TITLE

Global net biome CO_2 exchange predicted comparably well using parameter–environment relationships and plant functional types

RUNNING TITLE

Environmental filtering versus PFTs

AUTHORS

Caroline A. Famiglietti[1,*] (ORCID iD 0000-0002-6073-0457), Matthew Worden[1], Gregory R. Quetin[2], T. Luke Smallman[3], Uma Dayal[1], A. Anthony Bloom[4], Mathew Williams[3], & Alexandra G. Konings[1]

[1] Department of Earth System Science, Stanford University, Stanford, CA, USA
[2] Department of Geography, University of California at Santa Barbara, Santa Barbara, CA, USA
[3] School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, Edinburgh, Scotland, UK
[4] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

* Corresponding author: *cfamigli@stanford.edu*

ABBREVIATED ABSTRACT

Despite their importance for understanding the role of terrestrial ecosystems in a changing climate, forecasts of net biome CO_2 exchange are hindered by uncertainty in model parameters. Here, we compare the traditional plant functional type (PFT)-based parameterization approach to a novel top-down, machine learning-based “environmental filtering” (EF) approach. We find that the EF-based approach matches or outperforms the PFT-based approach at a narrow majority of vegetated pixels across the globe.

KEYWORDS

Terrestrial biosphere modeling; parametric uncertainty; environmental filtering; trait–environment relationships; machine learning; plant functional types

1. ABSTRACT

Accurate estimation and forecasts of net biome CO_2 exchange (NBE) are vital for understanding the role of terrestrial ecosystems in a changing climate. Prior efforts to improve NBE predictions have predominantly focused on increasing models' structural realism (and thus complexity), but parametric error and uncertainty are also key determinants of model skill. Here, we investigate how different parameterization assumptions propagate into NBE prediction errors across the globe, pitting the traditional plant functional type (PFT)-based approach against a novel top-down, machine learning-based "environmental filtering" (EF) approach. To do so, we simulate these contrasting methods for parameter assignment within a flexible model–data fusion framework of the terrestrial carbon cycle (CARDAMOM) at global scale. In the PFT-based approach, model parameters from a small number of select locations are applied uniformly within regions sharing similar land cover characteristics. In the EF-based approach, a pixel's parameters are predicted based on underlying relationships with climate, soil, and canopy properties. To isolate the role of parametric from structural uncertainty in our analysis, we benchmark the resulting PFT-based and EF-based NBE predictions with estimates from CARDAMOM's Bayesian optimization approach (whereby "true" parameters consistent with a suite of data constraints are retrieved on a pixel-by-pixel basis). When considering the mean absolute error of NBE predictions across time, we find that the EF-based approach matches or outperforms the PFT-based approach at 55% of pixels—a narrow majority. However, NBE estimates from the EF-based approach are susceptible to compensation between errors in component flux predictions, and predicted parameters can align poorly with the assumed "true" values. Overall, though, the EF-based approach is comparable to conventional approaches and merits further investigation to better understand and resolve these limitations. This work provides insight into the relationship between TBM performance and parametric uncertainty, informing efforts to improve model parameterization via PFT-free and trait-based approaches.

2. INTRODUCTION

The balance of carbon (C) fluxes entering and exiting the terrestrial biosphere—represented by net biome exchange, or NBE—directly influences the magnitude of future climate change by controlling how quickly carbon dioxide accumulates in the atmosphere *(Tans et al., 1990; Heimann & Reichstein, 2008)*. Projections of terrestrial ecosystems' behavior by process-based models can therefore play vital roles in setting future land management, conservation, and restoration priorities. However, such projections remain highly uncertain, as evidenced by the inability of most state-of-the-art terrestrial biosphere models (TBMs) to converge even on whether the land surface will act as a net sink or source of carbon by the end of the century *(Friedlingstein et al., 2013; Arora et al., 2020)*.

This spread in future TBM projections is the result of several factors, including uncertainty in the future trajectory of anthropogenic emissions and poor characterization of the climate system's internal variability. However, both are overshadowed by the role of model uncertainty itself *(Lovenduski & Bonan, 2017; Bonan & Doney, 2018)*. Indeed, how best to structure (*e.g.,* mathematically represent the functional forms of different ecological or hydrological processes and feedbacks; *Huntzinger et al., 2017)* and parameterize (*e.g.,* assign ecosystem "traits", such as leaf lifespan or leaf mass per area) a given model such that both realism and computational tractability are adequately preserved is a persistent and much debated challenge. (For simplicity, we will use the terms *traits* and *parameters* interchangeably throughout the remainder of this paper, although the former can be considered a subset of the latter, which encapsulates any time-invariant model coefficient.)

Most model development efforts have traditionally focused on increasing the realism of models' process representations (*e.g.,* by increasing structural complexity; *Luo et al., 2015; Fisher*

& Koven, 2020), but over-generalized and/or poorly determined model parameters also contribute to model uncertainty *(Prentice et al., 2015; Raczka et al., 2018)*. For example, in a model intercomparison across several biomes, *Famiglietti et al. (2021)* showed that making C cycle models more structurally realistic can actually decrease predictive accuracy if parameters are not accurately determined. Furthermore, using the ORCHIDEE TBM, *Mahmud et al. (2021)* found that optimizing parameters corrects the underestimation of modeled dryland net ecosystem CO_2 exchange. Parametric uncertainty has also been shown to dominate over structural uncertainty in model forecasts of both biomass and forest succession on regional scales *(Shiklomanov et al., 2020; Smallman et al., 2021)*. Thus, the need for improvements in model parameterization is becoming increasingly apparent. However, how best to do so remains opaque, in part because of the technical challenges and computational needs involved in optimizing parameters in complex land models *(MacBean et al., 2016; Ma et al., 2022)*.

Indeed, given the overwhelming inter- and intra-ecosystem diversity present across the land surface, parameterizing a global model requires making simplifying assumptions. Perhaps the most common parameterization assumption employed in nearly all current TBMs involves the use of plant functional types (PFTs), whereby parameters are assumed to be identical within regions sharing similar vegetation or land cover characteristics *(DeFries et al., 1995; Wullschleger et al., 2014; Poulter et al., 2015)*. This approach has clear and nontrivial benefits from a computational efficiency/tractability standpoint but is far from realistic. Research shows that actual plant traits can vary as much within a single PFT as between many different ones *(van Bodegom et al., 2012)*. Accordingly, carbon residence times and plant allocation strategies are poorly characterized by PFTs *(Bloom et al., 2016)*. While awareness of the uncertainties resulting from this overgeneralization is growing *(e.g., van Bodegom et al., 2014; Hartley et al., 2017; Thomas et al.,* *2019; Anderegg et al., 2021; C. G. Jung & Hararuk, 2022)*, underlying PFT-based hypotheses still remain ubiquitous in today's large-scale models.

Recently, novel approaches for generating spatially variable estimates of model parameters have been proposed to counter the limitations of static PFTs. In particular, the theory of "environmental filtering" (EF) posits that parameters are inherently predictable based on local climate, soil, and canopy properties—that is, the environment "filters" the vegetation traits that can exist in any particular place *(e.g., Joswig et al., 2022)*. Indeed, macroclimatic and biophysical factors like temperature, atmospheric aridity, water supply and nutrient availability strongly impact the strategies by which plants grow, allocate resources, and respond to stress *(e.g., Woodward, 1987)*. In practice, this concept—which broadly underlies certain large-scale predictive ecological frameworks like FLUXCOM *(M. Jung et al., 2020)*—is implemented by deriving mathematical relationships between community mean traits and environmental covariates *(e.g., Ordoñez et al., 2009; Chaney et al., 2016; Butler et al., 2017; Moreno-Martínez et al., 2018; Peaucelle et al., 2019; Boonman et al., 2020; Qian et al., 2021)*. However, while recent work focusing on a small subset of model parameters shows that these flexible, data-driven EF relationships can be feasibly implemented directly within large-scale TBMs *(Verheijen et al., 2013, 2015; Walker et al., 2017)*, the degree to which such an approach may impact the quality of simulated carbon fluxes—including NBE predictions—is not known. For example, although *Walker et al. (2017)* compared modeled photosynthesis rates resulting from an EF-based parameterization of the maximum photosynthetic carboxylation capacity (V_{cmax}) to three indirect proxies of gross primary productivity (GPP), those proxies are themselves highly uncertain, and only a single trait and a single carbon flux were considered.

