Optimizing Carbon Cycle Parameters Drastically Improves Terrestrial Biosphere Model Underestimates of Dryland Mean Net CO_2 Flux and its Inter-Annual Variability

Kashif Mahmud[1], Russell L. Scott[2], Joel A. Biederman[2], Marcy E. Litvak[3], Thomas Kolb[4], Tilden P. Meyers[5], Praveena Krishnan[5,6], Vladislav Bastrikov[7,8], and Natasha MacBean[1]

[1]Department of Geography, Indiana University, Bloomington, IN 47405, USA

[2]Southwest Watershed Research Center, United States Department of Agriculture, Agricultural Research Service, Tucson, AZ 85719, USA

[3]Department of Biology, University of New Mexico, Albuquerque, NM, 87131, USA

[4]School of Forestry, Northern Arizona University, Flagstaff, AZ, 86011, USA

[5]NOAA/ARL Atmospheric Turbulence and Diffusion Division, Oak Ridge, TN, 37830, USA

[6]Oak Ridge Associated Universities, Oak Ridge, TN, 37830, USA

[7]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, F-91191, France

[8]Now at: Science Partners, Paris, 75010, France

Corresponding author: Kashif Mahmud (kmahmud@iu.edu)

Key Points:

- ORCHIDEE terrestrial biosphere model drastically underestimates dryland mean annual net CO_2 fluxes and their inter-annual variability (IAV)
- Optimizing phenology, carbon allocation, and respiration parameters are crucial for capturing net CO_2 flux mean and IAV
- Models need to be optimized against dryland CO_2 flux data to achieve accurate predictions of dryland's role in global C cycle variability

Abstract

Drylands occupy ~40% of the land surface and are thought to dominate global carbon (C) cycle inter-annual variability (IAV). Therefore, it is imperative that global terrestrial biosphere models (TBMs), which form the land component of IPCC earth system models, are able to accurately simulate dryland vegetation and biogeochemical processes. However, compared to more mesic ecosystems, TBMs have not been widely tested or optimized using in situ dryland CO_2 fluxes. Here, we address this gap using a Bayesian data assimilation system and 89 site-years of daily net ecosystem exchange (NEE) data from 12 southwest US Ameriflux sites to optimize the C cycle parameters of the ORCHIDEE TBM. The sites span high elevation forest ecosystems, which are a mean sink of C, and low elevation shrub and grass ecosystems that are either a mean C sink or "pivot" between an annual C sink and source. We find that using the default (prior) model parameters drastically underestimates both the mean annual NEE at the forested mean C sink sites and the NEE IAV across all sites. Our analysis demonstrated that optimizing phenology parameters are particularly useful in improving the model's ability to capture both the magnitude and sign of the NEE IAV. At the forest sites, optimizing C allocation, respiration, and biomass and soil C turnover parameters reduces the underestimate in simulated mean annual NEE. Our study demonstrates that all TBMs need to be calibrated for dryland ecosystems before they are used to determine dryland contributions to global C cycle variability and long-term carbon-climate feedbacks.

Plain Language Summary

Drylands occupy ~40% of the land surface and are thought to dominate the inter-annual variability and long-term trend of the global carbon cycle. Therefore, it is imperative that global terrestrial biosphere models (TBMs) are able to accurately predict dryland vegetation and carbon cycle processes. However, models have not been widely tested or calibrated against in situ dryland ecosystem CO_2 fluxes. Here, we address this gap using a data assimilation system and daily net CO_2 flux data from 12 southwest US Ameriflux sites spanning forest, shrub and grass dryland ecosystems to optimize the carbon cycle related parameters of the ORCHIDEE TBM. We find that before parameter optimization, the model drastically underestimates both the mean annual magnitude and inter-annual variability of net CO_2 flux. By testing different optimization scenarios, we showed that optimizing model parameters related to phenology dramatically

improves the model's ability to capture the net CO_2 flux inter-annual variability. At the high elevation forested sites, optimizing parameters related to C allocation, respiration and biomass and soil C turnover reduces the model underestimate in simulated mean annual NEE. Our study demonstrates that all global TBMs need to be calibrated specifically for dryland ecosystems before they are used to determine dryland contributions to global carbon cycle variability and long-term carbon-climate feedbacks.

1 Introduction

Terrestrial ecosystems currently take up ~30% of anthropogenic CO_2 emissions, thus acting as a substantial global carbon (C) sink (Fu et al., 2017) and providing a critical reduction in the rate of global warming. However, while we know the magnitude of the global C sink to a good degree of certainty, our knowledge of other components of the global C cycle are more uncertain. One such knowledge gap is which ecosystems, and/or which processes, are driving inter-annual variability (IAV) in land net C uptake (Fu et al., 2017). Improving our understanding of the IAV characteristics of the global terrestrial C cycle is key to being able to forecast the future of the land C sink and long-term biosphere-climate feedbacks (Cox et al., 2013; Piao et al., 2019).

Recent studies have pointed to drylands (arid and semi-arid ecosystems) as the dominant driver of global terrestrial C cycle IAV (Ahlström et al., 2015; Humphrey et al., 2021; Poulter et al., 2014). High annual variability in net CO_2 exchange in response to plant-available moisture is observed in site-based flux studies in these regions (Biederman et al., 2017; Cleverly et al., 2016; Haverd et al., 2017; Scott et al., 2015). However, the global terrestrial biosphere models (TBMs) used in these recent C cycle IAV regional attribution studies (and which form the land component of the earth system models used in IPCC projections) have often only been extensively evaluated against data in more mesic ecosystems (e.g. (Peng et al., 2015; Piao et al., 2013; Raczka et al., 2013; Schaefer et al., 2012)), although studies have evaluated models against eddy covariance flux data from Australian dryland sites (Whitley et al., 2016; Teckentrup et al., in review). Similarly, TBM optimization (e.g. parameter calibration) studies have typically focused more on temperate and boreal site data (Kuppel et al., 2014; Raoult et al., 2016).

Therefore, there remains a relative gap in model benchmarking and optimization using dryland C cycle related data.

Model benchmarking and optimization studies that have been performed in dryland regions indicate considerable model-data discrepancies in vegetation dynamics, C and water fluxes (Dahlin et al., 2015; Exbrayat et al., 2018; Haverd et al., 2013; Lawal et al., 2019; MacBean et al., 2015; Renwick et al., 2019; Trudinger et al., 2016; Whitley et al., 2016; Teckentrup et al., in review; Traore et al., 2014; Yang et al., 2021). Whitley et al. (2016) evaluated six TBMs at five savanna flux tower sites along the Northern Australian Tropical Transect and found that accurately representing both tree and grass phenology in TBMs was crucial for simulating seasonal dynamics of leaf area index (LAI) and gross primary productivity (GPP). Traore et al. (2014) showed that long-term positive trends in GPP, fraction of absorbed photosynthetically active radiation (FAPAR – a measure of vegetation dynamics) and evapotranspiration (ET) were underestimated by the ORCHIDEE TBM across the Sahel and Southern Africa, even with a more mechanistic description of soil hydrology. MacBean et al. (2015) showed that calibrating the phenology parameters of the ORCHIDEE TBM (vAR5) using satellite NDVI at global scales could not account for model errors in dryland region seasonal cycle and long-term trends in vegetation dynamics. Forkel et al. (2019) also showed weaker model-data fit for GPP and FAPAR after parameter optimization in semiarid regions. In contrast, Forkel et al. (2014) used parameter optimization to improve seasonal dynamics and long-term trends in vegetation activity in water-limited (and other) biomes predicted by the LPJmL TBM. However, data-constrained modeled phenology only improved when the traditional phenology model schemes (based on growing degree days and water scaling factors) were replaced with a modified version of the empirical "growing season index" (GSI) model (Jolly et al., 2005) that predicts phenological status based on temperature, short-wave radiation and water availability. A recent model evaluation study by MacBean et al. (2021) demonstrated that all the global TBMs participating in the TRENDY v7 model intercomparison project (which have typically not been confronted against a wider variety of data for parameter calibration) drastically underestimate both the mean annual net ecosystem exchange (NEE) and its IAV at a suite of southwestern (SW) US dryland sites due to weak sensitivity of GPP to changing water availability. This analysis is corroborated by Renwick et al. (2019) who also showed that both model phenology parameter optimization and a new semi-deciduous shrub phenology scheme was necessary to

accurately predict the magnitude of GPP in a mixed shrub-grass dryland ecosystem. SW US hydrology modeling studies have also suggested that parameter calibration is needed to realistically represent semi-arid water fluxes because the default parameters hamper model performance (Hogue et al., 2005; MacBean et al., 2020; Unland et al., 1996). Given the lack of model parameter calibration studies that have included a number of dryland ecosystem sites in their optimizations, it remains to be seen whether model-data discrepancies in dryland ecosystem NEE simulations are due to inaccurate model processes or uncertain parameters. Parameter uncertainty may be higher for dryland ecosystems given parameter values were initially measured in the field and/or optimized for more mesic temperate and boreal ecosystems.

To address the gap in dryland site model parameter optimization, and to determine if parameter optimization can account for dryland model-data discrepancies in NEE observed across all TRENDY v7 TBMs (MacBean et al., 2021), we used a Bayesian data assimilation (DA) framework to optimize the photosynthesis, phenology, C allocation and turnover, and respiration parameters of the ORCHIDEE TBM using 89 site-years of daily NEE observations of 12 Ameriflux sites spanning SW US semi-arid grass, shrub and forest ecosystems. Following Biederman et al. (2017) and MacBean et al. (2021), we categorized sites based on their mean annual NEE: US-Vcm, US-Vcp, US-Mpj, US-Fuf, US-Wjs and US-Ses are mostly tree-dominated C sink sites; shrub and grass-dominated sites US-Wkg, US-SRG, US-Seg, US-SRM, and US-Whs "pivot" between a mean annual C sink and source; and the US-Aud grassland is a mean source of C. We used the well-established DA system designed for ORCHIDEE (ORCHIDAS: https://orchidas.lsce.ipsl.fr) (Kuppel et al., 2014; MacBean et al., 2018; Peylin et al., 2016), in which a cost function that represents the misfit between the model and the data – taking into account uncertainty in both – is iteratively minimized using the genetic algorithm (GA; see Methods and Data). Beyond investigating if the DA system could account for model-data discrepancies in dyland NEE simulations, our second objective was to identify which parameters (therefore, which processes) may be responsible for model errors. To address this objective, we performed multiple optimization tests with combinations of parameters related to different model processes in order to identify which processes were most influential in improving the model mean annual NEE and IAV. We focused in particular on which processes are responsible for model failure to capture NEE IAV. Our focus on improving NEE IAV was partly because of the dominant role dryland ecosystems are thought to play in controlling global

C cycle IAV, and partly because we expected that, with the exception of sites that are a strong C sink, eddy covariance estimates of mean annual NEE may be impacted by uncertainties in CO_2 flux partitioning. The methods and data are described in **Section 2** and the results are presented in **Section 3** and discussed in **Section 4**.

2 Methods and Data

2.1 Study sites

Twelve semi-arid eddy covariance flux sites in the southwestern US (SW US) have been utilized in this study, with a measurement period ranging between 2003 and 2014. These sites have a range of different vegetation types, climates, elevation and have been described in detail by Biederman et al. (2017), so we only provide a brief description here. We summarize the sites' description, dominant vegetation species, mean climate and corresponding vegetation plant functional types (PFTs), together with the observation period and disturbance history (**Table 1)**. The sites are listed consecutively based on their mean annual C balance in **Table 1**. The major regional IGBP vegetation classes represented include evergreen needleleaf forest, woody savanna, open and closed shrubland, and grassland. These sites typically experience monsoon rainfall during July to October, preceded by a hot, dry period in May and June. The SW US is characterized by water limitation at the annual scale, i.e. potential ET is greater than precipitation. The sites have large spatial gradients in mean annual precipitation (MAP 250–724 mm) and temperature (MAT 2.9 to 17.7°C) due to interactions among topography, latitude, wind patterns, and distance from oceans. For further site details, see references in **Table 1** and individual site pages on www.ameriflux.lbl.gov.

Table 1. Site descriptions, mean climate, observation years and corresponding vegetation plant functional types (PFTs) used in ORCHIDEE optimization. All years of site data were included in the assimilation with the exception of the final year, which was reserved for validation of the optimized parameters and fluxes (**Section 2.6**). PFT acronyms: BS = Bare soil (PFT=1); TeNE = Temperate Needleleaved Evergreen forest (PFT=4); TeBE = Temperate Broadleaved Evergreen forest (PFT=5); TeBD = Temperate Broadleaved Deciduous forest (PFT=6); C4G = C4 grass (PFT=11). Sites are given in order from largest mean annual C sink (US-Vcm) to mean annual C

175 source (US-Aud). IGBP = International Geosphere–Biosphere Programme; MAP = Mean
176 Annual Precipatation; MAT = Mean Annual Temperature.

Site ID	Description	Dominant species	IGBP class	PFT fractions	Koppen climate	Elevation (m)	MAP (mm)	MAT (°C)	Period of site data	Disturbance History	Site reference
US-Vcm	Valles Caldera mixed conifer forest	Picea engelmannii, Picea pugens, Abies lasiocarpa var. lasiocarpa, Abies concolor	Evergreen needleleaf forest	100% **TeNE**	Dfb	3042	724	2.9	2007–2012	Harvest 1960s	(Anderson-Teixeira et al., 2011)
US-Vcp	Valles Caldera ponderosa forest	Pinus ponderosa, Quercus gambeli	Evergreen needleleaf forest	100% **TeNE**	Dfb	2501	547	5.7	2007–2014	-	(Anderson-Teixeira et al., 2011)
US-Mpj	Heritage Land Conservancy pinyon-juniper	Pinus edulis, Juniperus monosperma	Savanna	20% BS; 60% **TeNE**; 20% **C4G**	Bsk	2200	423	9.6	2008–2014	-	(Anderson-Teixeira et al., 2011)
US-Fuf	Flagstaff unmanaged ponderosa	Pinus ponderosa	Evergreen needleleaf forest	100% **TeNE**	Csb	2215	607	7.1	2006–2010	Harvest 1910	(Dore et al., 2012)
US-Wjs	Tablelands juniper savanna	Juniperus monosperma, Bouteloua gracilis	Savanna	15% **TeNE**; 85% **C4G**	Bsk	1931	349	10.9	2008–2014	-	(Anderson-Teixeira et al., 2011)
US-Ses	Sevilleta creosote shrubland	Larrea tridentata, G. sarothrae	Open shrubland	20% BS; 55% **TeBE**; 25% **C4G**	Bsk	1610	252	12.6	2007–2014	-	(Petrie et al., 2015)

US-Wkg	Walnut Gulch Kendall grassland	Eragrostis lehmanniana, Bouteloua spp. Calliandra eriophylla	Grassland	60% BS; 3% TeBE; 37% **C4G**	Bsk	1529	386	15.8	2004–2013	Drought 2003-2005, non-native grass replacement 2007 onward, light grazing ongoing	(Scott, 2010)
US-SRG	Santa Rita grassland	Eragrostis lehmanniana	Savanna	45% BS; 11% **TeBD**; 44% **C4G**	Bsh	1292	494	16.7	2009–2014	Mesquite removal 1957, ongoing light grazing	(Scott et al., 2009, 2015)
US-Seg	Sevilleta grassland: burned 2009	Bouteloua eriopoda, Gutierrezia sarothrae, Ceratoides lanata	Grassland	40% BS; 60% **C4G**	Bsk	160	250	12.6	2007–2014	Burned 2009	(Petrie et al., 2015)
US-SRM	Santa Rita mesquite savanna	Prosopis velutina, Eragrostis lehmanniana	Woody savanna	50% BS; 35% **TeBD**; 15% **C4G**	Bsk	1122	421	17.7	2004–2014	Light grazing	(Scott et al., 2009)
US-Whs	Walnut Gulch Lucky Hills shrubland	Larrea tridentata, Parthenium incanum, Acacia constricta, Rhus microphylla	Open shrubland	57% BS; 40% **TeBE**; 3% C4G	Bsk	1376	352	16.8	2008–2014	Drought 2005-2006	(Scott, 2010)
US-Aud	Audubon grassland	Boutelou agracilis, B. curtipendula, Eragrostis spp.	Grassland	30% BS; 70% **C4G**	Bsk	1496	348	15.7	2004–2009	Burned 2002	(Krishnan et al., 2012)

2.2 ORCHIDEE terrestrial biosphere model

We used the ORCHIDEE (ORganizing Carbon and Hydrology In Dynamic EcosystEms) process-oriented land surface model version 2.2 that has been developed at the IPSL (Institut

Pierre Simon Laplace, France). The model is a state-of-the-art mechanistic terrestrial biosphere model (Krinner et al., 2005) and is the land surface component of the IPSLCM5 Earth System Model (Dufresne et al., 2013). The model describes the exchanges of water, carbon, and energy between biosphere and atmosphere at the smallest time scale (30 min), while the slow components of the terrestrial carbon cycle (including carbon allocation, autotrophic respiration, foliar onset and senescence, mortality and soil organic matter decomposition) are computed on a daily to annual basis. Version 2.2 is virtually identical to version 2.0, which is being used in the ongoing Coupled Modeling Intercomparison Project 6 (CMIP6) simulations, but includes few recent bug corrections and code enhancements. It has been updated since the "AR5" version used in CMIP5 (see Krinner et al., 2005) with the following developments: i) an 11-layer mechanistic description of soil hydrology and associated modifications as described in MacBean et al. (2020); ii) addition of a coupled carbon-nitrogen scheme (Vuichard et al., 2019); iii) an analytical solution for the set of equations for photosynthesis, stomatal conductivity and internal CO_2 concentration in the leaf (described in Vuichard et al., 2019), following Yin and Struik (2009); iv) an update of the soil thermal properties and extension of the soil depth for heat diffusion (Wang et al., 2016); v) a 3-layer snow scheme (Wang et al., 2013); vi) a spatially explicit observation-derived estimate for background albedo and optimized vegetation and snow albedo coefficients;; vii) a new reconstruction of global land cover history and wood harvest accounting following LUH2v2h maps (Hurtt et al., 2020) and PFT maps based on the European Space Agency Climate Change Initiative Land Cover product (Poulter et al., 2015).

