ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Dynamic global vegetation models underestimate net CO_2 flux mean and inter-annual variability in dryland ecosystems

IOP Publishing Environ. Res. Lett. 16 (2021) 094023 https://doi.org/10.1088/1748-9326/ac1a38

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
2 March 2021

REVISED
14 July 2021

ACCEPTED FOR PUBLICATION
3 August 2021

PUBLISHED
24 August 2021

LETTER

Dynamic global vegetation models underestimate net CO_2 flux mean and inter-annual variability in dryland ecosystems

Natasha MacBean[1,*], Russell L Scott[2], Joel A Biederman[2], Philippe Peylin[3], Thomas Kolb[4], Marcy E Litvak[5], Praveena Krishnan[6,7], Tilden P Meyers[6], Vivek K Arora[8], Vladislav Bastrikov[9], Daniel Goll[10], Danica L Lombardozzi[11], Julia E M S Nabel[12], Julia Pongratz[12,13], Stephen Sitch[14], Anthony P Walker[15], Sönke Zaehle[16] and David J P Moore[17]

[1] Department of Geography, Indiana University, Bloomington, IN 47405, United States of America
[2] Southwest Watershed Research Center, United States Department of Agriculture, Agricultural Research Service, Tucson, AZ 85719, United States of America
[3] Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette F-91191, France
[4] School of Forestry, Northern Arizona University, Flagstaff, AZ 86011, United States of America
[5] Department of Biology, University of New Mexico, Albuquerque, NM 87131, United States of America
[6] NOAA/ARL Atmospheric Turbulence and Diffusion Division, Oak Ridge, TN 37830, United States of America
[7] Oak Ridge Associated Universities, Oak Ridge, TN 37830, United States of America
[8] Canadian Centre for Climate Modelling and Analysis, Climate Research Division, Environment and Climate Change Canada, Victoria, BC, Canada
[9] Science Partners, Paris 75010, France
[10] Université Paris-Saclay, 91190 Saint-Aubin, France
[11] Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO 80302, United States of America
[12] Max Planck Institute for Meteorology, Bundesstraße 53, Hamburg 20146, Germany
[13] Ludwig-Maximilians-Universität Munich, Luisenstr. 37, 80333 Munich, Germany
[14] College of Life and Environmental Sciences, Exeter EX4 4QE, United Kingdom
[15] Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
[16] Max Planck Institute for Biogeochemistry, PO Box 600164, Hans-Knöll-Str. 10, 07745 Jena, Germany
[17] School of Natural Resources and the Environment, University of Arizona, Tucson, AZ 85721, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: nlmacbean@gmail.com

Keywords: Drylands, global carbon cycle, inter-annual variability, dynamic global vegetation models

Supplementary material for this article is available online

Abstract

Despite their sparse vegetation, dryland regions exert a huge influence over global biogeochemical cycles because they cover more than 40% of the world surface (Schimel 2010 *Science* **327** 418–9). It is thought that drylands dominate the inter-annual variability (IAV) and long-term trend in the global carbon (C) cycle (Poulter *et al* 2014 *Nature* **509** 600–3, Ahlstrom *et al* 2015 *Science* **348** 895–9, Zhang *et al* 2018 *Glob. Change Biol.* **24** 3954–68). Projections of the global land C sink therefore rely on accurate representation of dryland C cycle processes; however, the dynamic global vegetation models (DGVMs) used in future projections have rarely been evaluated against dryland C flux data. Here, we carried out an evaluation of 14 DGVMs (TRENDY v7) against net ecosystem exchange (NEE) data from 12 dryland flux sites in the southwestern US encompassing a range of ecosystem types (forests, shrub- and grasslands). We find that all the models underestimate both mean annual C uptake/release as well as the magnitude of NEE IAV, suggesting that improvements in representing dryland regions may improve global C cycle projections. Across all models, the sensitivity and timing of ecosystem C uptake to plant available moisture was at fault. Spring biases in gross primary production (GPP) dominate the underestimate of mean annual NEE, whereas models' lack of GPP response to water availability in both spring and summer monsoon are responsible for inability to capture NEE IAV. Errors in GPP moisture sensitivity at high elevation forested sites were more prominent during the spring, while errors at the low elevation

shrub and grass-dominated sites were more important during the monsoon. We propose a range of hypotheses for why model GPP does not respond sufficiently to changing water availability that can serve as a guide for future dryland DGVM developments. Our analysis suggests that improvements in modeling C cycle processes across more than a quarter of the Earth's land surface could be achieved by addressing the moisture sensitivity of dryland C uptake.

1. Introduction

Terrestrial ecosystems act as a global sink of carbon, C, absorbing ~30% of anthropogenic emissions. However, projections of the future fate of this land C sink are uncertain [1]. To improve our predictions of whether the land will remain a sink of C, we need to better understand how terrestrial C cycle related processes respond to climate variability. Several studies have examined which processes, and which regions, are contributing most to net CO_2 flux (e.g. net ecosystem exchange—NEE) interannual variability (IAV) [2–6]. Analyses of atmospheric CO_2 inversions, satellite data, and dynamic global vegetation model (DGVM) simulations indicate that dryland ecosystems dominate both the trend and IAV in the global C sink due to the sensitivity of vegetation growth to changes in water availability [7–11]. While well-tested in mesic ecosystems [12–16], DGVMs (which often form the land component of earth system models, ESMs, used in IPCC climate change projections) have been rarely tested against net CO_2 flux data from dryland regions (though see [17–19] for evaluations of modeled gross CO_2 uptake). DGVMs have performed poorly in comparison to satellite-based observations of seasonal to decadal trends in dryland vegetation dynamics [20–22], suggesting that DGVM estimates of gross CO_2 uptake (and therefore net CO_2 exchange) may be inaccurate.

Model evaluation and testing of gross and net CO_2 fluxes is needed to ensure dryland C cycle processes are well represented in DGVMs before they can be reliably used to predict the role of dryland ecosystems in the global C cycle. Dryland ecosystems encompass a wide range of ecosystems from semi-arid forests to shrublands, savannas and grasslands [23]. At intra- to inter-annual timescales, variability in dryland C fluxes is mostly caused by variability in climate drivers [3, 24]. Unlike mesic systems, the timing of moisture inputs strongly controls GPP [24, 25]; therefore, dryland ecosystem influence on the global C cycle IAV may be mediated through their high sensitivity to moisture availability [5, 7, 11]. Model failure to capture dryland C fluxes could result from misspecification of meteorological drivers or poor performance during critical precipitation seasons [26]. To date, the moisture sensitivity of C fluxes in DGVMs has not been well tested at water-limited dryland sites.

A network of 12 long running, eddy covariance Ameriflux flux tower sites spanning grassland, shrub, and forested ecosystems in the semi-arid southwestern US (hereafter, SW US) [24, table S1 and figure S1A] provides a rare opportunity to evaluate a suite of DGVMs at dryland sites dominated by changing water availability. In this study, we used this network of sites to evaluate seasonal and annual NEE and gross CO_2 fluxes (gross primary productivity, GPP, total ecosystem respiration, R_{eco}, and evapotranspiration, ET) simulated by 14 process-based DGVMs against *in situ* flux tower data. Simulations from the 14 DGVMs were taken from the TRENDY model intercomparison project (MIP) [27]; https://sites.exeter.acuk/trendy) version 7, which contributed to the Global Carbon Project's annual Global Carbon Budget [28]. Our primary objective was to evaluate: (1) whether the models could reproduce the observed annual net CO_2 flux (NEE) dynamics across this range of dryland sites. The final three objectives of our study were designed to diagnose the causes of any discrepancies in modeled dryland annual NEE. More specifically, we aimed to determine: (2) whether model-data discrepancies could be alleviated by running a DGVM with site-level forcing and vegetation and soil characteristics?; (3) which seasons, and which of the gross CO_2 fluxes, were responsible for model discrepancies in predicting dryland mean annual NEE and its variability, and whether the season and gross CO_2 flux responsible was different for high elevation forested sites versus low elevation shrub and grassland sites?; and (4) whether incorrect model sensitivity to climate drivers, and particularly those related to moisture availability, is causing the discrepancies in the key seasons and gross CO_2 fluxes identified in (3)? The analysis associated with the last objective allows us to evaluate which C cycle related processes may need improvement before DGVMspt can be used for reliably predicting the role of dryland ecosystems in the global C cycle.

