

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

Athanasios Paschalis <sup>1\*</sup>, Simone Fatichi <sup>2</sup>, Jakob Zscheischler <sup>3,4</sup>, Philippe Ciais <sup>5</sup>, Michael Bahn <sup>6</sup>, Lena Boysen <sup>7</sup>, Jinfeng Chang <sup>5</sup>, Martin De Kauwe <sup>8</sup>, Marc Estiarte <sup>9,10</sup>, Daniel Goll <sup>5,11</sup>, Paul J. Hanson <sup>12</sup>, Anna B. Harper <sup>13</sup>, Enqing Hou <sup>14</sup>, Jaime Kigel <sup>15</sup>, Alan K. Knapp <sup>16</sup>, Klaus Steenberg Larsen <sup>17</sup>, Wei Li <sup>5,18</sup>, Sebastian Lierert <sup>3,4</sup>, Yiqi Luo <sup>14</sup>, Patrick Meir <sup>19</sup>, Julia E. M. S. Nabel <sup>7</sup>, Romà Ogaya <sup>9,10</sup>, Anthony J Parolari <sup>21</sup>, Changhui Peng <sup>22</sup>, Josep Peñuelas <sup>9,10</sup>, Julia Pongratz <sup>23</sup>, Serge Rambal <sup>9,10</sup>, Inger Kappel Schmidt <sup>17</sup>, Hao Shi <sup>24</sup>, Marcelo Sternberg <sup>25</sup>, Hanqin Tian <sup>24</sup>, Elisabeth Tschumi <sup>3,4</sup>, Anna Ukkola <sup>8</sup>, Sara Vicca <sup>26</sup>, Nicolas Viovy <sup>5</sup>, Ying-Ping Wang <sup>27</sup>, Zhuonan Wang <sup>24</sup>, Karina Williams <sup>28</sup>, Donghai Wu <sup>29</sup>, Qiuan Zhu <sup>30</sup>

<sup>1</sup> Department of Civil and Environmental Engineering, Imperial College London, UK

<sup>2</sup> Institute of Environmental Engineering, ETH Zurich, Switzerland

<sup>3</sup> Climate and Environmental Physics, University of Bern, Switzerland;

<sup>4</sup> Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

<sup>5</sup> Laboratoire des Sciences du Climat et de l'Environnement, Gif sur Yvette, France

<sup>6</sup> Department of Ecology, University of Innsbruck, Austria

<sup>7</sup> Max Planck Institute for Meteorology, Hamburg, Germany

<sup>8</sup> ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, NSW, Australia

<sup>9</sup> CSIC, Global Ecology Unit CREAM-CSIC-UAB, 08193 Bellaterra, Catalonia, Spain

<sup>10</sup> CREAM, 08193 Cerdanyola del Vallès, Catalonia, Spain

<sup>11</sup> Department of Geography, University of Augsburg, Germany

<sup>12</sup> Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

<sup>13</sup> Department of Mathematics, University of Exeter, Exeter, UK

<sup>14</sup> Department of Biological Sciences, Northern Arizona University

<sup>15</sup> Institute for Plant Sciences and Genetics, Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot Campus, Israel

<sup>16</sup> Graduate Degree Program in Ecology, Department of Biology, Colorado State University

<sup>17</sup> Dept. of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark

<sup>18</sup> Ministry of Education Key Laboratory for Earth System modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China

<sup>19</sup> Research School of Biology, Australian National University, Australia

<sup>20</sup> School of Geosciences, University of Edinburgh, United Kingdom

<sup>21</sup> Department of Civil, Construction, and Environmental Engineering, Marquette University, Milwaukee, WI, USA

<sup>22</sup> Department of Biology Sciences, University of Quebec at Montreal, Canada

<sup>23</sup> Department of Geography, Ludwig Maximilian University of Munich, Germany

<sup>24</sup> International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, USA.

<sup>25</sup> School of Plant Sciences and Food Security, Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

<sup>26</sup> Centre of Excellence PLECO (Plants and Ecosystems), Biology Department, University of Antwerp, Belgium

<sup>27</sup> CSIRO Marine and Atmospheric Research and Centre for Australian Weather and Climate Research

<sup>28</sup> Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, Devon, UK

<sup>29</sup> Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University

<sup>30</sup> 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

Preprint submitted to Global Change Biology

## 1 Abstract

2 Changes in rainfall amounts and patterns have been observed and are expected to continue in the near future  
3 with potentially significant ecological and societal consequences. Modelling vegetation responses to

4 changes in rainfall is thus crucial to project water and carbon cycles in the future. In this study, we present  
5 the results of a new model-data intercomparison project, where we tested the ability of ten terrestrial  
6 biosphere models to reproduce observed sensitivity of ecosystem productivity to rainfall changes at ten  
7 sites across the globe, in nine of which, rainfall exclusion and/or irrigation experiments had been performed.

8 The key results are:

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

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

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

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

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

## 25 **1 Introduction**

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

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

44 To understand the ecosystem responses to changes in rainfall amounts and patterns at the local scale, rainfall  
45 manipulations experiments have been conducted. Typically, such experiments change the overall rainfall  
46 amount by exclusion (Estiarte *et al.*, 2016; Martin-Stpaul *et al.*, 2013; Limousin *et al.*, 2009) or irrigation  
47 (Collins *et al.*, 2012) and responses are commonly quantified by changes in Aboveground Net Primary  
48 Production (ANPP). In some experiments such as the Amazon rainfall exclusion experiment, (Nepstad *et al.*,  
49 2007) additional detailed data quantifying changes in forest structure and composition have been  
50 obtained. There are a small number of experiments where the structure of rainfall pulses is modified e.g.  
51 (Fay *et al.*, 2008; Heisler-White, *et al.*, 2008; Vicca *et al.*, 2014). Rainfall manipulation experiments have  
52 been conducted in a range of ecosystems, spanning from semiarid shrublands (Báez *et al.*, 2013), to  
53 temperate (Hanson and Wullschleger, 2003) and tropical forests (Fisher *et al.*, 2007; Nepstad *et al.*, 2007),  
54 even though most of the experiments have focused on grasslands or low-stature vegetation due to the  
55 difficulties in setting up experiments. Those experiments have identified a strong correlation between  
56 rainfall changes and vegetation productivity (e.g. Heisler-White *et al.*, 2009; Stuart-Haëntjens *et al.*, 2018),  
57 phenology (e.g. Peñuelas *et al.*, 2004), plant community structure e.g. (Miranda *et al.*, 2011; Zhang *et al.*,  
58 2019) and belowground carbon dynamics e.g. (Vicca *et al.*, 2014; Hagedorn *et al.*, 2016; Hasibeder *et al.*,  
59 2015). Despite the important findings derived from these field experiments, these studies have strong spatial  
60 and temporal limitations; they reported only few variables and it is challenging to extrapolate information  
61 beyond the specific design of the experiment. Extrapolation and mechanistic understanding related to  
62 vegetation responses to changes in precipitation can be better achieved by combining model and data driven  
63 approaches (e.g. Kayler *et al.*, 2015).

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

79 A large discrepancy of predicted model responses has direct consequences for the uncertainties related to  
80 the fate of terrestrial carbon under a changing climate (Zscheischler, *et al.*, 2014b; Ahlström *et al.*, 2015;  
81 Humphrey *et al.*, 2018). This is the case because the terrestrial vegetation and thus the terrestrial land carbon  
82 sink introduces the largest uncertainties of the global carbon cycle (Le Quéré *et al.*, 2018). In this context,  
83 large epistemic model uncertainties can have considerable impacts on our ability to forecast the growth rate  
84 of atmospheric CO<sub>2</sub>. Additionally, vegetation responses to water stress can influence land-atmosphere  
85 coupling (Koster, 2004; Seneviratne *et al.*, 2013; Lemordant *et al.*, 2016; Gentine *et al.*, 2019), since  
86 vegetation cover and canopy conductance affect land surface energy balance. This will have a large impact  
87 on our skill to model the coupled hydrological, plant physiological and meteorological processes and thus  
88 robustly projecting climate change (Miralles *et al.*, 2018).