As in most TBMs, the vegetation is grouped into several plant functional types (PFTs), with 14 different types of vegetation plus bare soil in the case of ORCHIDEE v2.2. The original 13 PFTs are reported in Krinner et al. (2005). Since ORCHIDEE v2.0 there are now two extra PFTs included: C3 grasses are now split into three groups - tropical, temperate and boreal. The equations governing individual processes are generic with PFT specific parameters, except for the phenology models (see Appendix A in MacBean et al. (2015)). In this study, ORCHIDEE was mainly used in a "grid-point mode" at each site location and forced with the corresponding local 30-minute gap-filled meteorological forcing data. Before performing the optimizations the modelled C stocks were brought to equilibrium in the spin-up phase by cycling the available site

meteorological forcing over a long period (1300 years) with the default parameters of the model, which ensures a net carbon flux close to zero over annual-to-decadal time scales.

2.3 ORCHIDEE data assimilation system

The ORCHIDEE Data Assimilation System (ORCHIDAS) has been described in detail in previous studies (Bastrikov et al., 2018; Kuppel et al., 2014; MacBean et al., 2018; Peylin et al., 2016), and hence we only briefly define the method here. ORCHIDAS uses a variational data assimilation method to optimize the model parameters, accounting for uncertainties in the observations, the model, and the prior parameters. It relies on a Bayesian framework that uses new information in the observations to update the prior parameter estimates (default values of ORCHIDEE). We find the optimized parameters by minimizing the following cost function *J(x)* (Tarantola, 2005):

$$\boldsymbol{J(x) = \frac{1}{2}[(H(x) - y)^T . \mathbf{R^{-1}} . (H(x) - y) + (x - x^b)^T . \mathbf{B^{-1}} (x - x^b)]} \quad (1)$$

where *x* represents the parameters and *H(x)* the model contingent on the parameters, and *y* the observations. The cost function contains both the misfit between observations,and corresponding model outputs (first term on the right hand side of Eq. 1), and the misfit between a priori parameter values x_b and optimized parameters *x* (second term on the right hand side of the Eq. 1. **R** is the observation error covariance matrix (including measurement and model errors), and **B** is the prior parameter error covariance matrix. Both matrices (**B** and **R**) are diagonal since observation and model errors are assumed to be uncorrelated in space and time, and parameters are assumed to be independent. The cost function is iteratively minimized using the genetic algorithm (GA), which is a meta-heuristic optimization algorithm and follows the principles of genetics and natural selection (Goldberg et al., 1989; Haupt et al., 2004). The GA algorithm has been applied previously with ORCHIDAS tool and described in details by Bastrikov et al. (2018). Briefly, the algorithm works iteratively and considers the vector of parameters as a chromosome and each parameter as a gene on that chromosome. The method fills a set of *n* chromosomes at every iteration, having the starting pool as a randomly perturbed parameter pool. The chromosomes at each subsequent iteration are chosen from randomly selected chromosomes of the previous iteration by either “crossover” or “mutation” process. Santaren et

al. (2014) showed that the performance of the algorithm is highly sensitive to its specific configuration and found the best configuration based on computational efficiency after testing different options. Here, we used the same configuration (i.e. number of chromosomes in the pool total number of parameters optimized; the number of iterations is 40; crossover/mutation ratio is 4:1; the number of gene blocks exchanged during crossover is 2 and the number of genes perturbed during mutation is 1) applied by Santaren et al. (2014) and Bastrikov et al. (2018). The algorithm does not assume prior knowledge of Gaussian probability distribution functions (PDFs) for the observation and parameter uncertainties; however, we do assume Gaussian errors for both **R** and **B** in this study. Given we do not fully know the model uncertainty, we set the prior observation uncertainty as the RMSE between the model and the observations following Kuppel et al. (2014). The prior parameter uncertainties are listed in **Table S1**.

The posterior error covariance matrix of the parameters (**A**) can be estimated by:

$$\mathbf{A} = [\boldsymbol{H}^{\boldsymbol{T}}\mathbf{R}^{-1}\boldsymbol{H} + \mathbf{B}^{-1}]^{-1} \quad (2)$$

This computes error correlations between parameters with the assumption of Gaussian prior errors and linearity of the model in the vicinity of the solution.

2.4 Flux tower measurements

At all twelve SW US sites, flux tower instruments collect 30-minutes measurements of meteorological forcing data and eddy covariance measurements of net surface energy and carbon exchanges, which are available from the AmeriFlux data portal (http://ameriflux.lbl.gov). Meteorological forcing data included air temperature and surface pressure, precipitation, incoming long and shortwave radiation, wind speed, and specific humidity. To run the ORCHIDEE model, we partitioned the in-situ precipitation into rain and snowfall using a temperature threshold of 0°C. The site-level meteorological forcing data were gap filled utilizing downscaled and corrected ERA-Interim data following the approach of Vuichard & Papale (2015). Gross primary productivity (GPP) and the ecosystem respiration (R_{eco}) were estimated from the net ecosystem exchange (NEE) via the flux partitioning method described in Biederman et al. (2016). We acknowledge that GPP and R_{eco} are not fully independent data with respect to NEE and are essentially model-derived estimates, but these concerns have been largely discussed

in previous studies e.g., Desai et al. (2008). Note that in this study, negative NEE refers to net CO_2 uptake into the ecosystem. In order to exclude the influence of the short-term variations in the fluxes on the model optimization, the daily averaged observations smoothed with a 15-day running mean were used in the assimilation as per Bastrikov et al. (2018).

2.5 Parameters optimized

The full set of parameters included in the assimilations optimized are described in Table S1 with their prior values, prior uncertainty, and upper and lower bounds for different plant functional types based on literature analysis, parameter databases and expert knowledge of the model equations. Prior values are the default parameter values used in all non-optimized ORCHIDEE simulations. In the most past ORCHIDAS studies with previous versions of ORCHIDEE, only subsets of ORCHIDEE C cycle parameters have been optimized (Bastrikov et al., 2018; Kuppel et al., 2012, 2014; MacBean et al., 2015; MacBean et al., 2018; Santaren et al., 2007; Verbeeck et al., 2011). In this study, we considered all possible C cycle related ORCHIDEE parameters to fully explore all sources of parameter uncertainty that is contributing to uncertainties in modeled net and gross CO_2 fluxes. We further allowed weak constraints in the DA system (i.e., large prior parameter bounds, albeit within realistic limits) because the main objective of our study was to determine if parameter calibration can improve the model-data fit within the existing model structure and to use our assimilation scenario tests to identify which processes are responsible for model-data errors.

We identified three main groups of parameters: parameters related to 1) phenology; 2) parameters related to photosynthesis; and 3) parameters related to all process calculations that occur after gross C uptake (i.e. C allocation, autotrophic and heterotrophic respiration, biomass and soil C turnover and a scalar on the active soil C pool; hereafter grouped as “post C uptake” parameters). We split the parameters into these three groups because GPP has been shown to be the dominant control on dryland NEE IAV (MacBean et al., 2021); therefore, we expected that optimizing parameters related to one or both of the main two processes controlling GPP (i.e., phenology and photosynthesis) will result in the strongest improvements in NEE IAV. However, optimizing all parameters related to processes that occur after gross C uptake can also influence NEE; therefore, we included these parameters as a third category. The parameters included in

each assimilation scenario are: P1 - all parameters, including all three phenology, photosynthesis and post C uptake parameters; P2 - phenology and photosynthesis parameters; P3 - phenology and post C uptake; P4 - photosynthesis and post C uptake; P5 - phenology parameters only; P6 - photosynthesis only; and P7 - post C uptake only. See Table 2 for a description of all parameters and to which category they belong.

We selected all 102 parameters related to all of the above mentioned processes and divided them into four categories. This resulted in 31 parameters related to photosynthesis, 42 to phenology, 16 to post C uptake (C allocation, respiration, biomass and soil turnover), and 13 related to conductance. In a preliminary study, we tested at several SW US sites (US-Vcp, US-Mpj, US-Fuf, US-Wkg, US-Whs, US-Seg) the sensitivity of the ecosystem fluxes (NEE, GPP and R_{eco}) when optimizing all model parameters and when we just optimized subsets of the parameters related to each of the main processes. This test showed no significant optimization improvement by adding the conductance related parameters (results not shown here), and thus we did not include those parameters for all final optimizations presented in this study, leaving a total of 89 optimized parameters for each site and three process-based parameter categories: 1) phenology; 2) photosynthesis; and 3) post C uptake. Documentation on the parameters can be accessed via ORCHIDEE webpage (https://forge.ipsl.jussieu.fr/orchidee/wiki/Documentation/OrchideeParameters, last access: 04 January 2021). The prior uncertainty was set to 40% of the bounds for each parameter following previous ORCHIDAS studies (Kuppel et al., 2012; MacBean et al., 2015).

2.6 Assimilation Scenarios

We conducted seven different assimilation scenarios to identify which processes (and their related parameters) are potentially causing model-data discrepancies (listed in Table 2). We grouped the optimizations based on various parameters set to optimize. The parameters included in each assimilation scenario are: P1 - all parameters, including all three phenology, photosynthesis and post C uptake parameter groups (89 parameters in total), whereas each consecutive scenario (P2 – P7) optimized different subsets of parameters related to each of the main C cycle processes (Table 2). P2 - phenology and photosynthesis parameters; P3 - phenology and post C uptake; P4 - photosynthesis and post C uptake; P5 - phenology parameters

only; P6 - photosynthesis only; and P7 - post C uptake only. See Table 2 for a description of all parameters and to which category they belong. The parameters that were not optimized were set to their default (prior) value. Comparing the P1 to P7 assimilation scenarios allows us to determine which sets of parameters (i.e. specific processes) are contributing most to the improvement in fluxes as a result of the parameter optimizations and therefore provides insight into which model processes may need further modification or development. See **Table S1** for groupings of model parameters according to specific processes. We did not include the last year of data for each site in the assimilations and used the final year to validate the impact of the calibrated parameter values on net and gross CO_2 fluxes (**Section 2.7**).

Table 2. Description of the different assimilation scenarios conducted in this study. The included parameter group(s) and numbers of parameters for each assimilation scenario are given. Parameters of each subgroup are listed in Table S1.

Optimization	Parameters included	Number of parameters
P1	All parameters (Phenology, Photosynthesis and Post C uptake)	87
P2	Phenology and Photosynthesis	72
P3	Phenology and Post C uptake	57
P4	Photosynthesis and Post C uptake	45
P5	Phenology only	42
P6	Photosynthesis only	30
P7	Post C uptake only	15

2.7 Post-optimization analysis

For all assimilation scenarios we compared the prior simulation (before parameter optimization) to the posterior simulations (after parameter optimization, with different parameter groupings for the different assimilation scenarios) by evaluating the simulations against the site data using standard goodness of fit metrics (root mean square error, RMSE and correlation coefficient, r) at daily, monthly and inter-annual timescales. We further attributed what might be causing model-data misfits by decomposing the daily mean squared deviation (MSD) into its component phase, variance and bias contributions following the approach of Kobayashi and Salam (2000). The bias, variance and phase indicate the mean difference in flux magnitude, the

mismatch in terms of the magnitude of fluctuations, and the seasonality in flux time series, respectively (Kobayashi & Salam, 2000). We calculated the MSD between daily model and observed time series and decompose it following the equation:

$$\boldsymbol{MSD} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\boldsymbol{x_i} - \boldsymbol{y_i})^{\mathbf{2}} = (\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}})^{\mathbf{2}} + (\boldsymbol{\sigma_x} - \boldsymbol{\sigma_y})^{\mathbf{2}} + \mathbf{2}\boldsymbol{\sigma_x}\boldsymbol{\sigma_y}(\mathbf{1} - \boldsymbol{R}) \qquad (\mathbf{3})$$

where x is the model and y is the observations, σ is the standard deviation and $\boldsymbol{R}$ is the correlation coefficient. The first term specifies the bias between model simulation and observation (squared). The second "variance" term measures their differences in terms of variability (i.e., the difference between the magnitude of the modeled and observed fluctuations). The third term in Eq. 3 generally demonstrates the lack of correlation between model and observations weighted by their standard deviations, which can be deemed a measure of their disagreement in terms in phase (Bacour et al., 2019; Gauch et al., 2003). We further calculated the contribution of each component (bias, variance and phase) to the overall MSD by dividing each component by the total MSD. Model evaluation metrics are presented in one of three ways: i) for each site; ii) grouped across all sites; and iii) sites grouped according to their mean net annual CO_2 flux characteristics across the observed time period as in Biederman et al. (2017). For the latter, the net CO_2 "sink" sites are US-Vcm, US-Vcp, US-Mpj, US-Fuf and US-Wjs; the "pivot" sites are US-Ses, US-Wkg, US-SRG, US-SRM, US-Whs, US-Seg; and the "source" site is US-Aud.

We performed a temporal validation at each site and for each assimilation scenario by comparing the modeled and observed daily NEE, GPP and R_{eco} RMSE and R during the validation window (using the final year of data that was excluded from the assimilation). The impact of the temporal validation tests is presented in **Section 3.1** and **3.2**. All other model evaluation metrics refer to the assimilation time window. We also evaluated the posterior parameter values using the limited available trait data close to the sites from the TRY database (Kattge et al., 2020. We searched within 0.5° of the latitude and longitude of each site for trait data related to the species present at each site. We found one estimate of SLA for *pinus ponderosa* close to US-Fuf, and $V_{c,max}$ and leaf longevity estimates for *larrea tridentata* (creosote

shrubs) close to US-Ses. The comparison of these trait values to posterior parameter estimates is presented in **Section 3.3**.

