



RESEARCH ARTICLE

10.1029/2019MS001889

Key Points:

- We compare CARDAMOM, a simple model with assimilated observations, to structurally complex terrestrial biosphere models (CMIP5 and TRENDY)
- Simple CARDAMOM model with flexible parameters is capable of capturing dynamics in CMIP5 and TRENDY
- CARDAMOM shows a greater range of possible outcomes, particularly for mean net biome exchange

Supporting Information:

- Supporting Information S1

Correspondence to:

G. R. Quetin,
gquetin@stanford.edu

Citation:

Quetin, G., Bloom, A. A., Bowman, K. W., & Konings, A. (2020). Carbon flux variability from a relatively simple ecosystem model with assimilated data is consistent with terrestrial biosphere model estimates. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001889. <https://doi.org/10.1029/2019MS001889>

Received 3 SEP 2019

Accepted 4 MAR 2020

Accepted article online 12 MAR 2020

Carbon Flux Variability From a Relatively Simple Ecosystem Model With Assimilated Data Is Consistent With Terrestrial Biosphere Model Estimates

Gregory R. Quetin¹, A. Anthony Bloom², Kevin W. Bowman^{2,3}, and Alexandra G. Konings¹

¹Department of Earth System Science, Stanford University, Stanford, CA, USA, ²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, ³Joint Institute for Regional Earth System Science and Engineering (JIFRESSE), University of California, Los Angeles, CA, USA

Abstract Modeling the net carbon balance is challenging due to the knowledge gaps in the variability and processes controlling gross carbon fluxes. Terrestrial carbon cycle modeling is susceptible to several sources of bias, including meteorological uncertainty, model structural uncertainty, and model parametric uncertainty. To determine the impact of these uncertainties, we compare three model-derived representations of the global terrestrial carbon balance across 1997–2009: (1) observation-constrained model-data fusion (CARBON Data Model Framework, CARDAMOM), (2) the reanalysis-driven Trends in Net Land-Atmosphere Carbon Exchange (TRENDY) land biosphere model ensemble, and (3) the Coupled Model Intercomparison Project 5 (CMIP5) Earth System Model ensemble. We consider the spread in carbon cycle simulations attributable primarily to parametric uncertainty (CARDAMOM), structural uncertainty (TRENDY), and combined structural and simulated meteorological uncertainty (CMIP5). We find that the spread across the CARDAMOM ensemble long-term mean—produced by parameter uncertainty—is larger than the spread of TRENDY and CMIP5 for net biosphere exchange (NBE) but similar for gross primary productivity (GPP). The carbon flux dynamics of CARDAMOM compares to models in TRENDY as well as models in TRENDY compare to each other in many regions for NBE seasonal (nine of 12), NBE interannual (11 of 12), and GPP seasonal variability (7 of 12), although not for GPP interannual variability (2 of 12). The simple model structure of CARDAMOM and systematic assimilation of observations is sufficient to produce carbon dynamics within the range of more complex models. These results are encouraging for the use of model-data fusion products with empirically estimated uncertainty for global carbon cycle studies.

Plain Language Summary There is uncertainty in where carbon is going into or leaving the land biosphere on a net basis. Models are used to investigate this movement of carbon, but it is unclear how different approaches compare: using a complex model for which the best parameter sets are difficult to determine or using a simpler model for which it is easier to find the right parameters sets that best match observations. This paper compares collections of relatively complex models with a different approach that systematically combines remote sensing observations with a simpler model. Comparisons across mean values, seasonal variations, and interannual variations of photosynthesis and carbon movement show that the simple model—constrained by observations—is able to recreate most dynamics but also shows a broader range of possibilities than the collections of more complex models. The results suggest that more attention needs to be paid to parameter uncertainty in model settings, not just what processes they include. Including this uncertainty creates a larger spread of carbon cycle outcomes for the present and possibly for the future.

1. Introduction

Modeling the net biosphere exchange (NBE) of carbon for the terrestrial biosphere is difficult due to the numerous processes driving variability of considerably larger gross fluxes (photosynthesis, respiration, and fires), and their overall effect on the net carbon balance (Piao et al., 2020). There are still large knowledge

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

gaps in how best to represent and parameterize such processes for the terrestrial biosphere models (TBMs) that contain the model of the carbon cycle embedded in Earth System Models (ESMs). These knowledge gaps are particularly apparent in the spread in future carbon cycle projections under rising atmospheric CO₂ concentrations and climate change. Across ESMs contained in the Coupled Model Intercomparison Project (CMIP5) archive, it is uncertain whether the present-day carbon sink will become stronger, weaken, or even become a source under future climate change (Friedlingstein et al., 2014; Fatichi et al., 2019; Le Quéré et al., 2018). Carbon cycle response to climatic change is the second largest source of uncertainty in predictions of global temperature rise due to anthropogenic CO₂ emissions, behind only climate sensitivity (e.g., cloud formation processes), but greater than aerosols, initial conditions, or anthropogenic emissions scenarios (Booth et al., 2012; Bodman et al., 2013; Bonan & Doney, 2018; Huntzinger et al., 2017). Thus, reducing the uncertainty in carbon cycle models remains a priority for improving long-term predictions of the Earth system. Terrestrial carbon cycle modeling approaches are susceptible to several sources of error, stemming largely from uncertainties in model processes and meteorological drivers. Model uncertainty can be further separated into structural uncertainty (e.g., fidelity and complexity of process representation to reality) and parametric uncertainty (specification accuracy of the parameter values in any given model) (Collins, 2007). While previous research has shown that overall model uncertainty dominates long-term carbon cycle uncertainty compared to initial conditions or human emission scenarios (Bonan & Doney, 2018), it is not clear whether the structural or parametric component dominates model uncertainty (Huntzinger et al., 2017).

In developing TBMs, considerable attention has been paid to structural uncertainty by refining and adding process representation to the carbon cycle, for example, nutrient dynamics processes (Fisher et al., 2014) and vegetation demography (Fisher et al., 2018). By contrast, parametric uncertainty has received less attention. For many carbon-related processes, the parameters of these models are still assigned based on plant functional types (PFTs) (Bonan, 2019). Furthermore, because the models are so complex, the process of deriving the PFT-specific values to be used for each parameter is often manual and opaque. However, the variability in vegetation parameters can be as large within a PFT as between them (Butler et al., 2017; Konings & Gentine, 2017; van Bodegom et al., 2014), suggesting that assuming uniform parameter values per PFT could be a significant source of error. Furthermore, the combination of complex models with simple parameter choices can lead to compensating errors. These compensating errors can lead to models having similar short-term climatic sensitivities yet very different long-term flux predictions (Huntzinger et al., 2017). Such compensating errors among parameters (also known as equifinality Beven & Freer, 2001) are a particular concern for projecting the carbon cycle into novel environments such as occur during long-term climate simulations. It has been shown that the uncertainty in parameters alone can drive a 461 ppm spread in predictions of atmospheric CO₂, larger than the 418 ppm uncertainty across concentration scenarios (Booth et al., 2012). Additionally, as a regional example, the high spread of gross primary productivity (GPP) in ESMs has been traced back to a high spread in emergent specific leaf area for the East Asian monsoon region (Cui et al., 2019). Thus, the difficulty of parameterization of structurally complex models leads to uncertainty when predicting long-term carbon fluxes.

So-called data assimilation or model-data fusion techniques can be used to compare model outputs to observations in order to constrain parameter values and model states (Rayner et al., 2019; Wikle & Berliner, 2007). Furthermore, the uncertainty in these parameters (and associated states of the carbon pools and fluxes) can be used to explicitly quantify the role of parametric uncertainty on carbon cycle variables. At regional to global scales, remote sensing data can be systematically assimilated as observational constraints to optimize model parameters, either within a PFT-based framework (MacBean et al., 2018; Norton et al., 2019) or for deriving parameters on a pixel-by-pixel basis (Bloom et al., 2016). In general, there is a trade-off in model-data fusion studies between the complexity of the underlying carbon cycle model and the complexity of the assimilation method used. Studies using relatively complex carbon cycle models requiring simplifying assumptions, such as Gaussianity and a differentiable cost function to ensure computational tractability (e.g., MacBean et al., 2016; Scholze et al., 2007). By contrast, assimilation methods using Markov chain Monte Carlo approaches allow uncertainty estimation without any assumptions on the shapes of distributions but require large (e.g., millions) of model runs. They are therefore only tractable for relatively parsimonious models that can be run for numerous iterations with current computer resources (e.g., Keenan et al., 2012; Williams et al., 2005). Using parsimonious models also helps to reduce equifinality, which can remain significant in model-data fusion studies, especially those using more complex models (Castro-Morales et al., 2019). Note that this trade-off between model complexity and the degree to which

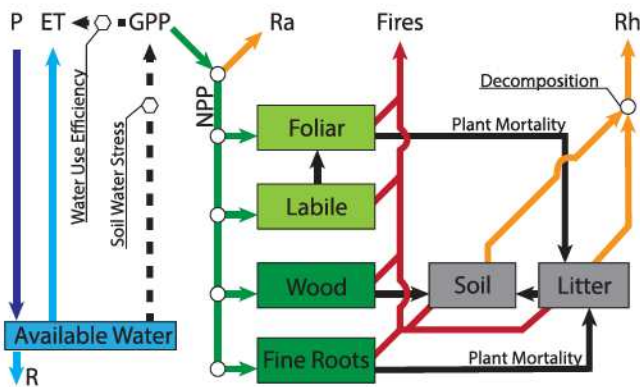


Figure 1. Diagram of the DALEC carbon cycle model in CARDAMOM. Adapted from (Bloom et al., 2016).

parameter choices can be rigorously optimized to be most consistent with observational data extends beyond model-data fusion. More generally, there is a choice between the systematically optimized parameters from model-data fusion and using the expert manual parameter tuning common in the development of more structurally complex TBMs.