89 To reduce this source of epistemic uncertainty and understand the reasons for model disagreement, a  
90 detailed comparison between the responses of different modelling schemes with respect to plant water  
91 availability is essential. Rainfall manipulation experiments assessing vegetation responses to water  
92 limitation are particularly useful in this regard. Arguably, this is an extremely important test to evaluate the  
93 structure and parameter values of a model and its capability to reproduce responses to environmental  
94 changes. A model should be able to reproduce the observed dynamics under control and manipulated  
95 conditions in order to be considered robust, especially for climate change simulations (Medlyn *et al.*, 2015).  
96 Despite the importance of this comparison, there are only few examples that have compared terrestrial

97 biosphere models and global change manipulation experiments (De Kauwe *et al.*, 2013, 2017; Fatichi and  
98 Leuzinger, 2013; Powell *et al.*, 2013; Zaehle *et al.*, 2014; Medlyn *et al.*, 2015). Recently, Wu *et al.* (2018)  
99 compared 14 models under different idealized rainfall scenarios for three grassland experiments sites and  
100 showed a fair reproduction of spatial sensitivities of ANPP to rainfall but large differences in the modelled  
101 asymmetric response of ANPP to interannual i.e. temporal rainfall variability at a given site. Wu *et al.*  
102 (2018) were not able to evaluate the modelled responses with respect to actual experiments because they  
103 used idealized rainfall changes that did not exactly mimic the site treatments. In this study we perform such  
104 an evaluation. We make use of ten sites with diverse climates and biomes, where multi-year rainfall  
105 manipulation experiments took place to evaluate ten terrestrial biosphere models, representing an  
106 unprecedented data-model intercomparison effort focused on ecosystem responses to water limitation. This  
107 data-model intercomparison will address the following questions: (a) Can models reproduce the observed  
108 responses to precipitation variability at rainfall manipulation sites? (b) Do models accurately reproduce the  
109 spatial (across-sites) and temporal (within-site) dependence of vegetation productivity to precipitation? (c)  
110 Which are the underlying reasons for model disagreement? Answering those questions will provide insights  
111 on the robustness of Earth System model projections with respect to the global carbon cycle.

## 112 **2 Data and Methods**

### 113 **2.1 Sites**

114 Ten different sites with contrasted climates and biomes and sufficiently long records were considered here.  
115 For all analyses presented in this study, the sites are termed: Lahav, Matta, SGS, Prades, Garraf, Konza  
116 (AmeriFlux ID: US-Kon), Puèchabon (FluxNet ID: FR-Pue), Brandbjerg, Walker Branch (Walker Branch;  
117 AmeriFlux ID: US-WBW) and Stubai (Table 1). The sites are in ascending order in terms of wetness index  
118  $WI$  defined as the average ratio of annual precipitation to annual potential evapotranspiration during the  
119 study period. For our analysis the sites are split in three wetness categories ( $WI < 0.4$  [Lahav, Matta,  
120 SGS];  $0.4 \leq WI < 1$  [Prades, Garraf, Konza, Puèchabon];  $WI \geq 1$  [Brandbjerg, Walker Branch, Stubai]).

121 The sites are in the USA (Konza, SGS, Walker Branch), Israel (Lahav, Matta), Spain (Garraf), France  
122 (Prades, Puèchabon), Austria (Stubai) and Denmark (Brandbjerg) and span a precipitation gradient from  
123 253-1440 mm y<sup>-1</sup> and include grasslands shrublands and forested ecosystems (Table 1). In eight sites  
124 rainfall exclusion experiments were carried out, and in four irrigation experiments. The experiment duration  
125 considered in this study was from 5 up to 32 years. The average experiment duration was 13.3 years.

126

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
<b>Lahav</b>	34.9/31.38	19.1	253	0.19	590	Annual grasses and shrubs, mostly <i>Sarcopoterium spinosum</i>	22.6% Sand, 39.7% Silt, 37.7% Clay	-30% rainfall for the entire year	+30% rainfall for the entire year	2002-2014	Tielbörger <i>et al.</i> , 2014
<b>Matta</b>	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 <i>et al.</i> , 2014
<b>SGS</b>	-104.75/40.81	8.4	304	0.35	1650	C4 grasses, primarily ( <i>Bouteloua gracilis</i> (H.B.K.) Lag. Ex Steud., <i>Buchloe dactyloides</i> (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 <i>et al.</i> , 2009
<b>Prades</b>	0.91/41.21	11.43	522	0.4	950	Mixed composition of <i>Quercus ilex</i> L., <i>Phillyrea latifolia</i> L., <i>Arbutus unedo</i> L., <i>Erica arborea</i> L., <i>Juniperus oxycedrus</i> L., <i>Cistus albidus</i> L. <i>Sorbus torminalis</i> (L.) Crantz and <i>Acer monspessulanum</i> L.	48% Sand, 32% Silt and 20% Clay	-20% rainfall for the entire year	None	1999-2012	Ogaya and Peñuelas, 2007
<b>Garraf</b>	1.82/ 41.3	15.04	580	0.48	210	<i>Erica multiflora</i> , <i>Globularia alypum</i>	41% Sand, 39% Silt and 18% Clay,	-50% in spring and fall	None	2000-2004	Beier <i>et al.</i> , 2009
<b>Konza</b>	-96.6/ 39.1	12.8	830	0.7	342	Mixed C3( <i>Solidago canadensis</i> , <i>Aster ericoides</i> , <i>Salix missouriensis</i> ) C4( <i>Andropogon gerardii</i> , <i>Sorghastrum nutans</i> , <i>Panicum virgatum</i> ) Grassland	10% Sand, 35% Clay	None	irrigation +20% was provided at two sites termed lowland and upland	1982-2013	Collins <i>et al.</i> , 2012
<b>Puèchabon</b>	43.74/3.6	13.8	969	0.87	270	Overstory ( <i>Quercus ilex</i> ); Understory ( <i>Buxus sempervirens</i> , <i>Phyllirea latifolia</i> , <i>Pistacia terebinthus</i> and <i>Juniperus oxycedrus</i> )	26% Sand, 35% Silt, 39% Clay	-30% throughfall exclusion for the entire year	None	2004-2013	Limousin <i>et al.</i> , 2009
<b>Brandbjerg</b>	11.97/55.89	9.59	757	1.1	39	70% grasses (mostly <i>Deschampsia flexuosa</i> ); 30% dwarf shrubs ( <i>Calluna vulgaris</i> )	88-95% Sand, 2-9% Silt, 1-2% Clay	rainfall exclusion for 4-6 weeks during spring and summer	None	2007-2012	Kongstad <i>et al.</i> , 2012

<b>WB</b>	-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 <i>et al.</i> , 2004
<b>Stubai</b>	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 <i>et al.</i> , 2014; Hasibeder <i>et al.</i> , 2015

128

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

## 137 **2.2 Participating models and simulation protocol**

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

153 photosynthetic rates (T&C, JULES, TECO), the maximum rate of carboxylation  $V_{cmax}$  (ORCHIDEE, ORC  
154 MICT, ORC CNP), stomatal conductance (JSBACH, DLEM), or a combination of all such parameters  
155 (CABLE), based on plant water availability. LPX uses a supply and demand driven approach to water  
156 limitation. If water demand exceeds supply, photosynthesis is downregulated until they match. None of the  
157 models simulates plant hydraulics and thus xylem cavitation in response to water stress.

158 For each site, we conducted a control simulation corresponding to the observed climate without  
159 manipulation, and simulations representative of each rainfall manipulation experiment (rainfall exclusion  
160 and/or irrigation) with the same timing and magnitude of water input as in the real experiment. For all  
161 experiments the common data distributed to all modelling groups included hourly values of incoming  
162 shortwave and longwave radiation, vapour pressure deficit, air temperature, wind speed, atmospheric  
163 pressure and ambient CO<sub>2</sub> concentration. Model set-up was performed by each modelling group separately  
164 based on common information for each site that included, apart from the meteorological input, species  
165 composition, vegetation cover, soil and root depth and soil textural properties. Reported model outputs  
166 included gross primary productivity, net primary productivity and aboveground net primary productivity  
167 (GPP, NPP, ANPP), evapotranspiration (ET) and its partition in evaporation (soil evaporation plus  
168 evaporation from interception) and transpiration respectively, soil moisture, leaf area index (LAI) and  
169 biomass carbon pool (below and above ground) dynamics. Some models additionally reported the water  
170 stress factor ( $\beta$ ) used in the model.  $\beta$  is a model parameter that quantifies the effects of plant physiological  
171 stress due to limitations in soil water availability.  $\beta$  is not identical between models and the description of  
172 the  $\beta$  meaning for each model can be found at the supplementary material of Wu *et al.* (2018). Initial  
173 conditions for all simulations were obtained after a spin-up period long enough to equilibrate the  
174 biogeochemical cycles.