2. Materials and methods

2.1. Southwestern US semi-arid sites

We evaluated the TRENDY v7 models against net and gross CO_2 fluxes from 12 semi-arid AmeriFlux sites in the SW US that spans tree, shrub, and grass dominated sites and elevations ranging from ~150 m to ~3050 m (AmeriFlux Network, 2021 - figure S1A and see table S1 (available online at

stacks.iop.org/ERL/16/094023/mmedia) for information on vegetation and soil characteristics for each site plus observation time period and site DOI) [24]. This part of the SW US is within the North American Monsoon region; therefore, these sites typically experience much of their rainfall during July to October, preceded by a hot, dry period in May and June ([24] figures 2(d)–(f)). The lower elevation (⩽1610 m) C3 shrub- and C4 grass-dominated sites have mean annual temperatures of 10 °C–20 °C and are predominantly driven by summer monsoon precipitation; however, winter and spring rains can also contribute to more ephemeral spring growing seasons at these sites [24, 25, 29, 30]. The high elevation (⩾1930 m) forested (conifer) sites experience cooler mean annual temperatures of <10 °C, and also have bi-modal growing seasons with available moisture coming from winter precipitation and summer monsoon rainfall [30, 31]. Sites are categorized throughout based on their mean annual NEE (see figure S1B): High elevation forest-dominated sites (US-Vcm, US-Vcp, US-Mpj, US-Fuf, US-Wjs and US-Ses) are a mean annual sink of C; whereas low elevation shrub- and grass-dominated sites (US-Wkg, US-SRG, US-Seg, US-SRM and US-Whs) 'pivot' between being a mean annual C sink or source, depending on annual water availability (figure S1B). One low elevation grassland site is a mean annual source of C to the atmosphere (US-Aud) [24].

2.2. *In situ* CO_2 and water flux data processing and analysis

Eddy covariance flux tower instruments at all sites collect half-hourly measurements of surface energy fluxes and NEE. NEE was partitioned into GPP and R_{eco} by each site PI. Gross CO_2 fluxes (GPP and R_{eco}) were calculated from the net CO_2 flux using the relationship between nighttime NEE and temperature [24]. Note that in this study a negative NEE implies a net CO_2 uptake into the ecosystem. Eddy covariance latent heat flux data were processed to provide evapotranspiration (ET). ET gaps were filled using a modified look-up table approach based on [32], with ET predicted from meteorological conditions within a 5 d moving window. We calculated seasonal CO_2 and water fluxes by summing the daily fluxes for the following months: November to February inclusive for the cool (winter) period; March to June hotter pre-monsoon (spring) period; and July to October for the monsoon. Note that the spring could be further split into the warm, moist months of March and April followed by the hotter, drier months of May and June; however, this entire period is marked by relatively warm and dry, moisture limited conditions compared with the winter or monsoon. To determine which seasons and gross CO_2 fluxes may be responsible for underestimate in NEE IAV, we examined the R^2 values obtained from the linear regression between the observed annual and seasonal C fluxes and the annual NEE.

2.3. TRENDY MIP v7 models and simulation protocol

Section 2.3.2 and tables 4 and A1 in [28, and references therein] provide details on the 14 DGVMs participating in TRENDY v7, as well as a description of the TRENDY MIP v7 protocol, including forcing data, simulation set-up, vegetation map, atmospheric CO_2 concentration data and land use change (LUC) datasets used. Simulated variables were downloaded from https://sites.exeter.acuk/trendy/. The models were forced by monthly CRU or the merged 6 hourly CRU–JRA-55 climate reanalysis datasets at $0.5 \times 0.5°$ resolution that start in the year 1901 and have been updated to 2017 [33]. The protocol was as follows: (a) first a spinup was performed by cycling the climate forcing over the 1901–1920 period with a fixed global PFT map and atmospheric CO_2 concentration level from the year 1700 (276.59 ppm); (b) a transient simulation from 1700 to 1900 with changing atmospheric CO_2 based on proxy and measured data, transient land-use changes, and cycling of climate forcing over 1901–1920; and (c) a historical simulation with climate forcing from 1901 to 2017 with changing atmospheric CO_2, nitrogen deposition (if used) and land-use changes (S3 simulation). The spinup is run to ensure the C stocks reach equilibrium, which can be a different time period for each model but effectively allows for several thousand simulation years. Atmospheric CO_2 concentrations are based on ice-core proxy data (pre-1958) and measured atmospheric mole fraction data from the Mauna Loa and South Pole Observatory stations (post-1958) provided by the NOAA Earth System Research Laboratory [28—section 2.4.1]. The models either prescribe static PFT fractions per grid cell or simulate dynamic vegetation changes over long timescales (annual to millennial); however, historical LUC is imposed in the models. LUC is based on gross land use transitions from the land use harmonization v2 (LUH2) dataset [34] and net transitions from the HYDE (History Database of the Global Environment) v3.2 [35]. Each modeling group has its own protocol for merging this LUC data with their own PFT descriptions. Individual modeling groups also use their own expert judgment for setting their model parameter values, other required forcing data streams, and soil texture and permeable depth information. Model grid cells corresponding to each site location were extracted from the global simulations for our analysis.

2.4. ORCHIDEE DGVM site simulation set-up

Typically, model-data discrepancies are expected when comparing coarse grid (∼50 × 50 km) scale climate reanalysis forcing data used in TRENDY model simulations against *in situ* observations, which are

representative of an area of ~1–2 km^2. Therefore, we tested whether scale mismatch was responsible for model-data misfits by running the ORCHIDEE v2.0 DGVM (equivalent to 'ORCHIDEE v2.0' used in TRENDY v7) with site-based meteorological forcing, vegetation fractional cover and type, and soil texture fractions (hereafter referred to as ORCHIDEE_SL for 'site level'). Meteorological forcing data used to run the ORCHIDEE_SL simulations for each of the 12 sites included 2 m air temperature and surface pressure, precipitation, incoming long and shortwave radiation, wind speed, and specific humidity. These data were downloaded from the AmeriFlux data portal for each site (http://ameriflux.lbl.gov). The meteorological forcing data were gap-filled using downscaled and corrected ERA-Interim reanalysis data [36]. *In situ* measured precipitation was partitioned into rainfall and snowfall using a temperature threshold of 0 °C. We followed the same protocol for the site simulations that was used in TRENDY v7 (see section 2.3), with the following exceptions: (a) we performed an analytical 400 year spinup [37] by cycling over the available gap-filled forcing data for each site (table S1) followed by the transient and historical simulations; and (b) plant functional types (PFT) and soil texture fractions and maximum LAI were prescribed in the model based on the current data and literature for each site and did not change during the spinup, transient or historical simulations (table S1).