While implementations of the EF hypothesis represent a promising avenue for introducing more realistic trait variation within TBMs, they face several key challenges *(Anderegg et al., 2021)*. First, the consistency of these relationships across taxonomic and ecological scales has been questioned *(Anderegg et al., 2018)*, and their ability to capture true ecological niche differences may be limited *(Kraft et al., 2015)*. A second issue involves the representativeness of the trait observations used to derive the EF relationships themselves. In situ parameter observations are useful but not a panacea. Measurements of plant traits are sparse relative to the heterogeneity and extent of terrestrial ecosystems *(Sandel et al., 2015)*, and some measurements are not compatible or easily reconcilable with model structure (*i.e.,* limited model representations of natural vertical heterogeneity, functional diversity, and more can make direct comparison nearly impossible). Other parameters are physically unobservable *(e.g.,* empirical coefficients such as the fraction of carbon lost to growth respiration; *Shiklomanov et al., 2020; Smith et al., 2020)*. Because of this, prior studies—which we classify as "bottom-up" *(e.g., Verheijen et al., 2013, 2015)*—were restricted by the availability and coverage of training data needed for model development. Most built EF relationships using in situ trait measurements from the TRY database *(Kattge et al., 2020)*, which, while expansive and ever-growing, contains significant spatial and species-related biases (*e.g.,* relatively few observations in the tropics and boreal regions; *Sandel et al., 2015; Schimel et al., 2015)*. Thus, it is not immediately clear whether EF-based predictions can reliably outperform those resulting from more classical PFT-based assumptions. Addressing this question, however, is necessary to determine if and how EF approaches can support the development of the next generation of TBMs.

To do so, we leverage the CARbon Data MOdel framework (CARDAMOM; *Bloom & Williams, 2015; Bloom et al., 2016)*, a Bayesian model–data fusion system built around an intermediate-complexity ecosystem model (DALEC; *Williams et al., 1997, 2005)* that is conceptually like most TBMs and produces similar carbon dynamics *(Quetin et al., 2020)*. Here, CARDAMOM provides dual benefits. First, CARDAMOM's flexible structure allows for straightforward implementation of either PFT-based or EF-based parameterization assumptions into DALEC. Second, it can retrieve the model's "true", or optimal, parameters at every pixel across the land surface—specifically, those consistent with a suite of remotely sensed and other global observational constraints (and their uncertainties) synthesized in a Bayesian inversion approach. CARDAMOM therefore provides a set of realistic "top-down", observationally informed parameter estimates across the globe, avoiding the large spatial biases of bottom-up trait datasets. Taken together, these two features allow us to benchmark PFT-based and EF-based DALEC models using CARDAMOM's wall-to-wall parameter retrievals and corresponding monthly, observationally constrained NBE predictions over the period 2000–2015. Hereafter, we refer to these CARDAMOM-derived benchmarks as "optimal". Because DALEC's model structure and forcing data remain fixed across all three simulations, NBE errors can be interpreted as wholly attributable to differences in parameterization. Overall, this study tests the dependence of C cycle prediction accuracy on parameterization assumption in a global context and demonstrates the potential of trait-based and PFT-free alternatives for reducing parametric uncertainty.

3. MATERIALS & METHODS

3.1. Overview. Using a set of realistic, observationally informed parameter retrievals and corresponding optimal C cycle stock and flux estimates, we performed a global, multi-decadal simulation experiment (4°×5° spatial resolution over the record 2000–2015) that tested the predictive capacity of PFT- and EF-based DALEC models to estimate NBE.

3.2. Modeling framework and parameter optimization. We used CARDAMOM *(Bloom & Williams, 2015; Bloom et al., 2016)* to conduct our parameterization experiments. CARDAMOM is a model–data fusion (MDF) system that uses a Bayesian inversion approach to constrain the parameters and initial conditions of an intermediate-complexity terrestrial ecosystem model with a suite of available satellite remote sensing observations (Table 1). CARDAMOM's underlying ecosystem model is called Data Assimilation Linked Ecosystem Carbon (DALEC; *Williams et al., 1997)*. Here, we use DALEC version C2 *(Bloom et al., 2020; Quetin et al., 2020; Famiglietti et al., 2021)* as the basis for our analysis. The model includes a coupled water cycle and uses 33 parameters governing ecosystem processes and defining the initial conditions of four live biomass pools and two dead organic matter pools. Further details of the model's structure are provided in *Famiglietti et al. (2021).*

CARDAMOM's MDF approach is summarized by Bayes' theorem:

$$p(\boldsymbol{y}|\boldsymbol{O}) \propto p(\boldsymbol{y}) \cdot p(\boldsymbol{O}|\boldsymbol{y}), \qquad (1)$$

where $p(\boldsymbol{y}|\boldsymbol{O})$ is the posterior probability distribution of model parameters $\boldsymbol{y}$ as informed by observations $\boldsymbol{O}$, $p(\boldsymbol{y})$ is the prior probability distribution of parameters $\boldsymbol{y}$, and $p(\boldsymbol{O}|\boldsymbol{y})$ is proportional to the likelihood of the observations $\boldsymbol{O}$ given $\boldsymbol{y}$. The posterior distribution $p(\boldsymbol{y}|\boldsymbol{O})$ is sampled using an adaptive proposal Metropolis-Hastings Markov Chain Monte Carlo (MCMC) approach. The prior distribution $p(\boldsymbol{y})$ encapsulates each model parameter's prior probability density function alongside a set of ecological and dynamical constraints (EDCs) that impose conditions on inter-relationships between parameters based on known ecological theory, as described in *Bloom & Williams (2015)* and *Famiglietti et al. (2021)*. The likelihood is derived such that

$$p(\boldsymbol{O}|\boldsymbol{y}) = e^{-\frac{1}{2}\sum_{i=1}^{n}(M_i-O_i)^2/\sigma_i^2}, \qquad (2)$$

where O_i is the *i*th observation, M_i is the corresponding modeled quantity at timestep *i*, and σ_i^2 is the *i*th error variance for each observation.

The set of observational constraints used in this analysis (*i.e.,* for the retrieval of DALEC's optimal model parameters), along with corresponding uncertainties, is listed in Table 1. It consists of several independent datasets aimed to constrain different carbon fluxes and pools. These include net biome exchange (NBE) estimates from the CMS-Flux atmospheric inversion system *(J. Liu et al., 2017, 2021)*, leaf area index (LAI) from MODIS *(Myneni et al., 2002)*, solar induced fluorescence (SIF) from GOSAT *(Frankenberg et al., 2011)*, soil organic matter (SOM) from SoilGrids *(Poggio et al., 2021)*, above- and below-ground biomass (ABGB) from *Saatchi et al. (2011)*, and fire C emissions from an inversion approach *(Bowman et al., 2017; Worden et al., 2017)*. Our analysis is performed at 4°×5° spatial resolution (928 total land pixels), which is the scale of the CMS-Flux NBE dataset. We chose to include the CMS-Flux dataset at the expense of higher spatial resolution because NBE integrates all aspects of the carbon cycle and, due to its connection to several model processes, is expected to exert a primary control over CARDAMOM's parameter retrievals and corresponding carbon fluxes *(Famiglietti et al., 2021)*.

To characterize the observational uncertainty of the NBE data, we took a novel approach compared to previous CARDAMOM studies *(e.g., Bloom et al., 2020; Quetin et al., 2020)*.

Rather than assigning a single, global average value to represent the observational uncertainty of NBE, here we introduced an additional model "parameter" to retrieve pixel-by-pixel uncertainty values (bringing the total number of parameters to 34). Further details of the uncertainty retrieval approach are provided in the supporting information (Text S1).

CARDAMOM typically runs in a two-stage process. First, in the "parameter assignment" stage, CARDAMOM retrieves location-specific optimal parameters (with uncertainty) for the DALEC model according to a suite of data constraints, as described above. Second, in the "forward run" stage, it produces monthly time series of carbon fluxes and pools by running DALEC forward in time with those parameter ensembles (*i.e.,* 1000 parameter samples from $p(\boldsymbol{y}|\boldsymbol{O})$). The forward runs are forced by a set of meteorological drivers from the combined data sets from Climate Research Unit (CRU) and reanalysis data from National Centers for Environmental Prediction (NCEP), or CRUNCEP *(Kalnay et al., 1996)*.

As described, CARDAMOM's inversion approach allows for the robust retrieval of a range of C cycle outcomes integrating the information content, quantity, and quality of its available data constraints. Due to this dependence, however, there is potential for its estimates to be poorly constrained when observations are temporally sparse and/or uncertain. For this reason, we introduced a filter requiring that the 25^{th}-75^{th} percentile range of a given pixel's optimal NBE ensemble not exceed the local NBE variability (*i.e.,* standard deviation across time). Pixels not satisfying this filter were omitted from the analysis (n = 138).

We used the resulting, strongly constrained optimal NBE predictions as benchmarks for estimates from alternatively parameterized (*i.e.,* EF-based and PFT-based) DALEC models. To derive and implement the EF-based and PFT-based parameterization assumptions, we amended CARDAMOM's first stage (parameter assignment) as described in the following sections, and then conducted additional forward runs with those alternative parameter sets.