## 175 **2.3 Statistical Analyses**

### 176 ***Data-Model Comparison***

177 First, we compare the models' ability to accurately reproduce the relationship between ANPP and  
178 precipitation (P) across sites (i.e. spatial dependence) and within each site (i.e. temporal dependence) at the  
179 annual scale. At all sites, observations of ANPP were based on biomass estimates (e.g. using above ground  
180 biomass harvesting for grasslands, and a carbon budget approach for forested sites combining litterfall  
181 observations with allometric equation for aboveground biomass growth) rather than carbon fluxes, therefore  
182 discrepancy between observed and modelled ANPP is expected (detailed bias quantification are reported in  
183 the Supplementary Material).

184 Model skill in reproducing the spatial dependence of ANPP to P was quantified as the root mean squared  
185 error (RMSE) and the coefficient of determination ( $R^2$ ) between the modelled and observed annual ANPP,  
186 averaged over the entire period, across sites for the control case. Model performance in capturing the  
187 magnitude of interannual variability of ANPP was assessed by comparing the standard deviation ( $\sigma$ ) of  
188 annual ANPP between models and observations for all sites. Model skill with respect to single-site  
189 interannual dependence of ANPP to P was quantified using an estimate of the sensitivity of annual ANPP  
190 to annual P. Specifically, we fitted a linear model  $ANPP = a_0 + a_1P + a_2T$ , where  $P$  is annual  
191 precipitation and  $T$  annual temperature. To increase the sample size and robustness of the fit, precipitation  
192 from both the control and the rainfall manipulation experiments were used. Additional covariates such as  
193 vapour pressure deficit and radiation could not be added due to the small sample size, making the linear fit  
194 over constrained. Preliminary analyses (not reported here) showed that  $P$  and  $T$  were the most important  
195 covariates. Model skill was evaluated by estimating the differences between observed and simulated  
196 sensitivities of ANPP with respect to  $P$  (i.e.  $a_1 = \frac{\partial ANPP}{\partial P}$ ). Observation uncertainty of the sensitivity metric  
197 was quantified as the 90% confidence interval of the linear model fit.

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

203 Responses due to rainfall manipulation were quantified at the annual scale using the response ratio for a  
204 variable  $X$  (e.g. ANPP) defined as the ratio  $RR = X_M^{(y)} / X_C^{(y)}$ , where the subscript  $M$  stands for manipulation  
205 and  $C$  for the control scenario.  $(y)$  indicates the annual scale. In this study, we focused on the simulated  
206 RRs of ANPP and ecosystem water use efficiency (WUE) calculated at the annual scale as the ratio of  
207 annual gross primary productivity (GPP) to annual actual evapotranspiration (ET). To quantify whether the  
208 simulated response ratios have a statistically significant different mean value from the observations, a two-  
209 sample t-test was performed for every model and the respective observed responses. For the two-sample t-  
210 test, the sample size for each site is equal to the number of years in the observations and simulations.  
211 Response ratios were assumed normally distributed and independent at the annual scale. The test's null  
212 hypothesis was that modelled and observed response ratios have the same mean. The analysis was also  
213 performed using the commonly used logarithm of  $RR$  yielding identical results, and thus not further shown  
214 here.

215 **Model agreement**

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

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

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

225 
$$ANPP = A_b \cdot CUE \cdot GPP_u \cdot \beta \cdot LAI_p \cdot LAI_r.$$

226 The term  $A_b$  is the aboveground fraction of carbon allocation,  $CUE$  is the carbon use efficiency,  $GPP_u$  is a  
 227 potential (unstressed) rate of GPP per unit of leaf area,  $\beta$  is the annually averaged value of the water stress  
 228 factor,  $LAI_p$  is the peak  $LAI$  during the year, and  $LAI_r$  is a proxy of the growing season length, defined as  
 229 the integral of  $LAI$  during the year divided by  $LAI_p$ . Considering that water stress and  $LAI$  dynamics,  
 230 determine most of the interannual variation of ANPP, assuming that  $A_b$ ,  $CUE$ , and  $GPP_u$  vary less between  
 231 treatments, then, the annual response ratio of ANPP can be estimated by the response ratios of  $\beta$ ,  $LAI_p$  and

232  $LAI_r$  (e.g.  $\frac{ANPP_M^{(y)}}{ANPP_C^{(y)}} \approx \frac{\beta_M^{(y)}}{\beta_C^{(y)}} \cdot \frac{LAI_{pM}^{(y)}}{LAI_{pC}^{(y)}} \cdot \frac{LAI_{rM}^{(y)}}{LAI_{rC}^{(y)}}$ , where the subscript  $M$  stands for manipulation and  $C$  for the  
 233 control scenario and  $(y)$  indicates the annual scale). If the response ratios of  $\beta$  and  $LAI_p$  and  $LAI_r$  are  
 234 independent at the annual scale, then

235 
$$\overline{\left(\frac{ANPP_M^{(y)}}{ANPP_C^{(y)}}\right)} \approx \overline{\left(\frac{\beta_M^{(y)}}{\beta_C^{(y)}}\right)} \cdot \overline{\left(\frac{LAI_{pM}^{(y)}}{LAI_{pC}^{(y)}}\right)} \cdot \overline{\left(\frac{LAI_{rM}^{(y)}}{LAI_{rC}^{(y)}}\right)},$$

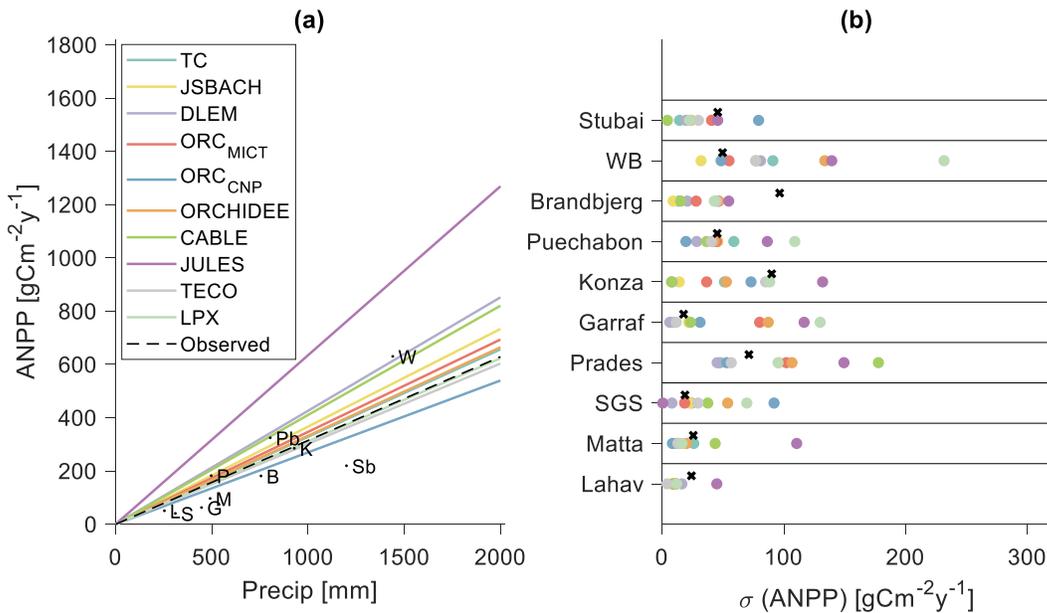
236 where overbars indicate average values for all years. This approximation is well supported by the results of  
 237 our simulations (Supplementary material, Figure S6), even though data evidence suggests that  $CUE$  may  
 238 change significantly under changes in water stress (Rowland *et al.*, 2014). Using this decomposition in the  
 239 model results, the average ANPP response ratio can be decomposed as the product of the average response  
 240 ratios of  $\beta$ ,  $LAI_r$ ,  $LAI_p$ . Based on these considerations, we can attribute the changes of the modelled  $ANPP$   
 241 among models to differences in simulated water stress,  $LAI$  dynamics, and phenological changes. Since  
 242 only six (T&C, CABLE, JULES, TECO, DLEM, JSBACH) of the ten participating models reported the

243 water stress  $\beta$  factor, this analysis was performed using this subset of models. All statistical analyses were  
 244 performed in MATLAB 2019a.