2.5. Model evaluation analysis

Several different metrics are used to evaluate the model annual NEE, GPP and R_{eco}, including mean bias error (MBE), the slope of the linear regression between model and observations, coefficient of determination and concordance correlation (ρ_c) [38]. Beyond our initial evaluation of the models' annual NEE using standard model evaluation metrics, we expanded our analysis in the following three additional steps: First, we examined the mean bias error (MBE) and bias contribution to the mean squared deviation (MSD—see below for MSD decomposition method) between modeled and observed NEE, GPP and R_{eco} to diagnose factors causing model discrepancies in mean annual NEE at the C sink and source sites. Second, we analyzed the slope of the linear relationship between model and observations, as well as the mean squared variation (MSV) contributions to the MSD (including phase and variance components—see below), to identify what is causing models' failure to capture NEE IAV. The analyses in these two steps were performed at sub-annual timescales in order to determine which seasons, and which of the gross CO_2 fluxes, are responsible for model-data discrepancies in annual NEE; thus, these two steps were intended to address objective 3 outlined in the introduction. Finally, for the key seasons controlling discrepancies NEE IAV we tested whether the models captured CO_2 flux response to variability in a range of climate drivers and measures of water availability (including observed air temperature, precipitation (rainfall and snowfall), vapor pressure deficit (VPD), soil moisture, and evapotranspiration, ET) given that variability in these drivers is a dominant control on intra- to interannual C flux variability [3, 24]. This final step was designed to address objective 4 of our study. We included an evaluation of the sensitivity of C fluxes to measured ET as a proxy of plant water availability because in water-limited ecosystems, ET is a more integrated metric of plant water availability (as opposed to soil moisture or precipitation) [24, 39] because plants have developed a range of strategies for dealing with limited water availability, including the ability to capture water from either deep or lateral root systems. Thus, ET accounts for water inputs, runoff losses and changes in soil moisture or snowpack storage, as well as differences in plant stomatal regulation and other vegetation characteristics that affect the energy balance. Another advantage of using ET is that it is measured at the same temporal scale, over the same flux footprint, and by the same instruments as NEE, whereas soil moisture is measured differently at every site.

We partitioned MSD between the model and the observations into bias, phase, and variance components [40]:

$$\mathrm{MSD} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 = (\bar{x} - \bar{y}) + \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \bar{x}) - (y_i - \bar{y})\right]^2 \quad (1)$$

where x is the model estimate and y is the observations. We calculated a separate MSD for each of the TRENDY models and each variable of interest (e.g. NEE and GPP). The first term on the right-hand side of equation (1) represents the squared bias. The second term represents the MSV. The MSV represents the ability of the model to simulate variability about the mean. The MSV can be further partitioned into phase and variance components:

$$\mathrm{MSV} = \left(\sigma_x - \sigma_y\right)^2 + 2\sigma_x\sigma_y(1 - R) \quad (2)$$

where σ is the standard deviation and R is the correlation coefficient between the model and observations. The first term on the right-hand side in equation (2) indicates the difference between the model and observations in the magnitude of variability (i.e. the variance component, or in the magnitude of the variations) [40, 41]. The second term on the right-hand side of equation (2) represents a lack of correlation weighted by standard deviations [40]; thus, this final term can be thought of as a difference in phase between the model and observation [41]. The contributions of individual components to the overall MSD (or MSV) are calculated by dividing the bias,

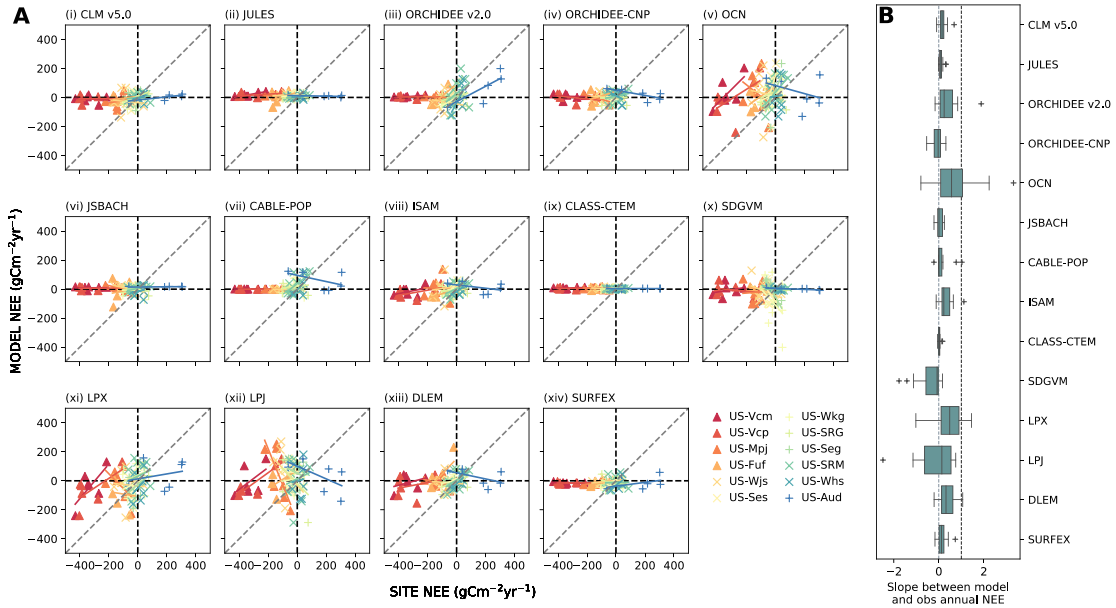

Figure 1. Panel (A): Scatter plots comparing modeled annual NEE at each site to observations for each TRENDY v7 DGVM (see table S1 for observation period) with linear regression trendlines included. Negative NEE values indicate a carbon sink. Ideally, all points—whether a sink, source or pivot site—should lie on the 1:1 grey dashed diagonal line with a regression slope of 1 (i.e. vertical dashed line in (B)) if the models capture IAV well. Sites (site descriptions in table S1) are ordered according to mean annual NEE (figure S1B) from sink (US-Vcm) to source (US-Aud). Triangles depict tree dominated sites; crosses depict shrub-dominated sites and plus signs depict grass-dominated sites. Red and orange colors show the high elevation mostly-forested sink sites; yellow, green and blue colors show the low elevation shrub and grass dominated mostly pivot sites. Panel (B): Boxplots of the slope of the linear regression between modeled and observed annual NEE across all sites for each TRENDY DGVM. The vertical black dashed line shows the ideal slope of 1, and the grey dashed line shows a slope of zero.

phase, and variance by the MSD (or MSV). We calculated all the above metrics at both annual and seasonal timescales.

3. Results

3.1. Evaluation of DGVM annual NEE

Across all sites, the DGVMs generally underestimate the mean annual NEE (as seen by model values clustering around a y-axis value of 0.0 in figure 1(a) instead of on the 1:1 line, median slope values of 0 in figure 1(b), and high mean bias errors, MBE, in figure S2(a)). The models only simulate a weak C sink or source, irrespective of the strength of the observed sink/source. Similarly, the majority of models underestimate the magnitude of NEE IAV and fail to capture the correct IAV sign (as seen by slopes $\ll 1$ in figures 1(a) and (b), green and white colors in figure S2(b), R^2 values $\ll 1$ in figure S2(c) and concordance correlation, ρ_c, <0.5 in figure S2(d)).

Decomposition of the annual NEE MSD into its component parts (see section 2.5) allows us to determine whether the MSD is dominated by the mean bias versus model inability to capture the variability about the mean (i.e. high MSV). The high elevation C sink sites have the strongest positive biases (model minus observations) in annual NEE (MBE of 100–400 g $Cm^{-2}\,yr^{-1}$—figure S2(a)), while the mean C source site (US-Aud) has a negative MBE. Low elevation pivot sites have low biases because their mean annual NEE is close to zero. Biases therefore dominate the mean sink and source site contributions to MSD (bias contribution to MSD > 0.5 in figure S3(a)), whereas the pivot sites are dominated by model inability to capture variability about the mean (phase plus variance contributions to MSD > 0.5—figures S3(b) and (c)). Across all sites, models' inability to replicate the annual variability is generally more related to inaccurate phase rather than model failure to capture the magnitude of fluctuations (variance component) (cf figures S3(c) with S3(b) although there is considerable spread across models.

