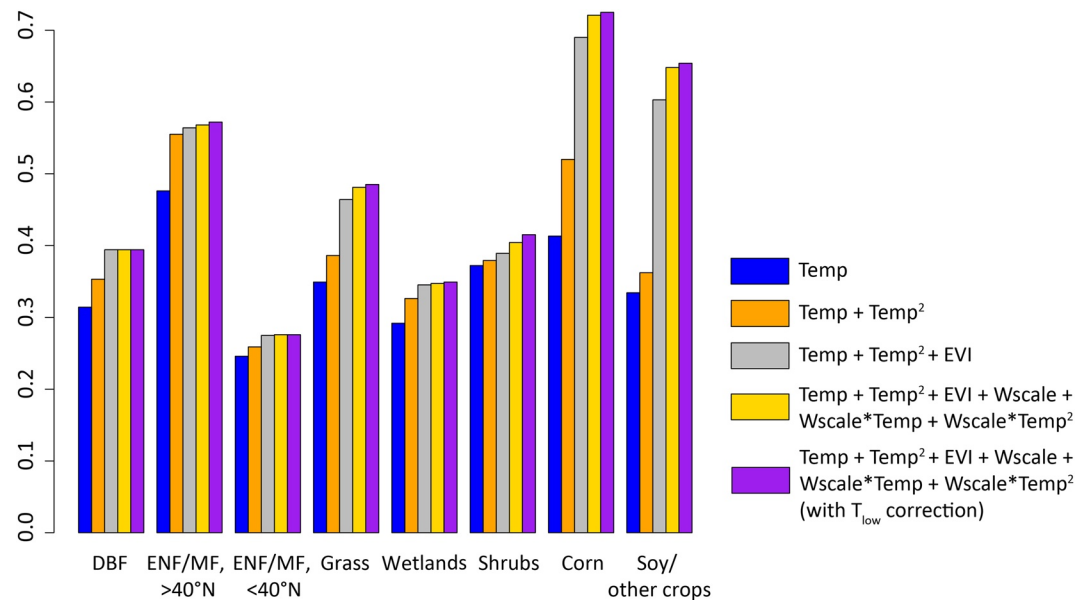
Figure 5. Adjusted R^2's from regressions predicting night-time daily mean NEE with site-specific meteorological and remote-sensing data for each PFT. Each bar (by PFT) includes additional predictor variables into the model, with the last bar adding the low air temperature correction into the full model with all variables (Equation 5).

model's ability to capture small-scale variability, while also penalizing model fits that have particularly biased flux estimates.

3. Results and Discussion

3.1. VPRM Model Development and Evaluation With Flux Tower Data

For the new VPRM respiration model (Equation 5), EVI was seen to be the single most important factor in improving model fit to hourly nighttime NEE observations, particularly for grass and croplands (difference between orange and gray in Figure 5). This is consistent with the fact that canopy development during the growing season in the temperate, humid climate that covers most of this domain induces large seasonal increases in autotrophic respiration. The second most important factor was the squared temperature term, particularly for northern evergreen & mixed forests and corn (difference between blue and orange). For crops and more water-limited ecosystems (i.e., grass and shrublands), the water stress scaling factor and its interactions with temperature also marginally help to improve model performance (difference between grey and yellow). Across PFTs, the new respiration model explains the most variability for crops (i.e., with adjusted R^2's of 0.73 for corn and 0.65 for other crops compared to 0.28–0.57 for other PFTs), perhaps due to sharper seasonal transitions in biomass (captured by EVI) or a lack of natural disturbances which have a stronger impact on other unmanaged ecosystems.

Overall, the new respiration model in VPRM$_{new}$ (green points in Figure 6, top row, Figure S5 in Supporting Information S1) substantially improves the fit of predicted R_e fluxes to nighttime NEE observations, especially in capturing the spread at high temperatures, compared to the fit using a simple linear model with the four other VPRM cases (VPRM$_{ann.orig}$, VPRM$_{ann}$, VPRM$_{seas}$, or VPRM$_{ann.ND}$). The relative increase in performance for VPRM$_{new}$ (green dots) is particularly dramatic for corn and soybean/other crops, with more of a linear relationship remaining for forested ecosystems. As expected, VPRM$_{seas}$ (purple lines in Figure 6, top row) has better performance than VPRM$_{ann}$ (red lines) across all PFTs (but not as good as VPRM$_{new}$), due to its ability to capture seasonal changes in baseline respiration. Interestingly with VPRM$_{seas}$, a flat or negative relationship in summer is inferred between air temperature and respiration for the grassland and soybean/other crop PFTs (Figure S5 and Table S2 in Supporting Information S1), pointing to soil moisture limitations at the height of the growing season for these ecosystems. VPRM$_{ann}$ also has generally better performance than VPRM$_{ann.orig}$, due to more even sampling across the diurnal cycle in the resampled flux tower observations (carried through to the other cases as well).
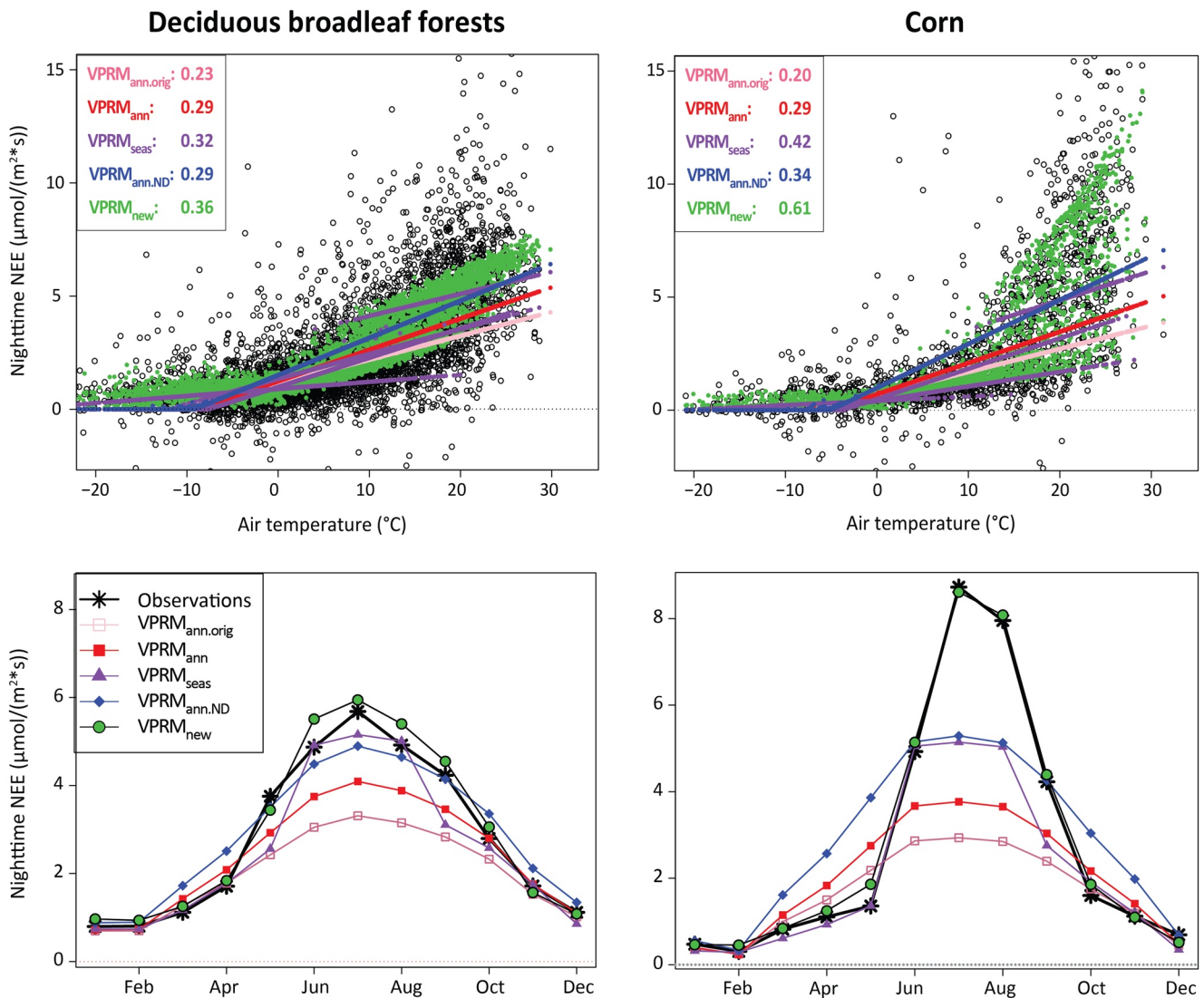
Figure 6. Scatter plots of observed air temperature versus night-time daily mean NEE (top row) for out-of-sample flux tower observations reserved from parameter optimization, and model fits for VPRM$_{ann.orig}$ VPRM$_{ann}$, VPRM$_{ann.ND}$, VPRM$_{seas}$, and VPRM$_{new}$, with four lines for VPRM$_{seas}$ corresponding to each season. NSC values are shown comparing each model to observations. Also shown are seasonal cycles of monthly mean night-time NEE (bottom row) comparing out-of-sample observations to model predictions from the same VPRM cases. Results are shown for two PFTs (representing ~40% of total land cover in the domain): deciduous broadleaf forests (27% + 5% urban, left column) and maize crops (8%, right column). Plots for other PFTs are included in Figure S6 in Supporting Information S1.

VPRM$_{new}$ also better captures the seasonal cycle in monthly mean nighttime NEE relative to the other VPRM cases (Figure 6, bottom row, Figure S6 in Supporting Information S1), as it is the only model to capture the magnitude of peak summertime respiration, especially in crop ecosystems. VPRM$_{seas}$ and VPRM$_{ann.ND}$ only slightly underestimate peak respiration in most PFTs but have other limitations. VPRM$_{seas}$ (purple) has sharp seasonal transitions every 3 months, given how the parameters are defined, which could potentially be alleviated by using monthly parameters in future work. The higher respiration year-round for VPRM$_{ann.ND}$ (blue) relative to VPRM$_{ann}$ (red) shows that separately estimating GEE and R_e parameters can improve the models' ability to match peak summertime R_e; however, VPRM$_{ann.ND}$ is shown to overestimate respiration in the spring and fall for all PFTs, and especially for crops. VPRM$_{ann}$ (red) and VPRM$_{ann.orig}$ (pink) have the lowest peak summertime respiration across cases, with underestimates for VPRM$_{ann}$ of ~20%–60% relative to observations across PFTs. The further ~15%–30% less peak respiration for VPRM$_{ann.orig}$ relative to VPRM$_{ann}$ again shows the benefit of the resampling procedure for capturing peak R_e fluxes.

Across the full diurnal cycle at the height of the growing season (shown on average from June to August in Figure 7, top sub-plot), the VPRM cases are shown to mainly differ at night, with VPRM$_{new}$ (green) the most unbiased from 6:30 p.m. to 4:30 a.m. and with a smaller spread across towers compared to the other cases. However, all cases have similar positive daytime biases from 8:30 a.m. to 2:30 p.m. local time of ~1–2 μmol*m^{-2}*s^{-1} on average across towers. It is unclear what could cause this similar bias across VPRM setups using different optimization procedures, but it could be related to morning venting of pooled nighttime NEE in the flux tower observations, which does not have a biophysical explanation in the model equations. This would also point to some amount of unobserved R_e during nighttime hours in the flux tower observations due to insufficient turbulence, which is instead observed later the next day. These diurnal cycle biases associated with flux tower observations remain to be investigated (and hopefully eliminated) in future work.

Given the positive bias in daytime NEE (and unbiased nighttime NEE), VPRM$_{new}$ has a small positive mean bias on a 24-h basis during the growing season (Figure 7, bottom sub-plot). VPRM$_{ann.orig}$ (red) and VPRM$_{ann}$ (pink) monthly mean NEE are biased low from June to September, due to their strong underestimation of nighttime R_e in the growing season (stronger in magnitude than the weak positive daytime bias), whereas VPRM$_{seas}$ (purple) NEE is biased high in June and low in September, likely due to a static baseline respiration value across 3-monthly periods. VPRM$_{ann.ND}$ (blue) is biased high in the shoulder months of March, April, May, October, and November, given that temperature alone is not a good predictor of the R_e seasonal cycle. All cases are relatively unbiased compared to flux tower NEE observations (with a smaller spread across towers) in the winter months of December, January, and February. The large spread across towers during the growing season for all cases could point to difficulties in predicting spatial variation, although this topic remains to be investigated further in future work (H. Meyer et al., 2019).

