AGU Advances

RESEARCH ARTICLE
10.1029/2020AV000310

Midwest US Croplands Determine Model Divergence in North American Carbon Fluxes

Wu Sun[1], Yuanyuan Fang[1,2], Xiangzhong Luo[3,4,5], Yoichi P. Shiga[1,6], Yao Zhang[3,4], Arlyn E. Andrews[7], Kirk W. Thoning[7], Joshua B. Fisher[8], Trevor F. Keenan[3,4], and Anna M. Michalak[1]

[1]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA, [2]Bay Area Air Quality Management District, San Francisco, CA, USA, [3]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [4]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA, [5]Department of Geography, National University of Singapore, Singapore, [6]Universities Space Research Association, Mountain View, CA, USA, [7]Global Monitoring Laboratory, Earth System Research Laboratories, National Oceanic and Atmospheric Administration, Boulder, CO, USA, [8]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Key Points:

- Space-time variability in North American carbon balance is better resolved in models informed by remote sensing inputs than otherwise
- Bottom-up models with strong growing-season carbon uptake in croplands are more consistent with atmospheric CO_2 observations
- Most terrestrial biosphere models misrepresent the timing of peak net carbon uptake in croplands

Supporting Information:

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

Correspondence to:

W. Sun,
wsun@carnegiescience.edu

Citation:

Sun, W., Fang, Y., Luo, X., Shiga, Y. P., Zhang, Y., Andrews, A. E., et al. (2021). Midwest US croplands determine model divergence in North American carbon fluxes. *AGU Advances*, *2*, e2020AV000310. https://doi.org/10.1029/2020AV000310

Received 10 NOV 2020
Accepted 14 APR 2021

Author Contributions:

Conceptualization: Wu Sun, Yuanyuan Fang, Anna M. Michalak
Data curation: Wu Sun, Yuanyuan Fang, Xiangzhong Luo, Yoichi P. Shiga, Yao Zhang, Arlyn E. Andrews, Kirk W. Thoning, Joshua B. Fisher, Trevor F. Keenan
Formal analysis: Wu Sun, Yuanyuan Fang, Anna M. Michalak
Funding acquisition: Joshua B. Fisher, Trevor F. Keenan, Anna M. Michalak
Investigation: Wu Sun, Yuanyuan Fang, Xiangzhong Luo, Yoichi P. Shiga,

Abstract Large uncertainties in North American terrestrial carbon fluxes hinder regional climate projections. Terrestrial biosphere models (TBMs), the essential tools for understanding continental-scale carbon cycle, diverge on whether temperate forests or croplands dominate carbon uptake in North America. Evidence from novel photosynthetic proxies, such as those based on chlorophyll fluorescence, has cast doubt on the "weak cropland, strong forest" carbon uptake patterns simulated by most TBMs. However, no systematic evaluation of TBMs has yet been attempted to pin down space-time patterns that are most consistent with regional CO_2 observational constraints. Here, we leverage atmospheric CO_2 observations and satellite-observed photosynthetic proxies to understand emergent space-time patterns in North American carbon fluxes from a large suite of TBMs and data-driven models. To do so, we evaluate how well the atmospheric signals resulting from carbon flux estimates reproduce the space-time variability in atmospheric CO_2, as is observed by a network of continuous-monitoring towers over North America. Models with gross or net carbon fluxes that are consistent with the observed CO_2 variability share a salient feature of growing-season carbon uptake in Midwest US croplands. Conversely, the remaining models place most growing-season uptake in boreal or temperate forests. Differences in model explanatory power depend mainly on the simulated annual cycles of cropland uptake—especially, the timing of peak uptake—rather than the distribution of annual mean fluxes across biomes. Our results suggest that improved model representation of cropland phenology is crucial to robust, policy-relevant estimation of North American carbon exchange.

Plain Language Summary In North America, land ecosystems have been known to act as a net sink of carbon, but how carbon fluxes are distributed in space and time remains uncertain. A key unresolved question about space-time patterns of North American carbon exchange is whether croplands or temperate forests show the strongest uptake rate during the growing season. In this study, we evaluate a large suite of land biosphere models and models that are driven by remotely sensed vegetation indices or ground-based flux observations to pin down the space-time patterns in North American carbon exchange. Models are assessed based on how well their carbon flux estimates capture the variability in the atmospheric CO_2 record observed from a network of towers. We find that models with carbon flux estimates that reproduce the observed CO_2 variability well show a hotspot of strong growing-season carbon uptake in the Midwest US croplands. In contrast, models that reproduce the observed atmospheric CO_2 variability less effectively place the strongest growing-season uptake in forests. Our findings reveal that peak cropland carbon uptake is underestimated in most models and that a better representation of cropland processes will be needed to obtain robust estimates of North American carbon fluxes.

1. Introduction

Terrestrial ecosystems in North America act as a strong carbon sink to mitigate human-induced carbon emissions (Hayes et al., 2018). Despite extensive observations, major uncertainties abound in understanding

Yao Zhang, Joshua B. Fisher, Trevor F. Keenan, Anna M. Michalak
Methodology: Wu Sun, Yuanyuan Fang, Xiangzhong Luo, Yoichi P. Shiga, Yao Zhang, Arlyn E. Andrews, Kirk W. Thoning, Joshua B. Fisher, Trevor F. Keenan, Anna M. Michalak
Project Administration: Joshua B. Fisher, Trevor F. Keenan, Anna M. Michalak
Resources: Anna M. Michalak
Supervision: Joshua B. Fisher, Trevor F. Keenan, Anna M. Michalak
Writing – review & editing: Wu Sun, Yuanyuan Fang, Xiangzhong Luo, Yoichi P. Shiga, Yao Zhang, Joshua B. Fisher, Trevor F. Keenan, Anna M. Michalak

the space-time variability of North American terrestrial carbon fluxes (Fang et al., 2014; Hilton et al., 2017; Huntzinger et al., 2012; King et al., 2015; Normile, 2017). Bottom-up modeling approaches to quantifying carbon fluxes disagree on how gross and net carbon fluxes are distributed seasonally and across North American biomes. Notably, most state-of-the-art models favor strong uptake in temperate forests and weaker uptake in croplands (Anav et al., 2015; Huntzinger et al., 2012; Normile, 2017; Zhou et al., 2020), while a small subset of models simulate otherwise. There is also widespread divergence in the timing and amplitude of seasonal cycles in major biomes across models (Normile, 2017; Richardson et al., 2012; Schwalm et al., 2010).

Multiple lines of observational evidence have cast doubt on the "strong forest, weak cropland" carbon uptake patterns in North America, however. Although croplands are typically not a major long-term carbon sink due to harvesting that exports carbon offsite and the respiration of residues (Ciais et al., 2010; Wolf et al., 2015), they are known to be highly productive ecosystems in terms of growing-season CO_2 uptake (Schulze et al., 2010). In addition, the outsized influence of cropland intensification on the increasing CO_2 seasonal amplitude in the Northern hemisphere (Gray et al., 2014; N. Zeng et al., 2014) appears to be incompatible with the weak cropland uptake in most model simulations. Boreal ecosystems are also implicated in the increased seasonal amplitude, however, based on model estimates of carbon fluxes (Forkel et al., 2016; Graven et al., 2013; Piao et al., 2018). Furthermore, evidence from novel photosynthetic proxies, such as chlorophyll fluorescence and carbonyl sulfide (COS), suggests that the most prominent growing-season photosynthesis occurs not in temperate forests, but in croplands (Guanter et al., 2014; Hilton et al., 2017). Atmospheric CO_2 inversions also show strong growing-season net uptake in mid-continental croplands (Cui et al., 2021; Peters et al., 2007; Schuh et al., 2013; Shiga et al., 2018).

Amid an abundance of observational constraints that underscore the widespread divergence in carbon flux estimates across models, a robust understanding of the space-time variability in North American carbon fluxes remains elusive. For example, how the magnitude and spatial patterns of cropland carbon uptake compare to those of temperate forests remains uncertain. It is also unclear to what extent the discrepancies in the distribution of gross primary productivity (GPP), as informed by photosynthetic proxies, are representative of the space-time variability of net ecosystem exchange (NEE). A systematic model evaluation using regional-scale observational constraints could inform the divergence in the space-time patterns of North American carbon fluxes across models.

A variety of models exist for estimating carbon fluxes at regional scales. Terrestrial biosphere models (TBMs) simulate carbon fluxes based on mechanistically resolved carbon exchange and vegetation phenology (Bonan, 2019; Fisher et al., 2014), and are essential tools for forecasting the state of the terrestrial carbon cycle (Bonan & Doney, 2018). In contrast to TBMs, data-driven models simulate fluxes based on vegetation phenology represented by remotely sensed indices—for example, leaf area index (LAI) and enhanced vegetation index (EVI). For data-driven models, the relationship between vegetation indices and carbon fluxes can be empirically parameterized, as in remote sensing models, or learned implicitly through algorithms, as in machine learning-based models (Beer et al., 2010). Machine learning-based models are typically trained to upscale plot-scale fluxes from eddy covariance observations to regional scales (e.g., Jung et al., 2020).

Divergence among model estimates of carbon fluxes has persisted through model generations (Friedlingstein et al., 2006, 2014), despite increasingly sophisticated mechanistic representations of ecosystem and hydrologic processes (Prentice et al., 2015), as well as the wider use of observational constraints in parameterization (e.g., Bonan et al., 2012; MacBean et al., 2015) and benchmarking (Collier et al., 2018; Luo et al., 2012). Model evaluation to diagnose the causes of the divergence in simulated carbon flux components has mostly focused on comparing modeled values with reference values, and more recently, on functional relationships between the fluxes and drivers (Collier et al., 2018; Schwalm et al., 2019). These benchmarks help to isolate processes that contribute to the spread in model outcomes (e.g., response of GPP to LAI). However, due to process complexity (Bonan, 2019) and equifinality (Tang & Zhuang, 2008) in model parameterization, locally optimized process representations may not translate into more realistic overall outcomes in terms of the regional space-time variability in carbon fluxes (Prentice et al., 2015).

Model evaluation that targets regional-scale patterns of carbon fluxes is, therefore, needed to diagnose model divergence. Instead of comparing model estimates of carbon fluxes against plot-scale observations as in

a typical model benchmarking setting (Normile, 2017; Schwalm et al., 2010; Zhou et al., 2020) or with inversions aggregated over a continent or large subcontinental domains (Hayes et al., 2018; King et al., 2015), here, we aim to understand the dominant regional-scale features of variability that explain model divergence in the North American carbon balance within and across major biomes by leveraging the scale-relevant constraints from atmospheric and satellite observations.

Atmospheric CO_2 observations have increasingly been used to constrain regional-scale carbon balance and evaluate carbon flux estimates from models. Unlike eddy covariance flux measurements that focus at the plot scale (~1 km^2; Schuepp et al., 1990), observations of atmospheric CO_2 concentrations integrate flux influences on regional scales (from ~10^3 to 10^5 km^2) and can therefore, inform carbon flux variability at these larger scales (Gloor et al., 2001; Michalak et al., 2004). With transport models linking fluxes to concentrations, atmospheric CO_2 observations can also be used to evaluate the regional-scale variability in carbon fluxes simulated by TBMs (Fang & Michalak, 2015; Fang et al., 2014), or to provide inverse estimates of NEE independent of process-based assumptions (Gourdji et al., 2012; Michalak et al., 2004; Shiga et al., 2018) after careful consideration of relevant sources of uncertainty.

