An observation-driven approach to improve vegetation phenology in a global land surface model

J. Kolassa,[1,2] R. H. Reichle,[2] R.D. Koster,[2] Q. Liu,[3,2] S. Mahanama,[3,2] F. Zeng,[3,2]

[1]Universities Space Research Association, Columbia, MD

[2]Global Modeling and Assimilation Office, NASA Goddard Spaceflight Center, Greenbelt, MD

[3]Science Systems and Applications Inc., Lanham, MD

Key Points:

- Plant functional subtypes are introduced and separately calibrated against MODIS FPAR observations to improve modeled vegetation estimates
- The calibration globally reduces the model error and bias with respect to observations with a neutral impact on the correlation skill
- Changes to the modeled vegetation activity propagate into the broader model eco-hydrology

Corresponding author: Jana Kolassa, jana.kolassa@nasa.gov

Abstract

An empirical model calibration approach is presented that aims to approximate missing biosphere processes in a global land surface model without the need for substantial model structural changes. The strategy is implemented here using the NASA Catchment-CN land surface model and Moderate Resolution Imaging Spectroradiometer (MODIS) observations of the fraction of absorbed photosynthetically active radiation (FPAR). Existing plant functional types (PFTs) of the Catchment-CN model are divided into 3 subtypes, based on the bias between the model simulated and MODIS observed FPAR. Separate sets of vegetation parameters for each subtype are then calibrated at a small number of grid cells with homogeneous, single-PFT land cover, using MODIS FPAR reference observations from 2003-2009. The effectiveness of the empirical approach at improving the realism of modeled vegetation dynamics is investigated with two global model simulations for the period 2010-2016, one using the newly calibrated parameter values and the other using the original values. Globally, the calibrated parameters reduce the RMSE of the modeled FPAR with respect to MODIS by 0.029 (approx. 10%) on average. In some regions, substantially larger RMSE reductions are achieved. RMSE reductions are primarily driven by model bias reductions, with neutral effects on the temporal correlation skill. While the empirical approach is suitable for achieving consistent model improvements, it is shown to be sensitive to the characteristics of the model error, specifically a dominance of the bias component in the case of Catchment-CN. Ultimately, more fundamental model structural changes may be required to achieve better improvements.

Plain Language Summary

Plants impact the exchange of water, energy, and carbon between the land surface and the atmosphere and are thus one of the key factors controlling land-atmosphere interactions. Because vegetation also evolves more slowly than the atmosphere, being able to correctly model vegetation activity is important to make accurate predictions of atmospheric behavior, for example for weather or seasonal forecasts. This study presents an approach to introduce more vegetation types in a land surface model and to use satellite observations of vegetation activity to calibrate the parameters that describe the behavior of each vegetation type. We show that using this approach results in better model simulations of vegetation activity globally compared to observations. We also show that changing the vegetation has wide-reaching consequences on other model components, including the water and carbon cycle at the land-atmosphere boundary.

1 Introduction

In the context of global Earth system modeling, one of the primary functions for land surface models is to provide accurate estimates of terrestrial water, energy and carbon fluxes to serve as boundary conditions for atmospheric model simulations (e.g. (Holtslag & Steeneveld, 2011)). The land states impact the atmosphere via two primary pathways: (1) through direct fluxes from the soil to the atmosphere (primarily controlled by soil moisture) and (2) via the surface vegetation. The longer memory of land surface states, such as soil moisture and vegetation and their related surface fluxes, can contribute to improved predictions of mean atmospheric behavior weeks or even months in advance.

The link between soil moisture and atmospheric states has been extensively investigated (e.g., (R. D. Koster et al., 2004; Dirmeyer et al., 2018)) and dedicated soil moisture satellite missions such as the Soil Moisture and Ocean Salinity (SMOS, (Kerr et al., 2012)) and Soil Moisture Active Passive (SMAP, (Entekhabi et al., 2010)) missions have contributed to improved understanding and modeling of these processes. By comparison, the role of the vegetation has received little attention, despite being the largest contributor to surface fluxes (Good et al., 2015). Through photosynthesis plants move water from the root zone to the atmosphere (transpiration), while at the same time fixing atmospheric carbon dioxide. The efficiency of carbon uptake and transpiration depends on the ecosystem sensitivity to climatic drivers (e.g., (Poulter et al., 2013)), particularly during extreme events (Tardieu & Simonneau, 1998). Furthermore, vegetation helps determine the surface albedo and surface roughness, intercepts precipitation, and reduces surface runoff (e.g., (Zhang et al., 2015)). In short, the vegetation state exerts a key control on surface water and energy partitioning and the associated response of terrestrial water, energy and carbon fluxes to environmental drivers. Through this control, the influence of vegetation propagates to other parts of the Earth system, such as near-surface atmospheric states (temperature, humidity), the planetary boundary layer (De Kauwe et al., 2017), and precipitation.

The current generation of land surface models typically uses Dynamic Vegetation Models (DVMs) to simulate the dynamic response of vegetation to environmental drivers, with variations in sensitivity to those drivers characterized through so-called Plant Functional Types (PFTs) (e.g., (Cox, 2001; Sitch et al., 2003; Oleson et al., 2010)). An example of this is the NASA Catchment-CN land model, which combines the land surface energy and water budget calculations of the Catchment land model with the carbon and nitrogen modules of the Community Land Model (CLM) (section 2.1). Initial simulations show that Catchment-CN can discern some of the variability in plant responses to environmental drivers, though with some strong biases with respect to observations (Figure 1(a)), and generally only produces a fraction of the observed temporal variability (Figure 1(c); see also R. D. Koster et al. (2014) and Lee et al. (2018)). The model also underestimates the observed variability *within* vegetation types, a known issue associated with the characterization of vegetation responses through PFTs (e.g., (Konings & Gentine, 2017)). Such errors in the modeled vegetation phenology can result in modeled surface water and energy fluxes that are out of phase or biased. This can lead to substantial biases in near surface states, such as air temperature and relative humidity, as well as a misrepresentation of soil moisture dynamics – problems that persist at the monthly and longer timescales often associated with vegetation anomalies.

Generally, there are three strategies for addressing errors in the modeled vegetation phenology. The first involves changes to the model structure, for example by including missing model processes or replacing existing vegetation parameterizations with more sophisticated process representations. Examples are the inclusion of plant hydraulics (Sperry & Love, 2015; Anderegg, 2015), vegetation mortality (McDowell et al., 2018), or legacy effects after extreme events (Anderegg et al., 2015). This approach, while appropriately grounded in the understanding of vegetation physics, tends to be time and labor intensive and poses the risk of increased model complexity, which is not always desirable. The second strategy, employed here, is to optimize the values of the model parameters within the existing modeling framework, the idea being that the existing framework, deficient as it might be, could be made to work better. This approach is simpler than the first, though it is less tailored toward correcting specific model deficiencies. The third strategy to improve model estimates is through the assimilation of observations. This approach, of course, can only impact model skill within a finite spatial and temporal interval defined by the availability of observations.

The objectives of the present study are (i) to optimize the Catchment-CN vegetation parameters to extract as much physical realism as possible from the existing modeling framework and (ii) to quantify the resulting impact on the accuracy of modeled vegetation dynamics across the globe. Specifically, we use a two-pronged approach. First, we use the bias of modeled vegetation with respect to observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) to divide the model's existing PFTs into *subtypes*. Secondly, we use the same MODIS observations as a reference to calibrate the model's vegetation parameters for each new subtype at a small number of model grid cells with a mostly homogeneous, single-PFT land cover. In effect, we increase the number of model PFTs by a factor of three, with the parameters for each PFT independently calibrated in such a way as to allow the model to reproduce, to the fullest extent possible, observed vegetation variability and sensitivity to climatic drivers at the calibration sites. We assess the overall effectiveness of our procedure through the evaluation, against MODIS observed vegetation dynamics, of two 7-year global model simulations – one using the newly expanded catalogue of PFTs and calibrated parameters and the other using the original Catchment-CN model. We use the same global simulations to assess the broader hydrological impacts of our changes.

We should emphasize that while the approach is applied here to the Catchment-CN model, the findings have broader significance. Our overall strategy for optimizing model parameters could, in principle, be applied to any DVM. That is, the success (or lack of success) we find here should have implications for DVM development in general.

The paper is structured as follows. Section 2 describes the Catchment-CN land surface model and MODIS observations used to develop and test the proposed empirical approach. Section 3 explains the methodology used to introduce the PFT subtypes, calibrate the vegetation parameters, and evaluate the modified model. Section 4 presents the results of the model calibration and evaluation, and section 5 discusses the findings of this study and presents conclusions regarding the efficacy of the proposed approach.

2 Model and Data

2.1 Catchment-CN

In this study we use the NASA Catchment-CN land model (R. D. Koster et al., 2014) to investigate how the introduction and calibration of PFT subtypes can lead to more realistic simulations of vegetation behavior. Catchment-CN is a hybrid of the NASA Catchment model (R. D. Koster et al., 2000; Ducharne et al., 2000) and CLM, version

4 (Oleson et al. (2010)). Simply put, the hybrid merges the Catchment model's water and energy budget framework with the CLM carbon and nitrogen dynamics for carbon reservoir and carbon flux calculations and with CLM's photosynthesis-based estimates of canopy conductance. The hybrid nature of Catchment-CN has a distinct impact on all modeled quantities, enough so that Catchment-CN is not simply equivalent to CLM in its simulation of vegetation and carbon behavior. Specifically, Catchment-CN inherits Catchment's advanced treatment of land surface hydrology, which ties the subgrid variability of soil moisture and temperature to a description of the topographic variability (cf. Fig. 2 in R. D. Koster et al. (2014)), allowing Catchment-CN to differentiate between vegetation properties in valley bottoms, hillsides, and hilltops. Carbon prognostic variables in Catchment-CN are followed independently in these three hydrological zones, allowing the vegetation in the different zones to respond to zone-specific hydrological states. Catchment-CN was exercised extensively in a coupled land-atmosphere system by R. Koster and Walker (2015).

Catchment-CN specifies 19 PFTs (Table 1) derived from ESA GLOBCOVER data (Arino et al., 2007; Bontemps et al., 2011), with up to 4 coexisting PFTs in each model grid element. The default number of types we would apply is two per element, but initial model simulations showed that the temperature stress control for the cropland, temperate grassland, and temperate shrubland PFTs led to an unrealistic vegetation shutdown during cold spells. To mitigate this, these three PFTs were split (in the standard Catchment-CN implementation) into two subtypes each, with one subject to a "moisture stress control" only and the other subject to a "moisture stress and deciduous control". The "moisture stress only" PFTs occur exclusively equatorward of 32°, and the "moisture and deciduous stress" PFTs occur exclusively poleward of 42°. A linear blend of both types is applied between 32° and 42°.

Model simulations with Catchment-CN are driven with surface meteorological data at $0.5° \times 0.625°$ resolution from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; (Gelaro et al., 2017)). The precipitation forcing data are corrected using global gauge-based observations from the NOAA Climate Prediction Center Unified (CPCU; (Xie et al., 2007)) product (except over Africa and poleward of 62.5°N/S latitude), after scaling the CPCU product to the climatology of the Global Precipitation Climatology Project (GPCP) v2.2 pentad precipitation product (Adler et al., 2003). Details of the precipitation correction approach are discussed in Reichle et al. (2017).

The MERRA-2 background precipitation is also scaled to the GPCP v2.2 climatology (Reichle et al., 2019).

Here we run model simulations on the 9-km resolution Equal-Area Scalable Earth grid version 2 (EASE v2; (Brodzik et al., 2012)), with up to 4 PFTs in each grid cell, and diagnose daily estimates of model Fraction of absorbed Photosynthetically Active Radiation (FPAR) – estimated from modeled incident and absorbed photosynthetically active radiation.