### 245 3 Results

#### 246 3.1 Control Scenario

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



255

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

264

265

266

267

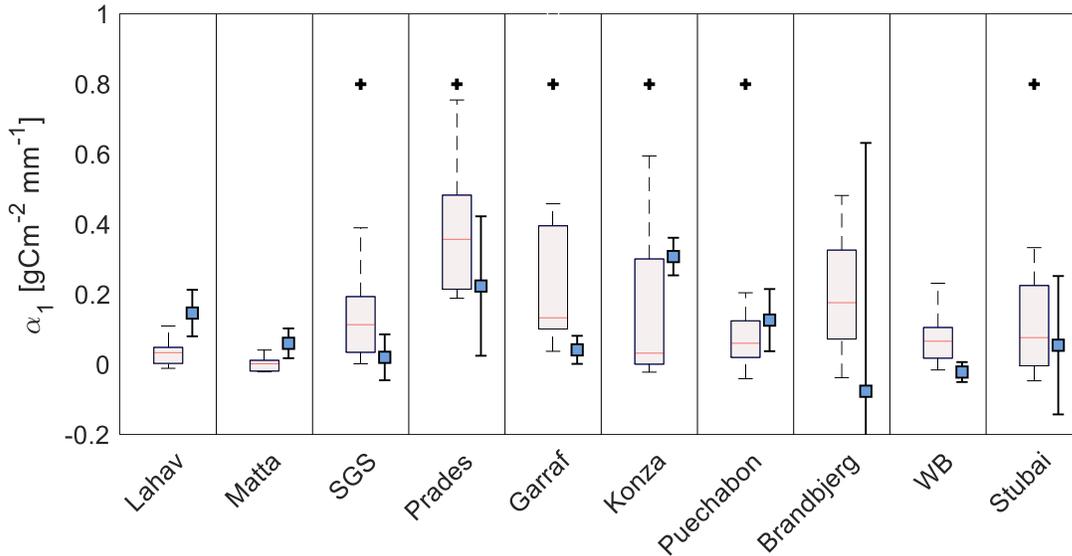
268

269

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 <sup>-2</sup> y <sup>-1</sup> )	ANPP - Normalized RMSE [-]	ANPP - R <sup>2</sup> [-]	ANPP - Bias (gCm <sup>-2</sup> y <sup>-1</sup> )	σ(ANPP) - Bias (gCm <sup>-2</sup> y <sup>-1</sup> )	LAI - RMSE [m <sup>2</sup> m <sup>2</sup> ]	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

270



271

272

273

274

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

275 *observed ANPP to precipitation (blue squares) and the corresponding 90% confidence*  
 276 *intervals (bar length) of the fit of the linear model. Crosses indicate the sites for which*  
 277 *the mean value of the distribution of simulated sensitivities is not statistically different*  
 278 *from the observed with 90% confidence. Sites are ranked from left to right in order of*  
 279 *ascending wetness.*

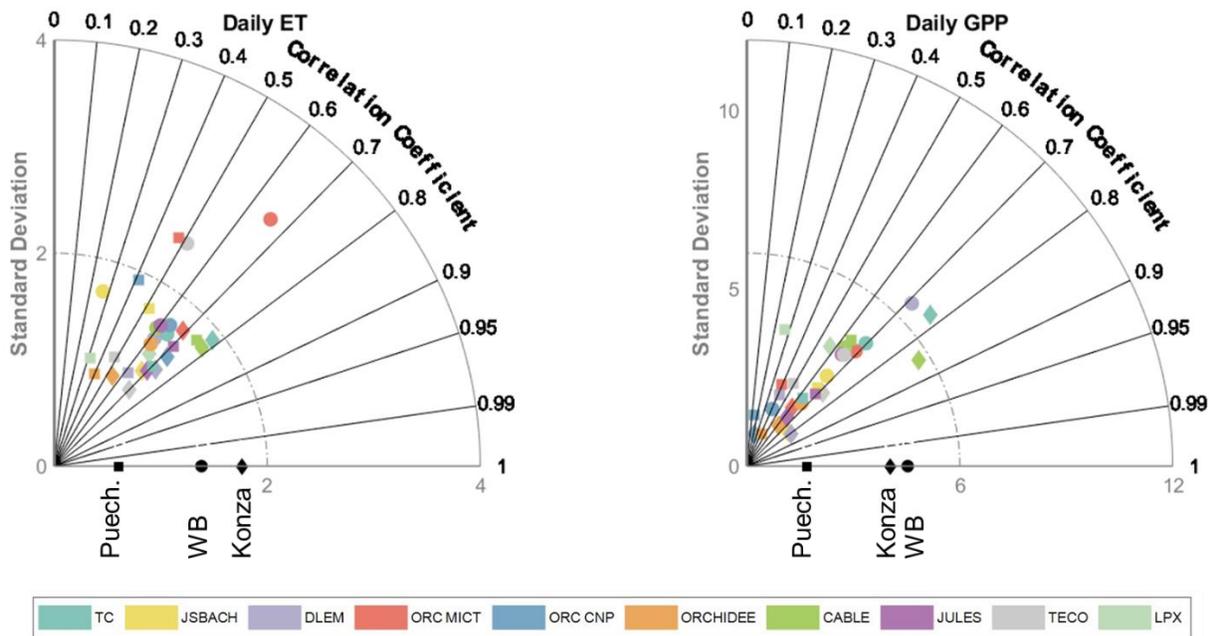
280 Both models and observations support a larger sensitivity of annual ANPP to interannual variation in  
 281 precipitation at sites with intermediate wetness conditions (e.g. Garraf, Prades, Puèchabon, Konza; Figure  
 282 2). Specifically, in sites with a wetness index  $WI < 0.4$  models(observations) have mean sensitivity  $\bar{a}_1 =$   
 283  $0.058(0.076) \text{ gCm}^{-2}\text{mm}^{-1}$ , in sites with  $0.4 \leq WI < 1$ ,  $\bar{a}_1 = 0.22(0.18) \text{ gCm}^{-2}\text{mm}^{-1}$  and in sites with  
 284  $WI > 1$ ,  $\bar{a}_1 = 0.13(-0.013) \text{ gCm}^{-2}\text{mm}^{-1}$ . At the most arid sites, annual precipitation explains a large  
 285 fraction of the observed and modelled variability of annual ANPP, but the sites are not highly productive  
 286 (i.e. Absolute productivity values are low; Figure 1), yielding a low average sensitivity  $a_1$ . At the opposite  
 287 end, mesic sites have higher productivity, but they are not water limited during the observation period,  
 288 resulting also in a low modelled sensitivity  $\bar{a}_1$ . Modelled sensitivity uncertainty was largest for intermediate  
 289 precipitation regimes due to a larger model disagreement for those sites. For sites with a  $WI < 0.4$ , the  
 290 average uncertainty, quantified here as the standard deviation between models of modelled  $a_1$  was  
 291  $\sigma_{a_1|dry} = 0.08 \text{ gCm}^{-2}\text{mm}^{-1}$ , for intermediate sites  $\sigma_{a_1|inter} = 0.24 \text{ gCm}^{-2}\text{mm}^{-1}$  and for wet sites  
 292  $\sigma_{a_1|wet} = 0.14 \text{ gCm}^{-2}\text{mm}^{-1}$ .

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

302 Simulated daily ET for the control simulations was substantially different regarding its day-to-day  
 303 variability from measured ET at all three eddy sites. Correlation coefficients were in the range 0.27-0.78  
 304 with an average value between all models and sites of  $\sim 0.60 \pm 0.13$  (mean  $\pm$  standard deviation) (Figure 3).  
 305 Simulated variability of ET, expressed in terms of standard deviation at the daily scale, deviated  
 306 substantially from the measured variability of ET. In particular, simulated variability from most models  
 307 was lower than observed at Konza (observed  $\sigma_{ET} = 1.76 \text{ mm d}^{-1}$ , modelled  $\sigma_{ET} = 1.40 \pm 0.3 \text{ mm d}^{-1}$ ),

308 and higher than observed at Puèchabon (observed  $\sigma_{ET} = 0.61 \text{ mm d}^{-1}$ , modelled  $\sigma_{ET} = 1.86 \pm 0.50$   
 309  $\text{mm d}^{-1}$ ). For WB, the modelled ET variability was higher than observed, and inter-model agreement was  
 310 low (observed  $\sigma_{ET} = 1.39 \text{ mm d}^{-1}$ , modelled  $\sigma_{ET} = 1.51 \pm 0.45 \text{ mm d}^{-1}$ ). Seasonality of ET was well  
 311 reproduced by all models (Figure S2), partially explaining the high correlation coefficients (Figure 3). One  
 312 pronounced exception is in Puèchabon, where the observed late summer reduction of ET and increase in  
 313 early fall was reproduced only by a small subset of models (Figure S2).