Satellite-observed photosynthetic proxies have routinely been used to constrain GPP. Absorbed photosynthetically active radiation (APAR) at the chlorophyll is the first-order driver of GPP (Monteith, 1972). A small fraction of the APAR is typically re-emitted as fluorescence, which can inform photosynthetic activity and canopy structure (Magney et al., 2019; Y. Zeng et al., 2019). Solar-induced chlorophyll fluorescence (SIF) observed from remote sensing platforms provides much-needed GPP constraints at regional and global scales (Frankenberg et al., 2011; Parazoo et al., 2014; Sun et al., 2017). Recently, SIF has been shown to capture the space-time variability of the NEE in North America better than most TBMs do (Shiga et al., 2018).

Here, we leverage atmospheric CO_2 observations and satellite-observed photosynthetic proxies to evaluate regional-scale carbon flux estimates from a large suite of TBMs and data-driven models. We assess how well the transported atmospheric signals of carbon flux estimates capture the space-time variability in atmospheric CO_2 concentrations, as observed by a network of continuous-monitoring towers over North America (Figure 1). Specifically, for each estimate of carbon flux or SIF, we evaluate how much variance in the observed biospheric CO_2 drawdown or enhancement is explained by the transported signal. Since the carbon flux and SIF estimates examined here are monthly averaged, we also look at the amount of observed atmospheric CO_2 variability that can be explained by the monthly averaged APAR and by the NEE estimated using a geostatistical inverse model (GIM NEE). These two pieces of information provide a context for interpreting the amount of observed atmospheric CO_2 variability that can be explained using the ensemble of TBMs and data-driven models examined here, because these models would be expected to perform at least as well as the APAR. On the other hand, the amount of observed atmospheric CO_2 variability that can be explained using inversion-derived fluxes indicates how much of this variability can, in principle, be explained despite the monthly resolution of the data sets and the presence of atmospheric transport model errors. To examine the proximate causes of differences in model performance, we examine how biome-level spatial distribution and seasonality affect the explanatory power of model estimates.

2. Methods

2.1. Atmospheric CO_2 Observations

Atmospheric CO_2 observations were obtained from the ObsPack CO_2 GLOBALVIEWplus v3.2 data product (Cooperative Global Atmospheric Data Integration Project, 2017; Masarie et al., 2014). We used in situ three-hourly averaged afternoon CO_2 measurements from 44 continuous-monitoring towers across North America during 2007–2010 (see Text S6 in the Supporting Information [SI] for an additional analysis using 2007–2012 measurements), with averaging centered at 15:00 local time for most sites, and 16:00 or 17:00 for the remaining few sites. Afternoon observations were chosen, as is common practice in regional inversions (e.g., Gourdji et al., 2012), because transport model errors arising from model representations of planetary boundary layer dynamics are expected to be lower in the afternoons relative to other times of the day (Lin et al., 2017). Urban sites were excluded. We filtered out (a) extreme outliers, (b) data that exceeded the background values by more than 30 ppmv (Fang & Michalak, 2015), (c) data overly sensitive to ocean fluxes

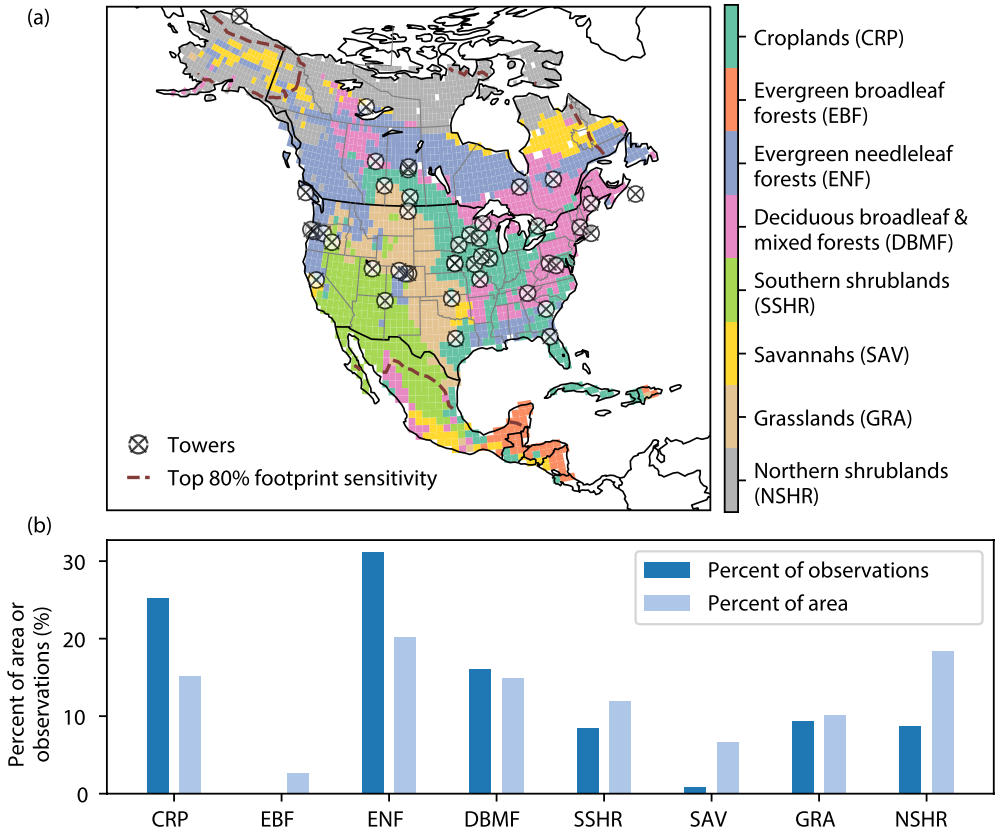

Figure 1. (a) A map of North American biomes with sites of the CO_2 observational network, after Shiga et al. (2018). Enclosed within the dashed dark red contours are grid cells of which the total footprint sensitivity summed during 2007–2010 is among the top 80% over continental North America. (b) Percent coverage of observations (dark blue) sensitive to biomes during 2007–2010, compared with the area percentages of these biomes (light blue). Croplands (CRP), evergreen needleleaf forests (ENF), and deciduous broadleaf and mixed forests (DBMF) were the most extensively observed biomes. In contrast, savannahs (SAV) and tropical evergreen broadleaf forests (EBF) received little coverage in the observing system. Other biomes were underrepresented in terms of observational coverage.

(Gourdji et al., 2012), and (d) those susceptible to transport errors (Gourdji et al., 2012; Shiga et al., 2014). Of the approximately 57,700 available mid-afternoon observations during 2007–2010, approximately 18,400 (32%) were filtered out (Table S1). The remaining ~39,300 observations were then processed into biospheric drawdown or enhancement signals by subtracting background values and fossil fuel enhancements. Background CO_2 mixing ratios were calculated by sampling the endpoints of the back trajectories from an empirical background, similar to that of S. Jeong et al. (2013) and the "EBG" method in Hu et al. (2019). Next, fossil fuel CO_2 contributions were estimated from the Fossil Fuel Data Assimilation System (FFDAS v2; Asefi-Najafabady et al., 2014). The FFDAS v2 data product, originally at the hourly, 0.1° × 0.1° resolution, was upscaled to a three-hourly, 1° × 1° resolution, and transported using the transport footprints (Section 2.2) to derive fossil fuel CO_2 enhancements for subtraction. Overall, CO_2 data processing was consistent with that in Fang et al. (2014) and Shiga et al. (2018). See Table S2 in the SI for tower coordinates, heights, and data providers.

2.2. Transport Footprints

The sensitivities of CO_2 drawdowns or enhancements to surface carbon fluxes in the upwind regions (ppmv $[\mu mol\, m^{-2} s^{-1}]^{-1}$), also known as "footprints," were quantified from simulations of the Weather Research and Forecasting (WRF) model (Skamarock & Klemp, 2008) coupled with the Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al., 2003; Nehrkorn et al., 2010), with three-hourly temporal resolution and 1° × 1° spatial resolution over the North American domain. These footprints were created as part

of the NOAA CarbonTracker-Lagrange regional inverse modeling framework (https://www.esrl.noaa.gov/gmd/ccgg/carbontracker-lagrange/). We represented the footprints in a matrix form to link CO_2 variation with surface fluxes. The footprints used here covered the period of 2007–2010. The WRF-STILT model has been used extensively to estimate greenhouse gas fluxes (Gourdji et al., 2010, 2012; Hu et al., 2019; S. Jeong et al., 2013; Miller et al., 2013; Shiga et al., 2018) and to evaluate TBM estimates of ecosystem CO_2 fluxes (Fang & Michalak, 2015; Fang et al., 2014) at regional and continental scales.

2.3. Model Estimates of Terrestrial Carbon Fluxes and Satellite-Observed Photosynthetic Proxies

We evaluated a diverse set of model estimates of the NEE and GPP and satellite-observed photosynthetic proxies over North America. Although the net biospheric CO_2 drawdown or enhancement reflects the NEE not GPP, we postulated that modeled GPP and NEE along with remotely sensed photosynthetic proxies explain a portion of the space-time variability of the "true unknown NEE." This premise holds if (a) model estimates of GPP and photosynthetic proxies capture the variability in the true GPP, and (b) the true GPP explains a substantial fraction of variability in the NEE. The plausibility of these requirements is supported by the fact that SIF as a GPP proxy performs well as the leading explanatory variable of the NEE (Shiga et al., 2018).

Models to be evaluated included eight remote sensing models, three machine-learning models from the FluxCom ensemble, and 29 prognostic TBMs sourced from two widely cited model intercomparison projects—the Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP, version 2; Huntzinger et al., 2013, 2017, 2018, 2020; Wei et al., 2014a, 2014b) and the Trends in Net Land-Atmosphere Exchange (TRENDY, version 6; Le Quéré et al., 2018; Sitch et al., 2008, 2015). Four SIF data products were also included as proxies to represent space-time patterns in carbon fluxes for comparison against model estimates of carbon fluxes (GOME-2A SIF [Joiner et al., 2013, 2014, 2016], spatially and temporally kriged GOME-2A SIF [Tadić et al., 2017], RSIF [Gentine & Alemohammad, 2018], and CSIF [Zhang et al., 2018]). Overall, this yielded $N = 82$ data sets to be evaluated.

