AGU PUBLICATIONS

Global Biogeochemical Cycles

RESEARCH ARTICLE

10.1002/2016GB005480

Key Points:
- Compared to observations, vegetation in CMIP5 models responds stronger to precipitation anomalies
- Over-response of models to hydrological anomalies is smaller if evaluated based on soil moisture anomalies instead of precipitation indices
- Further improvements in plant-available water and response time scales are beneficial for future model development

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. Huang,
yyhuang@ufl.edu

Citation:
Huang, Y., S. Gerber, T. Huang, and J. W. Lichstein (2016), Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data, *Global Biogeochem. Cycles*, *30*, doi:10.1002/2016GB005480.

Received 11 JUL 2016
Accepted 2 DEC 2016
Accepted article online 8 DEC 2016

Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data

Yuanyuan Huang[1,2], Stefan Gerber[1], Tongyi Huang[3], and Jeremy W. Lichstein[3]

[1]Soil and Water Sciences Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, Florida, USA, [2]Now at Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma, USA, [3]Department of Biology, University of Florida, Gainesville, Florida, USA

Abstract Realistic representation of vegetation's response to drought is important for understanding terrestrial carbon cycling. We evaluated nine Earth system models from the historical experiment of the Coupled Model Intercomparison Project Phase 5 for the response of gross primary productivity (GPP) and leaf area index (LAI) to hydrological anomalies. Hydrological anomalies were characterized by the standardized precipitation index (SPI) and surface soil moisture anomalies (SMA). GPP and LAI in models were on average more responsive to SPI than in observations revealed through several indicators. First, we find higher mean correlations between global annual anomalies of GPP and SPI in models than observations. Second, the maximum correlation between GPP and SPI across 1–24 month time scales is higher in models than observations. And finally, we found stronger excursions of GPP to extreme dry or wet events. Similar to GPP, LAI responded more to SPI in models than observations. The over-response of models is smaller if evaluated based on SMA instead of SPI. LAI responses to SMA are inconsistent among models, showing both higher and lower LAI when soil moisture is reduced. The time scale of maximum correlation is shorter in models than the observation for GPP, and the markedly different response time scales among models for LAI indicate gaps in understanding how variability of water availability affects foliar cover. The discrepancy of responses derived from SPI and SMA among models, and between models and observations, calls for improvement in understanding the dynamics of plant-available water in addition to how vegetation responds to these anomalies.

1. Introduction

Terrestrial ecosystems have been absorbing ~25–30% of anthropogenic CO_2 emissions over the past several decades, acting as an important negative feedback mechanism in the carbon-climate system [*Le Quere et al.*, 2009]. Earth system models (ESMs) are vital tools to understand the strength and variability of the global terrestrial carbon sink and its feedback to future climate change. Models used in the most recent coupled carbon-climate experiment, the Coupled Model Intercomparison Project Phase 5 (CMIP5), incorporate diverse mechanisms driving vegetation dynamics and terrestrial carbon cycling in response to climate change and variability, increasing CO_2 and, in a few models, nitrogen limitation [*Taylor et al.*, 2011]. Considering the large spread among these models and the uncertainty with respect to associated processes, it is important to critically evaluate these models against observations [*Friedlingstein et al.*, 2014].

Water is a primary resource limiting plant uptake of atmospheric CO_2. Water-related climate extremes, such as drought, have been reported to significantly alter large-scale vegetation processes, reduce the terrestrial carbon sink strength, or even convert terrestrial ecosystems into temporary carbon sources [*Ciais et al.*, 2005; *Phillips et al.*, 2009; *Zhao and Running*, 2010]. Fluctuations in vegetation CO_2 uptake contribute markedly to the interannual atmospheric CO_2 variability, and much of those fluctuations are attributable to plant water availability [*Reichstein et al.*, 2013; *Zscheischler et al.*, 2013; *Poulter et al.*, 2014; *Zscheischler et al.*, 2014a, 2014b]. The importance of drought impacts on vegetation processes, and terrestrial carbon cycling is widely recognized and has been studied in the context of manipulative drought experiments, eddy-covariance tower networks, remote sensing, and process-based models [*Schwalm et al.*, 2010; *Zhao and Running*, 2010; *Potter et al.*, 2011; *Beier et al.*, 2012; *Chen et al.*, 2013; *Vicente-Serrano et al.*, 2013; *Liu et al.*, 2014; *Zscheischler et al.*, 2014b; *Knapp et al.*, 2015; *Lei et al.*, 2015]. However, the degree to which CMIP5 models accurately represent drought impacts remains largely unknown. *Piao et al.* [2013] reported an overestimation of the slope of gross primary productivity (GPP) versus precipitation from 10 process-based terrestrial biosphere models. However, additional analyses are needed to gain a more complete understanding of how models compare to real ecosystems in terms such as time scale, vulnerability, and resistance that reflect different facets of drought response.

Responses to drought incorporate various aspects of vegetation activity from molecular, physiological to phenological, or ecological processes [*Reyer et al.*, 2013; *Niu et al.*, 2014] which can be characterized by different metrics. For example, the response of vegetation to drought is often delayed [*Reichstein et al.*, 2013; *Frank et al.*, 2015]. Lagged tree mortality is reported some period after severe droughts [*Bigler et al.*, 2007; *Phillips et al.*, 2010; *Anderegg et al.*, 2013], and grasslands appear to have lower aboveground net primary production following dry years despite normal precipitation [*Sala et al.*, 2012]. The lagged or legacy effect of drought can directly alter vegetation status after the removal of drought stress as well as indirectly affect the response of vegetation to other environmental variables [*Frank et al.*, 2015]. Time lags between the initialization of water scarcity and the detection of its impacts [*Vicente-Serrano et al.*, 2014] are therefore an important metric that reflects the time-dependent characteristic of drought response. Vulnerability refers to the susceptibility of vegetation to adverse effects caused by water anomalies [*Reyer et al.*, 2013], while the resistance describes the degree to which vegetation tolerates drought stress [*Farooq et al.*, 2009]. For process-based models, the drought response is an emergent property that integrates various processes in modeled carbon and water cycles and a realistic representation of different aspects is important for modeling drought responses [*Reichstein et al.*, 2013; *Niu et al.*, 2014].

Drought is generally described as an extended period of water scarcity. Drought responses of the carbon cycle are complex partly due to the multifaceted nature of drought. Precipitation is one of the main determinants of plant water availability. However, simple precipitation metrics do not provide a complete characterization of drought, such as their duration, extent, cumulative severity, and timing [*McKee et al.*, 1993; *Chen et al.*, 2013]. Droughts are frequently quantified by drought indices that adjust precipitation to better reflect water availability to plants. The Palmer drought severity index (PDSI) [*Palmer*, 1965] and the standardized precipitation index (SPI) [*McKee et al.*, 1993] are two of the mostly widely used indices. PDSI is based on soil water balance by accounting for water supply (precipitation) and demand (potential evapotranspiration); however, PDSI is a complex index that sometimes yields inconsistent results among alternative calculation methods [*Trenberth et al.*, 2014]. In contrast, SPI is based solely on precipitation anomalies and has become increasingly popular in studies of ecosystem's drought response [*Sims et al.*, 2002; *Ji and Peters*, 2003; *World Meteorological Organization* (*WMO*), 2012; *Orlowsky and Seneviratne*, 2013] because it provides a simple and consistent characterization of droughts at multiple temporal scales and to compare locations that have different precipitation regimes.

