DR. ATHANASIOS PASCHALIS (Orcid ID : 0000-0003-4833-9962)

DR. MARTIN GERARD DE KAUWE (Orcid ID : 0000-0002-3399-9098)

DR. PAUL HANSON (Orcid ID : 0000-0001-7293-3561)

DR. WEI LI (Orcid ID : 0000-0003-2543-2558)

DR. YIQI LUO (Orcid ID : 0000-0002-4556-0218)

PROF. JOSEP PENUELAS (Orcid ID : 0000-0002-7215-0150)

DR. JULIA PONGRATZ (Orcid ID : 0000-0003-0372-3960)

DR. SARA VICCA (Orcid ID : 0000-0001-9812-5837)

MR. DONGHAI WU (Orcid ID : 0000-0002-4638-3743)

Article type : Primary Research Articles

# Rainfall-manipulation experiments as simulated by terrestrial biosphere models: where do we stand?

Athanasios Paschalis [1*], Simone Fatichi [2], Jakob Zscheischler [3,4], Philippe Ciais [5], Michael Bahn [6], Lena Boysen [7], Jinfeng Chang [5], Martin De Kauwe [8], Marc Estiarte [9,10], Daniel Goll [5,11], Paul J. Hanson [12], Anna B. Harper [13], Enqing Hou [14], Jaime Kigel [15], Alan K. Knapp [16], Klaus Steenberg Larsen [17], Wei Li [5,18], Sebastian Lienert [3,4], Yiqi Luo [14], Patrick Meir [19], Julia E. M. S. Nabel [7], Romà Ogaya [9,10], Anthony J Parolari [21], Changhui Peng [22], Josep Peñuelas [9,10], Julia Pongratz [23], Serge Rambal [9,10], Inger Kappel Schmidt [17], Hao Shi [24], Marcelo Sternberg

[25], Hanqin Tian [24], Elisabeth Tschumi [3,4], Anna Ukkola [8], Sara Vicca [26], Nicolas Viovy [5], Ying-Ping Wang [27], Zhuonan Wang [24], Karina Williams [28], Donghai Wu [29], Qiuan Zhu [30]

[1] *Department of Civil and Environmental Engineering, Imperial College London, UK*

[2] *Institute of Environmental Engineering, ETH Zurich, Switzerland*

[3] *Climate and Environmental Physics, University of Bern, Switzerland;*

[4] *Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland*

[5] *Laboratoire des Sciences du Climat et de l'Environnement, Gif sur Yvette, France*

[6] *Department of Ecology, University of Innsbruck, Austria*

[7] *Max Planck Institute for Meteorology, Hamburg, Germany*

[8] *ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, NSW, Australia*

[9] *CSIC, Global Ecology Unit CREAF-CSIC-UAB, 08193 Bellaterra, Catalonia, Spain*

[10] *CREAF, 08193 Cerdanyola del Vallès, Catalonia, Spain*

[11] *Department of Geography, University of Augsburg, Germany*

[12] *Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA*

[13] *Department of Mathematics, University of Exeter, Exeter, UK*

[14] *Department of Biological Sciences, Northern Arizona University*

[15] *Institute for Plant Sciences and Genetics, Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot Campus, Israel*

[16] *Graduate Degree Program in Ecology, Department of Biology, Colorado State University*

[17] *Dept. of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark*

[18] *Ministry of Education Key Laboratory for Earth System modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China*

[19] *Research School of Biology, Australian National University, Australia*

[20] *School of Geosciences, University of Edinburgh, United Kingdom*

[21] *Department of Civil, Construction, and Environmental Engineering, Marquette University, Milwaukee, WI, USA*

[22] *Department of Biology Sciences, University of Quebec at Montreal, Canada*

[23] *Department of Geography, Ludwig Maximilian University of Munich, Germany*

[24] *International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, USA.*

[25] *School of Plant Sciences and Food Security, Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel*

[26] *Centre of Excellence PLECO (Plants and Ecosystems),Biology Department, University of Antwerp, Belgium*

[27] *CSIRO Marine and Atmospheric Research and Centre for Australian Weather and Climate Research*

[28] *Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, Devon, UK*

[29] *Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University*

[30] *Center for Ecological Forecasting and Global Change, College of Forestry, Northwest A&F University, China*

[*] *Corresponding Author:a.paschalis@imperial.ac.uk, +44 (0)207 594 6004*

**Abstract**

Changes in rainfall amounts and patterns have been observed and are expected to continue in the near future with potentially significant ecological and societal consequences. Modelling vegetation responses to changes in rainfall is thus crucial to project water and carbon cycles in the future. In this study, we present the results of a new model-data intercomparison project, where we tested the ability of ten terrestrial biosphere models to reproduce observed sensitivity of ecosystem productivity to rainfall changes at ten sites across the globe, in nine of which, rainfall exclusion and/or irrigation experiments had been performed.

The key results are:

(a) Inter-model variation is generally large and model agreement varies with time scales. In severely water limited sites, models only agree on the interannual variability of evapotranspiration and to a smaller extent gross primary productivity. In more mesic sites model agreement for both water and carbon fluxes is typically higher on fine (daily-monthly) time scales and reduces on longer (seasonal-annual) scales.

(b) Models on average overestimate the relationship between ecosystem productivity and mean rainfall amounts across sites (in space) and have a low capacity in reproducing the temporal (interannual) sensitivity of vegetation productivity to annual rainfall at a given site, even though observation uncertainty is comparable to inter-model variability.

(c) Most models reproduced the sign of the observed patterns in productivity changes in rainfall manipulation experiments but had a low capacity in reproducing the observed magnitude of productivity changes. Models better reproduced the observed productivity responses due to rainfall exclusion than addition.

(d) All models attribute ecosystem productivity changes to the intensity of vegetation stress and peak leaf area, whereas the impact of the change in growing season length is negligible. The relative contribution of the peak leaf area and vegetation stress intensity was highly variable among models.

**Keywords**: drought, irrigation, terrestrial biosphere models, rainfall manipulation experiment

# 1 Introduction

Understanding the impact of rainfall changes on ecosystem functioning and vegetation dynamics is crucial for accurately predicting responses of vegetation structure, composition and dynamics under present or future conditions. Changes in both rainfall intensity and variability have been measured in the last decades (Trenberth, 2011; IPCC, 2013). Changes in precipitation extremes have also been observed (Alexander *et al.*, 2006) and according to climate model projections, such changes will intensify as we progress through the 21st century (IPCC, 2012; Knutti and Sedláček, 2013).

Changes in rainfall can affect energy and carbon fluxes at the land surface (Green *et al.*, 2017). Rainfall changes modify soil water dynamics, alter plant water status and consequently the terrestrial biogeochemical cycles (Heisler-White, *et al,.* 2008; Allan *et al.*, 2014) through changes in plant productivity or plant mortality (Allen, *et al.*, 2015). The importance of plant water limitation has been highlighted by the fact that semi-arid regions, which typically experience drought, control part of the global interannual variability of the terrestrial carbon sink (Ahlström *et al.*, 2015), with an increasing sensitivity during the last decades (Poulter *et al.*, 2014). The importance of water limitation on carbon fluxes will likely increase soon, since terrestrial vegetation is thought to operate close to its critical hydraulic thresholds across a wide range of ecosystems (Choat *et al.*, 2012), even though the full implications of this result are still debated (Klein *et al.*, 2014; Körner, 2019). As a direct consequence, minor changes in plant water availability worldwide can lead to significant impacts on the terrestrial carbon sink (Allen *et al.*, 2010; Zhao and Running, 2010; Reichstein *et al.*, 2013; Frank *et al.*, 2015; Humphrey *et al.*, 2018; Green *et al.*, 2019).

To understand the ecosystem responses to changes in rainfall amounts and patterns at the local scale, rainfall manipulations experiments have been conducted. Typically, such experiments change the overall rainfall amount by exclusion (Estiarte et al., 2016; Martin-Stpaul et al., 2013; Limousin et al., 2009) or irrigation (Collins et al., 2012) and responses are commonly quantified by changes in Aboveground Net Primary Production (ANPP). In some experiments such as the Amazon rainfall exclusion experiment, (Nepstad *et al.*, 2007) additional detailed data quantifying changes in forest structure and composition have been obtained. There are a small number of experiments where the structure of rainfall pulses is modified e.g. (Fay *et al.*, 2008; Heisler-White, *et al.,* 2008; Vicca *et al.*, 2014). Rainfall manipulation experiments have been conducted in a range of ecosystems, spanning from semiarid shrublands (Báez *et al.*, 2013), to temperate (Hanson and Wullschleger, 2003) and tropical forests (Fisher *et al.*, 2007; Nepstad *et al.*, 2007), even though most of the experiments have focused on grasslands or low-stature vegetation due to the difficulties in setting up experiments. Those experiments have identified a strong correlation between rainfall changes and vegetation

productivity (e.g. Heisler-White et al., 2009; Stuart-Haëntjens et al., 2018), phenology (e.g. Peñuelas et al., 2004), plant community structure e.g. (Miranda et al., 2011; Zhang et al., 2019) and belowground carbon dynamics e.g. (Vicca et al., 2014; Hagedorn et al., 2016; Hasibeder et al., 2015). Despite the important findings derived from these field experiments, these studies have strong spatial and temporal limitations; they reported only few variables and it is challenging to extrapolate information beyond the specific design of the experiment. Extrapolation and mechanistic understanding related to vegetation responses to changes in precipitation can be better achieved by combining model and data driven approaches (e.g. Kayler et al., 2015).

Modelling vegetation responses to changes in water availability is a challenging task (Xu *et al.*, 2013). Despite strong evidence that modelling responses to drought is a significant factor affecting terrestrial carbon dynamics (Trugman *et al.*, 2018), a commonly accepted parameterization of water limitation does not exist (Egea, *et al.,* 2011; Zhou *et al.*, 2013; Fatichi, *et al.,* 2016; Medlyn, *et al.,* 2016a; Hu *et al.*, 2018). Plant water stress simulated in terrestrial biosphere models can affect various processes but is commonly a function of either volumetric soil water content e.g. (Clark *et al.*, 2011) or soil water potential e.g. (Fatichi, *et al.,* 2012; Manzoni *et al.*, 2013; Lawrence *et al.,* 2019), integrated over the root zone. Examples of how water limitation affects plant functions include a decline in stomatal conductance affecting photosynthesis (Egea, *et al.,* 2011; Fatichi, *et al*., 2012; De Kauwe, *et al.*, 2015), changes in the photosynthetic parameters $V_{cmax}$ and $J_{max}$ e.g. (Krinner *et al.*, 2005), and/or accelerated senescence of plant tissues, especially leaves (Thurner *et al.*, 2017) leading to drought-induced deciduousness. Recently, significant efforts have been made to include more detailed plant hydraulics, to better describe water flow within the soil-plant-water continuum (Bonan *et al.*, 2014; Mirfenderesgi *et al.*, 2016; Eller *et al.*, 2018; Kennedy *et al.*, 2019; Lawrence *et al.,* 2019) and to include dynamics of non-structural carbohydrates to simulate consequences of water stress for carbon allocation and carbon starvation (reviewed in Fatichi *et al.*, 2019).

A large discrepancy of predicted model responses has direct consequences for the uncertainties related to the fate of terrestrial carbon under a changing climate (Zscheischler, *et al.*, 2014b; Ahlström *et al.*, 2015; Humphrey *et al.*, 2018). This is the case because the terrestrial vegetation and thus the terrestrial land carbon sink introduces the largest uncertainties of the global carbon cycle (Le Quéré *et al.*, 2018). In this context, large epistemic model uncertainties can have considerable impacts on our ability to forecast the growth rate of atmospheric $CO_2$. Additionally, vegetation responses to water stress can influence land-atmosphere coupling (Koster, 2004; Seneviratne *et al.*, 2013; Lemordant *et al.*, 2016; Gentine *et al.,* 2019), since vegetation cover and canopy conductance affect land surface energy balance. This will have a large impact on our skill to model

the coupled hydrological, plant physiological and meteorological processes and thus robustly projecting climate change (Miralles *et al.*, 2018).

To reduce this source of epistemic uncertainty and understand the reasons for model disagreement, a detailed comparison between the responses of different modelling schemes with respect to plant water availability is essential. Rainfall manipulation experiments assessing vegetation responses to water limitation are particularly useful in this regard. Arguably, this is an extremely important test to evaluate the structure and parameter values of a model and its capability to reproduce responses to environmental changes. A model should be able to reproduce the observed dynamics under control and manipulated conditions in order to be considered robust, especially for climate change simulations (Medlyn *et al.*, 2015). Despite the importance of this comparison, there are only few examples that have compared terrestrial biosphere models and global change manipulation experiments (De Kauwe *et al.*, 2013, 2017; Fatichi and Leuzinger, 2013; Powell *et al.*, 2013; Zaehle *et al.*, 2014; Medlyn *et al.*, 2015). Recently, Wu *et al.* (2018) compared 14 models under different idealized rainfall scenarios for three grassland experiments sites and showed a fair reproduction of spatial sensitivities of ANPP to rainfall but large differences in the modelled asymmetric response of ANPP to interannual i.e. temporal rainfall variability at a given site. Wu *et al.* (2018) were not able to evaluate the modelled responses with respect to actual experiments because they used idealized rainfall changes that did not exactly mimic the site treatments. In this study we perform such an evaluation. We make use of ten sites with diverse climates and biomes, where multi-year rainfall manipulation experiments took place to evaluate ten terrestrial biosphere models, representing an unprecedented data-model intercomparison effort focused on ecosystem responses to water limitation. This data-model intercomparison will address the following questions: (a) Can models reproduce the observed responses to precipitation variability at rainfall manipulation sites? (b) Do models accurately reproduce the spatial (across-sites) and temporal (within-site) dependence of vegetation productivity to precipitation? (c) Which are the underlying reasons for model disagreement? Answering those questions will provide insights on the robustness of Earth System model projections with respect to the global carbon cycle.

## 2 Data and Methods

### 2.1 Sites

Ten different sites with contrasted climates and biomes and sufficiently long records were considered here. For all analyses presented in this study, the sites are termed: Lahav, Matta, SGS, Prades, Garraf, Konza (AmeriFlux ID: US-Kon), Puèchabon (FluxNet ID: FR-Pue), Brandbjerg, Walker Branch (Walker Branch; AmeriFlux ID: US-WBW) and Stubai (Table 1). The sites are in ascending order in terms of wetness index $WI$

defined as the average ratio of annual precipitation to annual potential evapotranspiration during the study period. For our analysis the sites are split in three wetness categories ($WI < 0.4$ [Lahav, Matta, SGS]; $0.4 \leq WI < 1$ [Prades, Garraf, Konza, Puèchabon]; $WI \geq 1$ [Brandbjerg, Walker Branch, Stubai]).