3.3. EF-based parameterization approach. The environmental filtering approach relies on the expectation that climate, soil, and canopy properties determine the distribution of ecological traits—and therefore model parameters—across space, so that they can be used as predictors in a statistical model. Here, we implemented such assumptions across the globe using climate data from CRUNCEP, soil information from the SoilGrids project, and remotely sensed canopy and other data (Table 2). These predictors, or input features, are chosen to describe as many aspects of ecosystem structure and function as possible, and largely align with those used in previous environmental filtering applications *(e.g., Verheijen et al., 2013, 2015)*.

We aimed to produce highly skilled EF predictions that could result from complex, potentially nonlinear inter-relationships between features and targets. Accordingly, we trained a set of random forest regression models to learn the relationships between these environmental covariates and model parameters. Here, each of DALEC's model parameters was predicted independently (*e.g.,* one random forest model per parameter). Although these relationships may not be sufficiently parsimonious for straightforward inclusion in TBMs, they represent a meaningful upper bound on the potential complexity of EF-based assumptions and predictive schemes. Furthermore, this approach also reduces the need to rigorously determine the optimal balance between an EF-based model's tractability and predictive skill, which is beyond the scope of this study.

For each parameter (regression model), our model selection approach consisted of a feature selection analysis, a grid search-based hyperparameter tuning step, and a 10-fold cross-validation procedure. The feature selection analysis allowed us to assess train/test error as a function of the number of features available to the EF model (Fig. S1). Specifically, we determined the optimal number of features for each regression model. For example, the minimum test error for the SOM turnover rate parameter is observed when 18 features are included in the model. To derive an EF-based parameter set for a given pixel, we extracted the corresponding parameter prediction from each optimal regression model. Note that initial conditions for each carbon or water pool, as well as dates of leaf onset and leaf fall, are treated differently than other parameters, as described in Sec. 3.5.

3.4. PFT-based parameterization approach. Plant functional types (PFTs) are broad groupings of vegetation into classes with similar characteristics (*e.g.,* needle-leaf evergreen, broad-leaf deciduous, tundra, and so on; *DeFries et al., 1995)*. Here we emulated a common approach for PFT-based parameterization in large-scale models, whereby ecosystem parameters observed at a select number of ground locations are assumed to be sufficiently representative of the entire PFT *(e.g., Bonan et al., 2012)*. Specifically, we employed what we refer to as a "representative pixel" approach, using the European Space Agency's GlobCover land cover map (V2.3) as the basis for our PFTs.

The GlobCover product, available at 300m spatial resolution, provides a discrete classification of each land surface pixel into one of 23 land cover classes, or PFTs *(Arino et al., 2012)*. To more closely align with the level of detail in many current TBMs *(Bastrikov et al., 2018; Harper et al., 2018; Reick et al., 2021)*, we reduced these 23 classes to 9 broad groupings (Table S2). We first determined each 4°×5° pixel's fractional PFT composition by summing the (aggregated) GlobCover classifications contained within it. That is, we computed PFT fractions for each coarse-scale analysis pixel based on the fine-resolution GlobCover data. We then identified the pixels with the largest fractional cover of each PFT. For example, for the evergreen needleleaf forest class, we found the CARDAMOM pixels with the greatest percentage of area covered by evergreen needleleaf vegetation. We refer to these relatively homogeneous locations as “representative pixels”. In the main results of this study, we used a maximum of 5 representative pixels for each PFT. The representative pixels’ relevant PFT fractions generally ranged between 60-100%—a strong majority (Fig. S2). The sole exception is the mixed forest class, whose representative pixels contained only 30-45%; we only used those pixels containing a plurality of mixed forest. The mixed forest class is relatively rare, comprising less than 10% of any given pixel (not shown).

Finally, following the assumption that parameter estimates can be retrieved locally and applied broadly among similar sites, we aggregated CARDAMOM’s observationally constrained ensembles by randomly sampling 1000 members (with the exception of initial conditions and phenological dates; see Sec. 3.5) across each group of representative pixels to yield parameter sets for each PFT. Note that our approach can be viewed as relatively generous given that it relies on pixel homogeneity (rather than on ground data availability, as in a typical TBM, which may not ensure representativeness) for the assignment of PFT-based parameter sets.

3.5. Calculation of initial conditions (ICs) and phenological dates for EF and PFT approaches.

Most land surface and terrestrial biosphere models set the initial states of their carbon pools based on a “spin-up” to steady state, which can be unrealistic and introduce uncertainty *(Schwalm et al., 2019; Bonan et al., 2021)*. To isolate only the effects of alternative parameterization approaches on NBE predictions, here we leveraged CARDAMOM's ability to statistically derive realistic ICs for any set of model parameters. Specifically, after developing the EF- and PFT-based parameter sets as described in Secs. 3.3 and 3.4, we performed additional CARDAMOM optimization runs while holding all non-IC parameters constant at their EF- or PFT-predicted values (*i.e.,* so that only ICs are estimated; 7 of DALEC's 34 parameters). For the EF case, this amounted to one optimization run at each pixel, but for the PFT case, this necessitated one optimization run per PFT at each pixel.

We also took the same approach to re-optimize each pixel's leaf onset and leaf fall date parameters, which influence DALEC's simulation of phenology, in both the PFT- and EF-based models. To understand why this is necessary, consider the case in which two representative pixels for a given PFT exist in different hemispheres. Simply aggregating leaf onset or leaf fall dates (numeric values between 0 and 365) across these two pixels would be problematic due to the reversal of growing seasons between hemispheres.

Overall, then, the remaining 25 of DALEC's 34 parameters (74%) are the result of a random forest prediction (in the EF-based model) or an aggregation across representative pixels (in the PFT-based model).

3.6. ***Analysis.*** Monthly NBE time series used in our analysis were created by running DALEC forward with the retrieved ICs and corresponding optimal, EF-based, or PFT-based parameter set for all vegetated pixels satisfying the ensemble range filter (Sec. 3.2) across the land surface over the period 2000–2015. We defined vegetated pixels as those containing less than 50% barren or sparse land cover. In the PFT case, we took weighted averages of the resulting flux predictions based on each pixel's PFT fractions to yield the final time series for analysis. Schematic diagrams summarizing the EF-based (Sec. 3.3) and PFT-based (Sec. 3.4) modeling approaches are presented in Fig. 1, and an example of a pixel's simulated NBE time series resulting from the optimal, EF-based, and PFT-based approaches is shown in Fig. 2.

To parse the relative strengths and weaknesses of the alternatively parameterized models, we first evaluated the mean absolute error and Pearson correlation of a pixel's NBE time series (relative to the optimal predictions rather than to observations, so that errors are attributable only to parametric and not structural uncertainties). We also performed time series decomposition analyses using moving averages (implemented using Python's StatsModels package) to compare the ability of each model to capture features like the interannual variability, trend, and seasonal cycle of NBE. We then investigated several potential controls on the models' NBE error distributions across space using measures of variance explained (*i.e.,* coefficient of determination in a regression framework). These controls included parameter prediction accuracy (for the EF-based model), as well as the uncertainty of CARDAMOM's retrievals. For this analysis, we decomposed NBE into its component fluxes to understand the frequency and mode of interacting errors (*i.e.,* whether errors in component flux predictions tend to compound or compensate in yielding the net flux, NBE). Note that NBE in DALEC and other ecosystem models is determined by calculating the difference between R_{eco} (carbon release through both autotrophic and heterotrophic respiration) and GPP (carbon uptake by plants), along with the potential flux of carbon to the atmosphere resulting from fires.

4. ***RESULTS***

4.1. Skill of EF-based parameter prediction

We observed significant variability in the ability of EF to predict CARDAMOM's optimal model parameters. Across all parameters, the average percent RMSE for EF predictions (relative to the optimal parameter retrievals) is 44% with a standard deviation of 33% (average $R^2 = 0.41 \pm 0.18$). This relatively high average error is largely driven by parameters describing fire and combustion, which, at ~84%, are nearly twice as poorly predicted as any other parameter (Fig. 3a). These error-prone parameters include combustion fractions for DALEC's different carbon pools, which the model couples with observations of burned area to predict total fire carbon emissions *(Quetin et al., 2020).* By contrast, parameters related to phenology (*e.g.,* leaf lifespan), canopy structure (*e.g.,* leaf carbon mass per area) and canopy function (*e.g.,* canopy efficiency, a proxy for nitrogen use efficiency) are the most predictable, with errors on the order of 20%. Parameters describing soil respiration, carbon allocation, water cycling, and turnover are predicted with intermediate skill (*i.e.,* in the range of 30-50%). These patterns across parameters and parameter groups reflect the differential descriptiveness of available environmental covariates used as predictors in the random forest framework (Sec. 3.3). Still, only a minority of parameters are predicted with $R^2 \geq 0.5$ (Fig. 3b), which is consistent with prior EF studies. *Verheijen et al. (2013)* achieved an average adjusted R^2 of 0.40 for bottom-up predictions of specific leaf area (SLA), V_{cmax}, and the maximum electron transport rate (J_{max}) across 8 vegetation types, while *Butler et al. (2017)* found an average pseudo-R^2 of 0.34 when predicting SLA, leaf nitrogen concentration, and leaf phosphorus concentration with 9 increasingly complex predictive models. Taken together, these results demonstrate that EF-based parameter errors can remain stubbornly large despite comprehensive training information and a nonlinear predictive scheme.