Although meteorological drought indices, such as SPI, are useful indicators of water deficits and surpluses, dryness or wetness experienced by plants is more directly linked to soil moisture conditions. Factors other than meteorological conditions, such as soil and vegetation characteristics, affect soil moisture dynamics and thus vegetation processes [*Porporato et al.*, 2004; *Weng and Luo*, 2008]. Previous global-scale studies were constrained by the availability of observations, as soil moisture is highly variable across space and time and is difficult to measure on a large scale [*Seneviratne et al.*, 2010]. Recent soil moisture products from spaceborne passive and active microwave sensors have become available for 30+ years [*Dorigo et al.*, 2014]. Despite being limited to the surface soil column (less than 10 cm depth) [*Liu et al.*, 2011], which may be "decoupled" from the root zone soil moisture in some cases [*Capehart and Carlson*, 1997; *Seneviratne et al.*, 2010], remotely sensed soil moisture products provide an additional valuable source of information for studying drought-carbon responses.

In this paper, we evaluate drought responses of CMIP5 models and compare these responses to those estimated from observation-based data products. We quantify different facets of the response of GPP and leaf area index (LAI) to drought, including correlations, time scales, cumulative anomalies, and sensitivities to extreme dry/wet events. We use the meteorological drought index SPI at time scales from 1 to 24 months to characterize dry/wet anomalies. We also test the response of GPP and LAI to both modeled and data-derived soil moisture anomalies. Our study goes beyond previous CMIP5 evaluation studies which either focus on hydrology [*Orlowsky and Seneviratne*, 2013] or terrestrial carbon cycling alone, or the relationship of carbon cycling with climate variables such as precipitation and temperature [*Shao et al.*, 2013] by exploring

Table 1. CMIP5 Models Used in This Study With Complete Model Expansions

Models	Model Expansion
	Earth System Models
BCC-CSM1.1	Beijing Climate Center, Climate System Model version 1.1
CanESM2	Second Generation Canadian Earth System Model
CESM1-BGC	Community Earth System Model, version 1.0-Biogeochemistry
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory Earth System Model with GOLD ocean component (ESM2G)
HadGEM2-ES	Hadley Centre Global Environmental Model, version 2 (Earth System)
INM-CM4.0	Institute of Numerical Mathematics Coupled Model, version 4.0
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5A, coupled with NEMO, low resolution
MIROC-ESM	Model for Interdisciplinary Research on Climate, Earth System Model
NorESM1-ME	Norwegian Earth System Model, version 1 (intermediate resolution)
	Land Surface Models or Vegetation Models
BCC-AVIM1.0	Beijing Climate Center Atmosphere Vegetation Interaction Model Version 1.0
CLASS	Canadian Land Surface Scheme
CTEM	Canadian Terrestrial Ecosystem Model
CLM4	Community Land Model, version 4
CLM4-CN	Community Land Model, version 4, with the Coupled Carbon–Nitrogen Cycle
LM3	Land Model, version 3
LM3V	Land Model, version 3, with Vegetation and Carbon Cycling
JULES	Joint United Kingdom Land Environment Simulator
TRIFFID	Top-down Representation of Interactive Foliage and Flora Including Dynamics
LSM	Land Surface Model
ORCHIDEE	Organizing Carbon and Hydrology in Dynamic Ecosystems
MATSIRO	Minimal Advanced Treatments of Surface Interaction and Runoff
SEIB-DGVM	Spatially Explicit Individual-Based Dynamic Global Vegetation Model

the response of terrestrial carbon cycling to drought by including time scales of responses and by evaluating the response to extreme events.

2. Methods

2.1. Data

2.1.1. CMIP5 Models and Experiments

CMIP5 provides a standard set of model simulations based on common protocols [*Taylor et al.*, 2011]. We selected nine ESMs from the historical experiment, which prescribe changing climate and greenhouse gas concentrations that are consistent with observations from 1850 to 2005 (Tables 1 and 2). ESMs were chosen based on the availability of monthly GPP, LAI, and relevant climate and soil moisture data required for our analysis. Where ESMs were represented in CMIP5 by multiple model versions, we randomly selected one model version for inclusion in our analysis, because we expected greater variation among different ESMs than among the different versions of each ESM. For models with multiple available realizations, only the first ensemble member was used ("r1i1p1," which indicates the first run under the first set of initial conditions with the first set of physical parameters [*Taylor et al.*, 2011]). Model output data were downloaded from the Program for Climate Model Diagnosis and Intercomparison data server.

The terrestrial components of these ESMs differ in their representations of plant functional types (PFTs), land use change, soil characteristics, carbon and nitrogen dynamics, and the spatial resolution (Table 2). Despite differences in model structure and parameterization, the models share many similarities in their treatments of terrestrial carbon and water cycles. Plant species with similar characteristics are represented in aggregate by PFTs. For each PFT, GPP is the accumulation of gross photosynthesis over a given time period. Photosynthesis is simulated according to modified versions of Farquhar's biochemical model for C3 [*Farquhar et al.*, 1980; *Collatz et al.*, 1991] and *Collatz et al.* [1992] for C4 plant, except in Spatially Explicit Individual-Based Dynamic Global Vegetation Model (SEIB-DGVM) where photosynthesis is simplified as a Michaelis-Menten type function of photosynthetically active radiation [*Sato et al.*, 2007]. Net primary production (the difference between GPP and autotrophic respiration) is allocated to different vegetation tissues, such as leaves, roots, and stems. LAI depends on carbon allocation to leaves, leaf lifespan and phenology, and specific leaf area ($m^2\,kg\,C^{-1}$), which is a PFT-specific constant in most models, but varies across the vertical canopy gradient in some cases (e.g., Community Land Model, version 4, with the Coupled Carbon–

Table 2. Primary Characteristic of the Land Carbon Cycle Component of the Nine Participating Models in This Study[a]

ESMs	BCC_CSM1.1	CanESM2	CESM1-BGC	GFDL_ESM2G	HadGEM2-ES	INM-CM4.0	IPSL-CM5A-LR	MIROC-ESM	NorESM1-ME
Abbreviation in this study	BCC	CAN	CESM	GFDL	HAD	INMCM4	IPSL	MIROC	NOR
Land model	BCC-AVIM1.0	CLASS	CLM4	LM3	JULES	LSM	ORCHIDEE	MATSIRO	CLM4
Vegetation model	-	CTEM	CLM4CN	LM3V	TRIFFID	-	-	SEIB-DGVM	CLM4CN
DGVM	No	Yes	No	Yes	Yes	Yes	No	Yes	No
N cycle	No	No	Yes	No	No	No	No	No	Yes
No. PFTs	15	9	15	5	5	11	13	13	15
No. soil layers	10	3	15	20	4	23	7	6	15
Soil depth	3.4	4.1	43.7	10	3	15	3.9	14	43.7
Fire	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Human activities	Crop	Crop	Crop, pasture, wood harvest	Crop, pasture, wood harvest, deforest	Crop, pasture	Deforestation	Crop, pasture	Crop, pasture, wood harvest	Crop, pasture, wood harvest
Land use emissions	Prescribed	Computed	Computed	Computed	Computed, prescribed	Prescribed	Computed	Computed	Computed
Historical span	1850–2012	1850–2005	1850–2005	1861–2005	1860–2005	1850–2005	1850–2005	1850–2005	1850–2005
Resolution (lat × lon)	2.8° × 2.8°	2.8° × 2.8°	0.9° × 1.3°	2.0° × 2.5°	1.3° × 1.9°	1.5° × 2°	1.9° × 3.8°	2.8° × 2.8°	0.9° × 1.3°
Reference	[*Ji et al.*, 2008; *Wu et al.*, 2013]	[*Arora and Boer*, 2010]	[*Thornton et al.*, 2002; *Gent et al.*, 2011]	[*Shevliakova et al.*, 2009; *Dunne et al.*, 2012]	[*Cox*, 2001; *Jones et al.*, 2011]	[*Bonan*, 1996; *Volodin et al.*, 2010]	[*Krinner et al.*, 2005; *Dufresne et al.*, 2013]	[*Sato et al.*, 2007; *Watanabe et al.*, 2011]	[*Thornton et al.*, 2002; *Tjiputra et al.*, 2013]