We also used the APAR (Text S2), calculated from the MODIS fAPAR (Myneni, 2020; Myneni et al., 2002) and the PAR derived from the North American Regional Reanalysis (Mesinger et al., 2006), and the monthly GIM NEE (Text S3; Shiga et al., 2018) as the lower and upper bounds on the amount of the observed atmospheric CO_2 variability that we would expect bottom-up models and GPP data sets to capture. All data products of carbon fluxes and photosynthetic proxies were standardized to monthly time scales and $1° \times 1°$ spatial resolution for a fair comparison, although their native resolutions in time and space could be higher. We also used the three-hourly GIM NEE as a counterfactual reference to show what the upper bound of explanatory power would be, had the evaluated variables been available at three-hourly instead of monthly resolution. A detailed description of the data products being evaluated can be found in Text S1–S4 and Table S3.

2.4. Model Benchmarking

Benchmarking of simulated carbon fluxes and SIF data products was posed as a regression problem (Shiga et al., 2018). Using the footprints derived from the WRF-STILT model, the space-time patterns of carbon fluxes or photosynthetic proxies can be translated into effects on atmospheric CO_2:

$$\mathbf{z} = \mathbf{HX}\boldsymbol{\beta} + \epsilon \tag{1}$$

where $\mathbf{z} \in \mathbb{R}^{n\times 1}$ is a column vector of n atmospheric CO_2 observations, $\mathbf{H} \in \mathbb{R}^{n\times m}$ is the footprint matrix that represents a linearization of the transport model (WRF-STILT), with m being the total number of discretized grid cells in space *and* in time, $\mathbf{X} \in \mathbb{R}^{m\times p}$ is the design matrix that consists of a constant term and an explanatory variable that are both discretized in space and time (here, $p = 2$), $\boldsymbol{\beta} \in \mathbb{R}^{p\times 1}$ is a column vector of linear coefficients that links the transported signals of the explanatory variable ($\mathbf{HX}$) to observed CO_2 concentrations (z), and $\epsilon \in \mathbb{R}^{m\times 1}$ is the total error that incorporates errors in observation, atmospheric transport, data representation, and bias due to the portion of fluxes not explained by the explanatory variable. We assumed that the total error $\boldsymbol{\varepsilon}$ would follow a Gaussian error model.

The performance of one or more explanatory variables (X) relative to the benchmark of observed atmospheric CO_2 concentrations (z) is measured here by the coefficient of determination (R^2) of the linear regression z∼HX. The R^2 metric is deemed a model's explanatory power. The confidence intervals around R^2 values were calculated from bootstrapping 1,000 times (i.e., resampling observations with replacement to recompute R^2 values). For a detailed description of the model benchmarking, see Text S5.

Given the oversampling of observations sensitive to croplands with respect to other biomes (Figure 1), we also conducted resampling tests with subsets of observations balanced among the six biomes with enough observations (excluding evergreen broadleaf forests and savannahs; Text S7) and found that the rank of models by explanatory power (R^2) did not differ from that derived while using all the observations (Figure S4).

In addition, we performed regressions to test the influences of biome-level flux distribution and seasonality on model explanatory power by encoding biome, month, or the biome–month combination as categorical predictors based on biomes as defined in Figure 1a. With the coefficients $\hat{\boldsymbol{\beta}}$ estimated from the regression $\mathbf{z}\sim\mathbf{HX}$, we then calculated the predicted NEE estimates: $\hat{\mathbf{s}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. Similar regression-adjusted NEE estimates were also calculated for biome and/or month-encoded regressions. A detailed description for the regressions encoded with biomes and/or months can be found in Text S10.

To further attribute differences in model performance to the (mis)representation of the seasonal cycle in specific major biomes, we also performed single-biome encoded regressions. These regressions tested how adjusting the seasonal cycles individually in each of the three major biomes—croplands, evergreen needleleaf forests, and deciduous broadleaf and mixed forests (Figure 1)—would increase model explanatory power (R^2). The single-biome encoded regressions are described in Text S11.

2.5. Analysis of Common Patterns

Principal component analysis (PCA) was used to extract the first principal component (PC1) flux patterns of model groups that were categorized based on R^2 skills. For example, thw NEE models that scored an R^2 value higher than the R^2 of the APAR were grouped as "high-R^2 NEE models." For each model group, we aligned fluxes of different models from different years as column vectors, while preserving spatial and monthly indices. This means that for the same model, simulated fluxes in 2007 were treated as a different predictor than the simulated fluxes in 2008. All months, from January to December, were included. This approach was used to extract the dominant mode of seasonally varying spatial patterns shared across different models and also across different years. Due to the short period of investigation (2007–2010), we have focused on the seasonality of geographic patterns in fluxes rather than on the trends in carbon uptake across years. The calculation was performed using the Scikit-learn package (Pedregosa et al., 2011). A detailed description is presented in Text S8, and Figures S7–S10 present monthly PC1 patterns.

3. Results and Discussion

3.1. Fidelity to Observed Atmospheric CO_2 Benchmarks

The fractions of space-time variability in CO_2 data explained by monthly carbon flux and SIF variables, expressed as R^2 values, range from low to moderate (0.03–0.45, Figure 2). Surprisingly, many models explain less of the observed CO_2 variability than does the APAR ($R^2 = 0.33$). The overall low R^2 values of the evaluated models and data sets are in part attributable to their monthly resolution, because the best model estimates of the NEE still achieve explanatory powers that are about 80% of the explanatory power of monthly averaged GIM NEE ($R^2 = 0.53$). This is further supported by the gap in the explanatory power between three-hourly and monthly GIM NEE estimates ($R^2 = 0.84$ vs. 0.53). It is beyond the scope of this study, however, to quantify how sub-monthly variability in carbon fluxes and SIF would affect other models' explanatory power, because most of the evaluated models and data sets are not available at diurnal or sub-diurnal resolution.

GPP estimates from remote sensing models, GPP and NEE estimates from FluxCom models, and SIF data products all explain broadly similar fractions of space-time variability in atmospheric CO_2 observations, which are close to or slightly above the fraction of variability explained by the APAR (Figure 2). The APAR as a GPP driver explains 33% of variability in atmospheric CO_2 drawdown, indicating that it correlates with

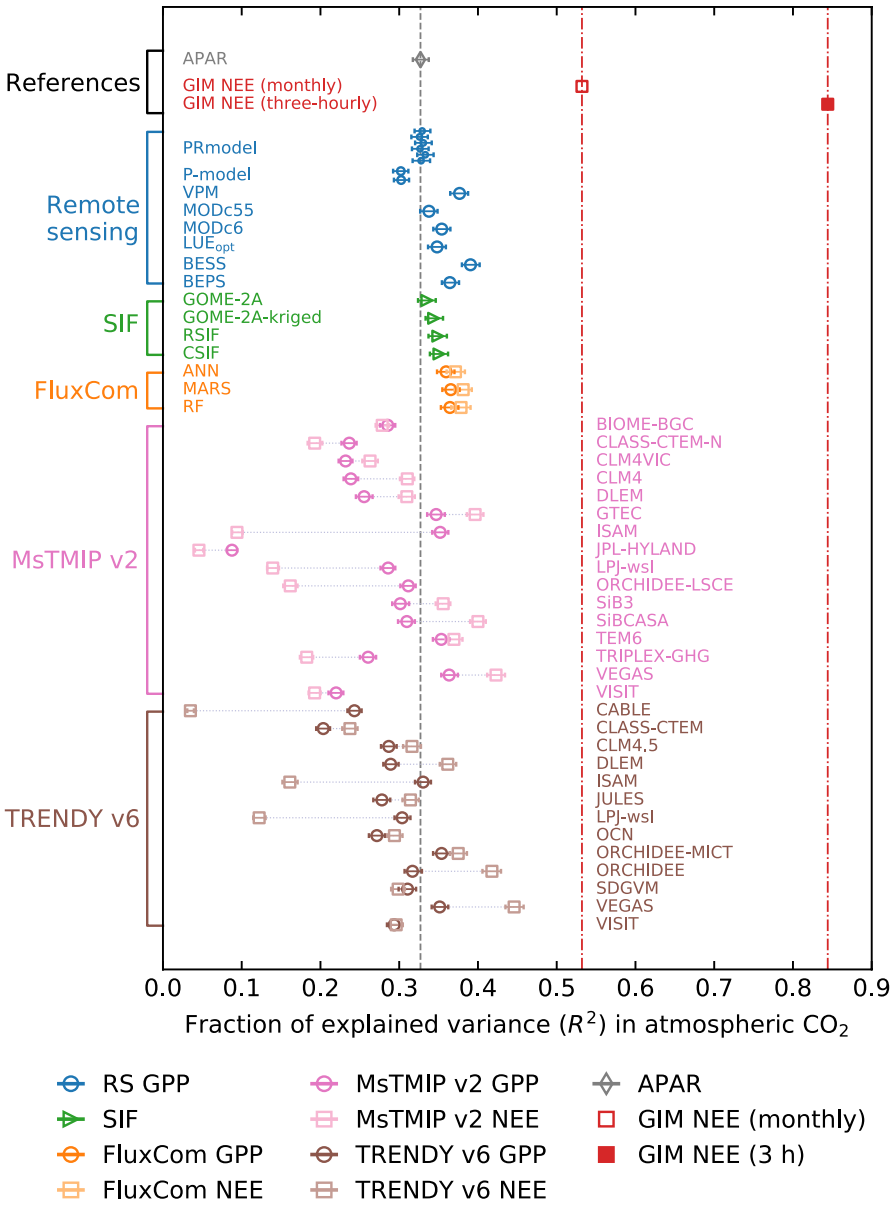

Figure 2. Fraction of variance (R^2) in atmospheric CO_2 explained by modeled carbon fluxes (GPP and NEE), SIF data products, and the APAR during 2007–2010, with error bars showing 95% confidence intervals obtained from bootstrapping. Model groups are color-coded and indicated in the legend. For each model with both GPP and NEE estimates, data points that denote R^2 values of GPP and NEE of that model are linked by a thin dashed line. The vertical lines denote R^2 values of three references (from left to right): APAR and monthly and three-hourly geostatistical inverse estimates of NEE (GIM NEE). APAR, absorbed photosynthetically active radiation; GIM, geostatistical inverse model; GPP, gross primary productivity; NEE, net ecosystem exchange; SIF, solar-induced chlorophyll fluorescence.

regional patterns of the NEE. Underlying this result is the fact that the GPP explains a substantial amount of variability in the NEE (Baldocchi, 2008; Janssens et al., 2001). Since SIF is a proxy of the GPP (Frankenberg et al., 2011; Guanter et al., 2014) and contains information about the APAR (Yang et al., 2015; Y. Zeng et al., 2019), it is not surprising to find similar explanatory powers among all four SIF data products (R^2 = 0.34–0.35; Figure 2), which are slightly above that of the APAR (R^2 = 0.33).

Despite different parameterizations of photosynthesis (Table S3)—namely, light use efficiency (e.g., MODIS) versus enzyme kinetics (e.g., BESS) approaches—GPP estimates from remote sensing models show broadly similar explanatory powers (R^2 = 0.31–0.39; Figure 2). Since remote sensing models use vegetation indices

to represent phenology and calculate the APAR, the similar explanatory powers of their GPP estimates reflect similar space-time patterns in key model inputs (e.g., fAPAR and EVI). This further indicates that, among remote sensing models, model explanatory power is more sensitive to inputs than to parameterizations. For example, multiple GPP estimates from the PRmodel (Keenan et al., 2016), generated by perturbing the aridity limitation function or from null models (Table S4), all explain virtually the same amount of variability in atmospheric CO_2 observations as does the APAR (Figure 2).