4.2. Effects of EF-based and PFT-based parameterization assumptions on NBE performance

On average, the EF-based assumptions yield comparable modeled NBE performance to the PFT-based assumptions, based on mean absolute error (MAE) relative to a given pixel's optimal predictions across the entire time series (Fig. 4). The global average NBE MAE is 0.42 ± 0.34 gC m^{-2} day^{-1} for the EF-based model and 0.39 ± 0.28 gC m^{-2} day^{-1} for the PFT-based model. The two approaches produce some similar error hotspots, such as in Northeast China and parts of the eastern United States (Fig. 4a-b). Indeed, NBE errors tend to scale with gradients of climate and vegetation (Fig. S3). Specifically, higher errors are observed in warmer, wetter places, and errors increase as the variability in month-to-month temperature and radiation declines. Ecosystems with denser vegetation (*e.g.,* greater average LAI and ABGB) are also more error prone. These patterns align with the error hotspots observed across tropical Africa, for instance (Fig. 4a-b). Overall, using the MAE metric, the EF-based model can match or outperform the PFT-based model at 55% of pixels, while it produces strictly less accurate NBE predictions at 45% of pixels (Fig. 4c-d). This behavior is mirrored when considering the Pearson correlation between a given EF-based or PFT-based monthly NBE time series and the optimal estimate; in that case, the EF-based model produces comparable or superior predictions at 63% of pixels and poorer estimates at 37% of pixels (Fig. S4). That is, for any given vegetated pixel and across multiple metrics, NBE simulated using an EF-based approach is likely to capture the optimally parameterized NBE fluxes just as well as—if not better than—that simulated using a PFT-based approach.

To better understand the nature of each model’s performance, we decomposed the NBE time series at each pixel, isolating its overall mean, interannual variability, and de-trended seasonal cycle. First, by comparing the “position” (*i.e.,* nearest percentile) of an EF-based or PFT-based mean NBE estimate within that pixel’s optimal mean NBE ensemble, we found that both the EF-based model and PFT-based model are likely to accurately capture the mean across the time series. While any given EF-based or PFT-based mean NBE estimate often aligns with the center of the optimal ensemble, indicating high accuracy, the EF-based approach is more likely to underestimate mean NBE (Fig. 5a, greater density below $x = 50$). The PFT-based model also approximates the interannual variability of NBE—calculated as the standard deviation of the annually averaged fluxes—more closely than the EF-based model, which is slightly too variable from year to year (Fig. 5b).

Both the EF-based and PFT-based models capture annual average NBE moderately well (Fig. 5c), and the seasonal cycle almost perfectly (Fig. 5d). To see this, we computed the Pearson correlation between a given pixel’s annually averaged optimal NBE or de-trended seasonal cycle and its EF-based or PFT-based counterpart. We find that there are pixels for which both EF-based and PFT-based NBE annual averages negatively correlate with those from the optimal model (bottom left quadrant in Fig. 5c); many such pixels align spatially with the models’ MAE hotspots (Fig. S5). However, the opposite is far more likely. Generally, both the EF-based and PFT-based estimates of annually averaged NBE correlate positively with the optimal one (greater density of points in top right quadrant than in all other quadrants in Fig. 5c). Additionally, both model variants nearly always capture the optimal model’s seasonal cycle correctly (very high point density in top right quadrant of Fig. 5d; shown across space in Fig. S6).

4.3 Controls on EF-based and PFT-based model errors

Although the EF-based model shows comparable or better performance than the PFT-based model across several dimensions, the relative skill of the two models shows significant spatial variability, the driving factors of which are not clear. That is, Fig. 4d begs the question: what factors determine variations in the EF-based and PFT-based models' relative performance, particularly across space? Understanding where—and why—the EF-based model falters in predicting NBE can help to inform future iterations of the approach.

To do so, we tested two (potentially overlapping) hypotheses as possible controls on the models' variable performance across different pixels. These hypotheses involve (a) how precisely the EF-based model's parameter predictions match the "truth" (*i.e.,* the optimal parameters) at a given location, and (b) how uncertain CARDAMOM's optimal retrievals themselves are. For this analysis, we expanded our lens to also consider the predictability of NBE's component fluxes, which critically influence the dynamics of the net flux. We focused on GPP and R_{eco} fluxes, given that errors in predicting fluxes from fires are far smaller in magnitude (Fig. S7).

First, it seems feasible that the more a given EF-based parameter set differs from the "true" values, the less accurate any of its resulting model predictions will be. Contrary to this hypothesis, though, we find no direct relationship between the EF model's GPP, R_{eco}, or NBE performance and the average precision of a given EF-based parameter set (relative to the corresponding optimal parameter set), suggesting that individual parameter accuracy is a necessary but insufficient control on its performance. Indeed, a multiple linear regression with access only to information on the quality of the EF-based prediction for each parameter across pixels explains at most 7% of the variance in the EF-based model's GPP, R_{eco}, and NBE errors (coefficient of determination, R^2) (Fig. 6). Here, quality is measured through the "position", or closest percentile, of an EF-based parameter prediction within the corresponding optimal posterior distribution, where proximity to the median indicates high accuracy—a measure chosen to normalize parameter error across different pixels even as the true parameter value varies.

Second, how strongly are model errors dictated by uncertainty in CARDAMOM's optimal retrievals? That is, because CARDAMOM has its own limitations in determining the "true" NBE (*e.g.,* the availability and accuracy of data constraints used in the optimization can vary across space, and flows of carbon may be inherently less predictable at some pixels than others), our assessment of the alternatively parameterized models' predictions may reflect this uncertainty.

For GPP and R_{eco} fluxes, both the EF-based and PFT-based models perform more poorly when CARDAMOM's optimal retrievals are less strongly constrained and more uncertain (*i.e.,* when the ensemble of optimal flux predictions is wider). The mean interquartile range (IQR; 25th-75th percentile) of CARDAMOM's optimal GPP ensembles across pixels explains 45% of the variance in the EF-based model's GPP errors and 64% of that in the PFT-based model's GPP errors (Fig. 7a). Similarly, CARDAMOM's R_{eco} IQR explains 41% of the variance in the EF-based model's R_{eco} errors and 60% of that in the PFT-based model's R_{eco} errors (Fig. 7b). Importantly, neither model appears significantly more sensitive than the other to CARDAMOM's IQR; for each flux, the slopes of the two regression lines are statistically indistinguishable (Fisher's z-test; $p < 0.01$). This indicates that our interpretation of model errors is not biased by the relationship between model performance and

CARDAMOM uncertainty. In effect, some pixels with large errors for GPP (or R_{eco}) may simply be those where the optimal GPP (R_{eco}) is so uncertain that mismatches between the EF-based or PFT-based GPP (R_{eco}) prediction and the GPP (R_{eco}) considered optimal are as much due to uncertainty in the latter as due to imperfect parameterization in the former.

When considering NBE, however, this relationship weakens markedly, with CARDAMOM's NBE IQR explaining only 21% and 34% of the EF-based model's and PFT-based model's NBE MAE variance, respectively (Fig. 7c). The discrepancy between the predictability of component versus net fluxes in the two models suggests the occurrence of significant compensating errors (Fig. 8). GPP and R_{eco} errors are generally larger in magnitude than NBE errors for both models (Fig. 8a), suggesting a greater absolute mismatch between component flux predictions than net flux predictions across approaches (Fig. 8b). On one hand, this is not unexpected given the relative sizes of the fluxes themselves. However, the skill of the EF-based model relative to the PFT-based model also declines when considering component fluxes (Fig. 8c). That is, while the EF-based model matches or outperforms the PFT-based model when predicting NBE at 55% of vegetated pixels, it does so at only 49% when predicting either GPP or R_{eco}—no longer a majority of pixels.

Taken together, these findings indicate a persistent error compensation effect, whereby larger errors in component flux predictions tend to "cancel out" to yield comparatively smaller errors in NBE (Fig. 8d). This effect is far more prevalent than the converse, whereby smaller errors in GPP and R_{eco} can compound to yield larger NBE errors. This suggests that the spatial pattern of NBE errors is strongly influenced by the frequency and degree of compensation between component fluxes. Critically, though, this behavior appears to impact the relative skill of EF-based and PFT-based predictions slightly differently. Indeed, the fraction of pixels showing equivalent or superior performance shrinks for the former model when considering net versus component fluxes, and grows for the latter model (Fig. 8c).