The sites are in the USA (Konza, SGS, Walker Branch), Israel (Lahav, Matta), Spain (Garraf), France (Prades, Puèchabon), Austria (Stubai) and Denmark (Brandbjerg) and span a precipitation gradient from 253-1440 mm $y^{-1}$ and include grasslands shrublands and forested ecosystems (Table 1). In eight sites rainfall exclusion experiments were carried out, and in four irrigation experiments. The experiment duration considered in this study was from 5 up to 32 years. The average experiment duration was 13.3 years.

*Table 1: Site Description*

| Site | Lon/Lat | annual T [°C] | annual P [mm] | WI | Altitude [m] | Species | Soil Type | Drought Treatment | Irrigation Treatment | Years | Key References |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Lahav** | 34.9/31.38 | 19.1 | 253 | 0.19 | 590 | Annual grasses and shrubs, mostly Sarcopoterium spinosum | 22.6% Sand, 39.7% Silt, 37.7% Clay | -30% rainfall for the entire year | +30% rainfall for the entire year | 2002-2014 | Tielbörger *et al.*, 2014 |
| **Matta** | 35.07/31.71 | 17.94 | 498 | 0.33 | 620 | Similar to Lahav | 19% Sand, 29.2% Silt, 51.8% Clay | -30% rainfall for the entire year | +30% rainfall for the entire year | 2002-2014 | Tielbörger *et al.*, 2014 |
| **SGS** | -104.75/40.81 | 8.4 | 304 | 0.35 | 1650 | C4 grasses, primarily (Bouteloua gracilis (H.B.K.) Lag. Ex Steud., Buchloe dactyloides (Nutt) Engelm., mixed with varying amounts of C3 grasses, cactus, shrubs and forb. | 14% Sand, 58% Silt, 28% Clay | None | None | 1986-2009 | Heisler-White *et al.*, 2009 |
| **Prades** | 0.91/41.21 | 11.43 | 522 | 0.4 | 950 | Mixed composition of Quercus ilex L., Phillyrea latifolia L., Arbutus unedo L., Erica arborea L., Juniperus oxycedrus L., Cistus albidus L. Sorbus torminalis(L.) Crantz and | 48% Sand, 32% Silt and 20% Clay | -20% rainfall for the entire year | None | 1999-2012 | Ogaya and Peñuelas, 2007 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Acer monspessulanum L. | | | | | |
| **Garraf** | 1.82/ 41.3 | 15.04 | 580 | 0.48 | 210 | Erica multiflora, Globularia alypum | 41% Sand, 39% Silt and 18% Clay, | -50% in spring and fall | None | 2000-2004 | Beier *et al.*, 2009 |
| **Konza** | -96.6/ 39.1 | 12.8 | 830 | 0.7 | 342 | Mixed C3(Solidago canadensis, Aster ericoides, Salix missouriensis) C4(Andropogon gerardii, Sorghastrum nutans, Panicum virgatum) Grassland | 10% Sand, 35% Clay | None | irrigation +20% was provided at two sites termed lowland and upland | 1982-2013 | Collins *et al.*, 2012 |
| **Puèchabon** | 43.74/3.6 | 13.8 | 969 | 0.87 | 270 | Overstory (Quercus ilex); Understory (Buxus sempervirens, Phyllirea latifolia, Pistacia terebinthus and Juniperus oxycedrus) | 26% Sand, 35% Silt,39% Clay | -30% throughfall exclusion for the entire year | None | 2004-2013 | Limousin *et al.*, 2009 |
| **Brandbjerg** | 11.97/55.89 | 9.59 | 757 | 1.1 | 39 | 70% grasses (mostly Deschampsia flexuosa); 30% dwarf shrubs (Calluna vulgaris) | 88-95% Sand, 2-9% Silt, 1-2% Clay | rainfall exclusion for 4-6 weeks during spring and summer | None | 2007-2012 | Kongstad *et al.*, 2012 |
| | | | | | | | | | | | |
| **WB** | -84.29/35.96 | 14.7 | 1440 | 1.1 | 343 | Mixed composition of Quercus spp; Quercus prinus L., Quercus alba L., Quercus rubra L., Acer rubrum L., Acer saccharum, Liriodendron tulipifera L., Nyssa sylvatica Marsh. and Oxydendrum arboretum (L.) | 28% Sand, 60% Silt, 12% Clay | -30% throughfall exclusion for the entire year | +33% rainfall for the entire year | 1995-2005 | Hanson *et al.*, 2004 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Stubai** | 11.32/47.12 | 6.8 | 1382 | 1.7 | 970 | C3 Grassland (Agrostis capillaris, Festuca rubra, Ranunculus montanus, Trifolium pratense, Trifolium repens) | 42.2% Sand, 47% Silt, 10.8% Clay | rainfall exclusion for 8 weeks of summer rainfall | None | 2009-2013 | Fuchslueger *et al.*, 2014; Hasibeder *et al.*, 2015 |

125

126 For all sites, aboveground NPP estimates (ANPP) were recorded for most of the experimental years derived by
127 either biomass harvesting (grasslands) or biomass increase estimates derived from allometric relations and
128 simultaneous observations of stem diameter, leaf area changes, plus litterfall (e.g., shrublands and forests). Leaf
129 area index was quantified using the MODIS (MCD15A2H v006) estimate of the pixel containing each site.
130 MODIS data were interpreted with caution as they are an indirect measurement, valid at typically larger scales,
131 and prone to large uncertainties. For three sites, Konza, Puèchabon and WB, ET and gross primary productivity
132 (GPP) were obtained at the half hourly scale by the Fluxnet2015 database and aggregated to the daily scale.

133 **2.2 Participating models and simulation protocol**

134 For all sites, we conducted simulations using ten terrestrial biosphere models: CABLE r54482.0 (Wang *et al.*,
135 2011), DLEM v2.0 (Tian *et al.*, 2010), JULES v5.2 (Clark *et al.*, 2011), JSBACH v3.2 (Mauritsen *et al.*, 2019;
136 Kaminski *et al.*, 2013), LPX v1.4 (Lienert and Joos, 2018), ORCHIDEE rev5150 (Krinner *et al.*, 2005),
137 ORCHIDEE MICT rev5308 (Guimberteau *et al.,* 2018), ORCHIDEE CNP rev4520 (Goll *et al.*, 2017), T&C
138 v1.0 (Fatichi, *et al.,* 2012; Paschalis *et al.*, 2017) and TECO v2.0 (Huang *et al.*, 2017). All models include a
139 land surface scheme, a hydrological component, and a dynamic vegetation module. Soil moisture dynamics are
140 simulated in multiple vertical layers by either solving the 1D Richards equation or simplified hydrological
141 "bucket-type" models. Some models can simulate vegetation succession; however, this feature was disabled in
142 the current study. Five models included nutrient dynamics. CABLE, DLEM, JSBACH and LPX simulated
143 nitrogen and ORCHIDEE CNP nitrogen/phosphorus cycles. Hydrological and biogeochemical processes are
144 simulated with a variable degree of complexity (for a detailed model description see the supplementary
145 material of (Wu *et al.*, 2018)). As there is no commonly accepted way to simulate water limitation, each model
146 has adopted significantly different approaches (Medlyn et al., 2016b). Water stress in all models but T&C is a
147 function of an average root zone soil moisture and in T&C, water stress is a function of the integrated root zone
148 soil water potential. Specifically, models alter either photosynthetic rates (T&C, JULES, TECO), the maximum
149 rate of carboxylation $V_{cmax}$ (ORCHIDEE, ORC MICT, ORC CNP), stomatal conductance (JSBACH, DLEM),
150 or a combination of all such parameters (CABLE), based on plant water availability. LPX uses a supply and
151 demand driven approach to water limitation. If water demand exceeds supply, photosynthesis is downregulated

until they match. None of the models simulates plant hydraulics and thus xylem cavitation in response to water stress.

For each site, we conducted a control simulation corresponding to the observed climate without manipulation, and simulations representative of each rainfall manipulation experiment (rainfall exclusion and/or irrigation) with the same timing and magnitude of water input as in the real experiment. For all experiments the common data distributed to all modelling groups included hourly values of incoming shortwave and longwave radiation, vapour pressure deficit, air temperature, wind speed, atmospheric pressure and ambient $CO_2$ concentration. Model set-up was performed by each modelling group separately based on common information for each site that included, apart from the meteorological input, species composition, vegetation cover, soil and root depth and soil textural properties. Each modelling group translated independently this information into model specific parameters. Dependent on the model, species composition and vegetation cover were used to either choose between prescribed plant functional types (PFTs) or plant specific model parameters. Soil and root depth were used by all modelling groups to set-up the simulation domain, and the vertical discretization of the simulation was decided by each modelling group independently. Soil textural properties were used to select soil hydraulic properties. All information concerning the simulation set-up of each model and the common site properties provided to all modelling groups can be found at a free access data repository (see Data Sharing and Accessibility statement). Reported model outputs included gross primary productivity, net primary productivity and aboveground net primary productivity (GPP, NPP, ANPP), evapotranspiration (ET) and its partition in evaporation (soil evaporation plus evaporation from interception) and transpiration respectively, soil moisture, leaf area index (LAI) and biomass carbon pool (below and above ground) dynamics. Some models additionally reported the water stress factor ($\beta$) used in the model. $\beta$ is a model parameter that quantifies the effects of plant physiological stress due to limitations in soil water availability. $\beta$ is not identical between models and the description of the $\beta$ meaning for each model can be found at the supplementary material of Wu *et al.* (2018). Initial conditions for all simulations were obtained after a spin-up period long enough to equilibrate the biogeochemical cycles.

## 2.3 Statistical Analyses

### *Data-Model Comparison*

First, we compare the models' ability to accurately reproduce the relationship between ANPP and precipitation (P) across sites (i.e. spatial dependence) and within each site (i.e. temporal dependence) at the annual scale. At all sites, observations of ANPP were based on biomass estimates (e.g. using above ground biomass harvesting

for grasslands, and a carbon budget approach for forested sites combining litterfall observations with allometric equation for aboveground biomass growth) rather than carbon fluxes, therefore discrepancy between observed and modelled ANPP is expected (detailed bias quantification are reported in the Supplementary Material).

Model skill in reproducing the spatial dependence of ANPP to P was quantified as the root mean squared error (RMSE) and the coefficient of determination ($R^2$) between the modelled and observed annual ANPP, averaged over the entire period, across sites for the control case. Model performance in capturing the magnitude of interannual variability of ANPP was assessed by comparing the standard deviation ($\sigma$) of annual ANPP between models and observations for all sites. Model skill with respect to single-site interannual dependence of ANPP to P was quantified using an estimate of the sensitivity of annual ANPP to annual P. Specifically, we fitted a linear model $ANPP = a_0 + a_1 P + a_2 T$, where $P$ is annual precipitation and $T$ annual temperature. To increase the sample size and robustness of the fit, precipitation from both the control and the rainfall manipulation experiments were used. Additional covariates such as vapour pressure deficit and radiation could not be added due to the small sample size, making the linear fit over constrained. Preliminary analyses (not reported here) showed that $P$ and $T$ were the most important covariates. Model skill was evaluated by estimating the differences between observed and simulated sensitivities of ANPP with respect to $P$ (i.e. $a_1 = \frac{\partial ANPP}{\partial P}$). Observation uncertainty of the sensitivity metric was quantified as the 90% confidence interval of the linear model fit.

For the control simulations, modelled ET and GPP were compared with eddy covariance high frequency observations from Walker Branch, Puèchabon and Konza. In these three locations, flux-tower data were available in the proximity and with the same vegetation cover as the rainfall exclusion/addition experiment. Comparison at the daily scale was performed by means of Taylor diagrams (Taylor, 2001). The magnitude and seasonal pattern of the fluxes were also analysed (Supplementary material Figures S2-S4).

Responses due to rainfall manipulation were quantified at the annual scale using the response ratio for a variable $X$ (e.g. ANPP) defined as the ratio $RR = X_M^{(y)}/X_C^{(y)}$, where the subscript $_M$ stands for manipulation and $_C$ for the control scenario. $(y)$ indicates the annual scale. In this study, we focused on the simulated RRs of ANPP and ecosystem water use efficiency (WUE) calculated at the annual scale as the ratio of annual gross primary productivity (GPP) to annual actual evapotranspiration (ET). To quantify whether the simulated response ratios have a statistically significant different mean value from the observations, a two-sample t-test was performed for every model and the respective observed responses. For the two-sample t-test, the sample size for each site is equal to the number of years in the observations and simulations. Response ratios were

assumed normally distributed and independent at the annual scale. The test's null hypothesis was that modelled and observed response ratios have the same mean. The analysis was also performed using the commonly used logarithm of $RR$ yielding identical results, and thus not further shown here.

***Model agreement***

Model agreement across time scales was quantified by estimating the Pearson correlation coefficient ($\rho$) between all pairs of models for ET and GPP at the daily, monthly and annual scale. In the supporting material (Figure S7), the analysis is expanded for a wider range of scales by estimating the wavelet coherence between all pairs of models for ET and GPP.

To quantify agreement with respect to modelled changes in ANPP and WUE due to rainfall alterations, a two-sample t-test for the response ratios of both ANPP for all model pairs was performed and presented in the Supplementary material (Tables S2-S3).

To attribute the variability of ANPP to its causes we proceeded similarly to De Kauwe *et al.* (2017) who found that the annual ANPP could be approximated by the product

$$ANPP = A_b \cdot CUE \cdot GPP_u \cdot \beta \cdot LAI_p \cdot LAI_r.$$

The term $A_b$ is the aboveground fraction of carbon allocation, $CUE$ is the carbon use efficiency, $GPP_u$ is a potential (unstressed) rate of GPP per unit of leaf area, $\beta$ is the annually averaged value of the water stress factor, $LAI_p$ is the peak $LAI$ during the year, and $LAI_r$ is a proxy of the growing season length, defined as the integral of $LAI$ during the year divided by $LAI_p$. Considering that water stress and LAI dynamics, determine most of the interannual variation of ANPP, assuming that $A_b$, $CUE$, and $GPP_u$ vary less between treatments, then, the annual response ratio of ANPP can be estimated by the response ratios of $\beta$, $LAI_p$ and $LAI_r$ (e.g. $\frac{ANPP_M^{(y)}}{ANPP_C^{(y)}} \approx \frac{\beta_M^{(y)}}{\beta_C^{(y)}} \cdot \frac{LAI_{p_M}^{(y)}}{LAI_{p_C}^{(y)}} \cdot \frac{LAI_{r_M}^{(y)}}{LAI_{r_C}^{(y)}}$, where the subscript $_M$ stands for manipulation and $_C$ for the control scenario and $(y)$ indicates the annual scale). If the response ratios of $\beta$ and $LAI_p$ and $LAI_r$ are independent at the annual scale, then

$$\overline{\left(\frac{ANPP_M^{(y)}}{ANPP_C^{(y)}}\right)} \approx \overline{\left(\frac{\beta_M^{(y)}}{\beta_C^{(y)}}\right)} \cdot \overline{\left(\frac{LAI_{p_M}^{(y)}}{LAI_{p_C}^{(y)}}\right)} \cdot \overline{\left(\frac{LAI_{r_M}^{(y)}}{LAI_{r_C}^{(y)}}\right)},$$

where overbars indicate average values for all years. This approximation is well supported by the results of our simulations (Supplementary material, Figure S6), even though data evidence suggests that $CUE$ may change

significantly under changes in water stress (Rowland et al., 2014). Using this decomposition in the model results, the average ANPP response ratio can be decomposed as the product of the average response ratios of $\beta$,$LAI_r$,$LAI_p$. Based on these considerations, we can attribute the changes of the modelled $ANPP$ among models to differences in simulated water stress, LAI dynamics, and phenological changes. Since only six (T&C, CABLE, JULES, TECO, DLEM, JSBACH) of the ten participating models reported the water stress $\beta$ factor, this analysis was performed using this subset of models. All statistical analyses were performed in MATLAB 2019a.