We find no substantial difference in the explanatory power between data-driven models that are trained or calibrated with flux tower observations—FluxCom, P-model, and LUE_{opt}—and those that are not directly informed by flux tower observations (the rest of remote sensing models; Figure 2). The explanatory power of the LUE_{opt} model ($R^2 = 0.35$) is similar to those of the MODc55 ($R^2 = 0.34$) and MODc6 ($R^2 = 0.35$). The FluxCom GPP and NEE estimates show explanatory powers ($R^2 = 0.36$–0.37 for GPP estimates, and $R^2 = 0.37$–0.38 for NEE estimates) comparable to GPP estimates from remote sensing models. Such similarity is, perhaps, not surprising given the relative scarcity of flux towers in some biomes, for example, croplands. Some models, such as the P-model, in fact, use no cropland sites in their calibration (Stocker et al., 2020). In addition, the small footprint of flux towers (~1 km^2) limits the degree to which flux tower observations inform regional patterns.

It is, however, unexpected that FluxCom NEE estimates do not explain a significantly higher portion of the observed CO_2 variability than the FluxCom GPP estimates. This may be caused by the fact that though FluxCom NEE estimates were trained with tower NEE observations, the upscaling used the same set of remote sensing inputs as those for the GPP (Tramontana et al., 2016) and thereby, preserved the space-time features in the input variables. Overall, the similar explanatory powers of the GPP and/or NEE estimates from remote sensing and machine learning models likely resulted from similar space-time patterns in the remotely sensed inputs.

In contrast to remote sensing and machine learning models, GPP and NEE estimates from prognostic models (MsTMIP and TRENDY ensembles) diverge strongly in how much variability in CO_2 drawdown they explain ($R^2 = 0.09$–0.36 for GPP estimates, and $R^2 = 0.03$–0.45 for NEE estimates; Figure 2). Most GPP and NEE estimates from TBMs explain less variability in CO_2 drawdown than the APAR alone, indicating that the prognostic phenology modules adopted by these models do not reproduce the vegetation dynamics seen by satellite remote sensing platforms.

These results highlight the strong divergence in the space-time patterns of the GPP and NEE estimates from prognostic models, despite the fact that climate drivers are kept consistent within each model ensemble (Huntzinger et al., 2013; Sitch et al., 2015). Functional benchmarks have revealed that model prediction skill for the GPP and NEE depends on process parameterizations, such as nitrogen limitation and light thresholds for photosynthesis, and initial conditions of carbon fluxes and pools (Schwalm et al., 2019). Compared with diagnostic models and machine learning models that prescribe phenology using remotely sensed inputs, prognostic models simulate phenology based on their representations of underlying processes. Errors in simulated phenology, coupled with shortcomings in the functions that map phenology to carbon fluxes (e.g., LAI–GPP relationship), can cause greater divergence in the explanatory power of the space-time patterns of model estimates of the GPP and NEE.

Given that prognostic TBMs are routinely compared with flux observations at the plot scale (Normile, 2017; Schaefer et al., 2012; Stöckli et al., 2008) and with inverse flux estimates at continental or sub-continental scales (Hayes et al., 2018; King et al., 2015), but less often with observations that inform intermediate regional-scale fluxes (but see Guanter et al., 2014; Hilton et al., 2017), the lack of constraints on regional-scale flux patterns may also contribute to model divergence. Our evaluation of a large suite of TBMs, data-driven models, and photosynthetic proxies suggests widespread misrepresentation of space-time variability in carbon fluxes among TBMs, given that most data-driven models that are informed by regional-scale constraints explain atmospheric CO_2 variability relatively well (Figure 2). To understand the proximate causes of misrepresented space-time variability in regional carbon flux estimates, we look into emergent features in carbon flux patterns that determine model divergence.

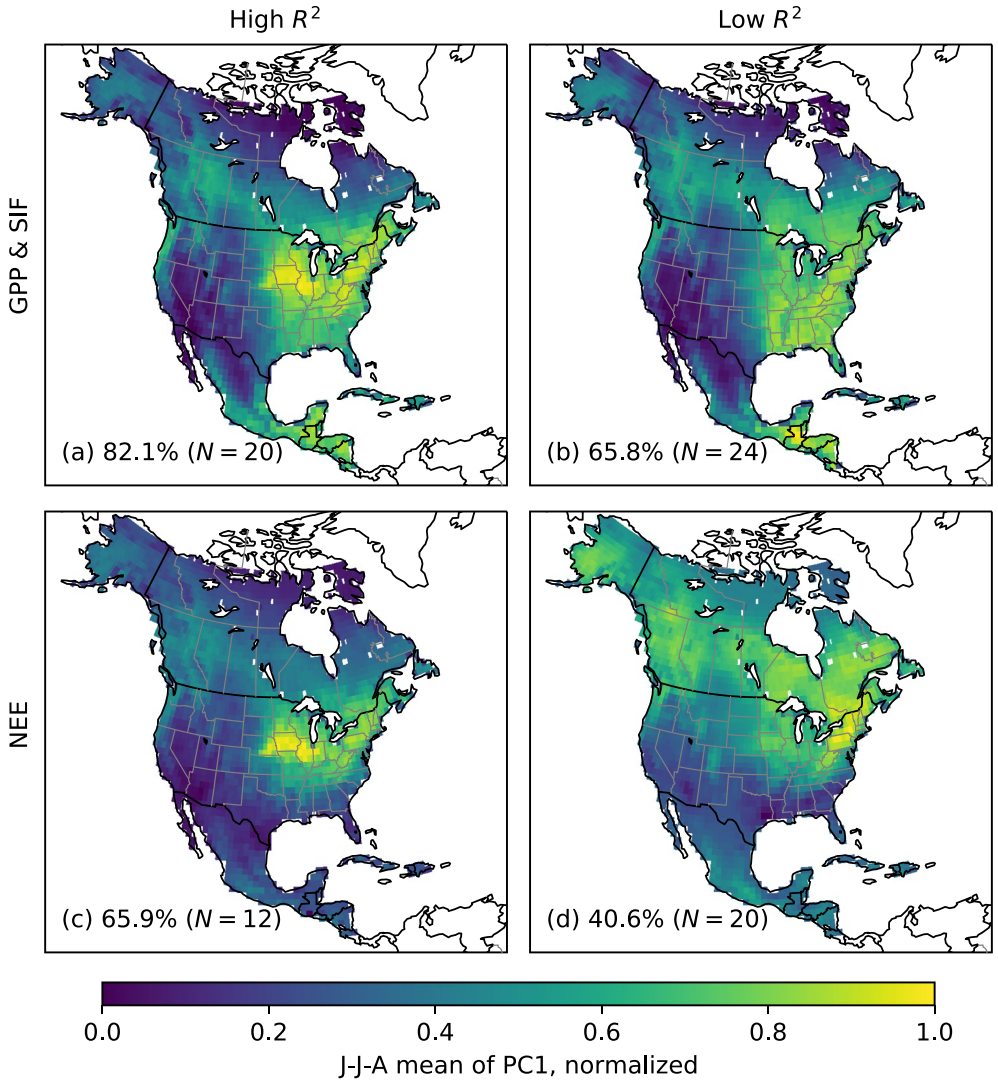

Figure 3. The June–July–August (JJA) mean patterns of the extracted first principal component (PC1) among (a) GPP and SIF models of high R^2 (0.330, 0.391), (b) GPP models of low R^2 (0.088, 0.317), (c) NEE models of high R^2 (0.356, 0.446), and (d) NEE models of low R^2 (0.035, 0.316), where the numbers in parentheses represent the range of R^2 within a group. All PC1 patterns are in normalized unit for visual comparison. Percentage numbers represent the fraction of variance in each group explained by the PC1, whereas N indicates the number of models in each group. Models of high R^2 and of low R^2 are separated using the R^2 value of APAR as a threshold. Note that the PRmodel ensemble members are not included in either GPP group (a and b), because they are similar to each other and to APAR. For both GPP and SIF (a) and NEE (c), models with high R^2 show common features of strong summer uptake in the mid-continental cropland region. For the PC1 patterns across a full year, see Figures S7–S10. APAR, absorbed photosynthetically active radiation; GPP, gross primary productivity; NEE, net ecosystem exchange; SIF, solar-induced chlorophyll fluorescence.

3.2. Models With Strong Growing-Season Carbon Uptake in Croplands Are More Consistent With Atmospheric CO_2 Observations

We find that the GPP and SIF estimates that outperform the APAR in explaining CO_2 variability (number of models $N = 20$) show a common pattern of strong growing-season photosynthetic activity in croplands (Figure 3a). Note that all SIF variables belong to this group with high explanatory power (Figure 2). In contrast, the GPP estimates with lower explanatory power than the APAR ($N = 24$) place the growing-season photosynthetic activities mainly in the eastern deciduous forests and the northern evergreen needleleaf forests (Figure 3b). The fact that the first principal component (PC1) of the group of models with high explanatory power explains a high fraction (82%) of variance in the GPP across both models and years (Figure 3a)

indicates that this is a robust pattern that emerges independently of model structure. The six PRmodel members were not included in either group because their explanatory power was very similar to that of the APAR (Figure 2) and had highly correlated flux patterns among each other (Figures S11 and S12).

Similarly, the NEE estimates with higher explanatory power than the APAR ($N = 12$) show the strongest growing-season net uptake in croplands (Figure 3c), and this strong feature of net uptake is also absent in the NEE estimates with weaker explanatory power than the APAR ($N = 20$; Figure 3d). The PC1 extracted from the former group also explains a high fraction of inter-model and interannual variability (~66%) in the NEE.

These results reveal an emergent relationship that links the overall model performance in North America to how models simulate the growing-season carbon uptake in mid-continental croplands. Although the geographic extent of specific biomes may vary somewhat among models, the pronounced uptake features observed in Figures 3a and 3c correspond well with the extent of croplands, as defined in the current analysis (Figure 1). This finding is corroborated by other independent regional-scale constraints on the GPP only, namely, COS and SIF, both of which suggest strong growing-season photosynthesis in mid-continental croplands that is only captured by a minority of models (Guanter et al., 2014; Hilton et al., 2017; Shiga et al., 2018). Furthermore, the inclusion of a large suite of models and photosynthetic proxies in the model evaluation supports the robustness of this feature, as they are representative of a diverse (though by no means exhaustive) model space. In other words, the large ensemble of models examined here, together with the regional-scale constraints from atmospheric and satellite observations point to a robust finding of strong growing-season net carbon uptake in the Midwest US croplands relative to forested regions.