5. DISCUSSION

5.1. Implications for TBMs

The top-down EF-based hypotheses implemented here yielded NBE errors that matched or outperformed those from traditional PFTs at a sizable fraction of pixels (55%; Fig. 4d), suggesting that the introduction of more realistic trait variability in large-scale TBMs can help to improve predictions of its future behavior, as previously hypothesized *(Scheiter et al., 2013; van Bodegom et al., 2014; Matheny et al., 2017; Xu & Trugman, 2021)*. Overall, our findings support the growing paradigm shift away from the representation of static PFTs and towards the incorporation of realistic trait variability into large-scale TBMs *(van Bodegom et al., 2014; Bloom et al., 2016; Berzaghi et al., 2020; C. G. Jung & Hararuk, 2022; Y. Liu et al., 2022)*. EF-based hypotheses represent one promising and flexible approach for doing so, although they are not a panacea—PFT-based assumptions are still superior at nearly half of vegetated pixels in our analysis (45%; Fig. 4d). Although the drawbacks of PFTs are well-known, they are relatively easy to implement and have been used with reasonable success in TBMs for decades.

The close performance we observed between models nevertheless suggests that EF-based assumptions merit further investigation, particularly because implementing an EF-based parameterization in a TBM would require solving several open questions. These include whether and to what degree trait covariations *(e.g., Peaucelle et al., 2019)* should be explicitly preserved; whether different traits should be predicted based on fully independent filters; how complex or parsimonious EF regression models themselves should be; which environmental covariates are most relevant for predicting which traits; whether EF relationships should be included even if they contain little theoretical support; whether all traits benefit from EF-based assumptions or if a hybrid, super-predictive EF- and PFT-based approach can improve simulations; and so on. An additional consideration involves the mathematical interpretability and/or generality of EF relationships *(Kyker-Snowman et al., 2022)*, which depends on the specific predictive framework selected for analysis (*i.e.,* a machine learning-based approach is less interpretable than a simple linear regression). It is also not clear whether EF relationships developed offline can be used directly in different TBMs with unique structures and dependencies, or whether the parameters of the EF relationships themselves would need local tuning for each specific TBM to avoid compensating errors *(Koster et al., 2009; J-F Exbrayat et al., 2013)*.

5.2. Model performance

Unlike previous (bottom-up) implementations of EF, which focused on only a select few measurable traits and still maintained a generalized PFT paradigm *(Verheijen et al., 2013, 2015; Butler et al., 2017)*, our satellite-based machine learning approach predicts every one of DALEC's dozens of parameters independently and simultaneously. This is an extreme case in the context of large-scale TBMs, for which a step-by-step implementation may be more realistic. Indeed, it is possible that our efforts—which served as a "stress test" to understand the integrated feasibility of the EF approach—may overestimate the appropriate levels of complexity and nonlinearity required for optimal EF predictions. For example, while any potential errors stemming from inaccuracies in the EF-based parameter predictions (Fig. 3) were not substantial enough to consistently limit the skill of the EF-based model below that of the PFT-based one when predicting NBE, these inaccuracies indicate significant room for improvement regarding the characterization and predictability of environmental controls on parameter variability. Indeed, model parsimony *(Famiglietti et al., 2021)* should remain an important consideration in EF contexts, given that compensating errors can occur not only between component fluxes (Fig. 8) but also between parameters themselves *(Wu et al., 2019)* and/or between different modeled environmental feedbacks *(Huntzinger et al., 2017)*.

Still, the retention of skill at the flux level by the EF-based model (Figs. 4-5) despite its parameter errors is especially notable given that the simulation of PFTs implemented here is relatively generous. For example, our PFT-based parameterization relies on pixel homogeneity rather than ground data availability for the fundamental representativeness assumption (Sec. 3.4). It also includes arguably more degrees of freedom than what may be observed in a typical TBM; that is, the total number of "representative pixels" used in the aggregation includes a relatively broad sample of locations within each PFT, although the total number of PFTs considered here ($n = 9$) aligns reasonably well with current approaches *(Bastrikov et al., 2018; Harper et al., 2018; Reick et al., 2021)*.

5.3. Spatial error distributions & component flux compensation

We found a strong relationship between CARDAMOM's ensemble range (*i.e.,* uncertainty) and the predictive skill of both alternatively parameterized models (Fig. 7). On one hand, this demonstrates consistency between modeling approaches; places where even an optimally parameterized model is under-constrained are also those where the EF-based and PFT-based models perform poorly. This helps to explain why both alternatively parameterized models share error hotspots (Fig. 4a-b). Notably, though, it also indicates the sensitivity of the EF-based approach to training data quality and that of the PFT-based approach to assumed representative parameters.

In our study, both such factors are a direct function of the uncertainty of observational constraints used in the optimization, many of which are necessarily broad and uniform across space (Table 1). Additional attention and improvements to direct constraints on modeled GPP and R_{eco} performance, for instance, may be needed to ensure that changes to a given model's underlying parameterization indeed map to improvements in NBE. However, directly constraining R_{eco} requires information on its own component parts (autotrophic and heterotrophic respiration fluxes, R_a and R_h). Such data are particularly challenging to assemble across large scales due to their sparsity *(Bond-Lamberty, 2018)*. Accordingly, neither R_a nor R_h was directly constrained in the optimization approach used here. More broadly, for bottom-up studies (or traditional PFTs), this relationship relies on the accuracy, representativeness, and coverage of in situ trait measurements (*e.g., Sandel et al., 2015; Kattge et al., 2020)*. Overall, alternative model parameterization approaches would benefit significantly from targeted increases in observational data that can be used for training.

Still, the frequency of error compensation between GPP and R_{eco} fluxes in our models—as also observed more broadly by *Caen et al. (2021)* in the JULES and INLAND land surface models—indicates that improvements in parameter realism also have the potential to yield unintended consequences, such that increases in the predictability of net fluxes are not guaranteed. Indeed, the role of error interactions appeared as strong or stronger than other potential controls on NBE performance, including parameter precision (Figs. 6-7). In particular, the performances of both the EF-based and PFT-based model were influenced by error compensation in our study (Fig. 8) despite the greater realism of the former's parameterization. Thus, neither model's NBE performance can be interpreted independently from compensation (Fig. 4d). Accordingly, a focus on validating gross rather than net fluxes and on simultaneous testing with multiple independent observational datasets of different fluxes and pools (with well-defined uncertainties) is recommended when implementing novel EF-based assumptions in TBMs to reduce the effects of possible error compensation.

5.4. *Remaining uncertainties & limitations*

Fire- and combustion-related parameters and processes were particularly poorly characterized in our study (Fig. 3), despite the inclusion of data describing burned area (average and variability) within the feature space. Given the critical importance of fire in explaining the evolution and trajectory of the land carbon sink *(Jean-François Exbrayat et al., 2018; Yin et al., 2020)*, we expect the accuracy of long-term EF-based NBE forecasts to increase with an improved representation of fire-related processes. This need dovetails with recent efforts to generate fine-resolution maps of variables describing fire risk and vulnerability *(e.g., Forzieri et al., 2021)*, for example, which could be sourced as additional environmental covariates in future implementations of EF-based assumptions.

An additional uncertainty relates to the fact that several of DALEC's parameters are biophysically inter-related (*e.g.,* leaf lifespan and leaf mass per area; *Wright et al., 2004)* and thus co-vary, potentially indicating limitations of our EF-based approach to predict each parameter independently. Here, we derived a unique trait–environment relationship for each model parameter using a random forest regression (Sec. 3.3). This means that a true biophysical inter-relationship between parameters could theoretically be violated when EF schemes are fitted independently, leaving one parameter estimate incongruent with another.

Accordingly, future studies should consider a multi-dimensional predictive framework, wherein dependencies between parameters are inherently preserved, or an alternative approach to maintaining covariation between parameters *(Peaucelle et al., 2019)*.

Finally, despite our efforts to robustly assign initial conditions consistent with each EF-based or PFT-based parameter set, our implementation still has limitations. Given that small disparities in initial states (that is, carbon pool sizes) can produce significantly different trajectories *(Hawkins & Sutton, 2009; Bonan & Doney, 2018)*, it is possible that remaining initial condition uncertainty—perhaps along with the influence of other poorly determined parameters—may partially explain the sometimes divergent relationships we observed between the alternative models' and optimal model's annual average NBE (Figs. 5c and S5a-b). Such uncertainties, however, are also far from resolved in large-scale TBMs *(Hurtt et al., 2010; Thurner et al., 2014)*, where initial conditions are generally calculated based on spinning up the model to steady state, even though this assumption is likely unrealistic *(Sierra et al., 2017)*.