# 3 Results

## 3.1 Control Scenario

Models captured the increasing trend of observed average ANPP to average P across sites (Figure 1a). The RMSE between simulated and observed ANPP was in the range 23-354 g $C$ m$^{-2}$ y$^{-1}$. Normalized RMSE of ANPP was weakly but positively correlated ($R^2 = 0.36, p - value = 0.067$) with the RMSE of normalized LAI (i.e. LAI divided by its maximum value). All models were positively biased. Positive biases can be partially attributed to model shortcomings but can be also explained by experimental underestimations in ANPP measurements (see Figure S1). Relative absolute biases (i.e.$|\text{relBias}| = \frac{|ANPP_{Mod} - ANPP_{obs}|}{ANPP_{obs}}$) are typically larger at the driest sites ($\frac{\partial |relBias|}{\partial P} = -6.3 \cdot 10^{-4}$ mm$^{-1}$, estimated using ordinary least squares method).

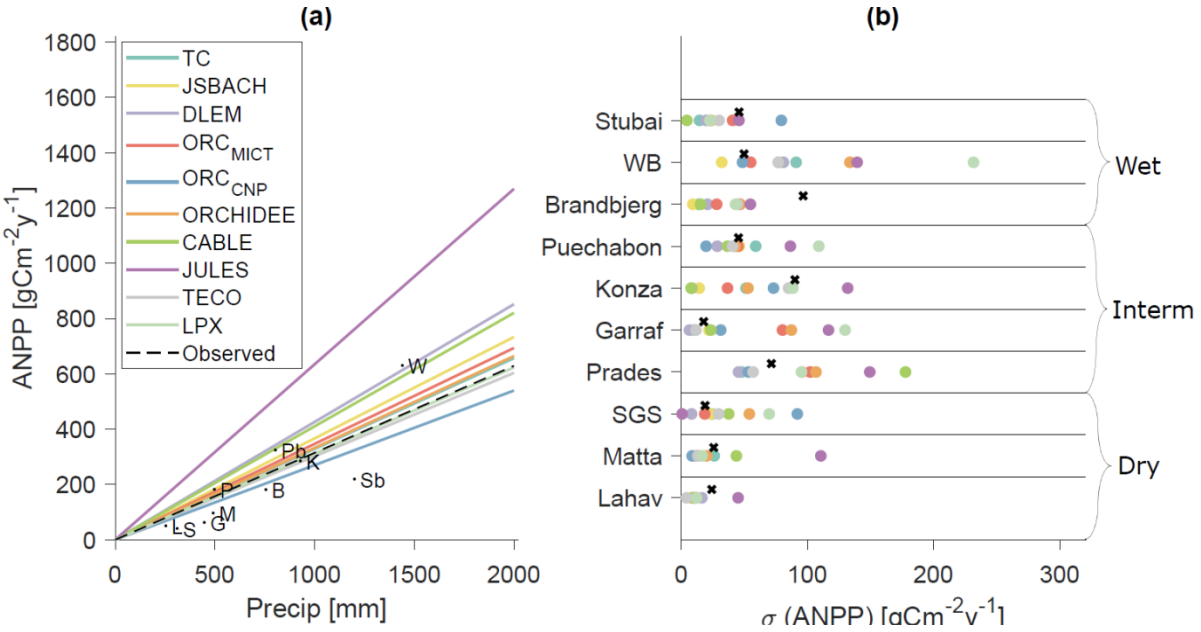


*Figure 1: (a) Dependence of mean annual ANPP to average annual precipitation during the study period. Letters indicate observed values (L: Lahav, M:Matta, S:SGS, P: Prades, G:Garraf, K: Konza, Pb: Puèchabon, B:Brandbjerg, W:WB, Sb: Stubai). Lines indicate, for each model, a least square fit of a linear relationship:* $ANPP(P) = \alpha P$ *between the modelled mean annual ANPP and mean annual Precipitation for all sites. (b) Standard deviation of modelled annual ANPP (circles) and observed annual ANPP (crosses) for all sites and models. Each model has a unique color indicated in the legend.*

*Table 2: Model skill across sites in terms of root mean square error (RMSE) for annual ANPP, normalized root mean square error (NRMSE) for annual ANPP, coefficient of determination for annual ANPP, average bias of ANPP, average bias of the standard deviation of annual ANPP, RMSE for daily LAI and RMSE for daily normalized LAI (i.e.* $\frac{LAI}{max\,(LAI)}$*).*

| Model | ANPP - RMSE ($gCm^{-2}y^{-1}$) | ANPP – Normalized RMSE [-] | ANPP - $R^2$ [-] | ANPP - Bias ($gCm^{-2}y^{-1}$) | σ(ANPP) - Bias ($gCm^{-2}y^{-1}$) | LAI - RMSE [$m^2m^2$] | LAI normalized RMSE [-] |
|---|---|---|---|---|---|---|---|
| TC | 76.318 | 0.368 | 0.8295 | 30.7907 | -13.5738 | 1.2399 | 0.2956 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| JSBACH | 233.0982 | 1.1239 | 0.2379 | 79.3713 | -19.6096 | 1.2972 | 0.4276 |
| DLEM | 202.8963 | 0.9783 | 0.7732 | 96.935 | -23.7873 | 1.2038 | 0.356 |
| ORC MICT | 121.7962 | 0.5872 | 0.6131 | 51.5792 | -5.7495 | 1.1895 | 0.3966 |
| ORC CNP | 210.5444 | 1.0151 | 0.041 | 15.0756 | -1.0366 | 1.1451 | 0.4198 |
| ORCHIDEE | 113.8664 | 0.549 | 0.6489 | 44.8288 | 9.3944 | 1.2675 | 0.3505 |
| CABLE | 215.6812 | 1.0399 | 0.4728 | 115.9473 | -5.1951 | 2.147 | 0.3437 |
| JULES | 354.0429 | 1.707 | 0.4399 | 278.4353 | 39.4962 | 1.4164 | 0.449 |
| TECO | 23.3013 | 0.1123 | 0.982 | 5.3858 | -9.3174 | 1.1347 | 0.3462 |
| LPX | 113.6602 | 0.548 | 0.5956 | 36.4618 | 33.2501 | 1.3886 | 0.4317 |

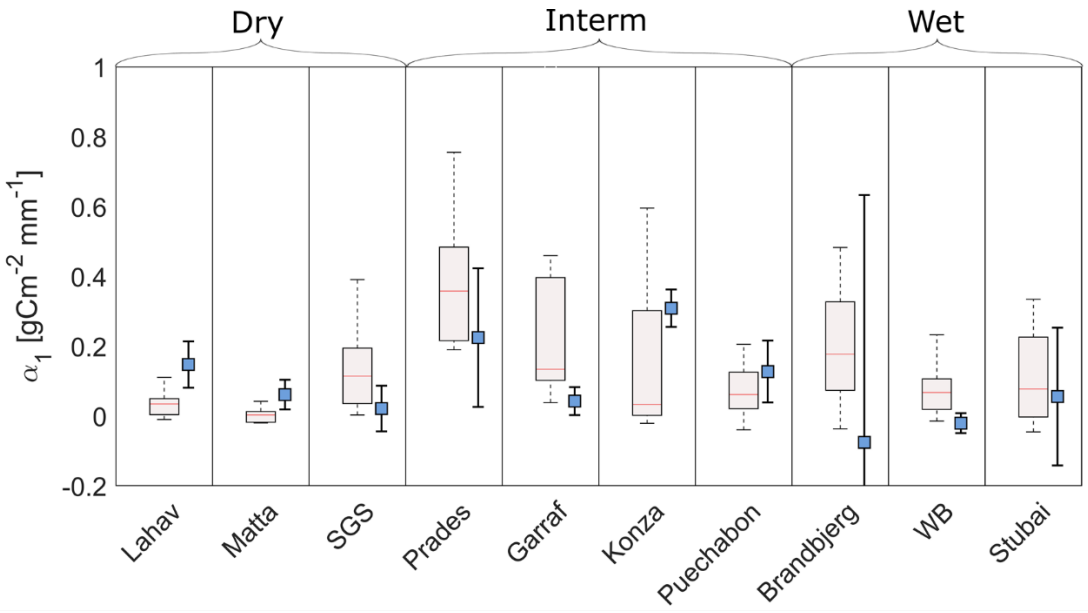


*Figure 2: Simulated and observed sensitivity of annual ANPP to annual precipitation ($\alpha_1 = \frac{\partial ANPP}{\partial P}$). For each site, boxplots indicate the distribution of the simulated sensitivity of ANPP to precipitation by all models. Error bars show the sensitivity of observed ANPP to precipitation (blue squares) and the corresponding 90% confidence intervals (bar length) of the fit of the linear model. Crosses indicate the sites for which the mean value of the distribution of simulated sensitivities is not statistically different from the observed with 90% confidence. Sites are ranked from left to right in order of ascending wetness.*

Both models and observations support a larger sensitivity of annual ANPP to interannual variation in precipitation at sites with intermediate wetness conditions (e.g. Garraf, Prades, Puèchabon, Konza; Figure 2).

Specifically, in sites with a wetness index WI $< 0.4$ models(observations) have mean sensitivity $\overline{a_1} = 0.058$ $(0.076)$ g$C$m$^{-2}$mm$^{-1}$, in sites with $0.4 \leq WI < 1$, $\overline{a_1} = 0.22(0.18)$ g$C$m$^{-2}$mm$^{-1}$ and in sites with $WI > 1$, $\overline{a_1} = 0.13(-0.013)$ g$C$m$^{-2}$mm$^{-1}$. At the most arid sites, annual precipitation explains a large fraction of the observed and modelled variability of annual ANPP, but the sites are not highly productive (i.e. Absolute productivity values are low; Figure 1), yielding a low average sensitivity $a_1$. At the opposite end, mesic sites have higher productivity, but they are not water limited during the observation period, resulting also in a low modelled sensitivity $\overline{a_1}$. Modelled sensitivity uncertainty was largest for intermediate precipitation regimes due to a larger model disagreement for those sites. For sites with a WI $< 0.4$, the average uncertainty, quantified here as the standard deviation between models of modelled $a_1$ was $\sigma_{a_1|dry} = 0.08$ g$C$m$^{-2}$mm$^{-1}$, for intermediate sites $\sigma_{a_1|inter} = 0.24$ g$C$m$^{-2}$mm$^{-1}$ and for wet sites $\sigma_{a_1|wet} = 0.14$ g$C$m$^{-2}$mm$^{-1}$.

On average, the modelled sensitivity of ANPP to precipitation within sites was lower ( ~0.15 g$C$m$^{-2}$mm$^{-1}$) than (~0.37 g$C$m$^{-2}$mm$^{-1}$; estimated as the average slope of the linear models reported in Figure 1a) between sites. However, the uncertainty of the estimated temporal sensitivity from observations, as quantified by the 90% confidence limits of the linear model, is very high in most sites (0.29 g$C$m$^{-2}$mm$^{-1}$, averaged across all sites) and comparable to the uncertainty between models ($\overline{\sigma_{a_1}} = \acute{\sigma}_{a_1} = 0.14$ g$C$m$^{-2}$mm$^{-1}$, averaged across all sites). A large uncertainty is related to either a small sample size, or low skill of the linear model. As a result, it is not possible to robustly quantify whether the modelled temporal sensitivities are statistically different from the observed ones, but overall only six out of ten sites had mean modelled that were not non-statistically scientifically different than the one observed (Figure 2).

Simulated daily ET for the control simulations was substantially different regarding its day-to-day variability from measured ET at all three eddy sites (Konza, Puèchabon, WB). Correlation coefficients were in the range 0.27-0.78 with an average value between all models and sites of ~0.60±0.13 (mean $\pm$ standard deviation) (Figure 3). Simulated variability of ET, expressed in terms of standard deviation at the daily scale, deviated substantially from the measured variability of ET. In particular, simulated variability from most models was lower than observed at Konza (observed $\sigma_{ET}$= 1.76 mm d$^{-1}$ , modelled $\sigma_{ET} = 1.40 \pm 0.3$ mm d$^{-1}$), and higher than observed at Puèchabon (observed $\sigma_{ET}$= 0.61 mm d$^{-1}$, modelled $\sigma_{ET} = 1.86 \pm 0.50$ mm d$^{-1}$). For WB, the modelled ET variability was higher than observed, and inter-model agreement was low (observed $\sigma_{ET}$ = 1.39 mm d$^{-1}$ , modelled $\sigma_{ET} = 1.51 \pm 0.45$ mm d$^{-1}$). Seasonality of ET was well reproduced by all models (Figure S2), partially explaining the high correlation coefficients (Figure 3). One pronounced exception is in

Puèchabon, where the observed late summer reduction of ET and increase in early fall was reproduced only by a small subset of models (Figure S2).

Simulated daily GPP had a correlation (~0.59 ±0.17) with observed daily GPP for all models (Figure 3). A large fraction of the GPP correlation can be attributed to seasonality. However, the modelled variability was significantly different from the observed for all sites. Most models underestimated the daily variation of GPP at Konza (observed $\sigma_{GPP} = 4.04\ \mathrm{gC m^{-2} d^{-1}}$, modelled $\sigma_{GPP} = 2.87 \pm 1.88.\ \mathrm{gC m^{-2} d^{-1}}$) and WB (observed $\sigma_{GPP} = 4.53\ \mathrm{gC m^{-2} d^{-1}}$, modelled $\sigma_{GPP} = 4.01 \pm 1.26\ \mathrm{gC m^{-2} d^{-1}}$) and overestimated the variability of daily GPP at Puèchabon (observed $\sigma_{GPP} = 1.68\ \mathrm{gC m^{-2} d^{-1}}$, modelled $\sigma_{GPP} = 2.67 \pm 1.01\ \mathrm{gC m^{-2} d^{-1}}$) (Figure 3). Large model differences between observed and simulated GPP can be partially attributed to an incorrect representation of the magnitude of LAI. There is, indeed, a large disagreement between the modelled LAI across models (Figure 4). Modelled LAI is also significantly different than observed, even though LAI derived via remote sensing is also uncertain (Fang *et al.*, 2013).