2.2 MODIS FPAR

As a reference to calibrate the Catchment-CN vegetation parameters, we use observations of FPAR from MODIS flown aboard the Terra (since 1999) and Aqua (since 2002) satellites. The decision to use FPAR observations is driven by the fact FPAR is more directly linked to the raw MODIS reflectances than the leaf area index. This means that FPAR values and - by extension - differences between observed and modeled FPAR are less dependent on specific assumptions in the satellite retrieval algorithm, thus reducing the risk of the algorithm-related differences being aliased into the parameter values during the calibration. Furthermore, the computation of FPAR within the Catchment-CN model is straightforward, making FPAR an attractive choice for the calibration variable.

Here we use the MODIS Collection 6 joint Aqua/Terra FPAR product (MCD15A2H, (Myneni et al., 2015)), provided as 8-day composites with a 500-m resolution on a sinusoidal grid. The FPAR is estimated with a look-up table-based procedure that uses MODIS red (648nm) and near-infrared (858nm) reflectances in addition to vegetation and viewing geometry ancillary information. Details about the processing of the MODIS FPAR observations for the Catchment-CN parameter calibration are provided in section 3.2.

3 Methodology

Central to our modification of the Catchment-CN vegetation module are (1) the introduction of multiple subtypes per PFT, and (2) the calibration of vegetation parameter sets for each subtype using MODIS FPAR observations as a reference. We will introduce 3 subtypes for each PFT, reflective of the model's FPAR bias with respect to MODIS. The idea is simple – we assume that the original PFT designations used in the

Catchment-CN model are too limiting, as they do not account for significant within-PFT variations in behavior. For instance, the PFT designation "needleleaf evergreen temperate trees" does not capture the differing behaviors of pine trees and spruce trees. Within a given PFT, we group together those locations for which the FPAR bias is positive (negative), assuming that the vegetation in these locations indeed behaves similarly and is distinct, in terms of optimal parameter values, from that in locations with a different direction of bias. In other words, we assume that the bias with respect to MODIS reflects intra-PFT differences in vegetation characteristics and can be used to inform the introduction of new PFT sub-types.

Specifically, the first subtype we define uses the default vegetation parameters of the current Catchment-CN; this type is assigned to locations for which the absolute bias between the model and MODIS FPAR falls below a threshold of 0.05 and for which the model performance is thus already considered acceptable. The second/third subtype is assigned to those locations where the model is positively/negatively biased with respect to MODIS. Vegetation parameter sets are separately calibrated for these latter two subtypes. Hereafter, we will refer to the first, second and third subtypes as the 'neutral subtype', 'positive subtype' and 'negative subtype', respectively, and – when discussing the calibration exercise – use 'subtypes' as a summary term for the positive and negative subtypes.

In summary, we first identify suitable Catchment-CN vegetation parameters to be included in the calibration (section 3.1). Next, we prepare the MODIS reference observations (section 3.2). The parameter calibration then consists of 3 steps for each calibrated PFT: (1) select suitable calibration locations (section 3.3.1), (2) calibrate the Catchment-CN parameters separately at all selected locations (section 3.3.2), and (3) from the calibrated locations, select the parameter sets that achieves the largest general error reduction (section 3.3.3). The parameter calibration is conducted using the MODIS data period 2003-2009, a period independent of that used later to evaluate the calibration at the global scale (section 3.4).

3.1 Calibration Parameter Selection

To select which model vegetation parameters to calibrate, we analyzed the sensitivity of the modeled FPAR estimates to perturbations in a number of candidate Catchment-

CN parameters (not shown). Only parameters that led to a significant change in modeled FPAR when varied across their physically reasonable range (as informed by literature) were included in the calibration exercise. These parameters, which are listed in Table 2, can be broadly grouped into three categories: (i) parameters controlling the timing of vegetation activity through leaf-out, senescence, and carbon and nitrogen storage, which impacts the model's correlation skill; (2) parameters controlling photosynthetic efficiency, which impacts the model bias, and (3) optical parameters controlling the vegetation radiative properties.

3.2 MODIS Pre-processing

To prepare the MODIS FPAR observations for their use as a reference in the Catchment-CN parameter calibration we implement 5 pre-processing steps, as listed below. For the MODIS observations used to validate the global model simulation (section 3.4), only steps 1 and 2 are implemented.

1. **Quality Control.** The MODIS FPAR observations are quality controlled using the flags provided with the product. Following the approach of Stöckli et al. (2011), we remove observations impacted by dead detectors, cloud presence or unclear conditions, as well as those marked with flags indicating a failed retrieval, non-land pixels, snow or ice, internal cloud mask, or cloud shadow (Myneni & Park, 2015).
2. **Aggregation to 9-km grid.** The MODIS FPAR observations are aggregated from their original 500-m resolution to the model's 9-km resolution through a simple averaging of all MODIS data points whose center falls within a 9-km model grid cell.
3. **Spectral Noise Correction.** We use a spectral filter to remove residual noise in the MODIS observations still present after the quality control and aggregation steps. The filter employs a Fast Fourier Transform to compute the power spectrum of the MODIS time series separately for each calibration location. Next, all frequencies with a power less than the 99th percentile are removed in order to retain only those MODIS signatures reflective of the seasonal phenology. Finally, the inverse FFT is applied to obtain a noise-corrected MODIS time series.
4. **Computation of Mean Seasonal Cycle.** The objective to improve the Catchment-CN simulated phenological cycle drives our decision to calibrate the model against

the MODIS FPAR mean seasonal cycle rather than against the raw yearly-varying MODIS data. In each calibration location, we estimate a 366-point mean seasonal cycle by averaging all MODIS FPAR observations from the same day-of-year in the calibration period 2003-2009, and assuming that the MODIS 8-day composite FPAR represents the FPAR for each of the 8 days. We subsequently smooth the 366-day mean seasonal cycle using a 5-day moving window centered on the current day.

5. **Scaling of MODIS Mean Seasonal Cycle for deciduous tree PFTs.** Catchment-CN does not simulate vegetation in the understory that is nevertheless observed by MODIS. This may explain large differences between simulated and MODIS-based FPAR during winter for deciduous tree PFTs, a difference that cannot be rectified without a fundamental change in model structure. In fact, forcing the calibrated FPAR to the nonzero value observed by MODIS during winter for a PFT that lacks leaves in the winter months, would result in parameter values that are not optimal for the growing season months. To prevent this, we scale the MODIS mean seasonal cycle for deciduous tree PFTs to match the model simulated FPAR minimum in the winter months. Specifically, the dynamic range of the MODIS mean seasonal cycle $[MOD_{min}, MOD_{max}]$ is scaled to the new range $[CCN_{min}, MOD_{max}]$, where MOD_{max} and MOD_{min} are the maximum and minimum, respectively, of the MODIS FPAR mean seasonal cycle at each location and CCN_{min} is the minimum of the Catchment-CN FPAR mean seasonal cycle. This scaling step is applied only for deciduous trees (PFTs 3, 7, and 8; Table 1), which in the model are photosynthetically inactive during winter.

3.3 Calibration

3.3.1 Selection of Calibration Locations

The objective of our study is to introduce separately calibrated PFT subtypes that are able to improve the global simulation of vegetation behavior. The parameter set for a given positive (or negative) PFT subtype is calibrated at 10 locations that: (i) are found in areas homogeneously covered by the PFT, (ii) provide a reasonable sample of the geographic areas where the PFT occurs, and (iii) are representative of the corresponding median positive (or negative) bias between the original model and the MODIS observations across the globe.

The selection of the calibration locations is schematically illustrated for one positive subtype in Figure 2 and described in more detail as follows. For a given PFT, we first identify all locations with the same dominant PFT in both the model and MODIS-derived PFTs, requiring both datasets to indicate a homogeneous PFT cover, defined here as an area fraction in excess of 95% (Step 1 in Figure 2). The resulting set of candidate locations is then subjected to a 5-element k-means clustering of their geographic location (latitude and longitude) to obtain 5 clusters that represent different geographic regions (Step 2 in Figure 2). From each of the 5 clusters, we then select two locations that are representative of the median positive bias of the original model and two more locations that are representative of the median negative bias (Step 3 in Figure 2). Two locations are chosen for both the positive and negative sub-type to provide a robust number of 10 calibration locations for each sub-type, while also keeping the overall computational cost of the calibration at a feasible level.

To estimate the median positive and negative bias for each PFT, we first compute the global model vs. MODIS bias distribution separately for each PFT. At each model grid cell with multiple PFTs, we use a small number of neighboring grid cells to unmix the bias contribution of each PFT (described in detail in Appendix A). The resulting global, PFT-specific bias distributions (Figure 3) are then used to estimate the median positive and negative bias value for each PFT, after removing (absolute) bias values smaller than 0.05, the threshold for the neutral subtype.

Finally, after determining this typical, PFT-specific bias for each PFT subtype, we now choose the two locations from each of the 5 geographic clusters of candidate locations such that the model bias at these two locations is closest to the typical bias of the PFT-subtype in question (pink diamonds in Step 3 of Figure 2; plotted on map in Step 4 of Figure 2). Owing to the rarity of some of the PFTs, it was not always possible to identify 10 calibration locations that fulfilled all of the above selection requirements, and as a result, fewer than 10 calibration locations were used for some subtypes of PFTs 7, 8, 15 and 16 (Table 1). No suitable calibration locations could be identified for PFT 9, which was therefore not calibrated at all. In addition, crops are strongly influenced by human-driven processes such as irrigation and harvesting, which are observed by MODIS but not represented in Catchment-CN. We do not expect the calibration to perform sensibly over cropland, and thus exclude crop PFTs 18 and 19 from the calibration. Figure 4 shows a map of the final set of calibration locations.

3.3.2 Calibration at Selected Locations

Here we use the Particle Swarm Optimization (PSO) algorithm developed by Kennedy and Eberhart (1995) to calibrate the Catchment-CN vegetation parameters at the selected calibration locations. The PSO algorithm employs an ensemble of trajectories (particles) to explore the solution space. The state of each particle is characterized by its position p, i.e., values of a possible solution (here, possible values of Catchment-CN vegetation parameters), and its velocity v, i.e., a combination of the direction and step size in which that particle explores the solution space. After initialization, the particle states are iteratively updated until an optimal solution (position) is found. Relatively unique to the PSO approach is that the update to a particle's state on each iteration is guided by the particle's own best solution encountered (cognitive aspect) as well as the best solution encountered globally across all particles (social aspect). Specifically, for particle n and iteration i, the new position and velocity are computed using:

$$v_{n,i+1} = w_i v_{n,i} + c_1 r_{n,1}(p_{best,n} - p_{n,i}) + c_2 r_{n,2}(p_{best,global} - p_{n,i}) \tag{1}$$

$$p_{n,i+1} = p_{n,i} + v_{n,i} \tag{2}$$

where $p_{best,n}$ is the best solution encountered by particle n on its trajectory, $p_{best,global}$ is the best global solution encountered across all particle trajectories, and $r_{n,1}$ and $r_{n,2}$ are random numbers. The second and third terms in Eq. 1 are the cognitive and social term, with their respective weights c_1 and c_2. The values chosen for c_1 and c_2 allow a trade-off between a thorough exploration of the solution space (with a higher weight given to the cognitive aspect) and a fast convergence (with a higher weight given to the social aspect). Finally, the term w_i is the inertia weight that constrains the velocity term and provides control over a particle's wide-spread exploration of the solution space (for large w_i) or thorough exploration around possible solutions (for small w_i).