6. CONCLUSIONS

Overall, the top-down EF relationships and corresponding parameter predictions shown here represent a significant step forward in the characterization of trait–environment associations independent of in situ measurement availability. The results of this study highlight the potential for EF approaches to reduce NBE prediction errors and may inform efforts to incorporate increasingly diverse parameter representations into next-generation TBMs and future iterations of widely used multi-model ensembles. Expansions in the quantity and quality of Earth observation data from satellite remote sensing *(Schimel et al., 2019)*, advancements in the development of explainable/interpretable physics-based machine learning techniques for Earth system science *(Reichstein et al., 2019)*, and increases in computational resource efficiency *(Gupta et al., 2021)* may work in tandem to foster this transition.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Aeronautics and Space Administration under Grant No. 80NSSC21K1593 issued through the Future Investigators in NASA Earth and Space Science and Technology (FINESST) program. CAF was also supported by an ARCS graduate fellowship. AGK and MWo were supported by NSF DEB grant 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.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

7. REFERENCES

Anderegg, L. D. L., Berner, L. T., Badgley, G., Sethi, M. L., Law, B. E., & HilleRisLambers, J. (2018). Within-species patterns challenge our understanding of the leaf economics spectrum. *Ecology Letters*, *21*(5), 734–744.

Anderegg, L. D. L., Griffith, D. M., Cavender-Bares, J., Riley, W. J., Berry, J. A., Dawson, T. E., & Still, C. J. (2021). Representing plant diversity in land models: An evolutionary approach to make 'Functional Types' more functional. *Global Change Biology*.

Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P., & Van Bogaert, E. (2012). Global land cover map for 2009 (GlobCover 2009). European Space Agency (ESA) & Université catholique de Louvain (UCL ….

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., … Ziehn, T. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, *17*(16), 4173–4222. doi: 10.5194/bg-17-4173-2020

Bastrikov, V., MacBean, N., Bacour, C., Santaren, D., Kuppel, S., & Peylin, P. (2018). Land surface model parameter optimisation using in situ flux data: comparison of gradient-based versus random search algorithms \hack{\break} (a case study using ORCHIDEE v1.9.5.2). *Geoscientific Model Development*, *11*(12), 4739–4754. doi: 10.5194/gmd-11-4739-2018

Berzaghi, F., Wright, I. J., Kramer, K., Oddou-Muratorio, S., Bohn, F. J., Reyer, C. P. O., … Hartig, F. (2020). Towards a New Generation of Trait-Flexible Vegetation Models. *Trends in Ecology & Evolution*, *35*(3), 191–205. doi: https://doi.org/10.1016/j.tree.2019.11.006

Bloom, A. A., Bowman, K. W., Liu, J., Konings, A. G., Worden, J. R., Parazoo, N. C., … Schimel, D. S. (2020). Lagged effects regulate the inter-annual variability of the tropical carbon balance. *Biogeosciences*, *17*(24), 6393–6422. doi: 10.5194/bg-17-6393-2020

Bloom, A. A., Exbrayat, J.-F., 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 LP-1290. doi: 10.1073/pnas.1515160113

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), 1299–1315. doi: 10.5194/bg-12-1299-2015

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. doi: 10.1126/science.aam8328

Bonan, G. B., Lombardozzi, D. L., & Wieder, W. R. (2021). The signature of internal variability in the terrestrial carbon cycle. *Environmental Research Letters*, *16*(3), 34022.

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: Biogeosciences*, *117*(G2).

Bond-Lamberty, B. (2018). New techniques and data for understanding the global soil respiration flux. *Earth's Future*, *6*(9), 1176–1180.

Boonman, C. C. F., Benítez-López, A., Schipper, A. M., Thuiller, W., Anand, M., Cerabolini, B. E. L., … Higuchi, P. (2020). Assessing the reliability of predicted plant trait distributions at the global scale. *Global Ecology and Biogeography*, *29*(6), 1034–1051.

Bowman, K. W., Liu, J., Bloom, A. A., Parazoo, N. C., Lee, M., Jiang, Z., … Gurney, K. R. (2017). Global and Brazilian carbon response to El Niño Modoki 2011–2010. *Earth and Space Science*, *4*(10), 637–660.

Butler, E. E., Datta, A., Flores-Moreno, H., Chen, M., Wythers, K. R., Fazayeli, F., … Reich, P. B. (2017). Mapping local and global variability in plant trait distributions. *Proceedings of the National Academy of Sciences*, *114*(51), E10937 LP-E10946. doi: 10.1073/pnas.1708984114

Caen, A., Smallman, T. L., de Castro, A. A., Robertson, E., von Randow, C., Cardoso, M., & Williams, M. (2021). Evaluating two land surface models for Brazil using a full carbon cycle benchmark with uncertainties. *Climate Resilience and Sustainability*, e10.

Chaney, N. W., Herman, J. D., Ek, M. B., & Wood, E. F. (2016). Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. *Journal of Geophysical Research: Atmospheres*, *121*(22), 13–218.

DeFries, R. S., Field, C. B., Fung, I., Justice, C. O., Los, S., Matson, P. A., … Prentice, K. (1995). Mapping the land surface for global atmosphere-biosphere models: Toward continuous distributions of vegetation's functional properties. *Journal of Geophysical Research: Atmospheres*, *100*(D10), 20867–20882.

Exbrayat, J.-F., Pitman, A. J., Zhang, Q., Abramowitz, G., & Wang, Y.-P. (2013). Examining soil carbon uncertainty in a global model: response of microbial decomposition to temperature, moisture and nutrient limitation. *Biogeosciences*, *10*(11), 7095–7108.

Exbrayat, J.-F., 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*(12), 1776–1789. doi: https://doi.org/10.1029/2018GB005925

Famiglietti, C. A., Smallman, T. L., Levine, P. A., Flack-Prain, S., Quetin, G. R., Meyer, V., … Konings, A. G. (2021). Optimal model complexity for terrestrial carbon cycle prediction. *Biogeosciences*, *18*(8), 2727–2754. doi: 10.5194/bg-18-2727-2021

Famiglietti, Caroline (2022): Data for: Global net biome CO2 exchange predicted comparably well using parameter–environment relationships and plant functional types. figshare. Dataset. https://doi.org/10.6084/m9.figshare.21753935

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of Land Surface Models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4). doi: 10.1029/2018MS001453

Forzieri, G., Girardello, M., Ceccherini, G., Spinoni, J., Feyen, L., Hartmann, H., … Mauri, A. (2021). Emergent vulnerability to climate-driven disturbances in European forests. *Nature Communications*, *12*(1), 1–12.

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

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2013). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *Journal of Climate*, *27*(2), 511–526. doi: 10.1175/JCLI-D-12-00579.1

Gupta, S., Aga, D., Pruden, A., Zhang, L., & Vikesland, P. (2021). Data Analytics for Environmental Science and Engineering Research. *Environmental Science & Technology*, *55*(16), 10895–10907.

Harper, A., Wiltshire, A., Cox, P., Friedlingstein, P., Jones, C., Mercado, L., … Duran-Rojas, C. (2018). Vegetation distribution and terrestrial carbon cycle in a carbon-cycle configuration of JULES4.6 with new plant functional types. *Geoscientific Model Development Discussions*, 1–34. doi: 10.5194/gmd-2017-311

Hartley, A. J., MacBean, N., Georgievski, G., & Bontemps, S. (2017). Uncertainty in plant functional type distributions and its impact on land surface models. *Remote Sensing of Environment*, *203*, 71–89. doi: https://doi.org/10.1016/j.rse.2017.07.037

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108.

Heimann, M., & Reichstein, M. (2008). Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*, *451*(7176), 289–292. doi: 10.1038/nature06591

Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., … Zhao, F. (2017). Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions. *Scientific Reports*, *7*(1), 4765. doi: 10.1038/s41598-017-03818-2

Hurtt, G. C., Fisk, J., Thomas, R. Q., Dubayah, R., Moorcroft, P. R., & Shugart, H. H. (2010). Linking models and data on vegetation structure. *Journal of Geophysical Research: Biogeosciences*, *115*(G2).

Joswig, J. S., Wirth, C., Schuman, M. C., Kattge, J., Reu, B., Wright, I. J., … Mahecha, M. D. (2022). Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nature Ecology & Evolution*, *6*(1), 36–50. doi: 10.1038/s41559-021-01616-8

Jung, C. G., & Hararuk, O. (2022). Assimilation of NEON observations into a process-based carbon cycle model reveals divergent mechanisms of carbon dynamics in temperate deciduous forests. *Journal of Geophysical Research: Biogeosciences*, e2021JG006474.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., … Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. doi: 10.5194/bg-17-1343-2020

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. doi: 10.1175/1520-

0477(1996)077<0437:TNYRP>2.0.CO;2

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., … Wirth, C. (2020). TRY plant trait database – enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188. doi: 10.1111/gcb.14904

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J. (2009). On the nature of soil moisture in land surface models. *Journal of Climate*, *22*(16), 4322–4335.