314 Simulated daily GPP had a correlation ( $\sim 0.59 \pm 0.17$ ) with observed daily GPP for all models (Figure 3). A  
 315 large fraction of the GPP correlation can be attributed to seasonality. However, the modelled variability  
 316 was significantly different from the observed for all sites. Most models underestimated the daily variation  
 317 of GPP at Konza (observed  $\sigma_{GPP} = 4.04 \text{ gCm}^{-2}\text{d}^{-1}$ , modelled  $\sigma_{GPP} = 2.87 \pm 1.88 \text{ gCm}^{-2}\text{d}^{-1}$ ) and WB  
 318 (observed  $\sigma_{GPP} = 4.53 \text{ gCm}^{-2}\text{d}^{-1}$ , modelled  $\sigma_{GPP} = 4.01 \pm 1.26 \text{ gCm}^{-2}\text{d}^{-1}$ ) and overestimated the  
 319 variability of daily GPP at Puèchabon (observed  $\sigma_{GPP} = 1.68 \text{ gCm}^{-2}\text{d}^{-1}$ , modelled  $\sigma_{GPP} = 2.67 \pm$   
 320  $1.01 \text{ gCm}^{-2}\text{d}^{-1}$ ) (Figure 3). Large model differences between observed and simulated GPP can be  
 321 partially attributed to an incorrect representation of the magnitude of LAI. There is, indeed, a large  
 322 disagreement between the modelled LAI across models (Figure 4). Modelled LAI is also significantly  
 323 different than observed, even though LAI derived via remote sensing is also uncertain (Fang *et al.*, 2013).

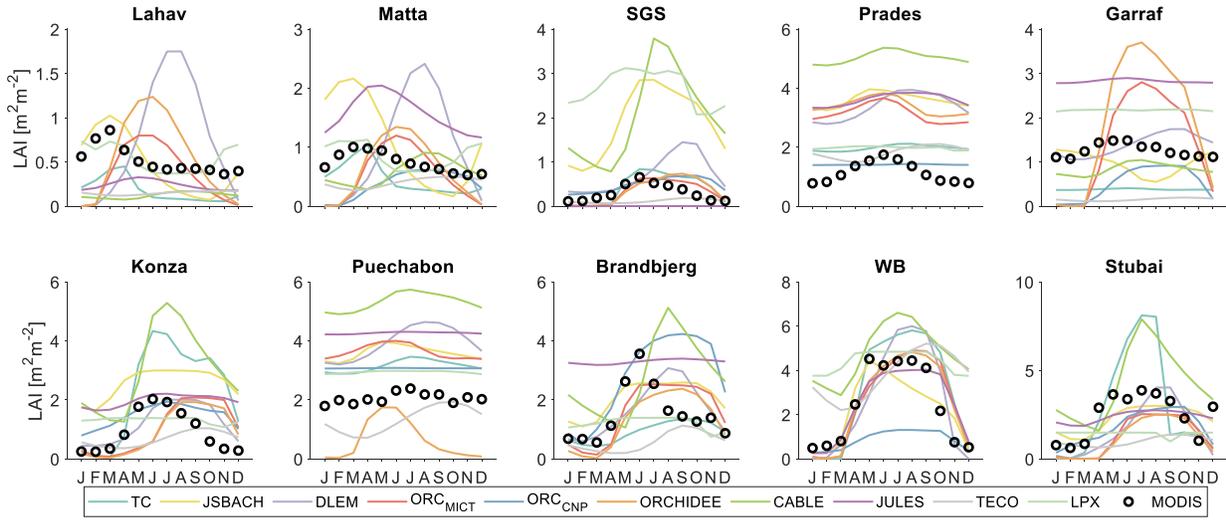


324

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

329  
330

*reproducing precisely the data) would lie on the black markers, each corresponding to different sites.*



331

332  
333  
334

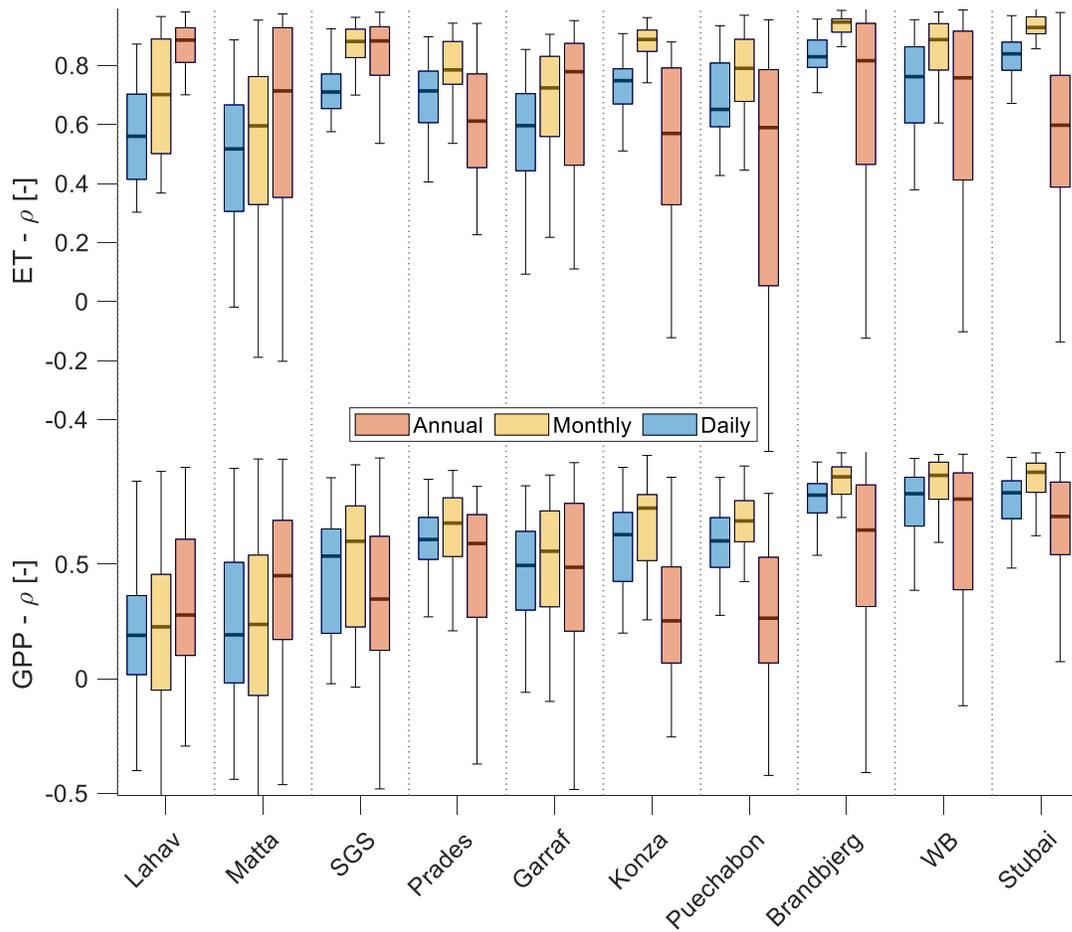
*Figure 4: Simulated average monthly LAI by all models for all sites for the control case simulation. Dots indicate the long-term monthly LAI averages of the nearest MODIS pixel in the area.*

335  
336  
337  
338  
339  
340  
341  
342  
343

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  $\rho_{ET|dry}^y = 0.75$ ; for GPP  $\rho_{GPP|dry}^y = 0.35$ ). 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 ( $\rho_{ET|dry}^d = 0.58$ ,  $\rho_{GPP|dry}^d = 0.30$ ). The opposite picture occurs in mesic sites ( $WI > 1$ ), where models agree better at the daily time scale for ET ( $\rho_{ET|wet}^d = 0.79$ ), but their agreement is significantly lower at the annual scale ( $\rho_{ET|wet}^y = 0.61$ ). A similar pattern is also valid for GPP ( $\rho_{GPP|wet}^d = 0.77$ ;  $\rho_{GPP|wet}^y = 0.60$ ) (Figure 5).