Each iteration of the PSO algorithm consists of 3 steps: (1) evaluating the skill of each particle position in terms of a defined cost function, (2) updating the best individual and global solution (if applicable), (3) updating each particle's state according to Eq. 1 and 2.

In our study, the solution space is spanned by the Catchment-CN vegetation parameters defined in section 3.1, i.e., each element of the particle position vector corresponds to one of the calibration parameters. The particle positions are initialized by sampling from a uniform distribution spanning the range of physically reasonable values for each parameter, defined from literature and expert knowledge. The cost function to be minimized is the root mean-squared error (RMSE) between the model simulated and MODIS observed FPAR mean seasonal cycles. The choice for the PSO algorithm weights are guided by the PSO performance analysis of Trelea (2003), as well as by the similar calibration exercise of De Lannoy et al. (2013); the values for w_i, c_1 and c_2 are set to 0.8, 0.7 and 1.3, respectively. Following the findings of Engelbrecht (2006) that ≈30 particles are sufficient for calibrating hydrological models, we choose 28 particles to allow for optimal parallelization of the calibration algorithm on the computing platform used for this study. We enforce a minimum number of 10 iterations and introduce a stopping criterion that halts the calibration when the global best position $p_{best,global}$ has not been updated for a consecutive 10 iterations.

Parameters are optimized separately for each of the calibration locations introduced above. Seven year runs of the model (2003-2009) produce a model simulated mean seasonal cycle resolved by day that is compared to the MODIS mean seasonal cycle for that location on each iteration of the PSO algorithm.

3.3.3 Selection of Best Parameter Sets

The calibration phase yields up to 10 new parameter sets for each calibrated PFT subtype. The remaining task is to select, separately for each PFT-subtype, the single calibrated parameter set that has the greatest promise of improving the model skill globally, i.e., the optimal parameter set. Figure 5 illustrates our approach. First, we use each parameter set in a 7-year model simulation (2003-2009) at all other calibration locations used for the same subtype. Next, the resulting FPAR simulations are evaluated against MODIS observations at each location using the RMSE, bias and Pearson correlation coefficient. Finally, average skill metrics for each parameter set are computed by averaging the skill at individual locations, using weights that correspond to the size of the geographic cluster from which a given location originates (see section 3.3.1). We confirmed in a separate analysis (not shown here) that this approach resulted in the same relative skill for each parameter set as evaluating the skill of each set in a global model simulation (using all locations rather than just the calibration locations) and thus does successfully select the parameter set that works best.

The selection of the optimal parameter set for each PFT subtype uses a hybrid metric based on both RMSE and correlation. This metric addresses the degradation in the correlation skill that was observed for parameter sets with large RMSE reductions (see section 4). To achieve this we ranked all parameter sets once according to their corresponding RMSE reductions (with rank 1 assigned to the parameter set corresponding to the largest average RMSE reduction) and then again in terms of their correlation increase (with rank 1 being assigned to the parameter set with the largest correlation increase). The optimal parameter set is chosen as the one for which the sum of both ranks is minimized.

3.4 Global Model Validation

We use a global, multi-year simulation of vegetation dynamics to assess the impact of introducing PFT subtypes with separately calibrated vegetation parameters. In contrast to the calibration locations, most model grid cells in this global simulation contain non-negligible contributions from two or more PFTs. The simulation thus also tests the interactions (via shared soil moisture within a given land surface element) of the calibrated parameters for different PFTs, an aspect that was – by construction – not addressed in the calibration. The model skill is quantified as the RMSE, bias and correlation of simulated FPAR with respect to MODIS observed FPAR, and this skill is compared to that of the original (uncalibrated) Catchment-CN. FPAR is simulated for the period 2003-2016 with the last seven years (2010-2016) – i.e., the years not used in the calibration – used to evaluate the model skill. The model simulation is conducted offline at 9-km resolution (see section 2.1) producing daily average estimates of FPAR, which we then aggregate to 8-day composites to match the MODIS temporal resolution.

4 Results

4.1 Performance of the PSO algorithm

As a sanity-check for our PSO algorithm, we first investigate the change in model skill at the individual calibration locations for the calibration period (2003-2009). The calibration algorithm is able to reduce the model RMSE with respect to MODIS at all calibration locations (Figure 6(a)), confirming that the PSO algorithm works as intended. The RMSE reductions are largest for PFTs 10 and 11; these PFTs featured the largest uncalibrated model RMSE and thus the largest potential for error reduction. Generally, the RMSE reductions are proportional to the uncalibrated model RMSE (not shown).

For most PFTs, the RMSE reductions are primarily driven by reductions in the model bias (Figure 6(b)). Changes in the correlation skill (Figure 6(c)) and unbiased RMSE (ubRMSE (Figure 6(d))), i.e., the metrics that are more sensitive to the vegetation dynamics, are generally small and can be positive or negative . In particular, evergreen trees (PFTs 1 and 2), for which the seasonal cycle against which we calibrate is small, show only marginal changes in the correlation skill. The ubRMSE changes are generally smaller in magnitude than the RMSE changes, with notable exceptions for PFTs 7, 10, 13 and 14.

4.2 Parameter Set Generalization Skill

In this subsection, we assess the ability of each of the calibrated parameter sets to reproduce the observed vegetation phenology at all of the other calibration locations used for a given subtype during the calibration period (2003-2009). Recall that all calibration locations have nearly or completely homogeneous land cover (section 3.3.1). Also recall that the parameter set for a given PFT subtype was selected based on the largest rank skill improvement for RMSE and correlation (section 3.3.3).

Generally, very few parameter sets are able to come close to the ideal rank sum of 2, indicating that the parameter sets that minimize the RMSE and maximize the correlation are typically not the same set (Figure 7). This is consistent with findings from section 4.1 that hinted at a trade-off between the RMSE and correlation skill. Nevertheless, for most subtypes it is possible to identify an optimal parameter set with a relatively low value (≤ 10 for subtypes with a full set of 10 calibration locations) for the selection metric, indicating a reasonable compromise between RMSE and correlations skill (red bars in Figure 7).

When applied to all calibration locations of the same subtype, the selected parameter sets (i.e., those producing the red bars in Figure 7) are able to consistently reduce the model's RMSE with respect to MODIS (Figure 8(a)). The largest RMSE reductions, covering a range of PFTs, are observed in South America, Central Africa and Central Eurasia. Very few locations show a degradation of the RMSE skill, suggesting that (1) the calibration does not over-fit the parameters to the specific calibration locations and

(2) the calibration locations are generally representative of each other. That is, parameters calibrated at one location generally yield skill improvements when applied at other calibration locations of the same PFT subtype.

As before, the RMSE reduction is largely driven by a reduction of the model bias, evidenced by the near one-to-one correspondence of the spatial patterns of the RMSE and bias changes (compare Figures 8(a) and (b)). The impact of the new parameter sets on the correlation skill is for the most part neutral (Figure 8(c)). Exceptions include small correlation increases in Central Africa and Central Eurasia at sites belonging mostly to grassland PFTs (cf. Figure 4.) A degradation of the correlation skill is observed for several sites belonging to evergreen PFTs in the Northern US and Canada, Eastern Russia and China, as well as the tropical forests of South America and Africa. As noted before, the small amplitude of the seasonal cycle of the evergreen PFTs impedes the calibration of their timing parameters.

The full set of calibrated parameters for each PFT subtype is provided in the supplementary material.

4.3 Global Model Validation

In this subsection, we evaluate the skill of the calibrated model for the independent period from 2010 to 2016 across the global domain, which consists primarily of grid cells with a mix of PFTs and also samples a much greater variety of climate conditions and vegetation characteristics compared to the handful of calibration locations with homogeneous land cover.

Compared to the original model, the calibrated version introduced here consistently reduces the RMSE with respect to MODIS observations globally (Figure 9(a)), with an average reduction of 0.029, corresponding to $\sim$10% of the uncalibrated model RMSE (Figure 1(b)). Locally, the RMSE reductions can reach significantly higher values of around 0.1, for example in the broadleaf deciduous forests (PFT 6, BDtT) of Central Africa, as well as in the needleleaf evergreen forests (PFT 2, NEbT) and cold C3 grasslands (PFT 14, CC3) of the Northern US and Canada. In particular the BDtT and NEbT types showed considerable bias in the original model (Figure 1(a)). An increase in the model RMSE is observed in a narrow stretch in Northeastern Eurasia that coincides with the imposed land cover boundary between needleleaf evergreen and deciduous forests (PFTs 2 and

3). In the model's GLOBCOVER-based land cover (section 2.1), this boundary is located to the north of the equivalent boundary in the MODIS-based land cover classification (not shown here), thus causing the discrepancy seen here.

As before, the change in RMSE appears to be dominated by the change in the model's bias with respect to MODIS (Figure 9(b)), suggesting that the parameter calibration primarily adjusts the photosynthetic efficiency parameters controlling the mean FPAR magnitude. The effect of the calibrations on the model's correlation skill is largely neutral (Figure 9(c)), with a global average decrease of 0.005. This is likely an effect of the timing parameters taking a secondary role during the calibration process (section 5). Exceptions include small correlation skill improvements in Amazonia (PFT 4) and northwest Australia (PFT 17). The correlation skill with respect to MODIS is slightly degraded in eastern China and parts of the southeastern US and Mexico, where temperate needleleaf evergreen forests (PFT 1, NETT) are dominant. Likely, this degradation is a side-effect of calibrating against the MODIS mean seasonal cycle in regions where the seasonal cycle is generally small.

The parameter calibration reduces the model's ubRMSE with respect to MODIS (Figure 9(b)) by 0.004 in the global average, corresponding to ∼3% of the uncalibrated model ubRMSE (Figure 1(d)). Compared to the RMSE and bias, the ubRMSE reductions are more localized, with the largest changes in the C3 grasslands of the Northern US and Canada. Smaller ubRMSE reductions are evident for the deciduous broadleaf forests of Europe (PFT 7) as well as the deciduous needleleaf forests of Siberia (PFT 3). These regions correspond to vegetation types for which the calibration was found above to be effective at reducing the model ubRMSE (Figure 6(d)). The fact that these localized ubRMSE reductions appear in the absence of corresponding improvements in correlation skill suggests that the ubRMSE improvements mainly derive from an improved model FPAR amplitude compared to MODIS. The large RMSE reductions observed over Central Africa have no correspondence on the ubRMSE map, suggesting that these improvements are dominated by a model bias reduction. Overall, the relatively smaller changes in the model ubRMSE indicate that the calibration is more effective at improving the model's long-term mean vegetation states than at improving the vegetation dynamics.

4.4 Impact on Ecohydrology

Changes to the model's vegetation phenology resulting from the parameter calibration are expected to propagate through the system and impact the model's overall simulation of hydrologic and vegetation states as well as its simulation of energy, water, and carbon fluxes. Here we quantify the changes in these quantities over the validation period (2010-2016).

The new parameter values generally reduce the modeled FPAR (Figure 10(a)), thereby counteracting the overestimation of vegetation activity observed for the uncalibrated model (Figure 1(a)). The most pronounced impacts are observed in large parts of Africa, eastern Amazonia, and the high latitudes of North America, all regions with large initial model errors (Figure 1(b)). In west China and in India, where the uncalibrated model RMSE was also large, the FPAR changes are relatively minor because the crop PFTs that are dominant in both regions were excluded from the calibration (section 3.3.1).