[a]DGVM refers to the dynamic change of vegetation coverage with plant competition; No. PFTs stands for the number of plant functional types implemented in models; No. soil layers refers to the number of soil layers.

Nitrogen Cycle (CLM4-CN) [*Thornton and Zimmermann*, 2007]). Carbon from mortality (including disturbance) and from leaf, root, and stem turnover enters litter or soil pools that are subjected to decomposition. Decomposition is represented as a first-order decay process that is modified by soil moisture and temperature [*Todd-Brown et al.*, 2013].

Coupling between vegetation dynamics and precipitation in these models depends on how soil hydrology is represented (e.g., soil depth and number of vertical soil layers; Table 2), the down-regulation of LAI and/or leaf-level photosynthesis due to soil water stress, and the stomatal conductance, which regulates the land-atmosphere exchange of carbon and water vapor (transpiration) and which depends on the atmospheric water demand and the surface energy budget. The formulation of stomatal conductance generally follows the Ball-Woodrow-Berry model [*Ball et al.*, 1987] or a modified version [*Leuning*, 1995]. Soil water stress functions that down-regulate GPP vary among models, as detailed in Text S1 in the supporting information. Effects of water availability on other vegetation processes—such as phenology, establishment, carbon allocation, respiration, and mortality—also differ among models.

2.1.2. Observation-Based Data Products

We compiled global observation-based data products to estimate the responses of GPP and LAI to precipitation anomalies and soil moisture anomalies. These observation-based GPP and LAI responses provide means to evaluate the corresponding responses in CMIP5 models. Below, we describe the observation-based data products we used in this work: GPP, LAI, precipitation, and soil moisture.

As an observation-based GPP data set, we used the 1982–2005 model tree ensemble (MTE) global gridded GPP product (Table 3), which is derived from the FLUXNET global eddy-covariance tower network [*Jung et al.*, 2011]. The MTE upscaling approach consists of a set of trained regression trees and 29 candidate predictors, such as the fraction of absorbed photosynthetically active radiation (FAPAR, derived from remote sensing), climate, and land cover. Although there is uncertainty in the MTE-GPP gridded time series, this data-derived product has gained acceptance in global GPP analyses [*Anav et al.*, 2013; *Piao et al.*, 2013; *Zscheischler et al.*,

Table 3. Observation-Based Data Sets Used for CMIP5 Model Evaluations

Variable	Description	Temporal Resolution	Spatial Resolution	Reference
MTE_GPP	Gross primary productivity derived from FLUXNET observations	Monthly, 1982–2011	0.5° × 0.5°	[*Jung et al.*, 2011]
GLASS_LAI	Satellite derived leaf area index from GLASS	8 days, 1981–2012	0.05° × 0.05°	[*Liang and Xiao*, 2012; *Xiao et al.*, 2014]
GLOBMAP_LAI	Satellite derived leaf area index from GLOBMAP	Half month/8 days, 1981–2011	0.07° × 0.07°	[*Liu et al.*, 2012]
CRU_PRE	Precipitation from CRU TS 3.21 (Climatic Research Unit at the University of East Anglia)	Monthly, 1901–2012	0.5° × 0.5°	[*Jones and Harris*, 2013]
DELA_PRE	Precipitation from University of Delaware v3.01	Monthly, 1901–2010	0.5° × 0.5°	[*Willmott and Matsuura*, 2012]
GPCP_PRE	Precipitation from GPCP v2.2 (Global Precipitation Climatology Project)	Monthly, 1979–current	2.5° × 2.5°	[*Adler et al.*, 2003]
GPCC_PRE	Precipitation from GPCC v6 (Global Precipitation Climatology Centre)	Monthly, 1901–2010	0.5° × 0.5°	[*Schneider et al.*, 2011]
ECV_SM	Soil moisture from ECV_SM v02.1	Daily, 1978–2013	0.05° × 0.05°	[*Liu et al.*, 2011]

2014a] and provides a longer temporal record than satellite-derived GPP products [*Mao et al.*, 2012]. Although the MTE-GPP product does not explicitly account for cumulative effects of forcing variables, and therefore may underestimate time lags in the GPP-drought response, cross validation of GPP monthly anomalies indicated a reasonable degree of predictive power (e.g., correlation of 0.72 when predicting site-level data that is excluded from the training data). In section 4, we further discuss how limitations of the MTE-GPP product may affect our results.

We used two global LAI data sets in our analysis: GLOBMAP [*Liu et al.*, 2012] and GLASS [*Xiao et al.*, 2014]. The GLOBMAP LAI product (available from http://www.globalmapping.org/globalLAI/) fuses the Moderate Resolution Imaging Spectroradiometer (MODIS) and advanced very high resolution radiometer LAI products, which are extracted from MOD09A1 (the MODIS land surface reflectance data set) and Global Inventory Modeling and Monitoring Study normalized difference vegetation index, respectively. The second global LAI product we used, GLASS, is generated with physical inversion techniques by using the general regression neural networks (GRNNs) method [*Xiao et al.*, 2014]. The GRNNs are trained with fused MODIS and CYCLOPES LAI products [*Friedl et al.*, 2002; *Baret et al.*, 2007] and the MODIS reflectance values for each MODIS biome. GLASS LAI is then retrieved from MODIS reflectance data based on the trained GRNNs.

To quantify the GPP response to precipitation (which we converted to the standardized precipitation index; see section 2.2), we used the monthly Global Precipitation Climatology Centre (GPCC) precipitation data set (Table 3), because this is the precipitation data set used to create the gridded MTE-GPP data product. By using the same precipitation data set used to create MTE-GPP, we preserve (as close as possible) the information content in the original FLUXNET data set and minimize the effects of precipitation uncertainty on our GPP analysis (see section 2.4 for further explanation). In contrast, uncertainty in actual precipitation is expected to affect our analysis of observation-based LAI (see section 2.4). Thus, we used four different global monthly precipitation data sets based on rain gauge measurements, remote sensing, or combination of these to derive observation-based precipitation indexes for LAI analyses (Table 3).