This continental-scale feature of strong cropland uptake also resonates with existing studies that find far-reaching global impacts of cropland productivity on the seasonal amplitudes of CO_2 in the Northern hemisphere (Gray et al., 2014; N. Zeng et al., 2014). On the other hand, a subset of the TBMs examined here were also used in attributing the increased CO_2 seasonal amplitudes to boreal ecosystems (Forkel et al., 2016; Graven et al., 2013; Piao et al., 2018), but many of them do not explain the observed CO_2 variability over North America well (Figure 2) and lack the strong cropland uptake that is more consistent with atmospheric observations (Figure 3). Therefore, the attribution of the observed increase of the seasonal amplitude of CO_2 may warrant further examination using a subset of models that capture the salient feature of cropland carbon uptake, at least for the North American domain.

3.3. Most Models Misrepresent Seasonality of Cropland Carbon Uptake

Among the models with GPP or NEE estimates with lower explanatory power, the lack of strong cropland carbon uptake could in principle be attributed to misrepresentation of either (a) the overall magnitude of the cropland GPP or NEE relative to other major biomes or (b) the seasonal cycle of fluxes *within* the cropland biome. If the misrepresentation of flux magnitudes among biomes is the major cause, we would expect that allowing linear coefficients ($\boldsymbol{\beta}$) in the regressions for model evaluation to vary by biome would lead to large increases in model explanatory power. Similarly, if the misrepresentation of seasonal cycles of fluxes dominates, allowing linear coefficients ($\boldsymbol{\beta}$) in the regressions for model evaluation to vary by month would lead to a relatively large increase in model explanatory power (see Text S10 for implementation details).

Based on biome and month encoded regressions conducted as described above and in Section 2.4, we find that the differences in model performance are mainly attributable to the models' ability to represent the seasonal cycle of fluxes across biomes, and only to a lesser extent to differences in the ability to represent spatial differences in fluxes across biomes (Figure 4). When the linear coefficients ($\boldsymbol{\beta}$) are allowed to vary by biome, R^2 values show only marginal improvements across all GPP, SIF (Figures 4a and 4b), and NEE estimates (Figures 4e and 4f). In contrast, allowing $\boldsymbol{\beta}$ to vary by month leads to a substantial improvement in R^2 values, to the point that the gap between the models with the highest and lowest explanatory powers is greatly reduced within each model group (Figures 4c and 4g). Furthermore, letting $\boldsymbol{\beta}$ vary both by biome and by month virtually eliminates the difference in performance across models (Figures 4d and 4h).

Hence, we find that the spatial variability within biomes is not a major contributor to the differences in models' ability to explain observed atmospheric CO_2 variability. This is further corroborated by the result that even a null model that has spatially constant fluxes within biomes (but for which fluxes can vary

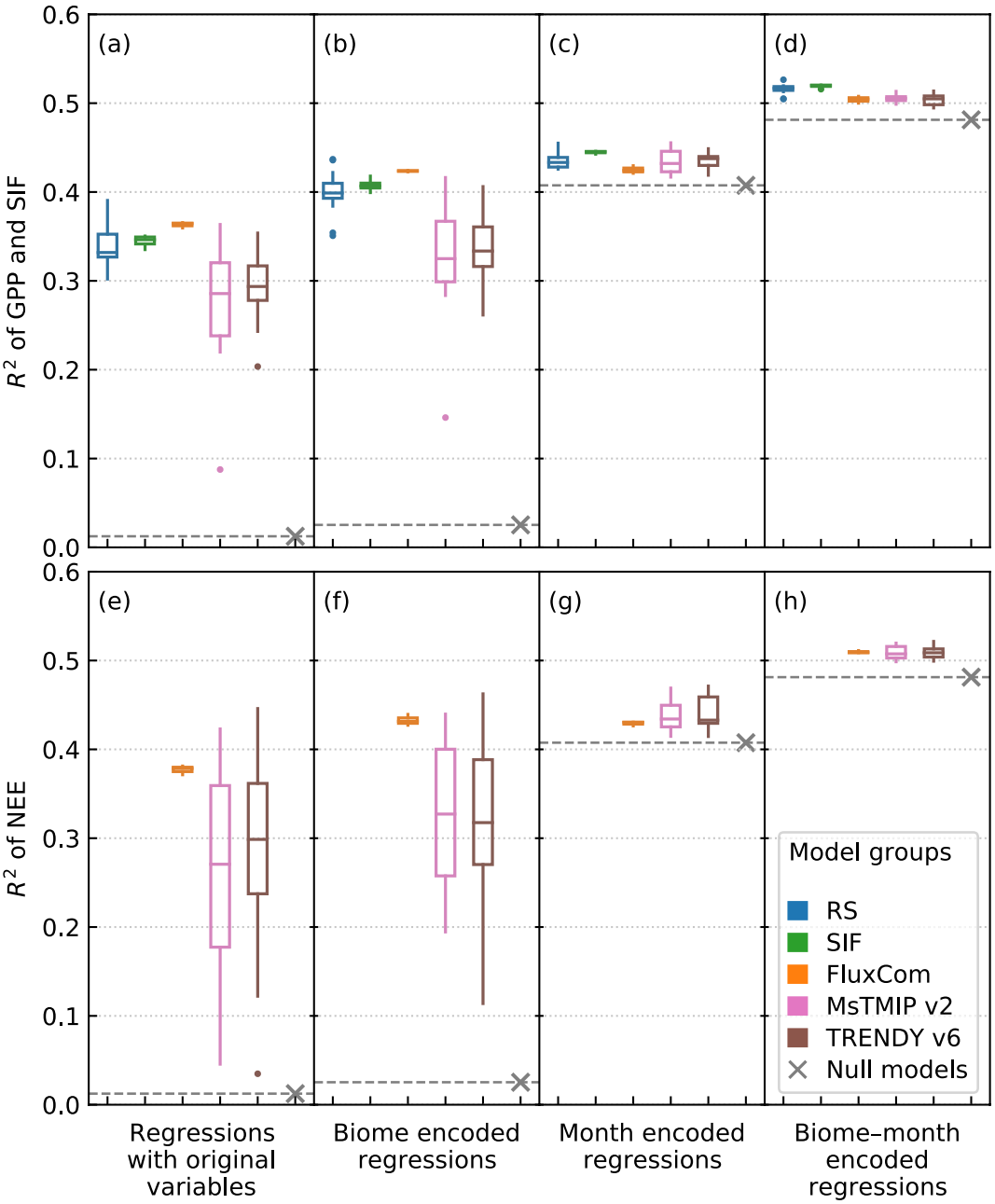

Figure 4. Increase in model explanatory power (R^2) of (a–d) GPP models and SIF data products and of (e–h) NEE models after encoding biome and/or month as categorical predictors, shown as boxplots. From left to right are: R^2 values from (a and e) regressions using model estimates alone ($\mathbf{z}\sim\mathbf{HX}$), (b and f) regressions encoded with the biome variable ($\mathbf{z}\sim\mathbf{HX}_{\text{biome}}$), (c and g) regressions encoded with the month variable ($\mathbf{z}\sim\mathbf{HX}_{\text{month}}$), and (d and h) regressions encoded with the Cartesian product of biome and month variables ($\mathbf{z}\sim\mathbf{HX}_{\text{biome}\times\text{month}}$). Note that R^2 values of the corresponding null models are shown as gray crosses for reference. GPP, gross primary productivity; SIF, solar-induced chlorophyll fluorescence.

between biomes and between months) can explain almost as much space-time variability in atmospheric CO_2 drawdown as the other models (Figures 4d and 4h).

The regression-adjusted cropland NEE estimates further confirm that the observed atmospheric CO_2 variability is sensitive to the annual seasonal cycles of cropland carbon fluxes, and in particular, the timing of peak growing-season uptake (Figures 4g and 5), rather than the magnitude of peak uptake or the

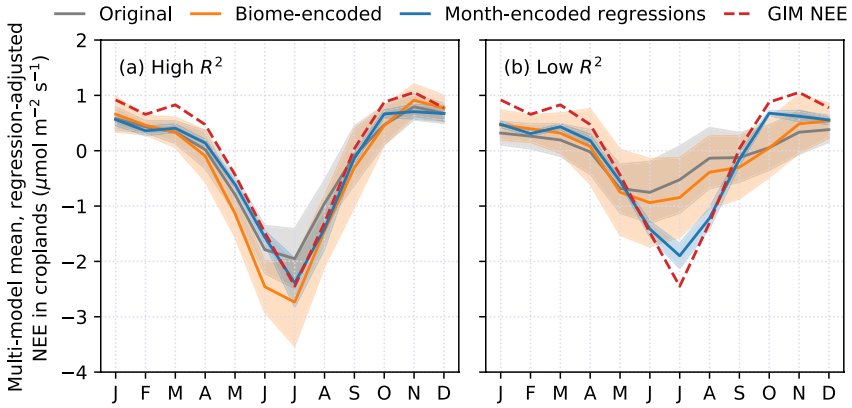

Figure 5. Multi-model mean, regression-adjusted NEE seasonal cycles (2007–2010) in croplands. For each explanatory variable, we calculate the predicted NEE estimates ($\hat{\mathbf{s}} = \mathbf{X}\hat{\boldsymbol{\beta}}$) from regressions with the original explanatory variable matrix $\mathbf{X}$ (gray lines), the biome-encoded explanatory variable matrix $\mathbf{X}_{\text{biome}}$ (orange lines), and the month-encoded explanatory variable matrix $\mathbf{X}_{\text{month}}$ (blue lines). The seasonal cycles are averaged by model groups with multi-model standard deviations ($\pm 1\sigma$) shown as the shading: (a) NEE models of high R^2, and (b) NEE models of low R^2 (these groups correspond exactly to those in Figures 3c and 3d, respectively). The mean seasonal cycles of GIM NEE over croplands are shown in red dashed lines for comparison. GIM, geostatistical inverse model; NEE, net ecosystem exchange.

distribution of flux magnitudes across biomes (Figures 4f and 5). For the NEE variables with low explanatory power (Figure 5b), month-encoded regressions attempt to adjust NEE seasonal cycles to be more similar to those of the NEE estimates with high explanatory power (Figure 5a), leading to a shift in the timing of peak uptake from June to July among the former (Figure 5b). In contrast, biome-encoded regressions are unable to fix the misaligned seasonal cycles of cropland NEE in this group of models. This feature is even more striking in the regression-adjusted NEE seasonal cycles of some representative individual models (Figures S13 and S14). For the NEE variables with high explanatory power, biome- and month-encoded regressions do not noticeably change the cropland NEE seasonal cycle with respect to the regressions with original explanatory variables (Figure 5a), which suggests that cropland NEE estimates from these models are robust in terms of both the overall magnitude relative to that of temperate forests and seasonal cycles. The widespread and substantial bias in the timing of peak uptake in croplands that we find here is consistent with the findings from plot-scale model evaluations at sub-monthly resolutions, for example, the 1- to 2-month lead of simulated phenology of corn and soybean in CLM4 over observed phenology at Midwestern cropland sites (Chen et al., 2015, 2018).