The reduced photosynthetic activity in the calibrated model leads to substantial changes in the terrestrial carbon fluxes. For example, the gross primary productivity (GPP) is reduced overall (Figure 10(b)), indicating that less carbon is fixed by the ecosystem. The spatial pattern of the GPP changes closely mirrors that of the FPAR changes, with the largest average reductions of ~ 5 gC m^{-2} d^{-1} over Africa. The large FPAR changes over Africa also impact the net ecosystem exchange (NEE), increasing it by as much as ~ 2 gC m^{-2} d^{-1} (Figure 10(c)), indicating that the land acts as less of a carbon sink. The correspondence of the spatial patterns of the NEE and FPAR change is smaller than was the case for GPP and FPAR, suggesting that the NEE response may be modulated by additional factors. In areas bordering the Amazon and southeast Asian tropical rainforests, the new model parameters result in slightly reduced NEE. This is a result of slight increases in the model FPAR (and GPP) that are too small to be seen in Figures 10(a) and 10(b).

We can the model GPP changes against global, daily, 0.5 degree resolution GPP estimates from FluxCom, generated through a machine learning based upscaling of eddy covariance tower observations and available for the period 2010-2013 (Jung et al., 2020). The parameter calibration reduces the RMSE of the modeled GPP versus FluxCom data by 0.24 gC m^{-2} day^{-1} in the global average with the largest GPP RMSE reductions of 3 - 4 gC m^{-2} day^{-1} occurring in the tropical forests of Amazonia, Central Africa and

Southeast Asia (not shown). When limiting this evaluation to model grid cells that contain an eddy covariance tower, i.e., locations where the FluxCom data are expected to be more reliable, the average RMSE reduction is 0.17 gC m^{-2} day^{-1}.

The reduced vegetation activity in the calibrated model also propagates into the model's hydrology. This is evident in a reduced transpiration, particularly over regions with large changes in FPAR (Figure 11(a)). For example, in large parts of Africa, the average transpiration rate is reduced by as much as 2 mm/day. Two mechanisms drive the transpiration changes. First, the vegetation that is present is less photosynthetically active and thus a similar leaf area in the calibrated model will transpire less than in the uncalibrated model. Second, over time, the reduced vegetation activity leads to a reduction of the leaf area index (not shown) and thus the plant surface area that *can* transpire.

The above changes to the surface water fluxes are also evident in the simulated soil moisture. The reduced transpiration of a less active vegetation means that less water is removed from the root-zone, resulting in a generally increased root-zone soil moisture (Figure 11(c)). Similarly, the reduced interception (as evidenced by the reduced evaporation of intercepted water), means that more precipitation reaches the soil surface, leading to an increased surface soil moisture (not shown) and – through infiltration – increased root-zone soil moisture. As expected, the spatial patterns of the soil moisture changes mirror those of the transpiration changes, with substantial soil moisture increases (on the order of $\sim$0.1 m^3m^{-3}) over Africa and some smaller increases over North America and southeast Asia. Evaluated against version 6 of the SMAP radiometer Level-2 36-km surface soil moisture product (O'Neill et al., 2019), the surface soil moisture changes in Central Africa correspond to an RMSE reduction of around 0.1 m^3m^{-3} (not shown). However, since the spatial extent of the soil moisture changes is limited, the global average change in model RMSE with respect to SMAP observations is neutral.

The combination of reduced plant transpiration and evaporation from interception leads also to an overall reduced flux of water from the land surface to the atmosphere, as characterized by a reduced total evapotranspiration (ET, Figure 11(d)). It is evident that the ET changes are dominated by changes to the plant transpiration, reflected in the similar magnitude and mostly similar spatial patterns compared to the transpiration changes. Exceptions to this occur over North America and northeast Africa, where the reduced transpiration is balanced by increased soil evaporation (not shown) resulting from the increased soil moisture. The effects of the leaf area reduction are also clearly visible in the reduced evaporation of intercepted precipitation (Figure 11(b)), which generally decreases with decreasing leaf area. When evaluated against evapotranspiration estimates from the Global Land Evaporation Amsterdam Model (GLEAM) version 3.3b (Gonzalez Miralles et al., 2011; Martens et al., 2017) for the period 2010-2017, the changes in model transpiration correspond to a global average RMSE increase of 0.05 mm/day (not shown). The corresponding changes in the interception loss represent a global average RMSE reduction of 0.04 mm/day. The overall impact on the modeled evapotranspiration RMSE with respect to GLEAM is thus neutral.

Overall, the impact of the modified vegetation phenology on the model's ecohydrology is consistent with expectations, in terms of both the sign and the magnitude of the changes. The evaluation against independent observations additionally shows that the ecohydrology changes generally go in the right direction, i.e., the calibration generally reduces the model error with respect to observations. The ecohydrological impact is investigated here using long-term annual means (i.e., including the dormant season). As a result, instantaneous changes in the model's ecohydrology can be substantially larger than those seen in Figures 10 and 11.

5 Discussion and Conclusions

In this study we present a simple approach for introducing observations-based vegetation variability into a traditional PFT-based dynamic vegetation model – an approach that does not require changing the existing model framework. First, PFTs were split into three subtypes based on the local bias between modeled and MODIS-observed FPAR. Next, vegetation parameters relating to photosynthetic efficiency and phenology were calibrated against the observed FPAR for two of the three subtypes (the neutral subtype was left uncalibrated) using a Particle Swarm Optimization (PSO) algorithm. The optimal parameter set for each subtype was selected as the set with the best generalization ability and the best compromise between a reduction of the RMSE and an increase in correlation skill. Finally, skill changes obtained with the newly calibrated Catchment-CN relative to the original model were evaluated in a global 7-year model simulation.

The use of the PSO approach to calibrate Catchment-CN vegetation parameters led to improved vegetation simulations (against observations) at the calibration locations (section 4.1). The effectiveness of the calibration generally scaled with the original model error, leading to larger improvements where the uncalibrated model was less skillful. The calibrated parameters also proved transferable, improving the model's skill against observations outside the particular calibration locations (section 4.2).

Overall, the newly calibrated parameters were able to reduce the global average model RMSE with respect to MODIS FPAR (section 4.3). Locally, the reduction in RMSE was largely driven by a reduction in the model bias, with a mostly neutral impact on the correlation skill. We attribute this behavior to the large bias of the uncalibrated model. A large bias may have emphasized the calibration of the parameters controlling photosynthetic efficiency, such as VC_{max} or $fcur$ (Table 2), over that of parameters controlling the dynamics of the phenology, which were found to be less effective at reducing the bias (and thus the RMSE) during the calibration. A deemphasis on the timing parameters, which do have a large impact on correlation skill, may have resulted in their non-optimal calibration. From this we infer that the parameter calibration might be more effective if it were performed in two phases: an initial calibration of the photosynthetic efficiency parameters to reduce the large model bias, followed by a calibration of the timing parameters to adjust the modeled phenology. A more focused calibration effort would in fact be in line with previous studies that successfully calibrated model parameters targeted at specific aspects of vegetation dynamics (e.g., (Knorr et al., 2010; Forkel et al., 2014; MacBean et al., 2015)).

A caveat worth mentioning involves the multi-century spin-up time generally required to bring carbon reservoirs to full equilibrium in dynamic vegetation models. Obviously, computational constraints make such equilibration impossible during the PSO-based calibration procedure. We are thus assuming here that our calibrated parameters and their impacts on the model's ecohydrology would basically be the same if full equilibration underlay each step of the PSO algorithm. Naturally, once a set of optimal parameter values has been identified, a proper multi-century spin-up using these parameter values would be appropriate.

It is important to keep in mind that differences in modeled and observed FPAR are driven by a multitude of factors beyond error in parameter values, including biases in the meteorological forcing data, uncertainties in the vegetation canopy radiative transfer scheme or inaccuracies in the modeled vegetation processes. During any model parameter calibration, the differences caused by some of these factors will be aliased into the parameter values. That is, the model parameters might be adjusted to compensate for structural deficiencies, which can cause problems in other parts of the model. A similar consideration holds for the MERRA-2 meteorological forcing data, which – although constrained by observations – have residual uncertainties that may lead to compensating errors in the parameter values. The ideal approach would thus be to simultaneously address these model structural deficiencies and forcing data uncertainties in tandem with adjusting the parameter values. However, while it is important to be cognizant of the interplay between parameters, model structure and forcing data, a more comprehensive approach is beyond the scope of this study.

Again, the overall goal of this study was to improve the realism of the Catchment-CN's simulated phenology by calibrating existing model parameters only, that is, without making fundamental, structural changes to the model's dynamic vegetation module. Our approach, we must emphasize, is not specific to the Catchment-CN model; it could in principle be applied to any DVM, and thus our results here have broad significance. The approach does result in simulated phenology that is in better agreement with observations globally. The average skill changes, however, are small. This suggests that the relatively simple sub-typing introduced here is insufficient to fully capture the complex processes that lead to observed intra-PFT vegetation variability. Apparently, more comprehensive model changes are required to realistically simulate vegetation dynamics and observed intra-PFT variability. Useful model changes might include, for example, the explicit treatment of plant hydraulics (Konings & Gentine, 2017) and enhanced representations of anthropogenic processes, such as burning, deforestation, harvesting and irrigation. Secondary processes that could further improve the realism of simulated vegetation include tree mortality (McDowell et al., 2018) or legacy effects from droughts (Anderegg et al., 2015).

Another avenue for improving the simulation of vegetation processes is through the merging of model and observation information in a data assimilation framework (e.g., (Fairbairn et al., 2017; Albergel et al., 2017; Fox et al., 2018; Kumar et al., 2019)). Traditional data assimilation (DA) typically requires the removal of biases between the observations and the model; however, several studies argue that in the case of phenology observations, the bias is informative of model deficiencies and should be retained (e.g. Fairbairn et al. (2017)). Nevertheless, DA is only able to correct the model when and where observations are available is thus not well-suited to correct the more systematic model errors found for the uncalibrated Catchment-CN. However, DA does represent an attractive tool to inform the day-to-day vegetation dynamics during periods when observations are available, perhaps alleviating the often poor correlation skill found with respect to MODIS observations.

Acknowledgments

This work has been funded under the NASA Science Utilization of the SMAP Mission (NNH15ZDA001N-SUSMAP) program. Computational resources for this study were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at the Goddard Space Flight Center.

The MODIS data for this research are available through Myneni et al. (2015) and can be downloaded from https://lpdaac.usgs.gov/products/mcd15a2hv006/.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., ... others (2003). The version-2 global precipitation climatology project (gpcp) monthly precipitation analysis (1979–present). *Journal of hydrometeorology*, *4*(6), 1147–1167.

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., ... others (2017). Sequential assimilation of satellite-derived vegetation and soil moisture products using surfex_v8. 0: Ldas-monde assessment over the euro-mediterranean area. *Geoscientific Model Development*, *10*(10), 3889–3912.

Anderegg, W. R. (2015). Spatial and temporal variation in plant hydraulic traits and their relevance for climate change impacts on vegetation. *New Phytologist*, *205*(3), 1008–1014.

Anderegg, W. R., Schwalm, C., Biondi, F., Camarero, J. J., Koch, G., Litvak, M., ... others (2015). Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. *Science*, *349*(6247), 528–532.

Arino, O., Gross, D., Ranera, F., Leroy, M., Bicheron, P., Brockman, C., ... others (2007). Globcover: Esa service for global land cover from meris. In *2007 ieee international geoscience and remote sensing symposium* (pp. 2412–2415).

Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., Kalogirou, V., & Perez, J. R. (2011). Globcover 2009-products description and validation report. *URL: http://ionia1. esrin. esa. int/docs/GLOBCOVER2009_Validation_Report_2*, *2*.

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H. (2012). Ease-grid 2.0: Incremental but significant improvements for earth-gridded data sets. *ISPRS International Journal of Geo-Information*, *1*(1), 32–45.