3.2. Site-based ORCHIDEE DGVM NEE simulations and evaluation

The site-level simulations with ORCHIDEE v2.0 (see section 2.4) yield a similar picture: ORCHIDEE_SL underestimated mean annual NEE and the magnitude of the NEE IAV at all sites (figure S4(a)). Given most of the models have similar representations of the main C cycle processes (see [28] and references therein), we suggest that these site level simulations with ORCHIDEE v2.0 demonstrate that inaccurate forcing or vegetation and soil characteristics are likely not the cause of DGVM underestimates in mean annual NEE and IAV; therefore, we need to analyze the models further to determine the root causes of these model-data discrepancies. The median slope of regression between simulated and observed annual

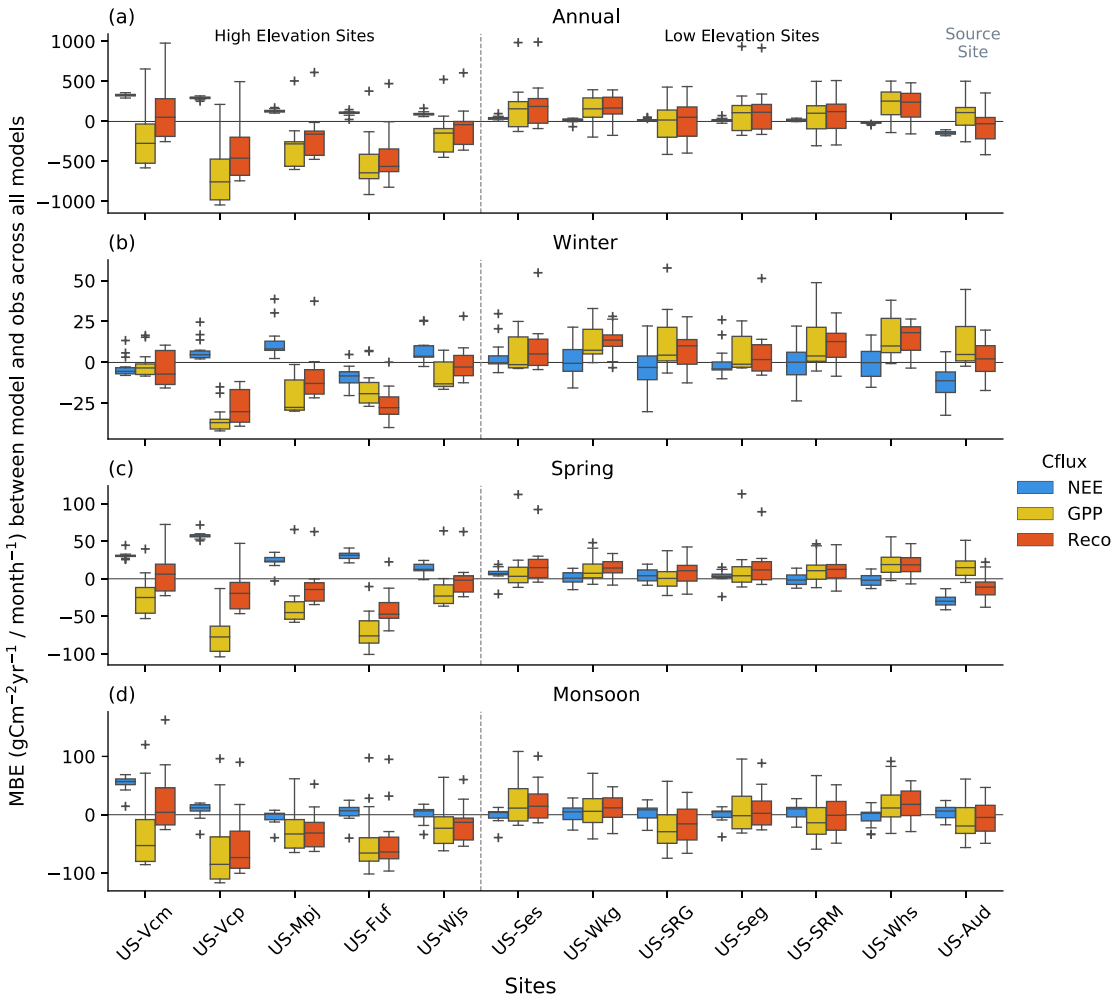

Figure 2. Box and whisker plots of the mean bias error (MBE) between simulated and observed net and gross CO_2 fluxes spread across all TRENDY v7 DGVMs ($n = 14$) for each site at annual and seasonal timescales. The MBE is the mean of the differences for each timepoint (or alternatively, the difference in the means). MBE is calculated as model minus observations; therefore, for NEE, underestimates in NEE at the sink sites result in positive MBEs because NEE is negative for net CO_2 uptake at sink sites, whereas at the C source site, underestimates in NEE result in negative MBEs. For both GPP and R_{eco}, underestimates result in negative MBEs (and vice versa for overestimates). Annual units are g Cm^{-2} yr^{-1} and seasonal units are g Cm^{-2} month^{-1}. Seasonal fluxes are summed over the following months: November to February inclusive for the cool (winter) period; March to June pre-monsoon (spring) period; and July to October for the summer monsoon growing season. Sites (site descriptions in table S1) are ordered according to mean annual NEE (figure S1B) from sink (US-Vcm) to source (US-Aud).

NEE was 0.24 for both ORCHIDEE_SL and ORCHIDEE v2.0 in TRENDY v7, albeit with differences across sites, with a range from −0.11 to 1.2 for the site level simulations and −0.17–1.88 for TRENDY. Similarly, the median R^2 was 0.3 for both site level and ORCHIDEE_SL and ORCHIDEE TRENDY simulations, with a range from 0.01 to 0.87 for ORCHIDEE_SL and 0.004–0.8 for TRENDY. In both coarse grid and site-level simulations, the strongest biases were seen at high C sink (and source) sites (figure S4). The median slope of regression between simulated and observed annual NEE was 0.24 for both ORCHIDEE_SL and ORCHIDEE TRENDY, albeit with differences across sites (figure S4). Similarly, the median R^2 was 0.3 for both site level and ORCHIDEE_SL and ORCHIDEE TRENDY simulations.

3.3. Seasonal and gross CO_2 flux contributions to model-data discrepancies in mean annual NEE and IAV

To diagnose the processes that are responsible for model-data discrepancies, we examined the seasonal and gross CO_2 flux contributions to biases in the mean annual NEE. Separating annual net and gross fluxes into component seasonal fluxes (see section 2.2) allows us to identify that model underestimates of mean annual NEE at sink and source sites (positive and negative NEE MBEs, respectively) are most clearly associated with biases in spring NEE (cf figures 2(a) and (c)), with the exception of US-Vcm (strongest sink site), for which biases during the summer monsoon NEE also play a key role (figure 2(d)). Comparing the model to observed