3 Results

3.1 Impact of optimization of all parameters (P1) on model net and gross CO_2 fluxes

Across all sites, the prior ORCHIDEE simulations (i.e. before parameter optimization) fail to capture both the mean annual NEE at mean C sink and source sites and the NEE IAV across all sites (**Figure 1a**) - as also seen for all TRENDY TBMs in MacBean et al (2021). Across all sites, optimizing all C cycle-related parameters (phenology, photosynthesis and post C uptake - assimilation scenario P1) with NEE data dramatically increases the ability of the model to capture both the mean C source/sink behavior and the IAV (**Figure 1b**). C sink and source sites show significant improvement in terms of both mean annual NEE and IAV. At the pivot sites there was no strong bias in NEE with either the prior or posterior parameters (given their mean annual NEE is close to zero); therefore the optimisation has impacted the simulation of NEE IAV rather than the mean annual NEE (as represented by the correlation and slope values shown in inset figures in **Figures 1a and b**).

Improvement of the model-data fit resulting from the assimilation is evident across all sites, with a reduction of daily NEE RMSE between 0.05 to 0.7 $gCm^{-2}d^{-1}$ (**Figure S1)**, with slightly lower reductions in daily GPP and R_{eco} RMSE (**Table S2**). Moreover, the temporal dynamics are well captured for all the sites: when optimizing all parameters, the median Pearson correlation coefficients (R) increase by 0.45, 0.45, and 0.3 for daily, monthly and annual modeled NEE, respectively and posterior median slope increase by ≥ 0.35 at all timescales (**Figure S2a and d)**. GPP temporal dynamics are also much improved by the P1 assimilation with a higher median value and tighter range in posterior R and slope values at all timescales (**Figure S2b and e**). In contrast, there is less improvement in R_{eco} temporal dynamics although the median R and slope values are higher after the optimization with the exception of the annual values (**Figure S2c and f**).

The median daily NEE RMSE and R for the temporal validation analysis indicates that the optimized parameters maintain an improved model-data fit outside the assimilation window when compared to the prior (**Figure S3**). The median value of daily NEE RMSE is 0.1 higher for

the validation test compared to the assimilation for the P1 assimilation scenario; however, the maximum to minimum range of RMSE values in the validation is very similar to original optimization and much less than the prior simulation (**Figure S3a**). Similarly, the median R of daily NEE is slightly less for the validation test than the original optimization for P1 (**Figure S3b**). The daily GPP RMSE and R show similar model-data fit for the validation analysis of the P1 optimized parameters as the original optimizations (results not shown). However, the median daily R_{eco} is the same for the prior, optimization and validation, while there is an increasing improvements in the median daily R_{eco} R for both the P1 optimization and validation (results not shown).

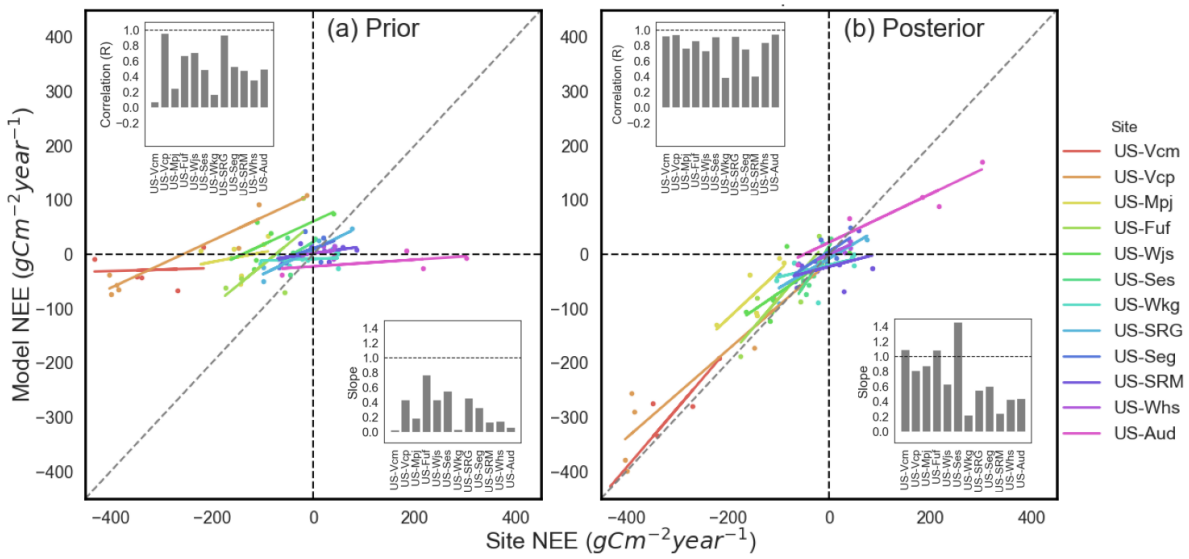

Figure 1. Comparison between modeled and observed annual NEE when assimilating NEE data and optimizing all phenology, photosynthesis and post C uptake parameters (P1) in the same assimilation. (**a**) Prior annual NEE simulation before parameter optimization, and (**b**) Posterior annual NEE after optimization. The trendline and slope value for the linear regression between the model and observations (bottom right inset figures) is shown for each site, together with their Pearson correlation coefficient, R (top left inset figures). The middle of the trend line should sit on the 1:1 line if the accurate mean annual source/sink behavior for a site is well captured by the model. A slope value close to or equal to 1 demonstrates the model is better at capturing the IAV. Colored points and trend lines represent all twelve sites, ordered from the largest mean sink (US-Vcm) to the largest mean source (US-Aud). The sink sites are: US-Vcm, US-Vcp, US-Mpj, US-Fuf, US-Wjs and US-Ses; the pivot sites are: US-Wkg, US-SRG, US-Seg, US-SRM and US-Whs; and the only source site is: US-Aud.

Across the majority of the sites, the prior model simulates a depressed seasonal NEE amplitude and/or is unable to capture the observed bi-modal seasonality (**Figure 2**). The NEE amplitude and bi-modal seasonality generally improved when optimizing all parameters (blue curves in **Figure 2**), although the posterior simulations struggle to reach the exact magnitude of the spring and monsoon NEE troughs (net CO_2 uptake) for several sites (e.g. US-Mpj, US-Wjs, US-Ses, US-Seg, US-Wkg and US-Whs). Accurately capturing the seasonal peaks and troughs is important for replicating observed NEE IAV because variability in summer monsoon season fluxes are the dominant driver of NEE IAV (MacBean et al., 2021). While posterior seasonal NEE peaks and troughs are generally well captured, the assimilation of NEE alone often fails to capture the correct peaks in gross CO_2 fluxes (**Figure S4**), likely due to compensating errors in both GPP and R_{eco}. We note however that the mean seasonal cycle for the gross CO_2 fluxes is generally much improved, especially for low-elevation "pivot" sites with a clear bi-modal growing season (e.g., US-Wkg, US-SRM, US-SRG, and US-Whs: **Figure S4**). At the C source site (US-Aud) the model also fails to simulate the accurate peak in mean springtime net carbon release (**Figure 2**). This is due to the fact that at US-Aud, TBMs tend to overestimate spring GPP and underestimate the earlier rise in spring R_{eco} (**Figure S4**). The optimization only partially corrects these model biases, suggesting that other missing processes may ultimately be responsible for the model-data misfit (such as disturbance following a fire that occurred at the site in 2002, which is not implemented in the current version of ORCHIDEE).

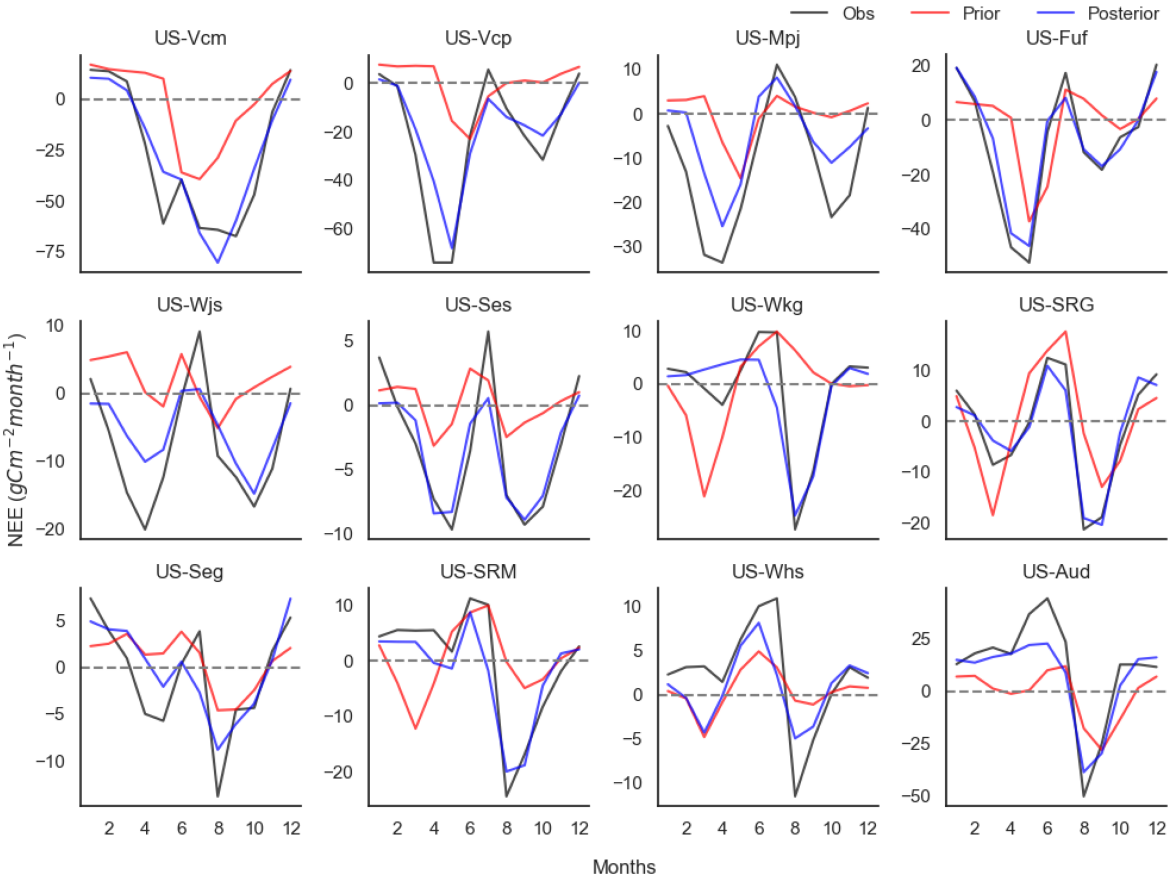

Figure 2. Mean monthly NEE seasonal cycles for each site comparing prior (red curve) and posterior (blue curve) ORCHIDEE simulations with observations (black curve). Posterior simulation after assimilation of NEE data and optimization of all parameters: phenology, photosynthesis and post C uptake (P1). The sites are listed in order from largest mean annual C sink (US-Vcm) to mean annual C source (US-Aud).

Decomposing the daily NEE MSD between model and observations into bias, variance and phase components shows that across all sites, all three components contribute to prior NEE model-data discrepancies (**Figure 3a** left of vertical dashed line). The prior daily NEE MSD at the C sink sites are dominated by both phase and bias components (**Figure 3a** top panel). The fact that the sink sites' NEE MSD is also dominated by bias is unsurprising given that at those sites the prior model does not capture the mean annual C sink (**Figure 1a**). Note that, if we decompose the *annual* NEE MSD into the constituent bias, phase and variance components then bias overwhelmingly dominates the MSD at sink (and source) sites given their large underestimate of mean annual NEE (**Figure S5** top and bottom rows). In contrast, at the C pivot

and source sites, the highest contribution to the prior daily NEE MSD is from the phase component (**Figures 3**d and g), indicating that the default model does a poor job of representing the timing of dryland C cycle related processes. Across all sites, optimizing all parameters (P1) dramatically reduces the bias, variance and phase components of the daily NEE MSD, with phase remaining the strongest contributor to daily NEE MSD (**Figures 3a, d and g** right of dashed line).

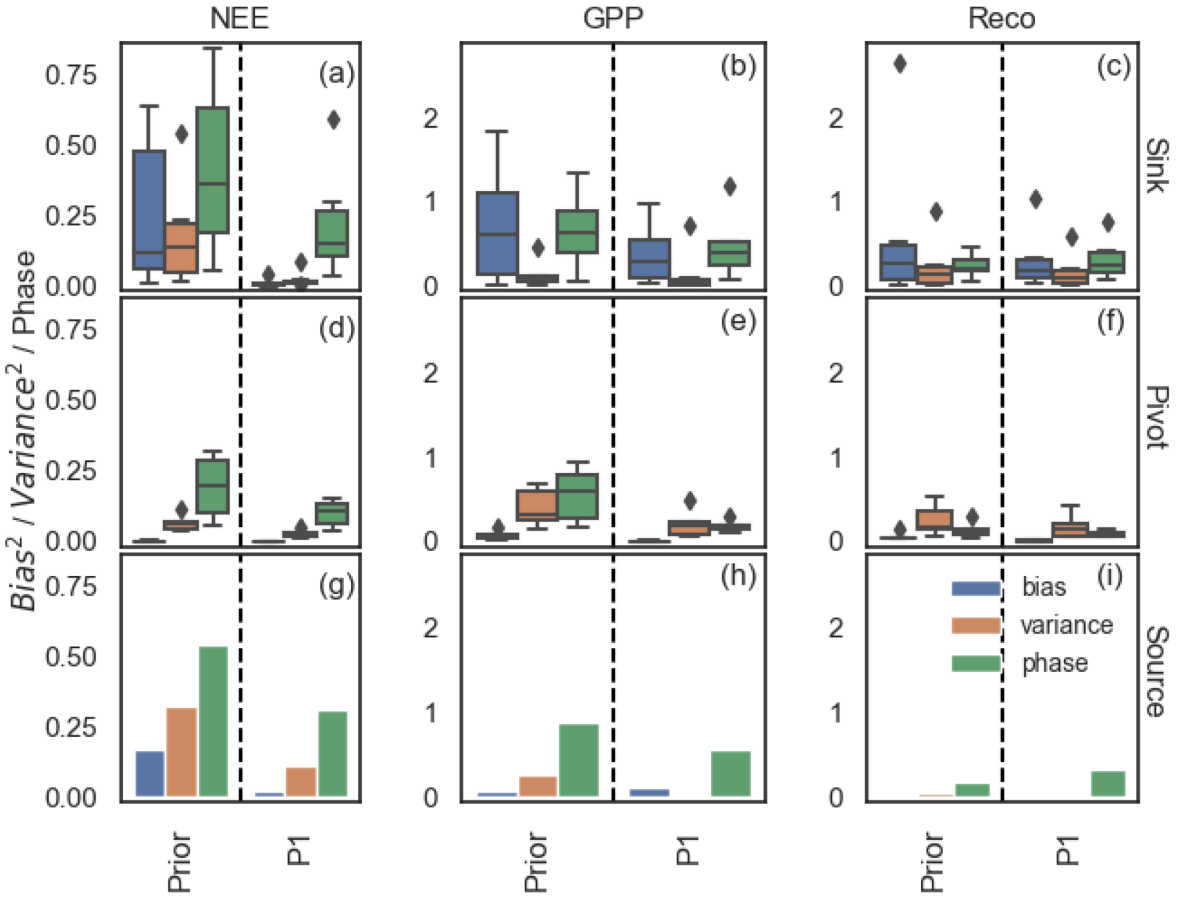

Figure 3. Daily NEE, GPP and R_{eco} mean square deviation (MSD) decomposition into bias, variance, and phase between simulations and observations for assimilating NEE observations and optimizing all phenology, photosynthesis and post C uptake parameters (P1). Blue, orange and green boxplots for bias, variance and phase components, respectively. Different rows separate the sites as sink (a-c), pivot (d-f) and source (g-i) based on total annual C flux. The sink sites are: US-Vcm, US-Vcp, US-Mpj, US-Fuf, US-Wjs and US-Ses; the pivot sites are: US-Wkg, US-SRG, US-Seg, US-SRM and US-Whs; and the source site ia: US-Aud. The x axes display the optimization scenarios (Prior and P1). The box whiskers show the spread of bias, variance and phase for all 12 sites considered in this study. The bias, variance and phase indicate the mean difference in flux magnitude, the mismatch in terms of flux fluctuation magnitude scales with the

476 mean seasonal amplitude, and the seasonality in flux time series, respectively. Note that the y
477 axis limits for both gross fluxes (GPP and R_{eco}) are the same.