3.2. Gridded NEE Spatial Patterns Across TBMs

Gridded monthly mean NEE flux estimates in January, June, July, and August (at 0.1° resolution) are shown in Figure 8 for VPRM$_{new}$, CASA, and SiB4. Given different model structures and input data sets, CASA and SiB4 spatial patterns are relatively independent from those of VPRM$_{new}$, thereby showing the range of uncertainty here across plausible TBMs. In contrast, VPRM$_{seas}$ and VPRM$_{ann}$ monthly mean NEE spatial patterns are highly correlated with those of VPRM$_{new}$, with differences primarily in their flux magnitude; thus, their spatial patterns are compared in Figure S7 of the Supporting Information S1. All TBMs show net sources in winter and net sinks in summer months, with proportionally higher net uptake in cropping areas at the peak of the season. But spatial patterns also differ across TBMs due to differences in underlying land cover maps, relative strength of component fluxes (i.e., GEE and R_e) within each PFT, and phenology.

In January, when R_e dominates NEE across most parts of the domain (Figure 8, first column), VPRM$_{new}$ has weaker sources to the atmosphere than either CASA or SiB4. Also, VPRM$_{new}$ and CASA have stronger sources in the south of the domain compared to the north, whereas SiB4 shows the opposite north-to-south gradient, due to a relatively stronger GEE in the south. Wintertime fluxes for VPRM$_{new}$ and SiB4 are also relatively smooth across the landscape compared to CASA, which estimates zero fluxes in urban areas (more noticeable in the United States with 500 m resolution) and anomalously low fluxes in parts of Ontario (likely related to the 5 km spatial resolution and dominant land cover formulation in this model).

During the growing season (i.e., June to August, second to fourth columns in Figure 8), all TBMs estimate the strongest net uptake in the Corn Belt relative to other parts of the domain, although the magnitude and timing of peak uptake differs across models. For example, in states like Illinois and Indiana, SiB4 estimates by far the strongest peak NEE (i.e., maximum pixel-wise July monthly mean of −9.6 μmol*m^{-2}*s^{-1}), followed by VPRM$_{new}$ (−7 μmol*m^{-2}*s^{-1}) and then CASA (−6.2 μmol*m^{-2}*s^{-1}). CASA has the weakest net uptake in cropping areas, in part because it does not distinguish between C3 and C4 crops (Table S4 in Supporting Information S1), unlike SiB4 and VPRM$_{new}$. The strong differences in July NEE between VPRM and SiB4 in the Corn Belt could be due to the daytime NEE bias in VPRM discussed in Section 3.1, and/or differences in the underlying crop maps between the models and crop-specific parameterizations.

The TBMs also clearly differ in terms of crop phenology. For example, in June, SiB4 net uptake is already strong in the Corn Belt, whereas crop uptake is just starting for CASA or not planted yet for VPRM$_{new}$. In states like Illinois and Indiana, peak crop uptake occurs in August for VPRM$_{new}$, but earlier in July for CASA and SiB4. In
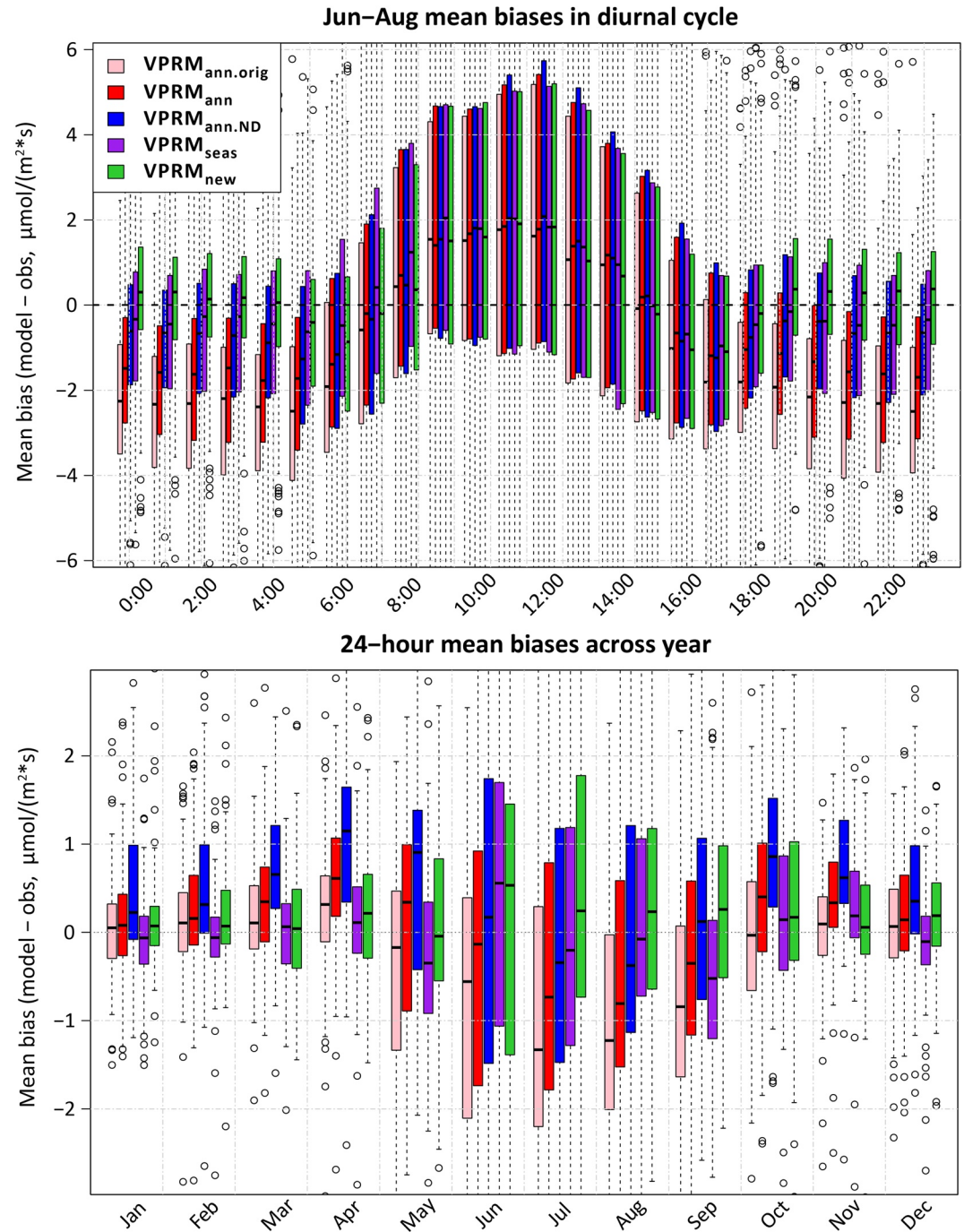
Figure 7. Seasonal mean biases (from June to August) compared to reserved flux tower observations for each hour in the diurnal cycle (in local time; top row) and 24-h monthly mean biases (bottom row), with biases defined as VPRM NEE—observed NEE. Results are shown for each of the VPRM cases described in Table 1. Boxplots show the spread across individual flux towers, with each box representing the interquartile range (IQR), the median indicated as a horizontal line inside each box, the error bars indicating the 25th percentile—the IQR (minimum) and the 75th percentile + the IQR (maximum), bounded by the range of the data, and dots as outliers falling outside of the error bars.

the Mississippi River Valley, which has substantial soybean production, SiB4 also shows an earlier start to the cropping season (in spring, results not shown) and an earlier end in July. In addition, SiB4 and CASA have more extensive cropping areas than in VPRM, with additional pixels in Michigan, Wisconsin, and the Carolinas (SiB4), and Canada (CASA; Figure S8 in Supporting Information S1).
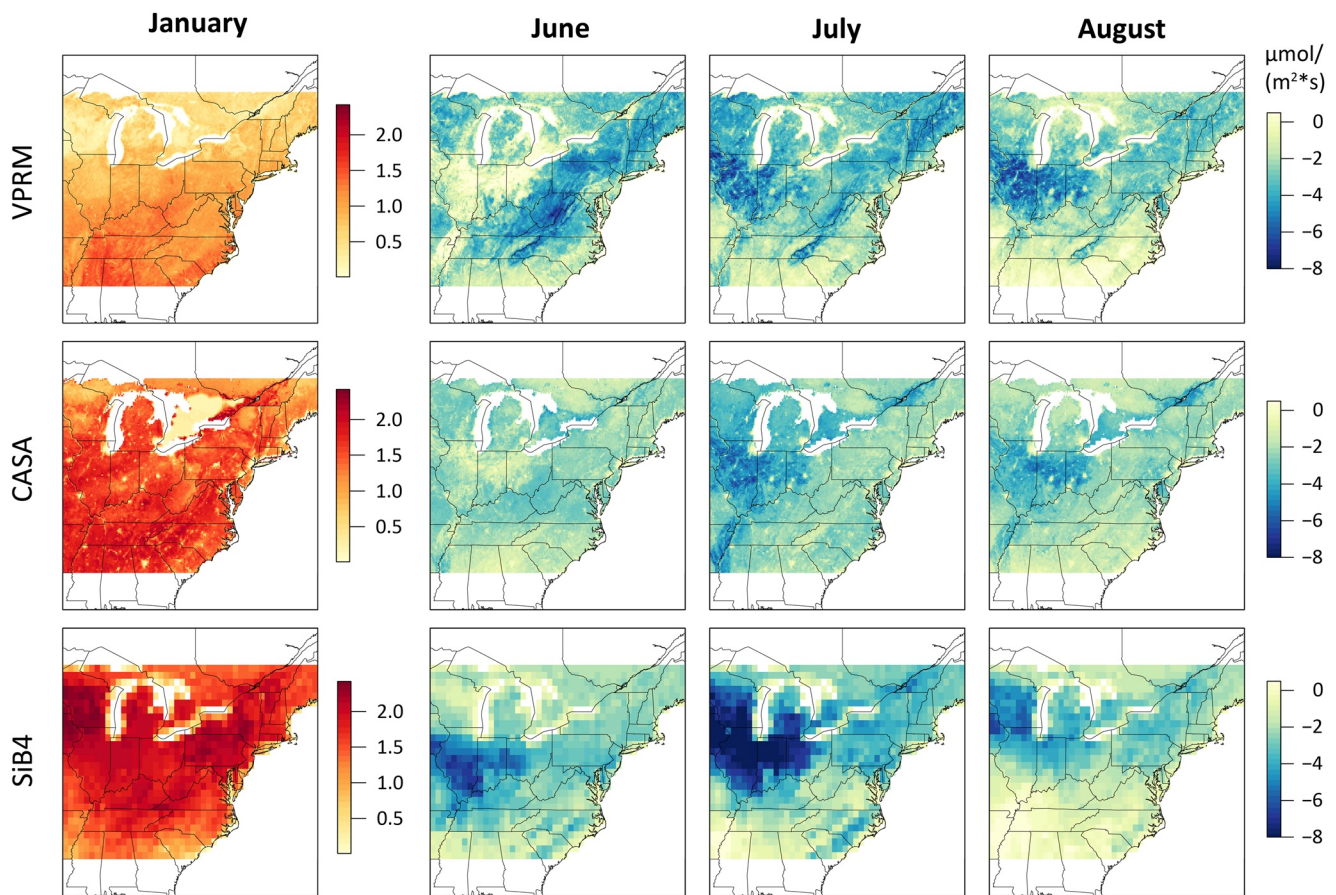
Figure 8. Monthly mean gridded NEE at 0.1° for VPRM$_{new}$, CASA, and SiB4 in January, June, July, and August. Corresponding plots comparing VPRM$_{new}$, VPRM$_{seas}$, and VPRM$_{ann}$ are shown in Figure S7 in Supporting Information S1.

For deciduous broadleaf forests during the growing season, VPRM$_{new}$ estimates stronger net uptake than CASA and SiB4, particularly in June, although CASA has a much larger spatial extent for this PFT than VPRM and SiB4 (Figure S8 in Supporting Information S1). For the VPRM parameter optimization, most of the deciduous broadleaf forest flux towers are in the north of the domain, with fewer in the Appalachian deciduous forests of Kentucky, West Virginia, and Pennsylvania, where most of this land cover exists (Figure 1), which could bias the resulting parameters and flux estimates. In July and August. VPRM$_{new}$ and SiB4 also estimate near-neutral fluxes in the southern part of the domain (i.e., northern halves of Alabama, Mississippi, and Georgia, containing primarily a mix of evergreen, mixed and deciduous broadleaf forests) compared to the net sinks in CASA. This could be because VPRM$_{new}$ and SiB4 better capture late-summer water stress (i.e., weaker GEE) in these areas at this time of year. Differences in phenology in forested PFTs are more apparent in spring and fall months (rather than June to August shown in Figure 8), as will be discussed further in Section 3.3.