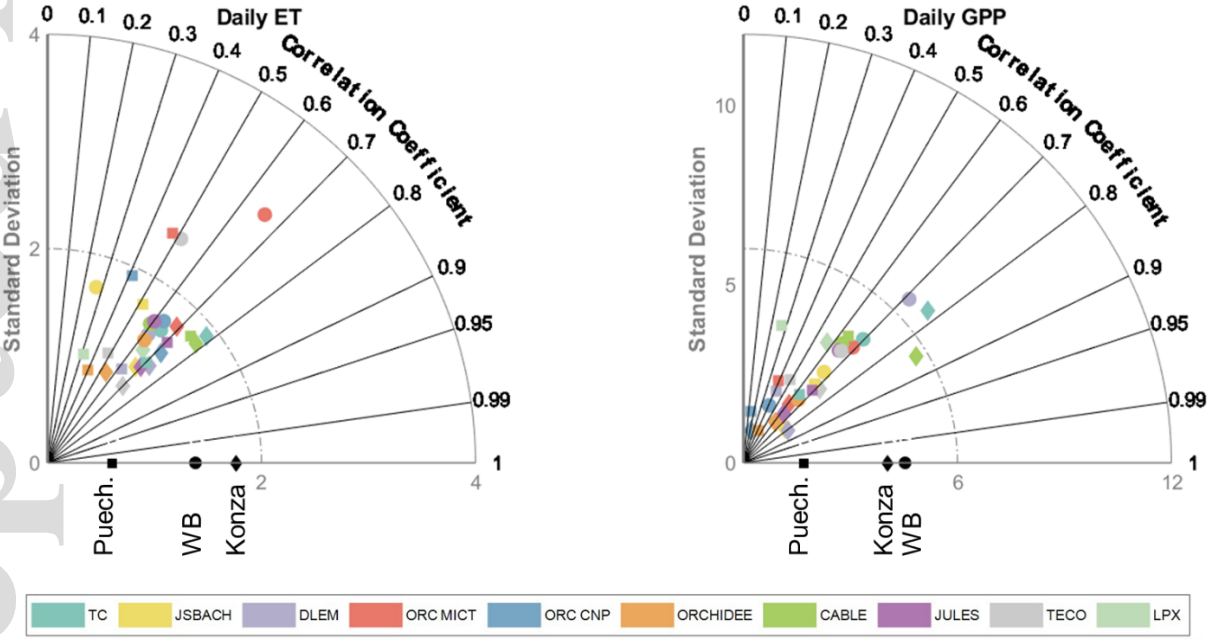


*Figure 3: Taylor Diagrams for daily evapotranspiration (ET) and gross primary productivity (GPP) for all models and all sites with available flux tower data. Models are indicated with different colors according to the legend. Each site has a different marker (diamond for Konza, circle for WB and square for Puèchabon). The ideal model (i.e. reproducing precisely the data) would lie on the black markers, each corresponding to different sites.*

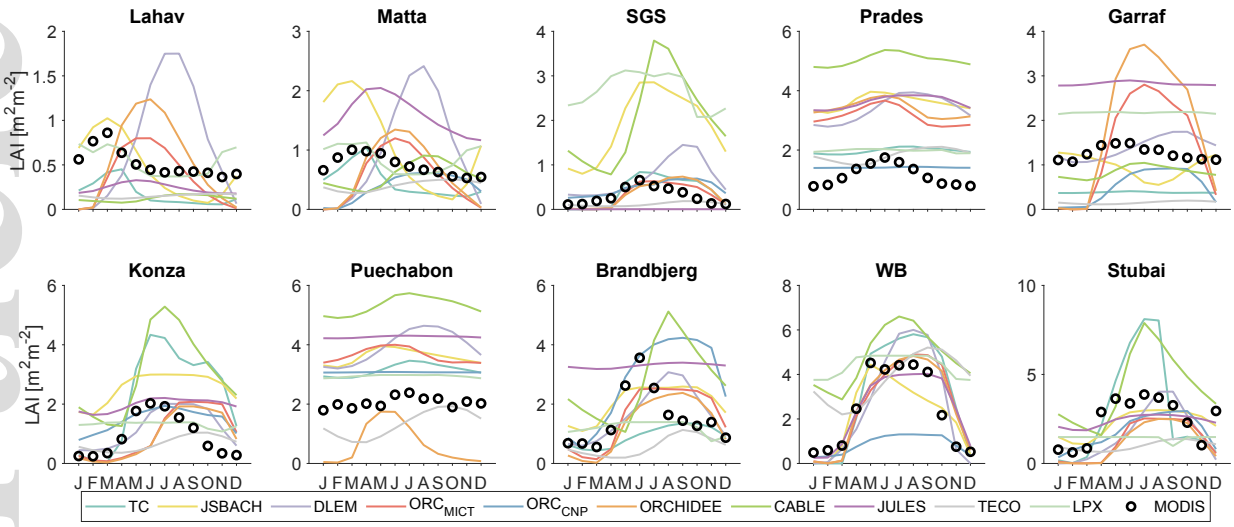


*Figure 4: Simulated average monthly LAI by all models for all sites for the control case simulation. Dots indicate the long-term monthly LAI averages of the nearest MODIS pixel in the area.*

Model agreement in terms of ET and GPP varies also with time scale (Figure 5). In the driest sites (e.g. Lahav, Matta, SGS; WI $< 0.4$), models agree mostly with each other on the interannual variability of ET (average corr. coef. $\rho$ for ET at the annual ($^{y}$) scale $\rho^{y}_{ET|dry} = 0.75$; for GPP $\rho^{y}_{GPP|dry} = 0.35$ ). This is expected since at those sites annual ET almost equals the total amount of rainfall. However, a significant model disagreement occurs at the daily ($^{d}$) scale ($\rho^{d}_{ET|dry} = 0.58$, $\rho^{d}_{GPP|dry} = 0.30$). The opposite picture occurs in mesic sites (WI>1), where models agree better at the daily time scale for ET ($\rho^{d}_{ET|wet} = 0.79$), but their agreement is significantly lower at the annual scale ($\rho^{y}_{ET|wet} = 0.61$). A similar pattern is also valid for GPP ($\rho^{d}_{GPP|wet} = 0.77$; $\rho^{y}_{GPP|wet} = 0.60$) (Figure 5).

Model agreement with regards to the dependence of the water stress factor $\beta$ on root averaged soil moisture $\theta($ $Z_r$) is also low (Figure 6). On average model agreement was highest for sites with a large percentage of sand (Brandbjerg 88-95% sand, Prades 48% sand) and lowest in sites with soils rich in more fine material (e.g. Lahav 22% sand, Matta 19% sand, SGS 14% sand, Konza 10% sand).

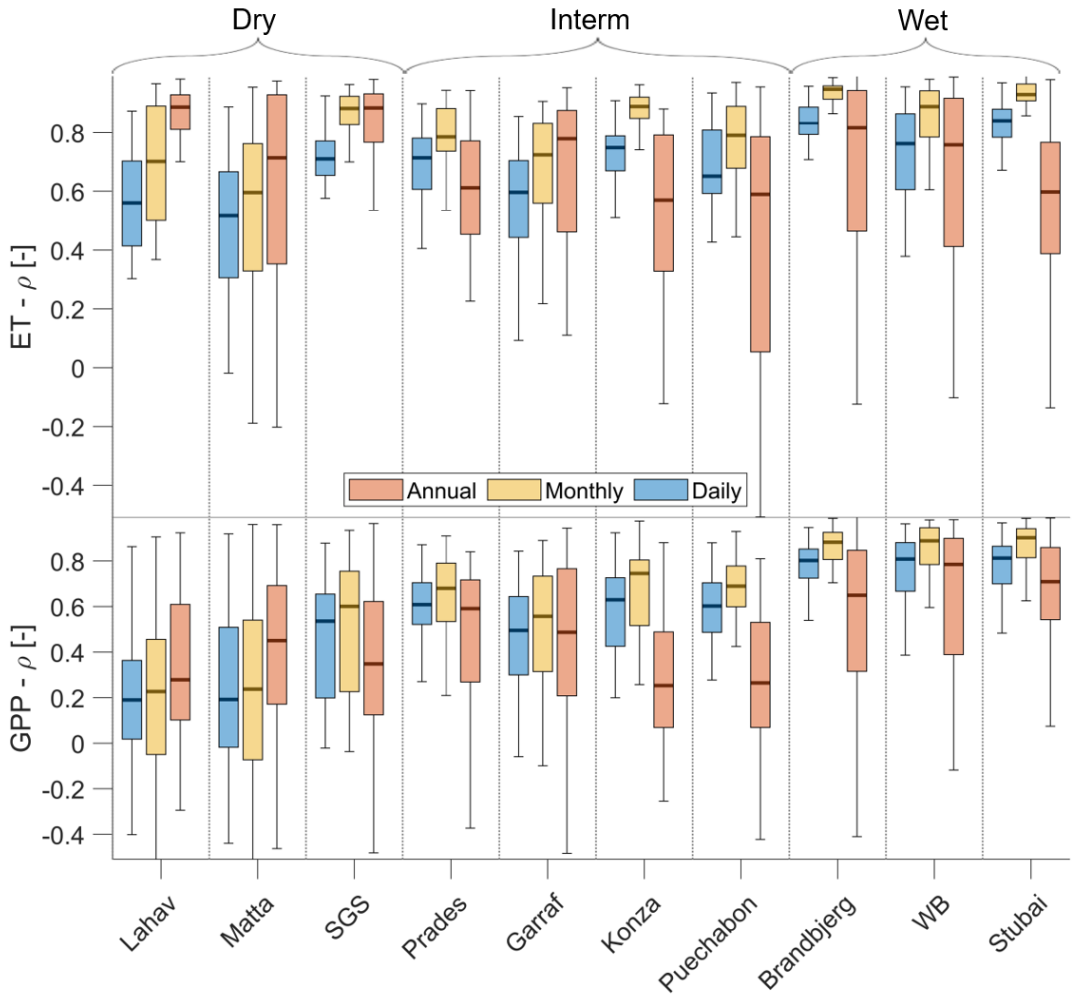


*Figure 5: Boxplots of Pearson correlation coefficients between simulated ET and GPP for all pairs of models for three time scales (daily, monthly, and annual) for all ten sites. Scales are indicated with different colors according to the legend.*

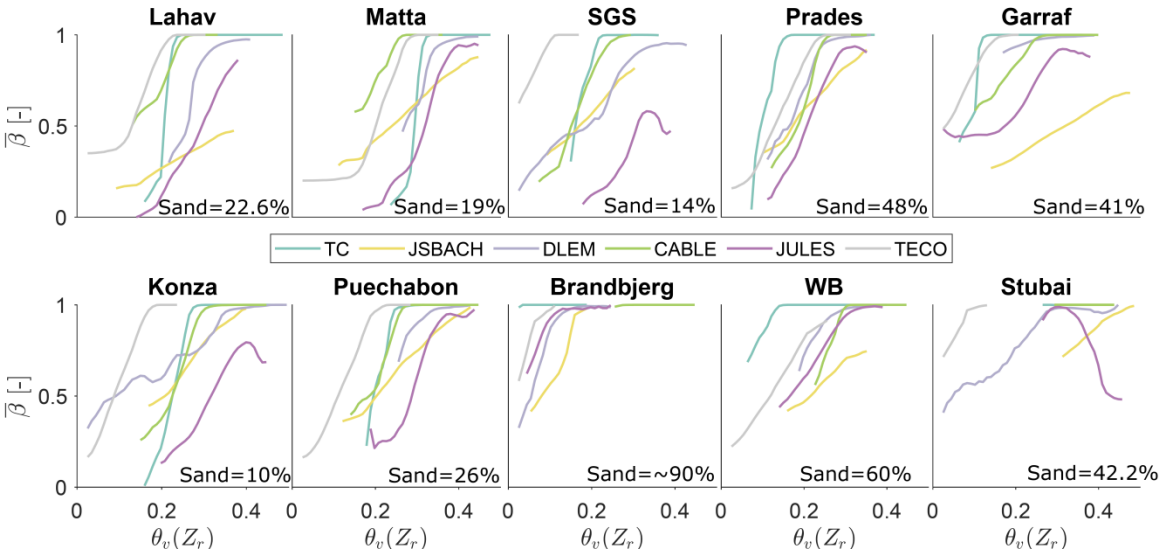


*Figure 6: Average simulated water stress factor β as a function of root zone averaged soil moisture. For all sites and models $\overline{\beta}$corresponds to the simulated average value of β at the daily scale for overlapping bins with soil moisture width 0.05.*

### 3.2 Manipulation Experiments

Models were tested for their skill at reproducing changes in ANPP due to rainfall manipulations (Figure 6). Most models (75% for model-site-treatment combinations) correctly predicted the sign of the change in ANPP. However only 54% of the models for the drought treatment (10 models × 8 sites) and 43% for the irrigation treatment (10 models × 4 sites) have a mean response that is statistically similar in magnitude with the observed, highlighting a better model performance for rainfall exclusion than addition. The worst performance of the models was obtained for both the drought and irrigation experiments in Lahav and in the irrigation experiment in Konza where almost no model was able to capture the correct magnitude of the response ratio.

Even though observed ANPP estimated from biomass should be close to modelled ANPP (Figure S1) several uncertainties related to observations, such as the choice of biomass harvest date, the use of specific allometric equations, and specific local conditions could affect our results. For instance, the observed response to irrigation in Lahav and Matta is considerably different despite the two sites having similar vegetation and climate. Those differences are either due to measurement uncertainties, or due the large effect of some local properties (e.g. soil composition, nutrient availability (Golodets *et al.*, 2013, 2015)) causing significant changes in the ecosystem dynamics. Overall, the magnitude of responses is similar amongst models except CABLE, JULES and TECO, which show a larger sensitivity of ANPP to rainfall manipulation. Modelled interannual

variability of the responses was in most cases similar in magnitude to the observed for the rainfall exclusion experiments, and lower for the irrigation experiments (for the drought experiments: average modelled standard deviation of the response ratios was $\acute{\sigma}_{RD}^{m} = 0.18$; and observed $\acute{\sigma}_{RD}^{o} = 0.178$. For irrigation experiments modelled standard deviation was $\acute{\sigma}_{RI}^{m} = 0.25$; and observed $\acute{\sigma}_{RI}^{o} = 0.42$). Outliers with regards to both the magnitude and the interannual variability of response ratios occurred for the most water-limited sites.