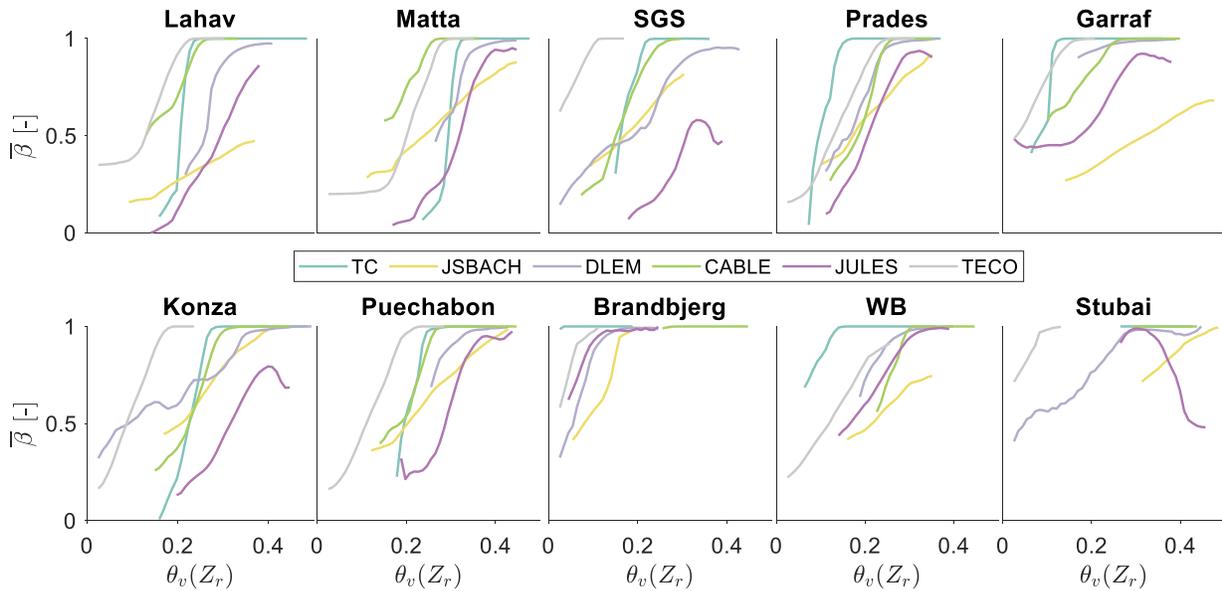
344  
345  
346  
347

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).



348

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



352

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

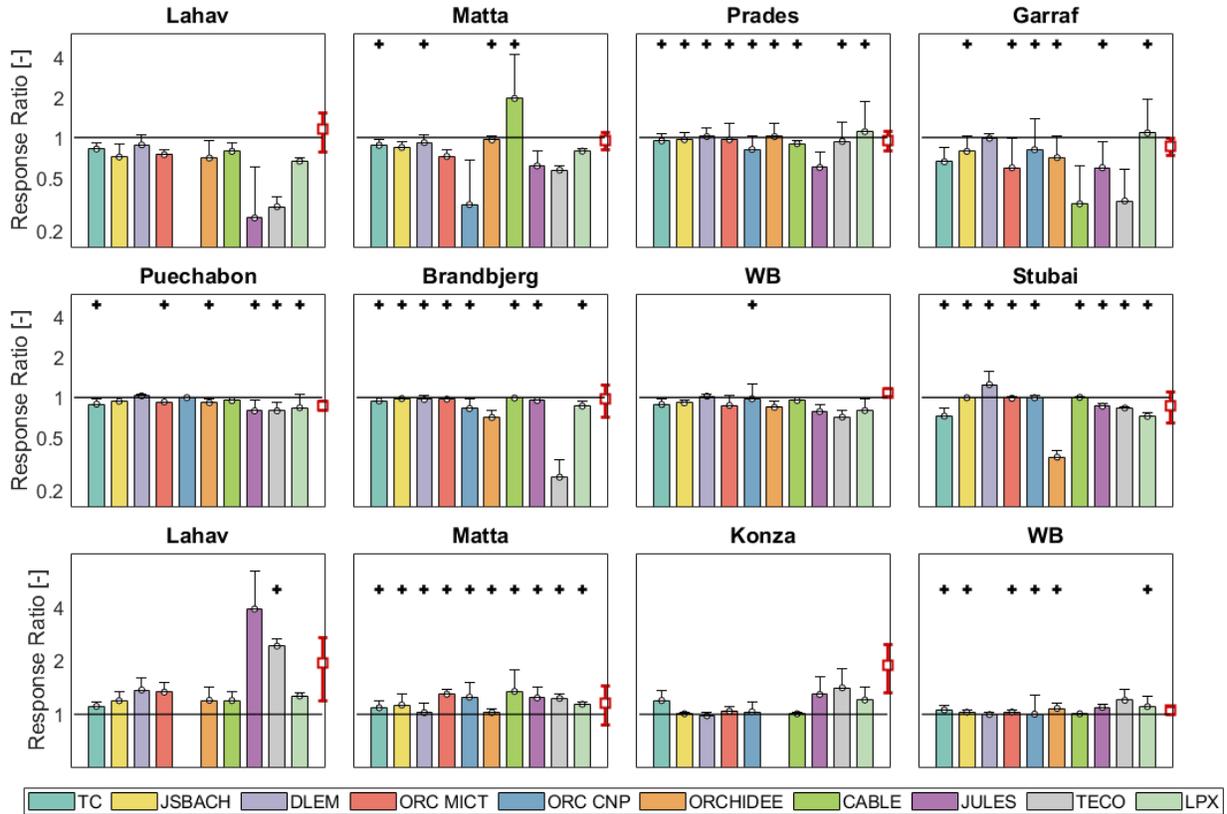
### 356 3.2 Manipulation Experiments

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

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

374 the observed for the rainfall exclusion experiments, and lower for the irrigation experiments (for the drought  
 375 experiments: average modelled standard deviation of the response ratios was  $\sigma_{R_D}^m = 0.18$ ; and observed  
 376  $\sigma_{R_D}^o = 0.178$ . For irrigation experiments modelled standard deviation was  $\sigma_{R_I}^m = 0.25$ ; and observed  
 377  $\sigma_{R_I}^o = 0.42$ ). Outliers with regards to both the magnitude and the interannual variability of response ratios  
 378 occurred for the most water-limited sites.

379

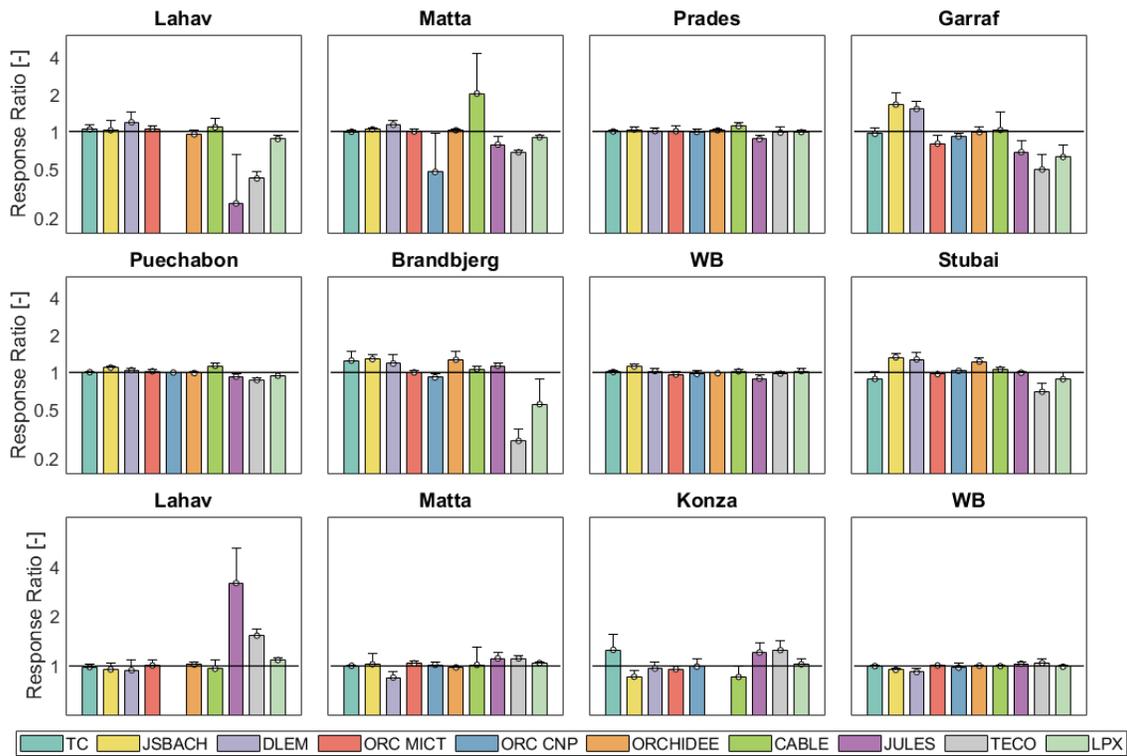


380

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

389 Besides carbon assimilation, changes in rainfall can simultaneously modify ET and thus the land surface  
 390 energy balance. The coupling between ET and GPP depends heavily on the parametrizations of water stress  
 391 and how this affects stomatal conductance and the reduction of photosynthesis. It further depends on

392 vegetation dynamics such as a shift of carbon allocation from leaves to roots or leaf shedding due to water  
 393 stress. To quantify the responses of the ET and GPP coupling, we compute the relative changes of water  
 394 use efficiency (WUE) for the various cases (Figure 8). Most models predict relatively small changes in  
 395 WUE (i.e.  $R \sim 1$ ) for both drought ( $R_D^m = 0.98$ ) and irrigation ( $R_I^m = 1.08$ ) treatments, indicating a change  
 396 of comparable magnitudes for both ET and GPP. CABLE, JULES and TECO occasionally simulate larger  
 397 changes, in both positive and negative directions, in WUE for the most water limited sites. This larger  
 398 change can be attributed to a more sensitive response of GPP to water stress than ET.