478

479 The bias and phase are the dominant contributors to prior daily GPP MSD for the sink
480 sites (left of vertical dashed line in **Figure 3b**), and phase only for the pivot and source sites
481 (**Figures 3e and h**). For R_{eco}, a different MSD component is dominant depending on the mean C
482 behavior of a site: bias dominates the prior daily R_{eco} MSD at the sink sites, variance at the pivot
483 sites, and phase at the source sites (**Figures 3c, f and i**). Overall, assimilating NEE data in the P1
484 assimilation scenario reduces all gross CO_2 flux MSD components (right of dashed line in
485 **Figure 3** middle and right columns), with phase remaining the strongest contributor to daily
486 gross CO_2 flux MSD at sink and source sites. However, unlike for the NEE, at the C sink sites
487 phase *and* bias remain strong contributors to posterior GPP MSD (**Figure 3b**).

488

489 3.2 Impact of different processes (assimilation scenarios) on optimization results

490 Across all sites, modeled annual and seasonal NEE are improved the most in the P1
491 assimilation scenario, although all scenarios result in some improvement (**Figures S6, S7a and**
492 **d,** and seasonal cycles in **Figure S8**). In general, there is less improvement in R_{eco} compared to
493 NEE and GPP (**Figure S7**). Examining the daily NEE median RMSE for the temporal validation
494 analysis for the P2 to P7 assimilation scenarios shows that the optimized parameters have
495 improved the model-data fit outside the assimilation window when compared to the prior, with
496 the exception of scenarios that include photosynthesis or post C uptake parameters (e.g., P2, P4,
497 P6 and P7 - **Figure S3a**). However, the range of RMSE values from the validation tests is similar
498 to the original optimization and much less than the prior simulation for all optimization
499 scenarios. Similarly, the median R between modeled and observed daily NEE for the validation
500 test is higher than the prior for all assimilation scenarios and is close to the optimized median R
501 (within ±0.1) for P2, P3, P5 and P7 (**Figure S3b**). The 25^{th} percentile of the range in R values is
502 generally higher than the prior for P2, P3, and P5, but not for P4, P6, and P7 (**Figure S3b**),
503 which, again, are scenarios that include photosynthesis or post C uptake parameters but not
504 parameters related to phenology.

Comparing the MSD decomposition results for the various assimilation scenarios (P1-P7) can help to identify which processes may be causing the prior model-discrepancies in mean annual NEE and NEE IAV. At the source and sink sites, the bias component (blue bars in **Figure 4a and c**) reduced dramatically (median squared bias across sink sites reduced by 90% and the source site by 80%) by all optimization tests that include the post C uptake parameters related to C allocation, respiration, and aboveground biomass and soil C turnover (P1, P3, P4 and P7). For the sink sites, assimilation scenarios that also include photosynthesis parameters (P2 and P6) also result in a strong reduction in bias (median squared bias reduction of 50%). This decrease in mean bias is also shown by the fact that the midpoints of the linear regression trendline between model and observations at forested sink sites (US-Vcm, US-Vcp, US-Mpj, and US-Fuf) and low-elevation source site (US-Aud) with optimization scenarios P1 to P4, P6 and P7 parameters all lie much closer to the 1:1 (grey dashed) line compared to P5 (**Figure S6**).

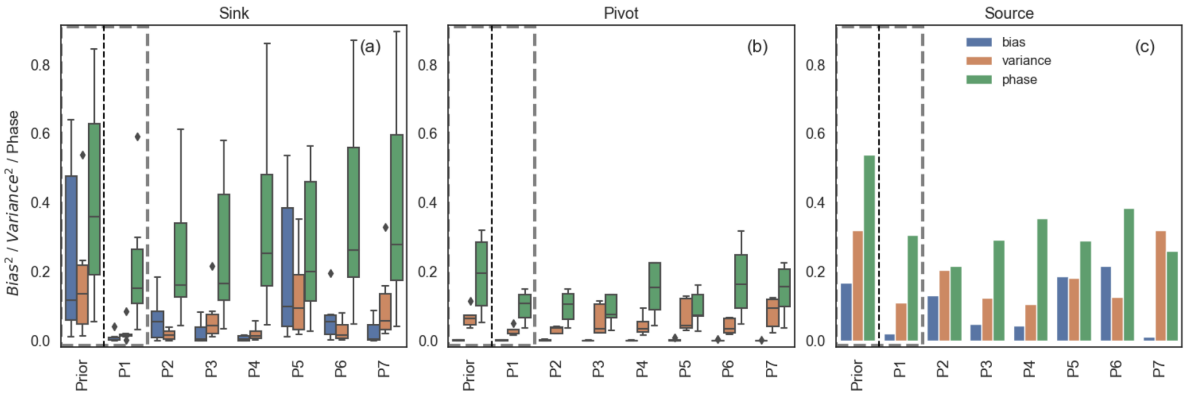

Figure 4. Daily NEE MSD decomposition into bias, variance, and phase components when assimilating NEE observations for different assimilation scenarios (P1-P7). Different panels separate the sites as sink (a), pivot (b) and source (c) based on total annual C flux. The C sink sites are: US-Vcm, US-Vcp, US-Mpj, US-Fuf, US-Wjs and US-Ses; the C pivot sites are: US-Wkg, US-SRG, US-Seg, US-SRM and US-Whs; and the C source site is: US-Aud. The grey dashed boxes highlight results repeated from **Figure 3(a,d,g)** to have better comparison of different process parameters side-by-side. The parameters included in each optimization are: P1: all parameters; P2: phenology and photosynthesis; P3: phenology and post C uptake; P4: photosynthesis and post C uptake; P5: phenology; P6: photosynthesis and P7: post C uptake. The boxplots show the median and interquartile range of the bias, variance and phase across all 12 sites considered in this study. US-Aud is the only C source site; therefore, the barplots in (c) show the bias, phase, and variance components of the MSD for that one site. The bias, variance and phase indicate the mean difference in flux magnitude, the difference in the magnitude of flux variations, and the difference in the correlations weighted by the standard deviations, respectively (see Methods).

533

534 Across all sites the difference in phase between the model and observations (green bars in
535 **Figure 4**), which, as already noted, is the largest contribution to the prior NEE MSD across all
536 sites, is mostly reduced by assimilation scenarios that include phenology parameters (i.e. P1, P2,
537 P3 and P5). The P4 assimilation (photosynthesis and post C uptake parameters) also does well in
538 reducing phase contributions to NEE MSD at forested C sink sites (**Figure 4a**). However, the
539 phase component is not reduced as much as the bias in any of the assimilation scenarios; thus, for
540 all sites and all assimilation scenarios the phase remains the largest component of the posterior
541 daily NEE MSD (**Figure 4**). Including parameters related to photosynthesis or post C uptake
542 with the phenology parameters (i.e. assimilation scenarios P2 and P3) helps to slightly reduce the
543 median phase discrepancy at sink sites compared with phenology parameters alone (P5) (**Figure**
544 **4a**). Examining the spread in slope and R values across all sites, we see that the annual
545 variability (median slope and R values) is improved the most for assimilation scenarios with at
546 least two parameter sets (P1 to P4 - **Figure S7a and d**). The persistence of phase as the dominant
547 component of the posterior daily NEE suggests further model improvement in processes related
548 to dryland vegetation temporal dynamics (e.g. phenology and all associated processes) is needed
549 before TBMs can correctly reproduce NEE seasonality and IAV.

550 The variance component of the daily NEE MSD (orange bars in **Figure 4**), which shows
551 a modest contribution to daily NEE MSD at the sink and source sites, is mostly reduced at the
552 sink sites with assimilation scenarios that include photosynthesis parameters (i.e. P1, P2, P4 and
553 P6 - **Figure 4a**). At US-Aud, which had a larger prior variance component than bias, the
554 posterior variance was reduced by assimilation scenarios that tended to include photosynthesis or
555 post C uptake parameters (i.e. P1, P3, P4 and P6) (**Figure 4c**).

556 While the post C uptake parameters are key for reducing bias in forested sink site NEE,
557 biases in GPP and R_{eco} at these sites are reduced by optimizing photosynthesis parameters (P1,
558 P2, P4, and P6 - blue boxes **Figure S9b and c**). The GPP and R_{eco} bias components at the sink
559 sites are not reduced as strongly as NEE biases for any assimilation scenario; thus, bias remains a
560 key contributor to posterior gross CO_2 flux MSD. Similarly to NEE, parameter subsets that
561 include phenology parameters (P1, P2, P3 and P5) are key for reducing the daily GPP MSD sink
562 and phase component at pivot sites (green boxes in **Figure S9e**; also see median GPP slope and

563 R values in **Figures S7b and e**). With the exception of P1 and P2 for GPP, the GPP and R_{eco}
564 variance components are not reduced much by any of the assimilation scenarios and remain a
565 considerable component of the MSD for both GPP and R_{eco} at the pivot sites, and for R_{eco} at the
566 sink sites (**Figures S9b,c,e,f**).

567 3.3 Constraint on parameters

568 For all assimilation scenarios, we found significant parameter deviations from prior
569 values for numerous phenology, photosynthesis and post C uptake related parameters (**Figures**
570 **5a and S10a**), which is consistent with the fact that all parameter subsets are needed to improve
571 model mean annual NEE and IAV. Parameter deviation was calculated using the difference
572 between the posterior and prior parameter value normalized by the total parameter variation used
573 in the optimization. Finally, the median value was taken as the mean deviation from all 12 sites.
574 We did not find that parameters deviate more, or the uncertainty reduction (calculated as 1 –
575 (posterior parameter uncertainty / prior parameter uncertainty)) was much different, when only
576 one subset or two parameter subsets were included in the optimization instead of all three (e.g.
577 cf. P2 with P1), although posterior values are different for each assimilation scenario (**Figure**
578 **S10**). In particular, most of the post C uptake parameters deviate strongly from the prior median
579 deviations (>20% of total parameter bound). There are also significant uncertainty reductions
580 (>50%) for most of the parameters which show strong deviations from their prior value: for P1
581 for example there are 10 for phenology (out of 42), 6 for photosynthesis (out of 31) and 8 for
582 post C uptake (out of 16) (denoted by asterisks in **Figure S10**). By grouping all parameters
583 according to their respective processes we found that phenology and post C uptake parameters
584 had the strongest uncertainty reductions across all assimilation scenarios, while spread in
585 reduction in photosynthesis parameter uncertainty is high (**Figure S11**). The error correlations
586 between the estimated parameters are usually minimal except for parameters involved in the
587 empirical calculation of the moisture stress function on soil C decomposition (e.g.,
588 "moist_coeff_") (see example for one site in **Figure S12**).

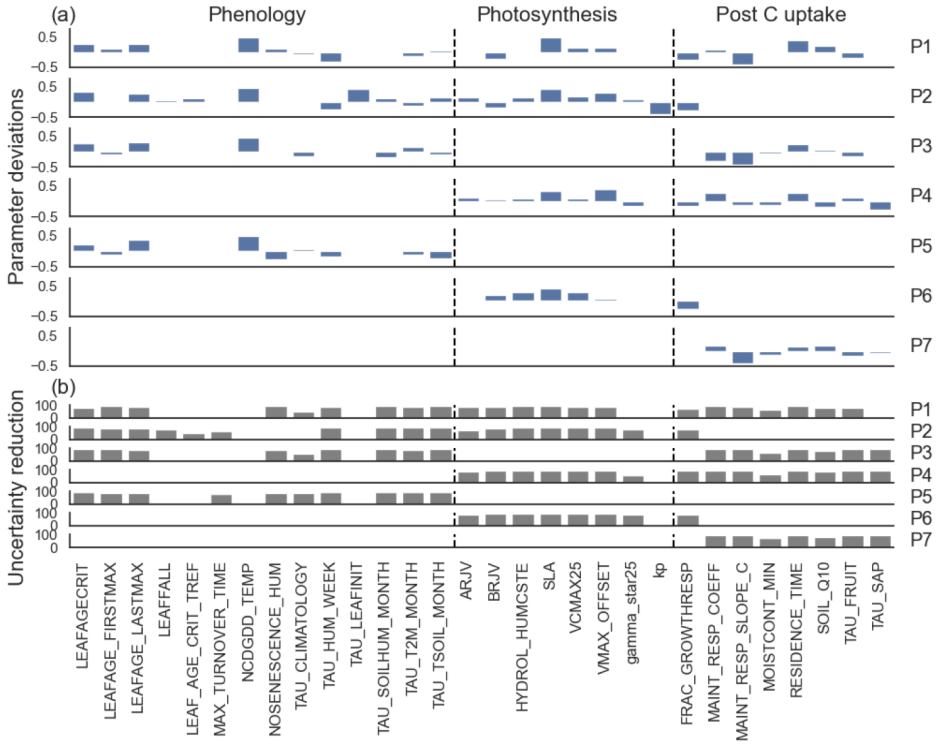

Figure 5. Optimized median parameter deviations [(posterior - prior) / (max - min)] (blue bars) and associated median parameter uncertainty reductions (grey bars) for parameters (having deviations > ±0.3 or uncertainty reduction > 50%) controlling phenology, photosynthesis and post C uptake assimilating NEE data (P1-P7). Bars represent the median across all 12 sites. The asterisks above blue bars indicate the parameters that have larger than 50% uncertainty reduction. Each line corresponds to a specific optimization test (shown on the right axis). The parameters are given on the bottom axis. The vertical dashed lines separate the different parameter subsets (phenology, photosynthesis and post C uptake). **Table S1** details the prior and posterior parameter values and their uncertainty for all parameters together with the maximum and minimum bounds used in the optimizations.

Certain phenology parameters are important across all assimilation scenarios: i) parameters related to leaf age; ii) a parameter related to the critical temperature threshold for the start of deciduous shrub leaf growth (ncdgdd_temp); iii) moisture thresholds that govern C4 grass senescence (nosenescence_hum); and iv) various parameters that control the time scales used in phenology schemes (e.g. tau_climatology, tau_hum_week) (**Figure 5**).The phenology models are highly dependent on such empirical parameters, which likely need to be optimized for each site. Key photosynthesis related parameters are SLA, parameters involved in the calculation of $V_{c,max}$ (the maximum carboxylation rate, which has been shown to be a highly sensitive model parameter in previous studies, e.g. Kuppel et al., 2014), and the parameter that

represents the root profile in the empirical calculation of leaf water stress (hydrol_humcste), which downregulates photosynthesis and stomatal conductance in the dry season. The most important post C uptake parameters are fairly similar across assimilation scenario tests, and are related to: i) the calculation of the growth and maintenance respiration as a fraction of biomass; ii) aboveground biomass residence time and various turnover rates for biomass and litter pools; and finally, iii) the Q10 parameter involved in the temperature dependence of soil C decomposition (**Figure 5**). We compared the posterior parameter values for all relevant assimilation scenarios to the available trait data from the TRY database (**Figure S13**). The two photosynthesis traits (*SLA* for *pinus ponderosa* close to US-Fuf and $V_{c,max}$ for *larrea tridentata* close to US-Ses were over- and underestimated by the posterior values across all assimilation scenarios, respectively, whereas the leaf longevity for *larrea tridentata* measured close to US-Ses was well captured by the P3 assimilation scenario. There is not enough trait information to perform a rigorous validation of the posterior parameter values. The existing measurements may differ from the model due to the fact the traits were not mentioned at the same location. However, mismatches between the posterior parameter values and traits presented here highlights that we need to collect more trait data with which to evaluate the optimized parameters, in addition to using the DA framework to explore how parameters may vary over space and time. We discuss this further in **Section 4**.