To quantify GPP and LAI responses to soil moisture dynamics, we used a recently generated multidecadal soil moisture data set, ECV_SM, which is a merged product from passive and/or active microwave remote sensing data [*Liu et al.*, 2011]. We judged ECV_SM to be the best available option, as a global soil moisture data set based on in situ measurement is not available. ECV_SM has relatively fine spatiotemporal coverage and preserves the short term (e.g., seasonal and interannual) as well as long-term dynamics of the microwave remote sensing data from which it is derived. ECV_SM has been evaluated against the Global Land Data Assimilation System-Noah land surface model, the land surface component of Max-Planck-Institut-ESM (JSBACH), and the ERA-Interim reanalysis data with respect to the trend (1988–2010) and anomalies [*Dorigo et al.*, 2012; *Loew et al.*, 2013], and against ground measurements for the interannual and intra-annual dynamics in different regions [*Liu et al.*, 2011; *Dorigo et al.*, 2014; *Pratola et al.*, 2014; *Zeng et al.*, 2015]. ECV_SM (v02.1) used in this study combines retrievals of soil moisture from six passive (scanning multichannel microwave radiometer, Special Sensor Microwave Imager, Tropical Rainfall Measuring Mission Microwave Imager, Advanced Microwave Scanning Radiometer–EOS (AMSR-E), WindSat, and AMSR2) microwave sensors and the scatterometers onboard ERS-1, ERS-2, and METOP-A into a global data set covering the period of 1979–2013. We chose this merged product as our observation-based surface (<10 cm depth) soil moisture [*Liu et al.*, 2011].

2.2. Drought Indices

We used monthly series of two standardized drought indices to quantify meteorological and soil moisture anomalies for both CMIP5 models and observation-based data sets. We used the standardized precipitation index (SPI) to quantify meteorological anomalies over different time scales. We used the algorithm described in *Lloyd-Hughes and Saunders* [2002] to calculate SPI: for each month, precipitation over the current and the previous ($k-1$) months is summed, where k is the time scale of the SPI (1–24 months in this study). A two-parameter Gamma distribution is then fitted to these derived monthly precipitation time series (one monthly time series for each time scale). Each fitted Gamma distribution is subsequently transformed into a standard normal distribution. Thus, the resulting SPI time series (one for each time scale) have a mean of 0 and a standard deviation of 1.

To quantify soil moisture dynamics, we calculated standardized surface soil moisture anomalies (SMA), where "surface" roughly corresponds to 0–10 cm depth (see below). First, we deseasonalized the observation-based and simulation results by subtracting the mean monthly values in the soil moisture time series $x_{y,m}$, where m (1,2,…12) indicates the month and y the year. The deseasonalized monthly time series $x'_{y,m}$ is given by

$$x'_{y,m} = x_{y,m} - \frac{1}{n}\sum_{i=1}^{n} x_{i,m} \tag{1}$$

where n is the number of years in the analysis. The $x'_{y,m}$ is further standardized to unit standard deviation for each grid cell to yield SMA. The standardization harmonizes soil moisture between observation-based and ESM simulations and allows a direct comparison with SPI. Months with soil temperature below 0°C are excluded from the SMA analysis because the remote-sensing-based soil moisture data (ECV_SM) are constrained to temperatures above 0°C. To similarly exclude months with temperature below 0°C from the CMIP5 models, which have different soil layers and depths (Table 2), we calculated the depth-weighted average temperature to the layer closest to 10 cm. We chose 10 cm depth for consistency with ECV_SM, which represents soil moisture no deeper than 10 cm.

2.3. Quantifying Drought Responses

We quantified relationships between response variables (GPP and LAI) and explanatory variables (SPI at 1–24 month time scales and SMA at the 1 month time scale) at both global and regional scales. All analyses were performed for both observation-based data sets and CMIP5 model output. We paired observation-based GPP and LAI data sets with observation-based SPI and SMA data sets (see section 2.1.2), and we paired GPP and LAI from each CMIP5 model with SPI and SMA derived from the model's forcing data (SPI) and output (SMA). To avoid correlations with confounding variables (e.g., trends in atmospheric CO_2 concentration), we detrended GPP, LAI, SPI, and SMA prior to analysis to remove long-term linear trends, thereby focusing our analysis on short-term (e.g., monthly or interannual) anomalies. All analyses covered the period of 1982–2005. All global CMIP5 model output and observation-based data set were re-gridded to a common spatial resolution of 1° × 1° and a common land mask (derived from the observation-based GPP data set described above) using the nearest neighbor method, which we assume conserves drought responses. Most of the data sets were available at a monthly temporal resolution; variables reported at finer temporal resolution were upscaled to monthly values.

At the global scale (all grid cells combined), we calculated the correspondence between annual GPP and SPI anomalies (time scales from $k = 1$–24 months) for the period of 1982–2005. Monthly GPP and SPI series (time scales from $k = 1$–24 months) were averaged over a year to obtain annual GPP and SPI anomalies. The Pearson correlation coefficient was used to quantify correlations between GPP and SPI anomalies. A similar procedure was applied to LAI for both CMIP5 models and observation-based data sets.

To understand the regional pattern of vegetation in response to water anomalies and the characteristic time scale of responses, we quantified responses of GPP and LAI to SPI for different geographic regions and also arid versus humid grid cells defined by the Köppen-Geiger climate classification [*Kottek et al.*, 2006] (Figure S1 in the supporting information). We separated GPP (or LAI) anomalies into 12 series (one per month) and correlated each series with 1 to 24 month SPIs, respectively. In each grid cell, we only included data points where monthly GPP or LAI was at least 25% of the maximum monthly value (across years) so that our analysis focuses on months with relatively high GPP or LAI corresponding roughly to the growing season in each grid cell and year. This yielded a maximum of 288 (24 SPI time scales by 12 months per year) correlation analyses per grid cell. For each grid cell, we used the maximum correlation coefficient (out of a maximum possible of 288) as an indicator of the sensitivity of vegetation activity to water anomalies, where "sensitivity," broadly reflects "vulnerability" and "resistance" to drought (see section 1). As an index of the time scale of vegetation's response to water anomalies, we determined for each grid cell the SPI time scale at which the greatest GPP and LAI sensitivities occurred.

In order to quantify sensitivities of GPP and LAI to extreme droughts or extreme wet events, we defined extreme events as 3 month SPI less than -2 (extreme dry) and greater than 2 (extreme wet) [*McKee et al.*, 1993]. The 3 month SPI can be considered as a short-term drought indicator and is a common temporal scale in drought assessment [*Ji and Peters*, 2003; *WMO*, 2012; *Zscheischler et al.*, 2014b]. We further aggregated the extreme dry (or wet) conditions that are adjacent in space and time into extreme dry (or wet) event clusters (three-dimensional, longitude × latitude × time) following *Lloyd-Hughes* [2012] and *Zscheischler et al.* [2014b]. By "adjacent," we refer to any of the 26 neighbors in three-dimensional (latitude × longitude × time) space. The size of an extreme event cluster is the integral of SPI over the spatiotemporal domain of the event cluster (event in short), and the impact of the event is the corresponding integral of GPP (or LAI) anomalies. The GPP (or LAI) anomalies used here are absolute deviations; i.e., they are detrended and deseasonalized, but are not standardized so as to preserve the relative magnitude of GPP (or LAI) in different months of the year. In the following, we use "volume" (km^2 month instead of km^3) interchangeably with the size of an extreme event. Each extreme event may span different spatial and temporal scales. We sorted these extreme events by volume in descending order and calculated their cumulative impacts by adding each individual extreme event's impact. For example, the cumulative impact of the 100 largest extreme events on GPP is the sum of all GPP anomalies over the spatiotemporal domain spanned by these 100 largest extreme events. The mean ratio (for all global events combined) between GPP (or LAI) anomalies and SPI in extreme dry events is used as an indicator of the global drought sensitivity, and the mean ratio for extreme wet events is used an indicator for the global wet sensitivity.