Cox, P. M. (2001). Description of the" triffid" dynamic global vegetation model. *Hadley Centre technical note*, *24*.

De Kauwe, M. G., Medlyn, B. E., Knauer, J., & Williams, C. A. (2017). Ideas and perspectives: how coupled is the vegetation to the boundary layer? *Biogeosciences (Online)*, *14*(19).

De Lannoy, G. J., Reichle, R. H., & Pauwels, V. R. (2013). Global calibration of the geos-5 l-band microwave radiative transfer model over nonfrozen land using smos observations. *Journal of Hydrometeorology*, *14*(3), 765–785.

Dirmeyer, P. A., Chen, L., Wu, J., Shin, C.-S., Huang, B., Cash, B. A., ... others (2018). Verification of land–atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations. *Journal of hydrometeorology*, *19*(2), 375–392.

Ducharne, A., Koster, R. D., Suarez, M. J., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 2. parameter estimation and model demonstration. *Journal of Geophysical Research: Atmospheres*, *105*(D20), 24823–24838.

Engelbrecht, A. P. (2006). Fundamentals of computational swarm intelligence. 2005. *Hoboken: John Wiley & Sons, Ltd*.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ... others (2010). The soil moisture active passive (smap) mission. *Proceedings of the IEEE*, *98*(5), 704–716.

Fairbairn, D., Barbu, A., Napoly, A., Albergel, C., Mahfouf, J.-F., & Calvet, J.-C. (2017). The effect of satellite-derived surface soil moisture and leaf area index land data assimilation on streamflow simulations over france. *Hydrology and Earth System Sciences*, *21*(4), 2015–2033.

Forkel, M., Carvalhais, N., Schaphoff, S., Bloh, W. v., Migliavacca, M., Thurner, M., & Thonicke, K. (2014). Identifying environmental controls on vegetation greenness phenology through model-data integration. *Biogeosciences*, *11*(23), 7025–7050.

Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., ... Moore, D. J. (2018). Evaluation of a data assimilation system for land surface models using clm4. 5. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2471–2494.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... others (2017). The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, *30*(14), 5419–5454.

Gonzalez Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., & Dolman, A. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, 453–469.

Good, S. P., Noone, D., & Bowen, G. (2015). Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. *Science*, *349*(6244), 175–177.

Holtslag, A. A. M., & Steeneveld, G.-J. (2011). Single column modeling of atmospheric boundary layers and the complex interactions with the land surface. In R. A. Meyers (Ed.), *Extreme environmental events: Complexity in forecasting and early warning* (pp. 844–857). New York, NY: Springer New York. Retrieved from `https://doi.org/10.1007/978-1-4419-7695-6_45` doi: 10.1007/978-1-4419-7695-6_45

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

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of icnn'95-international conference on neural networks* (Vol. 4, pp. 1942–1948).

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., … others (2012). The smos soil moisture retrieval algorithm. *IEEE transactions on geoscience and remote sensing*, *50*(5), 1384–1403.

Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R., & Mathieu, P.-P. (2010). Carbon cycle data assimilation with a generic phenology model. *Journal of Geophysical Research: Biogeosciences*, *115*(G4).

Konings, A. G., & Gentine, P. (2017). Global variations in ecosystem-scale isohydricity. *Global change biology*, *23*(2), 891–905.

Koster, R., & Walker, G. (2015). Interactive vegetation phenology, soil moisture, and monthly temperature forecasts. *Journal of Hydrometeorology*, *16*(4), 1456–1465.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., … others (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 1. model structure. *Journal of Geophysical Research: Atmospheres*, *105*(D20), 24809–24822.

Koster, R. D., Walker, G. K., Collatz, G. J., & Thornton, P. E. (2014). Hydroclimatic controls on the means and variability of vegetation phenology and carbon uptake. *Journal of climate*, *27*(14), 5632–5652.

Kumar, S. V., M. Mocko, D., Wang, S., Peters-Lidard, C. D., & Borak, J. (2019).

Assimilation of remotely sensed leaf area index into the noah-mp land surface model: Impacts on water and carbon fluxes and states over the continental united states. *Journal of Hydrometeorology*, *20*(7), 1359–1377.

Lee, E., Zeng, F.-W., Koster, R. D., Weir, B., Ott, L. E., & Poulter, B. (2018). The impact of spatiotemporal variability in atmospheric co2 concentration on global terrestrial carbon fluxes. *Biogeosciences*, *15*(18), 5635–5652.

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

Martens, B., Gonzalez Miralles, D., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., ... Verhoest, N. (2017). Gleam v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925.

McDowell, N., Allen, C. D., Anderson-Teixeira, K., Brando, P., Brienen, R., Chambers, J., ... others (2018). Drivers and mechanisms of tree mortality in moist tropical forests. *New Phytologist*, *219*(3), 851–869.

Myneni, R., Knyazikhin, Y., & Park, T. (2015). *Mcd15a2h modis/terra+ aqua leaf area index/fpar 8-day l4 global 500m sin grid v006. nasa eosdis land processes daac.*

Myneni, R., & Park, Y. (2015). *Modis collection 6 (c6) lai/fpar product user's guide.* Feb.

Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., ... others (2010). Technical description of version 4.0 of the community land model (clm). *Note NCAR/TN-4781STR.*

O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T., Bindlish, R., & Chaubell, J. (2019). *Smap l2 radiometer half-orbit 36 km ease-grid soil moisture, version 6.* (Accessed: 2020-01-15) doi: 10.5067/R50VUC07OM4W

Poulter, B., Pederson, N., Liu, H., Zhu, Z., D'Arrigo, R., Ciais, P., ... others (2013). Recent trends in inner asian forest dynamics to temperature and precipitation indicate high sensitivity to climate change. *Agricultural and Forest Meteorology*, *178*, 31–45.

Reichle, R. H., Liu, Q., Koster, R. D., Crow, W. T., De Lannoy, G. J., Kimball, J. S., ... others (2019). Version 4 of the smap level-4 soil moisture algorithm and data product. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3106–3130.

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P., & Partyka, G. S. (2017). Land surface precipitation in merra-2. *Journal of Climate*, *30*(5), 1643–1664.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., … others (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the lpj dynamic global vegetation model. *Global change biology*, *9*(2), 161–185.

Sperry, J. S., & Love, D. M. (2015). What plant hydraulics can tell us about responses to climate-change droughts. *New Phytologist*, *207*(1), 14–27.

Stöckli, R., Rutishauser, T., Baker, I., Liniger, M., & Denning, A. (2011). A global reanalysis of vegetation phenology. *Journal of Geophysical Research: Biogeosciences*, *116*(G3).

Tardieu, F., & Simonneau, T. (1998). Variability among species of stomatal control under fluctuating soil water status and evaporative demand: modelling isohydric and anisohydric behaviours. *Journal of experimental botany*, 419–432.

Trelea, I. C. (2003). The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information processing letters*, *85*(6), 317–325.

Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., & Yang, S. (2007). A gauge-based analysis of daily precipitation over east asia. *Journal of Hydrometeorology*, *8*(3), 607–626.

Zhang, L., Wang, J., Bai, Z., & Lv, C. (2015). Effects of vegetation on runoff and soil erosion on reclaimed land in an opencast coal-mine dump in a loess area. *Catena*, *128*, 44–53.

Appendix A Estimation of PFT-specific global bias distribution

As part of the calibration location selection process, we compute the median bias with respect to MODIS for each PFT subtype. To this end, we first estimate the global model vs. MODIS bias distribution separately for each PFT (Figure 2). For grid cells containing a single PFT, this is simply the long-term bias between modeled and MODIS-observed FPAR computed over the period 2003-2010. In mixed grid cells, we disaggregate the bias contribution from each PFT through a multi-linear regression applied over a center grid cell and its 24 neighbors (Step 1 in Figure A1). We make the assumption that similar environmental conditions lead to similar model biases for a given PFT, with variations in the total model bias across neighboring grid cells driven by the varying area fraction of the contributing PFTs (Step 2 in Figure A1). The (additional) implicit assumption here is that within the area defined by the neighbor grid cells no mixing of PFT subtypes occurs. Based on these assumptions, we can construct the following system of equations:

$$\begin{aligned} b_{t,1} &= a_{1,1}b_1 + a_{2,1}b_2 + a_{3,1}b_3 + a_{4,1}b_4 \\ &\vdots \\ b_{t,25} &= a_{1,25}b_1 + a_{2,25}b_2 + a_{3,25}b_3 + a_{4,25}b_4 \end{aligned} \tag{A1}$$

where $b_{t,n}$ is the total bias in grid cell n, b_p is the bias of PFT p, assumed constant over the 25 grid cells, and $a_{p,n}$ is the known area fraction of PFT p in grid cell n. We require at least 5 data points $b_{t,n}$ to attempt a solution. The above system can then be solved for the PFT-dependent biases b_p, resulting in a global bias map for each PFT (Figure A1 Step 3). As a last step, the bias distribution for each PFT can be used to identify the median positive and negative bias after removing bias values that fall into the 'neutral' zone [-0.05 0.05].

Table 1. Catchment-CN plant functional types

Nr.	Abbreviation	Name	# calibration locations
1	NETT	needleleaf evergreen temperate trees	20
2	NEbT	needleleaf evergreen boreal trees	20
3	NDbT	needleleaf deciduous boreal trees	20
4	BEtT	broadleaf evergreen tropical trees	20
5	BETT	broadleaf evergreen temperate trees	20
6	BDtT	broadleaf deciduous tropical trees	20
7	BDTT	broadleaf deciduous temperate trees	12
8	BDbT	broadleaf deciduous boreal trees	3
9	BETS	broadleaf evergreen temperate shrubs	n/a
10	BDTS	broadleaf deciduous temperate shrubs	20
11	BDTSm	broadleaf deciduous temperate shrubs (moisture stress only)	20
12	BDbS	broadleaf deciduous boreal shrubs	20
13	AC3	arctic C3 grasses	20
14	CC3	cold C3 grasses	19
15	CC3m	cold C3 grasses (moisture stress only)	12
16	WC4	warm C4 grasses	3
17	WC4m	warm C4 grasses (moisture stress only)	20
18	CROP	crops	n/a
19	CROPm	crops (moisture stress only)	n/a

Table 2. Catchment-CN vegetation parameters included in the calibration. Shown are the parameter name, description, and units. Also shown is the parameter characterization as timing parameters (t), photosynthetic efficiency parameters (e), and optical parameters influencing the vegetation radiative properties (o).