4 Discussion and Conclusions

4.1 Further testing and developments needed to improve modeling of dryland C cycling

In this study, we have shown that it is possible to account for model discrepancies in both the mean annual NEE and NEE IAV at a range of semi-arid SW US sites via optimization of C cycle parameters within a Bayesian DA framework. We used weak prior constraints (i.e. large prior parameter bounds) to give the assimilation the maximum chance to correct any model errors. Our goal was not to identify the ideal "correct" set of C cycle parameters for capturing semi-arid vegetation and C cycle dynamics, but rather to identify whether, within the current model representation, we could account for model-data mismatches. Looking at the individual parameter plots for the P1 assimilation scenario (**Figure S14)**, we find that at some sites several posterior parameters are "edge-hitting" (e.g. soil Q10). Given we chose weak prior constraints in the assimilation, the fact that some posterior parameters are hitting their bounds suggests that the

optimization may be aliasing model structural error onto the parameters (as demonstrated in MacBean et al., 2016 and Wutzler and Carvalhais, 2014) and/or that the model cannot improve further via parameter optimization. This suggests that further model developments are likely needed to address structural uncertainties and missing processes, which will then need to be followed up with additional parameter DA experiments to ensure increasing complexity does not degrade model skill (Famiglietti et al., 2021). We know for example that certain important processes for sparsely vegetated, mixed shrub- and grass-dominated dryland ecosystems, such as wildfires (Exbrayat et al., 2018; Lasslop et al., 2016; Whitley et al., 2017) and biological soil crust C cycling (Belnap et al., 2016), are currently not represented in most TBMs. Exbrayat et al. (2018) showed using a Bayesian parameter DA experiment that model simulations with fire had faster carbon turnover times and increased C allocation to wood and root pools (rather than foliage) than the simulations without fire – all of which resulted in changes to GPP, net primary productivity, biomass and carbon use efficiency. Their results neatly demonstrate that errors due to missing model processes can be aliased onto the posterior parameter values.

Hypotheses as to which processes might be responsible for model inability to capture semi-arid CO_2 flux IAV – and therefore which processes need further development in the model – are numerous and will take time to explore fully. MacBean et al. (2021) suggested that the following processes might be causing model errors in capturing semi-arid C cycle dynamics: the lack of drought-deciduous shrub phenology schemes in TBMs (Renwick et al., 2019); the lack of deep tap roots for trees and shrubs that draw up groundwater needed for growth during drier periods (Gibbens and Lenz, 2001; Kerhoulas, et al., 2012); the lack of dynamic root growth or hydraulic redistribution as soil moisture changes with depth (De Kauwe et al., 2015; Fu et al., 2016; Lee et al., 2018; Li et al., 2012; Whitley et al., 2016; 2017); and inaccurate nutrient limitation in dryland ecosystems (Sun et al., 2021; Hooper & Johnson, 1999). Future studies need to systematically test all these options to determine which, if any, can explain the observed model-data discrepancies. Here, we aimed to facilitate our understanding of which processes may be responsible for errors in modeling of semi-arid C fluxes by using the different assimilation scenarios as tests of which parameter sets (and therefore, which processes) most improve the model-data mismatch. The assimilation with all C cycle and vegetation parameters (P1) performed the best in terms of correcting underestimates in modeled mean annual NEE and IAV. However, the additional assimilation scenarios (P2 to P7) further demonstrated that

671 phenology parameters are likely key for improving semi-arid ecosystem NEE IAV. Issues with
672 semi-arid phenology in TBMs have been documented elsewhere (Traore et al., 2014; Dahlin et
673 al., 2015; MacBean et al., 2015; Renwick et al., 2019; Whitley et al., 2016; Teckentrup et al., in
674 review). In addition, Wu et al. (2018) found that TBMs underestimate vegetation productivity
675 responses to increased precipitation at grassland sites. Further evidence for inadequate TBM
676 phenology schemes comes from MacBean et al. (2020), who noted that while the ORCHIDEE
677 model can capture evapotranspiration (ET) fluxes extremely well, even without parameter
678 optimization, the model simulates a delayed increase in transpiration/ET (T/ET) ratios during the
679 summer monsoon when compared to two independent T/ET estimates. This suggests that the
680 model is getting ET right for the wrong reasons – i.e. the partitioning of ET into its component
681 fluxes of T and bare soil evaporation is incorrect. This lagged response of T to increasing rainfall
682 is consistent with the results of MacBean et al. (2021) who found across a suite of TBMs
683 (TRENDY v7) too weak ecosystem-scale water use efficiency (WUE) – i.e. a too weak response
684 of GPP to increasing ET – during the monsoon was likely the cause of their inability to capture
685 NEE IAV. Put simply, the models simulate too weak a response of vegetation growth to pulses
686 of moisture availability. Thus, the evidence from all these studies, including our results presented
687 here, is pointing to issues with processes controlling seasonal vegetation dynamics such as
688 phenology and plant hydraulics schemes that controls plant water stress.

689 Another source of error in the model NEE IAV simulations could be related to the
690 fractional cover (fCover) of different PFTs prescribed in the model. Although we used site-based
691 estimates of PFT fCover, these estimates typically represent the spatio-temporal average fCover
692 at each site (as is often the case in coarse-scale (>30m) satellite fCover estimates - Brandt et al.,
693 2016). In contrast, the PFT fCover prescribed in TBMs should be the maximum possible fCover:
694 The models then limit the growth of vegetation based on climate conditions and other resource
695 availability. In the lower elevation sites, the *in situ* fCover estimates suggest a high fraction of
696 bare soil at each site; however, in years with strong monsoon rainfall, growth of summer annual
697 C4 grasses will fill most of bare soil patches, resulting in a much lower bare soil fCover during
698 those periods. Therefore, the static PFT fCover prescribed in the models based on the *in situ*
699 estimates from each site likely prevent monsoon season growth of summer annual C4 grasses in
700 the interstitial bare soil patches that can vary year to year depending on monsoon rainfall
701 variability. It is possible that this issue of static PFT fractions based on spatio-temporally

averaged *in situ* estimates explains the model's inability to capture peak GPP fluxes for some sites, and the fact that even in the posterior simulations, the phase remains the strongest contribution to the NEE MSD. Errors in PFT fractions in sparsely vegetated regions have also been shown to propagate into large model errors in simulated carbon, water and energy fluxes (Hartley et al., 2017). The optimization of numerous phenology parameters with weak constraints in this study could be partially accounting for such a model error in spatially heterogeneous dryland ecosystems. Future simulations across all sites should be run with prescribed fCover that captures the maximum vegetation growth that is possible at the site, which will likely require new vegetation fCover classifications specifically for particularly wet time periods.

The same Bayesian DA system was used by MacBean et al (2015) to correct phenology model issues in a previous version of ORCHIDEE that was nonetheless identical in its representation of phenology. However, while they were able to correct the seasonal leaf dynamics in temperate and boreal ecosystems, they found the parameter optimization was unable to correct for phenology model issues in semi-arid ecosystems. While the data they used were different – normalized difference vegetation index (NDVI) from the MODIS satellite instrument as opposed to the flux tower NEE used here – they also used stronger prior constraints and fewer phenology parameters, suggesting that the additional degrees of freedom in the assimilations in this study (from weaker prior constraints and a greater number of phenology parameters) may have resulted in the improvements from the parameter optimization. In future studies we will test the combination of both NEE and NDVI, in addition to other proxy measurements of GPP such as solar induced chlorophyll fluorescence data, for improving ORCHIDEE vegetation dynamics in drylands. Still, as we noted above, the combination of weak prior constraints and edge-hitting posterior parameters suggests the assimilations are accounting for other structural errors in the model, and phase errors remain a strong source of NEE MSD even after optimization. As also noted, the phenology schemes in these models are highly dependent on a number of empirical parameters that require site calibration and which were typically not developed for dryland ecosystems. Future developments in this area should take account of the variety of different strategies in dryland plants for responding to highly variable water availability and water stress (Smith et al., 2012).

Our assimilation tests also showed that so-called "post C uptake" parameters related to maintenance respiration, biomass and litter turnover, and soil C decomposition are mainly responsible for reducing the strong model underestimate of mean annual NEE, particularly at the higher elevation forested C sink sites. Our key focus was not on correcting the mean annual NEE, and instead was more focused on correcting errors in NEE IAV, because the variability in eddy covariance measurements of NEE are more trusted than the absolute values due to errors in flux partitioning. Furthermore, for the semi-arid sites that pivot between a C source and sink, their mean sink versus source behavior may be a function of a time period involved. In particular, the only mean C source site (US-Aud) is likely a source because of a fire in 2002 from which the site was still recovering during the measurement period (Krishnan et al., 2012). As discussed, we know that even TBMs that include wildfire modules will likely not reproduce the specific impacts of an individual fire. Including an additional "$k_{biomass}$" parameter in the assimilations that, similar to k_{soilC} for soil C, scales the initial aboveground biomass pools could help to account for the impact unknown disturbances on changes vegetation cover and C flux dynamics. This needs to be tested in future DA experiments. Nevertheless, while we do not focus on the C source site, we do know that the high elevation forested sites in this study are consistently sinks of C, even during the drought period that has been affecting the SW US for most of this century (Scott et al., 2015). It is important that we are able to capture this dryland forested site C sink, particularly given these ecosystems have been shown to contribute to long-term trends in the global C cycle (Ahlstrom et al., 2015). Drylands are vulnerable to future increases in drought, which may reduce the C sink (Bodner and Robles, 2017). On the other hand, drought impacts on dryland vegetation could be mitigated by increases in WUE and vegetation growth under elevated CO_2 (e.g. Donohue et al., 2013). Thus, it is an important contribution that parameter optimizations presented here can account for these biases in modeling C sinks at high elevation forested sink sites. MacBean et al. (2021) postulated that TRENDY TBM underestimates in mean annual NEE at these sites was due to underestimates in spring GPP, possibly due to issues with model snow melt not providing enough moisture for spring growth. In contrast, the results presented here suggest that the biases at the high elevation forested sink sites may be more linked to processes that occur after the gross uptake of CO_2, such as growth and maintenance respiration, biomass turnover, and temperature limitation on soil C decomposition (**Figure S14**). It may be that TBMs can accurately capture dryland forested site

mean annual NEE if the parameters related to C respiration, allocation, biomass turnover and decomposition are better adapted for dryland PFTs, which simply requires more careful calibration across a range of dryland forest sites. Additional observations of: i) snow cover and snow melt; ii) autotrophic and heterotrophic respiration; and iii) above and belowground C stocks are needed to assess whether the parameter calibration is accounting for model biases in mean annual NEE. With this additional information we can start to tease apart if the different processes that contribute to the forested site mean C sink are well represented in the model.

4.2 DA for improving our understanding of dryland ecosystem processes

Our posterior parameter analysis points to the parameters that are most important for controlling net CO_2 flux dynamics (**Section 3.3**). Given the paucity of trait data for the SW US (**Section 2.7**), these results can guide further field trait data collection efforts that are needed to validate the posterior parameter values. The data assimilation experiments performed in this study can also provide additional information on how traits may vary across dryland species. The spread in posterior parameter values across all sites for each of the main PFTs at each site: evergreen trees, shrubs, and C4 grasses is considerable for almost all parameters **(Figure S16)**, particularly for phenology and post C uptake parameters and also for all C4 grass parameters. The lack of posterior parameter error correlations suggests that for the majority of parameters, unique parameters values have been found by the optimizations. However, before we analyse the spread in posterior values futher, synthetic DA experiments and further DA configuration tests (see **Section 4.3**) are needed to verify if the current DA set-up is able to find the correct posterior parameter values. If the spread in posterior parameter values across PFTs seen in the DA experiments presented here is real, it may mean that traditional PFT categories do not represent dryland species well. Instead, high spread in posterior parameter values grouped by current PFT groupings may indicate that new PFTs need to be developed specifically for dryland species, or that certain parameters vary more across different species, or across biomes, latitudes and continents within each PFT (Dahlin et al., 2017; Yang et al., 2021). In addition, high spread in posterior parameter values may point to temporal variation in traits. Barron-Gafford et al. (2012) demonstrated that maximum photosynthetic capacity and the optimum temperature for photosynthesis differed across seasons for both dryland woody plants and C4 grasses. Likewise, Cable et al. (2012) showed that different PFTs can alter the sensitivity of soil respiration to temperature and moisture. Future DA experiments can assess these different predictions of how

parameters and traits vary across species and over time by performing assimilation experiments grouping different vegetation types and allowing the parameters to vary over given timescales. These studies can in turn point to both soil and vegetative processes that are most important for governing C flux dynamics. For example, Verbeeck et al. (2011) used a time-varying parameter DA experiment to suggest that deep-root water access was crucial for Amazon forests to maintain high productivity during the dry season.

Our assimilation results and temporal validation analysis suggest that processes that control seasonal vegetation dynamics (e.g., phenology and plant water availability) are crucial for capturing NEE IAV (a conclusion shared by Whitley et al., 2017). This analysis mirrors Fu et al. (2019) who showed that, unlike mesic ecosystems in which NEE IAV is dominated by maximum carbon uptake (i.e., peak fluxes), in water-limited ecosystems NEE IAV is dominated more by the carbon uptake period. In other words, changes in the growing season length (GSL) may have an outsized impact on annual net CO_2 flux variability. The importance of phenology may be specifically related to the *timing* of plant growth rather than *duration* per se (Ogle and Reynolds, 2004). The timing of vegetation growth differs considerably among dryland species given their different strategies for accessing available water (Barron-Gafford et al., 2017; Cleverly et al., 2016; Guo et al., 2018; Krishnan et al., 2012; Reynolds et al., 2004; Scott et al., 2008; Wilcox et al., 2004) in addition to changing responses to rainfall different seasons (Biederman et al., 2018) and lagged effects that are also characteristic of dryland ecosystems (Barnes et al., 2016; Liu et al., 2019; Ogle and Reynolds, 2004; Shen et al., 2016). Clearly, there is a complexity of different vegetation responses to plant available water that need to be explored further and explicitly in terms of their contributions to NEE IAV. The various hypotheses and theoretical models for plant water use and growth (e.g., the “two-layer” and “pulse-reserve hypotheses, “threshold-delay model”, and hierarchical responses to rainfall pulses – Collins et al., 2014; Ogle and Reynolds, 2004; Schwinning and Sala, 2004 – among others) could be systematically tested within a TBM framework that is designed to incorporate the myriad different interacting vegetation, C and water cycle processes that contribute to ecosystem-scale net exchange of CO_2 between the surface and atmosphere.

In addition to testing the theoretical frameworks underpinning dryland biogeochemical cycling and vegetation dynamics, more research needs to be conducted into dryland phenology. Phenological drivers are not as well understood or modeled for dryland vegetation as they are for

825 more mesic ecosystems (Dahlin et al., 2017; Eamus and Prior 2001; Singh and Kushwaha, 2005;
826 Smith et al., 2012). Empirical studies are needed to develop an understanding of which
827 environmental cues are the most important for different dryland species. Research into different
828 strategies for accessing root zone soil moisture is more advanced (Shiqin et al., 2017; Wilcox et
829 al., 2004); however, model developments and further DA experiments could be used to assess
830 how site level understanding of these processes scales to impact ecosystem CO_2 fluxes that are
831 affected more by competition between species in spatially heterogeneous landscapes.

832 The analysis presented here has focused on site-level CO_2 fluxes. Once further DA
833 system configuration tests have been performed across all flux tower sites in the SW US and
834 other dryland ecosystems worldwide (see **Section 4.3**), regional-scale DA experiments will
835 provide us with data-constrained posterior simulations that can be used to address questions such
836 as which dryland regions are most responsible for global scale NEE IAV (Haverd et al., 2017),
837 which vegetation types account for the majority of C flux variability across different ecosystem
838 types (Haverd et al., 2013), or which processes dominate the NEE IAV (Haverd et al., 2016 ;
839 Humphrey et al., 2021).