For the soil moisture anomaly (SMA) analysis for each grid cell, we separated monthly GPP (or LAI) anomalies into 12 series (one per month) and correlated each series with the corresponding SMA series. As with SPI, we used the maximum correlation in each grid cell as an index of vegetation sensitivity to SMA. We implemented an analysis of extreme SMA events as explained above for SPI. We defined extreme dry conditions as SMA less than -2 and extreme wet conditions as SMA greater than 2. We then quantified the size of events (three-dimensional, longitude × latitude × time) and responses and calculated their global ratios.

2.4. Measurement Error Models

Most previous assessments of ecosystem response to water availability have ignored uncertainty in the explanatory variables. Errors in response variables (GPP and LAI in our study) introduce noise, but no systematic bias into standard regression analyses. In contrast, errors in explanatory variables (e.g., SPI) bias the estimated response slopes (e.g., the slope of LAI versus SPI) toward zero if these errors are ignored [*Fuller*, 1987; *Lichstein et al.*, 2014]. Errors in precipitation (and thus SPI) and other forcing data do not affect our analyses of CMIP5 models, because the environmental conditions experienced by vegetation in each CMIP5 model are known without error; i.e., although the forcing data include errors, the conditions experienced by the modeled vegetation are known, and thus, errors do not affect the GPP or LAI slopes estimated from CMIP5 model output. In contrast, these slopes will be biased toward zero in observation-based analyses if there is uncertainty in precipitation or soil moisture data products. As with CMIP5 models, uncertainty in precipitation does not affect our analysis of the MTE-GPP data product, because we used the same precipitation product (GPCC) used by *Jung et al.* [2011] to create the MTE-GPP product. In contrast, uncertainty in precipitation does affect our analysis of observation-based LAI data sets, because these are based on satellite reflectance rather than precipitation-based algorithms. Precipitation uncertainty is expected especially after 1991, when fewer meteorological stations are available compared to 1950–1990 [*Trenberth et al.*, 2014]. Uncertainty in soil moisture is also expected to affect our analyses of GPP and LAI. However, in contrast to precipitation—where multiple data sets provide a

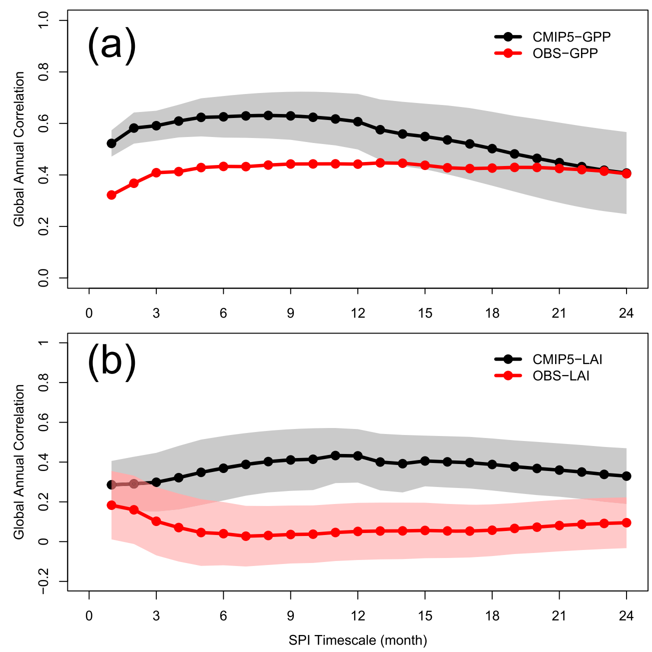

Figure 1. Pearson correlation coefficients between detrended global annual gross primary productivity (GPP, a), or leaf area index, (LAI, b) and the standardized precipitation index (SPI) over 1982–2005. The *x* axes are SPI time scales indicating drought severity at different temporal scales (1–24 months). The black dots represent result from CMIP5 models (multimodel mean) and the red dots from observations. The shaded areas correspond to the 95% confidence limits among CMIP5 models (nine in total) or for the observation-based LAI responses (eight in total, two LAI data sets by four precipitation data sets).

straightforward means of quantifying uncertainty (Table 3)—only a single soil moisture data set was available. Thus, we restricted our analysis of measurement errors to the response of observation-based LAI to SPI. For simplicity, we further restricted this analysis to global LAI (as opposed to grid-level LAI analyses). Although incomplete, this analysis allows an initial exploration of how errors in explanatory variables can affect estimated responses of vegetation activity to water anomalies, and how these errors can affect the comparison of observation-based and CMIP5 responses.

To account for errors in SPI when estimating global LAI responses from the observation-based data sets, we used measurement error models (MEMs) [*Fuller*, 1987]. To implement MEMs for the observation-based global SPI analyses, we estimated uncertainty in global SPI (all grid cells combined) by quantifying variation among multiple observation-based SPI data sets as follows: For each year, the uncertainty of global SPI is estimated as the standard deviation among the four global SPIs derived from GPCC, Global Precipitation Climatology Project (GPCP), University of Delaware (DELA), and Climatic Research Unit (CRU) data sets (Table 3). For simplicity, we assumed no measurement error for LAI in the MEMs.

3. Results

3.1. Sensitivity of Global GPP and LAI to SPI

CMIP5 models generally have higher correlations between global annual GPP and water anomalies compared to observation-based data sets (Figure 1a). However, the difference diminishes as SPI time scale increases and approaches zero with the SPI time scale of 24 months. In addition, differences among CMIP5 models—indicated by 95% confidence intervals—increase with SPI time scale.

As with GPP, correlations with SPI are higher for LAI in CMIP5 models compared to observation-based LAI data sets (Figure 1b). Global LAI is generally less responsive to water anomalies than GPP in both CMIP5 models and observation-based data sets, as indicated by lower correlation coefficients (Figure 1). Correlations between observation-based LAIs and SPIs are weak with the mean correlation coefficients less than 0.2 for each SPI time scale and most of the observation-based correlations showing nonsignificant ($P > 0.05$) relationships.

Accounting for errors in precipitation data sets using measurement error models (MEMs) increased the estimated slope of observation-based global annual LAI versus global annual SPI (Figure 2). For the GLASS LAI data set, MEM slopes were greater than CMIP5 slopes for all of the 1 to 3 month SPI time scales, whereas ordinary least squares (OLS) slopes that ignore errors in *x* were similar for GLASS and CMIP5. The GLOBMAP LAI data set yielded shallower slopes than GLASS, such that OLS slopes based on GLOBMAP were consistently shallower than for CMIP5, and the MEM slope from GLOBMAP was steeper than CMIP5 only for the 1 month SPI time scale (Figure 2).