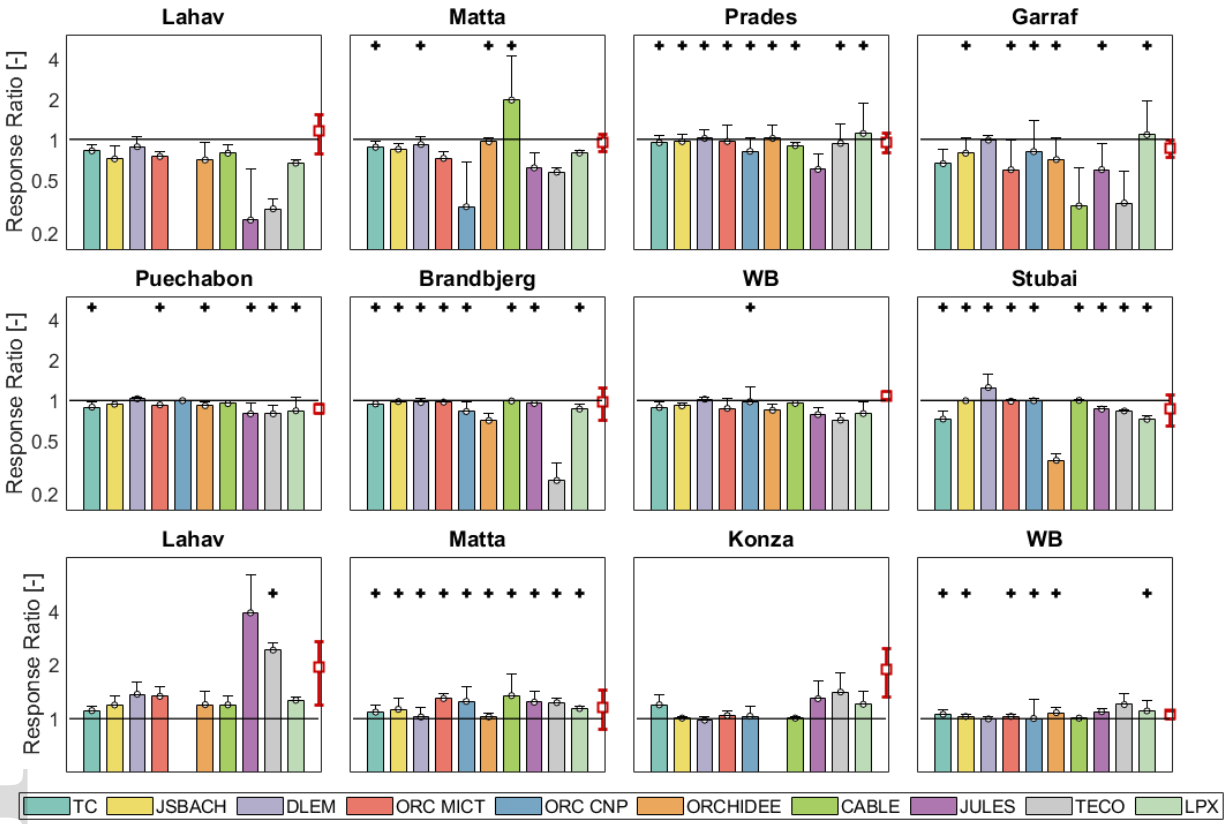


*Figure 7: Simulated and observed response ratios of annual ANPP due to rainfall exclusion (rows 1 and 2) and addition (irrigation) (row 3). Different models are presented with different colors according to the legend. Error bars represent the standard deviation for all years of treatment. Red error bars represent measured response ratios. Black crosses indicate models where the null hypothesis of the same mean between simulated and observed response ratios is not rejected based on a two sample t-test. Missing bars relate to spurious model output due to loss of vegetation survival.*

Besides carbon assimilation, changes in rainfall can simultaneously modify ET and thus the land surface energy balance. The coupling between ET and GPP depends heavily on the parametrizations of water stress and how this affects stomatal conductance and the reduction of photosynthesis. It further depends on vegetation dynamics such as a shift of carbon allocation from leaves to roots or leaf shedding due to water stress. To quantify the responses of the ET and GPP coupling, we compute the relative changes of water use efficiency (WUE) for the various cases (Figure 8). Most models predict relatively small changes in WUE (i.e. R ~ 1) for both drought ($R_D^{'m} = 0.98$) and irrigation ($R_I^{'m} = 1.08$) treatments, indicating a change of comparable magnitudes for both ET and GPP. CABLE, JULES and TECO occasionally simulate larger changes, in both positive and negative directions, in WUE for the most water limited sites. This larger change can be attributed to a more sensitive response of GPP to water stress than ET.

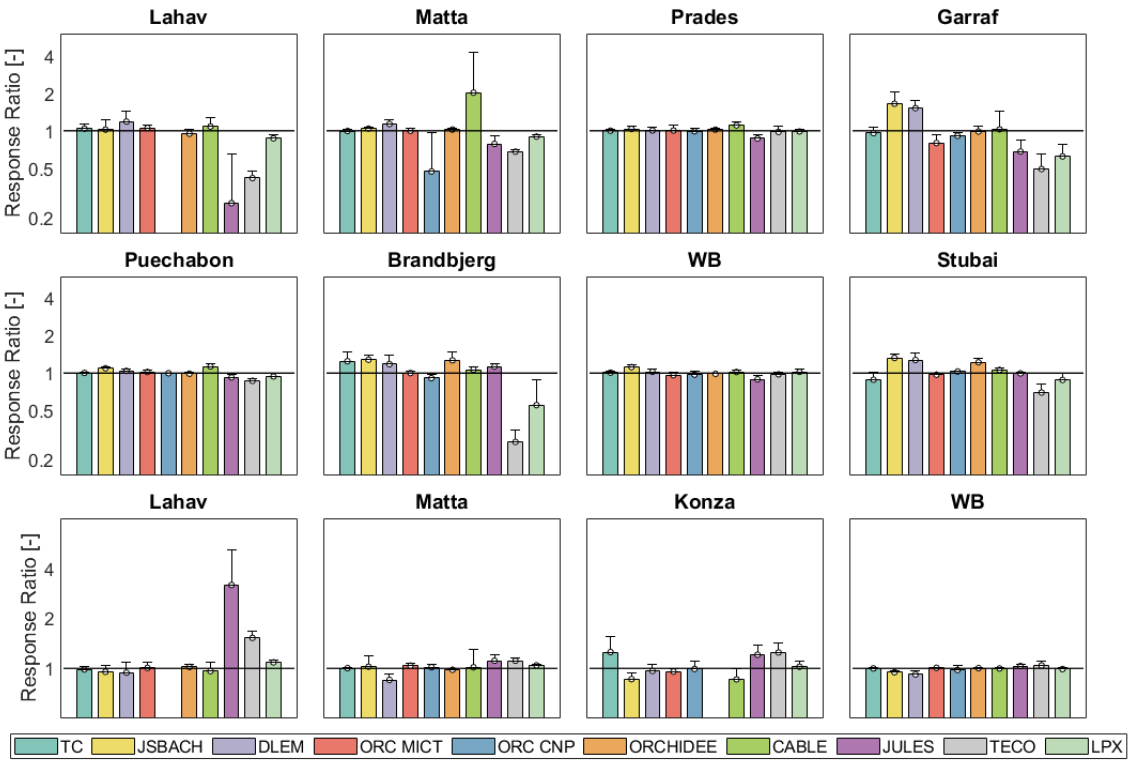


*Figure 8: Simulated response ratios of water use efficiency during treatment period per year due to rainfall exclusion (rows 1 and 2) and addition (irrigation) (row 3). Different models*

*are presented with different colors according to the legend. Error bars represent the standard deviation for all years of treatment.*

### 3.3 Response attribution

We partitioned the total response ratio of ANPP into relative changes of (a) the $\beta$ stress factor; (b) peak LAI ( $LAI_p$); and (c) the length of the growing season approximated by $LAI_r$ (Figure 9). Changes in simulated ANPP following rainfall manipulation can be almost exclusively attributed to changes in $\beta$ and $LAI_p$. The response ratio of $LAI_r$ was always close to unity ($R_{LAI_r} = 0.98 \pm 0.058$ (mean $\pm$ standard deviation) for the drought treatment and $R_{LAI_r} = 1.01 \pm 0.029$ for the irrigation treatment) contributing insignificantly to the response ratio of ANPP. Thus, no model predicted substantial changes in the length of the growing season. A reduction or enhancement of $\beta$ for the drought and irrigation experiments explained the largest fraction of ANPP responses at wet sites, but the uncertainty of the relative strengths of changes in $\beta$ and $LAI_p$ was high (Drought treatment for sites with $WI > 1$, $R_\beta = 0.95 \pm 0.08$, $R_{LAI_p} = 0.91 \pm 0.18$; Irrigation treatment for sites with $WI > 1$, $R_\beta = 1.05 \pm 0.06$,$R_{LAI_p} = 1.02 \pm 0.02$). For the driest sites both $LAI_p$ and $\beta$ explained a large fraction of the total response for the drought treatment, whereas $LAI_p$ was the dominant and simultaneously the most uncertain factor for the irrigation treatment (Drought treatment for sites with $WI < 0.4$, $R_\beta = 0.87 \pm 0.10$, $R_{LAI_p} = 0.77 \pm 0.24$; Irrigation treatment for dry sites with $WI < 0.4$, $R_\beta = 1.06 \pm 0.10$, $R_{LAI_p} = 1.49 \pm 0.86$). Differences in the simulated responses of both $\beta$ and $LAI_p$ amongst models was high as indicated by the standard deviations above. At the sites where rainfall exclusion was applied only in part of the year (Garraf, Brandbjerg) the response ratio of $LAI_p$ was larger than the reduction of $\beta$ ($R_\beta = 0.93 \pm 0.09$, $R_{LAI_p} = 0.78 \pm 0.27$), but given the large variability amongst models, it is not possible to conclude if this is a true signal. The variability was higher for the most water stressed sites, primarily because for those sites model disagreement on the estimated response ratio of ANPP was also the highest.

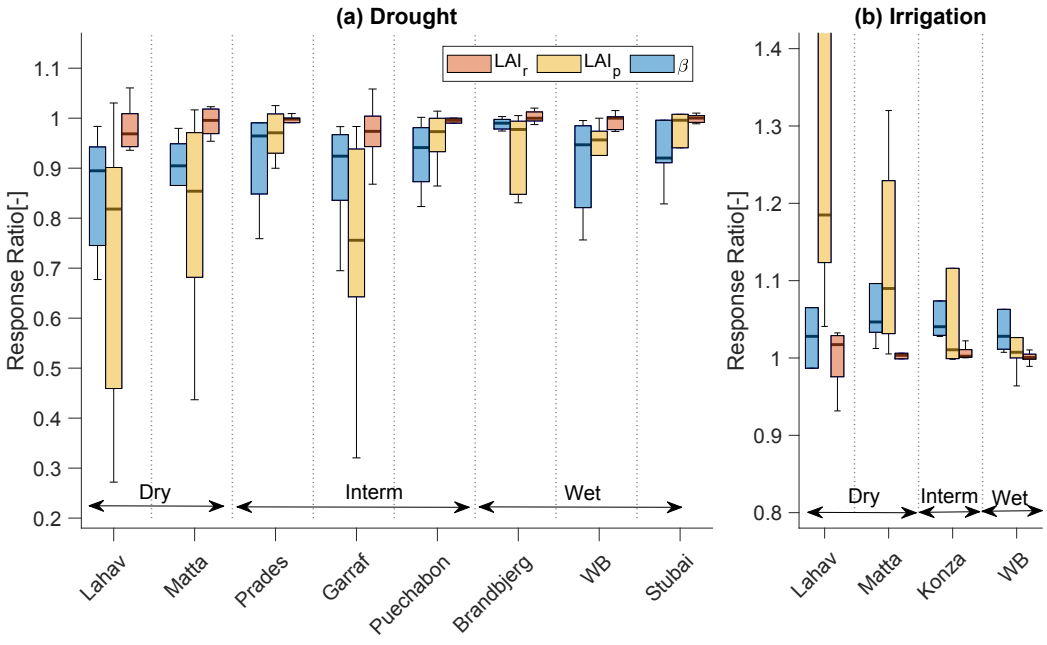


*Figure 9: Boxplots of the response ratios of the change of $\beta$,$LAI_p$ and $LAI_r$ as simulated by (T&C, JSBACH, DLEM, CABLE, JULES and TECO) for the drought experiments (a) and the irrigation experiments (b).*

## 4 Discussion

***Multi-site and local sensitivities to rainfall and the role of temporal scales***

Most models overestimated the relationship between mean annual precipitation and average annual ANPP observed across sites, but managed to capture well the overall trend, despite large site differences in terms of vegetation coverage and overall climatic regime (Figure 1). This result confirms that terrestrial biosphere models can capture spatial gradients of vegetation productivity relatively well (e.g. Wu *et al.*, 2018). Reproducing local (single-site) response of ANPP to interannual precipitation variability has been generally found to be more challenging (Fatichi and Ivanov, 2014). In fact, previous intercomparison studies have found that models have significant biases at various time scales, from subdaily (Matheny *et al.*, 2014) to decadal (Dietze *et al.*, 2011). Dietze *et al.*, (2011) found model errors to be largest at the annual scale. In agreement with such a result in our experiment, models differed greatly in their simulated sensitivity of local scale productivity to annual precipitation but were able to reproduce the previously reported stronger spatial than

temporal sensitivity of productivity to rainfall. A large model disagreement with regards to the magnitude of the interannual variability of ANPP also confirms the previously found difficulties of models to properly capture carbon dynamics at the annual scale (e.g. Dietze *et al.*, 2011, Paschalis *et al.,* 2015). Despite large model disagreement we found that the within site sensitivity of ANPP to precipitation is lower than across site sensitivity of ANPP to average precipitation, in agreement with a number of previous observational (Goward and Prince, 1995; Knapp and Smith, 2001; Huxman et al., 2004) and modelling results (Fatichi and Ivanov, 2014; Wu et al., 2018).