The CARBOn Data Model FraMEwork (CARDAMOM) is a model-data fusion framework (Bloom et al., 2016; Bloom & Williams, 2015) built around the Data Assimilation Linked Ecosystem Carbon (DALEC) model (v2.1). CARDAMOM uses an adaptive Metropolis-Hastings Markov chain Monte Carlo (MHMCMC) algorithm data assimilation method and Bayesian inference to update prior distributions of parameters—including initial conditions—in CARDAMOM to better match remote sensing observations. In order to further reduce equifinality, the MHMCMC incorporates known limits on relationships among carbon cycle processes (so-called ecological and dynamic constraints).

For example, the turnover rate of the soil organic matter pool is constrained to be lower than the turnover rate of litter (Bloom & Williams, 2015), which helps avoid unrealistic jumps in soil organic matter (Fox et al., 2009). Like all modeling systems, CARDAMOM's approach has several disadvantages (inability to represent biosphere-atmosphere interactions (Green et al., 2017), sensitivity to observational uncertainty, missing processes related to nutrient dynamics, land use change, etc.) that may be prohibitive for specific applications. Nevertheless, it has been used to investigate a variety of carbon cycle processes, including the effect of fire on turnover rate (Exbrayat et al., 2018), arctic carbon cycling (López-Blanco et al., 2019), and the pools and parameters of the carbon cycle across the globe (Bloom et al., 2016).

In this work, we aim to understand the trade-off between model structural uncertainty and parameter uncertainty for global-scale modeling of the terrestrial carbon cycle. That is, we address the following question: Can simplified models, with data-informed parameters (such as in CARDAMOM), simulate realistic observable carbon fluxes (i.e., NBE and GPP) and their uncertainty, relative to more structurally complex ecosystem models? NBE is the summation of the terrestrial carbon fluxes of GPP, autotrophic respiration, heterotrophic respiration, and fires. Note that it does not account for contributions from methane fluxes nor from lateral fluxes of carbon such as harvest products, or dissolved organic carbon in rivers, etc. We also evaluate the GPP, since it is the largest and most studied of the component fluxes of NBE (Fatichi et al., 2019).

To address this question, we intercompare the mean values, dynamics, and spread in simulated NBE and GPP fluxes of CARDAMOM with those of traditional ensembles of complex TBMs, such as the Trends in Net Land-Atmosphere Carbon Exchange (TRENDY) (Sitch et al., 2008) and Coupled Model Intercomparison Project 5 (CMIP5) (Taylor et al., 2012) ensembles. TRENDY is an ensemble of TBMs that are all forced by the same meteorology. The TBMs in the CMIP5 ensemble are similar but are coupled to a simulated atmosphere in a full ESM. Note that, unlike the estimate of parameter uncertainty in CARDAMOM, the intermodel range of TRENDY and CMIP5 does not statistically capture their complete model structural uncertainty. Nevertheless, consistent with common practice in the interpretation of these ensembles, the intermodel range is taken here to reflect, in part, the effects of different structural representation in individual member models. We further interpret CMIP5, which consists of coupled land-atmosphere models, to reflect additional uncertainty in meteorological drivers, compared to the reanalysis-driven TRENDY ensemble.

2. Methods

2.1. CARDAMOM

CARDAMOM is a model-data fusion system that uses a relatively simple carbon cycle model—DALEC model (Version 2.1)—and a data assimilation algorithm using adaptive MHMCMC (Haario et al., 2001). The underlying carbon cycle model—DALECV2.1—is a modified version of DALEC (Bloom & Williams, 2015) that consists of six pools of carbon and one plant-available water pool and depicted in Figure 1 and detailed in the supporting information. Additionally, CARDAMOM includes ecological and dynamical constraints that help reduce the parameter search space and guard against equifinality (Bloom et al., 2016). Meteorological forcings include temperature (monthly minimum and maximum), shortwave radiation, vapor pressure deficit, precipitation, atmospheric CO₂ concentrations, and burned area.

Table 1
Observation-Based Data Sets Assimilated Into the CARDAMOM Simulation

Observation	Data set description	Assumed uncertainty
Leaf area index (LAI)	2010–2015 MODIS LAI retrievals ^a	$\pm \log(1.2)$
Soil organic matter (SOM)	Soil C inventory (Hiederer & Köchy, 2011)	$\pm \log(1.5)$
Biomass	Derived from GLAS and machine learning (Saatchi et al., 2011) ^b	$\geq \pm \log(1.5)$
Solar-induced fluorescence (SIF)	2010–2015 GOSAT retrievals of fluorescence (Frankenberg et al., 2011) ^c	$\pm \log(2)$
Fire C emissions	2010–2015 Derived from inverse estimates of CO emissions (Worden et al., 2017) and CO ₂ :CO ratios (Bowman et al., 2017)	$\pm 20\%$
Net biosphere exchange (NBE)	2010–2013 CMS-Flux, GOSAT CO ₂ -derived 4° × 5° inverse estimates (Liu et al., 2017).	Seasonal: ± 1 g C/m ² /day, Annual: ± 0.02 g C/m ² /day

Note. Uncertainties denoted as $\pm \log()$ indicate log-transformed model and observed quantities.

^aOnly mean 2010–2015 LAI is assimilated into CARDAMOM, in order to mitigate the influence of seasonal LAI retrieval biases (Bi et al., 2015). ^bsee Bloom et al. (2016) for details. ^cTime-resolved SIF is assimilated as a relative constraint on the temporal variability of GPP.

Here, CARDAMOM was run with monthly time series from 1997–2015 of meteorological drivers from the combined data sets from Climate Research Unit (CRU) and reanalysis data from National Centers for Environmental Prediction (NCEP), the CRUNCEP data set (Harris et al., 2014; Kalnay et al., 1996; Viovy, 2018). Atmospheric CO₂ concentrations from Mauna Loa (Keeling et al., 2009) were also used as a driver. Lastly, burned area was obtained from the Global Fire Emissions Database Version 4.1s (Giglio et al., 2013; Randerson et al., 2012).

All assimilated data sets are described in Table 1. The cost function and uncertainties used in the assimilation are further described in detail in the supporting information. The data sets are first regridded to the 4° × 5° resolution GEOS-Chem grid (Bey et al., 2001), which corresponds to that of the coarsest assimilated data set, NBE. CARDAMOM is then run at this resolution. The dynamic assimilated data sets span the period 2010–2015, with solar-induced fluorescence (SIF)—photon emissions approximately proportional to GPP—and fire emission estimates spanning this entire period, while NBE constraints were only available from 2010 to 2013 (Mohammed et al., 2019). The modeled initial conditions for biomass in 1997 were constrained by a global, remote sensing-derived map (Saatchi et al., 2011), though some error may be induced by the fact that this biomass map represents average conditions over 2001–2003. Similarly, initial conditions for the soil organic matter pool were also constrained by a static map of soil organic carbon (Hiederer & Köchy, 2011). Note that these estimated initial conditions—based on the assimilated data—do not have to be in equilibrium in contrast to most TBMs retrieve their initial conditions from a long run that reaches an equilibrium state.

The uncertainty values applied to each observation represent a combination of both observational uncertainty (errors in measurement), and systematic model-observation error (i.e., model representation error or bias in the data). For lack of better knowledge on both error distributions at 4° × 5° resolution, we manually selected uncertainty values associated with each data set for this study to capture the predominant spatial and temporal variability of the data sets listed in Table 1.

The MHMCMC assimilation with CARDAMOM optimizes the model parameters to best match assimilated data and creates 250 ensemble members each with different model parameter sets. The distribution of the parameters and outputs represent an estimated uncertainty. We run the assimilation four times at each point to create four separate ensembles where the MHMCMC process is initialized differently. We combine these separate ensembles in our analysis to create one ensemble of 1,000 different parameter sets. All parameters converged for 38% of pixels based on a Gelman-Rubin convergence using a threshold of 1.2 (the between chain variance divided by the average variance of each chain) and across the whole data set, 77% of individual parameter solutions converged (Figure S1).

2.2. Comparison: Models, Time, and Regions

We compared estimates of NBE and GPP from CARDAMOM and the TRENDY and CMIP5 ensembles (Table 2). We used the TRENDY V2 output that covered 1901 to 2010. This included nine TBMs: TRIFFID, LPJ, SDGVM, Hyland, NCAR CLM4C, NCAR CLM4CN, OCN, ORCHIDEE, and Vegas (Le Quéré et al., 2018; Sitch et al., 2008). We used the S2 experiment, which includes historical climate from CRUNCEP, and

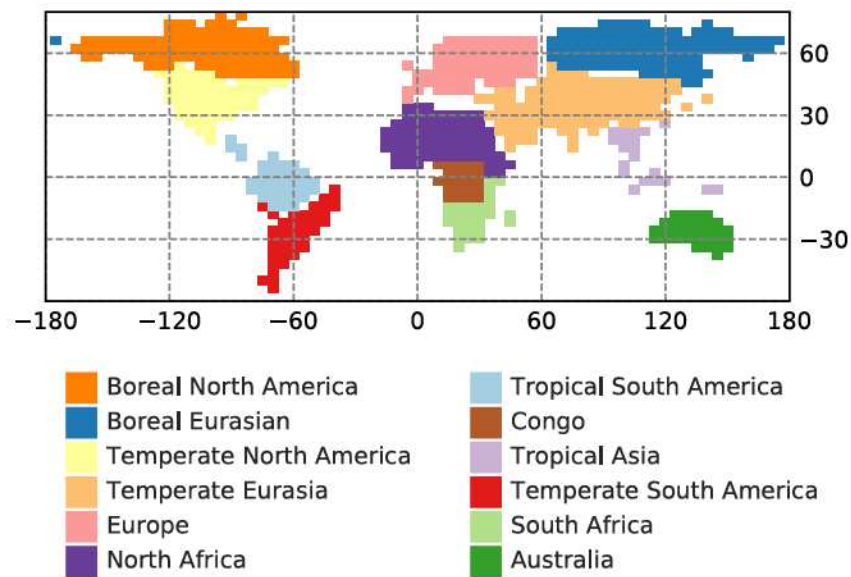


Figure 2. Spatial regions used to aggregate the models for comparison.

increasing atmospheric CO_2 . Note that not all TRENDY models had output on a monthly resolution. Where monthly resolution was necessary—monthly climatology comparisons—only the six models with monthly resolution output were used here: LPJ, Vegas, CLM4C, CLM4CN, TRIFFID, and OCN. We further used models from the CMIP5 archives that had C4MIP experiments (those with a predictive carbon cycle) and had both NBE and GPP publicly available (Taylor et al., 2012) (Table S2). The historical run and RCP8.5 run of the models were downloaded from the CMIP5 archive.