The dominant contribution of croplands to overall model performance is also corroborated by additional lines of evidence. First, single-biome encoded regressions (i.e., regressions that adjust the seasonal cycles in one biome only; Text S11 and Figure S16) applied to croplands (Figures S16d and S16h) increase explanatory power among prognostic models more than when applied to evergreen needleleaf forests (Figures S16c and S16g) or deciduous broadleaf and mixed forests (Figures S16b and S16f), and substantially reduce the variability in explanatory power across models (Figures S16d and S16h relative to Figures S16b and S16f and Figures S16c and S16g). Second, the increases in model explanatory power are similar when adjusting the seasonal cycles in croplands alone (Figures S16d and S16h vs. Figures S16a and S16e) as when uniformly adjusting the seasonal cycles in all biomes (Figures 4c and 4g vs. Figures 4a and 4e).

To eliminate the possibility that this dominant role is due to differences between model-specific delineations of cropland area and the uniform biome map used here (Figure 1a), we repeated the single-biome encoded regressions using model-specific biome distributions for a subset of TRENDY models for which such an analysis was possible. Results from four models (CABLE, OCN, ORCHIDEE-MICT, and VISIT) included information on plant functional types at a resolution of $1° \times 1°$ or higher. We find that adjusting the seasonal cycles of carbon uptake in model-specific cropland areas (or in the absence of a specific cropland type, the biome that most closely matches the geographic extent of croplands) yields results consistent with those shown in Figures S16e and S16h (results not shown), confirming that model differences in the delineation of cropland areas are not a driver of the observed effect.

Overall, these results confirm that the seasonal cycle of cropland carbon uptake—in particular, the timing of peak uptake, also known as the "phase" (Collier et al., 2018)—plays a dominant role in determining model performance. The mechanistic cause of the phase bias in the seasonal cycles of cropland carbon uptake appears to be rooted in vegetation phenology, as further analysis established a robust link between the explanatory power of model GPP estimates and how well model LAI estimates capture the variability in the remotely sensed MODIS LAI product over croplands, but not other major biomes, among MsTMIP v2 models (Figure S18).

Several causes may underlie the dominant role of cropland carbon uptake in the divergence of North American carbon flux estimates. First, prognostic TBMs rely on parameterizations of vegetation phenology that are specific to each plant functional type (PFT), whereas data-driven models can leverage the spatial and seasonal patterns in remotely sensed vegetation indices to capture regional-scale variability in carbon fluxes, even without an explicit representation of crops and cropland management activities. The lack of PFTs to resolve the dominant crops in North America—corn, soybean, and wheat—may prevent TBMs (e.g., BIOME-BGC and DLEM) from simulating the dominant feature of a strong mid-continental cropland uptake. Although MsTMIP enforced a standard biome map among its participating models, some models may not have accurate parameterization of physiological parameters that control carbon fluxes for crop PFTs (e.g., $V_{c,max}$, the maximum carboxylation activity for photosynthesis). Second, except for a few models equipped with advanced cropland modules that resolve management practices (e.g., irrigation, planting, harvesting, etc.), many TBMs (or at least their standard, non-crop-specific simulations included in model ensembles) are not designed to simulate croplands. Third, cropland management can be heavily affected by extreme weather events (e.g., 2019 Midwest flooding that delayed the growing season; Yin et al., 2020) that are not anticipated by even the most sophisticated models. While other deficiencies, such as the response of photosynthesis to temperature, have been implicated in the phenology bias in boreal ecosystems (e.g., S.-J. Jeong et al., 2017), their effects are difficult to isolate for croplands, because the poor representation of management practices appears to be a primary factor of model underperformance.

Overall, our results indicate an urgent need to bring about a more realistic representation of the seasonal cycles of cropland carbon uptake to prognostic models for better understanding of space-time patterns in North American carbon fluxes. Recent advances have shown promise in reducing phenology biases through improved mechanistic representation of leaf area index dynamics (Wu et al., 2016), the implementation of active management processes (Lombardozzi et al., 2020), crop-specific parameterization (Boas et al., 2021; Peng et al., 2018), and/or intensive calibration with flux tower data (Bilionis et al., 2015; Chen et al., 2018). With concerted model development, calibration, and evaluation, the next-generation of TBMs will undoubtedly be better positioned to deliver more robust estimation of North American carbon fluxes.

4. Conclusion

Using atmospheric CO_2 observations as regional-scale benchmarks on North American carbon exchange, we link the performance of carbon flux estimates and SIF to underlying regional patterns. We find that models with flux estimates that reproduce the observed CO_2 variability well all share pronounced growing-season uptake in the mid-continental croplands, and this feature is either absent or has an incorrect timing in the remaining models. Although croplands are typically not a long-term carbon sink, seasonal carbon uptake in croplands emerges as a salient regional-scale feature from a wide range of model estimates and photosynthetic proxies. Through regression analyses, we attribute this difference in model performance mainly to the models' ability to represent the seasonality of fluxes across and within biomes, and to a lesser degree to the biome-level spatial differences in fluxes.

These findings echo a long-standing challenge in model development, namely, the challenge of capturing the seasonality of carbon fluxes. To reduce the divergence in model estimates of the North American carbon exchange, future development needs to target phenology and carbon allocation, because these processes dictate the seasonality of cropland fluxes. More broadly, model benchmarking with atmospheric CO_2 observations must be streamlined as a routine practice to detect and diagnose discrepancies at regional to continental scales. Looking ahead, reducing model divergence in regional carbon fluxes is a crucial step

toward obtaining accurate and policy-relevant climate projections, and this process requires coordinated and integrated efforts in model design and benchmarking.

Conflict of Interest

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

Data Availability Statement

Model evaluation results presented in this study are available at https://gitlab.com/wusun/na-carbon-flux-eval, under a CC-BY 4.0 License. Geostatistical inverse estimates of NEE are available upon request addressed to Yoichi P. Shiga. All other data sets used in this study are available online and/or from original data contributors, as listed in Table S5.

Acknowledgments
This study was funded by the NASA Terrestrial Ecology Interdisciplinary Science (IDS) Award no. NNH17AE86I. Joshua B. Fisher contributed to this research at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. California Institute of Technology. Government sponsorship acknowledged. The authors thank all modelers and investigators who contributed to the Multi-scale synthesis and Terrestrial Model Intercomparison Project (MsTMIP; http://nacp.ornl.gov/MsTMIP.shtml): Deborah Huntzinger, M. Altaf Arain, Dominique Bachelet, Gwenaëlle Berthier, Robert Cook, Josh Fisher, Robert Grant, Dan Hayes, Maoyi Huang, Akihiko Ito, Atul Jain, Andy Jacobson, Jiafu Mao, Francesc Montane, Weile Wang, Yaxing Wei, Changhui Peng, Shushi Peng, Benjamin Poulter, Daniel Ricciuto, Kevin Schaefer, Christopher Schwalm, Hanqin Tian, and Ning Zeng. Funding for the MsTMIP activity was provided through NASA ROSES Grant no. NNX10AG01A. Data management support for preparing, documenting, and distributing model driver and output data was performed by the Modeling and Synthesis Thematic Data Center at Oak Ridge National Laboratory (ORNL; http://nacp.ornl.gov), with funding through NASA ROSES Grant no. NNH10AN681. Finalized MsTMIP data products are archived at the ORNL DAAC (http://daac.ornl.gov). The authors thank Stephen Sitch, Pierre Friedlingstein, and all modelers who have contributed to the Trends in Net Land–Atmosphere Exchange project (TRENDY; http://sites.exeter.ac.uk/trendy). The authors thank Benjamin Stocker for P-model simulations, Youngryel Ryu for BESS simulations, Jing-Ming Chen and Weimin Ju for BEPS simulations, Steven Running and Maosheng Zhao for MODIS GPP data (collections 5.5 and 6), Nima Madani

References

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., et al. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*(3), 785–818. https://doi.org/10.1002/2015RG000483

Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., et al. (2014). A multiyear, global gridded fossil fuel CO_2 emission data product: Evaluation and analysis of results. *Journal of Geophysical Research: Atmospheres*, *119*(17), 10213–10231. https://doi.org/10.1002/2013JD021296

Baldocchi, D. (2008). 'Breathing' of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*, *56*(1), 1–26. https://doi.org/10.1071/BT07151

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., et al. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Bilionis, I., Drewniak, B. A., & Constantinescu, E. M. (2015). Crop physiology calibration in the CLM. *Geoscientific Model Development*, *8*(4), 1071–1083. https://doi.org/10.5194/gmd-8-1071-2015. Retrieved from https://gmd.copernicus.org/articles/8/1071/2015/

Boas, T., Bogena, H., Grünwald, T., Heinesch, B., Ryu, D., Schmidt, M., et al. (2021). Improving the representation of cropland sites in the Community Land Model (CLM) version 5.0. *Geoscientific Model Development*, *14*(1), 573–601. https://doi.org/10.5194/gmd-14-573-2021. Retrieved from https://gmd.copernicus.org/articles/14/573/2021/

Bonan, G. B. (2019). Terrestrial biosphere models. In *Climate change and terrestrial ecosystem modeling* (pp. 1–24). Cambridge University Press. https://doi.org/10.1017/9781107339217.002

Bonan, G. B., & Doney, S. C. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, *359*(6375), eaam8328. https://doi.org/10.1126/science.aam8328

Bonan, G. B., Oleson, K. W., Fisher, R. A., Lasslop, G., & Reichstein, M. (2012). Reconciling leaf physiological traits and canopy flux data: Use of the TRY and FLUXNET databases in the Community Land Model version 4. *Journal of Geophysical Research*, *117*(G2). https://doi.org/10.1029/2011JG001913

Chen, M., Griffis, T. J., Baker, J. M., Wood, J. D., Meyers, T., & Suyker, A. (2018). Comparing crop growth and carbon budgets simulated across AmeriFlux agricultural sites using the Community Land Model (CLM). *Agricultural and Forest Meteorology*, *256–257*, 315–333. https://doi.org/10.1016/j.agrformet.2018.03.012. Retrieved from http://www.sciencedirect.com/science/article/pii/S0168192318300959

Chen, M., Griffis, T. J., Baker, J. M., Wood, J. D., & Xiao, K. (2015). Simulating crop phenology in the Community Land Model and its impact on energy and carbon fluxes. *Journal of Geophysical Research: Biogeosciences*, *120*(2), 310–325. https://doi.org/10.1002/2014JG002780

Ciais, P., Wattenbach, M., Vuichard, N., Smith, P., Piao, S. L., Don, A., et al. (2010). The European carbon balance. Part 2: Croplands. *Global Change Biology*, *16*(5), 1409–1428. https://doi.org/10.1111/j.1365-2486.2009.02055.x

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The International Land Model Benchmarking (ILAMB) system: Design, theory, and implementation. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2731–2754. https://doi.org/10.1029/2018MS001354

Cooperative Global Atmospheric Data Integration Project. (2017). *Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957–2016; obspack_co2_1_globalviewplus_v3.2_2017-11-02*. NOAA Earth System Research Laboratory, Global Monitoring Division. https://doi.org/10.15138/G3704H. Retrieved from http://www.esrl.noaa.gov/gmd/ccgg/obspack/data.php?id=obspack_co2_1_GLOBALVIEWplus_v3.2_2017-11-02