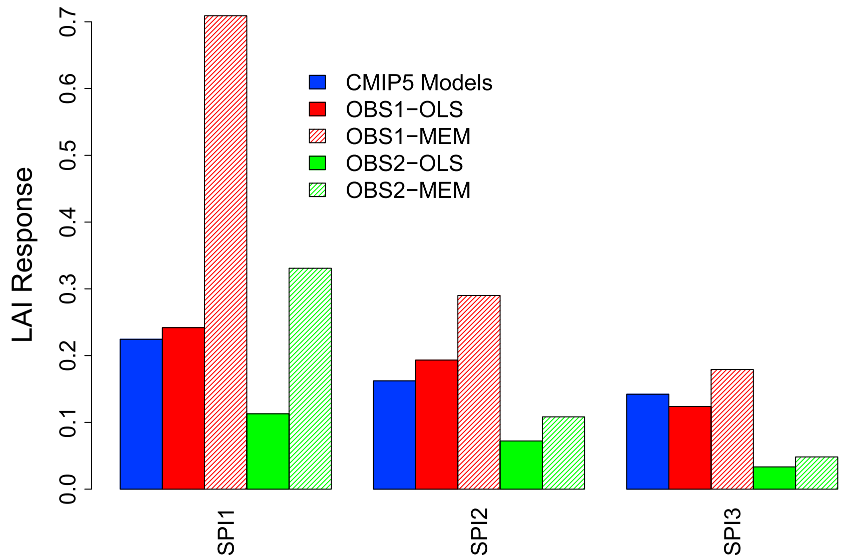

Figure 2. LAI responses to SPI. The *y* axis units are the change in LAI (m^2 m^{-2}) per SPI change (unitless). The CMIP5 model (filled in blue) indicates LAI responses with the multimodel mean ordinary least squares regression slopes from the nine CMIP5 models; OLS (filled in red or green) indicates the mean ordinary least squares regression slopes from the observation-based data sets (four global precipitation data sets); MEM (filled in slashed red or green) are slopes of observation-based data set estimated from the measurement error model taking into account error in *x* in SPI. The error of SPI is estimated as the standard deviation of the mean annual SPI among four versions of SPI derived from four precipitation data sets (Table 3). The observation-based LAI data sets are from GLASS (OBS1) and GLOBAMP (OBS2). Results are shown for SPI with 1 month (SPI1), 2 month (SPI2), and 3 month (SPI3) time scales.

3.2. Sensitivities and Time Scales of Grid-Cell GPP and LAI Responses to SPI

Maximum correlations (largest magnitude) between GPP (or LAI) and SPI out of up to 288 correlations per grid cell (24 SPI time scales × up to 12 growing season months, depending on location) provide indices of a location's sensitivity to water anomalies. These sensitivities in CMIP5 models tend to be stronger than in the observation-based MTE-GPP data set (Figures 3a and 3b and 4a). Higher GPP-SPI sensitivities in CMIP5 models compared to MTE-GPP are apparent across a large portion of the global land surface, where the multimodel mean correlation is greater than 0.75 (Figure 3a). Although MTE-GPP shows comparable or higher GPP-SPI sensitivity than CMIP5 models in some regions (e.g., Alaska; Figure 3b), sensitivities are higher for CMIP5 models when averaged over the global land surface and in both arid and humid regions (Figure 4 a). High GPP-SPI sensitivity is generally consistent across CMIP5 models despite some regional differences (Figure S2). Both CMIP5 models and MTE-GPP yielded some negative GPP-SPI correlations, particularly in the northern high latitudes (Figures 3a and 3b).

As with GPP, LAI-SPI sensitivities (i.e., maximum correlations) tended to be higher for CMIP5 models than for the observation-based data sets (two LAI data sets × four precipitation data sets) (Figures 3c and 3d and 4b). There was considerable variability in LAI-SPI sensitivities among CMIP5 models and among the observation-based data sets (Figure S3). LAI-SPI sensitivities in the GLOBMAP LAI data product were similar to those in some CMIP5 models and were higher than those in the GLASS LAI data product (Figure S3).

The observation-based data set for GPP exhibits more grid cells with both short (e.g., <3 months) and long (e.g., >18 months) SPI time scales compared to CMIP5 models (Figures 5a and 5b). When SPI time scales are averaged across the globe, time scales with maximum correlations are shorter in CMIP5 models compared to observation-based data set for GPP (CMIP5 models: mean, 7.06 months; observation-based GPP: 8.75 months; Figure 4). However, LAI has a longer average time scale in CMIP5 models (mean, 10.47 months) compared to the mean of the observation-based data sets (8.64 months) (Figure 4). The overall shorter mean time scale of modeled GPP response holds for both arid and humid regions (Figure 4), and the longer mean time scale in modeled LAI is also consistent between arid and humid regions (Figure 4). Individual CMIP5 models varied strongly with respect to LAI response time scale (Figure S5). The mean response time scale of LAI is as low as 3.7 months in the Institute of Numerical Mathematics Coupled Model, version 4.0