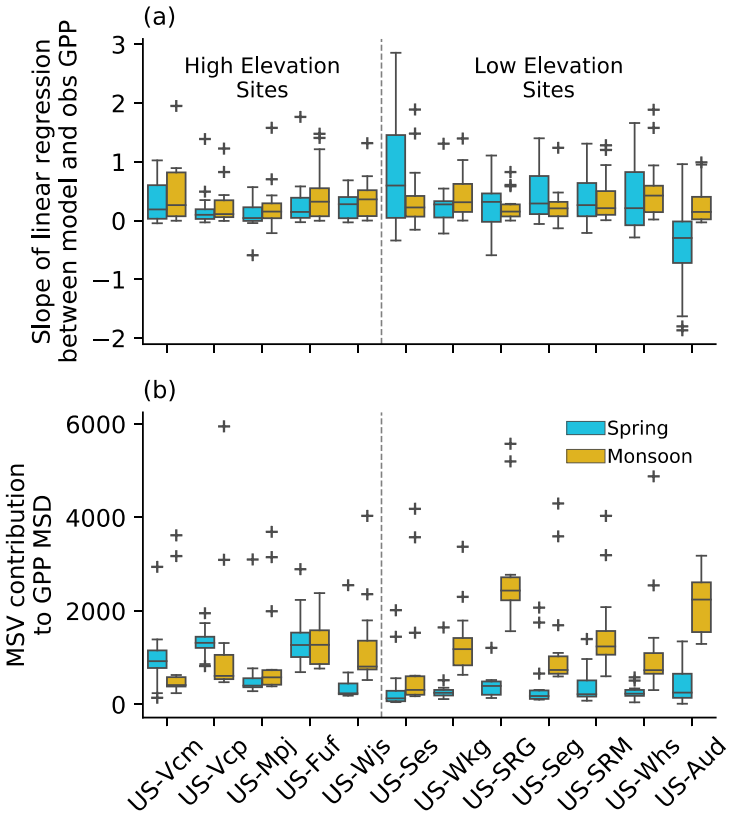

Figure 3. Box and whisker plots showing the median and spread across models ($n = 14$) for each site of (a) the MSVs contribution to GPP MSD in the spring (blue boxes) and monsoon (yellow boxes); and (b) the slope of the linear regression between modeled and observed spring and monsoon GPP. The vertical dashed grey line separates the high elevation forested sink sites (US-Vcm, US-Vcp, US-Mpj, US-Fuf, US-Wjs) and the low elevation shrub and grass dominated (mostly pivot) sites (US-Ses, US-Wkg, US-SRG, US-Seg, US-SRM, US-Whs, US-Aud). The MSV contribution shown in (a) comprises both the variance and phase contributions. See figure S9 for a breakdown into these two components of the MSD. See section 2.5 for more detail on the MSD decomposition into bias and MSV, and further into decomposition of MSV into phase and variance components. Units are g Cm^{-2} $month^{-1}$. Spring and monsoon fluxes are summed over the following months: March to June for the hot spring (pre-monsoon) period; and July to October for the monsoon. Sites (site descriptions in table S1) are ordered according to mean annual NEE (figure S1B) from sink (US-Vcm) to source (US-Aud).

gross CO_2 fluxes we see that at high elevation sites (US-Vcm, US-Vcp, US-Mpj, US-Fuf, and, to a lesser extent, US-Wjs), the underestimate in spring NEE (positive MBEs) are mostly due to an underestimate in spring GPP (rather than an overestimate in R_{eco}—figure 2(c)). In the summer monsoon growing season, both GPP and R_{eco} are underestimated by most models (figure 2(d)), which results in monsoon NEE MBE values closer to zero (except for US-Vcm). The negative MBE in annual NEE at US-Aud (underestimate in the strength of the mean C source) across all models is associated with the negative biases in winter and spring (figures 2(b) and (c)). The models fail to capture the earlier rise in R_{eco} than GPP in the winter and spring at US-Aud, followed by a later increase in GPP. Instead, all models have almost the same mean seasonal trajectory in both GPP and R_{eco} (figure S5l). However, US-Aud was still recovering from a fire that occurred in 2002 until late 2005 [28]. Even if the models contain representations of wildfire, these schemes generally capture broad spatial fire patterns and are not likely to capture specific fires at one flux tower site; therefore, it is unlikely that the 2002 fire at US-Aud was correctly simulated, possibly explaining model discrepancies in winter and spring GPP and R_{eco}. Finally, we note that partitioning of NEE into the component of gross CO_2 fluxes is subject to uncertainties [42] that likely impact the absolute magnitude of the C fluxes.

Preliminary model evaluation showed that there was no clear distinction between seasons or gross CO_2 fluxes in terms of the models' ability to capture NEE IAV: The models did not capture NEE IAV well in any season, or in either of the gross CO_2 fluxes (low slope values, and high MSV contributions to MSD across all seasons and fluxes—figure S6). Therefore, to focus our model evaluation analysis into only the seasons and gross CO_2 fluxes that are crucial for controlling NEE IAV, we performed a site-by-site analysis of the measured flux data. This analysis indicated that our model IAV evaluation efforts should focus on spring GPP at the high elevation sites, and monsoon GPP at the low elevation sites (figure S7

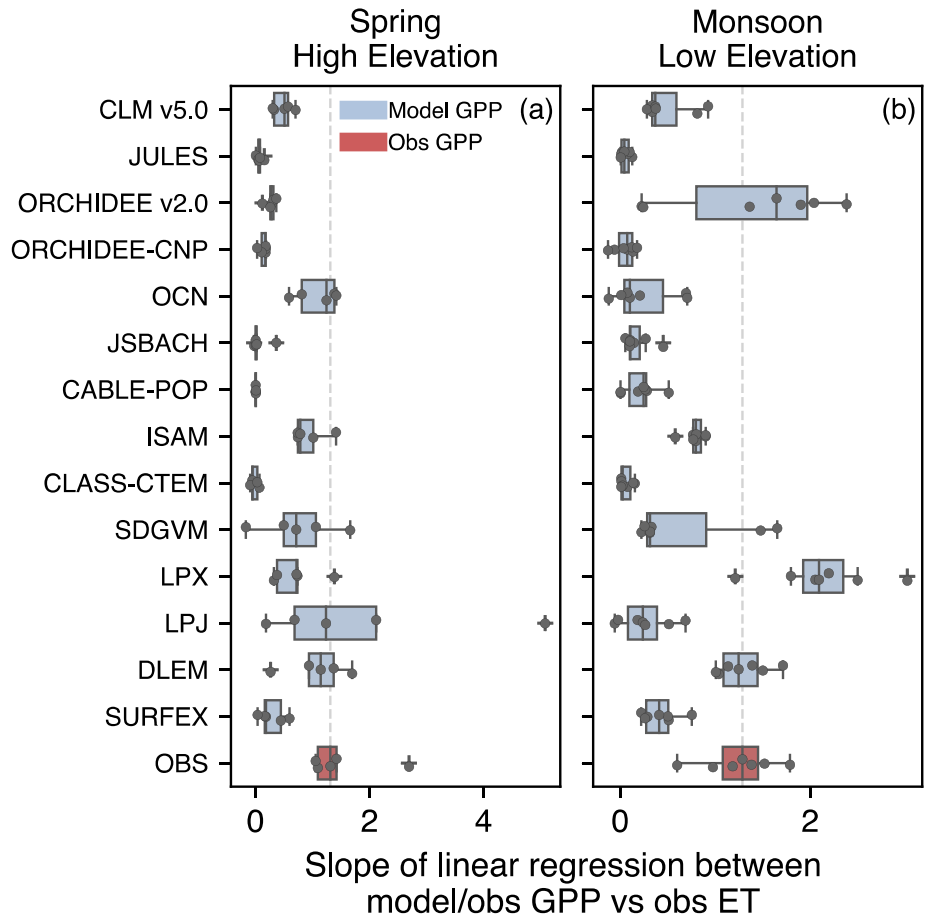

Figure 4. Boxplots of the slope of the linear relationship between monthly simulated GPP and monthly observed ET (here used a proxy for the GPP response to changing plant water availability) for each model (blue boxes) and the slope of the linear relationship between observed GPP and observed ET (red box) across (a) all high elevation (⩾1930 m) forested sink sites (US-Vcm, US-Vcp, US-Mpj, US-Fuf, US-Wjs) during the spring (March to June, inclusive); and (b) all low elevation (⩽1610 m) shrub and grass pivot and source sites (US-Ses, US-Wkg, US-SRG, US-Seg, US-SRM, US-Whs, US-Aud) during the monsoon (July to October, inclusive). Grey dashed line shows the median observed GPP/ET slope across each set of sites.