3.3. NEE Seasonal and Diurnal Cycles Across TBMs

The TBMs generally agree well in terms of timing for the monthly mean NEE seasonal cycle aggregated across crop and deciduous broadleaf forest pixels (Figure 9), except for SiB4 in croplands, which has a growing season shifted about a month earlier than for the other TBMs. During the growing season, VPRM$_{new}$, VPRM$_{seas}$ and CASA agree well in terms of flux magnitude for both ecosystems, whereas SiB4 and VPRMann have stronger peak net uptake in June and July. It is interesting to note that VPRM$_{ann}$ has stronger net uptake in summer months than VPRM$_{seas}$ and VPRM$_{new}$, despite having the lowest magnitude of component GEE and R_e fluxes (Figure S9 in Supporting Information S1), which is due to having the most depressed R_e relative to GEE during the growing season.
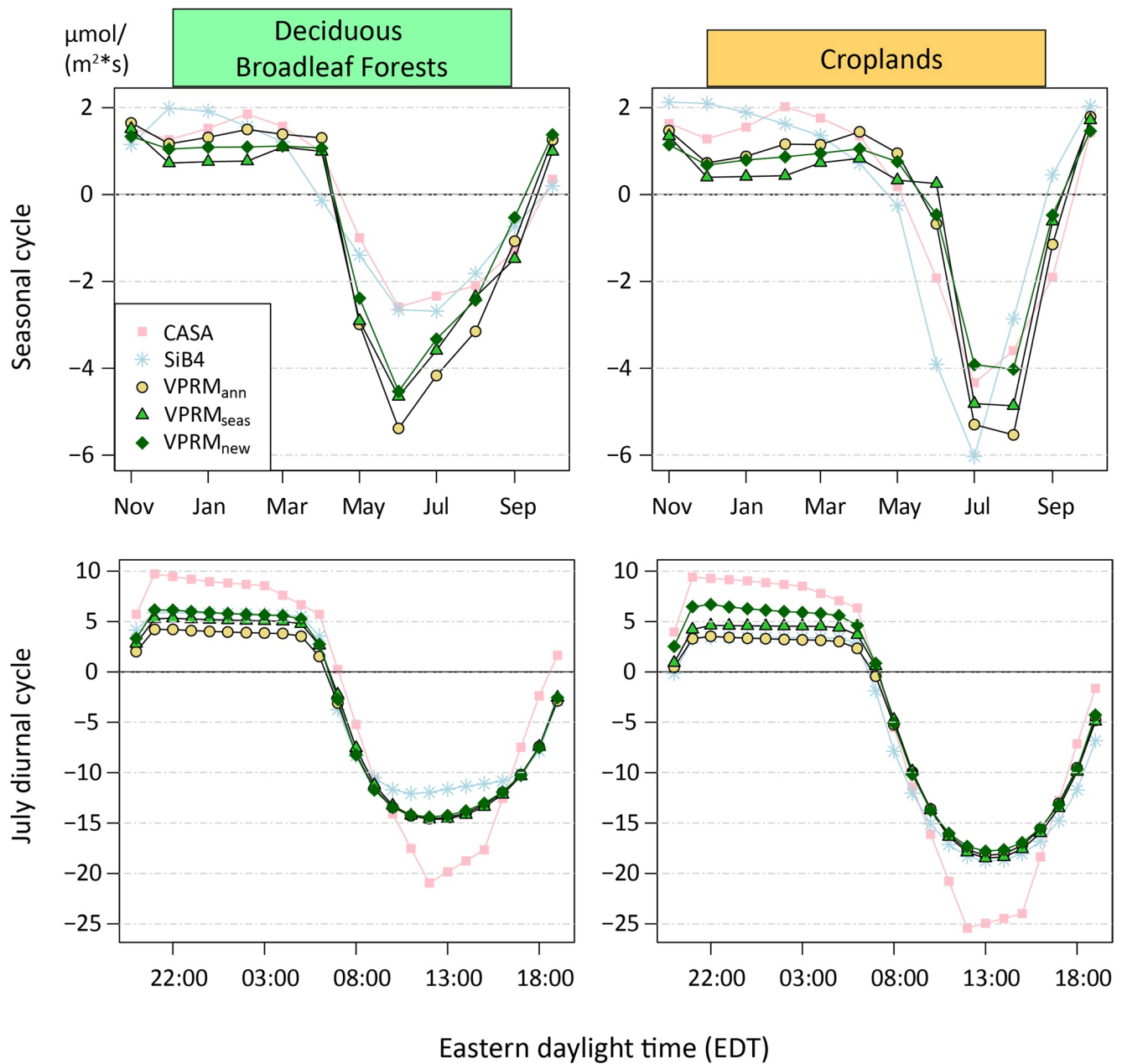
Figure 9. Seasonal cycle of monthly mean NEE fluxes (top row), and mean July diurnal cycle (bottom row), from the gridded runs for each TBM, spatially aggregated across pixels with predominantly deciduous broadleaf forests (left column) and croplands (right column), as indicated in Figure 1. The corresponding plots with GEE and R_e fluxes are shown in Figures S9 and S10 in Supporting Information S1.

It is interesting to note that VPRM$_{ann}$ has stronger net uptake in summer months for both ecosystems than VPRM$_{seas}$ and VPRM$_{new}$, despite having the lowest magnitude of component GEE and R_e fluxes (Figure S9 in Supporting Information S1); this is due to its having the most depressed R_e relative to GEE during the growing season. In general, GEE and R_e fluxes in VPRM$_{new}$ are stronger than in VPRM$_{seas}$, followed by VPRM$_{ann}$, in part due to separate optimization of nighttime and daytime parameters with VPRM$_{new}$.

CASA and SiB4 estimate much stronger component fluxes (i.e. GEE and R_e, Figure S9 in Supporting Information S1) than any of the VPRMs, although GEE and R_e fluxes in VPRM$_{new}$ are stronger than in VPRM$_{seas}$, followed by VPRM$_{ann}$, in part due to separate optimization of nighttime and daytime parameters with VPRM$_{new}$. In addition, CASA has longer growing seasons for GEE and R_e relative to both SiB4 and the VPRMs, especially for

R_e, which begins to ramp up about a month earlier in spring compared to the other TBMs. This was also seen in Messerschmidt et al. (2013), who showed that the timing in CASA was less consistent with atmospheric column CO_2 and flux tower observations than that in SiB. Given the difference in component flux timing, the similarity between CASA and the VPRM's in timing and flux magnitude of the NEE seasonal cycle during the growing season is likely due to some combination of both models using diagnostic phenology and calibration with flux tower NEE.

Stronger R_e fluxes year-round for CASA and SiB4 are also reflected in stronger wintertime NEE, when R_e dominates the total flux. CASA and SiB4 are neutral biosphere models that do not account for net annual sinks, which could lead to overestimation of R_e in forests recovering from disturbances (e.g., storms, insect attacks, or harvest) or in croplands where harvested products are transferred to other areas for consumption (Haynes, Baker, Denning, Stöckli, et al., 2019; Haynes, Baker, Denning, Wolf, et al., 2019; Zhou, Williams, Lauvaux, Davis, et al., 2020). (In contrast, net annual sources and sinks in VPRM primarily reflect the influence of the flux tower observations used in the parameter optimization, which in this case result in a domain-wide net annual sink about 4.5 times greater than for the other TBMs in this year.) VPRM may also underestimate R_e year-round due to insufficient nighttime turbulence and later morning venting of nighttime NEE in the flux tower observations, as mentioned previously (Barr et al., 2013). In addition, drainage loss due to horizontal advection during stratified stable nighttime conditions is non-negligible and represents flux to the atmosphere that is never measured (Aubinet, 2008; Nicolini et al., 2018), which could also lead to underestimated VPRM R_e fluxes (which are optimized with nighttime measurements).

For the mean July NEE diurnal cycle (Figure 9), the VPRM cases differ primarily at night (also seen in Figure 7), with similar afternoon net uptake (despite slightly stronger GEE for VPRM$_{new}$ relative to VPRM$_{seas}$ and VPRM$_{ann}$, Figure S10). Compared to the other TBMs, CASA has by far the strongest amplitude of the NEE diurnal cycle, with daytime and nighttime peaks about 50% stronger than for VPRM$_{new}$ in both ecosystems for this month. The SiB4 diurnal cycles are more similar to that of the VPRMs, with only slightly stronger net afternoon uptake and very similar nighttime sources as VPRM$_{seas}$. The longer growing seasons in CASA for GEE and R_e also lead to stronger diurnal cycles for CASA in shoulder seasons (spring and fall) compared to the other TBMs (results not shown).

3.4. Gridded NEE Model Evaluation Using Atmospheric CO_2 Observations

3.4.1. Monthly Mean CO_2 Seasonal Cycles at Four Towers

A comparison between monthly mean simulated and observed atmospheric CO_2 enhancements is shown for four towers in Figure 10. S01, influenced by corn production, shows the strongest observational peak drawdown in July and August compared to the other towers, whereas DVA, in the southern part of the domain, has the longest growing season from May through September. Peak uptake occurs in July for DNH and S01, whereas peak uptake occurs earlier in June at DVA (perhaps due to late-summer water stress) and in August at TPD (perhaps due to crops planted later at northern latitudes compared to S01). At DVA, all the TBMs are biased low in November, December, and April for both transport options (Figures S11 and S12 in Supporting Information S1) pointing to a systematic problem with fossil fuels or background conditions at this site in these months (although all TBM flux estimates could also be similarly biased). At other towers and times of year, observations generally fall within the spread of modeled enhancements from the different TBMs.

Simulated CO_2 enhancements from all the VPRMs are biased low at DNH during the growing season, whereas at the crop-influenced towers (TPD and S01), VPRM$_{new}$ and VPRM$_{seas}$ are biased high in terms of peak uptake. VPRM$_{ann}$ more closely matches peak uptake at the crop sites but overestimates net uptake at the forested towers (DNH and DVA). The better performance of VPRM$_{ann}$ at the crop sites is likely due to compensating errors (both underestimated R_e and GEE), rather than higher overall model skill relative to VPRM$_{new}$.

At the two cropland towers TPD and S01, SiB4 has the strongest peak drawdown in July across TBMs, with a negative bias relative to atmospheric observations (in particular, by ∼100% at TPD). At DNH, simulated SiB4 July uptake is also biased low relative to observations, showing that errors in cropland fluxes likely have an influence on CO_2 mole fractions far downwind. CASA simulated enhancements match observations well during the growing season at all towers except S01, where peak uptake is slightly weak in July and August. In winter
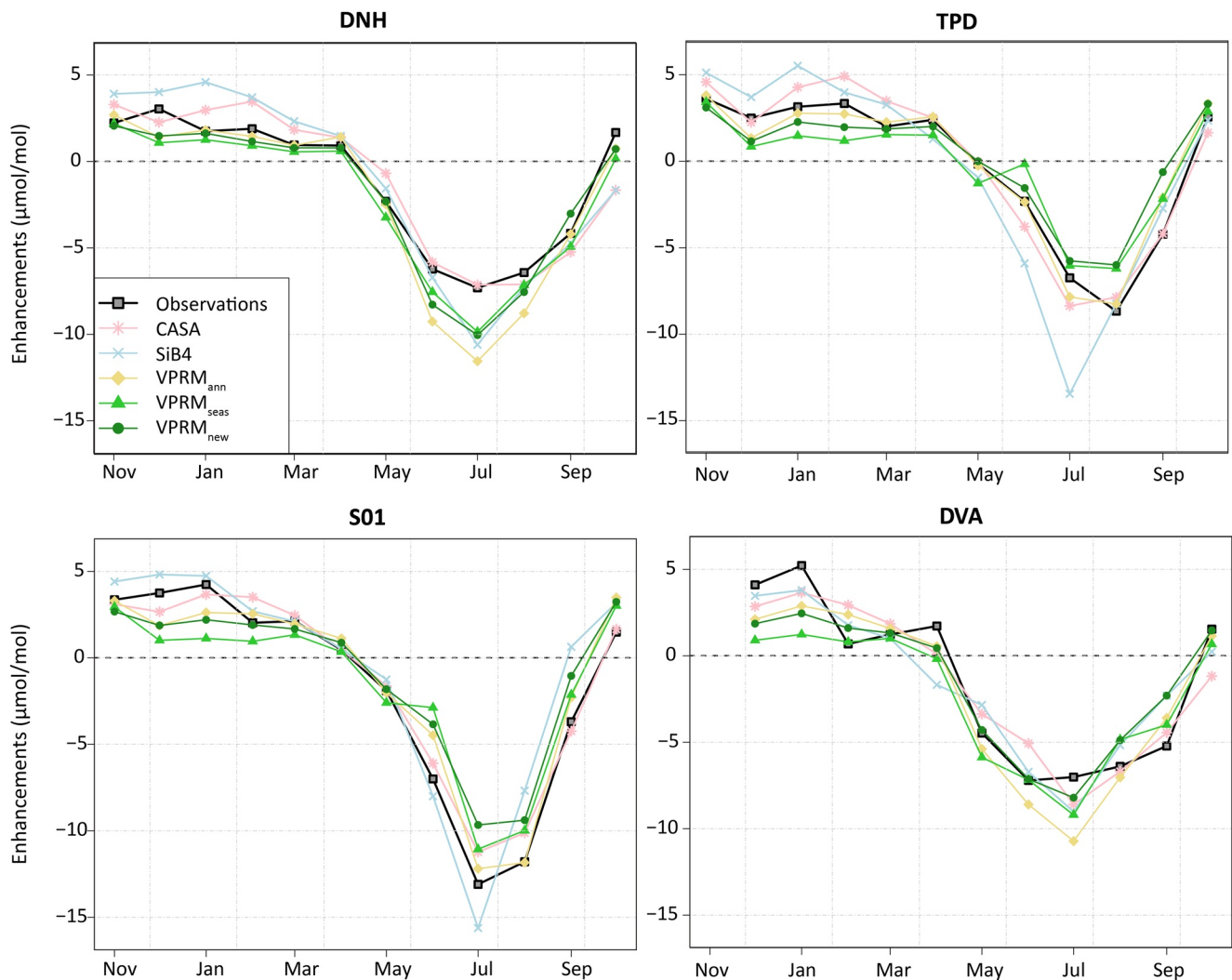
Figure 10. Monthly mean simulated and observed biological atmospheric CO_2 enhancements for VPRM$_{new}$, VPRM$_{seas}$, VPRM$_{ann}$, CASA, and SiB4 at four towers: DNH in New Hampshire, TPD in Ontario, Canada, S01 in Indiana, and DVA in Virginia. Enhancements are determined using averaged convolutions with WRF-STILT and NAMS-STILT transport, "optimal" monthly background conditions, and Vulcan 3.0 + FFDAS fossil fuel emissions. The same figures using WRF-STILT and NAMS-STILT transport alone are shown in Figures S11 and S12 in Supporting Information S1.

months (especially January to March) at DNH and TPD, both CASA and SiB4 are biased high relative to observed enhancements.