840 4.3 Caveats of the DA approach and perspectives for future dryland C cycle DA studies

841 In this study we focused on correcting parameters related to GPP, partly because
842 MacBean et al. (2021) found that GPP, and particularly summer monsoon season GPP, is the
843 dominant driver of NEE IAV. We also are obliged to focus on GPP parameters because the
844 number of model parameters is higher for GPP. In a follow up study, we are assessing how the
845 number of parameters linked to each different process affects the ability of the optimization to
846 correct for errors in those processes. We may find, for example, that the sheer number of
847 parameters related to phenology that are included here results in those parameters being the most
848 important for correcting NEE IAV. This then becomes an issue of wider model development
849 because we can only include parameters in the optimization that are in the model. Still, the fact
850 that the relatively few “post C uptake” parameters included in the assimilation tests carried out in
851 this study can account for biases in mean annual NEE suggests that the number of parameters
852 linked to each process does not prevent us from identifying which set of parameters (and

processes) are mostly causing model-data discrepancies. It is still possible that those parameters are accounting for other model structural errors, as we have discussed in **Section 4.1**.

The spin-up procedure used in these assimilations results in the model being at quasi-steady state, which could also explain the model underestimate in mean annual NEE. In future, we will implement a so-called "transient" simulation in the assimilation framework and test the impact on the NEE assimilation and resultant posterior parameter values. The transient simulation would occur after spin-up and before the assimilation and would simulate changing climate, rising CO_2, and land use history since the early 20th century. Given the k_{soilC} parameter, which acts as a scalar on the slow and passive C pools, did not change much during any of the assimilation experiments, we do not expect inclusion of a transient simulation would result in a considerable change in soil C. However, we do expect that we might see a stronger C sink at the forest sites due to the increased time since equilibrium conditions that are imposed by the end of the spin-up.

Further tests and validation of different DA configurations and optimizations at these, and other, dryland sites are needed to explore fully the potential of Bayesian DA systems for quantifying and reducing error in dryland ecosystem C fluxes. The specific DA configuration (e.g. type of data included, the number of parameters optimized and to which processes they are related, the data record length, and the model version) can lead to different posterior values and degree of improvement in model-data fit. Previous dryland DA studies have suggested that including hydrology related data streams in the assimilation can be beneficial in improving model root zone soil moisture, vegetation dynamics and C and water flux estimates due to the strong C-water interactions in these ecosystems (Barrett et al., 2005; Haverd et al., 2013; Tian et al., 2019a,b). Raoult et al. (2021) have demonstrated how *in situ* soil moisture DA can be used to improve model "drydowns" in water availability following rain events, which should help the model to capture C flux dynamics in these characteristically pulse-driven ecosystems (Huxman et al., 2004). However, we note that as we include multiple data streams in the assimilation the computational cost will increase; therefore, there will be a need for sensitivity analyses to select only the parameters to which assimilated variables are most sensitive. Improving estimates of NEE IAV may require additional terms in the cost-function that specifically estimate the model-data misfit at annual timescales (e.g., Desai, 2010), as opposed to evaluating the daily timestep as

we do in the cost function in this study. We also need to find better ways to estimate the combined observation and model structural errors (**R** matrix), including assessing observation temporal autocorrelation and error correlations between different datasets in future multiple data stream DA experiments. In future SW US C cycle DA studies we will also consider options for using the available trait information to better estimate prior parameter bounds and off-diagonal elements of the prior **B** matrix (e.g., Bloom and Williams, 2015). Few trait data are available for the wide variety of different plant species in dryland ecosystems (**Section 2.7**). Collaborative projects between modelers and empirical scientists could improve that situation, with model experiments better informing trait data collection needs) as well as additional DA experiments testing how parameters may vary over time (Barron-Gafford et al., 2012; Downton et al., 1984). More well-defined prior bounds on the parameters would reduce prior uncertainties and therefore provide a much stronger constraint on the parameter optimization. The result would be lower parameter error reductions, but increased efficiency in finding the most optimal parameter vector. In addition, the impact of posterior parameter error correlations and model equifinality on modeled dryland C flux uncertainty should be explored further (e.g., Trudinger et al., 2016). Equifinality – the situation in which multiple different posterior parameter vectors result in the same reduction in model-data misfit – can be ruled out via synthetic DA experiments. Synthetic experiments use model runs with default parameters to provide pseudo-observations with which to test the ability of the inversion to find the known "true" parameter values. Finally, we ultimately need one set of parameters for each PFT in order to run regional to global scale simulations; therefore we must test how well a multiple site assimilation that includes all sites for a given PFT performs in comparison to the single site optimizations (e.g., Kuppel et al., 2012). However, we may find that multi-site assimilations only perform well once we have developed new PFTs that can better represent the variety of dryland plant species (**Section 4.2**).

Despite the need for many more DA system tests at dryland sites, the assimilation experiments presented here already demonstrate that strong reductions in parameter uncertainty and dramatic improvements in model-data fit are possible using *in situ* dryland CO_2 fluxes. Our results clearly show that, in addition to model developments that may be needed for models to better represent dryland ecosystems, C cycle related parameters likely need optimizing by TBM groups before they can accurately model dryland CO_2 fluxes. Only by addressing these issues

will we be able to reliably use TBMs to accurately simulate regional to global-scale dryland contributions to IAV and long-term trends in the global C cycle.

Acknowledgments

Funding for AmeriFlux data resources and data collection at US-SRM, US-SRG, US-Wkg, and US-Whs was provided by the U.S. Department of Energy's Office of Science and the USDA (an equal-opportunity employer). Data collection at sites US-Vcp, US-Vcm, US-Mpj, US-Wjs, US-Seg, and US-Ses were funded by the U.S. Department of Energy EPSCoR (DE-FG02-08ER46506), and the Department of Energy Ameriflux Management Project (Subcontract 7074628), and the Sevilleta Long Term Ecological Research site (NSF-DEB LTER 1440478). The US-Fuf site was supported by grants from the North American Carbon Program/USDA CREES NRI (2004-35111-15057 and 2008-35101-19076), Science Foundation Arizona (CAA 0-203-08), the Arizona Water Institute, and the Mission Research Program, School of Forestry, Northern Arizona University (McIntire-Stennis/Arizona Bureau of Forestry). KM was funded by Indiana University Prepared for Environmental Change Grand Challenge. We would like to thank the ORCHIDEE team for development and maintenance of the ORCHIDEE code and for providing the ORCHIDEE version used in this study.

Code availability

The ORCHIDEE model is under a free software license (CeCILL; see http://www.cecill.info/index.en.html) and the source code is visible here: https://forge.ipsl.jussieu.fr/orchidee/browser/tags/ORCHIDEE_2_0 (Peylin et al., 2021). The ORCHIDEE model code is written in Fortran 90 and is maintained and developed under an SVN version control system at the Institute Pierre Simon Laplace (IPSL) in France. The ORCHIDAS code is currently in the process of being put on a GitHub repository but for now it is available on request to vladislav.bastrikov@lsce.ipsl.fr.

Data availability

Meteorological forcing data and eddy covariance measurements of net surface energy and carbon exchanges at 30-minutes intervals are available from the AmeriFlux data portal (http://ameriflux.lbl.gov). The model outputs from ORCHIDEE simulations and post-processing python scripts for manuscript figures and tables are freely available in a Git repository (https://github.com/kashifmahmud/SW_US_semiarid, last access: 20 September 2021) and on Figshare (Mahmud et al., 2021).

References

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S., & Zeng, N. (2015). Carbon cycle. The dominant role of semi-arid ecosystems in the trend and variability of the land CO_2 sink. *Science*, *348*(6237), 895–899.

Anderson-Teixeira, K. J., Delong, J. P., Fox, A. M., Brese, D. A., & Litvak, M. E. (2011). Differential responses of production and respiration to temperature and moisture drive the carbon balance across a climatic gradient in New Mexico. In *Global Change Biology* (Vol. 17, Issue 1, pp. 410–424). https://doi.org/10.1111/j.1365-2486.2010.02269.x

Bacour, C., Maignan, F., Peylin, P., MacBean, N., Bastrikov, V., Joiner, J., Köhler, P., Guanter, L., & Frankenberg, C. (2019). Differences Between OCO-2 and GOME-2 SIF Products From a Model-Data Fusion Perspective. In *Journal of Geophysical Research: Biogeosciences* (Vol. 124, Issue 10, pp. 3143–3157). https://doi.org/10.1029/2018jg004938

Barron-Gafford, G.A., Scott, R.L., Jenerette, G.D., Hamerlynck, E.P. and Huxman, T.E. (2012), Temperature and precipitation controls over leaf- and ecosystem-level CO2 flux along a woody plant encroachment gradient. Glob Change Biol, 18: 1389-1400. https://doi.org/10.1111/j.1365-2486.2011.02599.x

Barron-Gafford, G.A., Sanchez-Cañete, E.P., Minor, R.L., Hendryx, S.M., Lee, E., Sutter, L.F., Tran, N., Parra, E., Colella, T., Murphy, P.C., Hamerlynck, E.P., Kumar, P. and Scott, R.L. (2017), Impacts of hydraulic redistribution on grass–tree competition vs facilitation in a semi-arid savanna. New Phytol, 215: 1451-1461. https://doi.org/10.1111/nph.14693.

970 Barnes, M. L., Moran, M. S., Scott, R. L., Kolb, T. E., Ponce-Campos, G. E., Moore, D. J. P.,
971 Ross, M. A., Mitra, B., & Dore, S. (2016). Vegetation productivity responds to sub-annual
972 climate conditions across semiarid biomes. Ecosphere, 7(5), [e01339].
973 https://doi.org/10.1002/ecs2.1339.

974 Barrett, D. J., Hill, M. J., Hutley, L. B., Beringer, J., Xu, J. H., Cook, G. D., ... & Williams, R. J.
975 (2005). Prospects for improving savanna biophysical models by using multiple-constraints
976 model-data assimilation methods. Australian Journal of Botany, 53(7), 689-714.

977 Bastrikov, V., MacBean, N., Bacour, C., Santaren, D., Kuppel, S., & Peylin, P. (2018). Land
978 surface model parameter optimisation using in situ flux data: comparison of gradient-based
979 versus random search algorithms (a case study using ORCHIDEE v1.9.5.2). In *Geoscientific*
980 *Model Development* (Vol. 11, Issue 12, pp. 4739–4754). https://doi.org/10.5194/gmd-11-4739-
981 2018

982 Belnap, J., Weber, B., & Büdel, B. (2016). Biological soil crusts as an organizing principle in
983 drylands. In Biological soil crusts: an organizing principle in drylands (pp. 3-13). Springer,
984 Cham.

985 Biederman J.A., Scott, R. L., Arnone III J.A., Jasoni R.L., Litvak M.E., Moreo M.T., Papuga
986 S.A., Ponce-Campos G.E., Schreiner-McGraw A.P., Vivoni E.R. (2018). Shrubland carbon sink
987 depends upon winter water availability in the warm deserts of North America, Agricultural and
988 Forest Meteorology, Volume 249, Pages 407-419.

989 Biederman, J. A., Scott, R. L., Bell, T. W., Bowling, D. R., Dore, S., Garatuza-Payan, J., Kolb,
990 T. E., Krishnan, P., Krofcheck, D. J., Litvak, M. E., Maurer, G. E., Meyers, T. P., Oechel, W. C.,
991 Papuga, S. A., Ponce-Campos, G. E., Rodriguez, J. C., Smith, W. K., Vargas, R., Watts, C. J., …
992 Goulden, M. L. (2017). CO exchange and evapotranspiration across dryland ecosystems of
993 southwestern North America. *Global Change Biology*, *23*(10), 4204–4221.

994 Biederman, J. A., Scott, R. L., Goulden, M. L., Vargas, R., Litvak, M. E., Kolb, T. E., Yepez, E.
995 A., Oechel, W. C., Blanken, P. D., Bell, T. W., Garatuza-Payan, J., Maurer, G. E., Dore, S., &
996 Burns, S. P. (2016). Terrestrial carbon balance in a drier world: the effects of water availability
997 in southwestern North America. In *Global Change Biology* (Vol. 22, Issue 5, pp. 1867–1879).
998 https://doi.org/10.1111/gcb.13222

999 Bloom, A. A. and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited
1000 world: integrating ecological "common sense" in a model–data fusion framework,
1001 Biogeosciences, 12, 1299–1315, https://doi.org/10.5194/bg-12-1299-2015, 2015.

1002 Bodner, G. S., & Robles, M. D. (2017). Enduring a decade of drought: Patterns and drivers of
1003 vegetation change in a semi-arid grassland. *Journal of Arid Environments*, *136*, 1-14.

1004 Brandt, M., Hiernaux, P., Tagesson, T., Verger, A., Rasmussen, K., Diouf, A.A., Mbow, C.,
1005 Mougin, E., Fensholt, R. (2016). Woody plant cover estimation in drylands from Earth
1006 Observation based seasonal metrics. Remote Sensing of Environment, 172, 28-38.

1007 Cable, J. M., Barron-Gafford, G. A., Ogle, K., Pavao-Zuckerman, M., Scott, R. L., Williams, D.
1008 G., and Huxman, T. E. (2012), Shrub encroachment alters sensitivity of soil respiration to
1009 temperature and moisture, J. Geophys. Res., 117, G01001, doi:10.1029/2011JG001757.

1010 Cleverly, J., Eamus, D., Luo, Q., Coupe, N. R., Kljun, N., Ma, X., Ewenz, C., Li, L., Yu, Q., &
1011 Huete, A. (2016). The importance of interacting climate modes on Australia's contribution to
1012 global carbon cycle extremes. In *Scientific Reports* (Vol. 6, Issue 1).
1013 https://doi.org/10.1038/srep23113

1014 Cleverly, J., Eamus, D., Coupe, N. R., Kljun, Chen, C., Maes, W., Li, L., Faux, R., Santini, N.S.,
1015 Rumman, R., Yu, Q., & Huete, A. (2016). The importance of interacting climate modes on
1016 Australia's contribution to global carbon cycle extremes. In *Scientific Reports* (Vol. 6, Issue 1).
1017 https://doi.org/10.1038/srep23113.

1018 Collins S.L., Belnap J., Grimm N.B., Rudgers J.A., Dahm C.N., D'Odorico P., Litvak M., Natvig
1019 D.O., Peters D.C., Pockman W.T., Sinsabaugh R.L., Wolf B.O. (2014). A Multiscale,
1020 Hierarchical Model of Pulse Dynamics in Arid-Land Ecosystems. Annual Review of Ecology,
1021 Evolution, and Systematics, 45:1, 397-419.

1022 Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., & Luke,
1023 C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide
1024 variability. *Nature*, *494*(7437), 341–344.

1025 Dahlin, K. M., Fisher, R. A., & Lawrence, P. J. (2015). Environmental drivers of drought
1026 deciduous phenology in the Community Land Model. *Biogeosciences*, *12*(16), 5061-5074.

1027 Dahlin, K.M., Ponte, D.D., Setlock, E. and Nagelkirk, R. (2017), Global patterns of drought
1028 deciduous phenology in semi-arid and savanna-type ecosystems. Ecography, 40: 314-323.
1029 https://doi.org/10.1111/ecog.02443.

1030 De Kauwe, M. G., Zhou, S.-X., Medlyn, B. E., Pitman, A. J., Wang, Y.-P., Duursma, R. A., and
1031 Prentice, I. C.: Do land surface models need to include differential plant species responses to
1032 drought? Examining model predictions across a mesic-xeric gradient in Europe, Biogeosciences,
1033 12, 7503–7518, https://doi.org/10.5194/bg-12-7503-2015, 2015.