and see caption for more information). Models' failure to capture variability in spring GPP at high elevation forest sites (blue bars in figure 3), and variability in monsoon GPP across all low elevation grass- and shrub-dominated sites (orange bars in figure 3), are the predominant causes of their inability to replicate observed NEE IAV. High elevation sites tend to have lower slopes during the spring than the low elevation pivot sites (excluding US-Aud as a C source site—figure 3(a) cf blue bars to left of vertical grey dashed line for high elevation sites compared to those on the right for low elevation sites). Similarly, the spring GPP variance and phase components of the MSD (i.e. the MSV contribution to MSD) are generally larger at the higher elevation sink sites than at the lower elevation sites (figure 3(b)). We note that a higher MSV contribution to MSD indicates a poorer model performance in capturing variability in GPP. During the monsoon there is generally less differentiation in GPP slope values between high and low elevation sites (figure 3(a) yellow bars). Monsoon GPP slopes tend to be low (<0.5) across all site simulations, although a few exceptions exist: LPX and DLEM have GPP slopes close to 1 across at most sites, while ORCHIDEE v2.0 generally has robust GPP slope values for the low elevation sites, and LPJ for the high elevation sites (figure S8). The MSV contribution to monsoon GPP MSD is much higher (i.e. poorer performance for GPP variability) for low elevation sites (and much higher than in the spring—cf yellow bars to the right of the vertical dashed line in figure 3(b) for the low elevation sites compared to those on the left of the line for the high elevation sites). Differences in variance between the model and observations are more responsible for lack of spring GPP variability (high spring MSV) at high elevation sink sites (figure S9(a)), whereas phase differences dominate both the spring and monsoon GPP variability at low elevation pivot sites (figure S9(b)).

3.4. Evaluation of GPP sensitivity to climate drivers

To diagnose the possible cause(s) of biases in modeled spring and monsoon GPP variability, we examined the relationships between GPP and various climate drivers. Comparing the modeled and observed slope of the linear relationship between modeled GPP and

observed ET (as a proxy of plant water availability—see section 2.5) reveals that all models generally underestimate the GPP response to changing plant water availability in the spring at high elevation sink sites and during the monsoon at the low elevation sites (light blue model GPP/observed ET slope ≪ red observed GPP/observed ET slope—figures 4(a) and (b), respectively). Several exceptions exist: at high elevation sites, the response of spring GPP to changing plant water availability estimated by OCN and DLEM is stronger than other models and matches the observations fairly well (figure 4(a)). The same is true for ORCHIDEE v2.0 and DLEM during the monsoon at low elevation sites (figure 4(b)). OCN had the smallest (i.e. closest to 0.0) median spring GPP MBE across all high elevation sites (followed by DLEM—results not shown). ORCHIDEE v2.0 and DLEM are two of the models that also performed well at low elevation sites in comparison to observed monsoon GPP, with model-data slopes around ∼1 (figure S8—see section 3.3). We chose to highlight the spring GPP response to changing plant water availability (ET) for the high elevation sites and monsoon GPP response for the low elevation sites in figure 4 because our analysis in section 3.3 pointed us towards these two combinations of seasons and site elevations as the most problematic in terms of simulating observed variability in GPP (also noting the importance of spring biases in GPP for capturing mean annual NEE at high elevation sites). However, in section 3.3 (figure 3) we also pointed out that the slope values between modeled and observed GPP were low for all sites during the monsoon. If we examine the GPP response to changing ET in both the spring and monsoon seasons at both high and low elevation sites (figure S10), we can see that many of the models systematically underestimate monsoon GPP moisture sensitivity at the high elevation sites (figure S10(c)), in addition to the spring. This further adds weight to what we found in section 3.3—that simulated GPP is an issue across all sites in the monsoon. Low elevation sites generally fare better in representing spring GPP moisture sensitivity (figure S10(b)), but again a few of the models are drastically underestimating this response. We also assessed the model GPP (and R_{eco}) relationships to other climate drivers and measures of water availability (see section 2.5). This evaluation revealed that observed sensitivity to most other climate drivers is weak or unclear (results not shown), except for the relationship between GPP and ET shown here. Therefore, we focused solely on the GPP-ET relationship. This finding was not surprising because, as discussed in section 2.5, past studies have found that ET is an integrated measure of plant water availability [24, 38]. The integrated nature of ET is further evidenced here by the fact we found a strong relationship between observed GPP and ET and only weak relationships with precipitation inputs and surface soil moisture (as well as other climate drivers such as T_{air} and VPD—results not shown).

4. Discussion and conclusion

Why do models underestimate spring GPP response to changing plant water availability at the high elevation sink sites (figure 4(a))? Comparing TRENDY modeled ET to observations, we found that almost all high elevation site simulations underestimate spring ET and its variability (figures S11(a) and (b)). If models cannot capture changing plant water availability (i.e. ET variability) correctly it is unlikely that the subsequent range of processes that contribute to GPP and NEE (e.g. phenology, C allocation, water-limitation on photosynthesis, stomatal conductance, etc) will be accurately simulated. Thus, we first need to determine why models are not capturing variability in spring ET at the high elevation forested sites. However, without additional information on variables that contribute to the calculation of ET, it is difficult to determine why these ET biases are occurring. [43] did perform a multi-variable evaluation of ORCHIDEE v2.0 water stores and fluxes against *in situ* ET and soil moisture, as well as satellite-derived LAI and snow cover, at two of the high elevation forest sites included in this study (US-Fuf and US-Vcp) and found that spring upper layer soil moisture was generally underestimated. Spring ET at the US-Fuf and US-Vcp was both over- and under-estimated, respectively. They hypothesized that these model-data biases could be the result of issues with model deficiencies in timing of snowmelt and/or the representation of snow cover under forest canopies. For example, the overestimate in spring ET at US-Fuf could be related to the fact that the simulated snowpack melted too early, thus causing a premature rise in bare soil evaporation. The positive biases in spring ET could explain the negative biases we observe in modeled GPP moisture sensitivity. At US-Vcp, the negative bias in spring ET was thought to be due to an underestimate in modeled LAI [43]. Low simulated LAI could be explained by the low soil moisture values and in turn could explain both the reduced ET and underestimate in forest site spring GPP. Dryland woody plant species often have deep taproots [44, 45] for accessing groundwater during periods of limited water availability, such as in the hotter, drier spring period. However, current DGVMs generally do not have a representation of either groundwater or deep-rooted plants; therefore, these missing processes could also explain both model biases in spring GPP and ET as well as lack of spring contributions to NEE IAV.