Kraft, N. J. B., Godoy, O., & Levine, J. M. (2015). Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences*, *112*(3), 797 LP-802. doi: 10.1073/pnas.1413650112

Kyker-Snowman, E., Lombardozzi, D. L., Bonan, G. B., Cheng, S. J., Dukes, J. S., Frey, S. D., … Smith, N. G. (2022). Increasing the spatial and temporal impact of ecological research: A roadmap for integrating a novel terrestrial process into an Earth system model. Wiley Online Library.

Liu, J., Baskaran, L., Bowman, K., Schimel, D., Bloom, A. A., Parazoo, N. C., … Joiner, J. (2021). Carbon monitoring system flux net biosphere exchange 2020 (CMS-flux NBE 2020). *Earth System Science Data*, *13*(2), 299–330.

Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., … Eldering, A. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. *Science*, *358*(6360), eaam5690. doi: 10.1126/science.aam5690

Liu, Y., Flournoy, O., Zhang, Q., Novick, K. A., Koster, R. D., & Konings, A. G. (2022). Canopy Height and Climate Dryness Parsimoniously Explain Spatial Variation of Unstressed Stomatal Conductance. *Geophysical Research Letters*, *49*(15), e2022GL099339. doi: https://doi.org/10.1029/2022GL099339

Lovenduski, N. S., & Bonan, G. B. (2017). Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters*, *12*(4), 44020. doi: 10.1088/1748-9326/aa66b8

Luo, Y., Keenan, T. F., & Smith, M. (2015). Predictability of the terrestrial carbon cycle. *Global Change Biology*, *21*(5), 1737–1751. doi: 10.1111/gcb.12766

Ma, R., Xiao, J., Liang, S., Ma, H., He, T., Guo, D., … Lu, H. (2022). Pixel-level parameter optimization of a terrestrial biosphere model for improving estimation of carbon fluxes with an efficient model–data fusion method and satellite-derived LAI and GPP data. *Geosci. Model Dev.*, *15*(17), 6637–6657. doi: 10.5194/gmd-15-6637-2022

MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geosci. Model Dev.*, *9*(10), 3569–3588. doi: 10.5194/gmd-9-3569-2016

Mahmud, K., Scott, R. L., Biederman, J. A., Litvak, M. E., Kolb, T., Meyers, T. P., … MacBean, N. (2021). Optimizing Carbon Cycle Parameters Drastically Improves Terrestrial Biosphere Model Underestimates of Dryland Mean Net CO2 Flux and its Inter-Annual Variability. *Journal of Geophysical Research: Biogeosciences*, *126*(10), e2021JG006400.

Matheny, A. M., Mirfenderesgi, G., & Bohrer, G. (2017). Trait-based representation of hydrological functional properties of plants in weather and ecosystem models. *Plant*

Diversity, *39*(1), 1–12. doi: https://doi.org/10.1016/j.pld.2016.10.001

Moreno-Martínez, Á., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., … Running, S. W. (2018). A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sensing of Environment*, *218*, 69–88. doi: https://doi.org/10.1016/j.rse.2018.09.006

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., … Smith, G. R. (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.

Ordoñez, J. C., Van Bodegom, P. M., Witte, J.-P. M., Wright, I. J., Reich, P. B., & Aerts, R. (2009). A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. *Global Ecology and Biogeography*, *18*(2), 137–149. doi: 10.1111/j.1466-8238.2008.00441.x

Peaucelle, M., Bacour, C., Ciais, P., Vuichard, N., Kuppel, S., Peñuelas, J., … Viovy, N. (2019). Covariations between plant functional traits emerge from constraining parameterization of a terrestrial biosphere model. *Global Ecology and Biogeography*, *28*(9), 1351–1365. doi: 10.1111/geb.12937

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240. doi: 10.5194/soil-7-217-2021

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., … Peylin, P. (2015). Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative. *Geosci. Model Dev.*, *8*(7), 2315–2328. doi: 10.5194/gmd-8-2315-2015

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. *Atmos. Chem. Phys.*, *15*(10), 5987–6005. doi: 10.5194/acp-15-5987-2015

Qian, X., Liu, L., Croft, H., & Chen, J. (2021). Relationship between leaf maximum carboxylation rate and chlorophyll content preserved across 13 species. *Journal of Geophysical Research: Biogeosciences*, *126*(2), e2020JG006076.

Quetin, G. R., Bloom, A. A., Bowman, K. W., & Konings, A. G. (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*(3), e2019MS001889. doi: 10.1029/2019MS001889

Quetin, G. R., Famiglietti, C. A., Dadap, N., Bloom, A. A., Bowman, K., Diffenbaugh, N., Liu, J., Trugman, A., & Konings, A. G. Attributing past carbon fluxes to CO2 and climate change: respiration response to CO2 fertilization shifts regional distribution of the carbon sink. In revision.

Raczka, B., Dietze, M. C., Serbin, S. P., & Davis, K. J. (2018). What limits predictive certainty of long-term carbon uptake? *Journal of Geophysical Research: Biogeosciences*, *123*(12), 3570–3588.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat.

(2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. doi: 10.1038/s41586-019-0912-1

Reick, C. H., Gayler, V., Goll, D., Hagemann, S., Heidkamp, M., Nabel, J. E. M. S., … Wilkenskjeld, S. (2021). JSBACH 3-The land component of the MPI Earth System Model: documentation of version 3.2.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., … Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*(24), 9899 LP-9904. doi: 10.1073/pnas.1019576108

Sandel, B., Gutiérrez, A. G., Reich, P. B., Schrodt, F., Dickie, J., & Kattge, J. (2015). Estimating themissing species bias in plant trait measurements. *Journal of Vegetation Science*, *26*(5), 828–838. doi: 10.1111/jvs.12292

Scheiter, S., Langan, L., & Higgins, S. I. (2013). Next-generation dynamic global vegetation models: learning from community ecology. *New Phytologist*, *198*(3), 957–969. doi: 10.1111/nph.12210

Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., … Cox, P. (2015). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, *21*(5), 1762–1776. doi: 10.1111/gcb.12822

Schimel, D., Schneider, F. D., & Participants, J. P. L. C. and E. (2019). Flux towers in the sky: global ecology from space. *New Phytologist*, *224*(2), 570–584. doi: 10.1111/nph.15934

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

Shiklomanov, A. N., Bond-Lamberty, B., Atkins, J. W., & Gough, C. M. (2020). Structure and parameter uncertainty in centennial projections of forest community structure and carbon cycling. *Global Change Biology*, *n/a*(n/a). doi: 10.1111/gcb.15164

Sierra, C. A., Müller, M., Metzler, H., Manzoni, S., & Trumbore, S. E. (2017). The muddle of ages, turnover, transit, and residence times in the carbon cycle. *Global Change Biology*, *23*(5), 1763–1773.

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, *116*(G4). doi: 10.1029/2011JG001708

Smallman, T. L., Milodowski, D. T., Neto, E. S., Koren, G., Ometto, J., & Williams, M. (2021). Parameter uncertainty dominates C cycle forecast errors over most of Brazil for the 21st Century. *Earth Syst. Dynam. Discuss.*, *2021*, 1–52. doi: 10.5194/esd-2021-17

Smith, W. K., Fox, A. M., MacBean, N., Moore, D. J. P., & Parazoo, N. C. (2020). Constraining estimates of terrestrial carbon uptake: new opportunities using long-term satellite observations and data assimilation. *New Phytologist*, *225*(1), 105–112. doi: 10.1111/nph.16055

Tans, P. P., Fung, I. Y., & Takahashi, T. (1990). Observational Contrains on the Global Atmospheric Co2 Budget. *Science*, *247*(4949), 1431–1438. doi:

10.1126/science.247.4949.1431

Thomas, H. J. D., Myers-Smith, I. H., Bjorkman, A. D., Elmendorf, S. C., Blok, D., Cornelissen, J. H. C., … Prevéy, J. S. (2019). Traditional plant functional groups explain variation in economic but not size-related traits across the tundra biome. *Global Ecology and Biogeography*, *28*(2), 78–95.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., … Levick, S. R. (2014). Carbon stock and density of northern boreal and temperate forests. *Global Ecology and Biogeography*, *23*(3), 297–310.