3.4.2. Monthly Mean Atmospheric CO_2 Biases Across Towers

Across all 21 towers in the domain, VPRM$_{new}$ and VPRM$_{seas}$ are shown to produce relatively unbiased atmospheric CO_2 enhancements throughout the year (Figure 11, Table 2), whereas VPRM$_{ann}$ is biased low on average from June to August. The summertime bias in VPRM$_{ann}$ is due to not enough seasonal increase in R_e in this model, especially for the towers farthest from cropping areas. The summertime sink bias for VPRM$_{ann}$ is alleviated by using seasonal parameters in VPRM$_{seas}$ and the EVI covariate in the respiration model for VPRM$_{new}$, both of which allow for large realistic seasonal increases in R_e. This result also shows the strong influence of nighttime R_e fluxes on afternoon CO_2 observations (Figure S4 in Supporting Information S1) given that the VPRM versions differ primarily in terms of nighttime NEE (Figures 7 and 9), as also seen in X.-M. Hu et al. (2021) and Lauvaux et al. (2008, 2012).

During the summer months of June to August, VPRM$_{new}$ and CASA both produce relatively unbiased CO_2 enhancements, with CASA having the smallest spread across towers (Figure 11 and Figure S13 in Supporting
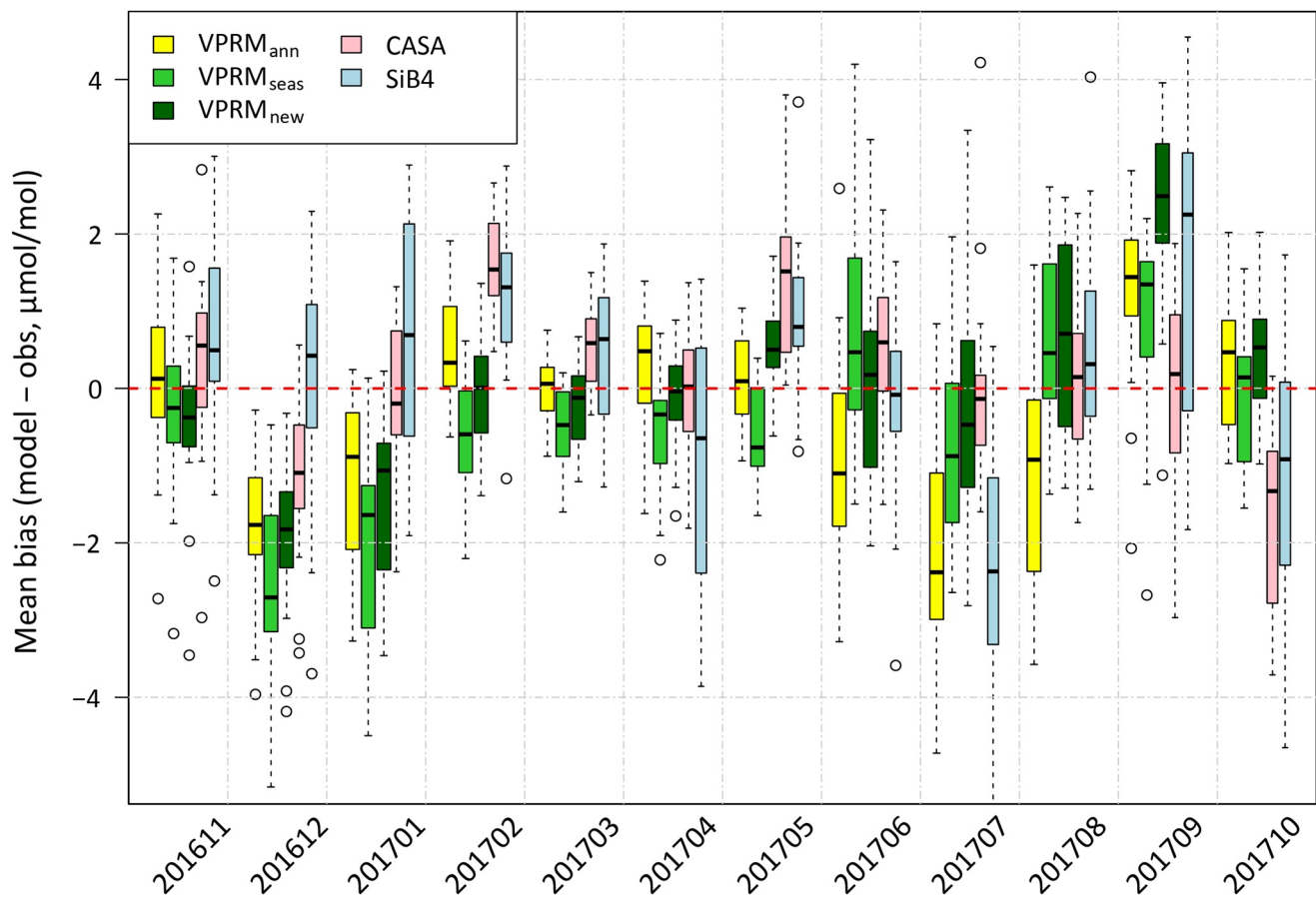
Figure 11. Boxplots of monthly mean biases (simulated - observed biospheric atmospheric CO_2 enhancements) across towers from November 2016 to October 2017 for each biospheric model, using mean of WRF-STILT and NAMS-STILT convolutions, "optimal" background conditions, and Vulcan 3.0 + FFDAS fossil fuel emissions. The same figures using WRF-STILT and NAMS-STILT transport alone are shown in Figure S13 in Supporting Information S1, as well as monthly mean biases for each individual tower with mean transport in Figure S14 in Supporting Information S1. The boxplots are the same as indicated in the caption to Figure 7.

Information S1, Table 2). However, in seasonal transition months like October and November and February to May, the VPRMs have the most skill in simulating phenological transitions, whereas CASA is biased low in October and high in May, pointing to difficulties in simulating CO_2 in transition periods. The higher temporal resolution diagnostic phenology (based on EVI) in VPRM relative to the monthly phenology (based on fPAR) in CASA can likely explain VPRM's higher skill in these months.

SiB4 tends to produce more biased CO_2 than VPRM and CASA throughout the year, with a larger spread across towers, particularly in January, April, July, and October, although it is relatively unbiased on average in June and August (Figure 11 and Table 2). The sensitivity test from May to October averaging all TBMs to the 0.5° scale before convolution with footprints showed that the relative biases among the TBMs changed very little with spatial resolution, implying that this was not the driving factor in explaining SiB4's poorer performance. (For example, the mean absolute error across towers in July for VPRM$_{new}$ went up from 1.2 μmol/mol at 0.1° to 1.3 μmol/mol at 0.5°, compared to 2.6 μmol/mol for SiB4 at 0.5°.) The SiB4 biases seen here are likely associated with errors in cropland phenology, flux magnitude, and spatial extent (i.e., for the Mississippi River Valley in April and the Corn Belt in July) which propagate downwind across the domain.

The VPRMs' underestimated enhancements in December and January and overestimation in September could be due to flux errors, but also potentially other components of the analysis. For example, in September, the positive biases for the VPRMs (especially VPRM$_{new}$) and SiB4 could be influenced by large background errors relative to small enhancements in this seasonal transition month (Figure 3b). The negative biases for VPRM in December and January could point to a bias in the flux tower nighttime NEE observations used for optimization, as discussed previously, and/or a systematic transport model error affecting both WRF-STILT and NAMS-STILT.

Table 2
Mean Absolute Error (MAE) Across Towers of Monthly Mean Biases Between Simulated and Observed Biospheric Atmospheric CO_2 Enhancements in µmol/mol, Shown for Each TBM and Month, Using Averaged WRF-STILT and NAMS-STILT Convolutions.

	$VPRM_{ann}$	$VPRM_{seas}$	$VPRM_{new}$	CASA	SiB4
201611	0.82	0.84	0.83	0.98	1.14
201612	1.84	2.59	1.93	1.22	1.15
201701	1.10	2.08	1.36	0.90	1.49
201702	0.68	0.77	0.62	1.58	1.33
201703	0.37	0.58	0.45	0.62	0.82
201704	0.75	0.75	0.51	0.70	1.45
201705	0.50	0.74	0.61	1.44	1.10
201706	1.48	1.17	1.11	0.92	0.96
201707	2.31	1.32	1.23	0.85	2.58
201708	1.34	1.01	1.22	0.94	1.05
201709	1.50	1.44	2.43	0.97	2.37
201710	0.81	0.79	0.80	1.68	1.60

Note. This metric indicates how well each TBM minimizes overall bias and captures spatial gradients in atmospheric CO_2 across towers within each month. These statistics also correspond to the monthly mean biases shown in Figure 11. MAE values ≤1.00 µmol/mol are highlighted in light yellow, with the TBM having an MAE≤1.00 µmol/mol and the minimal value across models is highlighted in orange.
(Corresponding tables using WRF-STILT or NAMS-STILT convolutions alone, and each set of background conditions are shown in table S6 in supporting information S1)

Evidence pointing towards a systematic transport model error is that CASA is also biased low in December and unbiased in January relative to atmospheric observations, whereas (Zhou, Williams, Lauvaux, Davis, et al., 2020) found overestimated wintertime R_e fluxes compared to flux tower data for this particular CASA implementation. In fact, R_e fluxes in SiB4 and CASA should be overestimated to some extent throughout the year due to the neutral biosphere assumptions discussed previously, which is seen for SiB4 in the atmospheric CO_2 comparisons from November to March, especially with WRF-STILT transport (Figure 10 and Figure S13 in Supporting Information S1).

3.4.3. Comparison of Hourly Atmospheric CO_2 Variability Across Towers

Comparisons of simulated and observed atmospheric CO_2 enhancements at the hourly timescale (across all towers by month) show that all the TBMs do a better job capturing high-resolution atmospheric CO_2 variability during the growing season (May to August) when the biospheric signal is stronger and the relative impact of other errors (transport, background, and fossil fuels) is smaller (Figure 12). In fact, in January, the mean biospheric enhancements across towers are of similar magnitude to the fossil fuel emission enhancements (Figure 3c), given the large extent over which R_e fluxes accumulate in this domain. The biospheric models also better explain hourly variability in atmospheric CO_2 throughout the year using the adjusted R^2 metric compared to the NSC, which penalizes biased flux estimates (as seen by negative NSC values for the VPRMs in December, CASA and SiB4 in February, and SiB4 in April, July, September, and October). However, NSC and adjusted R^2 values are no higher than 0.45 and 0.5, respectively, in any month, pointing to substantial unexplained variability in simulating atmospheric CO_2 variations from the biosphere throughout the year.