One of the main reasons for model disagreement originates from the differences in parametrization in schemes representing water limitation effects on water and carbon fluxes (e.g. Trugman *et al*., 2018), summarized here by the water stress parameter $\beta$ (Figure 6). Those parametrizations influence ecosystem dynamics at a wide range of temporal scales, complicating assessment of their skill. For instance, at shorter time scales (e.g. daily), in ecosystems with no water limitation, where temperature and radiation are the dominant controls for ET and GPP (Paschalis *et al.*, 2015), models had a high agreement (Figure 5), in terms of correlation. This highlights that parametrizations that impact the temporal changes of ET and GPP should be relatively consistent among models, at least during wet conditions (Ukkola *et al.*, 2016). Even though correlation between models was high, large variability between models with regards to the actual magnitude of the fluxes was pronounced (Figure S2-S4), primarily for carbon fluxes (e.g. GPP). This indicates that a "scaling" factor affecting GPP is significantly different amongst models. For our experiments, LAI could be this explanatory "scaling" factor (Figure 4), as models greatly differed regarding the seasonality and magnitude of LAI.

Significant changes emerge under drought, when water stress parametrizations influence the simulation of water and carbon fluxes. Different water stress parametrizations alter the water/carbon dynamics at different scales. In severely water-limited systems ($WI < 0.4$), model results diverge in terms of GPP and ET at short temporal scales (e.g. daily - Figure 5). Thus, parametrizations of how water stress impacts processes operating at daily and sub daily time scales are crucial, and highly diverging amongst models. Such parametrizations include stomatal regulations and downregulation of photosynthesis during drought. In general, plant hydraulic dynamics will also operate at these temporal scales, but none of the participating models simulated such processes in detail. In severely water limited ecosystems the amount of annual precipitation imposes a strong constraint on evapotranspiration (i.e. $ET \cong P$), leading to overall good agreement between models for annual ET. However, this agreement is not true for transpiration alone (Figure S8), highlighting the major importance of how stomatal limitations are implemented in models. Physical constraints for productivity are not as strong, and thus models have large disagreement with respect to GPP even at annual scales.

In intermediate wetness sites ($0.4 \leq WI < 1$), in our simulations, models disagree at intermediate scales (weeks-months) in terms of GPP (consistent with the wavelet coherence analysis presented at Figure S7). As mentioned before, at short (daily) temporal scales, temperature and radiation mostly determine water and carbon fluxes, when water is not a strong limiting factor, and due to the similar parametrizations among models (Wu *et al.*, 2018), we detect a substantial convergence in GPP. However, since such controls "fade" with increasing temporal scales, the effects of features linked to soil moisture dynamics, such as the soil moisture retention after a rainfall event, can manifest at longer temporal scales (Paschalis *et al.*, 2015). Those dynamics can be influenced by factors including both biotic and abiotic factors such as the parametrizations of soil properties that determine the temporal dynamics of soil moisture and the vertical distribution of root biomass, affecting how plants withdraw water from the soil. In fact, models were found to strongly disagree on how plants are affected by soil moisture (biotic factor – Figure 6) and on the soils' water holding capacity, as indicated by the range of accessible values of soil moisture (abiotic factor – Figure 6).

At the wettest sites (WI>1), strong model disagreement in terms of both water and carbon fluxes occurs at annual scales. A key factor for model disagreement for those sites is LAI (Figure 4). Model disagreement in LAI is a composite effect of the water stress impacts to LAI development and the overall model disagreement in leaf phonology and carbon allocation rules (Figure 4; Richardson *et al.*, 2012).

All those behaviours highlight further the need to correctly capture water/carbon dynamics at multiple time scales, from the scale of the individual rain pulse (Huxman et al., 2004a) up to interannual scales where drought legacies can have an important effect (Anderegg *et al.*, 2015). The need to understand in detail multi-scale dynamics linked to water stress and soil moisture dynamics is also exacerbated by the fact that model disagreement in terms of the sensitivity of ANPP to annual rainfall is highest for sites with intermediate wetness ($0.4 \leq WI < 1$). Those regions experience moderate water limitations, and the impact of water limitation to fast acting processes (changes in e.g. stomatal conductance, photosynthesis) can accumulate and impact longer time scales through slow acting processes (e.g. changes in LAI). Additionally, areas with intermediate wetness are expected to operate close to soil moisture thresholds inducing plant water stress. Sensitivity of the responses of ANPP to precipitation in those sites is concurrently the highest and most uncertain (Figure 2). This can have a large impact on our ability to model the fate of terrestrial $CO_2$, given that those areas are amongst the largest contributors to the interannual dynamics of the growth rate of $CO_2$ (Poulter *et al.*, 2014; Ahlström *et al.*, 2015). Understanding such dynamics across scales requires high quality and high frequency long-term measurements, not only for $CO_2$ and water fluxes but also soil moisture dynamics (Vicca

*et al*., 2012). Annual ANPP values alone are limiting our inference capabilities and even 10-20 years of annual ANPP data were not long enough to obtain a precise estimate of the sensitivity of ANPP to precipitation.

Uncertainties arise from the relatively short span of the record, but also due to the lack of data describing short-scale dynamics of carbon assimilation and growth in manipulation experiments. Annual precipitation has been found to be a relatively weak descriptor of the interannual variability of water and carbon fluxes in many locations worldwide (Fatichi and Ivanov, 2014). A better descriptor would be the time duration during a year when favourable meteorological conditions for photosynthesis occur under well-watered conditions (Fatichi and Ivanov, 2014; Zscheischler *et al.*, 2016a). As a result, a few bursts of positive extremes in terms of productivity can strongly modify the annual budget and long-term dynamics (Zscheischler, *et al.*, 2014a). Therefore, to quantify the interannual dynamics of vegetation productivity, detailed knowledge of water/carbon fluxes, meteorology, soil moisture and plant water status at fine temporal scales would be essential. In fact, previous research at the PHACE experiment, one of the few facilities that combined such high frequency measurement clearly identified the problems models have in reproducing sub-annual dynamics (De Kauwe *et al.*, 2017). Given the present limited availability of such data, new ways of combining existing data (e.g., combining different data-streams representing short and long-term-dynamics in multiple locations, such as Fluxnet sites for water and carbon fluxes at high frequencies, sites equipped with phonecams for high frequency phenology monitoring, soil moisture networks (e.g. COSMOS, the International Soil Moisture Network, the Long Term Ecological Research Network etc.), open access data archiving with common data formats to facilitate data exchange between research groups and the use of proxy data to extend the length of the time series (e.g., tree rings) are necessary to better inform models (Pappas *et al*., 2017; Babst *et al.*, 2018).

***Response to manipulation experiments***

The modelled sensitivities of vegetation dynamics to changes in rainfall are highly uncertain. On average, most models captured better the observed responses of vegetation to rainfall exclusion than addition (Figure 7). That behaviour can be associated with low skill in reproducing the asymmetric response of productivity to precipitation (Wu *et al.*, 2018), failing to capture the correct pattern of the productivity saturation effect associated with rainfall increase.

Even though, multiple models generated close vegetation productivity responses in the rainfall exclusion experiments, the underlying reasons are very different and at the same time highly uncertain (Figure 9). In the more water-limited ecosystems, both changes in LAI magnitude and the level of plan water limitation determine productivity responses. Variability of the relative strength of $\beta$ and $LAI_p$ between models is large.

Variability concerning $LAI_p$ is larger than $\beta$, which can be explained by the fact that $LAI_p$ integrates the model differences related to $LAI$ phenology, carbon allocation rules, and reductions in photosynthetic rates due to soil moisture limitations. Pinpointing which model best captures the relative strengths of changes in $\beta$ and $LAI_p$ would require simultaneous high frequency data, including soil moisture, regular measurements of stomatal conductance and leaf water potentials, high frequency photosynthetic rates, and regular LAI estimates. At more mesic sites, physiological effects of water stress (through $\beta$) are the main reason for productivity responses. The reason is that in such sites, induced water stress is mild. Productivity will be reduced during the imposed water stress due to rainfall exclusion, but this small increase in water stress cannot cause large changes in vegetation structure (Estiarte *et al*., 2016), or LAI.

Disagreement in irrigation experiments is primarily related to leaf area dynamics. The reason can be that in the simulations where water stress was relieved, model disagreement originates primarily from the leaf area dynamics simulated for the unstressed conditions. Those dynamics are related to the choice of carbon allocation and leaf phenology algorithms. Pronounced model differences related to those dynamics can be shown via the magnitude and seasonal patterns of LAI (Figure 4) as simulated by all models. Both the allocation and the phenology algorithms affect the dynamics of LAI. In our simulations (Figure 4) the range in modelled LAI is large and comparable with that reported by previous studies (Walker *et al*., 2014; De Kauwe *et al*., 2017). Parametrizations of carbon allocation and are also limited by generic plant functional types (PFTs) used by most models. Such a choice is generally very restrictive and cannot capture the natural variability of plant traits, which is relevant at the local scale.

In our analysis changes in growing season length were not evident and did not influence out results. This is not surprising, as all rainfall manipulation experiments decreased or increased the available water to the ecosystem, without altering its "pulse" structure, including the frequency of rainfall occurrence, and the time of storm arrival (Ross *et al.*, 2012). As vegetation phenology in water limited ecosystem is very sensitive to the pulse structure dynamics of rainfall (Heisler-While et al., 2009), evaluating in future experiments, whether models can properly capture the responses of vegetation to rainfall pulses in terms of productivity and drought deciduousness is very important. Changes in rainfall pulses will also strongly impact soil respiration dynamics, that will contribute significantly to the total carbon balance (Unger *et al*., 2010; Jarvis *et al*., 2007).

***Outlook for model developments and observations***

Our results highlight the need for a coordinated effort of new model development and data collection that could enable validations that are much more detailed than currently achievable here. Model discrepancies in the

present study were attributed to the $\beta$ stress factor, and long-term leaf area dynamics. The models used in this study implemented simple conceptual, yet vastly different (Wu *et al*., 2018) parametrizations of the effects of water limitation, neglecting plant hydraulics and thus impacts on the water transport system (xylem cavitation) that can lead to hydraulic failure or/and carbon starvation (McDowell, 2011; McDowell *et al.*, 2013; Bonan *et al.*, 2014; Xu *et al.*, 2016). This could be an important limitation. However, tree mortality is not a prominent feature of the manipulation experiments considered here and while it has attracted a lot of attention, models first need to better simulate mild to severe water stress before considering vegetation death. For instance, differences associated with the $\beta$ factor are not only related to plant physiological thresholds but are a complex function of the assumed soil textural properties. Those properties are translated into soil hydraulic parameters (Van Looy *et al* 2017), affecting soil moisture dynamics and ET and ultimately their interplay with the value of the $\beta$ factor. It is currently impossible or very difficult to identify which model is more realistic in this respect and each model can only "tune" all the above components at once. Specialized experiments measuring e.g. simultaneously high frequency water and carbon fluxes, soil moisture and plant water status in controlled environments could be designed to develop more informed parameterizations of $\beta$, and eventually expand to more detailed mechanistic representation of ecosystem scale plant hydraulics (Anderegg *et al.*, 2016; Konings and Gentine, 2017).

Correct modelling of leaf area dynamics is equally important as the plant physiological stress $\beta$ for quantifying the effect of rainfall changes in ecosystem functioning (Yang *et al.*, 2018). Simulation of LAI could be constrained better than currently done with available information, considering that high frequency LAI measurements in an experiment could be added with a relatively low budget. Observations of LAI, via indirect methods, are common at large scale. Extensive ground (Iio *et al.*, 2014) and remote sensing estimates (Zhu *et al.*, 2013) of LAI and phenology data from low cost cameras worldwide (Klosterman *et al.*, 2014; Brown *et al.*, 2016) can be used to further constrain phenology and carbon allocation. Regarding carbon allocation, below ground dynamics and their responses to water limitation should also be simultaneously quantified.

From an observational perspective, in order to improve models, we need to disentangle the effects on plant physiological stress from those on vegetation dynamics at the local scales. Since physiological effects of water stress manifest earlier than changes of LAI or carbon pools, a nearly continuous monitoring of photosynthesis, evapotranspiration, leaf and soil water potentials, sap flow and leaf area index would be essential to get further insights. These quantities are often observed (e.g. using eddy covariance systems, sap flow sensors, leaf porometers, hyperspectral cameras), but rarely in an integrated manner and associated with rainfall manipulation experiments. This should become a priority to foster model developments.

Finally, new streams of data via remote sensing can be also used for detailed model confirmation at larger scales. Satellite and airborne data related to vegetation structure, spanning from leaf chemistry to delineation of individual trees (Andersen, *et al.,* 2006; Gougeon and Leckie, 2006; Asner and Martin, 2009), high frequency photosynthesis through solar induced fluorescence (SIF), soil moisture (Liu *et al.*, 2011), and plant hydraulic status (Konings and Gentine, 2017) currently exist. Such data can help us to identify the mechanistic link between plant water stress and how it affects vegetation productivity from short term photosynthesis reduction to decadal scales involving plant mortality and composition shifts. Note however that estimates of photosynthetic activity during water stress purely based on remote sensing (light reflection signals) are often biased and need to be interpreted with care (De Kauwe *et al.*, 2016; Stocker *et al.*, 2019).

In conclusion, our key finding in this study is that current generation terrestrial biosphere models have major uncertainties related to simulating plant water stress, and its impact on the terrestrial carbon cycling. Those uncertainties arise from the model formulations related to both carbon allocation patterns and phenology and the representation of water stress frequency and magnitude on carbon assimilation. These two effects are inherently coupled at a wide range of scales. To decouple the two effects and constrain mechanistic representations of how water stress acts on multiple processes will require the close collaboration between experimentalists and modellers, for planning and implementing new "high frequency" experiments (Rineau *et al*., 2019). These experiments should observe across a range of temporal scales from hourly values of photosynthesis and ET, to daily and weekly LAI dynamics, up to arrive to annual changes in species composition (Halbritter *et al*., 2019).

*The authors declare no conflict of interest*

***Data Sharing and Accessibility statement:*** All meteorological input data and model outputs can be found at the zenodo data repository (https://zenodo.org/ doi:10.5281/zenodo.3627959)

**Acknowledgements**

We would like to thank Prof. Gil Bohrer and three anonymous reviewers for their constructive comments that help us improve the manuscript.. A.P. acknowledges financial support from NERC (grant no. NE/S003495/1). J.Z. acknowledges the Swiss National Science Foundation (Ambizione Grant 179876). DSG, PC, WL, ME, RO and JP are funded by the "IMBALANCE-P" project of the European Research Council (ERC-2013-SyG-610028). CP acknowledges the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discover Grant. YPW acknowledges the financial support from the National

Environmental Science Program for Earth System and Climate Change from the Australian Federal government. IKS and KSL acknowledge the financial support to the CLIMAITE project at Brandbjerg from the Villum Foundation. J.Po. was supported by the German Research Foundation's (DFG) Emmy Noether Program (PO 1751/1-1). L.B. was supported by the DFG's CE-LAND project. Computational resources were made available by the German Climate Computing Center (DKRZ) through support from the German Federal Ministry of Education and Research (BMBF). MB acknowledges the support of the Austrian Science Fund (FWF; P22214-B17), and the European Community's Seventh Framework Programme (FP7/2007-2013, project 'CARBO-Extreme', grant agreement no. 226701), the Austrian Academy of Sciences (OeAW; ClimLUC) and the Austrian Research Promotion Agency (FFG; LTER-CWN). We thank all site operators, MODIS and FLUXNET2015 for providing the data for this study.