We performed our model comparisons for the period between 1997 and 2009. The RCP8.5 run was used to extend the four years (2006 to 2009) that were not covered by the CMIP5 historical run. We began our comparison at 1997 since that is the earliest that Global Fire Emissions Database Version 4 was available as an input for CARDAMOM. The comparison ended in 2009 in order to keep the entirety of the comparison outside the time period where carbon fluxes were assimilated into CARDAMOM (2010 to 2015). That is, we chose the 1997–2009 test period to be different from the assimilation period in order to ensure that CARDAMOM's performance as a forward model is tested rather than its ability to match specific assimilation data.

Finally, all analyses were performed at each grid point and then aggregated into regions adapted from the Atmospheric Tracer Transport Model Intercomparison Project (Gurney et al., 2004) but with the addition of a region around the Congo rainforest, as carbon cycle dynamics there are expected to differ significantly from elsewhere in North and South Africa (Figure 2). Analysis was performed across regions to better match the informational content of the assimilated NBE from Carbon Monitoring System Flux framework (CMS-Flux) (Liu et al., 2014) and for ease of interpretation.

2.3. Calculation of NBE

NBE is the balance of carbon flux to the atmosphere and includes GPP, autotrophic respiration, heterotrophic respiration, fires, and land use (NBE > 0 to the atmosphere) (Fatichi et al., 2019). Note that the representation of fires and land use varies such that those fluxes are not explicitly represented in each of the

Table 2
Meteorology and Data Approaches of Models in Comparison

Model	Drivers	Land-atmosphere		Model type
		coupling	Observation constraints	
CARDAMOM	Reanalysis (CRUNCEP)	No	Land-surface and atmospheric C data sets	Process
TRENDY	Reanalysis (CRUNCEP)	No	None	Process
CMIP5	Simulated meteorology	Yes	None	Process

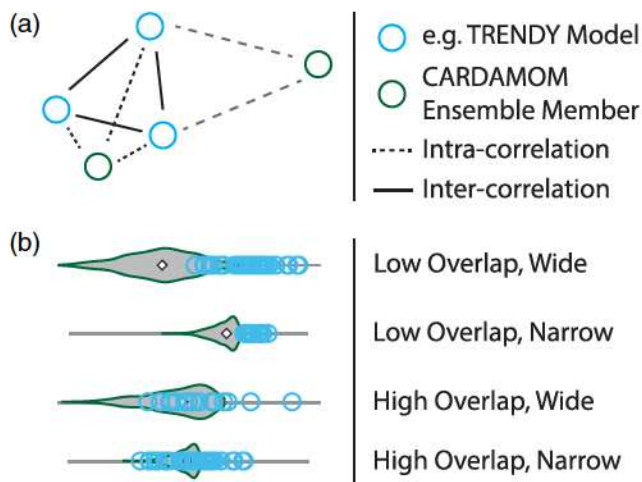


Figure 3. (a) Diagram of relationship among different ensemble members. (b) Schematic of possible patterns in cross-correlation distributions (with correlation on the x axis). Gray shaded violin is intra-correlation (e.g., correlation of TRENDY with CARDAMOM). Blue circles are inter-correlations (e.g., correlation between TRENDY models). Fraction of overlap depends on the uncertainty in both ensembles.

monthly mean)—of NBE and GPP were calculated for each CARDAMOM ensemble member, each model in TRENDY and CMIP5 for the period 1997 to 2009. The statistical difference among median values—across ensembles—was calculated by performing a Kruskal-Wallis H-test, which is a nonparametric test to falsify the null hypothesis that the medians of two distributions are the same.

Additionally, we designed an experiment specifically to investigate whether the dynamics—for example, the pattern of temporal variability—are similar among CARDAMOM and each of the two ensembles of opportunity. This is driven by the recognition that the above statistics could differ not just because of parametric and model uncertainty but also because of observational uncertainty. To address whether the relatively simplistic nature of the model underlying CARDAMOM (induced by the model-data fusion) prevents it from representing the same temporal dynamics as either the TRENDY or CMIP5 ensembles, the distribution of Pearson correlations among ensemble members of each model group (CARDAMOM, TRENDY, and CMIP5) was compared to that among members from the different ensembles. For example, each model in the TRENDY ensemble is correlated with each member of the CARDAMOM ensemble, the “intracorrelation” (1,000 correlations per TRENDY model) at each grid point (as illustrated in Figure 3). The resulting correlations are considered the TRENDY-CARDAMOM correlations. Additionally, each model in the TRENDY ensemble is also correlated with each other model in the TRENDY ensemble; these correlations are considered the TRENDY-TRENDY correlations or more generally the “intercorrelation” as in Figure 3.

Finally, to aggregate over each region, the distributions of correlations at the grid point were averaged together. Each grid point contains an independent CARDAMOM ensemble (i.e., there is not a fixed way to pick, which ensembles to aggregate across space). For example, the first ensemble member in Grid Point X is not necessarily related to the first ensemble member in Group Y. Here, we chose to combine the ensembles ranked by correlation value. For this reason, when averaging the TRENDY-CARDAMOM correlations, we first ranked the correlations from lowest to highest in each grid point and then averaged them together. As a result, the distribution that is reported for a region spans the mean of the lowest correlations to the mean of the highest correlations. This is in contrast to the TRENDY-TRENDY correlations (and CMIP5-CMIP5 correlations, for Figure S2) where a particular model has a solution across space. The result of this is that the regional averages are made up of the average correlation of Model X with Model Y.

We further created a simple univariate metric to summarize the amount of overlap among—using TRENDY and CARDAMOM distributions as an example again—the TRENDY-TRENDY agreement and the TRENDY-CARDAMOM agreement. A strong overlap of the distributions is defined by more than 75% of the TRENDY-CARDAMOM correlations falling above the median of the TRENDY-TRENDY correlations minus two times the interquartile distance of the TRENDY-TRENDY correlations. Similarly, a separate metric was

models considered in this paper. The fire carbon emissions in CARDAMOM are based off of observed burned area combined with combustion fractions (parameters that are optimized during assimilation). When comparing models outside the window of assimilated data, the emissions are thus a combination of prognostic and diagnostic calculations. In contrast, most TRENDY models and CMIP5 have prognostic fire emission, predicted from climate drivers. These differences in estimating fire are only pronounced in regions—here Congo, tropical Asia, and South Africa—where there is a large amount of burning (Giglio et al., 2013; Randerson et al., 2012). Additionally, CARDAMOM does not explicitly include land use; however, the parameter optimization—including states of carbon pools—is based on observations that do include the impact of historical land use change (Pongratz et al., 2009). The TRENDY S2 experiment we use here does not include land use change (Le Quéré et al., 2018; Sitch et al., 2008), while CMIP5 also includes land use change (Taylor et al., 2012). We expect GPP, respiration, and fires (where there is a large amount of burning) to dominate NBE over the broad regions and decadal time scale on which we compare the models.

2.4. Comparison Metrics

The long-term statistics—including the mean, interannual variability (standard deviation of annual values), and seasonality (climatology of

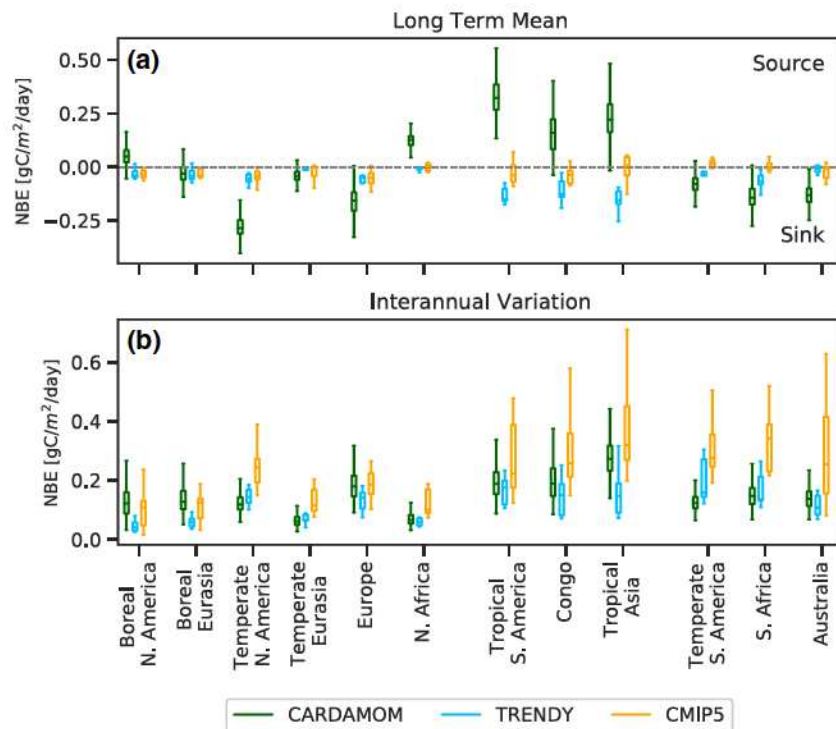


Figure 4. The mean (a) and the coefficient of interannual variation (b) of 1997–2009 NBE for CARDAMOM, TRENDY, and CMIP5. Boxplots reflect the uncertainty across the ensembles.

used to summarize whether the strongest correlations among TRENDY and CARDAMOM were as strong (or nearly as strong) as the strongest correlations among TRENDY models. For this metric, we calculated the difference between the maximum TRENDY-TRENDY correlation and the maximum TRENDY-CARDAMOM correlation. The maximums were judged as close if the maximum TRENDY-CARDAMOM correlation was no more than 0.05 less than the maximum TRENDY-TRENDY correlation. All correlation distributions were compared separately for NBE and GPP and separately for seasonal and interannual time scales. Note that both these metrics were only used to compare two groups of distributions in a binary way and do not have precise values. The binary determinations of the groups of distributions as overlapping or having equally high maxima are presented visually alongside the distributions below.