Across TBMs, $VPRM_{new}$ generally outperforms other models in reproducing hourly CO_2 variability using the adjusted R^2 metric (for 10 of 12 months) and during the growing season with the NSC (from April to August, but also February). In addition, the annual tower-specific adjusted R^2 (Table 3) comparing $VPRM_{new}$ to observations is highest or tied with other TBMs at all 21 towers, whereas NSC values with $VPRM_{new}$ are highest or tied at 18 of 21 towers (with S01 and the two southernmost towers, SMT and SCT as the only exceptions). Overall, these results point to remarkable skill for $VPRM_{new}$ in simulating fine-scale afternoon CO_2 variability across towers, likely due to a better representation of R_e relative to the other VPRMs, but also domain-specific optimized parameters and high temporal resolution diagnostic phenology common to all the VPRMs. This can be seen by
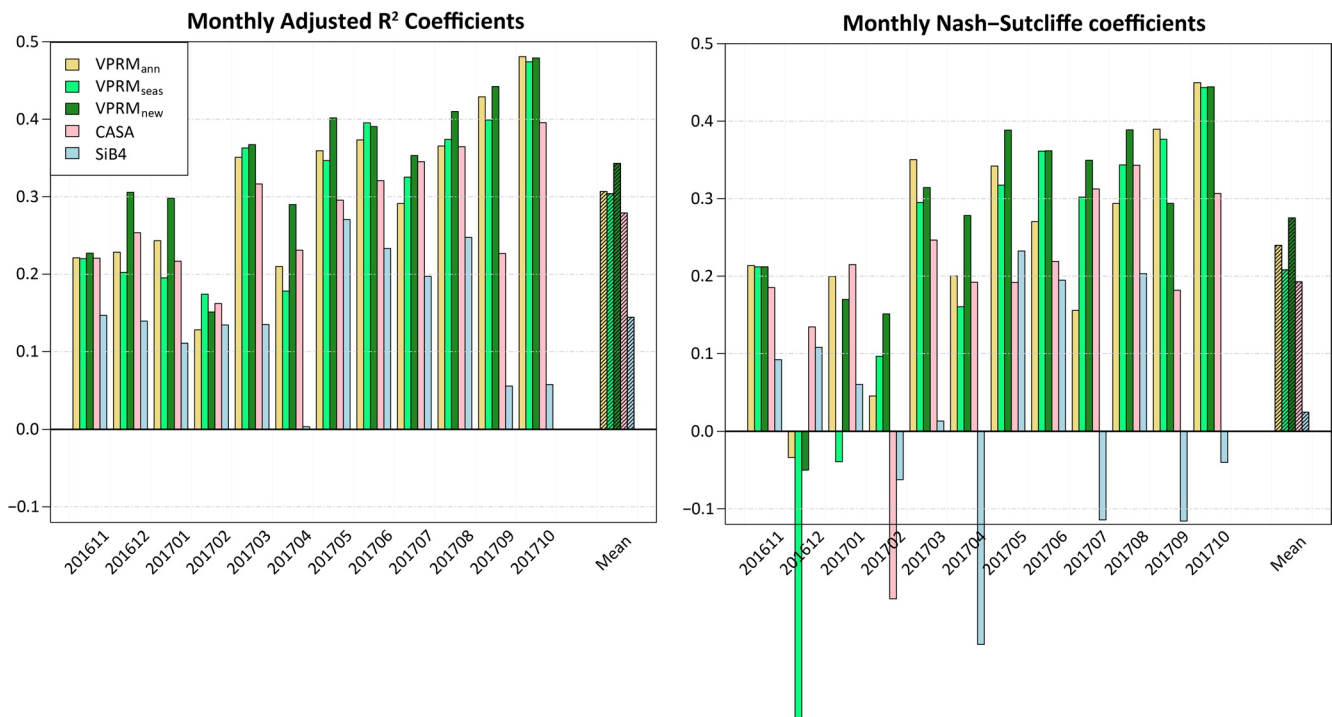
Figure 12. Monthly adjusted R^2 (bottom) and Nash-Sutcliffe coefficients (top) comparing simulated to observed biospheric atmospheric CO_2 enhancements at the hourly timescale across all towers by month. Averaged WRF-STILT and NAMS-STILT convolutions, Vulcan3.0 (+FFDAS in Canada) fossil fuel emissions, and "optimal" monthly background conditions are used for all comparisons. The same plots using WRF-STILT or NAMS-STILT transport alone are shown in Figure S15 in Supporting Information S1.

significantly higher adjusted R^2 and NSC values for all the VPRMs compared to CASA and SiB4 in March, May, June, September, and October, i.e., the seasonal transition months.

In most months, CASA shows only slightly lower adjusted R^2 and NSC values than the VPRMs, while SiB4 shows the lowest values, with negative NSC in February, April, July, September, and October and R^2s < 0.1 in April, September, and October. The sensitivity test averaging VPRM and CASA fluxes to 0.5° spatial resolution from May to October also had only marginal impact on the results shown here, indicating that SiB4 weaknesses in simulating hourly atmospheric CO_2 variability (as with the monthly mean biases) have less to do with spatial resolution than other model-specific factors.

3.4.4. Transport Model Uncertainty in Atmospheric CO_2 Analyses

The previous sections highlight lessons learned about the TBMs by comparing observed atmospheric CO_2 enhancements to simulations using the mean of two transport models (i.e., WRF-STILT and NAMS-STILT). However, we can also learn something about transport model uncertainty itself from these analyses. First, using WRF-STILT or NAMS-STILT convolutions alone is seen to have the most impact on monthly mean biases in the months of June, July, and August, when biospheric fluxes are strongest (as also seen in Feng, Lauvaux, Keller, et al., 2019; Figures S11–S13, Table S6 in Supporting Information S1).

Second, differences in footprint strength can make it difficult to properly evaluate flux magnitudes (e.g., in December and January), although transport model uncertainty appears to be low enough in this study to detect overall source or sink biases in flux estimates during the growing season. For example, in July, simulated enhancements for VPRM$_{ann}$ and SiB4 are both biased low relative to enhancements (indicating a sink bias in the fluxes), but these biases are roughly double with NAMS-STILT compared to WRF-STILT, making it difficult to assess the magnitude of the flux error. In contrast, in June and August, VPRM$_{new}$, VPRM$_{seas}$, and CASA convolutions are biased high with WRF-STILT, but are relatively unbiased with NAMS-STILT, making it difficult to ascertain whether a flux bias exists at all without additional information related to footprint quality for each model.

Table 3
Adjusted R^2 and NSC Metrics by Tower Comparing Modeled to Observed Biospheric Atmospheric CO_2 Enhancements at the Hourly Timescale for the Full Year

	VPRM$_{ann}$		VPRM$_{seas}$		VPRM$_{new}$		CASA		SiB4	
	r^2	NSC	r^2	NSC	r^2	NSC	r^2	NSC	r^2	NSC
LEF	**0.51**	**0.50**	0.50	0.48	0.51	0.49	0.45	0.43	0.41	0.38
AMT	0.39	0.32	0.40	0.38	**0.44**	**0.43**	0.34	0.29	0.32	0.27
DNH	0.60	0.45	0.62	0.56	**0.63**	**0.58**	0.55	0.51	0.44	0.27
UNY	0.56	0.38	0.57	0.51	**0.58**	**0.53**	0.50	0.42	0.40	0.17
TPD	0.51	0.51	0.49	0.48	**0.54**	**0.52**	0.52	0.50	0.36	0.15
HAF	0.53	0.47	0.54	**0.54**	**0.55**	0.53	0.50	0.43	0.39	0.21
MSH	0.28	0.10	0.29	0.22	**0.32**	0.27	0.31	**0.28**	0.20	0.01
MLD	0.62	0.46	0.62	0.57	**0.63**	**0.60**	0.54	0.50	0.47	0.24
BRI	0.66	0.65	0.63	0.63	**0.68**	**0.68**	0.66	0.66	0.53	0.49
HCT	0.41	**0.37**	0.39	0.32	**0.42**	0.36	0.38	0.36	0.37	0.36
SNJ	0.23	0.14	0.24	0.18	**0.25**	**0.22**	0.21	0.15	0.20	0.08
S01	0.67	**0.67**	0.64	0.63	**0.67**	0.64	0.67	0.66	0.57	0.55
TMD	0.39	0.30	0.41	0.39	**0.41**	**0.40**	0.31	0.25	0.26	0.21
BUC	0.51	0.50	0.49	0.48	**0.55**	**0.53**	0.51	0.50	0.41	0.40
SFD	0.41	0.38	0.47	**0.47**	**0.48**	0.47	0.44	0.43	0.28	0.28
RIC	0.60	0.59	0.60	0.58	**0.64**	**0.63**	0.56	0.56	0.46	0.46
SKY	0.51	0.49	0.53	**0.53**	**0.54**	0.53	0.46	0.35	0.27	0.26
DVA	0.56	0.53	0.55	0.52	**0.58**	**0.57**	0.51	0.50	0.42	0.41
MNC	0.52	0.52	0.49	0.48	**0.54**	**0.53**	0.51	0.51	0.40	0.39
SMT	0.18	0.12	0.15	-0.05	**0.24**	0.10	0.20	**0.14**	0.11	-0.10
SCT	0.40	0.39	0.40	0.37	0.41	0.38	**0.42**	**0.42**	0.26	0.25
mean	*0.48*	*0.42*	*0.48*	*0.44*	***0.51***	***0.48***	*0.46*	*0.42*	*0.36*	*0.27*

Note. Cells highlighted in gray are the highest across TBMs (within 0.01) for each tower, with the highest value for each metric and tower indicated in bold.

Third, we can also gather some clues about the relative quality of transport simulations by looking at systematic biases in atmospheric CO_2 across TBMs (although not a complete proof due to the possibility of similarly biased flux estimates across TBMs as well as errors in background conditions and fossil fuels). For example, simulated CO_2 for all TBMs is biased especially low with NAMS-STILT convolutions in July, and NSC values are also lower from June to September with NAMS-STILT compared to WRF-STILT (Figure S15 in Supporting Information S1), indicating that NAMS-STILT may have lower quality footprints than WRF-STILT during the growing season.

Finally, averaging convolutions from WRF-STILT and NAMS-STILT is shown to help improve correspondence with hourly atmospheric CO_2 variability. That is, both the NSC and adjusted R^2 metrics are higher for almost all months and TBMs when using averaged WRF-STILT and NAMS-STILT convolutions, compared to using convolutions with either transport model alone (Figure S15 in Supporting Information S1). For example, the mean year-round adjusted R^2 for VPRM$_{new}$ goes up to 0.34 with mean transport convolutions compared to 0.27 with WRF-STILT and 0.29 with NAMS-STILT, while the mean year-round NSC for CASA goes up to 0.19 compared to 0.12 with WRF-STILT and 0.11 with NAMS-STILT footprints (results not shown). This suggests that convolution averaging helps to reduce random transport errors and may also potentially cancel out some errors in footprint magnitude across transport models.

4. Synthesis, Future Work, and Conclusions

In this study, the VPRM biospheric model was customized for an eastern North American domain containing large tracts of forests and productive croplands interspersed with industrialized and urban areas. Parameters were optimized using flux tower observations from a large historical domain-specific database, using a re-evaluated PFT classification with separated northern and southern evergreen needleleaf (and mixed) forests, and corn

distinct from other crops. The respiration model was also modified to introduce more process-based realism (with EVI, quadratic temperature, and water stress factors) relative to the simple linear function of temperature in the original model. VPRM flux estimates were then evaluated a) at flux tower locations by comparing them to observations reserved from the parameter optimization, and b) across the landscape by comparison with output from other TBMs (i.e., CASA and SiB4) and atmospheric CO_2 observations at 21 towers in the domain.

Model-data comparisons at flux towers were used to evaluate the impact of the new VPRM respiration model, seasonal versus annual parameters, the parameter optimization technique (all 24 hour vs. night/day) and database resampling procedures. These comparisons showed that including EVI in the modified respiration model (i.e., $VPRM_{new}$) was the single most important factor in recovering realistic seasonality and fine-scale spatial variability for R_e fluxes across PFTs, but especially for crops, which have large seasonal changes in biomass. In contrast, the original respiration model (i.e., a simple linear function of temperature) with annual parameters ($VPRM_{ann}$) fails to capture peak R_e during the growing season, as well as sharp transitions in the spring and fall (Figures 6 and S5 in Supporting Information S1). Using seasonal parameters with the original respiration model ($VPRM_{seas}$) partially accounts for seasonal changes in baseline respiration, but not to the extent of $VPRM_{new}$. These comparisons also showed that the superior performance of the new respiration model in $VPRM_{new}$ was not only due to using a separate night-day optimization scheme for R_e and GEE parameters (compared to optimizing all parameters simultaneously), but also due to more physiological realism in the updated model equation. Finally, disproportionate filtering out of nighttime data based on u-star thresholds (to ensure adequate turbulence) can lead to flux estimates biased toward daytime uptake; thus, database resampling to ensure even coverage across the diurnal cycle may be warranted in future studies where biospheric model parameters are optimized using flux tower data.

For gridded flux estimates, the availability of continuous atmospheric CO_2 observations from a dense observing network in the domain allowed us to evaluate regional-scale fluxes across the landscape, and thus the implicit spatial scaling relationships embedded in PFT-specific parameters and land cover maps for VPRM. Gridded flux estimates from VPRM, CASA, and SiB4 were evaluated by first comparing their spatial patterns and seasonal and diurnal cycles, and then comparing simulated with observed atmospheric CO_2 enhancements within the context of uncertainty from transport models, background conditions, and fossil fuel emission estimates. The averaging of multiple transport models and a priori evaluation of multiple background condition products in this study hopefully helped to reduce the influence of these additional uncertainties on the analyses.