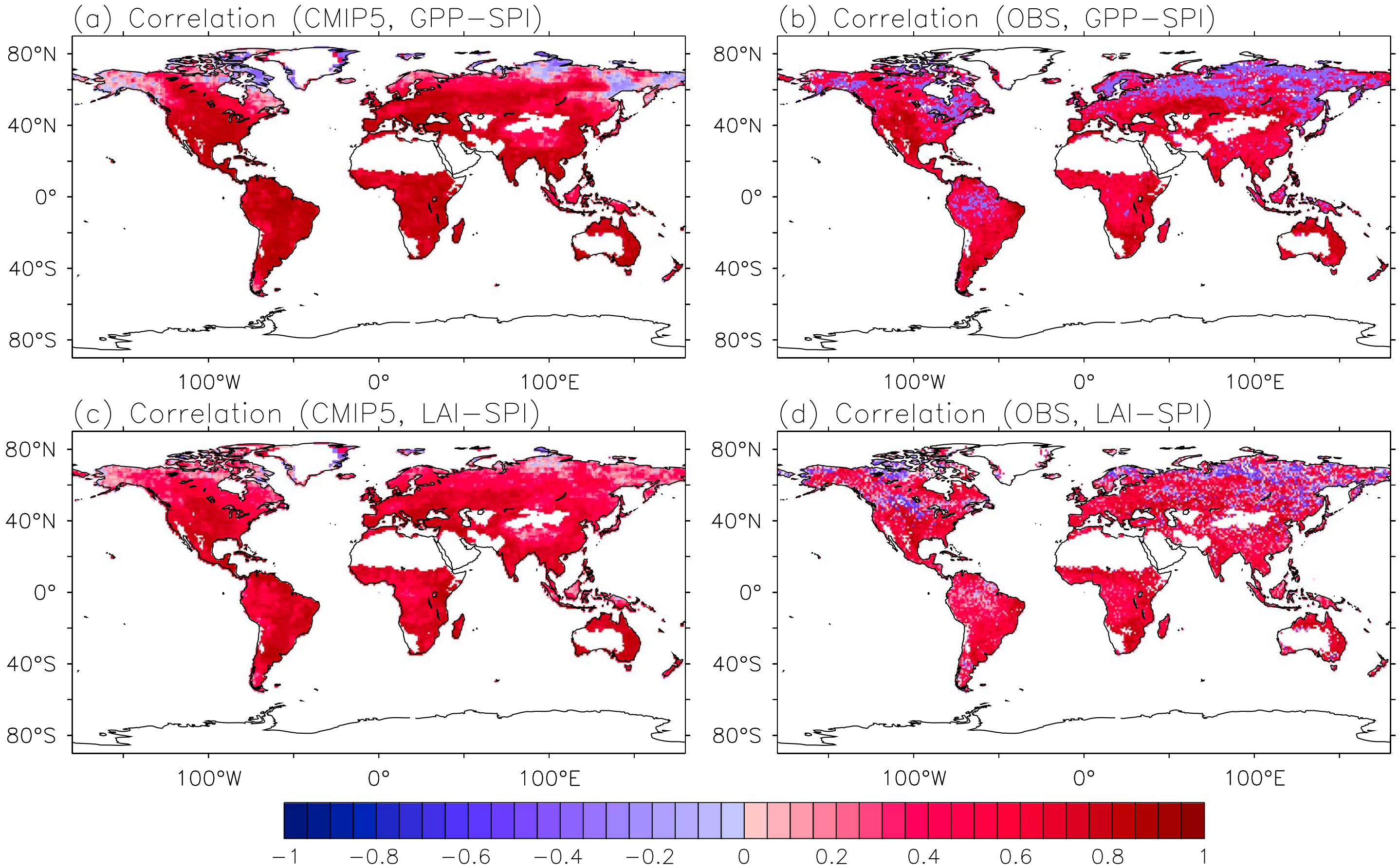

Figure 3. Spatial map of the maximum magnitude Pearson correlation obtained from up to 288 (24 SPI time scales × up to 12 growing-season months per year) possible statistically significant correlations ($P < 0.05$) between GPP (or LAI) anomalies and SPI. Shown are averages among models (a and c: CMIP5). For observations (OBS), results are shown for one analysis of (b) GPP and average among eight (d) LAI-SPIs (two LAI data products × four precipitation data products). The blank areas in each panel represent either polar frost or arid desert according to the Köppen-Geiger climate classification or locations with no significant correlations. The analysis is restricted to the growing season for each grid cell, defined as months where mean monthly GPP (or LAI) is at least 25% of its maximum.

(INMCM4) and as high as 14.3 months in NOR (NorESM1-ME), indicating large uncertainties in model-derived LAI response time scale.

3.3. Extreme Event Analysis Based on SPI

GPP was more responsive to extreme dry and wet events in CMIP5 models than in the observation-based data set. GPP reductions integrated over all extreme dry events (1982–2005) in CMIP5 models range from −5.30 to −30.28 GtC, which are larger in magnitude than that of observation-based GPP (−3.00 GtC) (Figure 6). The integrated increase of GPP during extreme wet events is also higher in CMIP5 models (ranging from 3.64 GtC to 14.90 GtC) compared to observation-based GPP (1.61 GtC) (Figure 6). Higher excursions of modeled GPP during extreme events compared to the observation-based product are not attributable to differences in the magnitude of extreme events. The mean total volume of extreme dry events is 16.36×10^8 km^2 month in CMIP5 models compared to 17.88×10^8 km^2 month for observation-based GPP during the period of 1982–2005. The mean total volume of extreme wet events is 11.83×10^8 km^2 month in CMIP5 models compared to 12.75×10^8 km^2 month for the observation-based GPP. Thus, the global dry/wet sensitivity (total GPP excursions divided by total volume of extreme dry or wet events) is smaller for the observation-based data set compared to each of the nine CMIP5 models under study (Figure S6). The apparent oversensitivity of CMIP5 modeled GPP to extreme events is also unlikely to be caused by differences between CMIP5 and the observation-based data set in the means or temporal variation of GPP, because CMIP5 oversensitivities are still apparent when GPP anomalies are normalized for each grid cell (Figures S7 and S8).

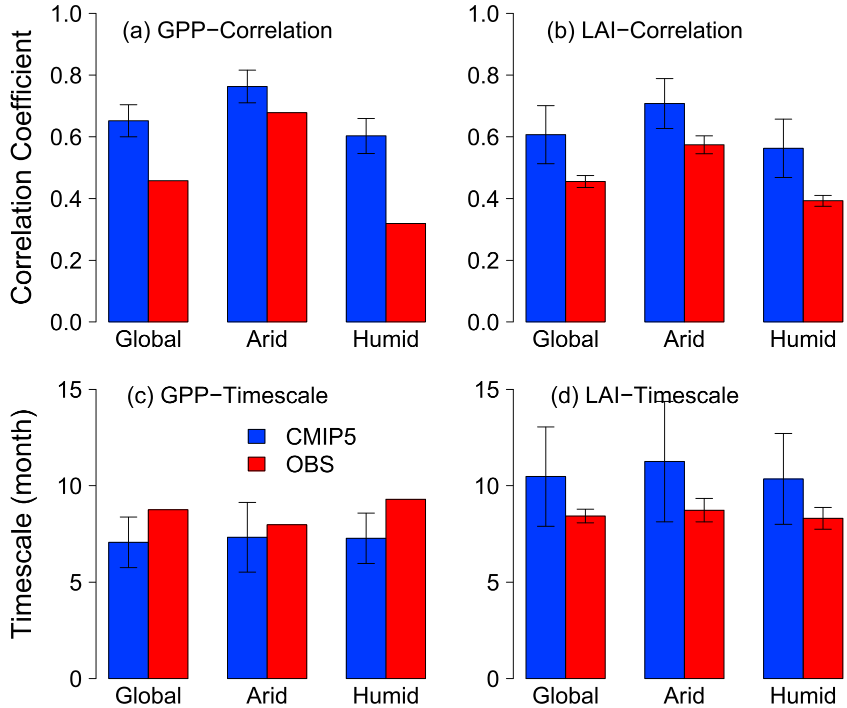

Figure 4. Global and regional averages of (a and b) the maximum correlations and (c and d) the SPI time scales (in months) at which the maximum correlations are obtained (excluding grid cells with no signification correlations). Maximum correlations are chosen from up to 288 (see Figure 3 for details) possible statistically significant correlation coefficients ($P < 0.05$) between GPP (or LAI) anomalies and SPI for each grid cell. Shown are means and 95% confidence limits among nine global averages for CMIP5 models or eight (two LAI data sets by four precipitation data sets) for the observation-based LAI responses. For the observation-based GPP (no error bars), only one data set is used. The differentiation between arid versus humid regions is based on the Köppen-Geiger climate classification [*Kottek et al.*, 2006] (see Figure S1).

As with GPP, LAI was more responsive in CMIP5 models to extreme dry/wet events than observation-based data sets. Globally, the reduction of LAI is -11.58×10^7 km^2 month if integrated over all extreme dry events, which is larger in magnitude than the mean of the observation-based data sets, -3.86×10^7 km^2 month (Figure 6). Further, the mean cumulative (total) increase of leaf area from all extreme wet events is 6.37×10^7 km^2 month, again higher than the observation-based of 1.56×10^7 km^2 month (Figure 6). Similar to GPP, the sensitivity of LAI to extreme dry/wet events is on average stronger in CMIP5 models compared to observation-based data sets if calculated based on all extreme events. As for GPP, through normalization, LAI from CMIP5 models still shows a higher mean sensitivity to extreme dry and wet events compared to that of observation-based data sets, although the differences are marginally nonsignificant ($P = 0.06$ for both dry and wet extreme events) (Figure S8). The difference in sensitivities derived for the two observation-based LAI data sets (GLOBMAP versus GLASS) is large, especially during extreme wet events, reflecting large uncertainties in observations (Figure S6). However, the sensitivity gap in observation-based LAI is reduced when LAI anomalies are normalized to 0 means and 1 standard deviations (Figures S6 and S8).