What could be the cause(s) of the models' failure to capture monsoon GPP response to changing plant water availability at the low elevation pivot sites (figure 4(b))? Our comparison of TRENDY

model ET to observations showed that most models do capture ET and its variability well at low elevation sites during the monsoon (figures S11(c) and (d)). This finding matches [43], who also evaluated ORCHIDEE v2.0 water stores and fluxes at a subset of the lower elevation shrub and grassland sites included here. The energy balance calculations and physically based description of soil hydrology in ORCHIDEE v2.0 are similar to the schemes implemented in most DGVMs that form the land component of ESMs [46]. We also validated the CRU-JRA v1.1 forcing data used in TRENDY v7 against the site meteorological data to test if the models are indeed seeing an increase in monsoon precipitation. This exercise showed that CRU-JRA captures monthly precipitation well, including the dramatic increase during the monsoon (figure S12). Therefore, models' inability to capture GPP variability during the monsoon is not because of inaccurate simulations of changing plant water availability (i.e. ET); instead, it is likely due to issues in the modeled GPP response to increases in plant available water (i.e. they underestimate ecosystem water use efficiency). Models' inability to respond to monsoon increases in water availability could be explained by several hypotheses: (a) inaccurate controls of soil moisture versus VPD on stomatal conductance [47]; (b) poor representation of desert plant hydraulic schemes; (c) model structural limits on maximum leaf area magnitude; (d) model photosynthetic parameters are not well adapted to dryland species [48], particularly for C4 plants [49]; (e) models inability to simulate dynamic changes in vegetation at seasonal timescales (e.g. to grow summer annuals in bare soil patches); (f) lack of drought-deciduous phenological strategies in models [18]; and (g) lack of model representation of biocrust contributions to biogeochemical cycles [50, 51]. Hypotheses related to phenology in particular could explain the mismatch in the phase component of MSV, while those related to photosynthesis, stomatal conductance, and biocrust activity may explain the mismatch in variance. These hypotheses may also explain models' failure to capture positive asymmetry at grassland sites in response to precipitation variability [52]. We also note that other resource limitations on biogeochemical cycling may be affecting models' ability to capture spring and monsoon GPP responses to increasing plant water availability. For example, OCN and DLEM, which did accurately simulate spring GPP (and ET) at high elevation sites, do include a representation of nitrogen (N) cycling, whereas many other models do not. Accurate estimates of N limitation on photosynthesis may reduce LAI, and thus result in greater overall soil moisture stores that can promote growth in more water-limited spring periods. On the other hand, ORCHIDEE-CNP, which contains both N and phosphorus (P) cycles, was unable to replicate variability in monsoon season GPP or ET (figures S8 and S11(d)), while ORCHIDEE v2.0, which does not account for N and P limitations on GPP and is the same model in all other respects, had a better performance in terms of capturing monsoon GPP variability. Clearly, models should account for nutrient limitation on GPP; therefore, this comparison suggests that ORCHIDEE v2.0 might have a better model-data fit but for the wrong reasons. A recent global evaluation of ORCHIDEE-CNP found that the model simulates too strong a nutrient limitation on GPP for tropical grasses (equivalent to the grasses in this study domain) [53]. Therefore, more development, parameterization, and testing of leaf and soil stoichiometry is likely needed to accurately simulate nutrient limitation on photosynthesis and all related carbon-water interactions in dryland ecosystems.

DGVMs drastically underestimate dryland ecosystem mean annual NEE and its IAV, likely due to deficiencies in vegetation response to changing moisture availability. Further testing, optimization, and development of models at site-level against daily C fluxes is needed before we can rely on them to accurately represent dryland ecosystem processes across 40% of the terrestrial surface. A targeted dryland multi model-data intercomparison and model optimization project would help to fully diagnose what is causing errors in modeled dryland C fluxes. Within such a project we could use the models to test competing hypotheses as to which processes are responsible for model deficiencies by evaluating different model formulations against a wider range of field data and manipulation experiments [e.g. 54] from different dryland ecosystems worldwide. Modelers should also collaborate more extensively with empirical scientists to make best use of existing data and to collect or derive new dryland C cycle related datasets where needed. Once this cycle of model hypothesis testing and development is complete, we may find that dryland regions play an even greater contribution to global C cycle IAV than previously thought.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: http://ameriflux.lbl.gov.

Acknowledgments

NM was supported by a grant from the Macrosystems Program in the Emerging Frontiers Section of the US National Science Foundation (NSF Award 1065790). DG received support from the French Agence Nationale de la Recherche (ANR) Convergence Lab Changement climatique et usage des terres

(CLAND). AK was supported by Oak Ridge National Laboratories (ORNL). ORNL is managed by UT-Battelle, LLC, for the DOE under Contract DE-AC05-1008 00OR22725. Partial funding for some of the AmeriFlux sites run by RLS and MEL was provided by the US Department of Energy's Office of Science, NSF funding to the Sevilleta LTER program, and the USDA. USDA is an equal opportunity employer. PK and TM were funded by the NOAA OAR/ARL Climate Research Program. SS was supported by NERC Driving-C project (NE/R00062X/1). SZ was supported by the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 821003 (4C-project). The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. DLL was supported by the U.S. Department of Agriculture NIFA Award 2015-67003-23485. We would like to thank the ORCHIDEE Project Team for developing and maintaining the ORCHIDEE code and for providing the ORCHIDEE version 2.0 used in this study. Finally, we thank the two anonymous referees for their comprehensive and useful reviews.

ORCID iDs

Natasha MacBean https://orcid.org/0000-0001-6797-4836
Thomas Kolb https://orcid.org/0000-0002-3829-4265
Praveena Krishnan https://orcid.org/0000-0002-5760-3254
Danica L Lombardozzi https://orcid.org/0000-0003-3557-7929

References

[1] Friedlingstein P *et al* 2019 Global Carbon Budget 2019 *Earth Syst. Sci. Data* **11** 1783–838
[2] Fu Z *et al* 2017 Climate controls over the net carbon uptake period and amplitude of net ecosystem production in temperate and boreal ecosystems *Agric. For. Meteorol.* **243** 9–18
[3] Jung M *et al* 2017 Compensatory water effects link yearly global land CO2 sink changes to temperature *Nature* **541** 516–20
[4] Marcolla B, Rödenbeck C and Cescatti A 2017 Patterns and controls of inter-annual variability in the terrestrial carbon budget *Biogeosciences* **14** 3815–29
[5] Humphrey V, Zscheischler J, Ciais P, Gudmundsson L, Sitch S and Seneviratne S I 2018 Sensitivity of atmospheric CO2 growth rate to observed changes in terrestrial water storage *Nature* **560** 628–31
[6] Liu Z, Ballantyne A P, Poulter B, Anderegg W R L, Li W, Bastos A and Ciais P 2018 Precipitation thresholds regulate net carbon exchange at the continental scale *Nat. Commun.* **9** 3596
[7] Poulter B *et al* 2014 Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle *Nature* **509** 600–3
[8] Ahlstrom A *et al* 2015 The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink *Science* **348** 895–9
[9] Zhang X, Wang Y-P, Peng S, Rayner P J, Ciais P, Silver J D, Piao S, Zhu Z, Lu X and Zheng X 2018 Dominant regions and drivers of the variability of the global land carbon sink across timescales *Glob. Change Biol.* **24** 3954–68
[10] Mekonnen Z A, Grant R F and Schwalm C 2017 Carbon sources and sinks of North America as affected by major drought events during the past 30 years *Agric. Forest Meteorol.* **244–245** 42–56
[11] Huang L, He B, Chen A, Wang H, Liu J, Lű A and Chen Z 2016 Drought dominates the interannual variability in global terrestrial net primary production by controlling semi-arid ecosystems *Sci. Rep.* **19** 24639
[12] Schwalm C R *et al* 2010 A model-data intercomparison of CO2 exchange across North America: results from the North American Carbon Program site synthesis *J. Geophys. Res.* **115** G00H05
[13] Keenan T F *et al* 2012 Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange *Glob. Change Biol.* **18** 1971–87
[14] Schaefer K *et al* 2012 A model-data comparison of gross primary productivity: results from the North American Carbon Program site synthesis *J. Geophys. Res.* **117** G03010
[15] Raczka B M *et al* 2013 Evaluation of continental carbon cycle simulations with North American flux tower observations *Ecol. Monogr.* **83** 531–56
[16] Peng S *et al* 2015 Benchmarking the seasonal cycle of CO2 fluxes simulated by terrestrial ecosystem models *Glob. Biogeochem. Cycles* **29** 46–64
[17] Whitley R, Beringer J, Hutley L, Abramowitz G, De Kauwe M G and Duursma R 2016 A model inter-comparison study to examine limiting factors in modelling Australian tropical savannas *Biogeosci.* **13** 3245–65
[18] Renwick K M, Fellows A, Flerchinger G N, Lohse K A, Clark P E, Smith W K, Emmett K and Poulter B 2019 Modeling phenological controls on carbon dynamics in dryland sagebrush ecosystems *Agric. For. Meteorol.* **15** 85–94
[19] Chen C, Cleverly J, Zhang L, Yu Q and Eamus D 2016 Modelling seasonal and inter-annual variations in carbon and water fluxes in an arid-zone acacia savanna woodland, 1981–2012 *Ecosystems* **19** 625–44
[20] Traore A *et al* 2014 1982–2010 trends of light use efficiency and inherent water use efficiency in African vegetation: sensitivity to climate and atmospheric CO2 concentrations *Remote Sens.* **6** 8923–44
[21] Dahlin K M, Fisher R A and Lawrence P J 2015 Environmental drivers of drought deciduous phenology in the Community Land Model *Biogeosciences* **12** 5061–74
[22] MacBean N, Maignan F, Peylin P, Bacour C and Bréon F-M C P 2015 Using satellite data to improve the leaf phenology of a global terrestrial biosphere model *Biogeosciences* **12** 7185–208
[23] Gaur M K and Squires V R 2018 Geographic extent and characteristics of the world's arid zones and their peoples *Climate Variability Impacts on Land Use and Livelihoods in Drylands* ed M K Gaur and V R Squires (Berlin: Springer) pp 3–20
[24] Biederman J A *et al* 2017 CO2 exchange and evapotranspiration across dryland ecosystems of southwestern North America *Glob. Change Biol.* **23** 4204–21
[25] Scott R L, Biederman J A, Hamerlynck E P and Barron-Gafford G A 2015 The carbon balance pivot point of southwestern U.S. semiarid ecosystems: insights from the 21st century drought *J. Geophys. Res.* **120** 2612–24