The results from atmospheric CO_2 comparisons show that $VPRM_{ann}$ has an overall sink bias during the growing season (only partially alleviated by using seasonal parameters in $VPRM_{seas}$). In contrast, $VPRM_{new}$ is relatively unbiased throughout the year (except in December and January), although CASA best captures spatial gradients across towers at the height of the growing season from June to August. The atmospheric CO_2 analyses also show that the high temporal resolution diagnostic phenology in all VPRM versions (with 8-day average EVI and 4-day average LSWI) is important for capturing seasonality in GEE and R_e, particularly in cropping areas. In contrast, CASA with monthly diagnostic phenology is more biased than the VPRM models in seasonal transition months like May and October, and SiB4 prognostic phenology appears to be too early in croplands for this year.

$VPRM_{new}$ also best explains atmospheric CO_2 enhancements at the hourly timescale throughout the year compared to other TBMs (Figure 12 and Table 3), showing the benefit of both the new respiration model as well as high-resolution diagnostic phenology for capturing fine-scale CO_2 variability in the atmosphere. Also, given the simpler process-based representation of R_e in $VPRM_{new}$ compared to CASA and SiB4, and the more empirical photosynthesis model in VPRM and CASA compared to SiB4, the superior performance of $VPRM_{new}$ in explaining fine-scale atmospheric CO_2 variability demonstrates that factors like high-quality gridded input data sets, diagnostic phenology and domain-specific parameters may be equally as important (or more) than process-based model complexity in simulating high-resolution gridded CO_2 fluxes (e.g., as also found in Raczka et al., 2013; Schwalm et al., 2010).

This study also shows the importance of adequately simulating agricultural fluxes in the Corn Belt and Mississippi River Valley for explaining atmospheric CO_2 in the growing season across this domain, due to their large flux magnitudes and upwind locations. For example, SiB4 has the largest cropland extent among the TBMs in its underlying land cover map, and along with errors in crop phenology for this year and a strong uptake for corn (too strong?), this results in a large sink bias in July for SiB4 compared to atmospheric CO_2 observations at almost all towers, which cannot be explained by its coarser spatial resolution in this study.

The underestimation of wintertime NEE for the VPRM's relative to atmospheric observations, as well as the smaller magnitudes of component fluxes (GEE and R_e) in VPRM relative to CASA and SiB4, remains an unresolved issue here. Given that the wintertime bias in VPRM NEE is not seen in the flux tower evaluation (Figure 7, bottom sub-plot), this could indicate a bias in the flux tower observations used for parameter optimization, due to nighttime drainage and morning venting (and hence under-observed nighttime R_e) and/or a systematic bias in both transport models (WRF-STILT and NAMS-STILT) in these months. Given the magnitude of the difference in component fluxes among the TBMs, and hence the shape of the diurnal cycle (especially for CASA), resolving this issue in future work is clearly warranted.

Other future potential improvements to VPRM include (a) incorporating SIF and/or NIRv into the GEE equation (e.g. (Badgley et al., 2019; Luus & Lin, 2015; Turner et al., 2020), to better simulate the impact of growing season transitions and water stress, and (b) further modifying the respiration equation to incorporate biomass maps, simultaneous GEE (to account for the large contribution of recently assimilated carbon to autotrophic respiration), disturbance maps, and/or week-to-week changes in EVI (to account for inputs to surface litter pools at the end of the growing season). In addition, new flux towers in additional locations, especially in the Appalachian deciduous broadleaf forests, the southern half of the domain, in urban areas, and across disturbance gradients, would help to improve the representativeness of flux estimates for all biospheric models with empirically derived parameters.

Overall, this study demonstrates the scientific value of using a dense atmospheric CO_2 observing network for evaluating TBM flux estimates at regional scales, and the analyses shown here form a prototype evaluation framework that can be used to guide future model improvements. Parallel and continuing development for each TBM, along with simultaneous evaluation using atmospheric measurements, can help to refine TBMs for the ultimate purpose of separating anthropogenic and biospheric signals to monitor fossil fuel CO_2 emissions using an atmospheric data constraint.

Data Availability Statement

Scripts to generate the results shown here, along with input data and summary files, are archived on the NIST server at https://data.nist.gov/od/id/mds2-2362. VPRM gridded flux estimates at 0.02°, hourly resolution from November 2016 to December 2020 are similarly archived at https://data.nist.gov/od/id/mds2-2382.

DOIs for other data sets used in the paper are listed below:

Atmospheric CO_2 Data: Karion, Anna, Prinzivalli, Steve, Fain, Clayton, Stock, Michael, DiGangi, Elizabeth, Biggs, Bryan, Draper, Charlie, Baldelli, Seth, Veseshta, Uran, Salameh, Peter, Callahan, William, Whetstone, James (2019), Observations of CO_2, CH_4, and CO mole fractions from the NIST Northeast Corridor urban testbed, National Institute of Standards and Technology, https://doi.org/10.18434/M32126.

Miles, N.L., S.J. Richardson, D.K. Martins, K.J. Davis, T. Lauvaux, B.J. Haupt, and S.K. Miller. 2018. ACT-America: L2 In Situ CO_2, CO, and CH_4 Concentrations from Towers, Eastern USA. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1568. Cooperative Global Atmospheric Data Integration Project: Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957–2018, obspack_co2_1_GLOBALVIEWplus_v5.0_2019_08_12, https://doi.org/10.25925/20190812, 2019.

Flux tower data:

Arain, M. Altaf (2003-) AmeriFlux CA-TP1 Ontario - Turkey Point 2002 Plantation White Pine. [Dataset]. https://doi.org/10.17190/AMF/1246009.

Arain, M. Altaf (2003-) AmeriFlux CA-TP3 Ontario - Turkey Point 1974 Plantation White Pine. [Dataset]. https://doi.org/10.17190/AMF/1246011.

Arain, M. Altaf (2012-) AmeriFlux CA-TPD Ontario - Turkey Point Mature Deciduous. [Dataset]. https://doi.org/10.17190/AMF/1246152.

Baker, John, Tim Griffis (2003–2010) AmeriFlux US-Ro3 Rosemount- G19. [Dataset]. https://doi.org/10.17190/AMF/1246093.

Baker, John, Tim Griffis (2003–2017) AmeriFlux US-Ro2 Rosemount- C7. [Dataset]. https://doi.org/10.17190/AMF/1418683.

Baker, John, Tim Griffis (2014-) AmeriFlux US-Ro4 Rosemount Prairie. [Dataset]. https://doi.org/10.17190/AMF/1419507.

Baker, John, Tim Griffis, Timothy Griffis (2003–2017) AmeriFlux US-Ro1 Rosemount- G21. [Dataset]. https://doi.org/10.17190/AMF/1246092.

Bernacchi, Carl (2004–2008) AmeriFlux US-Bo2 Bondville (Companion site). [Dataset]. https://doi.org/10.17190/AMF/1246037.

Biraud, Sebastien, Marc Fischer, Stephen Chan, Margaret Torn (2002-) AmeriFlux US-ARM ARM Southern Great Plains site- Lamont. [Dataset]. https://doi.org/10.17190/AMF/1246027.

Bohrer, Gil (2011–2016) AmeriFlux US-ORv Olentangy River Wetland Research Park. [Dataset]. https://doi.org/10.17190/AMF/1246135.

Bohrer, Gil, Janice Kerns (2015–2016) AmeriFlux US-OWC Old Woman Creek. [Dataset]. https://doi.org/10.17190/AMF/1418679.

Chen, Jiquan (2002-2002) AmeriFlux US-Wi8 Young hardwood clearcut (YHW). [Dataset]. https://doi.org/10.17190/AMF/1246023.

Chen, Jiquan (2002–2005) AmeriFlux US-Wi4 Mature red pine (MRP). [Dataset]. https://doi.org/10.17190/AMF/1246019.

Chen, Jiquan (2003-2003) AmeriFlux US-Wi1 Intermediate hardwood (IHW). [Dataset]. https://doi.org/10.17190/AMF/1246015.

Chen, Jiquan (2004-2004) AmeriFlux US-Wi5 Mixed young jack pine (MYJP). [Dataset]. https://doi.org/10.17190/AMF/1246020.

Chen, Jiquan (2004–2005) AmeriFlux US-Wi9 Young Jack pine (YJP). [Dataset]. https://doi.org/10.17190/AMF/1246024.

Chen, Jiquan (2005-2005) AmeriFlux US-Wi7 Red pine clearcut (RPCC). [Dataset]. https://doi.org/10.17190/AMF/1246022.

Chen, Jiquan, Housen Chu (2011–2013) AmeriFlux US-CRT Curtice Walter-Berger cropland. [Dataset]. https://doi.org/10.17190/AMF/1246156.

Chen, Jiquan, Housen Chu (2011–2013) AmeriFlux US-WPT Winous Point North Marsh. [Dataset]. https://doi.org/10.17190/AMF/1246155.

Chen, Jiquan, Housen Chu, Asko Noormets (2004–2013) AmeriFlux US-Oho Oak Openings. [Dataset]. https://doi.org/10.17190/AMF/1246089.

Clark, Ken (2004-) AmeriFlux US-Slt Silas Little- New Jersey. [Dataset]. https://doi.org/10.17190/AMF/1246096.

Clark, Ken (2005–2008) AmeriFlux US-Dix Fort Dix. [Dataset]. https://doi.org/10.17190/AMF/1246045.

Clark, Ken (2005-) AmeriFlux US-Ced Cedar Bridge. [Dataset]. https://doi.org/10.17190/AMF/1246043.

Desai, Ankur (1996-) AmeriFlux US-PFa Park Falls/WLEF. [Dataset]. https://doi.org/10.17190/AMF/1246090.

Desai, Ankur (1999-) AmeriFlux US-WCr Willow Creek. [Dataset]. https://doi.org/10.17190/AMF/1246111.

Desai, Ankur (2001-) AmeriFlux US-Los Lost Creek. [Dataset]. https://doi.org/10.17190/AMF/1246071.

Desai, Ankur (2001-) AmeriFlux US-Syv Sylvania Wilderness Area. [Dataset]. https://doi.org/10.17190/AMF/1246106.

Drake, Bert, Ross Hinkle (2000–2007) AmeriFlux US-KS2 Kennedy Space Center (scrub Oak). [Dataset]. https://doi.org/10.17190/AMF/1246070.

Gough, Christopher, Gil Bohrer, Peter Curtis (1999-) AmeriFlux US-UMB Univ. of Mich. Biological Station. [Dataset]. https://doi.org/10.17190/AMF/1246107.

Gough, Christopher, Gil Bohrer, Peter Curtis (2007-) AmeriFlux US-UMd UMBS Disturbance. [Dataset]. https://doi.org/10.17190/AMF/1246134.

Hadley, Julian, J. William Munger (2004-) AmeriFlux US-Ha2 Harvard Forest Hemlock Site. [Dataset]. https://doi.org/10.17190/AMF/1246060.

Heilman, Jim (2004-) AmeriFlux US-FR3 Freeman Ranch- Woodland. [Dataset]. https://doi.org/10.17190/AMF/1246055.

Hollinger, David (1996-) AmeriFlux US-Ho1 Howland Forest (main tower). [Dataset]. https://doi.org/10.17190/AMF/1246061.

Hollinger, David (1999-) AmeriFlux US-Ho2 Howland Forest (west tower). [Dataset]. https://doi.org/10.17190/AMF/1246062.

Hollinger, David (2000-) AmeriFlux US-Ho3 Howland Forest (harvest site). [Dataset]. https://doi.org/10.17190/AMF/1246063.

Lee, Xuhui (1999–2004) AmeriFlux US-GMF Great Mountain Forest. [Dataset]. https://doi.org/10.17190/AMF/1246057.

Matamala, Roser (2004-) AmeriFlux US-IB2 Fermi National Accelerator Laboratory- Batavia (Prairie site). [Dataset]. https://doi.org/10.17190/AMF/1246066.

Matamala, Roser (2005-) AmeriFlux US-IB1 Fermi National Accelerator Laboratory- Batavia (Agricultural site). [Dataset]. https://doi.org/10.17190/AMF/1246065.

McCaughey, Harry (2003-) AmeriFlux CA-Gro Ontario - Groundhog River, Boreal Mixedwood Forest. [Dataset]. https://doi.org/10.17190/AMF/1245996.

McFadden, Joe (2005–2009) AmeriFlux US-KUT KUOM Turfgrass Field. [Dataset]. https://doi.org/10.17190/AMF/1246145.

Meyers, Tilden (1995–1999) AmeriFlux US-WBW Walker Branch Watershed. [Dataset]. https://doi.org/10.17190/AMF/1246109.

Meyers, Tilden (1996-) AmeriFlux US-Bo1 Bondville. [Dataset]. https://doi.org/10.17190/AMF/1246036.

Meyers, Tilden (2002–2006) AmeriFlux US-Goo Goodwin Creek. [Dataset]. https://doi.org/10.17190/AMF/1246058.

Meyers, Tilden (2004-) AmeriFlux US-CaV Canaan Valley. [Dataset]. https://doi.org/10.17190/AMF/1246042.

Meyers, Tilden (2005-) AmeriFlux US-ChR Chestnut Ridge. [Dataset]. https://doi.org/10.17190/AMF/1246044.