Cui, Y. Y., Jacobson, A. R., Feng, S., Wesloh, D., Gerken, T., Barkley, Z. R., et al. (2021). *Evaluation of inverse estimates of North American net ecosystem exchange of CO_2 from different observing systems using ACT-America airborne observations*. ESSOAr. https://doi.org/10.1002/essoar.10505569.1

Fang, Y., & Michalak, A. M. (2015). Atmospheric observations inform CO_2 flux responses to enviroclimatic drivers. *Global Biogeochemical Cycles*, *29*(5), 555–566. https://doi.org/10.1002/2014GB005034

Fang, Y., Michalak, A. M., Shiga, Y. P., & Yadav, V. (2014). Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models. *Biogeosciences*, *11*(23), 6985–6997. https://doi.org/10.5194/bg-11-6985-2014

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

Forkel, M., Carvalhais, N., Rödenbeck, C., Keeling, R., Heimann, M., Thonicke, K., et al. (2016). Enhanced seasonal CO_2 exchange caused by amplified plant productivity in northern ecosystems. *Science*, *351*(6274), 696–699. https://doi.org/10.1126/science.aac4971. Retrieved from https://science.sciencemag.org/content/351/6274/696

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., et al. (2011). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, *38*(17). https://doi.org/10.1029/2011GL048738

for the optimum LUE derived GPP data, the FluxCom Initiative for FluxCom data, Ranga Myneni for MODIS LAI and FPAR data, Joanna Joiner for GOME-2A SIF data, Jovan Tadić for the kriged GOME-2A SIF data, and Pierre Gentine for CSIF and RSIF data. The authors thank Kevin Gurney for the FFDAS v2 data set. The authors acknowledge the NCEP North American Regional Reanalysis data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, obtained from their website at https://psl.noaa.gov/. The authors acknowledge the National Oceanic and Atmospheric Administration Earth System Research Laboratories (NOAA/ESRL) and site principal investigators for providing atmospheric CO_2 observations. The authors thank the following individuals for collecting and providing the atmospheric CO_2 data from the following sites: Arlyn Andrews (NOAA) for AMT, BAO, LEF, WBI, and WKT; Arlyn Andrews (NOAA) and Marc L. Fischer (LBNL) for WGC; Arlyn Andrews (NOAA) and Matt J. Parker (SRNL) for SCT; Arlyn Andrews (NOAA) and Stephan De Wekker (UVA) for SNP; Sebastien Biraud (LBNL) and Margaret Torn (LBNL) for SGP; Tim Griffis (UMN) for KCMP; Dan Jaffe (UW) and Arlyn Andrews (NOAA) for MBO; Beverly Law (Oregon State University) and the TERRA-PNW group for data from the seven Oregon sites, OFR, OMP, OMT, ONG, OSI, OWA, OYQ; Natasha Miles (PSU), Scott Richardson (PSU), and Ken Davis (PSU) for AAC, ACR, ACV, AME, AOZ, INX01, INX09, FPK, RCE, RGV, RKW, RMM, and RRL; Britton Stephens (NCAR) and the Regional Atmospheric Continuous CO_2 Network in the Rocky Mountains (RACCOON) for HDP, NWR, RBA, and SPL; Colm Sweeney (NOAA) for MVY; Colm Sweeney (NOAA), Arlyn Andrews (NOAA), John B. Miller (NOAA), and Kathryn McKain (NOAA) for CRV; Kirk Thoning (NOAA) and Pieter Tans (NOAA) for BRW; Steven Wofsy (Harvard) and J. William Munger (Harvard) for HFM; Doug Worthy (Environment Canada) for BCK, BRA, CBY, CDL, CHL, CHM, CPS, EGB, ESP, EST, ETL, FSD, INU, LLB, TPD, and WSA. Measurements at WGC were partially supported by grants from the California Energy Commission (CEC) Public Interest Environmental Research Program to the Lawrence Berkeley National Laboratory, which operates under US Department of Energy under contract DE-AC02-05CH11231. The authors thank the Atmospheric and Environmental Research, Inc. (AER)—in particular, Thomas Nehrkorn, John Henderson, and Jenusz Eluszkiewicz—for conducting WRF-STILT simulations and providing transport footprints. The authors thank the CarbonTracker-La-

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., et al. (2006). Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *Journal of Climate*, *19*(14), 3337–3353. https://doi.org/10.1175/JCLI3800.1

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*(2), 511–526. https://doi.org/10.1175/JCLI-D-12-00579.1

Gentine, P., & Alemohammad, S. H. (2018). Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. *Geophysical Research Letters*, *45*(7), 3136–3146. https://doi.org/10.1002/2017GL076294

Gloor, M., Bakwin, P., Hurst, D., Lock, L., Draxler, R., & Tans, P. (2001). What is the concentration footprint of a tall tower? *Journal of Geophysical Research*, *106*(D16), 17831–17840. https://doi.org/10.1029/2001JD900021

Gourdji, S. M., Hirsch, A. I., Mueller, K. L., Yadav, V., Andrews, A. E., & Michalak, A. M. (2010). Regional-scale geostatistical inverse modeling of North American CO2 fluxes: A synthetic data study. *Atmospheric Chemistry and Physics*, *10*(13), 6151–6167. https://doi.org/10.5194/acp-10-6151-2010. Retrieved from https://acp.copernicus.org/articles/10/6151/2010/

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., et al. (2012). North American CO_2 exchange: Inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion. *Biogeosciences*, *9*(1), 457–475. https://doi.org/10.5194/bg-9-457-2012

Graven, H. D., Keeling, R. F., Piper, S. C., Patra, P. K., Stephens, B. B., Wofsy, S. C., et al. (2013). Enhanced seasonal exchange of CO_2 by northern ecosystems since 1960. *Science*, *341*(6150), 1085–1089. https://doi.org/10.1126/science.1239207. Retrieved from https://science.sciencemag.org/content/341/6150/1085

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

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., et al. (2014). Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(14), E1327–E1333. https://doi.org/10.1073/pnas.1320008111. Retrieved from https://www.pnas.org/content/111/14/E1327

Hayes, D. J., Vargas, R., Alin, S., Conant, R. T., Hutyra, L. R., Jacobson, A. R., et al. (2018). Chapter 2: The North American carbon budget. In N. Cavallaro, et al. (Eds.), *Second state of the carbon cycle report (SOCCR2): A sustained assessment report* (pp. 71–108). U.S. Global Change Research Program. https://doi.org/10.7930/SOCCR2.2018.Ch2

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

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). https://doi.org/10.1126/sciadv.aaw0076. Retrieved from https://advances.sciencemag.org/content/5/6/eaaw0076

Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., et al. (2017). Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-03818-2

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

Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., et al. (2013). The North American carbon program multi-scale synthesis and terrestrial model intercomparison project—Part 1: Overview and experimental design. *Geoscientific Model Development*, *6*(6), 2121–2133. https://doi.org/10.5194/gmd-6-2121-2013. Retrieved from https://www.geosci-model-dev.net/6/2121/2013/

Huntzinger, D. N., Schwalm, C. R., Wei, Y., Cook, R. B., Michalak, A. M., Schaefer, K., et al. (2018). *NACP MsTMIP: Global 0.5-degree model outputs in standard format, version 1.0*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1225. Retrieved from https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1225

Huntzinger, D. N., Schwalm, C. R., Wei, Y., Shrestha, R., Cook, R. B., Michalak, A. M., et al. (2020). *NACP MsTMIP: Global 0.5-degree model outputs in standard format, version 2.0*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1599. Retrieved from https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1599

Janssens, I. A., Lankreijer, H., Matteucci, G., Kowalski, A. S., Buchmann, N., Epron, D., et al. (2001). Productivity overshadows temperature in determining soil and ecosystem respiration across European forests. *Global Change Biology*, *7*(3), 269–278. https://doi.org/10.1046/j.1365-2486.2001.00412.x

Jeong, S., Hsu, Y.-K., Andrews, A. E., Bianco, L., Vaca, P., Wilczak, J. M., & Fischer, M. L. (2013). A multitower measurement network estimate of California's methane emissions. *Journal of Geophysical Research: Atmospheres*, *118*(19), 339–11. https://doi.org/10.1002/jgrd.50854

Jeong, S.-J., Schimel, D., Frankenberg, C., Drewry, D. T., Fisher, J. B., Verma, M., et al. (2017). Application of satellite solar-induced chlorophyll fluorescence to understanding large-scale variations in vegetation phenology and function over northern high latitude forests. *Remote Sensing of Environment*, *190*, 178–187. https://doi.org/10.1016/j.rse.2016.11.021. Retrieved from http://www.sciencedirect.com/science/article/pii/S0034425716304680

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., et al. (2013). Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. *Atmospheric Measurement Techniques*, *6*(10), 2803–2823. https://doi.org/10.5194/amt-6-2803-2013

Joiner, J., Yoshida, Y., Guanter, L., & Middleton, E. M. (2016). New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: Simulations and application to GOME-2 and SCIAMACHY. *Atmospheric Measurement Techniques*, *9*(8), 3939–3967. https://doi.org/10.5194/amt-9-3939-2016. Retrieved from https://www.atmos-meas-tech.net/9/3939/2016/

Joiner, J., Yoshida, Y., Vasilkov, A. P., Schaefer, K., Jung, M., Guanter, L., et al. (2014). The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange. *Remote Sensing of Environment*, *152*, 375–391. https://doi.org/10.1016/j.rse.2014.06.022

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-17-1343-2020. Retrieved from https://www.biogeosciences.net/17/1343/2020/

Keenan, T. F., Prentice, I. C., Canadell, J. G., Williams, C. A., Wang, H., Raupach, M., & Collatz, G. J. (2016). Recent pause in the growth rate of atmospheric CO_2 due to enhanced terrestrial carbon uptake. *Nature Communications*, *7*(1). https://doi.org/10.1038/ncomms13428

grange project team for providing the WRF-STILT transport footprints. The authors thank Nina Randazzo for creating the North American land mask. Computation performed in this study was supported by the Scientific Computing Committee for High-Performance Computing at the Carnegie Institution for Science (https://hpc.carnegiescience.edu). The authors thank Floyd Fayton, Garret Huntress, and Maria Lopez for their help in computational work. The authors thank Leslie Willoughby for editorial assistance. The authors thank Martin Heimann and an anonymous reviewer for thoughtful input that helped to improve this work. This study is a contribution to the North American Carbon Program (https://www.nacarbon.org).