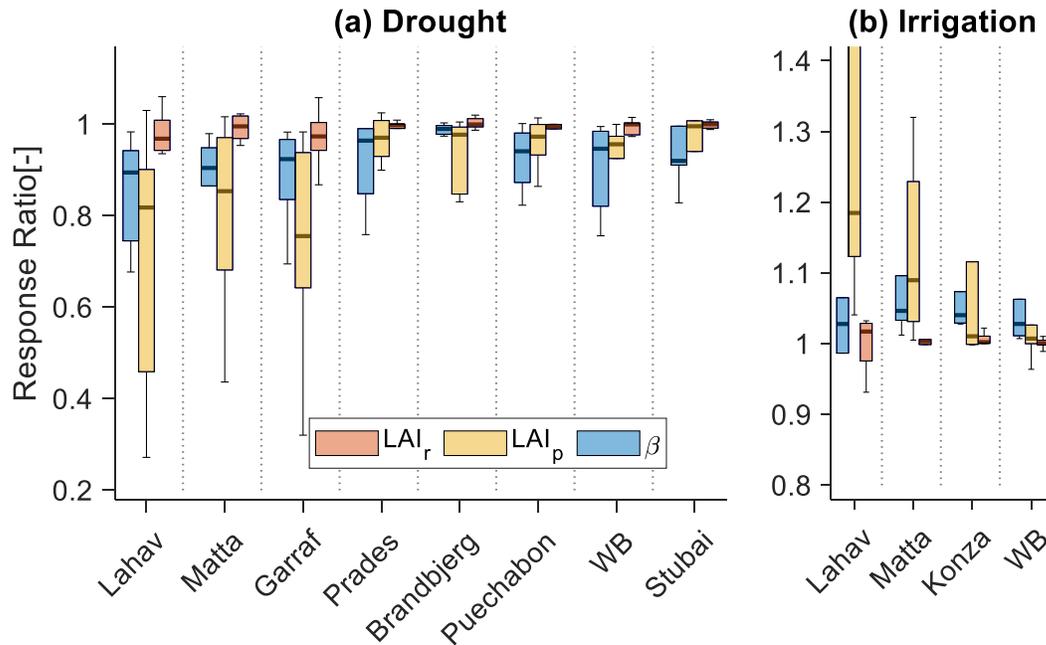
399  
 400 *Figure 8: Simulated response ratios of water use efficiency during treatment period due to rainfall exclusion (rows 1 and 2) and addition (irrigation) (row 3). Different*  
 401 *models are presented with different colors according to the legend. Error bars represent*  
 402 *the standard deviation for all years of treatment.*  
 403

404

### 405 3.3 Response attribution

406 We partitioned the total response ratio of ANPP into relative changes of (a) the  $\beta$  stress factor; (b) peak  
 407 LAI ( $LAI_p$ ); and (c) the length of the growing season approximated by  $LAI_r$  (Figure 9). Changes in  
 408 simulated ANPP following rainfall manipulation can be almost exclusively attributed to changes in  $\beta$  and

409  $LAI_p$ . The response ratio of  $LAI_r$  was always close to unity ( $R_{LAI_r} = 0.98 \pm 0.058$  (mean $\pm$ standard  
410 deviation) for the drought treatment and  $R_{LAI_r} = 1.01 \pm 0.029$  for the irrigation treatment) contributing  
411 insignificantly to the response ratio of ANPP. Thus, no model predicted substantial changes in the length  
412 of the growing season. A reduction or enhancement of  $\beta$  for the drought and irrigation experiments  
413 explained the largest fraction of ANPP responses at wet sites, but the uncertainty of the relative strengths  
414 of changes in  $\beta$  and  $LAI_p$  was high (Drought treatment for sites with  $WI > 1$ ,  $R_\beta = 0.92 \pm 0.09$ ,  $R_{LAI_p} =$   
415  $0.95 \pm 0.11$ ; Irrigation treatment for sites with  $WI > 1$ ,  $R_\beta = 1.05 \pm 0.06$ ,  $R_{LAI_p} = 1.02 \pm 0.02$ ). For the  
416 driest sites both  $LAI_p$  and  $\beta$  explained a large fraction of the total response for the drought treatment,  
417 whereas  $LAI_p$  was the dominant and simultaneously the most uncertain factor for the irrigation treatment  
418 (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  
419 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  
420 responses of both  $\beta$  and  $LAI_p$  amongst models was high as indicated by the standard deviations above. At  
421 the sites where rainfall exclusion was applied only in part of the year (Garraf, Brandbjerg) the response  
422 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  
423 large variability amongst models, it is not possible to conclude if this is a true signal. The variability was  
424 higher for the most water stressed sites, primarily because for those sites model disagreement on the  
425 estimated response ratio of ANPP was also the highest.



427

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

## 431 4 Discussion

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

433 Most models overestimated the relationship between mean annual precipitation and average annual ANPP  
 434 observed across sites, but managed to capture well the overall trend, despite large site differences in terms  
 435 of vegetation coverage and overall climatic regime (Figure 1). This result confirms that terrestrial biosphere  
 436 models can capture spatial gradients of vegetation productivity relatively well (e.g. Wu *et al.*, 2018).  
 437 Reproducing local (single-site) response of ANPP to interannual precipitation variability has been generally  
 438 found to be more challenging (Fatichi and Ivanov, 2014). In fact, previous intercomparison studies have  
 439 found that models have significant biases at various time scales, from subdaily (Matheny *et al.*, 2014) to  
 440 decadal (Dietze *et al.*, 2011). Dietze *et al.*, (2011) found model errors to be largest at the annual scale. In  
 441 agreement with such a result in our experiment, models differed greatly in their simulated sensitivity of  
 442 local scale productivity to annual precipitation but were able to reproduce the previously reported stronger  
 443 spatial than temporal sensitivity of productivity to rainfall. A large model disagreement with regards to the  
 444 magnitude of the interannual variability of ANPP also confirms the previously found difficulties of models  
 445 to properly capture carbon dynamics at the annual scale (e.g. Dietze *et al.*, 2011, Paschalis *et al.*, 2015).

446 Despite large model disagreement we found that the within site sensitivity of ANPP to precipitation is lower  
447 than across site sensitivity of ANPP to average precipitation, in agreement with a number of previous  
448 observational (Goward and Prince, 1995; Knapp and Smith, 2001; Huxman et al., 2004) and modelling  
449 results (Fatichi and Ivanov, 2014; Wu et al., 2018).

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

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

476 In intermediate wetness sites ( $0.4 \leq WI < 1$ ), in our simulations, models disagree at intermediate scales  
477 (weeks-months) in terms of GPP (consistent with the wavelet coherence analysis presented at Figure S7).

478 As mentioned before, at short (daily) temporal scales, temperature and radiation mostly determine water  
479 and carbon fluxes, when water is not a strong limiting factor, and due to the similar parametrizations among  
480 models (Wu *et al.*, 2018), we detect a substantial convergence in GPP. However, since such controls “fade”  
481 with increasing temporal scales, the effects of features linked to soil moisture dynamics, such as the soil  
482 moisture retention after a rainfall event, can manifest at longer temporal scales (Paschalis *et al.*, 2015).  
483 Those dynamics can be influenced by factors including both biotic and abiotic factors such as the  
484 parametrizations of soil properties that determine the temporal dynamics of soil moisture and the vertical  
485 distribution of root biomass, affecting how plants withdraw water from the soil. In fact, models were found  
486 to strongly disagree on how plants are affected by soil moisture (biotic factor – Figure 6) and on the soils’  
487 water holding capacity, as indicated by the range of accessible values of soil moisture (abiotic factor –  
488 Figure 6).