3. Results

The seasonal and interannual variation of NBE and GPP are intercompared for CARDAMOM, CMIP5, and TRENDY in detail below. Additionally, in supporting information section S4, we compare most of these fluxes to FluxCom, an empirical estimate of NBE and GPP based on flux tower observations (Tramontana et al., 2016). We include FluxCom to serve as an independent—entirely empirical—estimate to contrast with the three estimates that include process information in the form of models (supporting information Figures S9–S13).

3.1. Long-Term Mean and Interannual Standard Deviation

3.1.1. NBE

The median of the CARDAMOM ensemble is significantly different than other models ($p < 0.05$) in at least 9 of 12 regions for each of the other models and with significant overlap of the ensembles in only 2 of 12 regions. The largest differences in the long-term means of regional NBE across all modeling approaches occur in the tropics, with a relatively strong source in all three wet tropical regions for CARDAMOM (~ 0.25 g C/m²/day), a consistent weak sink in TRENDY models (~ -0.1 g C/m²/day), and consistently neutral CMIP5 models (Figure 4a). The disagreement in whether the tropical regions are a source or a sink parallels continued disagreement in the modeling literature (see overview in Fatichi et al., 2019). The source present in CARDAMOM is driven by the assimilated CMS-Flux NBE observations (as shown in Figure S4a).

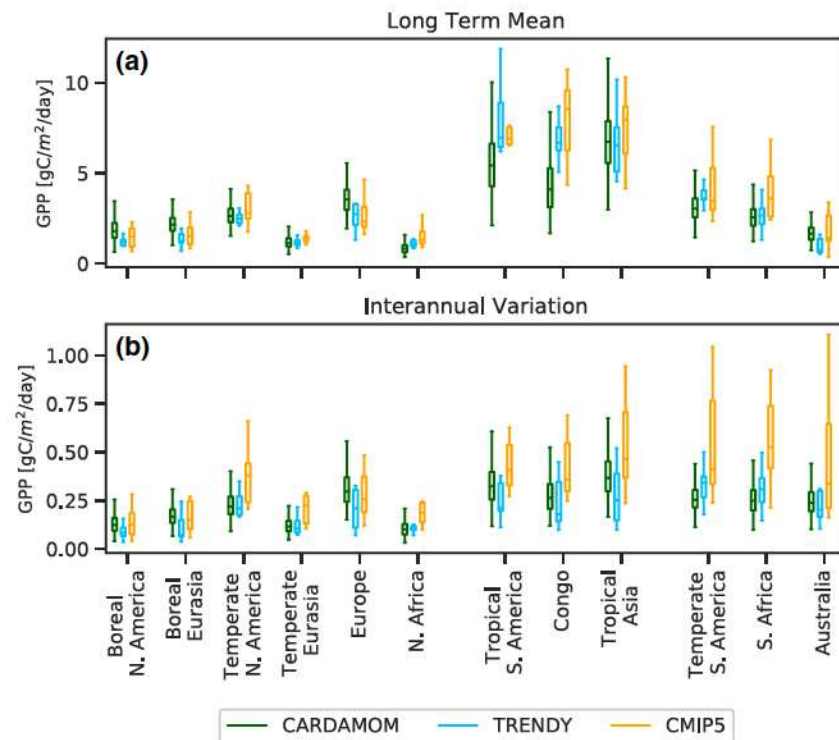


Figure 5. The mean (a) and the standard deviation of interannual variation (b) of 1997–2009 GPP for CARDAMOM, TRENDY, and CMIP5. Boxplots reflect the uncertainty across the ensembles.

By contrast, the weak sinks in TRENDY and CMIP5 are more consistent with the relatively neutral sink estimated by atmospheric inversion estimates driven primarily by extratropical CO_2 surface flask observations (Gaubert et al., 2019). A strong tropical source is a robust feature of fluxes derived from CO_2 satellite observations, which have substantial tropical coverage (Bowman et al., 2017; Crowell et al., 2019; Liu et al., 2017). We note that, as mentioned in section 2.2, the time period considered here (1997–2009) is before the training period with satellite data for CARDAMOM (2010–2015). Consequently, we must assume that the tropical dynamics have not fundamentally changed between those periods of time.

It is notable that even with an observational constraint on NBE, the CARDAMOM ensemble spread of mean NBE is larger than the spread of the other estimates in all regions. The spread in the CARDAMOM ensemble is larger than that of either TRENDY or CMIP5 (Figure 4a). In the tropics, the 95th percentile range of the ensemble spread includes CARDAMOM ensemble members that are nearly neutral, or even a weak sink (Figure 4a). In comparison, the CARDAMOM 95th percentile in temperate Eurasia—though still larger than either TRENDY or CMIP5—is about a third that of the tropics. This broad range in CARDAMOM outcomes highlights that the spread in CARDAMOM solutions are likely providing a better sampling of the full uncertainty than the limited ensembles of opportunity across all regions.

There is less disagreement across model ensembles for the standard deviation of the interannual variation in either median value or ensemble spread (Figure 4b). All models show a relatively large interannual variability and ensemble spread in the tropics, though they are not dramatically different than for some other regions. For example, Tropical South America has median and interquartile ranges that are only somewhat larger than those for Europe for CARDAMOM and TRENDY: median = $0.15 \text{ g C/m}^2/\text{day}$, IQR = $0.06 \text{ g C/m}^2/\text{day}$ in Tropical South America versus median = $0.14 \text{ g C/m}^2/\text{day}$, IQR = $0.06 \text{ g C/m}^2/\text{day}$ in Europe for CARDAMOM, and median = $0.17 \text{ g C/m}^2/\text{day}$, IQR = $0.08 \text{ g C/m}^2/\text{day}$ in Tropical South America versus median = $0.14 \text{ g C/m}^2/\text{day}$, IQR = $0.05 \text{ g C/m}^2/\text{day}$ in Europe for TRENDY. The median value and ensemble spread of CMIP5 interannual variation amplitude are consistently larger than that of TRENDY across all regions and generally larger than that of CARDAMOM except for boreal regions (Figure 4b). The large interannual variability in CMIP5 is consistent with findings of larger than observed meteorology and

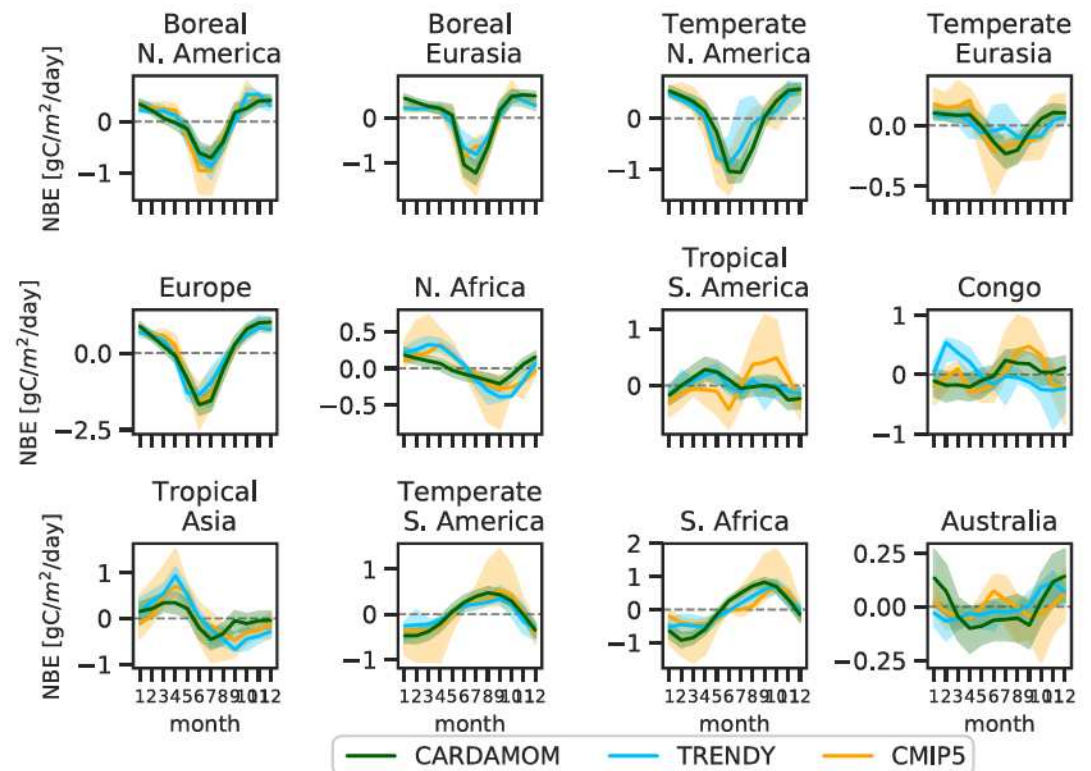


Figure 6. Comparison of the NBE mean-centered seasonal climatology of each region for CARDAMOM, TRENDY, and CMIP5. The 25th to 75th percentiles are shown for CARDAMOM, TRENDY, and CMIP5 as shading.