Munger, J. William (1991-) AmeriFlux US-Ha1 Harvard Forest EMS Tower (HFR1). [Dataset]. https://doi.org/10.17190/AMF/1246059.

Noormets, Asko (2005-) AmeriFlux US-NC2 NC_Loblolly Plantation. [Dataset]. https://doi.org/10.17190/AMF/1246083.

Noormets, Asko (2005–2013) AmeriFlux US-NC1 NC_Clearcut. [Dataset]. https://doi.org/10.17190/AMF/1246082.

Noormets, Asko (2013-) AmeriFlux US-NC3 NC_Clearcut#3. [Dataset]. https://doi.org/10.17190/AMF/1419506.

Novick, Kim, Rich Phillips (1999-) AmeriFlux US-MMS Morgan Monroe State Forest. [Dataset]. https://doi.org/10.17190/AMF/1246080.

Oishi, Chris, Kim Novick, Paul Stoy (2001–2008) AmeriFlux US-Dk1 Duke Forest-open field. [Dataset]. https://doi.org/10.17190/AMF/1246046.

Oishi, Chris, Kim Novick, Paul Stoy (2001–2008) AmeriFlux US-Dk2 Duke Forest-hardwoods. [Dataset]. https://doi.org/10.17190/AMF/1246047.

Oishi, Chris, Kim Novick, Paul Stoy (2001–2008) AmeriFlux US-Dk3 Duke Forest - loblolly pine. [Dataset]. https://doi.org/10.17190/AMF/1246048.

Prueger, John, Tim Parkin (2001-) AmeriFlux US-Br1 Brooks Field Site 10- Ames. [Dataset]. https://doi.org/10.17190/AMF/1246038.

Prueger, John, Tim Parkin (2001-) AmeriFlux US-Br3 Brooks Field Site 11- Ames. [Dataset]. https://doi.org/10.17190/AMF/1246039.

Richardson, Andrew, David Hollinger (2004-) AmeriFlux US-Bar Bartlett Experimental Forest. [Dataset]. https://doi.org/10.17190/AMF/1246030.

Sturtevant, Cove, David Durden, Stefan Metzger (2016-) AmeriFlux US-xSC NEON Smithsonian Conservation Biology Institute (SCBI). [Dataset]. https://doi.org/10.17190/AMF/1671900.

Sturtevant, Cove, David Durden, Stefan Metzger (2016-) AmeriFlux US-xSE NEON Smithsonian Environmental Research Center (SERC). [Dataset]. https://doi.org/10.17190/AMF/1617734.

Sturtevant, Cove, David Durden, Stefan Metzger (2017-) AmeriFlux US-xDL NEON Dead Lake (DELA). [Dataset]. https://doi.org/10.17190/AMF/1579721.

Sturtevant, Cove, David Durden, Stefan Metzger (2017-) AmeriFlux US-xGR NEON Great Smoky Mountains National Park, Twin Creeks (GRSM). [Dataset]. https://doi.org/10.17190/AMF/1634885.

Sturtevant, Cove, David Durden, Stefan Metzger (2017-) AmeriFlux US-xST NEON Steigerwaldt Land Services (STEI). [Dataset]. https://doi.org/10.17190/AMF/1617737.

Sturtevant, Cove, David Durden, Stefan Metzger (2017-) AmeriFlux US-xTA NEON Talladega National Forest (TALL). [Dataset]. https://doi.org/10.17190/AMF/1671902.

Sturtevant, Cove, David Durden, Stefan Metzger (2017-) AmeriFlux US-xTR NEON Treehaven (TREE). [Dataset]. https://doi.org/10.17190/AMF/1634886.

Sturtevant, Cove, David Durden, Stefan Metzger (2017-) AmeriFlux US-xUK NEON The University of Kansas Field Station (UKFS). [Dataset]. https://doi.org/10.17190/AMF/1617740.

Sturtevant, Cove, David Durden, Stefan Metzger (2017-) AmeriFlux US-xUN NEON University of Notre Dame Environmental Research Center (UNDE). [Dataset]. https://doi.org/10.17190/AMF/1617741.

Suyker, Andy (2001-) AmeriFlux US-Ne2 Mead - irrigated maize-soybean rotation site. [Dataset]. https://doi.org/10.17190/AMF/1246085.

Suyker, Andy (2001-) AmeriFlux US-Ne3 Mead - rainfed maize-soybean rotation site. [Dataset]. https://doi.org/10.17190/AMF/1246086.

Torn, Margaret (2005–2006) AmeriFlux US-ARc ARM Southern Great Plains control site- Lamont. [Dataset]. https://doi.org/10.17190/AMF/1246026.

Wood, Jeffrey, Lianhong Gu (2004-) AmeriFlux US-MOz Missouri Ozark Site. [Dataset]. https://doi.org/10.17190/AMF/1246081.

References

Amthor, J. S. (2000). The McCree–de Wit–Penning de Vries–Thornley respiration paradigms: 30 years later. *Annals of Botany*, *86*(1), 1–20. https://doi.org/10.1006/anbo.2000.1175

Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., et al. (2014). A multiyear, global gridded fossil fuel CO2 emission data product: Evaluation and analysis of results. *Journal of Geophysical Research: Atmospheres*, *119*(17, 10213–10231. https://doi.org/10.1002/2013JD021296

Atkin, O. K., Bloomfield, K. J., Reich, P. B., Tjoelker, M. G., Asner, G. P., Bonal, D., et al. (2015). Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. *New Phytologist*, *206*(2), 614–636. https://doi.org/10.1111/nph.13253

Aubinet, M. (2008). Eddy Covariance CO2 flux measurements in nocturnal conditions: An analysis of the problem. *Ecological Applications*, *18*(6), 1368–1378. https://doi.org/10.1890/06-1336.1

Badgley, G., Anderegg, L. D. L., Berry, J. A., & Field, C. B. (2019). Terrestrial gross primary production: Using NIRV to scale from site to globe. *Global Change Biology*, *25*(11), 3731–3740. https://doi.org/10.1111/gcb.14729

Barr, A. G., Richardson, A. D., Hollinger, D. Y., Papale, D., Arain, M. A., Black, T. A., et al. (2013). Use of change-point detection for friction–velocity threshold evaluation in eddy-covariance studies. *Agricultural and Forest Meteorology*, *171–172*, 31–45. https://doi.org/10.1016/j.agrformet.2012.11.023

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The rapid refresh. *Monthly Weather Review*, *144*(4), 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Bonan, G. B., Levis, S., Kergoat, L., & Oleson, K. W. (2002). Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, *16*(2), 55–123. https://doi.org/10.1029/2000GB001360

Buyantuyev, A., & Wu, J. (2009). Urbanization alters spatiotemporal patterns of ecosystem primary production: A case study of the Phoenix metropolitan region, USA. *Journal of Arid Environments*, *73*(4), 512–520. https://doi.org/10.1016/j.jaridenv.2008.12.015

Chandrasekar, K., Sai, M. V. R. S., Roy, P. S., & Dwevedi, R. S. (2010). Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product. *International Journal of Remote Sensing*, *31*(15), 3987–4005. https://doi.org/10.1080/01431160802575653

Chernick, M. R. (2007). *Bootstrap methods: A guide for Practitioners and Researchers* (2nd ed.). Hoboken, N.J: Wiley-Interscience.

Elder, J. (2018). Chapter 16 – The apparent paradox of complexity in ensemble modeling. In R. Nisbet, G. Miner, & K. Yale (Eds.), *Handbook of statistical analysis and data mining applications* (2nd ed., pp. 705–718). Boston: Academic Press. https://doi.org/10.1016/B978-0-12-416632-5.00016-5

Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta*, *149*(1), 78–90. https://doi.org/10.1007/BF00386231

Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., et al. (2019). Seasonal characteristics of model uncertainties from biogenic fluxes, transport, and large-scale boundary inflow in atmospheric CO2 simulations over North America. *Journal of Geophysical Research: Atmospheres*, *124*(24), 14325–14346. https://doi.org/10.1029/2019JD031165

Feng, S., Lauvaux, T., Keller, K., Davis, K. J., Rayner, P., Oda, T., & Gurney, K. R. (2019). A road map for improving the treatment of uncertainties in high-resolution regional carbon flux inverse estimates. *Geophysical Research Letters*, *46*(22), 13461–13469. https://doi.org/10.1029/2019GL082987

Fisher, J. B., Huntzinger, D. N., Schwalm, C. R., & Sitch, S. (2014). Modeling the terrestrial biosphere. *Annual Review of Environment and Resources*, *39*(1), 91–123. https://doi.org/10.1146/annurev-environ-012913-093456

Fisher, J. B., Sikka, M., Huntzinger, D. N., Schwalm, C., & Liu, J. (2016). Technical note: 3-hourly temporal downscaling of monthly global terrestrial biosphere model net ecosystem exchange. *Biogeosciences*, *13*(14), 4271–4277. https://doi.org/10.5194/bg-13-4271-2016

Flexas, J., Bota, J., Galmés, J., Medrano, H., & Ribas-Carbó, M. (2006). Keeping a positive carbon balance under adverse conditions: Responses of photosynthesis and respiration to water stress. *Physiologia Plantarum*, *127*(3), 343–352. https://doi.org/10.1111/j.1399-3054.2006.00621.x

Golubiewski, N. E. (2006). Urbanization increases grassland carbon pools: Effects of landscaping In Colorado's Front Range. *Ecological Applications*, *16*(2), 555–571. https://doi.org/10.1890/1051-0761(2006)016[0555:UIGCPE]2.0.CO;2

Gray, J. M., Frolking, S., Kort, E. A., Ray, D. K., Kucharik, C. J., Ramankutty, N., & Friedl, M. A. (2014). Direct human influence on atmospheric CO2 seasonality from increased cropland productivity. *Nature*, *515*(7527), 398–401. https://doi.org/10.1038/nature13957

Gurney, K. R., Liang, J., Patarasuk, R., Song, Y., Huang, J., & Roest, G. (2020). The Vulcan version 3.0 high-resolution fossil fuel CO2 emissions for the United States. *Journal of Geophysical Research: Atmospheres*, *125*(19), e2020JD032974. https://doi.org/10.1029/2020JD032974

Hardiman, B. S., Wang, J. A., Hutyra, L. R., Gately, C. K., Getson, J. M., & Friedl, M. A. (2017). Accounting for urban biogenic fluxes in regional carbon budgets. *The Science of the Total Environment*, *592*, 366–372. https://doi.org/10.1016/j.scitotenv.2017.03.028

Haynes, K. D., Baker, I. T., Denning, A. S., Stöckli, R., Schaefer, K., Lokupitiya, E. Y., & Haynes, J. M. (2019). Representing grasslands using dynamic prognostic phenology based on biological growth stages: 1. Implementation in the Simple Biosphere Model (SiB4). *Journal of Advances in Modeling Earth Systems*, *11*(12), 4423–4439. https://doi.org/10.1029/2018MS001540

Haynes, K. D., Baker, I. T., Denning, A. S., Wolf, S., Wohlfahrt, G., Kiely, G., et al. (2019). Representing grasslands using dynamic prognostic phenology based on biological growth stages: Part 2. Carbon cycling. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4440–4465. https://doi.org/10.1029/2018MS001541

Hilton, T. W., Davis, K. J., & Keller, K. (2014). Evaluating terrestrial CO2 flux diagnoses and uncertainties from a simple land surface model and its residuals. *Biogeosciences*, *11*(2), 217–235. https://doi.org/10.5194/bg-11-217-2014

Hilton, T. W., Davis, K. J., Keller, K., & Urban, N. M. (2013). Improving North American terrestrial CO2 flux diagnosis using spatial structure in land surface model residuals. *Biogeosciences*, *10*(7), 4607–4625. https://doi.org/10.5194/bg-10-4607-2013

Hilton, T. W., Whelan, M. E., Zumkehr, A., Kulkarni, S., Berry, J. A., Baker, I. T., et al. (2017). Peak growing season gross uptake of carbon in North America is largest in the Midwest USA. *Nature Climate Change*, *7*(6), 450–454. https://doi.org/10.1038/nclimate3272

Högberg, P., Nordgren, A., Buchmann, N., Taylor, A. F., Ekblad, A., Högberg, M. N., et al. (2001). Large-scale forest girdling shows that current photosynthesis drives soil respiration. *Nature*, *411*(6839), 789–792. https://doi.org/10.1038/35081058

Hu, L., Andrews, A. E., Thoning, K. W., Sweeney, C., Miller, J. B., Michalak, A. M., et al. (2019). Enhanced North American carbon uptake associated with El Niño. *Science Advances*, *5*(6), eaaw0076. https://doi.org/10.1126/sciadv.aaw0076

Hu, X.-M., Gourdji, S. M., Davis, K. J., Wang, Q., Zhang, Y., Xue, M., et al. (2021). Implementation of improved parameterization of terrestrial flux in WRF-VPRM improves the simulation of nighttime CO2 peaks and a daytime CO2 band ahead of a cold front. *Journal of Geophysical Research: Atmospheres*, *126*(10), e2020JD034362. https://doi.org/10.1029/2020JD034362