1034 Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., Falge, E.,
1035 Noormets, A., Papale, D., Reichstein, M., & Stauch, V. J. (2008). Cross-site evaluation of eddy
1036 covariance GPP and RE decomposition techniques. In *Agricultural and Forest Meteorology*
1037 (Vol. 148, Issues 6-7, pp. 821–838). https://doi.org/10.1016/j.agrformet.2007.11.012

1038 Desai, A. R. (2010), Climatic and phenological controls on coherent regional interannual
1039 variability of carbon dioxide flux in a heterogeneous landscape, J. Geophys. Res., 115, G00J02,
1040 doi:10.1029/2010JG001423.

1041 Donohue RJ, Roderick ML, McVicar TR, Farquhar GD. 2013. Impact of CO_2 fertilization on
1042 maximum foliage cover across the globe's warm, arid environments. Geophysical Research
1043 Letters 40: 3031–3035.

1044 Dore, S., Montes-Helu, M., Hart, S. C., Hungate, B. A., Koch, G. W., Moon, J. B., Finkral, A. J.,
1045 & Kolb, T. E. (2012). Recovery of ponderosa pine ecosystem carbon and water fluxes from
1046 thinning and stand-replacing fire. *Global Change Biology*, *18*(10), 3171–3185.

1047 Downton, W.J.S., Berry, J.A., Seemann, J.R. (1984). Tolerance of Photosynthesis to High
1048 Temperature in Desert Plants. Plant Physiology, 74(4), 786–790,
1049 https://doi.org/10.1104/pp.74.4.786.

1050 Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y.,
1051 Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule,
1052 P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., … Vuichard, N. (2013). Climate change
1053 projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. In *Climate*
1054 *Dynamics* (Vol. 40, Issues 9-10, pp. 2123–2165). https://doi.org/10.1007/s00382-012-1636-1

1055 Eamus, D. and Prior, L. 2001. Ecophysiology of trees of seasonally dry tropics: comparisons
1056 among phenologies. – In: Caswell, H. (ed.), Advances in ecological research. Academic Press,
1057 pp. 113–197.

1058 Exbrayat, J.-F., Bloom, A. A., Falloon, P., Ito, A., Smallman, T. L., and Williams, M.:
1059 Reliability ensemble averaging of 21st century projections of terrestrial net primary productivity
1060 reduces global and regional uncertainties, Earth Syst. Dynam., 9, 153–165,
1061 https://doi.org/10.5194/esd-9-153-2018, 2018.

1062 Famiglietti, C. A., Smallman, T. L., Levine, P. A., Flack-Prain, S., Quetin, G. R., Meyer, V.,
1063 Parazoo, N. C., Stettz, S. G., Yang, Y., Bonal, D., Bloom, A. A., Williams, M., and Konings, A.
1064 G.: Optimal model complexity for terrestrial carbon cycle prediction, Biogeosciences, 18, 2727–
1065 2754, https://doi.org/10.5194/bg-18-2727-2021, 2021.

1066 Forkel, M., Carvalhais, N., Schaphoff, S., v. Bloh, W., Migliavacca, M., Thurner, M., and
1067 Thonicke, K.: Identifying environmental controls on vegetation greenness phenology through
1068 model–data integration, Biogeosciences, 11, 7025–7050, https://doi.org/10.5194/bg-11-7025-
1069 2014, 2014.

1070 Forkel, M., Drüke, M., Thurner, M. *et al.* Constraining modelled global vegetation dynamics and
1071 carbon turnover using multiple satellite observations. *Sci Rep* **9,** 18757 (2019).
1072 https://doi.org/10.1038/s41598-019-55187-7.

1073 Fu, C., Wang, G., Goulden, M. L., Scott, R. L., Bible, K., and G. Cardon, Z.: Combined
1074 measurement and modeling of the hydrological impact of hydraulic redistribution using CLM4.5
1075 at eight AmeriFlux sites, Hydrol. Earth Syst. Sci., 20, 2001–2018, https://doi.org/10.5194/hess-
1076 20-2001-2016, 2016.

1077 Fu, Z., Dong, J., Zhou, Y., Stoy, P. C., & Niu, S. (2017). Long term trend and interannual
1078 variability of land carbon uptake—the attribution and processes. In *Environmental Research*
1079 *Letters* (Vol. 12, Issue 1, p. 014018). https://doi.org/10.1088/1748-9326/aa5685.

1080 Fu, Z, Stoy, PC, Poulter, B, et al. Maximum carbon uptake rate dominates the interannual
1081 variability of global net ecosystem exchange. Glob Change Biol. 2019; 25: 3381– 3394.
1082 https://doi.org/10.1111/gcb.14731.

1083 Gauch, H. G., Gene Hwang, J. T., & Fick, G. W. (2003). Model Evaluation by Comparison of
1084 Model-Based Predictions and Measured Values. In *Agronomy Journal* (Vol. 95, Issue 6, pp.
1085 1442–1446). https://doi.org/10.2134/agronj2003.1442

1086 Gibbens, R. P., & Lenz, J. M. (2001). Root systems of some Chihuahuan Desert plants. Journal
1087 of Arid Environments, 49(2), 221-263.

1088 Goldberg, D. E., David Edward, G., Goldberg, D. E. G., & Visiting Assistant Professor of
1089 History David E Goldberg. (1989). *Genetic Algorithms in Search, Optimization, and Machine*
1090 *Learning*. Addison-Wesley Publishing Company.

1091 Guo, J.S., B.A. Hungate, T.E. Kolb, G.W. Koch. 2018. Water source niche overlap increases
1092 with site moisture availability in woody perennials. Plant Ecology:
1093 https://doi.org/10.1007/s11258-018-0829-z.

1094 Hartley, A. J., MacBean, N., Georgievski, G., & Bontemps, S. (2017). Uncertainty in plant
1095 functional type distributions and its impact on land surface models. *Remote Sensing of*
1096 *Environment*, *203*, 71-89.

1097 Haupt, R. L., Haupt, S. E., & Haupt, S. E. A. (2004). *Practical Genetic Algorithms*. Wiley.

1098 Haverd, V., Ahlström, A., Smith, B., & Canadell, J. G. (2017). Carbon cycle responses of semi-
1099 arid ecosystems to positive asymmetry in rainfall. *Global Change Biology*, *23*(2), 793–800.

1100 Haverd, V., Raupach, M. R., Briggs, P. R., Canadell, J. G., Isaac, P., Pickett-Heaps, C.,
1101 Roxburgh, S. H., van Gorsel, E., Viscarra Rossel, R. A., & Wang, Z. (2013). Multiple
1102 observation types reduce uncertainty in Australia's terrestrial carbon and water cycles. In
1103 *Biogeosciences* (Vol. 10, Issue 3, pp. 2011–2040). https://doi.org/10.5194/bg-10-2011-2013

1104 Haverd, V., Smith, B., and Trudinger, C. (2016), Process contributions of Australian ecosystems
1105 to interannual variations in the carbon cycle. Environ. Res. Lett. 11 054013.

1106 Haverd, V., Ahlström, A., Smith, B. and Canadell, J.G. (2017), Carbon cycle responses of semi-
1107 arid ecosystems to positive asymmetry in rainfall. Glob Change Biol, 23: 793-800.
1108 https://doi.org/10.1111/gcb.13412.

1109 Hogue, T. S., Bastidas, L., Gupta, H., Sorooshian, S., Mitchell, K., & Emmerich, W. (2005).
1110 Evaluation and Transferability of the Noah Land Surface Model in Semiarid Environments. In
1111 *Journal of Hydrometeorology* (Vol. 6, Issue 1, pp. 68–84). https://doi.org/10.1175/jhm-402.1

1112 Hooper, D.U., Johnson, L. Nitrogen limitation in dryland ecosystems: Responses to geographical
1113 and temporal variation in precipitation. *Biogeochemistry* **46,** 247–293 (1999).
1114 https://doi.org/10.1007/BF01007582.

1115 Humphrey, V., Berg, A., Ciais, P. et al. Soil moisture–atmosphere feedback dominates land
1116 carbon uptake variability. Nature 592, 65–69 (2021). https://doi.org/10.1038/s41586-021-03325-
1117 5.

1118 Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C.,
1119 Fisk, J., Fujimori, S., Klein Goldewijk, K., Hasegawa, T., Havlik, P., Heinimann, A.,
1120 Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., … Zhang,
1121 X. (2020). Harmonization of global land use change and management for the period 850–2100
1122 (LUH2) for CMIP6. *Geoscientific Model Development*, *13*(11), 5425–5464.

1123 Huxman, T.E., Snyder, K.A., Tissue, D. et al. Precipitation pulses and carbon fluxes in semiarid
1124 and arid ecosystems. Oecologia 141, 254–268 (2004). https://doi.org/10.1007/s00442-004-1682-
1125 4.

1126 Jolly, W.M., Nemani, R. and Running, S.W. (2005), A generalized, bioclimatic index to predict
1127 foliar phenology in response to climate. Global Change Biology, 11: 619-
1128 632. https://doi.org/10.1111/j.1365-2486.2005.00930.x

1129 Kattge, J, Boenisch, G, Diaz, S, et al. TRY plant trait database - enhanced coverage and open
1130 access. Glob Change Biol. 2020; 26: 119-188. https://doi.org/10.1111/gcb.14904

1131 Kerhoulas L.P., Kolb, T.E. & Koch, G.W. (2012) Tree size, stand density, and the source of
1132 water used across seasons by ponderosa pine in northern Arizona. Forest Ecology and
1133 Management. 289: 425-433.

1134 Kobayashi, K., & Salam, M. U. (2000). Comparing Simulated and Measured Values Using Mean
1135 Squared Deviation and its Components. In *Agronomy Journal* (Vol. 92, Issue 2, p. 345).
1136 https://doi.org/10.1007/s100870050043

1137 Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais,
1138 P., Sitch, S., & Colin Prentice, I. (2005). A dynamic global vegetation model for studies of the
1139 coupled atmosphere-biosphere system. In *Global Biogeochemical Cycles* (Vol. 19, Issue 1).
1140 https://doi.org/10.1029/2003gb002199

1141 Krishnan, P., Meyers, T. P., Scott, R. L., Kennedy, L., & Heuer, M. (2012). Energy exchange
1142 and evapotranspiration over two temperate semi-arid grasslands in North America. In
1143 *Agricultural and Forest Meteorology* (Vol. 153, pp. 31–44).
1144 https://doi.org/10.1016/j.agrformet.2011.09.017

1145 Krishnan, P., Meyers, T.P., Scott, R.L., Kennedy, L., Heuer, M. (2012). Energy exchange and
1146 evapotranspiration over two temperate semi-arid grasslands in North America. Agricultural and
1147 Forest Meteorology, 153: 31-44.

1148 Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., & Richardson, A. D. (2012).
1149 Constraining a global ecosystem model with multi-site eddy-covariance data. In *Biogeosciences*
1150 (Vol. 9, Issue 10, pp. 3757–3776). https://doi.org/10.5194/bg-9-3757-2012

1151 Kuppel, S., Peylin, P., Maignan, F., Chevallier, F., Kiely, G., Montagnani, L., & Cescatti, A.
1152 (2014). Model–data fusion across ecosystems: from multisite optimizations to global
1153 simulations. In *Geoscientific Model Development* (Vol. 7, Issue 6, pp. 2581–2597).
1154 https://doi.org/10.5194/gmd-7-2581-2014

1155 Lasslop, G., Brovkin, V., Reick, C. H., Bathiany, S., & Kloster, S. (2016). Multiple stable states
1156 of tree cover in a global land surface model due to a fire-vegetation feedback. Geophysical
1157 Research Letters, 43(12), 6324-6331.

1158 Lawal, S., Lennard, C., Jack, C., Wolski, P., Hewitson, B., and Abiodun, B. (2019) The observed
1159 and model-simulated response of southern African vegetation to drought, Agricultural and Forest
1160 Meteorology, 279, 107698, https://doi.org/10.1016/j.agrformet.2019.107698.

1161 Lee, E., Kumar, P., Barron-Gafford, G. A., Hendryx, S. M., Sanchez-Canete, E. P., Minor, R. L.,
1162 et al. (2018). Impact of hydraulic redistribution on multispecies vegetation water use in a
1163 semiarid savanna ecosystem: An experimental and modeling synthesis. Water Resources
1164 Research, 54, 4009– 4027. https://doi.org/10.1029/ 2017WR021006.

1165 Li, L., Wang, Y.-P., Yu, Q., Pak, B., Eamus, D., Yan, J., van Gorsel, E., and Baker, I. T. (2012),
1166 Improving the responses of the Australian community land surface model (CABLE) to seasonal
1167 drought, J. Geophys. Res., 117, G04002, doi:10.1029/2012JG002038.

1168 Liu, Y., Schwalm, C.R., Samuels-Crow, K.E. and Ogle, K. (2019), Ecological memory of daily
1169 carbon exchange across the globe and its importance in drylands. Ecol Lett, 22: 1806-1816.
1170 https://doi.org/10.1111/ele.13363.

1171 MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., Köhler, P., Gómez-
1172 Dans, J., & Disney, M. (2018). Strong constraint on modelled global carbon uptake using solar-
1173 induced chlorophyll fluorescence data. *Scientific Reports*, *8*(1), 1973.

1174 MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., & Ciais, P. (2015). Using
1175 satellite data to improve the leaf phenology of a global terrestrial biosphere model. In
1176 *Biogeosciences* (Vol. 12, Issue 23, pp. 7185–7208). https://doi.org/10.5194/bg-12-7185-2015

1177 MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent
1178 assimilation of multiple data streams in a carbon cycle data assimilation system. In *Geoscientific*
1179 *Model Development* (Vol. 9, Issue 10, pp. 3569–3588). https://doi.org/10.5194/gmd-9-3569-
1180 2016

1181 MacBean, N., R. L. Scott, J. A. Biederman, P. Peylin, T. Kolb, M. Litvak, P. Krishnan, T.
1182 Meyers, V. Arora, V. Bastrikov, D. Goll, D. L. Lombardozzi, J. Nabel, J. Pongratz, S. Sitch, A.
1183 P. Walker, S. Zaehle, and D. J. P. Moore. (2021). Dynamic global vegetation models
1184 underestimate net CO2 flux mean and inter-annual variability in dryland ecosystems. *Environ.*
1185 *Res. Lett. 16 094023.*

1186 MacBean, N., Scott, R. L., Biederman, J. A., Ottlé, C., Vuichard, N., Ducharne, A., Kolb, T.,
1187 Dore, S., Litvak, M., & Moore, D. J. P. (2020). Testing water fluxes and storage from two
1188 hydrology configurations within the ORCHIDEE land surface model across US semi-arid sites.
1189 In *Hydrology and Earth System Sciences* (Vol. 24, Issue 11, pp. 5203–5230).
1190 https://doi.org/10.5194/hess-24-5203-2020

1191 Mahmud K., Biederman J., Scott R. L., Litvak M., Kolb T., Meyers T. P., Krishnan P., Bastrikov
1192 V., MacBean N. "Optimizing Carbon Cycle Parameters Drastically Improves Terrestrial
1193 Biosphere Model Underestimates of Dryland Mean Net CO2 Flux and its Inter-Annual

Variability" by Mahmud et al. (2021), figshare, Dataset, https://doi.org/10.6084/m9.figshare.14327489.v4.

Ogle, K., Reynolds, J.F. Plant responses to precipitation in desert ecosystems: integrating functional types, pulses, thresholds, and delays. Oecologia 141, 282–294 (2004). https://doi.org/10.1007/s00442-004-1507-5.