**References**

Ahlström, A. *et al.* (2015) 'The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink', *Science*, 348(6237), pp. 895–899. doi: 10.1126/science.aaa1668.

Alexander, L. V. *et al.* (2006) 'Global observed changes in daily climate extremes of temperature and precipitation', *Journal of Geophysical Research*, 111(D5), p. D05109. doi: 10.1029/2005JD006290.

Allan, R. P. *et al.* (2014) 'Physically Consistent Responses of the Global Atmospheric Hydrological Cycle in Models and Observations', *Surveys in Geophysics*, 35(3), pp. 533–552. doi: 10.1007/s10712-012-9213-z.

Allen, C. D. *et al.* (2010) 'A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests', *Forest Ecology and Management*, 259(4), pp. 660–684. doi: 10.1016/j.foreco.2009.09.001.

Allen, C. D., Breshears, D. D. and McDowell, N. G. (2015) 'On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene', *Ecosphere*, 6(8), p. art129. doi: 10.1890/ES15-00203.1.

Anderegg, W. R. L. *et al.* (2015) 'Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models', *Science*, 349(6247), pp. 528–531. doi: 10.1126/science.aab1833.

Anderegg, W. R. L. *et al.* (2016) 'Meta-analysis reveals that hydraulic traits explain cross-species patterns of drought-induced tree mortality across the globe', *Proceedings of the National Academy of Sciences*, 113(18), pp. 5024–5029. doi: 10.1073/pnas.1525678113.

Andersen, H.-E., Reutebuch, S. E. and McGaughey, R. J. (2006) 'A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods', *Canadian Journal of Remote Sensing*, 32(5), pp. 355–366. doi: 10.5589/m06-030.

Asner, G. P. and Martin, R. E. (2009) 'Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests', *Frontiers in Ecology and the Environment*, 7(5), pp. 269–276. doi: 10.1890/070152.

Babst, F. *et al.* (2018) 'When tree rings go global: Challenges and opportunities for retro- and prospective insight', *Quaternary Science Reviews*, 197, pp. 1–20. doi: 10.1016/j.quascirev.2018.07.009.

Báez, S. *et al.* (2013) 'Effects of experimental rainfall manipulations on Chihuahuan Desert grassland and shrubland plant communities', *Oecologia*, 172(4), pp. 1117–1127. doi: 10.1007/s00442-012-2552-0.

Beier, C. *et al.* (2009) 'Carbon and nitrogen balances for six shrublands across Europe', *Global Biogeochemical Cycles*, 23(4), p. n/a-n/a. doi: 10.1029/2008GB003381.

Bonan, G. B. *et al.* (2014) 'Modeling stomatal conductance in the Earth system: linking leaf water-use efficiency and water transport along the soil-plant-atmosphere continuum', *Geoscientific Model Development*, 7(3), pp. 3085–3159. doi: 10.5194/gmd-7-2193-2014.

Brown, T. B. *et al.* (2016) 'Using phenocams to monitor our changing Earth: toward a global phenocam network', *Frontiers in Ecology and the Environment*, 14(2), pp. 84–93. doi: 10.1002/fee.1222.

Choat, B. *et al.* (2012) 'Global convergence in the vulnerability of forests to drought', *Nature*, 491(7426), pp. 752–755. doi: 10.1038/nature11688.

Clark, D. B. *et al.* (2011) 'The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics', *Geoscientific Model Development*, 4(3), pp. 701–722. doi: 10.5194/gmd-4-701-2011.

Collins, S. L. *et al.* (2012) 'Stability of tallgrass prairie during a 19-year increase in growing season precipitation', *Functional Ecology*, 26(6), pp. 1450–1459. doi: 10.1111/j.1365-2435.2012.01995.x.

Dietze, M. C*., et al*.,. (2011). Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis. *Journal of Geophysical Research*, 116(G4), G04029. https://doi.org/10.1029/2011JG001661

Egea, G., Verhoef, A. and Vidale, P. L. (2011) 'Towards an improved and more flexible representation of water stress in coupled photosynthesis–stomatal conductance models', *Agricultural and Forest Meteorology*. Elsevier B.V., 151(10), pp. 1370–1384. doi: 10.1016/j.agrformet.2011.05.019.

Eller, C. B. *et al.* (2018) 'Modelling tropical forest responses to drought and El Niño with a stomatal optimization model based on xylem hydraulics', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1760), p. 20170315. doi: 10.1098/rstb.2017.0315.

Estiarte, M. *et al.* (2016) 'Few multiyear precipitation-reduction experiments find a shift in the productivity-precipitation relationship', *Global change biology*, 22(7), pp. 2570–2581. doi: 10.1111/gcb.13269.

Fang, H. *et al.* (2013) 'Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties', *Journal of Geophysical Research: Biogeosciences*, 118(2), pp. 529–548. doi: 10.1002/jgrg.20051.

Fatichi, S. *et al.* (2019) 'Modelling carbon sources and sinks in terrestrial vegetation', *New Phytologist*, 221(2), pp. 652–668. doi: 10.1111/nph.15451.

Fatichi, S. and Ivanov, V. Y. (2014) 'Interannual variability of evapotranspiration and vegetation productivity', *Water Resources Research*, 50(4), pp. 3275–3294. doi: 10.1002/2013WR015044.

Fatichi, S., Ivanov, V. Y. and Caporali, E. (2012) 'A mechanistic ecohydrological model to investigate complex interactions in cold and warm water-controlled environments: 1. Theoretical framework and plot-scale analysis', *Journal of Advances in Modeling Earth Systems*, 4(2), p. n/a-n/a. doi: 10.1029/2011MS000086.

Fatichi, S. and Leuzinger, S. (2013) 'Reconciling observations with modeling: The fate of water and carbon allocation in a mature deciduous forest exposed to elevated CO2', *Agricultural and Forest Meteorology*. Elsevier B.V., 174–175, pp. 144–157. doi: 10.1016/j.agrformet.2013.02.005.

Fatichi, S., Pappas, C. and Ivanov, V. Y. (2016) 'Modeling plant-water interactions: an ecohydrological overview from the cell to the global scale', *Wiley Interdisciplinary Reviews: Water*, 3(3), pp. 327–368. doi: 10.1002/wat2.1125.

Fay, P. a. *et al.* (2008) 'Changes in grassland ecosystem function due to extreme rainfall events: implications for responses to climate change', *Global Change Biology*, 14(7), pp. 1600–1608. doi: 10.1111/j.1365-2486.2008.01605.x.

Fisher, R. A. *et al.* (2007) 'The response of an Eastern Amazonian rain forest to drought stress: results and modelling analyses from a throughfall exclusion experiment', *Global Change Biology*, 13(11), pp. 2361–2378. doi: 10.1111/j.1365-2486.2007.01417.x.

Frank, D. D. *et al.* (2015) 'Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts', *Global Change Biology*, 21(January), pp. 2861–2880. doi: 10.1111/gcb.12916.

Fuchslueger, L. *et al.* (2014) 'Experimental drought reduces the transfer of recently fixed plant carbon to soil microbes and alters the bacterial community composition in a mountain meadow', *New Phytologist*, 201(3), pp. 916–927. doi: 10.1111/nph.12569.

Gentine, P., et al, (2019), Land–atmosphere interactions in the tropics – a review, Hydrol. Earth Syst. Sci., 23, 4171–4197, doi: 10.5194/hess-23-4171-2019, 2019.

Goll, D. S. *et al.* (2017) 'A representation of the phosphorus cycle for ORCHIDEE (revision 4520)', *Geoscientific Model Development*, 10(10), pp. 3745–3770. doi: 10.5194/gmd-10-3745-2017.

Golodets, C. *et al.* (2013) 'From desert to Mediterranean rangelands: will increasing drought and inter-annual rainfall variability affect herbaceous annual primary productivity?', *Climatic Change*, 119(3–4), pp. 785–798. doi: 10.1007/s10584-013-0758-8.

Golodets, C. *et al.* (2015) 'Climate change scenarios of herbaceous production along an aridity gradient: vulnerability increases with aridity', *Oecologia*, 177(4), pp. 971–979. doi: 10.1007/s00442-015-3234-5.

Gougeon, F. A. and Leckie, D. G. (2006) 'The Individual Tree Crown Approach Applied to Ikonos Images of a Coniferous Plantation Area', *Photogrammetric Engineering & Remote Sensing*, 72(11), pp. 1287–1297. doi: 10.14358/PERS.72.11.1287.

Goward, S. N. and Prince, S. D. (1995) 'Transient Effects of Climate on Vegetation Dynamics: Satellite Observations', *Journal of Biogeography*, 22(2/3), p. 549. doi: 10.2307/2845953.

Green, J. K. *et al.* (2017) 'Regionally strong feedbacks between the atmosphere and terrestrial biosphere', *Nature Geoscience*, 10(6), pp. 410–414. doi: 10.1038/ngeo2957.

Green, J. K. *et al.* (2019) 'Large influence of soil moisture on long-term terrestrial carbon uptake', *Nature*. Springer US, 565(7740), pp. 476–479. doi: 10.1038/s41586-018-0848-x.

Hagedorn F. *et al*. (2016), Recovery of trees from drought depends on belowground sink control, Nature Plants, 2 (16111), doi: 10.1038/nplants.2016.111.

Halbritter A. H. *et al.*, (2019) The handbook for standardized field and laboratory measurements in terrestrial climate change experiments and observational studies (ClimEx), *Methods in Ecology and Evolution*, doi: 10.1111/2041-210X.13331

Hanson, P. J. *et al.* (2004) 'Oak forest carbon and water simulations: model intercoparisons and evaluations against independent data', *Ecological Monographs*, 74(3), pp. 443–489. doi: 10.1890/03-4049.

Hanson, P. J. and Wullschleger, S. D. (eds) (2003) *North American Temperate Deciduous Forest Responses to*

*Changing Precipitation Regimes*. New York, NY: Springer New York (Ecological Studies). doi: 10.1007/978-1-4613-0021-2.

Hasibeder R. *et al*., (2014), Summer drought alters carbon allocation to roots and root respiration in mountain grassland, *New Phytologist*, 205(3), 1117-1127, doi: 10.1111/nph.13146

Heisler-White, J. L. *et al.* (2009) 'Contingent productivity responses to more extreme rainfall regimes across a grassland biome', *Global Change Biology*, 15(12), pp. 2894–2904. doi: 10.1111/j.1365-2486.2009.01961.x.

Heisler-White, J. L., Knapp, A. K. and Kelly, E. F. (2008) 'Increasing precipitation event size increases aboveground net primary productivity in a semi-arid grassland.', *Oecologia*, 158(1), pp. 129–140. doi: 10.1007/s00442-008-1116-9.

Hu, Z. *et al.* (2018) 'Joint structural and physiological control on the interannual variation in productivity in a temperate grassland: A data-model comparison', *Global Change Biology*, 24(7), pp. 2965–2979. doi: 10.1111/gcb.14274.

Huang Y., et al., (2017), Soil thermal dynamics, snow cover, and frozen depth under five temperature treatments in an ombrotrophic bog: Constrained forecast with data assimilation, *Journal of Geophysical Research: Biogeosciences,* 122(8), doi: 10.1002/2016JG003725.

Humphrey, V. *et al.* (2018) 'Sensitivity of atmospheric CO2 growth rate to observed changes in terrestrial water storage', *Nature*. Springer US, 560(7720), pp. 628–631. doi: 10.1038/s41586-018-0424-4.

Huxman, T. E. *et al.* (2004) 'Precipitation pulses and carbon fluxes in semiarid and arid ecosystems.', *Oecologia*, 141(2), pp. 254–268. doi: 10.1007/s00442-004-1682-4.

Iio, A. *et al.* (2014) 'Global dependence of field-observed leaf area index in woody species on climate: a systematic review', *Global Ecology and Biogeography*, 23(3), pp. 274–285. doi: 10.1111/geb.12133.

IPCC (2012) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Edited by C. B. Field et al. Cambridge Univ Press.

IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press Cambridge, UK.

Jarvis P. et al., (2007), Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the "Birch effect", *Tree Physiology*, 27(7), 929-940, doi: 10.1093/treephys/27.7.929

Kaminski, T. *et al.* (2013) 'The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges', *Journal of Geophysical Research: Biogeosciences*, 118(4), pp. 1414–1426. doi: 10.1002/jgrg.20118.

De Kauwe, M. G. *et al.* (2013) 'Forest water use and water use efficiency at elevated CO2 : a model-data intercomparison at two contrasting temperate forest FACE sites.', *Global change biology*, 19(6), pp. 1759–1779. doi: 10.1111/gcb.12164.

De Kauwe, M. G., Kala, J., *et al.* (2015) 'A test of an optimal stomatal conductance scheme within the CABLE land surface model', *Geoscientific Model Development*, 8(2), pp. 431–452. doi: 10.5194/gmd-8-431-2015.

De Kauwe, M. G., Zhou, S.-X. X., *et al.* (2015) 'Do land surface models need to include differential plant species responses to drought? Examining model predictions across a mesic-xeric gradient in Europe', *Biogeosciences Discussions*, 12(24), pp. 7503–7518. doi: 10.5194/bgd-12-12349-2015.

De Kauwe, M. G. *et al.* (2016) 'Satellite based estimates underestimate the effect of CO2 fertilization on net primary productivity', *Nature Climate Change*. Nature Publishing Group, 6(10), pp. 892–893. doi: 10.1038/nclimate3105.

De Kauwe, M. G. *et al.* (2017) 'Challenging terrestrial biosphere models with data from the long-term multifactor Prairie Heating and CO2 Enrichment experiment', *Global Change Biology*, 23(9), pp. 3623–3645. doi: 10.1111/gcb.13643.

Kayler Z. E., *et al.,* (2015). Experiments to Confront the Environmental Extremes of Climate Change. Frontiers in Ecology and the Environment, 13(4), 219-225, doi: 10.1890/140174

Kennedy, D. *et al.* (2019) 'Implementing Plant Hydraulics in the Community Land Model, Version 5', *Journal of Advances in Modeling Earth Systems*, 11(2), pp. 485–513. doi: 10.1029/2018MS001500.

Klein, T. *et al.* (2014) 'Towards an advanced assessment of the hydrological vulnerability of forests to climate change-induced drought', *New Phytologist*, 201(3), pp. 712–716. doi: 10.1111/nph.12548.