Name	Description	Units	parameter type
ndays_{on}	number of days to complete leaf-onset	days	t
ndays_{off}	number of days to complete leaf offset	days	t
crit. dayl.	critical day length for offset	seconds	t
gdd_{c1}	constant in calculation of critical onset growing degree-day sum	-	t
gdd_{c2}	constant in calculation of critical onset growing degree-day sum	-	t
swi_{on}	onset soil water index	-	t
swi_{off}	offset soil water index	-	t
ρ_l	leaf reflectance (visible)	-	o
ρ_s	stem reflectance (visible)	-	o
fcur	fraction of allocation that goes to currently displayed growth, remainder goes to storage	-	e
x_l	leaf/stem orientation index	-	o
τ_l	leaf transmittance (visible)	-	o
τ_s	stem transmittance (visible)	-	o
$\mathrm{vc}_{max,25}$	maximum rate of carboxylation at 25°C	μmol CO_2 m^{-2} s^{-1}	e
qe_{25}	quantum efficiency at 25°C	μmol CO_2 / μmol photon	e

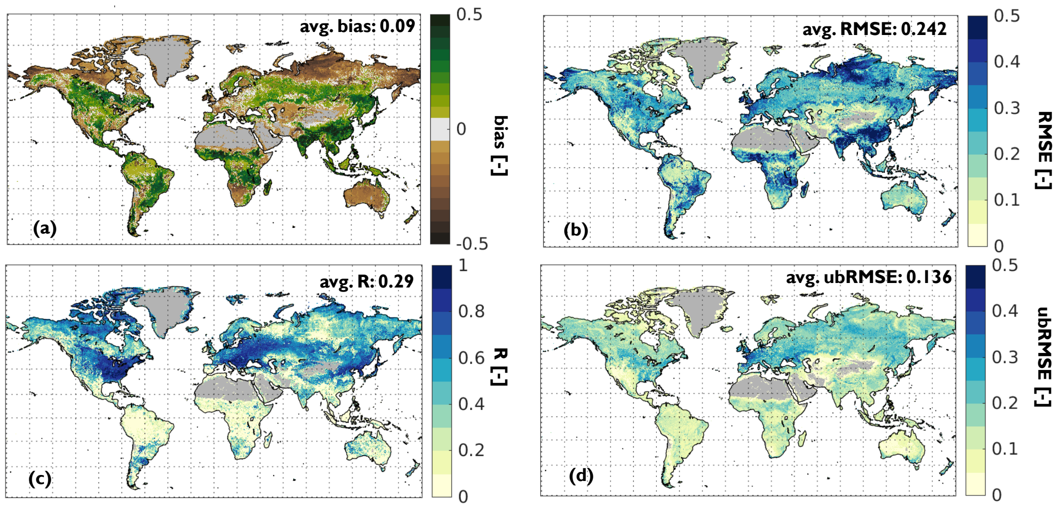


Figure 1. Skill of the uncalibrated Catchment-CN FPAR simulations evaluated against MODIS observed FPAR for 2010-2016 in terms of the (a) bias (model - MODIS), (b) RMSE bias (model - MODIS), (c) Pearson correlation coefficient and (d) the unbiased RMSE (ubRMSE).

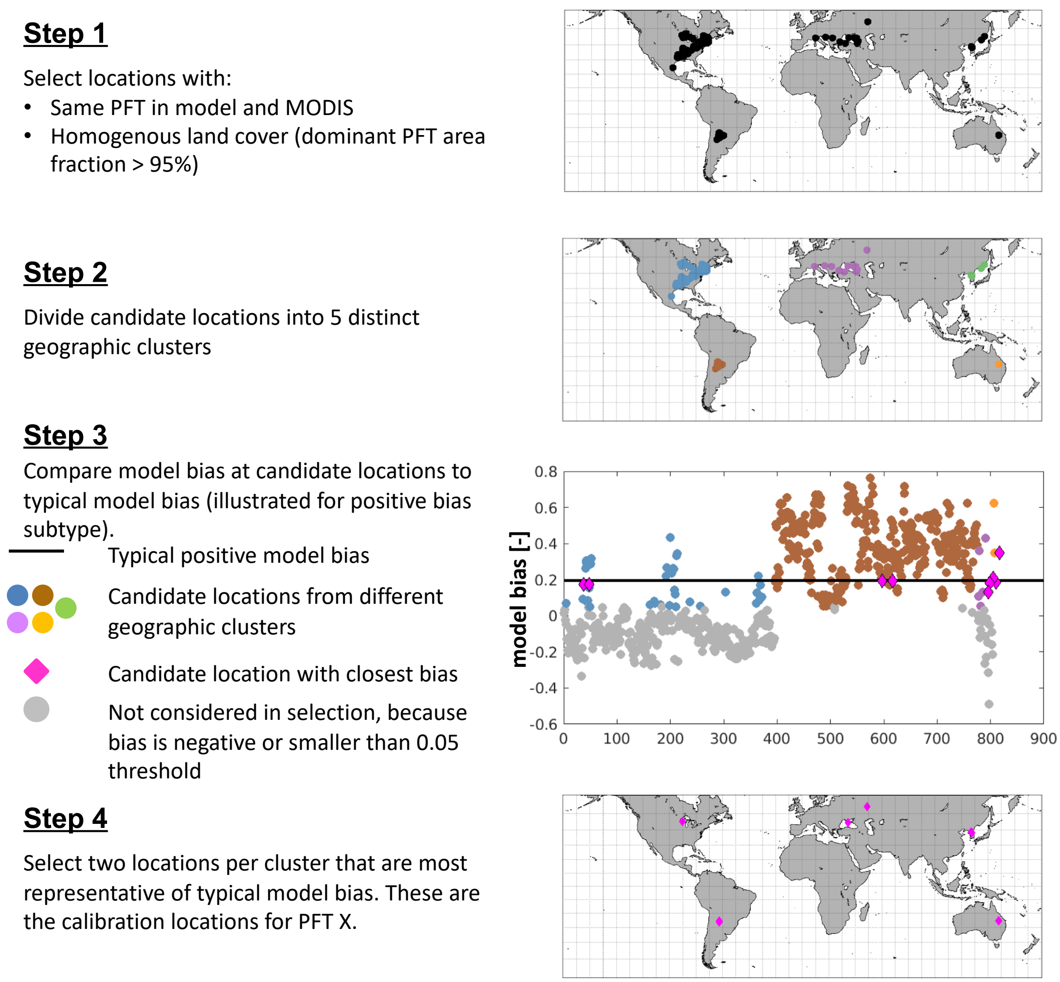
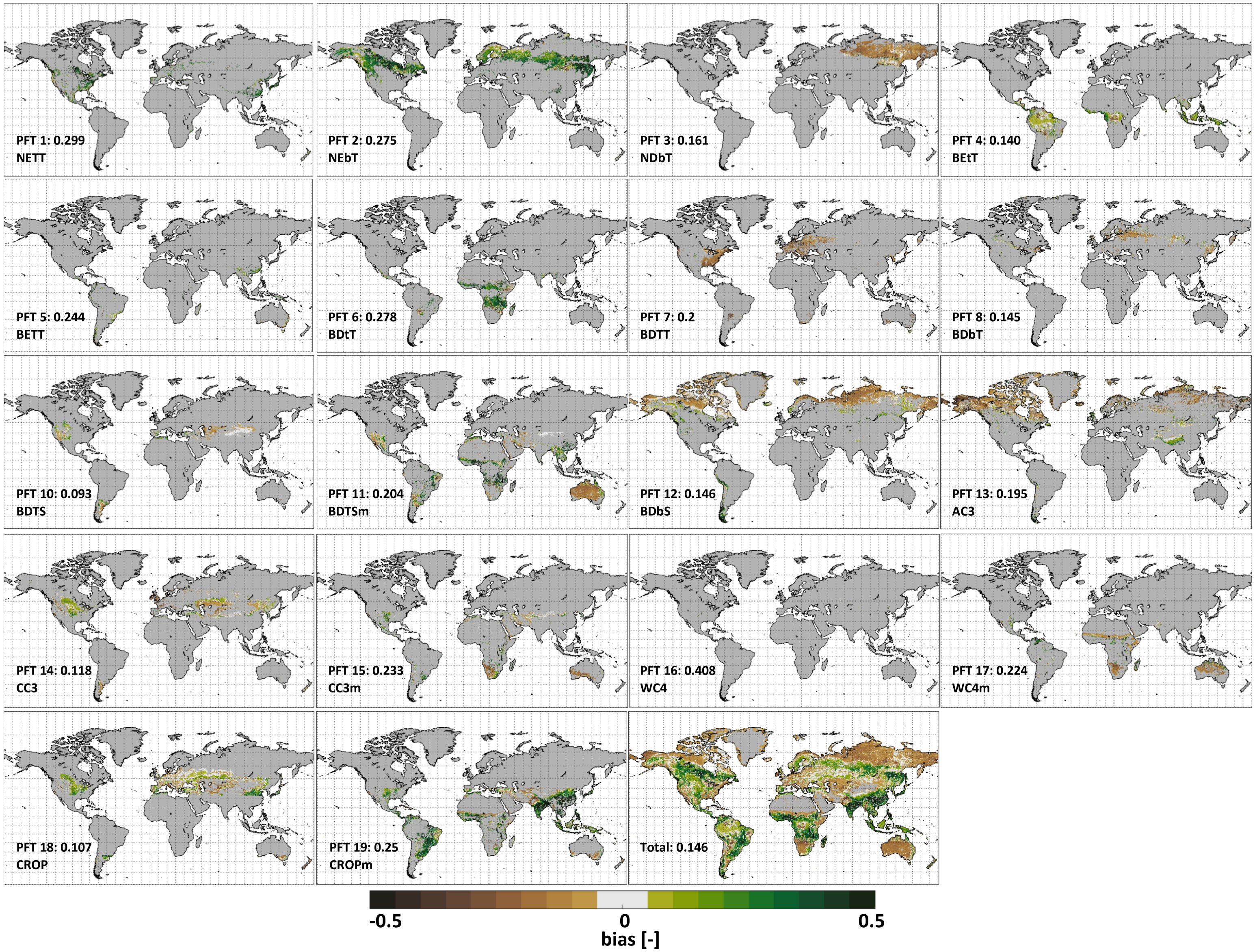


Figure 2. Schematic illustration of the calibration location selection for one (positive) PFT subtype.



[H]

Figure 3. PFT-specific distribution of bias between modeled and MODIS observed FPAR computed for 2003-2009, as well as global bias distribution. Indicated in each plot is the mean absolute bias for each PFT.

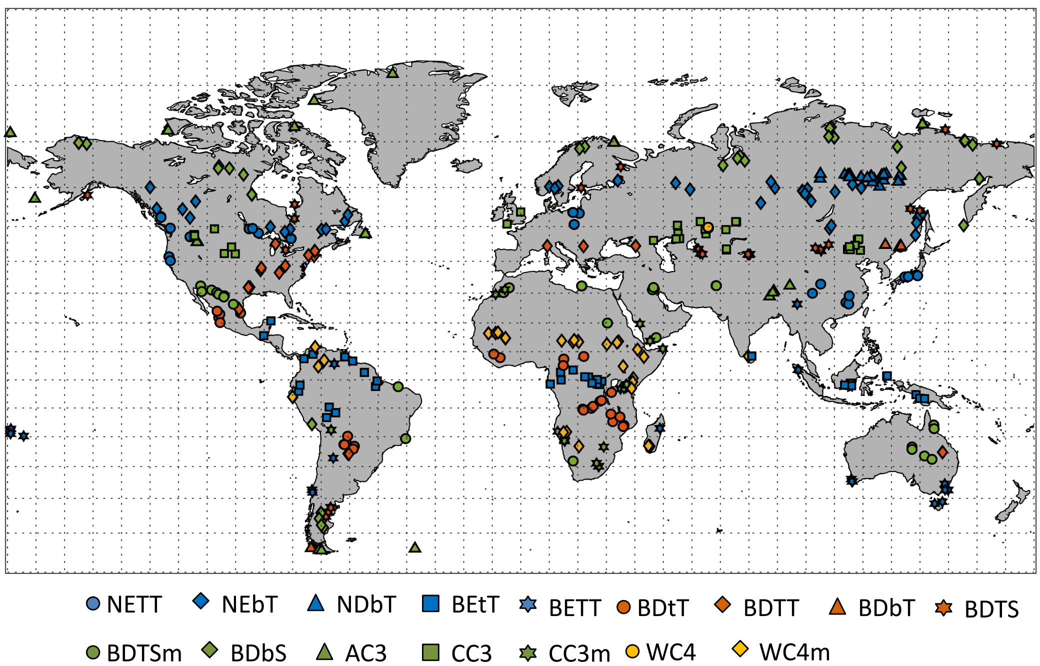


Figure 4. Calibration locations for each PFT. Abbreviations correspond to those introduced in Table 1. PFTs shown in gray are not included in the calibration.