[26] Barnes M L, Susan Moran M, Scott R L, Kolb T E, Ponce-Campos G E, Moore D J P, Ross M A, Mitra B and Dore S 2016 Vegetation productivity responds to sub-annual climate conditions across semiarid biomes *Ecosphere* **7** e01339
[27] Sitch S *et al* 2015 Recent trends and drivers of regional sources and sinks of carbon dioxide *Biogeosciences* **12** 653–79
[28] Quéré C L *et al* 2018 Global Carbon Budget 2018 *Earth Syst. Sci. Data* **10** 2141–94
[29] Krishnan P, Meyers T P, Scott R L, Kennedy L and Heuer M 2012 Energy exchange and evapotranspiration over two temperate semi-arid grasslands in North America *Agric. For. Meteorol.* **153** 31–44
[30] Anderson-Teixeira K J, Delong J P, Fox A M, Brese D A and 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 *Glob. Change Biol.* **17** 410–24
[31] Dore S, Kolb T E, Montes-Helu M, Eckert S E, Sullivan B W, Hungate B A, Kaye J P, Hart S C, Koch G W and Finkral A 2010 Carbon and water fluxes from ponderosa pine forests disturbed by wildfire and thinning *Ecol. Appl.* **20** 663–83
[32] Falge E *et al* 2001 Gap filling strategies for defensible annual sums of net ecosystem exchange *Agric. For. Meteorol.* **107** 43–69
[33] Harris I, Jones P D, Osborn T J and Lister D H 2014 Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset *Int. J. Climatol.* **34** 623–42
[34] Hurtt G C *et al* 2020 Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6 *Geosci. Model Develop.* **13** 5425–64
[35] Goldewijk K K, Beusen A, Doelman J and Stehfest E 2017 Anthropogenic land use estimates for the Holocene—HYDE 3.2 *Earth Syst. Sci. Data* **9** 927–53
[36] Vuichard N and Papale D 2015 Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis *Earth Syst. Sci. Data* **7** 157–71
[37] Lardy R, Bellocchi G and Soussana J-F 2011 A new method to determine soil organic carbon equilibrium *Environ. Model. Softw.* **26** 1759–63
[38] Lin L I-K 1989 A concordance correlation coefficient to evaluate reproducibility *Biometrics* **45** 255
[39] Scott R L, Knowles J F, Nelson J A, Gentine P, Li X and Barron-Gafford G 2021 Water availability impacts on evapotranspiration partitioning *Agric. For. Meteorol.* **297** 108251
[40] Kobayashi K and Salam M U 2000 Comparing simulated and measured values using mean squared deviation and its components *Agron. J.* **92** 345
[41] Bacour C, Maignan F, Peylin P, MacBean N, Bastrikov V and Joiner J 2019 Differences between OCO-2 and GOME-2 SIF products from a model-data fusion perspective *J. Geophys. Res.* **124** 3143–57
[42] Baldocchi D, Sturtevant C and Contributors F 2015 Does day and night sampling reduce spurious correlation between canopy photosynthesis and ecosystem respiration? *Agric. For. Meteorol.* **15** 117–26
[43] MacBean N, Scott R L, Biederman J A, Ottlé C, Vuichard N, Ducharne A, Kolb T, Dore S, Litvak M and Moore D J P 2020 Testing water fluxes and storage from two hydrology configurations within the ORCHIDEE land surface model across US semi-arid sites *Hydrol. Earth Syst. Sci.* **24** 5203–30
[44] Cable D R 1969 Competition in the semidesert grass-shrub type as influenced by root systems, growth habits, and soil moisture extraction *Ecology* **50** 27–38
[45] Heitschmidt R K, Ansley R J, Dowhower S L, Jacoby P W and Price D L 1988 Some observations from the excavation of honey mesquite root systems *J. Range Manage.* **41** 227
[46] Clark M P, Fan Y, Lawrence D M, Adam J C, Bolster D and Gochis D J 2015 Improving the representation of hydrologic processes in Earth System Models *Water Resour. Res.* **51** 5929–56
[47] Novick K A *et al* 2016 The increasing importance of atmospheric demand for ecosystem water and carbon fluxes *Nat. Clim. Change* **6** 1023–7
[48] Barron-Gafford G A, Scott R L, Darrel Jenerette G, 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** 1389–400
[49] Smith S D, Monson R K and Anderson J E 1997 *Physiological Ecology of North American Desert Plants* (New York: Springer) p 288
[50] Dettweiler-Robinson E, Nuanez M and Litvak M E 2018 Biocrust contribution to ecosystem carbon fluxes varies along an elevational gradient *Ecosphere* **9** e02315
[51] Coe K K, Belnap J and Sparks J P 2012 Precipitation-driven carbon balance controls survivorship of desert biocrust mosses *Ecology* **93** 1626–36
[52] Wu D *et al* 2018 Asymmetric responses of primary productivity to altered precipitation simulated by ecosystem models across three long-term grassland sites *Biogeosciences* **15** 3421–37
[53] Sun Y *et al* 2021 Global evaluation of the nutrient-enabled version of the land surface model ORCHIDEE-CNP v1.2 (r5986) *Geosci. Model Dev.* **14** 1987–2010
[54] Paschalis A *et al* 2020 Rainfall manipulation experiments as simulated by terrestrial biosphere models: where do we stand? *Glob. Change Biol.* **26** 3336–55
[55] Schimel D S 2010 Drylands in the Earth system *Science* **327** 418–9