Peng, S., Ciais, P., Chevallier, F., Peylin, P., Cadule, P., Sitch, S., Piao, S., Ahlström, A., Huntingford, C., Levy, P., Li, X., Liu, Y., Lomas, M., Poulter, B., Viovy, N., Wang, T., Wang, X., Zaehle, S., Zeng, N., Zhao, H. (2015). Benchmarking the seasonal cycle of CO2fluxes simulated by terrestrial ecosystem models. In *Global Biogeochemical Cycles* (Vol. 29, Issue 1, pp. 46–64). https://doi.org/10.1002/2014gb004931

Petrie, M. D., Collins, S. L., Swann, A. M., Ford, P. L., & Litvak, M. E. (2015). Grassland to shrubland state transitions enhance carbon sequestration in the northern Chihuahuan Desert. *Global Change Biology*, *21*(3), 1226–1235.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., Chevallier, F., Ciais, P., & Prunet, P. (2016). A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. In *Geoscientific Model Development* (Vol. 9, Issue 9, pp. 3321–3346). https://doi.org/10.5194/gmd-9-3321-2016

Peylin, P., Ghattas, J., Cadule, P., Cheruy, F., Ducharne, A., Guenet, B., Lathière, J., Luyssaert, S., Maignan, F., Maugis, P., Ottle, C., Polcher, J., Viovy, N., Vuichard, N., Bastrikov, V., Guimberteau, M., Lanso, A.-S., MacBean, N., Mcgrath, M., Tafasca, S., and Wang, F. (2021) The global land surface model ORCHIDEE – Tag2.0, available at: http://forge.ipsl.jussieu.fr/orchidee/browser/tags/ORCHIDEE_2_0.

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J. G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P. E., Li, J., Lin, X., Lomas, M. R., Lu, M., Luo, Y., … Zeng, N. (2013). Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends. *Global Change Biology*, *19*(7), 2117–2132.

1222 Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J.G., Ciais, P, Friedlingstein, P., and
1223 Sitch S. (2019) Interannual variation of terrestrial carbon cycle: Issues and perspectives. Glob
1224 Change Biol.; 26: 300– 318. https://doi.org/10.1111/gcb.14884.

1225 Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G.,
1226 Chevallier, F., Liu, Y. Y., Running, S. W., Sitch, S., & van der Werf, G. R. (2014). Contribution
1227 of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature*, *509*(7502),
1228 600–603.

1229 Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Bontemps, S.,
1230 Boettcher, M., Brockmann, C., Defourny, P., Hagemann, S., Herold, M., Kirches, G., Lamarche,
1231 C., Lederer, D., Ottlé, C., Peters, M., & Peylin, P. (2015). Plant functional type classification for
1232 earth system models: results from the European Space Agency's Land Cover Climate Change
1233 Initiative. *Geoscientific Model Development*, *8*(7), 2315–2328.

1234 Raczka, B. M., Davis, K. J., Huntzinger, D., Neilson, R. P., Poulter, B., Richardson, A. D., Xiao,
1235 J., Baker, I., Ciais, P., Keenan, T. F., Law, B., Post, W. M., Ricciuto, D., Schaefer, K., Tian, H.,
1236 Tomelleri, E., Verbeeck, H., & Viovy, N. (2013). Evaluation of continental carbon cycle
1237 simulations with North American flux tower observations. In *Ecological Monographs* (Vol. 83,
1238 Issue 4, pp. 531–556). https://doi.org/10.1890/12-0893.1

1239 Raoult, N. M., Jupp, T. E., Cox, P. M., and Luke, C. M.: Land-surface parameter optimization
1240 using data assimilation techniques: the adJULES system V1.0, Geosci. Model Dev., 9, 2833–
1241 2852, https://doi.org/10.5194/gmd-9-2833-2016, 2016.

1242 Raoult, N., Ottlé, C., Peylin, P., Bastrikov, V., & Maugis, P. (2021). Evaluating and Optimizing
1243 Surface Soil Moisture Drydowns in the ORCHIDEE Land Surface Model at In Situ Locations,
1244 Journal of Hydrometeorology, 22(4), 1025-1043.

1245 Renwick, K. M., Fellows, A., Flerchinger, G. N., Lohse, K. A., Clark, P. E., Smith, W. K.,
1246 Emmett, K., & Poulter, B. (2019). Modeling phenological controls on carbon dynamics in
1247 dryland sagebrush ecosystems. In *Agricultural and Forest Meteorology* (Vol. 274, pp. 85–94).
1248 https://doi.org/10.1016/j.agrformet.2019.04.003

1249 Reynolds, J.F., Kemp, P.R., Ogle, K. et al. Modifying the 'pulse–reserve' paradigm for deserts of
1250 North America: precipitation pulses, soil water, and plant responses. Oecologia 141, 194–210
1251 (2004). https://doi.org/10.1007/s00442-004-1524-4.

1252 Santaren, D., Peylin, P., Bacour, C., Ciais, P., & Longdoz, B. (2014). Ecosystem model
1253 optimization using in situ flux observations: benefit of Monte Carlo versus variational schemes
1254 and analyses of the year-to-year model performances. In *Biogeosciences* (Vol. 11, Issue 24, pp.
1255 7137–7158). https://doi.org/10.5194/bg-11-7137-2014

1256 Santaren, D., Peylin, P., Viovy, N., & Ciais, P. (2007). Optimizing a process-based ecosystem
1257 model with eddy-covariance flux measurements: A pine forest in southern France. In *Global*
1258 *Biogeochemical Cycles* (Vol. 21, Issue 2). https://doi.org/10.1029/2006gb002834

1259 Schaefer, K., Schwalm, C. R., Williams, C., Altaf Arain, M., Barr, A., Chen, J. M., Davis, K. J.,
1260 Dimitrov, D., Hilton, T. W., Hollinger, D. Y., Humphreys, E., Poulter, B., Raczka, B. M.,
1261 Richardson, A. D., Sahoo, A., Thornton, P., Vargas, R., Verbeeck, H., Anderson, R., … Zhou, X.
1262 (2012). A model-data comparison of gross primary productivity: Results from the North
1263 American Carbon Program site synthesis. In *Journal of Geophysical Research: Biogeosciences*
1264 (Vol. 117, Issue G3). https://doi.org/10.1029/2012jg001960

1265 Schwinning, S., Sala, O.E. Hierarchy of responses to resource pulses in arid and semi-arid
1266 ecosystems. Oecologia 141, 211–220 (2004). https://doi.org/10.1007/s00442-004-1520-8.

1267 Scott, R. L. (2010). Using watershed water balance to evaluate the accuracy of eddy covariance
1268 evaporation measurements for three semiarid ecosystems. In *Agricultural and Forest*
1269 *Meteorology* (Vol. 150, Issue 2, pp. 219–225). https://doi.org/10.1016/j.agrformet.2009.11.002

1270 Scott, R. L., Biederman, J. A., Hamerlynck, E. P., & Barron-Gafford, G. A. (2015). The carbon
1271 balance pivot point of southwestern U.S. semiarid ecosystems: Insights from the 21st century
1272 drought. In *Journal of Geophysical Research: Biogeosciences* (Vol. 120, Issue 12, pp. 2612–
1273 2624). https://doi.org/10.1002/2015jg003181

1274 Scott, R. L., Cable, W. L., and Hultine, K. R. (2008), The ecohydrologic significance of
1275 hydraulic redistribution in a semiarid savanna, Water Resour. Res., 44, W02440,
1276 doi:10.1029/2007WR006149.

1277 Scott, R. L., Darrel Jenerette, G., Potts, D. L., & Huxman, T. E. (2009). Effects of seasonal
1278 drought on net carbon dioxide exchange from a woody-plant-encroached semiarid grassland. In
1279 *Journal of Geophysical Research* (Vol. 114, Issue G4). https://doi.org/10.1029/2008jg000900

1280 Shen, W., Jenerette, G. D., Hui, D., and Scott, R. L.: Precipitation legacy effects on dryland
1281 ecosystem carbon fluxes: direction, magnitude and biogeochemical carryovers, Biogeosciences,
1282 13, 425–439, https://doi.org/10.5194/bg-13-425-2016, 2016.

1283 Singh, K. P., & Kushwaha, C. P. (2005). Emerging paradigms of tree phenology in dry tropics.
1284 Current Science, 964-975.

1285 Shiqin Xu, Xibin Ji, Bowen Jin, Jinglin Zhang, Root distribution of three dominant desert shrubs
1286 and their water uptake dynamics, Journal of Plant Ecology, Volume 10, Issue 5, October 2017,
1287 Pages 780–790, https://doi.org/10.1093/jpe/rtw079.

1288 Smith, S. D., Monson, R., & Anderson, J. E. (2012). Physiological Ecology of North American
1289 Desert Plants. Springer Science & Business Media.

1290 Sun, Y., Goll, D. S., Chang, J., Ciais, P., Guenet, B., Helfenstein, J., Huang, Y., Lauerwald, R.,
1291 Maignan, F., Naipal, V., Wang, Y., Yang, H., and Zhang, H.: Global evaluation of the nutrient-
1292 enabled version of the land surface model ORCHIDEE-CNP v1.2 (r5986), Geosci. Model Dev.,
1293 14, 1987–2010, https://doi.org/10.5194/gmd-14-1987-2021, 2021.

1294 Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*.
1295 https://doi.org/10.1137/1.9780898717921

1296 Teckentrup, L., De Kauwe, M. G., Pitman, A. J., Goll, D., Haverd, V., Jain, A. K., Joetzjer, E.,
1297 Kato, E., Lienert, S., Lombardozzi, D., McGuire, P. C., Melton, J. R., Nabel, J. E. M. S.,
1298 Pongratz, J., Sitch, S., Walker, A. P., and Zaehle, S.: Assessing the representation of the
1299 Australian carbon cycle in global vegetation models, Biogeosciences Discuss. [preprint],
1300 https://doi.org/10.5194/bg-2021-66, in review.

1301 Tian, S., Renzullo, L. J., van Dijk, A. I. J. M., Tregoning, P., and Walker, J. P. (2019a): Global
1302 joint assimilation of GRACE and SMOS for improved estimation of root-zone soil moisture and
1303 vegetation response, Hydrol. Earth Syst. Sci., 23, 1067–1081, https://doi.org/10.5194/hess-23-
1304 1067-2019.

1305 Tian, S., Van Dijk, A.I.J.M., and Tregoning, P. (2019b). Forecasting dryland vegetation
1306 condition months in advance through satellite data assimilation. Nat Commun 10, 469.
1307 https://doi.org/10.1038/s41467-019-08403-x.

1308 Traore, A., Ciais, P., Vuichard, N., MacBean, N., Dardel, C., Poulter, B., Piao, S., Fisher, J.,
1309 Viovy, N., Jung, M., & Myneni, R. (2014). 1982–2010 Trends of Light Use Efficiency and
1310 Inherent Water Use Efficiency in African vegetation: Sensitivity to Climate and Atmospheric
1311 CO2 Concentrations. In *Remote Sensing* (Vol. 6, Issue 9, pp. 8923–8944).
1312 https://doi.org/10.3390/rs6098923

1313 Trudinger, C. M., Haverd, V., Briggs, P. R., & Canadell, J. G. (2016). Interannual variability in
1314 Australia's terrestrial carbon cycle constrained by multiple observation types. In *Biogeosciences*
1315 (Vol. 13, Issue 23, pp. 6363–6383). https://doi.org/10.5194/bg-13-6363-2016

1316 Unland, H. E., Houser, P. R., Shuttleworth, W. J., & Yang, Z.-L. (1996). Surface flux
1317 measurement and modeling at a semi-arid Sonoran Desert site. In *Agricultural and Forest*
1318 *Meteorology* (Vol. 82, Issues 1-4, pp. 119–153). https://doi.org/10.1016/0168-1923(96)02330-1

1319 Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., and Ciais, P. (2011), Seasonal
1320 patterns of CO2 fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE
1321 model, J. Geophys. Res., 116, G02018, doi:10.1029/2010JG001544.

1322 Vuichard, N., Messina, P., Luyssaert, S., Guenet, B., Zaehle, S., Ghattas, J., Bastrikov, V., &
1323 Peylin, P. (2019). Accounting for carbon and nitrogen interactions in the global terrestrial
1324 ecosystem model ORCHIDEE (trunk version, rev 4999): multi-scale evaluation of gross primary
1325 production. In *Geoscientific Model Development* (Vol. 12, Issue 11, pp. 4751–4779).
1326 https://doi.org/10.5194/gmd-12-4751-2019

1327 Vuichard, N., & Papale, D. (2015). Filling the gaps in meteorological continuous data measured
1328 at FLUXNET sites with ERA-Interim reanalysis. In *Earth System Science Data* (Vol. 7, Issue 2,
1329 pp. 157–171). https://doi.org/10.5194/essd-7-157-2015

1330 Wang, F., Cheruy, F., & Dufresne, J.-L. (2016). The improvement of soil thermodynamics and
1331 its effects on land surface meteorology in the IPSL climate model. *Geoscientific Model*
1332 *Development*, *9*(1), 363–381.

1333 Wang, T., Ottlé, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S., & Peng, S.
1334 (2013). Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE
1335 land surface model: ORCHIDEE SNOW MODEL EVALUATION. *Journal of Geophysical*
1336 *Research*, *118*(12), 6064–6079.

1337 Whitley, R., Beringer, J., Hutley, L. B., Abramowitz, G., De Kauwe, M. G., Duursma, R., Evans,
1338 B., Haverd, V., Li, L., Ryu, Y., Smith, B., Wang, Y.-P., Williams, M., & Yu, Q. (2016). A model
1339 inter-comparison study to examine limiting factors in modelling Australian tropical savannas. In
1340 *Biogeosciences* (Vol. 13, Issue 11, pp. 3245–3265). https://doi.org/10.5194/bg-13-3245-2016

1341 Whitley, R., Beringer, J., Hutley, L. B., Abramowitz, G., De Kauwe, M. G., Evans, B., Haverd,
1342 V., Li, L., Moore, C., Ryu, Y., Scheiter, S., Schymanski, S. J., Smith, B., Wang, Y.-P., Williams,
1343 M., and Yu, Q.: Challenges and opportunities in land surface modelling of savanna ecosystems,
1344 Biogeosciences, 14, 4711–4732, https://doi.org/10.5194/bg-14-4711-2017, 2017.

1345 Wilcox, C.S., Ferguson, J.W., Fernandez, G.C.J., Nowak R.S. (2004). Fine root growth dynamics
1346 of four Mojave Desert shrubs as related to soil moisture and microsite. Journal of Arid
1347 Environments. 56(1), 129-148.

1348 Wu, D., Ciais, P., Viovy, N., Knapp, A. K., Wilcox, K., Bahn, M., Smith, M. D., Vicca, S.,
1349 Fatichi, S., Zscheischler, J., He, Y., Li, X., Ito, A., Arneth, A., Harper, A., Ukkola, A., Paschalis,
1350 A., Poulter, B., Peng, C., Ricciuto, D., Reinthaler, D., Chen, G., Tian, H., Genet, H., Mao, J.,
1351 Ingrisch, J., Nabel, J. E. S. M., Pongratz, J., Boysen, L. R., Kautz, M., Schmitt, M., Meir, P.,
1352 Zhu, Q., Hasibeder, R., Sippel, S., Dangal, S. R. S., Sitch, S., Shi, X., Wang, Y., Luo, Y., Liu,
1353 Y., and Piao, S.: Asymmetric responses of primary productivity to altered precipitation simulated
1354 by ecosystem models across three long-term grassland sites, Biogeosciences, 15, 3421–3437,
1355 https://doi.org/10.5194/bg-15-3421-2018, 2018.

1356 Wutzler, T., and Carvalhais, N. (2014), Balancing multiple constraints in model-data integration:
1357 Weights and the parameter block approach, J. Geophys. Res. Biogeosci., 119, 2112– 2129,
1358 doi:10.1002/2014JG002650.

1359 Yang, H., Ciais, P., Wang, Y., Huang, Y., Wigneron, J.P., Bastos, A., Chave, J., Chang, J.,
1360 Doughty, C.E., Fan, L., Goll, D., Joetzjer, E., Li, W., Lucas, R., Quegan, S., Toan, T.L. and Yu,

K. (2021). Variations of carbon allocation and turnover time across tropical forests. Global Ecol Biogeogr; 30: 1271– 1285. https://doi.org/10.1111/geb.13302.

Yin, X., and Struik, P. C.: C3 and C4 photosynthesis models: An overview from the perspective of crop modelling, NJAS - Wageningen Journal of Life Sciences, 57, 27-38, https://doi.org/10.1016/j.njas.2009.07.001, 2009.