vegetation variability in the fully coupled ESMs capturing the compounding effects of both meteorological and biogeochemical uncertainties (Anav, Friedlingstein et al., 2013; Anav, Murray-Tortarolo et al., 2013; Merrifield & Xie, 2016; Quetin & Swann, 2018). The spread of the amplitude of interannual variability in CMIP5 is consistently larger in the Southern Hemisphere than in the Northern Hemisphere. TRENDY also generally has a larger amplitude of interannual variation compared to the median CARDAMOM value (8 of 12 regions). However, the ensemble spread of CARDAMOM is equal or larger than TRENDYs in all but one region and shows a lot of overlap. The more closely shared values of TRENDY and CARDAMOM (relative to the larger differences between TRENDY and CMIP5) highlight the impact of sharing similar meteorology on carbon cycle dynamics.

3.1.2. GPP

Across model ensembles, there is strong agreement in the simulations of the mean GPP from 1997–2009 (Figure 5a). Using our metric of greater than 75% of CARDAMOM ensemble members within twice the interquartile distance, there is strong overlap among CARDAMOM ensemble members and alternative model ensembles across most regions (TRENDY [8 of 12 regions] and CMIP5 [11 of 12 regions]). Even with this strong overlap, the majority of regions have statistically distinct ensemble medians ($p < 0.05$). The spread across the CARDAMOM ensemble is larger than the spread across TRENDY in all regions and larger or equal to the spread across CMIP5 in all but temperate South America, South Africa, and Australia (Figure 5a). CARDAMOM has the largest GPP in the boreal regions and is consistently lower than both TRENDY and CMIP5 in tropical South America and the Congo, while there is general agreement in tropical Asia. The general difference between CARDAMOM and TRENDY/CMIP5—higher GPP in the high latitudes, lower in the tropics—is similar to shifts when TBMs are constrained by observations of SIF either as a data assimilation product (MacBean et al., 2018) or optimal integration with already run model ensembles (Parazoo et al., 2014).

The interannual variability of GPP is qualitatively similar across regions, especially relative to the regional variations in NBE interannual variation (Figure 5b). The ordering of magnitudes is generally—from lowest to highest—CARDAMOM (regional median = $0.19 \text{ g C/m}^2/\text{day}$), TRENDY (regional median = $0.20 \text{ g C/m}^2/\text{day}$), and CMIP5 (regional median = $0.35 \text{ g C/m}^2/\text{day}$) the same ordering as for NBE interannual

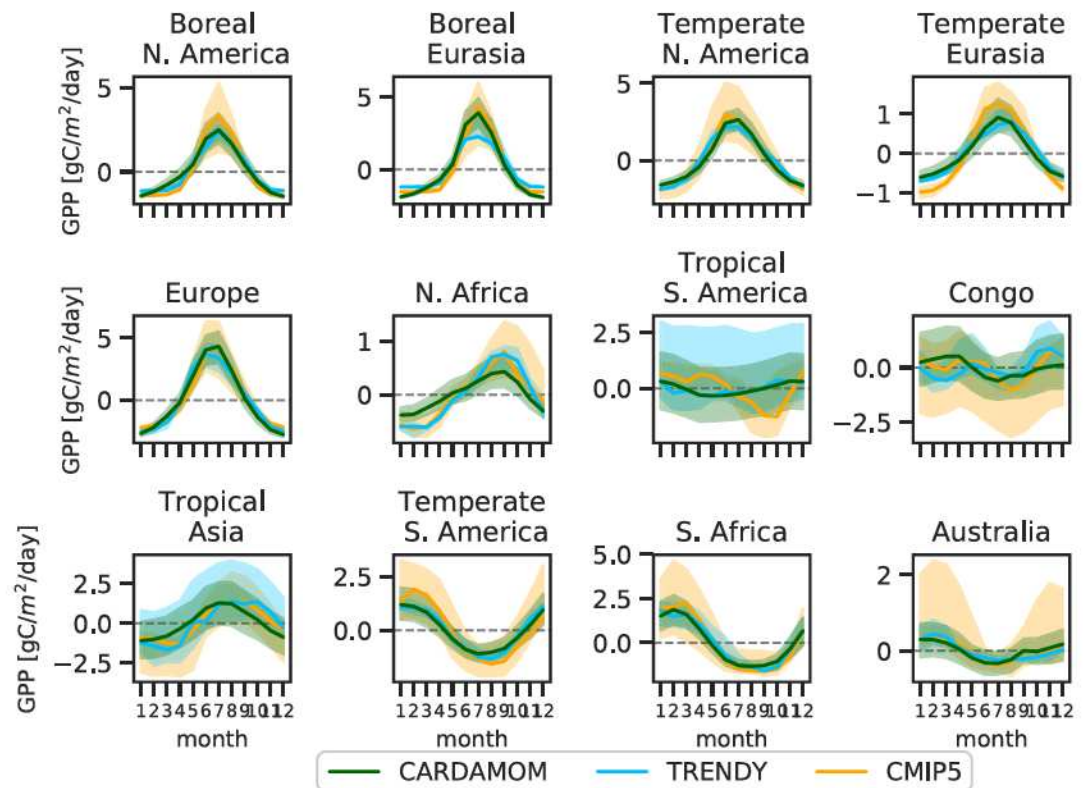


Figure 7. Comparison of the GPP mean-centered seasonal climatology of each region for CARDAMOM, TRENDY, and CMIP5. The line is the ensemble median, 25th to 75th percentiles are shown for CARDAMOM, TRENDY, and CMIP5 as shading.

variability. CARDAMOM and TRENDY are quite similar in most regions with strong overlap in all regions except the boreal regions, temperate South America, and the Congo. Additionally, the medians are statistically different ($p < 0.05$) in only half the regions (6 of 12 regions). As for the interannual variation of NBE, the magnitude and spread of the CMIP5 ensemble are larger than for both TRENDY and CARDAMOM, particularly in the Southern Hemisphere (Figure 5b) likely reflecting the impact of meteorological uncertainty in the CMIP5 models.

3.2. Seasonality

The seasonal cycle of both NBE and GPP has relatively consistent amplitudes across models, and generally the 25th to 75th percentile spread across ensembles is small compared to the median values (Figures 6 and 7).

3.2.1. NBE

The greatest disagreement in the seasonal cycle of NBE occurs in northern Africa, the tropical regions, and Australia (Figure 6). The greatest qualitative agreement occurs in the higher latitudes, where there are stronger seasonal cycles of climate. Though broadly consistent with all other models across most of the year, CARDAMOM consistently has a different shape during the initiation of the growing season. CARDAMOM has a gradual decrease of NBE (CO_2 drawn from the atmosphere) over the period January to April while TRENDY and CMIP5 are nearly flat (no change) until April; this is most apparent in the regions boreal N. America and Europe (Figure 6). Additionally, in the high latitudes TRENDY and CMIP5 agree strongly but CARDAMOM has a different growing season value of NBE, either a stronger sink (boreal Eurasia) or a weaker one (boreal N. America). Disagreement across models stands out in the temperate Eurasia and northern Africa regions where CARDAMOM, TRENDY, and CMIP5 have similar magnitudes.

3.2.2. GPP

There is generally less disagreement across the seasonal cycle of GPP than in the seasonal cycle of NBE (Figure 7). The spread across the ensembles is large in the tropics for CARDAMOM, TRENDY, and CMIP5 across the tropics and CMIP5 stands out as different from the other models in tropical S. America, while all models disagree in the Congo (Figure 7). Similarly to the seasonal cycle of NBE, the beginning of the growing season is more gradual in CARDAMOM than other models in the high latitudes, particularly in