Trabucco, A., & Zomer, R. (2019, January 18). Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. doi: 10.6084/m9.figshare.7504448.v3

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), 13733 LP-13738. doi: 10.1073/pnas.1304551110

van Bodegom, P. M., Douma, J. C., Witte, J. P. M., Ordoñez, J. C., Bartholomeus, R. P., & Aerts, R. (2012). Going beyond limitations of plant functional types when predicting global ecosystem–atmosphere fluxes: exploring the merits of traits-based approaches. *Global Ecology and Biogeography*, *21*(6), 625–636. doi: 10.1111/j.1466-8238.2011.00717.x

Verheijen, L. M., Aerts, R., Brovkin, V., Cavender-Bares, J., Cornelissen, J. H. C., Kattge, J., & van Bodegom, P. M. (2015). Inclusion of ecologically based trait variation in plant functional types reduces the projected land carbon sink in an earth system model. *Global Change Biology*, *21*(8), 3074–3086. doi: 10.1111/gcb.12871

Verheijen, L. M., Brovkin, V., Aerts, R., Bönisch, G., Cornelissen, J. H. C., Kattge, J., … van Bodegom, P. M. (2013). Impacts of trait variation through observed trait–climate relationships on performance of an Earth system model: a conceptual analysis. *Biogeosciences*, *10*(8), 5497–5515. doi: 10.5194/bg-10-5497-2013

Walker, A. P., Quaife, T., van Bodegom, P. M., De Kauwe, M. G., Keenan, T. F., Joiner, J., … Woodward, F. I. (2017). The impact of alternative trait-scaling hypotheses for the maximum photosynthetic carboxylation rate (Vcmax) on global gross primary production. *New Phytologist*, *215*(4), 1370–1386. doi: https://doi.org/10.1111/nph.14623

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Shaver, G. R., & Johnson, L. C. (1997). Predicting gross primary productivity in terrestrial ecosystems. *Ecological Applications*, *7*(3), 882–894. doi: 10.1890/1051-0761(1997)007[0882:PGPPIT]2.0.CO;2

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. doi: 10.1111/j.1365-2486.2004.00891.x

Woodward, F. I. (1987). *Climate and plant distribution*. Cambridge University Press.

Worden, J. R., Bloom, A. A., Pandey, S., Jiang, Z., Worden, H. M., Walker, T. W., … Röckmann, T. (2017). Reduced biomass burning emissions reconcile conflicting estimates of the post-2006 atmospheric methane budget. *Nature Communications*, *8*(1), 2227. doi: 10.1038/s41467-017-02246-0

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., … Diemer, M.

(2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827.

Wu, M., Scholze, M., Voßbeck, M., Kaminski, T., & Hoffmann, G. (2019). Simultaneous assimilation of remotely sensed soil moisture and FAPAR for improving terrestrial carbon fluxes at multiple sites using CCDAS. *Remote Sensing*, *11*(1), 27.

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

Xu, X., & Trugman, A. T. (2021). Trait-Based Modeling of Terrestrial Ecosystems: Advances and Challenges Under Global Change. *Current Climate Change Reports*, *7*(1), 1–13. doi: 10.1007/s40641-020-00168-6

Yin, Y., Bloom, A. A., Worden, J., Saatchi, S., Yang, Y., Williams, M., … Schimel, D. (2020). Fire decline in dry tropical ecosystems enhances decadal land carbon sink. *Nature Communications*, *11*(1), 1900. doi: 10.1038/s41467-020-15852-2

Table 1: *Observation-based datasets assimilated into CARDAMOM. Adapted from Quetin et al. (in revision).*

Observation	**Source**	**Years**	**Uncertainty**	**Reference**
Net biome exchange (NBE)	CMS-Flux	2010–2015	Optimized (prior range = 0.001-2 gC m^{-2} day^{-1})	*J. Liu et al., 2017, 2021*
Leaf area index (LAI)	MODIS	2010–2015	±log(1.2)	*Myneni et al., 2002*
Solar-induced fluorescence (SIF)	GOSAT	2010–2015	±log(2)	*Frankenberg et al., 2011*
Above- and below-ground biomass (ABGB)	Multiple	2000	≥±log(1.5)	*Saatchi et al., 2011*
Soil organic matter (SOM)	SoilGrids	2000	±log(1.5)	*Poggio et al., 2021*
Fire C emissions	MOPITT	2010–2015	±20%	*Bowman et al., 2017;* *Worden et al., 2017*

Table 2: *Environmental covariates used as features in the predictive EF model framework. For relevant time-varying covariates, the mean and standard deviation are computed over the analysis period (2000–2015).*

	Environmental covariate	**Source**
Climate	Minimum temperature *(mean, std. dev.)*	CRUNCEP
	Maximum temperature *(mean, std. dev.)*	CRUNCEP
	Shortwave radiation *(mean, std. dev.)*	CRUNCEP
	Vapor pressure deficit (VPD) *(mean, std. dev.)*	CRUNCEP
	Precipitation *(mean, std. dev.)*	CRUNCEP
	Burned area *(mean, std. dev.)*	CRUNCEP
	Aridity index	*Trabucco & Zomer, 2019*
Vegetation & soil	LAI *(mean, std. dev.)*	MODIS
	ABGB *(mean)*	*Saatchi et al., 2011*
	SOM *(mean)*	SoilGrids
	Soil water holding capacity	SoilGrids
	Soil pH	SoilGrids
	Soil clay fraction	SoilGrids
	Soil bulk density	SoilGrids
	Depth to bedrock	SoilGrids
	Canopy height	IceSat *(Simard et al., 2011)*

FIGURE CAPTIONS

Figure 1: Schematic diagrams of the (a) EF-based and (b) PFT-based parameterization approaches within CARDAMOM.

Figure 2: Example NBE time series (2000–2015) for one pixel (latitude = 54, longitude = -10), including the optimally parameterized model estimate with 25th-75th percentile range (green), the PFT-based model estimate (orange), and the EF-based model estimate (blue). Here, the mean absolute error (MAE) for the PFT-based model is 0.69 gC m^{-2} day^{-1}, while that for the EF-based model is 0.36 gC m^{-2} day^{-1}.

Figure 3: (a) Errors (calculated as normalized RMSE) and (b) R^2 values for EF parameter predictions relative to optimal parameters. Individual DALEC parameters (gray circles) are organized into broad functional groups (x-axis bins), with each group's mean shown as a black diamond (error bar indicating standard deviation).

Figure 4: Maps comparing NBE performance of the PFT-based and EF-based models. (a) Mean absolute error (MAE) for NBE predictions from the PFT-based model; (b) MAE for NBE predictions from the EF-based model; (c) difference between (a) and (b); (d) best-performing model at each pixel, based on lowest MAE. Dark gray pixels in (d) represent cases in which NBE_{PFT} MAE and NBE_{EF} MAE are within 5% of each other. Light gray pixels are excluded from analysis either due to the ensemble range filter (Sec. 3.2), land cover filter (Sec. 3.6) or unavailability of NBE data.

Figure 5: Results from time series decomposition analysis. (a) Distributions of the location of each pixel's PFT-based and EF-based mean NBE within the corresponding optimal NBE ensemble. A value of 50 indicates that the PFT-based or EF-based mean NBE estimate aligns with the median of the optimal ensemble and is considered the most accurate outcome. (b) Distributions of NBE IAV for the PFT-based, EF-based, and optimally parameterized model. (c) Heatmap comparing Pearson correlations between annually averaged NBE from the optimally parameterized model and annually averaged NBE from the PFT-based model (x-axis) with correlations between annually averaged NBE from the optimally parameterized model and annually averaged NBE from the EF-based model (y-axis). Points lying in the upper right-hand corner (first quadrant) have PFT-based and EF-based NBE annual averages that are both strongly correlated with those from the optimal model. (d) Heatmap comparing correlations between the de-trended NBE seasonal cycle from the optimally parameterized model and that from the PFT-based model (x-axis) with correlations between the de-trended NBE seasonal cycle from the optimally parameterized model and that from the EF-based model (y-axis). For subplots (c) and (d), coloration of grid cells corresponds to relative point density.

Figure 6: Role of parameter precision in controlling MAE. (a) Observed GPP MAE (resulting from EF versus optimal comparison) versus predicted MAE (resulting from multiple linear regression with information on EF parameter precision). (b) Same, but for R_{eco}. (c) Same, but for NBE. R^2 is the coefficient of determination. The thin black line denotes a 1:1 relationship.

Figure 7: Role of CARDAMOM's uncertainty in controlling MAE. (a) CARDAMOM's GPP ensemble interquartile range (IQR) versus predicted GPP MAE for both the EF-based (blue) and PFT-based (orange) models. (b) Same, but for R_{eco}. (c) Same, but for NBE. In each subplot, regression lines are

plotted in blue and orange (m represents the slope of each line; R^2 is the coefficient of determination). The thin black line denotes a 1:1 relationship.

Figure 8: Component flux prediction skill and error compensation. (a) Boxplots comparing the EF-based and PFT-based models' MAE across fluxes. (b) Distributions of the difference in MAE between PFT-based and EF-based predictions for GPP, R_{eco}, and NBE. (c) Bar charts showing the percentage of vegetated pixels for which each model's predictions were more accurate (lower MAE). (d) Bar chart showing the percentage of vegetated pixels for which errors between component fluxes either compound or compensate to yield NBE errors.