Klosterman, S. T. *et al.* (2014) 'Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery', *Biogeosciences*, 11(16), pp. 4305–4320. doi: 10.5194/bg-11-4305-2014.

Knapp, A. K. and Smith, M. D. (2001) 'Variation among biomes in temporal dynamics of aboveground primary production.', *Science*, 291(5503), pp. 481–484. doi: 10.1126/science.291.5503.481.

Knutti, R. and Sedláček, J. (2013) 'Robustness and uncertainties in the new CMIP5 climate model projections', *Nature Climate Change*, 3(4), pp. 369–373. doi: 10.1038/nclimate1716.

Kongstad, J. *et al.* (2012) 'High Resilience in Heathland Plants to Changes in Temperature, Drought, and CO2 in Combination: Results from the CLIMAITE Experiment', *Ecosystems*, 15(2), pp. 269–283. doi: 10.1007/s10021-011-9508-9.

Konings, A. G. and Gentine, P. (2017) 'Global variations in ecosystem-scale isohydricity', *Global Change Biology*, 23(2), pp. 891–905. doi: 10.1111/gcb.13389.

Körner, C. (2019) 'No need for pipes when the well is dry—a comment on hydraulic failure in trees', *Tree Physiology*. Edited by S. Sevanto, pp. 1–6. doi: 10.1093/treephys/tpz030.

Koster, R. D. (2004) 'Regions of Strong Coupling Between Soil Moisture and Precipitation', *Science*, 305(5687), pp. 1138–1140. doi: 10.1126/science.1100217.

Krinner, G. *et al.* (2005) 'A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system', *Global Biogeochemical Cycles*, 19(1). doi: 10.1029/2003GB002199.

Lawrence, D. M., *et al.,* ( 2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. Journal of Advances in Modeling Earth Systems, 11. https://doi.org/10.1029/2018MS001583

Lemordant, L. *et al.* (2016) 'Modification of land-atmosphere interactions by CO 2 effects: Implications for summer dryness and heat wave amplitude', *Geophysical Research Letters*, 43(19), pp. 10,240-10,248. doi: 10.1002/2016GL069896.

Lienert, S. and Joos, F. (2018) 'A Bayesian ensemble data assimilation to constrain model parameters and land-use carbon emissions', *Biogeosciences*, 15(9), pp. 2909–2930. doi: 10.5194/bg-15-2909-2018.

Limousin, J. M. *et al*., (2009), Long-term transpiration change with rainfall decline in a Mediterranean Quercus ilex forest. *Global Change Biology*, 15(9), 2163-2175, doi: 10.1111/j.1365-2486.2009.01852.x

Liu, Y. Y. *et al.* (2011) 'Developing an improved soil moisture dataset by blending passive and active

microwave satellite-based retrievals', *Hydrology and Earth System Sciences*, 15(2), pp. 425–436. doi: 10.5194/hess-15-425-2011.

Van Looy, K. *et al.* (2017) 'Pedotransfer Functions in Earth System Science: Challenges and Perspectives', *Reviews of Geophysics*, 55(4), pp. 1199–1256. doi: 10.1002/2017RG000581.

Manzoni, S. *et al.* (2013) 'Biological constraints on water transport in the soil–plant–atmosphere system', *Advances in Water Resources*, 51, pp. 292–304. doi: 10.1016/j.advwatres.2012.03.016.

Matheny, A. M., *et al*., (2014). Characterizing the diurnal patterns of errors in the prediction of evapotranspiration by several land-surface models: An NACP analysis. *Journal of Geophysical Research: Biogeosciences*, 119(7), 1458–1473. https://doi.org/10.1002/2014JG002623

Martin-Stpaul, N. K. *et al.* (2013) 'The temporal response to drought in a Mediterranean evergreen tree: Comparing a regional precipitation gradient and a throughfall exclusion experiment', *Global Change Biology*, 19(8), pp. 2413–2426. doi: 10.1111/gcb.12215.

Mauritsen T., *et al*., (2019), Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2, Journal of Advances in Modeling Earth Systems, 11(4), doi: 10.1029/2018MS001400.

McDowell, N. G. (2011) 'Mechanisms linking drought, hydraulics, carbon metabolism, and vegetation mortality.', *Plant physiology*, 155(3), pp. 1051–1059. doi: 10.1104/pp.110.170704.

McDowell, N. G. *et al.* (2013) 'Evaluating theories of drought-induced vegetation mortality using a multimodel-experiment framework', *New Phytologist*, 200(2), pp. 304–321. doi: 10.1111/nph.12465.

Medlyn, B. E. *et al.* (2015) 'Using ecosystem experiments to improve vegetation models', *Nature Climate Change*. Nature Publishing Group, 5(6), pp. 528–534. doi: 10.1038/nclimate2621.

Medlyn, B. E., De Kauwe, M. G. and Duursma, R. A. (2016a) 'New developments in the effort to model ecosystems under water stress', *New Phytologist*, 212(1), pp. 5–7. doi: 10.1111/nph.14082.

Medlyn, B. E., *et al.,* (2016b). Using models to guide field experiments: a priori predictions for the CO2 response of a nutrient- and water-limited native Eucalypt woodland. *Global Change Biology*, 22(8), 2834–2851. https://doi.org/10.1111/gcb.13268

Miralles, D. G. *et al.* (2018) 'Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges', *Annals of the New York Academy of Sciences*, 1436, pp. 19–35. doi: 10.1111/nyas.13912.

Miranda, J. D. *et al.* (2011) 'Climatic change and rainfall patterns: Effects on semi-arid plant communities of the Iberian Southeast', *Journal of Arid Environments*. Elsevier Ltd, 75(12), pp. 1302–1309. doi: 10.1016/j.jaridenv.2011.04.022.

Mirfenderesgi, G. *et al.* (2016) 'Tree level hydrodynamic approach for resolving aboveground water storage and stomatal conductance and modeling the effects of tree hydraulic strategy', *Journal of Geophysical Research: Biogeosciences*, 121(7), pp. 1792–1813. doi: 10.1002/2016JG003467.

Nepstad, D. C., *et al.,* (2007). Mortality of large trees and lianas following experimental drought in an amazon forest. *Ecology*, 88(9), 2259–2269. https://doi.org/10.1890/06-1046.1

Ogaya, R. and Peñuelas, J. (2007) 'Tree growth, mortality, and above-ground biomass accumulation in a holm oak forest under a five-year experimental field drought', *Plant Ecology*, 189(2), pp. 291–299. doi: 10.1007/s11258-006-9184-6.

Pappas C. *et al*., (2017) Ecosystem functioning is enveloped by hydrometeorological variability. *Nature ecology & evolution*, 1(9), 1263, doi: 10.1038/s41559-017-0277-5.

Paschalis, A. *et al.* (2015) 'Cross-scale impact of climate temporal variability on ecosystem water and carbon fluxes', *Journal of Geophysical Research: Biogeosciences*, 120(9), pp. 1716–1740. doi: 10.1002/2015JG003002.

Paschalis, A. *et al.* (2017) 'On the variability of the ecosystem response to elevated atmospheric CO2 across spatial and temporal scales at the Duke Forest FACE experiment', *Agricultural and Forest Meteorology*. Elsevier B.V., 232, pp. 367–383. doi: 10.1016/j.agrformet.2016.09.003.

Peñuelas, J. *et al.* (2004) 'Complex spatiotemporal phenological shifts as a response to rainfall changes', *New Phytologist*, 161(3), pp. 837–846. doi: 10.1111/j.1469-8137.2004.01003.x.

Poulter, B. *et al.* (2014) 'Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle', *Nature*. Nature Publishing Group, 509(7502), pp. 600–603. doi: 10.1038/nature13376.

Powell, T. L. *et al.* (2013) 'Confronting model predictions of carbon fluxes with measurements of Amazon

forests subjected to experimental drought', *New Phytologist*, 200(2), pp. 350–365. doi: 10.1111/nph.12390.

Le Quéré, C. *et al.* (2018) 'Global Carbon Budget 2018', *Earth System Science Data*, 10(4), pp. 2141–2194. doi: 10.5194/essd-10-2141-2018.

Reichstein, M. *et al.* (2013) 'Climate extremes and the carbon cycle.', *Nature*. Nature Publishing Group, 500(7462), pp. 287–295. doi: 10.1038/nature12350.

Richardson, A. D., *et al*., (2012). Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis. *Global Change Biology*, 18(2), 566–584. doi: 10.1111/j.1365-2486.2011.02562.x

Rineau F. *et al*., (2019) Towards more predictive and interdisciplinary climate change ecosystem experiments, *Nature Climate Change*, 9. 809-819, doi: 10.1038/s41558-019-0609-3.

Ross I. *et al*., (2012), How do variations in the temporal distribution of rainfall events affect ecosystem fluxes in seasonall water-limited Northern Hemisphere shrublands and forests?, *Biogeosciences* 9(9):1007-1024, doi: 10.5194/bg-9-1007-2012.

Rowland, L., *et al.,.* (2014). Evidence for strong seasonality in the carbon storage and carbon use efficiency of an Amazonian forest. *Global Change Biology*, 20(3), 979–991. https://doi.org/10.1111/gcb.12375

Seneviratne, S. I. *et al.* (2013) 'Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment', *Geophysical Research Letters*, 40(19), pp. 5212–5217. doi: 10.1002/grl.50956.

Stocker, B. D. *et al.* (2019) 'Drought impacts on terrestrial primary production underestimated by satellite monitoring', *Nature Geoscience*. Springer US, 12(4), pp. 264–270. doi: 10.1038/s41561-019-0318-6.

Stuart-Haëntjens E., et al., (2018), Mean annual precipitation predicts primary production resistance and resilience to extreme drought, Science of the total Environment, 636(15), 360-366, doi: 10.1016/j.scitotenv.2018.04.290.

Taylor, K. E. (2001) 'Summarizing multiple aspects of model performance in a single diagram', *Journal of Geophysical Research*, 106(D7), pp. 7183–7192. doi: 10.1029/2000JD900719.

Thurner, M. *et al.* (2017) 'Evaluation of climate-related carbon turnover processes in global vegetation models

for boreal and temperate forests', *Global Change Biology*, 23(8), pp. 3076–3091. doi: 10.1111/gcb.13660.

Tian, H. *et al.* (2010) 'Model estimates of net primary productivity, evapotranspiration, and water use efficiency in the terrestrial ecosystems of the southern United States during 1895–2007', *Forest Ecology and Management*, 259(7), pp. 1311–1327. doi: 10.1016/j.foreco.2009.10.009.

Tielbörger, K. *et al.* (2014) 'Middle-Eastern plant communities tolerate 9 years of drought in a multi-site climate manipulation experiment', *Nature Communications*, 5(1), p. 5102. doi: 10.1038/ncomms6102.

Trugman, A. T., *et al*. (2018). Soil Moisture Stress as a Major Driver of Carbon Cycle Uncertainty. Geophysical Research Letters, 45(13), 6495–6503. doi: 10.1029/2018GL078131

Ukkola, A. M. et al. (2016) 'Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts', Environmental Research Letters. IOP Publishing, 11(10), p. 104012. doi: 10.1088/1748-9326/11/10/104012.

Vicca, S. *et al*., (2012), Urgent need for a common metric to make precipitation manipulation experiments comparable, *New Phytologist*, 195(3), 518-522, doi: 10.1111/j.1469-8137.2012.04224.x

Vicca, S. *et al.* (2014) 'Can current moisture responses predict soil CO2 efflux under altered precipitation regimes? A synthesis of manipulation experiments', *Biogeosciences*, 11(11), pp. 2991–3013. doi: 10.5194/bg-11-2991-2014.

Vicca, S. *et al*., (2016) Remotely-sensed detection of effects of extreme droughts on gross primary production, *Scientific Reports*, 6(28269), doi: 10.1038/srep28269

Walker, A. P. *et al.* (2014) 'Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO 2 enrichment experiments: Model performance at ambient CO 2 concentration', *Journal of Geophysical Research: Biogeosciences*, 119(5), pp. 937–964. doi: 10.1002/2013JG002553.

Wang, Y. P. *et al.* (2011) 'Diagnosing errors in a land surface model (CABLE) in the time and frequency domains', *Journal of Geophysical Research*, 116(G1), p. G01034. doi: 10.1029/2010JG001385.

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(11), pp. 3421–3437. doi: 10.5194/bg-15-3421-2018.

Xu, C. *et al.* (2013) 'Our limited ability to predict vegetation dynamics under water stress', *New Phytologist*, 200(2), pp. 298–300. doi: 10.1111/nph.12450.

Xu, X. *et al.* (2016) 'Diversity in plant hydraulic traits explains seasonal and inter-annual variations of vegetation dynamics in seasonally dry tropical forests', *New Phytologist*, 212(1), pp. 80–95. doi: 10.1111/nph.14009.

Yang, J. *et al.* (2018) 'Applying the Concept of Ecohydrological Equilibrium to Predict Steady State Leaf Area Index', *Journal of Advances in Modeling Earth Systems*, 10(8), pp. 1740–1758. doi: 10.1029/2017MS001169.

Zaehle, S. *et al.* (2014) 'Evaluation of 11 terrestrial carbon-nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies', *New Phytologist*, 202(3), pp. 803–822. doi: 10.1111/nph.12697.

Zhang, W. *et al.* (2019) 'Ecosystem structural changes controlled by altered rainfall climatology in tropical savannas', *Nature Communications*. Springer US, 10(1), p. 671. doi: 10.1038/s41467-019-08602-6.

Zhao, M. and Running, S. W. (2010) 'Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009', *Science*, 329(5994), pp. 940–943. doi: 10.1126/science.1192666.

Zhou, S. *et al.* (2013) 'How should we model plant responses to drought? An analysis of stomatal and non-stomatal responses to water stress', *Agricultural and Forest Meteorology*. Elsevier B.V., 182–183, pp. 204–214. doi: 10.1016/j.agrformet.2013.05.009.

Zhu, Z. *et al.* (2013) 'Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2', *Remote Sensing*, 5(2), pp. 927–948. doi: 10.3390/rs5020927.

Zscheischler, J., Mahecha, M. D., *et al.* (2014a) 'A few extreme events dominate global interannual variability in gross primary production', *Environmental Research Letters*, 9(3), p. 035001. doi: 10.1088/1748-9326/9/3/035001.

Zscheischler, J., Michalak, A. M., *et al.* (2014b) 'Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data', *Global Biogeochemical Cycles*, 28(6), pp. 585–600. doi: 10.1002/2014GB004826.

Zscheischler, J. *et al.* (2016) 'Short-term favorable weather conditions are an important control of interannual variability in carbon and water fluxes', *Journal of Geophysical Research: Biogeosciences*, 121(8), pp. 2186–2198. doi: 10.1002/2016JG003503.