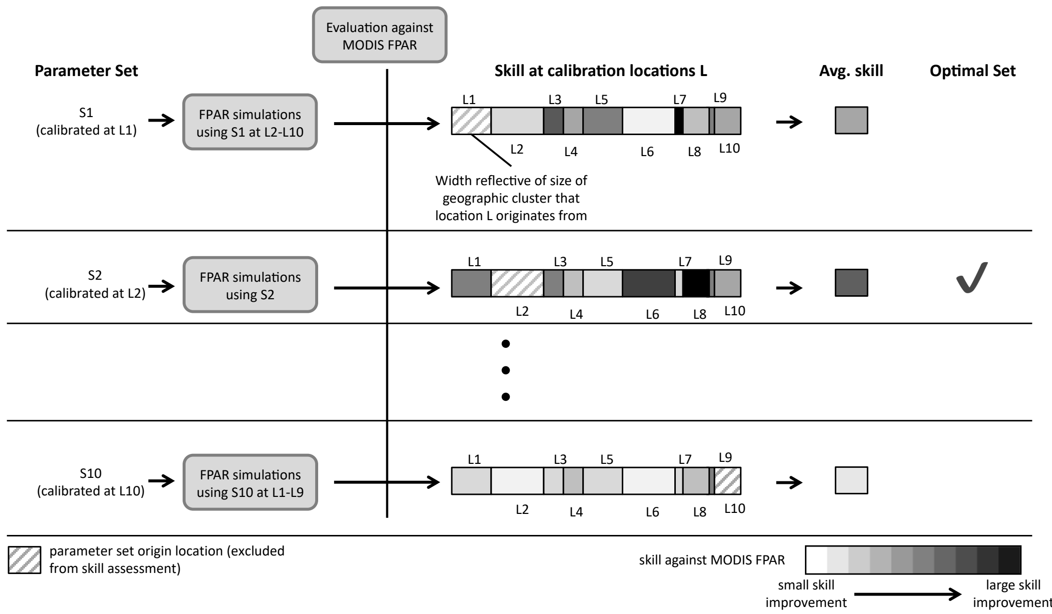


Figure 5. Schematic illustration of the optimal parameter set selection approach. Each parameter set S calibrated at a location L is applied at all other locations of the same subtype in a 7-year (2003-2009) model simulation. The model skill at each location is evaluated against MODIS FPAR observations and combined in a weighted average of the skill at individual locations, with weights corresponding to the size of the geographic origin cluster of each location (section 3.3.1). The parameter set resulting in the largest average skill is selected as the optimal parameter set.

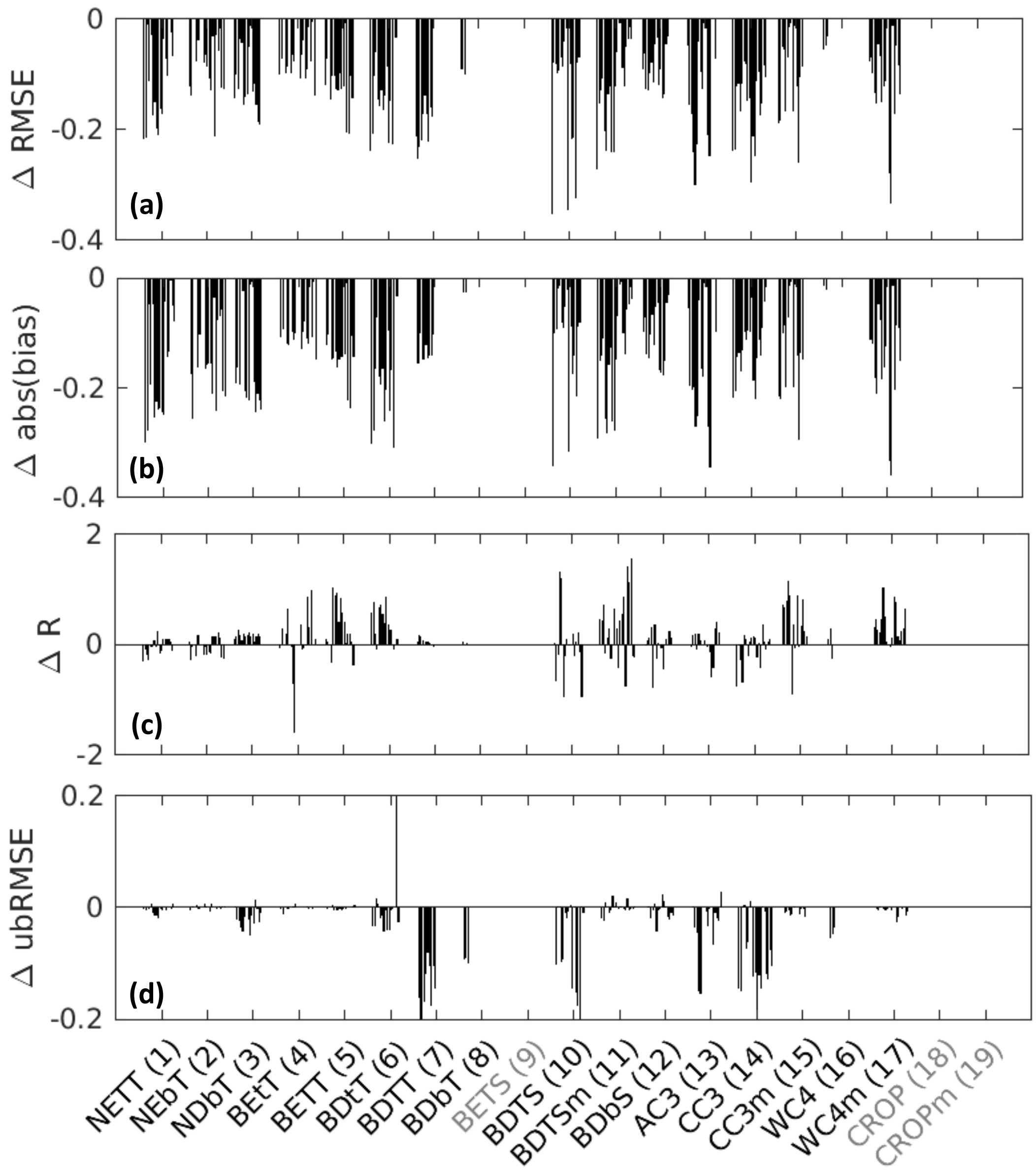


Figure 6. Performance of PSO vegetation parameter calibration at each calibration location showing the difference in (a) RMSE, (b) bias, (c) correlation R, and (d) ubRMSE for 2003-2009 relative to MODIS FPAR between the calibrated and uncalibrated model (calibrated – uncalibrated). PFTs shown in gray are not included in the calibration

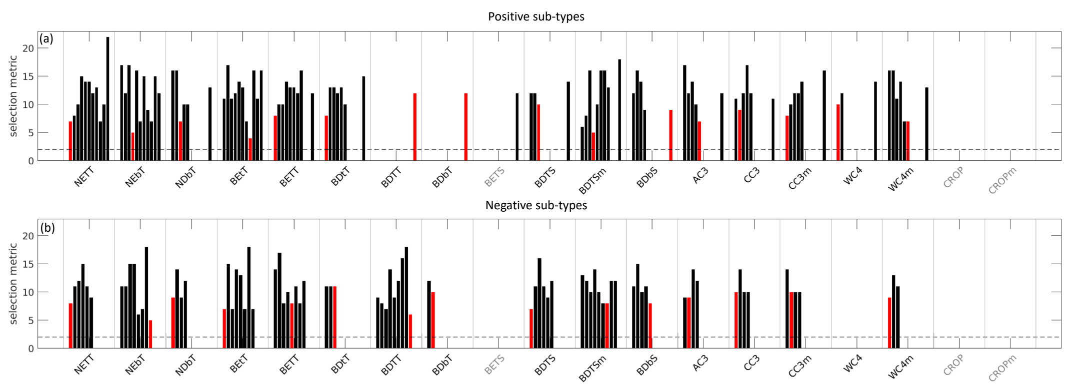


Figure 7. Performance of the parameter sets for the (a) positive and (b) negative subtypes in terms of the selection metric for 2003-2009. Each histogram bar in a given PFT's grouping shows the ability (measured as the sum of two rank skill values) of the parameter set derived at a particular calibration location to capture the observed FPAR behavior at the subtype's other calibration locations (section 3.3.3). Not plotted are the bars for parameter sets that degraded the model skill. The optimal (selected) parameter set for each PFT subtype is highlighted in red and the best possible value of the selection metric of 2 is indicated by the dashed lines.

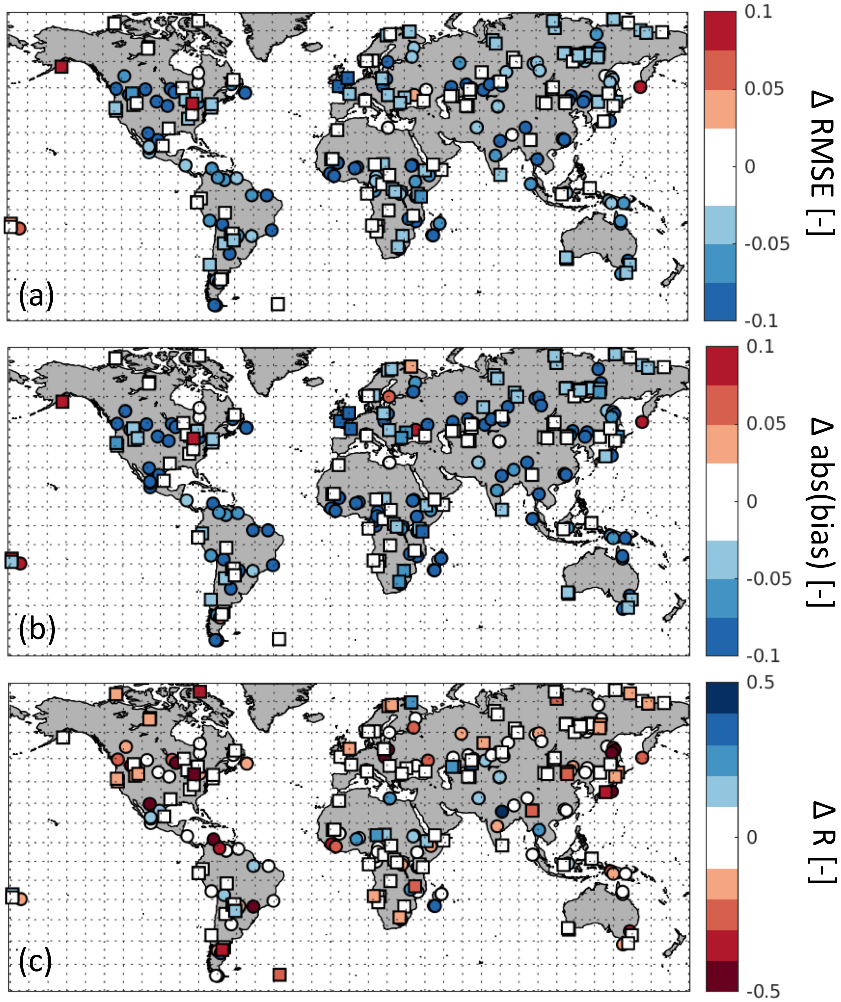


Figure 8. Performance of the selected parameters sets at calibration locations. Shown is the difference (calibrated - uncalibrated) in (a) RMSE, (b) bias, and (c) correlation with respect to MODIS FPAR for 2003-2009 . Blue colors indicate a skill improvement from the calibration, red colors indicate a degradation. The circles correspond to positive subtype calibration locations and squares represent negative subtype calibration locations.