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

493 All those behaviours highlight further the need to correctly capture water/carbon dynamics at multiple time  
494 scales, from the scale of the individual rain pulse (Huxman *et al.*, 2004a) up to interannual scales where  
495 drought legacies can have an important effect (Anderegg *et al.*, 2015). The need to understand in detail  
496 multi-scale dynamics linked to water stress and soil moisture dynamics is also exacerbated by the fact that  
497 model disagreement in terms of the sensitivity of ANPP to annual rainfall is highest for sites with  
498 intermediate wetness ( $0.4 \leq WI < 1$ ). Those regions experience moderate water limitations, and the impact  
499 of water limitation to fast acting processes (changes in e.g. stomatal conductance, photosynthesis) can  
500 accumulate and impact longer time scales through slow acting processes (e.g. changes in LAI).  
501 Additionally, areas with intermediate wetness are expected to operate close to soil moisture thresholds  
502 inducing plant water stress. Sensitivity of the responses of ANPP to precipitation in those sites is  
503 concurrently the highest and most uncertain (Figure 2). This can have a large impact on our ability to model  
504 the fate of terrestrial  $CO_2$ , given that those areas are amongst the largest contributors to the interannual  
505 dynamics of the growth rate of  $CO_2$  (Poulter *et al.*, 2014; Ahlström *et al.*, 2015). Understanding such  
506 dynamics across scales requires high quality and high frequency long-term measurements, not only for  $CO_2$   
507 and water fluxes but also soil moisture dynamics (Vicca *et al.*, 2012). Annual ANPP values alone are  
508 limiting our inference capabilities and even 10-20 years of annual ANPP data were not long enough to  
509 obtain a precise estimate of the sensitivity of ANPP to precipitation.

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

### 529 ***Response to manipulation experiments***

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

535 Even though, multiple models generated close vegetation productivity responses in the rainfall exclusion  
536 experiments, the underlying reasons are very different and at the same time highly uncertain (Figure 9). In  
537 the more water-limited ecosystems, both changes in LAI magnitude and the level of plant water limitation  
538 determine productivity responses. Variability of the relative strength of  $\beta$  and  $LAI_p$  between models is  
539 large. Variability concerning  $LAI_p$  is larger than  $\beta$ , which can be explained by the fact that  $LAI_p$  integrates  
540 the model differences related to LAI phenology, carbon allocation rules, and reductions in photosynthetic  
541 rates due to soil moisture limitations. Pinpointing which model best captures the relative strengths of  
542 changes in  $\beta$  and  $LAI_p$  would require simultaneous high frequency data, including soil moisture, regular

543 measurements of stomatal conductance and leaf water potentials, high frequency photosynthetic rates, and  
544 regular LAI estimates. At more mesic sites, physiological effects of water stress (through  $\beta$ ) are the main  
545 reason for productivity responses. The reason is that in such sites, induced water stress is mild. Productivity  
546 will be reduced during the imposed water stress due to rainfall exclusion, but this small increase in water  
547 stress cannot cause large changes in vegetation structure (Estiarte *et al.*, 2016), or LAI.

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

558 In our analysis changes in growing season length were not evident and did not influence our results. This is  
559 not surprising, as all rainfall manipulation experiments decreased or increased the available water to the  
560 ecosystem, without altering its “pulse” structure, including the frequency of rainfall occurrence, and the  
561 time of storm arrival (Ross *et al.*, 2012). As vegetation phenology in water limited ecosystem is very  
562 sensitive to the pulse structure dynamics of rainfall (Heisler-While *et al.*, 2009), evaluating in future  
563 experiments, whether models can properly capture the responses of vegetation to rainfall pulses in terms of  
564 productivity and drought deciduousness is very important. Changes in rainfall pulses will also strongly  
565 impact soil respiration dynamics, that will contribute significantly to the total carbon balance (Unger *et al.*,  
566 2010; Jarvis *et al.*, 2007).

### 567 ***Outlook for model developments and observations***

568 Our results highlight the need for a coordinated effort of new model development and data collection that  
569 could enable validations that are much more detailed than currently achievable here. Model discrepancies  
570 in the present study were attributed to the  $\beta$  stress factor, and long-term leaf area dynamics. The models  
571 used in this study implemented simple conceptual, yet vastly different (Wu *et al.*, 2018) parametrizations  
572 of the effects of water limitation, neglecting plant hydraulics and thus impacts on the water transport system  
573 (xylem cavitation) that can lead to hydraulic failure or/and carbon starvation (McDowell, 2011; McDowell  
574 *et al.*, 2013; Bonan *et al.*, 2014; Xu *et al.*, 2016). This could be an important limitation. However, tree

575 mortality is not a prominent feature of the manipulation experiments considered here and while it has  
576 attracted a lot of attention, models first need to better simulate mild to severe water stress before considering  
577 vegetation death. For instance, differences associated with the  $\beta$  factor are not only related to plant  
578 physiological thresholds but are a complex function of the assumed soil textural properties. Those properties  
579 are translated into soil hydraulic parameters (Van Looy *et al* 2017), affecting soil moisture dynamics and  
580 ET and ultimately their interplay with the value of the  $\beta$  factor. It is currently impossible or very difficult  
581 to identify which model is more realistic in this respect and each model can only “tune” all the above  
582 components at once. Specialized experiments measuring e.g. simultaneously high frequency water and  
583 carbon fluxes, soil moisture and plant water status in controlled environments could be designed to develop  
584 more informed parameterizations of  $\beta$ , and eventually expand to more detailed mechanistic representation  
585 of ecosystem scale plant hydraulics (Anderegg *et al.*, 2016; Konings and Gentine, 2017).

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

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

602 Finally, new streams of data via remote sensing can be also used for detailed model confirmation at larger  
603 scales. Satellite and airborne data related to vegetation structure, spanning from leaf chemistry to  
604 delineation of individual trees (Andersen, *et al.*, 2006; Gougeon and Leckie, 2006; Asner and Martin,  
605 2009), high frequency photosynthesis through solar induced fluorescence (SIF), soil moisture (Liu *et al.*,  
606 2011), and plant hydraulic status (Konings and Gentine, 2017) currently exist. Such data can help us to

607 identify the mechanistic link between plant water stress and how it affects vegetation productivity from  
608 short term photosynthesis reduction to decadal scales involving plant mortality and composition shifts. Note  
609 however that estimates of photosynthetic activity during water stress purely based on remote sensing (light  
610 reflection signals) are often biased and need to be interpreted with care (De Kauwe *et al.*, 2016; Stocker *et*  
611 *al.*, 2019).

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

622 *The authors declare no conflict of interest*

### 623 **Acknowledgements**

624 A.P. acknowledges financial support from NERC (grant no. NE/S003495/1). J.Z. acknowledges the Swiss  
625 National Science Foundation (Ambizione Grant 179876). DSG, PC, WL, ME, RO and JP are funded by the  
626 “IMBALANCE-P” project of the European Research Council (ERC-2013-SyG-610028). CP acknowledges  
627 the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC)  
628 Discover Grant. YPW acknowledges the financial support from the National Environmental Science  
629 Program for Earth System and Climate Change from the Australian Federal government. IKS and KSL  
630 acknowledge the financial support to the CLIMAITE project at Brandbjerg from the Villum Foundation.  
631 J.Po. was supported by the German Research Foundation’s (DFG) Emmy Noether Program (PO 1751/1-1).  
632 L.B. was supported by the DFG’s CE-LAND project. Computational resources were made available by the  
633 German Climate Computing Center (DKRZ) through support from the German Federal Ministry of  
634 Education and Research (BMBF). We thank all site operators, MODIS and FLUXNET2015 for providing  
635 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 CO<sub>2</sub> 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 CO<sub>2</sub>’, *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 intercomparisons 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 CO<sub>2</sub> 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 CO<sub>2</sub>: 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 CO<sub>2</sub> 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 CO<sub>2</sub> 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 CO<sub>2</sub> 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<sub>2</sub> 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 CO<sub>2</sub>, *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 CO<sub>2</sub> 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 CO<sub>2</sub> 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. <https://doi.org/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 seasonally 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. <https://doi.org/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 CO<sub>2</sub> 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<sub>2</sub> enrichment experiments: Model performance at ambient CO<sub>2</sub> 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 CO<sub>2</sub> 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.