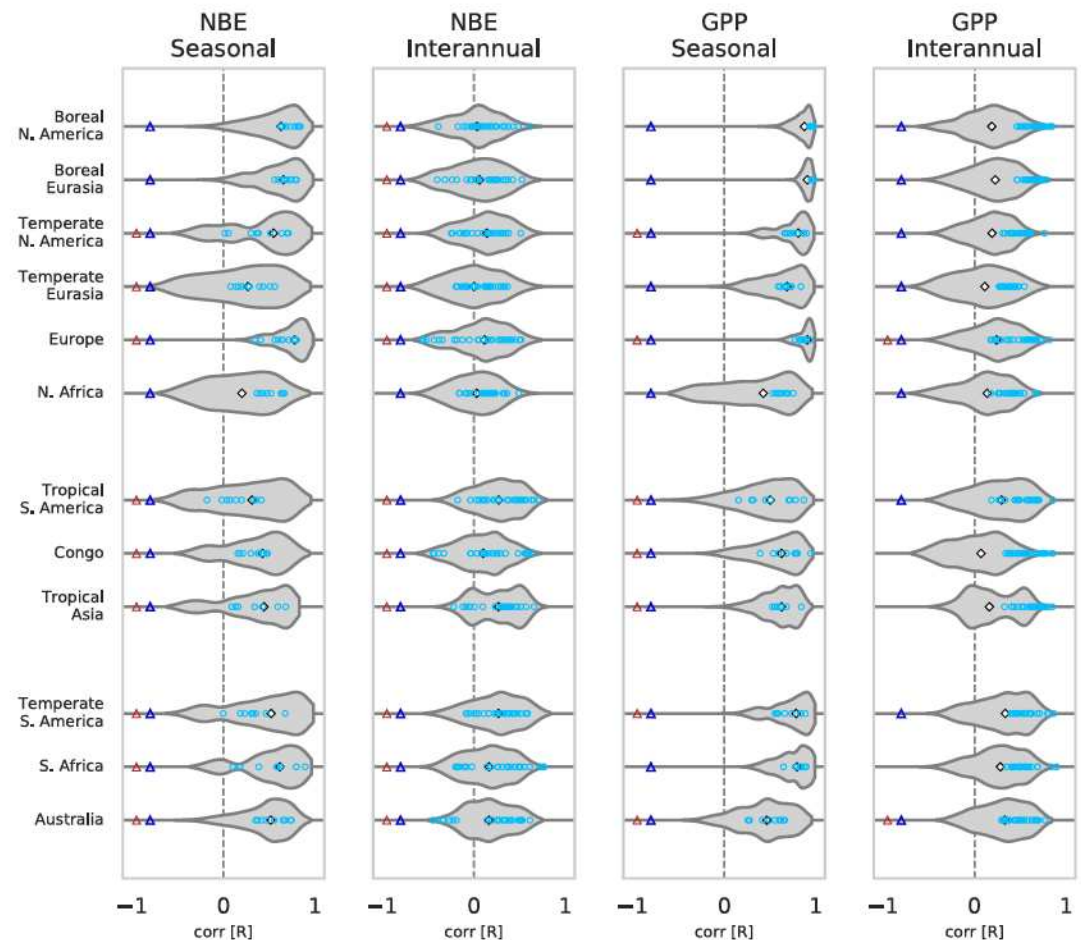


Figure 8. The correlation among TRENDY models and CARDAMOM ensemble members (gray violin, white diamond is median value) and cross correlation among TRENDY models (blue circles) for NBE seasonal climatology (a), NBE interannual variation (b), GPP seasonal climatology (c), and GPP interannual variation (d) across regions. Red triangles represent regions for which more than 75% of CARDAMOM-TRENDY correlations are within 2*interquartile from the median of TRENDY-TRENDY correlations. Blue triangles indicate regions for which the maximum CARDAMOM-TRENDY correlation is greater than or no more than 0.05 correlation less than the maximum TRENDY-TRENDY correlation.

the boreal regions. This slow increase after winter may signify a difference in the models in how GPP under near freezing temperatures is represented.

3.3. Comparing TRENDY and CARDAMOM Seasonal and Interannual Variability

We compare the agreement in variability across the TRENDY ensemble (using correlations of pairs of the different TRENDY models) with the agreement among TRENDY models and CARDAMOM ensemble members (using the correlation of each TRENDY model and each CARDAMOM ensemble member). However, because both data sets are based on the CRUNCEP meteorological forcing, their differences are solely due to model structure and the observations CARDAMOM assimilates. Thus, this comparison can determine whether the simplified structure of CARDAMOM is able to create similar dynamics as the TRENDY models.

We find strong overlap among TRENDY-TRENDY correlations and TRENDY-CARDAMOM correlations in most regions (high overlap denoted by red triangles in Figure 8) for both the NBE seasonal cycle (9 of 12 regions) and NBE interannual variability (11 of 12). High overlap can be due to the two distributions having similar medians or due to one of the distributions being relatively broad compared to the other, as shown in Figure 3b. For the NBE seasonal cycle, the high overlap is due to mediocre agreement among TRENDY models and most of the CARDAMOM ensemble members agreeing with the bulk of TRENDY models. For NBE interannual variation, the overlap of distributions is due to a lack of agreement among TRENDY models (and the resulting broad distribution of correlations). The TRENDY-TRENDY

correlations for NBE interannual variation are grouped around 0 and low correlation values and have a relatively broad distribution. This TRENDY-TRENDY disagreement is then reflected in a similar distribution of TRENDY-CARDAMOM correlations. We note that this disagreement is primarily in the TRENDY ensemble, as the CARDAMOM-CARDAMOM correlations for NBE interannual variability have a relatively strong correlation, though there is still a broad range of correlations (Figure S3).

In contrast, fewer regions show strong overlap between distributions for the GPP seasonal cycle (7 of 12) and for the GPP interannual variability (2 of 12) (Figure 8). The lack of overlap in GPP seasonal cycle is due to a strong agreement among TRENDY models—high average correlation and a small distribution—so that, even though there is high correlation in TRENDY-CARDAMOM, a large portion of it still falls outside the TRENDY-TRENDY distribution. The difference in GPP interannual variability is more systematic—with the bulk distribution of TRENDY-CARDAMOM correlations not agreeing with the relatively strong TRENDY-TRENDY correlations.

We further compared the strongest correlations in each distribution: Is at least one CARDAMOM ensemble member able to match a TRENDY model as well (or at least within 0.05) as two of the TRENDY models are able to match each other? From this perspective (denoted by blue triangles in Figure 8), the TRENDY-CARDAMOM contains at least one high correlation member in most regions for NBE seasonal (12 of 12), GPP seasonal (12 of 12), NBE interannual variation (12 of 12), and GPP interannual variation (9 of 12).

Overall, the TRENDY-CARDAMOM correlation distributions are broader than those for TRENDY-TRENDY for both NBE and GPP, across all the regions and for both the seasonal cycle and interannual variation (Figure 8). This broader distribution of relationships suggests that the CARDAMOM ensemble contains a more diverse set of patterns than the TRENDY ensemble. This is particularly apparent in the seasonal cycles of both variables and in the GPP interannual variability where the TRENDY-TRENDY correlations are relatively high, and the distribution is compact compared to the TRENDY-CARDAMOM distribution. By contrast, there is little agreement across the TRENDY ensemble on the interannual variability of NBE, where median TRENDY-TRENDY and TRENDY-CARDAMOM correlations are near 0, and the spread of the two distributions is similar in most regions.

4. Discussion

4.1. Simple Versus Complex Models: Dynamics of Carbon Cycle Fluxes

The match among CARDAMOM, TRENDY, and CMIP5 suggests that the relatively simple structure of CARDAMOM is sufficient (at least relative to current alternative models and uncertain observations) to simulate the amplitude of and much of the dynamics of the carbon cycle variation across regions of the globe (Figures 4–8 and S4–S8). Though the processes resulting in fluxes of carbon are complex, it is possible to capture their dynamics with a model simpler than typical TBMs. The simulation of NBE and GPP by CARDAMOM is informed by both the carbon cycle model and observational constraints included in this analysis. The benefits of assimilating multiple streams of observations to constrain carbon cycle estimates and provide empirical uncertainties for those estimates appear to largely compensate for any drawbacks of a simple model.

We focused our investigation on the dynamics—particularly the interannual variation—of TRENDY and CARDAMOM, as they are driven by meteorology from similar reanalysis data sets. Areas where there is consistent disagreement among the two suggest more detailed investigation is necessary to determine how either TRENDY or CARDAMOM (or both) might be improved. For example, we hypothesize that the consistent difference between CARDAMOM and TRENDY in the shape of the ramp up of GPP and NBE at the beginning of the boreal growing season is due to an oversimplified representation of leaf-out processes in CARDAMOM (Figures 6–7). In CARDAMOM, the day of leaf out (and leaf fall) is an optimized parameter for each pixel that does not change from year to year once set during assimilation. In reality, the leaf-out day is likely a function of climate (e.g., temperature and photoperiod) (Richardson et al., 2010, 2013; White et al., 1997, 2010). These observed differences can be investigated in future work through incremental tests of more complex models, especially as more observational constraints become available. It is possible that this process is also part of the reason for the systematic discrepancy between CARDAMOM and TRENDY in the pattern of interannual variation (Figure 8). Previous work has shown that interannual variation of GPP is closely tied to the variation in growing season duration (Keenan et al., 2012). At the same time, the

CARDAMOM-TRENDY mismatch in interannual variation could also be due to the use of the SIF constraint in CARDAMOM leading to a more realistic set of interannual variation in GPP. This hypothesis is supported by the fact that there are consistently CARDAMOM ensemble members—though only a few—that correlate highly with TRENDY models, at least showing the capability of the CARDAMOM structure to recreate patterns similar to those created by more structurally complex models (Figure 8). In future work, we can learn more about the processes governing these interannual variations by considering the difference between CARDAMOM ensemble members that match and do not match other models and observations where they are available.

4.2. TRENDY and CMIP5 Ensembles Underestimate the Range of Plausible Carbon Cycle Fluxes Due to Parameter Uncertainty

The spread across ensembles of TRENDY or CMIP5 provides an estimate of the range of outcomes that are possible from different models. The spread of these ensembles is not a formal uncertainty as the models are neither independent nor systematically constrained by their ability to predict observations (Knutti et al., 2013; Masson & Knutti, 2011). To constrain flux estimates, there have been a number of attempts to weight each ensemble member by the skill of reproducing observations (Abramowitz & Gupta, 2008; Parazoo et al., 2014; Sanderson et al., 2017). However, the result from CARDAMOM that even systematically constrained parameter uncertainty is large relative to the TRENDY and CMIP5 ensembles suggests that they are likely underestimating the plausible carbon cycle outcomes. Indeed, the CARDAMOM ensemble spread—driven solely by uncertainty in observationally constrained parameters—is as large or larger than the spread of TRENDY—driven by structural uncertainty—in all regions (and in most regions larger than the spread of CMIP5 as well) (Figures 4 and 5). We propose that CARDAMOM is likely capturing a more complete set of outcomes than the TRENDY or CMIP5 outcomes due to the importance—and uncertainty—of functional parameters. If this is true, estimates of outcomes based on TRENDY and CMIP5 used in hindcasts and forecasts are underestimating the possible range of flux outcomes.