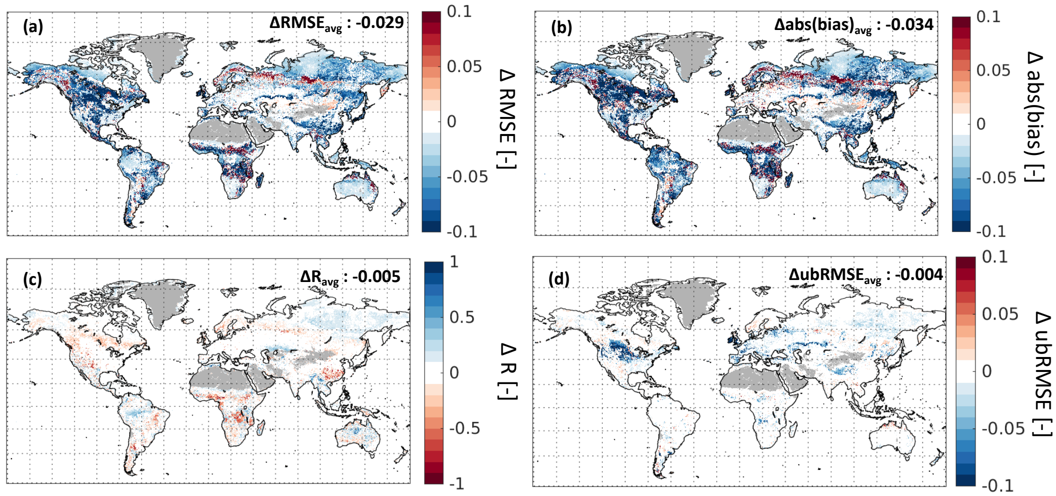


Figure 9. Change in (calibrated - uncalibrated) model (a) RMSE, (b) absolute bias, (c) correlation skill , and (d) ubRMSE versus MODIS FPAR observations for 2010-2016. Also shown is the global average skill change across all locations.

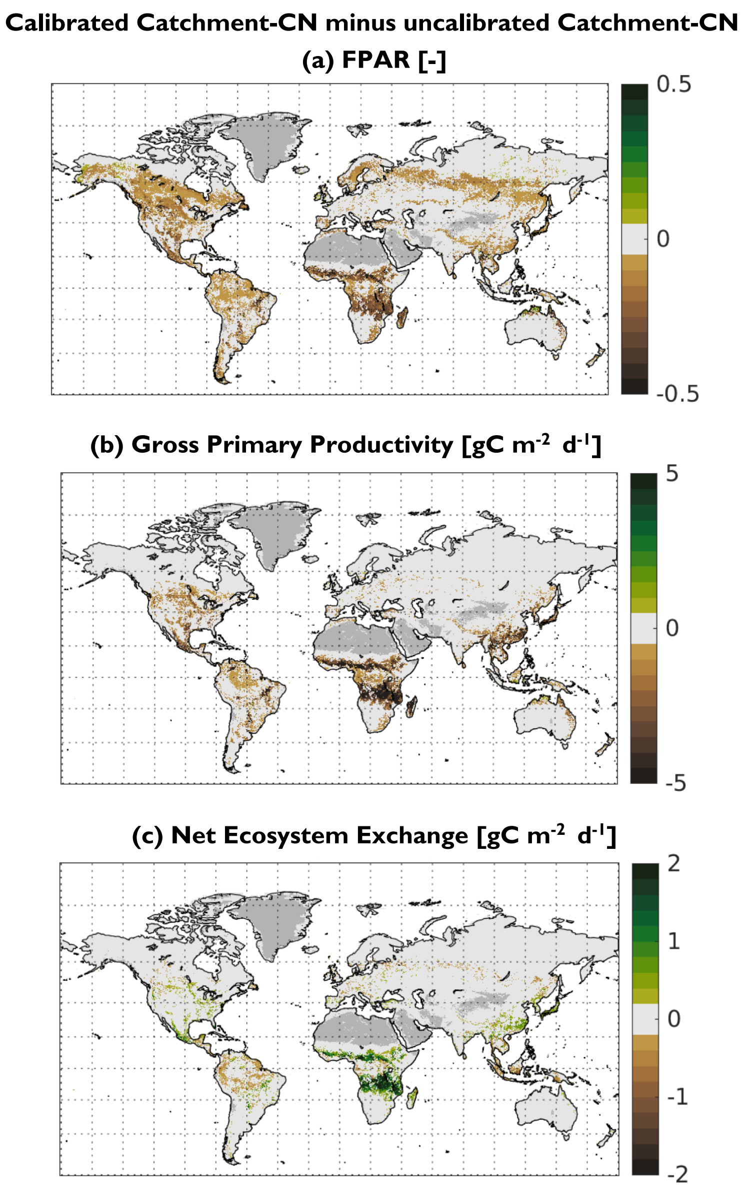


Figure 10. 2010-2016 mean difference in (calibrated minus uncalibrated) Catchment-CN (a) FPAR, (b) gross primary productivity (GPP) , and (c) net ecosystem exchange (NEE).

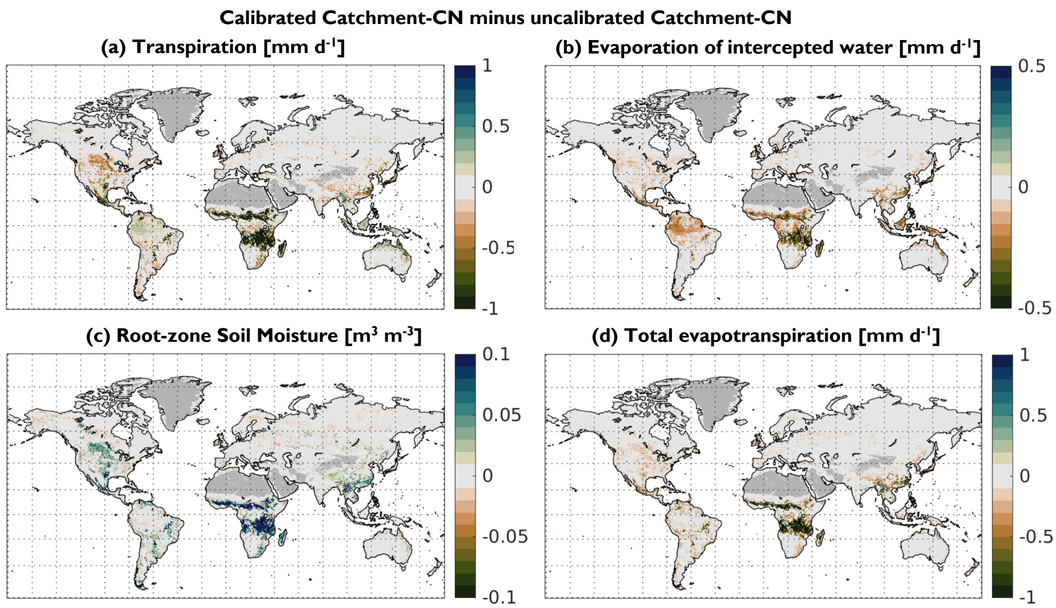


Figure 11. 2010-2016 mean difference in (calibrated minus uncalibrated) Catchment-CN (a) transpiration, (b) interception loss, (c) root-zone soil moisture, and (d) total evapotranspiration. Note that the value range plotted for the interception loss differs from that plotted for transpiration and total evapotranspiration.

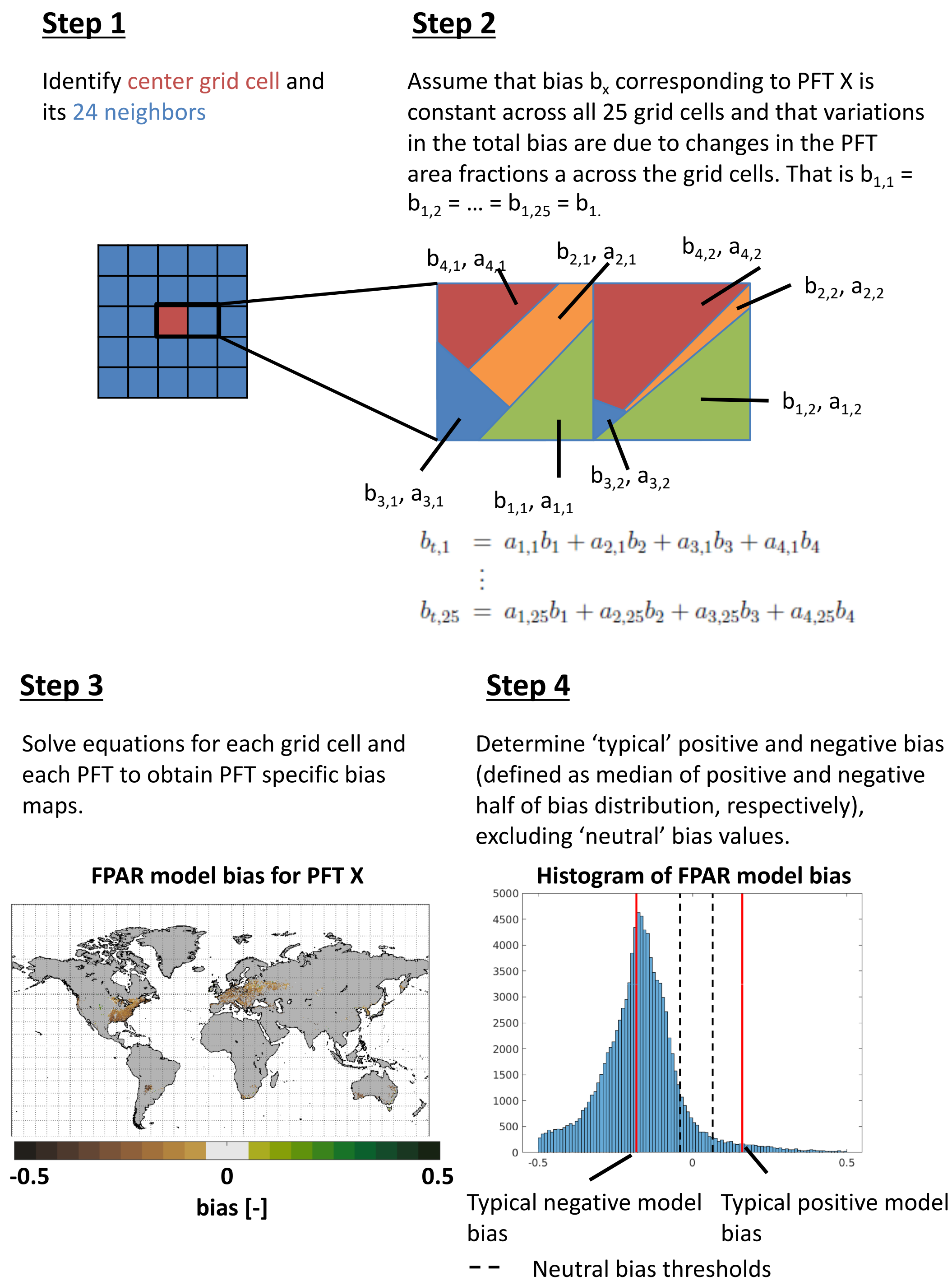


Figure A1. Schematic illustration of the procedure for computing the typical model bias for each subtype, which is used in the calibration location selection process.