King, A. W., Andres, R. J., Davis, K. J., Hafer, M., Hayes, D. J., Huntzinger, D. N., et al. (2015). North America's net terrestrial CO_2 exchange with the atmosphere 1990-2009. *Biogeosciences*, *12*(2), 399–414. https://doi.org/10.5194/bg-12-399-2015. Retrieved from https://bg.copernicus.org/articles/12/399/2015/

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., et al. (2018). Global carbon budget 2017. *Earth System Science Data*, *10*(1), 405–448. https://doi.org/10.5194/essd-10-405-2018. Retrieved from https://www.earth-syst-sci-data.net/10/405/2018/

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*, *108*(D16). https://doi.org/10.1029/2002jd003161

Lin, J. C., Mallia, D. V., Wu, D., & Stephens, B. B. (2017). How can mountaintop CO_2 observations be used to constrain regional carbon fluxes? *Atmospheric Chemistry and Physics*, *17*(9), 5561–5581. https://doi.org/10.5194/acp-17-5561-2017. Retrieved from https://acp.copernicus.org/articles/17/5561/2017/

Lombardozzi, D. L., Lu, Y., Lawrence, P. J., Lawrence, D. M., Swenson, S., Oleson, K. W., et al. (2020). Simulating agriculture in the Community Land Model version 5. *Journal of Geophysical Research: Biogeosciences*, *125*(8), e2019JG005529. https://doi.org/10.1029/2019JG005529

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., et al. (2012). A framework for benchmarking land models. *Biogeosciences*, *9*(10), 3857–3874. https://doi.org/10.5194/bg-9-3857-2012

MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., & Ciais, P. (2015). Using satellite data to improve the leaf phenology of a global terrestrial biosphere model. *Biogeosciences*, *12*(23), 7185–7208. https://doi.org/10.5194/bg-12-7185-2015

Magney, T. S., Bowling, D. R., Logan, B. A., Grossmann, K., Stutz, J., Blanken, P. D., et al. (2019). Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11640–11645.Retrieved from https://www.pnas.org/content/116/24/11640

Masarie, K. A., Peters, W., Jacobson, A. R., & Tans, P. P. (2014). ObsPack: A framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements. *Earth System Science Data*, *6*(2), 375–384. https://doi.org/10.5194/essd-6-375-2014

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., et al. (2006). North American regional reanalysis. *Bulletin of the American Meteorological Society*, *87*(3), 343–360. https://doi.org/10.1175/BAMS-87-3-343

Michalak, A. M., Bruhwiler, L., & Tans, P. P. (2004). A geostatistical approach to surface flux estimation of atmospheric trace gases. *Journal of Geophysical Research*, *109*(D14). https://doi.org/10.1029/2003JD004422

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., et al. (2013). Anthropogenic emissions of methane in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50), 20018–20022. https://doi.org/10.1073/pnas.1314392110. Retrieved from https://www.pnas.org/content/110/50/20018

Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, *9*(3), 747–766. https://doi.org/10.2307/2401901

Myneni, R. B. (2020). *MODIS Collection 6 (C6) LAI/FPAR product user's guide* (Technical Report). Retrieved from https://lpdaac.usgs.gov/documents/624/MOD15_User_Guide_V6.pdf

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., et al. (2002). Global products of vegetation leaf area and fraction absorbed par from year one of MODIS data. *Remote Sensing of Environment*, *83*(1–2), 214–231. https://doi.org/10.1016/S0034-4257(02)00074-3

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., & Freitas, S. (2010). Coupled weather research and forecasting-stochastic time-inverted lagrangian transport (WRF-STILT) model. *Meteorology and Atmospheric Physics*, *107*(1–2), 51–64. https://doi.org/10.1007/s00703-010-0068-x

Normile, C. P. (2017). *Assessment of uncertainties in atmospheric transport and surface flux of carbon from the North American terrestrial biosphere* (Doctoral dissertation). The Pennsylvania State University. https://doi.org/10.1126/science.aan7192. Retrieved from https://etda.libraries.psu.edu/files/final_submissions/16086

Parazoo, N. C., Bowman, K., Fisher, J. B., Frankenberg, C., Jones, D. B. A., Cescatti, A., et al. (2014). Terrestrial gross primary production inferred from satellite fluorescence and vegetation models. *Global Change Biology*, *20*(10), 3103–3121. https://doi.org/10.1111/gcb.12652

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peng, B., Guan, K., Chen, M., Lawrence, D. M., Pokhrel, Y., Suyker, A., et al. (2018). Improving maize growth processes in the community land model: Implementation and evaluation. *Agricultural and Forest Meteorology*, *250–251*, 64–89. https://doi.org/10.1016/j.agrformet.2017.11.012. Retrieved from http://www.sciencedirect.com/science/article/pii/S0168192317303854

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., et al. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(48), 18925–18930. https://doi.org/10.1073/pnas.0708986104. Retrieved from https://www.pnas.org/content/104/48/18925

Piao, S., Liu, Z., Wang, Y., Ciais, P., Yao, Y., Peng, S., et al. (2018). On the causes of trends in the seasonal amplitude of atmospheric CO_2. *Global Change Biology*, *24*(2), 608–616. https://doi.org/10.1111/gcb.13909

Prentice, I. C., Liang, X., Medlyn, B. E., & Wang, Y.-P. (2015). Reliable, robust and realistic: The three R's of next-generation land-surface modelling. *Atmospheric Chemistry and Physics*, *15*(10), 5987–6005. https://doi.org/10.5194/acp-15-5987-2015

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., et al. (2012). Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program site synthesis. *Global Change Biology*, *18*(2), 566–584. https://doi.org/10.1111/j.1365-2486.2011.02562.x

Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., et al. (2012). A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research*, *117*(G3). https://doi.org/10.1029/2012jg001960

Schuepp, P. H., Leclerc, M. Y., MacPherson, J. I., & Desjardins, R. L. (1990). Footprint prediction of scalar fluxes from analytical solutions of the diffusion equation. *Boundary-Layer Meteorology*, *50*(1–4), 355–373. https://doi.org/10.1007/BF00120530

Schuh, A. E., Lauvaux, T., West, T. O., Denning, A. S., Davis, K. J., Miles, N., et al. (2013). Evaluating atmospheric CO_2 inversions at multiple scales over a highly inventoried agricultural landscape. *Global Change Biology*, *19*(5), 1424–1439. https://doi.org/10.1111/gcb.12141

Schulze, E. D., Ciais, P., Luyssaert, S., Schrumpf, M., Janssens, I. A., Thiruchittampalam, B., et al. (2010). The European carbon balance. Part 4: Integration of carbon and other trace-gas fluxes. *Global Change Biology*, *16*(5), 1451–1469. https://doi.org/10.1111/j.1365-2486.2010.02215.x

Schwalm, C. R., Schaefer, K., Fisher, J. B., Huntzinger, D., Elshorbany, Y., Fang, Y., et al. (2019). Divergence in land surface modeling: Linking spread to structure. *Environmental Research Communications*, *1*(11), 111004. https://doi.org/10.1088/2515-7620/ab4a8a

Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., et al. (2010). A model-data intercomparison of CO_2 exchange across North America: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research*, *115*(G3), G00H05. https://doi.org/10.1029/2009JG001229

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 CO_2 measurements. *Geophysical Research Letters*, *41*(12), 4381–4388. https://doi.org/10.1002/2014GL059684

Shiga, Y. P., Tadić, J. M., Qiu, X., Yadav, V., Andrews, A. E., Berry, J. A., & Michalak, A. M. (2018). Atmospheric CO_2 observations reveal strong correlation between regional net biospheric carbon uptake and solar-induced chlorophyll fluorescence. *Geophysical Research Letters*, *45*(2), 1122–1132. https://doi.org/10.1002/2017GL076630

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., et al. (2015). Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, *12*(3), 653–679. https://doi.org/10.5194/bg-12-653-2015. Retrieved from https://www.biogeosciences.net/12/653/2015/

Sitch, S., Huntingford, C., Gedney, N., Levy, P. E., Lomas, M., Piao, S. L., et al. (2008). Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five dynamic global vegetation models (DGVMs). *Global Change Biology*, *14*(9), 2015–2039. https://doi.org/10.1111/j.1365-2486.2008.01626.x

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, *227*(7), 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037

Stocker, B. D., Wang, H., Smith, N. G., Harrison, S. P., Keenan, T. F., Sandoval, D., et al. (2020). P-model v1.0: An optimality-based light use efficiency model for simulating ecosystem gross primary production. *Geoscientific Model Development*, *13*(3), 1545–1581. https://doi.org/10.5194/gmd-13-1545-2020

Stöckli, R., Lawrence, D. M., Niu, G.-Y., Oleson, K. W., Thornton, P. E., Yang, Z.-L., et al. (2008). Use of FLUXNET in the community land model development. *Journal of Geophysical Research*, *113*(G1). https://doi.org/10.1029/2007jg000562

Sun, Y., Frankenberg, C., Wood, J. D., Schimel, D. S., Jung, M., Guanter, L., et al. (2017). OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science*, *358*(6360), eaam5747. https://doi.org/10.1126/science.aam5747. Retrieved from https://science.sciencemag.org/content/358/6360/eaam5747

Tadić, J. M., Qiu, X., Miller, S., & Michalak, A. M. (2017). Spatio-temporal approach to moving window block kriging of satellite data v1.0. *Geoscientific Model Development*, *10*(2), 709–720. https://doi.org/10.5194/gmd-10-709-2017

Tang, J., & Zhuang, Q. (2008). Equifinality in parameterization of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics. *Journal of Geophysical Research*, *113*(G4), G04010. https://doi.org/10.1029/2008JG000757

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016. Retrieved from https://www.biogeosciences.net/13/4291/2016/

Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., et al. (2014a). *NACP MsTMIP: Global and North American driver data for multi-model intercomparison*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1220. Retrieved from http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1220

Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., et al. (2014b). The North American carbon program multi-scale synthesis and terrestrial model intercomparison project - Part 2: Environmental driver data. *Geoscientific Model Development*, *7*(6), 2875–2893. https://doi.org/10.5194/gmd-7-2875-2014. Retrieved from https://www.geosci-model-dev.net/7/2875/2014/

Wolf, J., West, T. O., Le Page, Y., Kyle, G. P., Zhang, X., Collatz, G. J., & Imhoff, M. L. (2015). Biogenic carbon fluxes from global agricultural production and consumption. *Global Biogeochemical Cycles*, *29*(10), 1617–1639. https://doi.org/10.1002/2015GB005119

Wu, X., Vuichard, N., Ciais, P., Viovy, N., de Noblet-Ducoudré, N., Wang, X., et al. (2016). ORCHIDEE-CROP (v0), a new process-based agro-land surface model: Model description and evaluation over Europe. *Geoscientific Model Development*, *9*(2), 857–873. https://doi.org/10.5194/gmd-9-857-2016. Retrieved from https://gmd.copernicus.org/articles/9/857/2016

Yang, X., Tang, J., Mustard, J. F., Lee, J.-E., Rossini, M., Joiner, J., et al. (2015). Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. *Geophysical Research Letters*, *42*(8), 2977–2987. https://doi.org/10.1002/2015GL063201

Yin, Y., Byrne, B., Liu, J., Wennberg, P. O., Davis, K. J., Magney, T., et al. (2020). Cropland carbon uptake delayed and reduced by 2019 Midwest floods. *AGU Advances*, *1*(1), e2019AV000140. https://doi.org/10.1029/2019AV000140

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

Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., & Berry, J. A. (2019). A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence. *Remote Sensing of Environment*, *232*, 111209. https://doi.org/10.1016/j.rse.2019.05.028

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., & Gentine, P. (2018). A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences*, *15*(19), 5779–5800. https://doi.org/10.5194/bg-15-5779-2018

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