Additional efforts to improve model-data fusion systems such as CARDAMOM might therefore be useful toward better understanding the past and future of the terrestrial carbon cycle. These results are consistent with previous work showing the importance of parameter choice to model-simulated NBE and GPP. In Booth et al. (2012), a carbon cycle model is simplified to a few functional parameters, and the spread in future predictions is estimated by varying these functional parameters across a reasonable range of uncertainty—indeed, the results show the spread is as large as that of the whole CMIP5 ensemble. Similarly, in the diagnostic effort of Cui et al. (2019), a simple model of functional parameters is fit to global models of the Multiscale Synthesis and Terrestrial Model Intercomparison Project ensemble to show that the variation in functional parameters explains a large portion of the variation in GPP. Lovenduski and Bonan (2017) demonstrate that model uncertainty dominates the spread of future terrestrial biosphere projections and suggests that a multitude of modeling approaches is part of what is required for more reliable projections. Taken together with the results of the comparison performed in this work, these results suggest that future development of carbon cycle estimation systems should carefully consider parametric uncertainty.

4.3. CARDAMOM Permits Carbon Cycle Outcomes Not Found in TRENDY and CMIP5 Ensembles

There is no agreed upon observational arbiter to determine which simulations are more realistic when comparing the ensembles from CARDAMOM, TRENDY, and CMIP5. Although there are many candidate data sets to be used as a benchmark, all are either sparse (in the case of in situ measurements) or uncertain (in the case of remote sensing measurements Anav et al., 2015; Jiang et al., 2017). However, CARDAMOM (like other data assimilation systems) has the potential to expose additional carbon cycle outcomes when the uncertainty associated with observational constraints and the flexibility created by different realistic parameter values are considered. One example of such an outcome is the presence of a carbon source in the tropics in CARDAMOM not found in CMIP5 or TRENDY. This is driven by the CMS-Flux NBE observational constraint used. While the question of whether tropical land surface is a source or sink of carbon remains controversial (Baccini et al., 2017; Gaubert et al., 2019; Schimel et al., 2014), satellite CO₂ observations clearly indicate a source—and the CARDAMOM results reflect the impact of that data. However, the ability of CARDAMOM to match these observations (Figures 4 and S1) suggests that realistic parameter choices exist for which a source can exist in an integrated carbon cycle model (conserves mass) and are consistent with other observational constraints (e.g., SIF, leaf area index, biomass, and fire emissions).

As a second example of CARDAMOM's greater range of outcomes, it is notable that the ensemble spread of GPP in CARDAMOM, TRENDY, and CMIP5 are of similar magnitudes, while the ensemble spread of NBE in TRENDY and CMIP5 is much smaller relative to the ensemble spread of NBE in CARDAMOM (Figures 4 and 5). The NBE estimates from all the models—relatively small compared to GPP—demonstrate the close balance of GPP with respiration and fire. However, the relatively large ensemble spread of NBE in CARDAMOM suggests that either the NBE range in CARDAMOM is underconstrained (e.g., because of too wide an assumed uncertainty in observations) or that the NBE/GPP ratio in TRENDY and CMIP5 is overconstrained (e.g., through too limited a range of parameter scenarios considered). It is unclear which of the two, if not both, is true.

This study cannot conclusively attribute differences between CARDAMOM, TRENDY, and CMIP5, to failures of any of the three data sets. We argue that at present, no sufficiently high-fidelity observational benchmarks of carbon fluxes exist to determine whether either CARDAMOM or TRENDY and CMIP5 ensembles are more correct at times when they disagree, particularly in regions where observations are sparse, such as the tropics or boreal regions (Schimel et al., 2015). Nevertheless, insofar as CARDAMOM rigorously assimilates observational estimates of NBE and other constraints and their uncertainty, it might be expected to more closely approximate uncertain observational data sets and estimate an empirical uncertainty. Indeed, the problem of interpolating uncertain observations to determine a global benchmark is not unique to the carbon cycle. In the atmospheric sciences, long-term reanalyses of weather models—incorporating numerous assimilated data streams—have a strong history (Harms et al., 1992) being used as pseudo-observations to benchmark the simulated atmospheres in climate models (e.g., Gleckler et al., 2008). Relative to its application in atmospheric sciences and meteorology, the use of data assimilation within carbon cycle science is still relatively new. A number of open questions remain, including the role of structural error (Keenan et al., 2011; Wieder et al., 2015), choice of assimilation method (Fox et al., 2009), or combination of data sets used (Chevallier et al., 2019; McNeall et al., 2016; Smith et al., 2019), and consistency between performance during and outside the assimilation time window. Nevertheless, as model-data fusion systems such as CARDAMOM and other efforts continue to mature in the medium and long terms, they could eventually serve as a “carbon cycle reanalysis” to benchmark global vegetation.

5. Conclusions

We compared the NBE and GPP estimates from CARDAMOM—a relatively simple carbon cycle model with parameters optimized through data assimilation—and the complex TBMs in TRENDY and CMIP5. The assimilated CMS-Flux NBE in CARDAMOM-shaped distinct differences in carbon sources and sinks compared to both TRENDY and CMIP5. However, the dynamics of the models were similar, with only systematic differences between the CARDAMOM ensemble and TRENDY ensemble in GPP interannual variation. These comparisons provide evidence that the complexity of CARDAMOM is sufficient to simulate the dynamics of GPP and NBE fluxes, showing little systematic difference with more complex TBMs. Additionally, we propose that the spread across the TRENDY and CMIP5 ensembles does not capture the full spread of potential carbon cycle behavior, as the spread in CARDAMOM from optimized parameters alone is as large or larger.

Comparisons between CARDAMOM and other TBMs in TRENDY and CMIP5 have highlighted two areas to be investigated in future work: (1) the range of uncertainty in NBE and (2) the growing season parameterization in boreal regions. In the first, NBE appears highly constrained in complex land surface models compared to the spread in GPP. This leads to a weak sink that is not observed in CMS-Flux or estimated through CARDAMOM. In the second, the consistent difference between the growing season increase of complex models and CARDAMOM suggests that improvements to CARDAMOM can likely be made based on improved process representation of early growing season leaf out. The ability to optimize parameters based on remote sensing observations and constrain parameter uncertainty makes CARDAMOM a useful addition to the suite of models currently estimating global fluxes of the terrestrial carbon cycle.

References

- Abramowitz, G., & Gupta, H. (2008). Toward a model space and model independence metric. *Geophysical Research Letters*, 35, L05705. <https://doi.org/10.1029/2007GL032834>
- 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, 785–818. <https://doi.org/10.1002/2015RG000483>

Acknowledgments

All authors were funded by NASANNH16ZDA001N-IDS. K. W. B. and A. A. B. also acknowledge the support of NASANNH14ZDA001N-CMS. A. G. K. and G. R. Q. also acknowledge support from NSF under Award DEB-1942133. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The data used in this manuscript are available online (<https://figshare.com/s/d1e03b182ad506250815>).

- Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., et al. (2013). Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth System Models. *Journal of Climate*, 26(18), 6801–6843.
- Anav, A., Murray-Tortarolo, G., Friedlingstein, P., Sitch, S., Piao, S., & Zhu, Z. (2013). Evaluation of land surface models in reproducing satellite derived leaf area index over the high-latitude northern hemisphere. Part II: Earth system models. *Remote Sensing*, 5(8), 3637–3661.
- Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., & Houghton, R. A. (2017). Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science*, 358, 230–234.
- Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249, 11–29.
- Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research*, 106(D19), 23,073–23,095.
- Bi, J., Knyazikhin, Y., Choi, S., Park, T., Barichivich, J., Ciais, P., et al. (2015). Sunlight mediated seasonality in canopy structure and photosynthetic activity of Amazonian rainforests. *Environmental Research Letters*, 10(6), 064,014.
- Bloom, A. A., Exbrayat, J., van der Velde, I. R., Feng, L., & Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, 113(5), 1285–1290.
- Bloom, A. A., & Williams, M. (2015). Constraining ecosystem carbon dynamics in a data-limited world: Integrating ecological “common sense” in a model-data fusion framework. *Biogeosciences*, 12(5), 1299.
- Bodman, R. W., Rayner, P. J., & Karoly, D. J. (2013). Uncertainty in temperature projections reduced using carbon cycle and climate observations. *Nature Climate Change*, 3(8), 725–729.
- Bonan, G. B. (2019). Cambridge: Cambridge University Press.
- 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.
- Booth, B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., et al. (2012). High sensitivity of future global warming to land carbon cycle processes. *Environmental Research Letters*, 7(2), 024,002.
- Bowman, K. W., Liu, J., Bloom, A. A., Parazoo, N. C., Lee, M., Jiang, Z., et al. (2017). Global and Brazilian carbon response to El Niño Modoki 2011–2010: Brazilian carbon balance. *Earth and Space Science*, 4(10), 637–660.
- Butler, E. E., Datta, A., Flores-Moreno, H., Chen, M., Wythers, K. R., Fazayeli, F., et al. (2017). Mapping local and global variability in plant trait distributions. *Proceedings of the National Academy of Sciences*, 114(51), E10937.
- Castro-Morales, Karel, Schürmann, G., Köstler, C., Rödenbeck, C., Heimann, M., & Zaehele, S. (2019). Three decades of simulated global terrestrial carbon fluxes from a data assimilation system confronted to different periods of observations. *Biogeosciences Discussions*, 16, 3009–3032.
- Chevallier, F., Remaud, M., O'Dell, C. W., Baker, D., Peylin, P., & Cozic, A. (2019). Objective evaluation of surface- and satellite-driven carbon dioxide atmospheric inversions. *Atmospheric Chemistry and Physics Discussions*, 19(22), 14,233–14,233.
- Collins, M. (2007). Ensembles and probabilities: A new era in the prediction of climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857), 1957–1970.
- Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A., Chevallier, F., & Sweeney, C. (2019). The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network. *Atmospheric Chemistry and Physics Discussions*, 19, 9797–9831.
- Cui, E., Huang, K., Arain, M. A., Fisher, J. B., Huntzinger, D. N., Ito, A., et al. (2019). Vegetation functional properties determine uncertainty of simulated ecosystem productivity: A traceability analysis in the East Asian monsoon region. *Global Biogeochemical Cycles*, 33, 668–689. <https://doi.org/10.1029/2018GB005909>
- Exbrayat, J., Smallman, T. L., Bloom, A. A., Hutley, L. B., & Williams, M. (2018). Inverse determination of the influence of fire on vegetation carbon turnover in the pantropics. *Global Biogeochemical Cycles*, 32, 1776–1789. <https://doi.org/10.1029/2018GB005925>
- Faticchi, S., Pappas, C., Zscheischler, J., & Leuzinger, S. (2019). Modelling carbon sources and sinks in terrestrial vegetation. *New Phytologist*, 221(2), 652–668.
- 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.
- Fisher, R. A., Koven, C. D., Anderegg, W. R. L., Christoffersen, B. O., Dietze, M. C., Farrior, C. E., et al. (2018). Vegetation demographics in Earth System Models: A review of progress and priorities. *Global Change Biology*, 24(1), 35–54.
- Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., et al. (2009). The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agricultural and Forest Meteorology*, 149(10), 1597–1615.
- 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, L17706. <https://doi.org/10.1029/2011GL048738>
- 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.
- Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., et al. (2019). Global atmospheric CO₂ inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate. *Biogeosciences*, 16, 117–134.
- Giglio, L., Randerson, J. T., & van der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation Global Fire Emissions Database (GFED4): Analysis of burned area. *Journal of Geophysical Research: Biogeosciences*, 118, 317–328. <https://doi.org/10.1002/jgrg.20042>
- Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, 113, D06104. <https://doi.org/10.1029/2007JD008972>
- Green, J. K., Konings, A. G., Alemohammad, S. H., Berry, J., Entekhabi, D., Kolassa, J., et al. (2017). Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nature Geoscience*, 10(6), 410–414.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., et al. (2004). Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks: T3 seasonal results. *Global Biogeochemical Cycles*, 18, GB1010. <https://doi.org/10.1029/2003GB002111>
- Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2), 223.
- Harms, D. E., Raman, S., & Madala, R. V. (1992). An examination of four-dimensional data-assimilation techniques for numerical weather prediction. *Bulletin of the American Meteorological Society*, 73(4), 425–440.
- Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset: Updated high-resolution grids of monthly climatic observations. *International Journal of Climatology*, 34(3), 623–642.

- Hiederer, R., & Köchy, M. (2011). Global soil organic carbon estimates and the harmonized world soil database. *EUR*, 25225, 79.
- 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), 4765.
- Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., & Zhu, Z. (2017). Inconsistencies of interannual variability and trends in long-term satellite leaf area index products. *Global Change Biology*, 23(10), 4133–4146.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437–472.
- Keeling, R., Piper, S., Bollenbacher, A., & Walker, J. (2009). Atmospheric carbon dioxide record from Mauna Loa (1958–2008) (Tech. Rep.). Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National.
- Keenan, T. F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., et al. (2012). Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO₂ exchange. *Global Change Biology*, 18(6), 1971–1987.
- Keenan, T. F., Carbone, M. S., Reichstein, M., & Richardson, A. D. (2011). The model–data fusion pitfall: Assuming certainty in an uncertain world. *Oecologia*, 167(3), 587–597.
- Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W., & Richardson, A. D. (2012). Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. *Global Change Biology*, 18(8), 2555–2569.
- Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there: Climate model genealogy. *Geophysical Research Letters*, 40, 1194–1199. <https://doi.org/10.1002/grl.50256>
- Konings, A. G., & Gentine, P. (2017). Global variations in ecosystem-scale isohydricity. *Global Change Biology*, 23(2), 891–905.
- Le Quéré, C. L., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., et al. (2018). Global carbon budget 2018. *Earth System Science Data*, 10, 2141–2194.
- Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserrez, N., Brix, H., et al. (2014). Carbon monitoring system flux estimation and attribution: of Impact of ACOS-GOSAT X CO₂ sampling on the inference of terrestrial biospheric sources and sinks. *Tellus B: Chemical and Physical Meteorology*, 66(1), 22,486.
- Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., et al. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. *Science*, 358(6360).
- López-Blanco, E., Exbrayat, J.-F., Lund, M., Christensen, T. R., Tamstorf, M. P., Slevin, D., et al. (2019). Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system. *Earth System Dynamics*, 10(2), 233–255.
- Lovenduski, N. S., & Bonan, G. B. (2017). Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters*, 12(4), 044,020.
- MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., et al. (2018). Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data. *Scientific Reports*, 8(1), 1973.
- MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle dataassimilation system. *Geoscientific Model Development*, 9(10), 3569–3588.
- Masson, D., & Knutti, R. (2011). Climate model genealogy: Climate model genealogy. *Geophysical Research Letters*, 38, L08703. <https://doi.org/10.1029/2011GL046864>
- McNeill, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wiltshire, A., & Sexton, D. (2016). The impact of structural error on parameter constraint in a climate model. *Earth System Dynamics*, 7(4), 917–935.
- Merrifield, A. L., & Xie, S.-P. (2016). Summer U.S. surface air temperature variability: Controlling factors and AMIP simulation biases. *Journal of Climate*, 29(14), 5123–5139.
- Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., et al. (2019). Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sensing of Environment*, 231, 111,177.
- Norton, A. J., Rayner, P. J., Koffi, E. N., Scholze, M., Silver, J. D., & Wang, Y.-P. (2019). Estimating global gross primary productivity using chlorophyll fluorescence and a data assimilation system with the BETHY-SCOPE model. *Biogeosciences*, 16, 3069–3093. <https://doi.org/10.5194/bg-16-3069-2019>
- Parazoo, N. C., Bowman, K., Fisher, J. B., Frankenberg, C., Jones, Dylan 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.
- Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., et al. (2020). Interannual variation of terrestrial carbon cycle: Issues and perspectives. *Global Change Biology*, 26(1), 300–318.
- Pongratz, J., Reick, C. H., Raddatz, T., & Claussen, M. (2009). Effects of anthropogenic land cover change on the carbon cycle of the last millennium: Land cover change and carbon cycle. *Global Biogeochemical Cycles*, 23, GB4001. <https://doi.org/10.1029/2009GB003488>
- Quetin, G. R., & Swann, A. L. S. (2018). Sensitivity of leaf area to interannual climate variation as a diagnostic of ecosystem function in CMIP5 carbon cycle models. *Journal of Climate*, 31(20), 8607–8625.
- Randerson, J. T., Chen, Y., van der Werf, G. R., Rogers, B. M., & Morton, D. C. (2012). Global burned area and biomass burning emissions from small fires. *Journal of Geophysical Research*, 117, G04012. <https://doi.org/10.1029/2012JG002128>
- Rayner, P. J., Michalak, A. M., & Chevallier, F. (2019). Fundamentals of data assimilation applied to biogeochemistry. *Atmospheric Chemistry and Physics*, 19(22), 13,911–13,932.
- Richardson, A. D., Andy Black, T., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., et al. (2010). Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1555), 3227.
- Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., & Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, 169(0), 156–173.
- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108(24), 9899–9904.
- Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, 10(6), 2379–2395.
- Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., et al. (2015). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, 21(5), 1762–1776.
- Schimel, D., Stephens, B. B., & Fisher, J. B. (2014). Effect of increasing CO₂ on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences*, 112, 436–441.
- Scholze, M., Kaminski, T., Rayner, P., Knorr, W., & Giering, R. (2007). Propagating uncertainty through prognostic carbon cycle data assimilation system simulations. *Journal of Geophysical Research*, 112, D17305. <https://doi.org/10.1029/2007JD008642>
- 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.

- Smith, W. K., Fox, A. M., MacBean, N., Moore, D. J. P., & Parazoo, N. C. (2019). Constraining estimates of terrestrial carbon uptake: New opportunities using long-term satellite observations and data assimilation. *New Phytologist*, 225, 105–112.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498.
- 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.
- van Bodegom, P. M., Douma, J. C., & Verheijen, L. M. (2014). A fully traits-based approach to modeling global vegetation distribution. *Proceedings of the National Academy of Sciences*, 111(38), 13,733–13,738.
- Viovy, N. (2018). CRUNCEP Version 7—Atmospheric forcing data for the Community Land Model. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.
- White, M. A., Thornton, P. E., & Running, S. W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, 11(2), 217–234.
- Wieder, W. R., Cleveland, C. C., Lawrence, D. M., & Bonan, G. B. (2015). Effects of model structural uncertainty on carbon cycle projections: Biological nitrogen fixation as a case study. *Environmental Research Letters*, 10(4), 044016.
- Wikle, C. K., & Berliner, L. M. (2007). A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1–2), 1–16.
- Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., & Kurpius, M. R. (2005). An improved analysis of forest carbon dynamics using data assimilation. *Global change biology*, 11(1), 89–105.
- Worden, J. R., Bloom, A. A., Pandey, S., Jiang, Z., Worden, H. M., Walker, T. W., et al. (2017). Reduced biomass burning emissions reconcile conflicting estimates of the post-2006 atmospheric methane budget. *Nature Communications*, 8(1), 2227.