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., et al. (2012). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, *232*, 144–157. https://doi.org/10.1016/j.ecolmodel.2012.02.004

Jacobson, A. R., Schuldt, K. N., Miller, J. B., Oda, T., Tans, P., Andrews, A., et al. (2020). CarbonTracker CT2019B. https://doi.org/10.25925/20201008

Jassal, R. S., Black, T. A., Cai, T., Morgenstern, K., Li, Z., Gaumont-Guay, D., & Nesic, Z. (2007). Components of ecosystem respiration and an estimate of net primary productivity of an intermediate-aged Douglas-fir stand. *Agricultural and Forest Meteorology*, *144*(1), 44–57. https://doi.org/10.1016/j.agrformet.2007.01.011

Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., et al. (2019). Overall methodology design for the United States National Land Cover Database 2016 Products. *Remote Sensing*, *11*(24), 2971. https://doi.org/10.3390/rs11242971

Joiner, J., & Yoshida, Y. (2020). Satellite-based reflectances capture large fraction of variability in global gross primary production (GPP) at weekly time scales. *Agricultural and Forest Meteorology*, *291*, 108092. https://doi.org/10.1016/j.agrformet.2020.108092

Karion, A., Callahan, W., Stock, M., Prinzivalli, S., Verhulst, K. R., Kim, J., et al. (2020). Greenhouse gas observations from the Northeast Corridor tower network. *Earth System Science Data*, *12*(1), 699–717. https://doi.org/10.5194/essd-12-699-2020

Karion, A., Lopez-Coto, I., Gourdji, S. M., Mueller, K., Ghosh, S., Callahan, W., et al. (2021). Background conditions for an urban greenhouse gas network in the Washington, D.C. and Baltimore metropolitan region. *Atmospheric Chemistry and Physics Discussions*, 1–27. https://doi.org/10.5194/acp-2020-1256

Lauvaux, T., Gurney, K. R., Miles, N. L., Davis, K. J., Richardson, S. J., Deng, A., et al. (2021). Policy-relevant assessment of urban CO2 emissions, *Environmental Science & Technology*, *54*(16), 10237–10245. https://pubs.acs.org/doi/abs/10.1021/acs.est.0c00343

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., et al. (2012). Constraining the CO$_2$ budget of the corn belt: Exploring uncertainties from the assumptions in a mesoscale inverse system. *Atmospheric Chemistry and Physics*, *12*(1), 337–354. https://doi.org/10.5194/acp-12-337-2012

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., et al. (2008). Mesoscale inversion: First results from the CERES campaign with synthetic data. *Atmospheric Chemistry and Physics*, *8*(13), 3459–3471. https://doi.org/10.5194/acp-8-3459-2008

Li, D., Bou-Zeid, E., Barlage, M., Chen, F., & Smith, J. A. (2013). Development and evaluation of a mosaic approach in the WRF-Noah framework. *Journal of Geophysical Research: Atmospheres*, *118*, 11918–11935. https://doi.org/10.1002/2013JD020657

Li, X., Hu, X.-M., Cai, C., Jia, Q., Zhang, Y., Liu, J., et al. (2020). Terrestrial CO2 fluxes, concentrations, sources and budget in Northeast China: Observational and modeling studies. *Journal of Geophysical Research: Atmospheres*, *125*(6), e2019JD031686. https://doi.org/10.1029/2019JD031686

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., & Grainger, C. A. (2003). A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model. *Journal of Geophysical Research: Atmospheres*, *108*(D16). https://doi.org/10.1029/2002JD003161

Lokupitiya, E., Denning, S., Paustian, K., Baker, I., Schaefer, K., Verma, S., et al. (2009). Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands. *Biogeosciences*, *6*(6), 969–986. https://doi.org/10.5194/bg-6-969-2009

Luus, K. A., & Lin, J. C. (2015). The Polar Vegetation Photosynthesis and Respiration Model: A parsimonious, satellite-data-driven model of high-latitude CO$_2$ exchange. *Geoscientific Model Development*, *8*(8), 2655–2674. https://doi.org/10.5194/gmd-8-2655-2015

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., et al. (2008). A satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM). *Global Biogeochemical Cycles*, *22*(2). https://doi.org/10.1029/2006GB002735

Martin, C. R., Zeng, N., Karion, A., Mueller, K., Ghosh, S., Lopez-Coto, I., et al. (2019). Investigating sources of variability and error in simulations of carbon dioxide in an urban region. *Atmospheric Environment*, *199*, 55–69. https://doi.org/10.1016/j.atmosenv.2018.11.013

Meir, P., Metcalfe, D. b, Costa, A. c. l, & Fisher, R. a. (2008). The fate of assimilated carbon during drought: Impacts on respiration in Amazon rainforests. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1498), 1849–1855. https://doi.org/10.1098/rstb.2007.0021

Messerschmidt, J., Parazoo, N., Wunch, D., Deutscher, N. M., Roehl, C., Warneke, T., & Wennberg, P. O. (2013). Evaluation of seasonal atmosphere-biosphere exchange estimations with TCCON measurements. *Atmospheric Chemistry and Physics*, *13*, 5103–5115. https://doi.org/10.5194/acp-13-5103-2013

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Meyer, N., Welp, G., & Amelung, W. (2018). The temperature sensitivity (Q10) of soil respiration: Controlling factors and spatial prediction at regional scale based on environmental soil classes. *Global Biogeochemical Cycles*, *32*(2), 306–323. https://doi.org/10.1002/2017GB005644

Miles, N. L., Richardson, S. J., Martins, D. K., Davis, K. J., Lauvaux, T., Haupt, B. J., & Miller, S. K. (2018). ACT-America: L2 in situ CO2, CO, and CH4 concentrations from Towers, Eastern USA. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1568

Mitchell, L., Lin, J. C., Hutyra, L. R., Sargent, M., Wofsy, S. C., Miles, N. L., et al. (2019). NACP: Urban greenhouse gases across the CO2 urban synthesis and analysis network. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/1743

Molchanov, A. G. (2009). Effect of moisture availability on photosynthetic productivity and autotrophic respiration of an oak stand. *Russian Journal of Plant Physiology*, *56*(6), 769–779. https://doi.org/10.1134/S1021443709060065

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Mueller, K., Yadav, V., Lopez-Coto, I., Karion, A., Gourdji, S., Martin, C., & Whetstone, J. (2018). Siting background towers to characterize incoming air for urban greenhouse gas estimation: A case study in the Washington, DC/Baltimore Area. *Journal of Geophysical Research: Atmospheres*, *123*(5), 2910–2926. https://doi.org/10.1002/2017JD027364

NCEI, NWS, NOAA, & US DOC (2020). North American Mesoscale Forecast System (NAM) [12 km]. Retrieved from https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00630

Nicolini, G., Aubinet, M., Feigenwinter, C., Heinesch, B., Lindroth, A., Mamadou, O., et al. (2018). Impact of CO_2 storage flux sampling uncertainty on net ecosystem exchange measured by eddy covariance. *Agricultural and Forest Meteorology*, *248*, 228–239. https://doi.org/10.1016/j.agrformet.2017.09.025

NOAA Earth System Research Laboratory. (2019). Global Monitoring Division.Cooperative Global Atmospheric Data Integration Project-Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2018; obspack_co2_1_GLOBALVIEWplus_v5.0_2019_08_12. https://doi.org/10.25925/20190812

Nowak, D. J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution*, *116*(3), 381–389. https://doi.org/10.1016/S0269-7491(01)00214-7

Peters, W., Krol, M. C., Werf, G. R. V. D., Houweling, S., Jones, C. D., Hughes, J., et al. (2010). Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations. *Global Change Biology*, *16*(4), 1317–1337. https://doi.org/10.1111/j.1365-2486.2009.02078.x

Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., et al. (2011). High-resolution simulations of atmospheric CO_2 over complex terrain – Representing the Ochsenkopf mountain tall tower. *Atmospheric Chemistry and Physics*, *11*(15), 7445–7464. https://doi.org/10.5194/acp-11-7445-2011

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

Raciti, S. M., Hutyra, L. R., & Newell, J. D. (2014). Mapping carbon storage in urban trees with multi-source remote sensing data: Relationships between biomass, land use, and demographics in Boston neighborhoods. *The Science of the Total Environment*, *500–501*, 72–83. https://doi.org/10.1016/j.scitotenv.2014.08.070

Raczka, B. M., Davis, K. J., Huntzinger, D., Neilson, R. P., Poulter, B., Richardson, A. D., et al. (2013). Evaluation of continental carbon cycle simulations with North American flux tower observations. *Ecological Monographs*, *83*(4), 531–556. https://doi.org/10.1890/12-0893.1

Randerson, J. T., Thompson, M. V., Malmstrom, C. M., Field, C. B., & Fung, I. Y. (1996). Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO2. *Global Biogeochemical Cycles*, *10*(4), 585–602. https://doi.org/10.1029/96GB01981

Reitz, O., Graf, A., Schmidt, M., Ketzler, G., & Leuchner, M. (2021). Upscaling net ecosystem exchange over heterogeneous landscapes with machine learning. *Journal of Geophysical Research: Biogeosciences*, *126*(2), e2020JG005814. https://doi.org/10.1029/2020JG005814

Richardson, S. J., Miles, N. L., Davis, K. J., Lauvaux, T., Martins, D. K., Turnbull, J. C., et al. (2017). Tower measurement network of in-situ CO2, CH4, and CO in support of the Indianapolis FLUX (INFLUX) Experiment. *Elementa: Science of the Anthropocene*, *5*(59). https://doi.org/10.1525/elementa.140

Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., et al. (2010). A model-data intercomparison of CO2 exchange across North America: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research: Biogeosciences*, *115*(G3). https://doi.org/10.1029/2009JG001229

Sellers, P. J., Mintz, Y., Sud, Y. C., & Dalcher, A. (1986). A Simple Biosphere Model (SIB) for use within general circulation models. *Journal of the Atmospheric Sciences*, *43*(6), 505–531. https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., et al. (1996). A Revised Land Surface Parameterization (SiB2) for atmospheric GCMS. Part I: Model formulation. *Journal of Climate*, *9*(4), 676–705. https://doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., & Yadav, V. (2014). Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO2 measurements. *Geophysical Research Letters*, *41*(12), 4381–4388. https://doi.org/10.1002/2014GL059684

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., et al. (2008). *A description of the advanced research WRF version 3 [Application/pdf]* (p. 1002). UCAR/NCAR. https://doi.org/10.5065/D68S4MVH

Turner, A. J., Köhler, P., Magney, T. S., Frankenberg, C., Fung, I., & Cohen, R. C. (2020). A double peak in the seasonality of California's photosynthesis as observed from space. *Biogeosciences*, *17*(2), 405–422. https://doi.org/10.5194/bg-17-405-2020

Williams, C. A., Sundquist, E. T., Poulter, B., Hoffman, F. M., Davis, K. J., Brown, M., et al. (2021). *NACP science implementation plan* (p. 2021). Retrieved from https://nacarbon.org/nacp/implementation_plan.html

Wullschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., et al. (2014). Plant functional types in Earth system models: Past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. *Annals of Botany*, *114*(1), 1–16. https://doi.org/10.1093/aob/mcu077

Xiao, J., Davis, K. J., Urban, N. M., Keller, K., & Saliendra, N. Z. (2011). Upscaling carbon fluxes from towers to the regional scale: Influence of parameter variability and land cover representation on regional flux estimates. *Journal of Geophysical Research: Biogeosciences*, *116*(G3). https://doi.org/10.1029/2010JG001568

Xiao, X., Hollinger, D., Aber, J., Goltz, M., Davidson, E. A., Zhang, Q., & Iii, B. M. (2004). *Satellite-based modeling of gross primary production in an evergreen needleleaf forest*. Retrieved from https://www.nrs.fs.fed.us/pubs/6798

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., et al. (2018). A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 108–123. https://doi.org/10.1016/j.isprsjprs.2018.09.006

Zeng, N., Zhao, F., Collatz, G. J., Kalnay, E., Salawitch, R. J., West, T. O., & Guanter, L. (2014). Agricultural Green Revolution as a driver of increasing atmospheric CO2 seasonal amplitude. *Nature*, *515*(7527), 394–397. https://doi.org/10.1038/nature13893

Zhou, Y., Williams, C. A., Lauvaux, T., Davis, K. J., Feng, S., Baker, I., et al. (2020). A multiyear gridded data ensemble of surface biogenic carbon fluxes for North America: Evaluation and analysis of results. *Journal of Geophysical Research: Biogeosciences*, *125*(2), e2019JG005314. https://doi.org/10.1029/2019JG005314

Zhou, Y., Williams, C. A., Lauvaux, T., Feng, S., Baker, I. T., Wei, Y., et al. (2020). Atmospheric carbon and transport - America (ACT-America) ACT-America: Gridded ensembles of surface biogenic carbon fluxes, 2003-2019(version 1.1) [NetCDF], 0 MB. https://doi.org/10.3334/ORNLDAAC/1675