3.4. Correlations of GPP and LAI With Soil Moisture Anomalies

Maximum correlations (largest magnitude) between GPP and SMA (out of 12 possible correlations, one per growing season month) were higher for CMIP5 models (multimodel mean) than for the observation-based data set in some regions (e.g., southeastern United States, India, and southeastern China) (Figure 7). Maximum correlation coefficients vary among CMIP5 models and across regions (Figure S9). CMIP5 models had a negative GPP-SMA response for a substantial fraction of grid cells, particularly in the northern high latitudes (Figures 7 and S9). These negative correlations were less common in CMIP5 GPP-SPI responses (Figures 3 and S2).

The multimodel mean maximum LAI-SMA correlations from CMIP5 are smaller compared to those from observation-based data sets over a large portion of the land (Figures 7c and 7d), in contrast to LAI-SPI

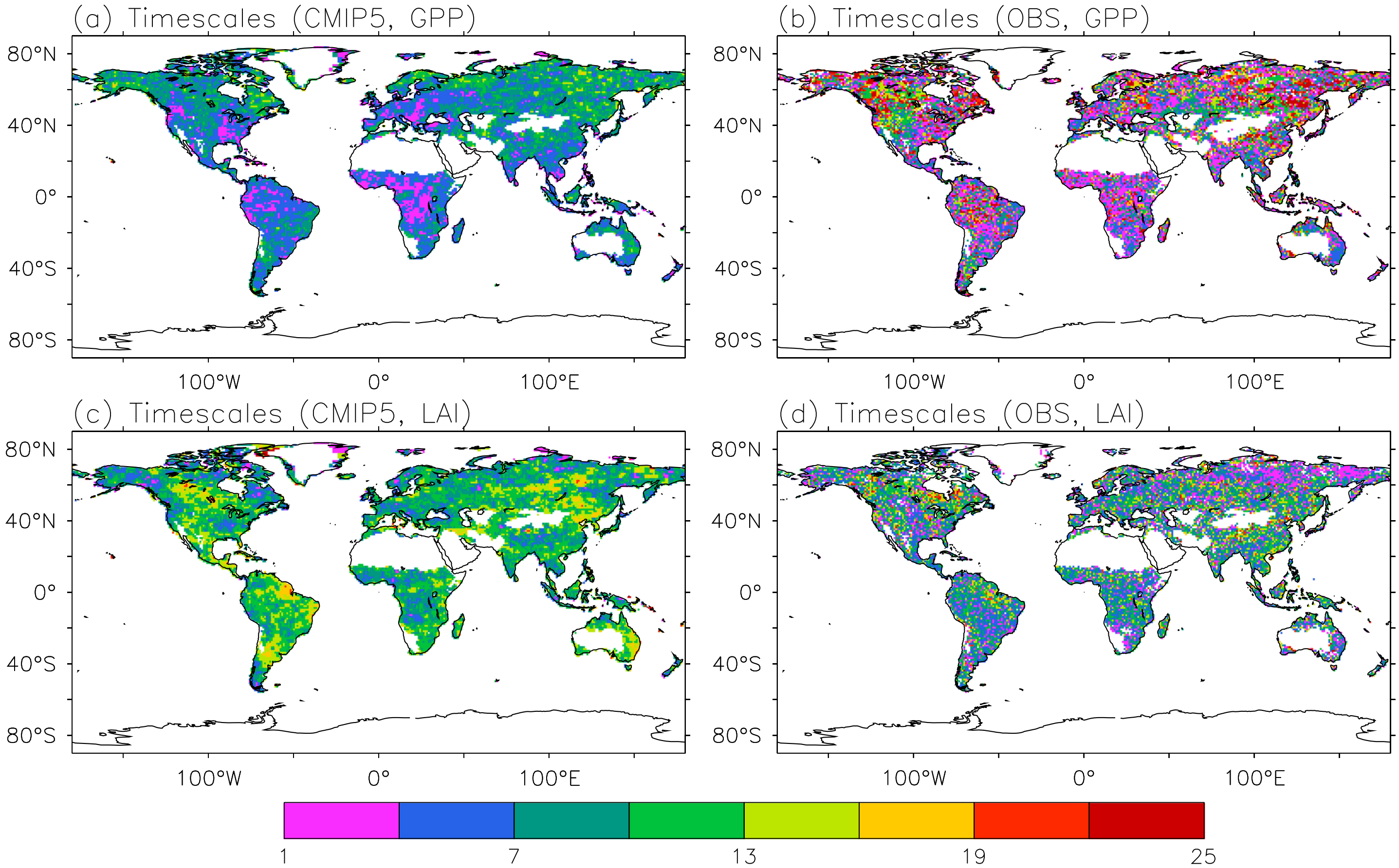

Figure 5. (a–d) SPI time scales (in months) at which the maximum correlations in Figure 3 are obtained.

correlations (Figures 3c and 3d). CMIP5 models vary greatly in the spatial pattern of the maximum LAI-SMA correlations (Figure S10). While maximum correlations from INMCM4 are predominantly positive and strong, Model for Interdisciplinary Research on Climate (MIROC) produces negative correlations over a large portion of the global land (Figure S10).

3.5. Extreme Event Analysis Based on Soil Moisture Anomalies

Modeled GPP reduction during extreme SMA dry events ranges from −0.82 to −25.93 GtC, which is more severe than the observation-based of −0.68 GtC (Figure 8a). Consequently, all of the CMIP5 models have higher sensitivity to drought (range from 1.5 to 19.34 gC m^{-2} $month^{-1}$) compared to the observation-based of 1.28 gC m^{-2} $month^{-1}$ (Figure S11a). The mean GPP increase in wet extreme events is higher, on average, in CMIP5 models than in observation-based data sets, but the difference is not significant (Figure 8b). When GPP anomalies are expressed as normalized anomalies, mean excursions in carbon budgets or sensitivities are still stronger in CMIP5 models compared to observation-based data set (Figures S12 and S13).

The mean response of LAI to extreme SMA dry/wet extremes is weaker in observation-based data compared to CMIP5 models (Figure 8). However, the differences between observation-based data and CMIP5 models are not statistically significant (within 95% confidence interval). The cumulative (total) leaf area changes over 1982–2005 range from an increase of 3.18×10^7 km^2 month to a reduction of -5.15×10^7 km^2 month during droughts and from an increase of 10.70×10^7 km^2 month to a reduction of -5.93×10^7 km^2 month during extreme wet conditions among CMIP5 models (Figure 8d). The inconsistent response of LAI in CMIP5 models to SMA extreme events (Figure 8) contrasts with the more consistent CMIP5 LAI responses to SPI extreme events (Figure 6). The overall pattern is similar in the analysis of sensitivity to droughts (total leaf area excursions divided by total volume of extreme dry/wet events) (Figure S11). However, GLASS (0.015 m^2 m^{-2})

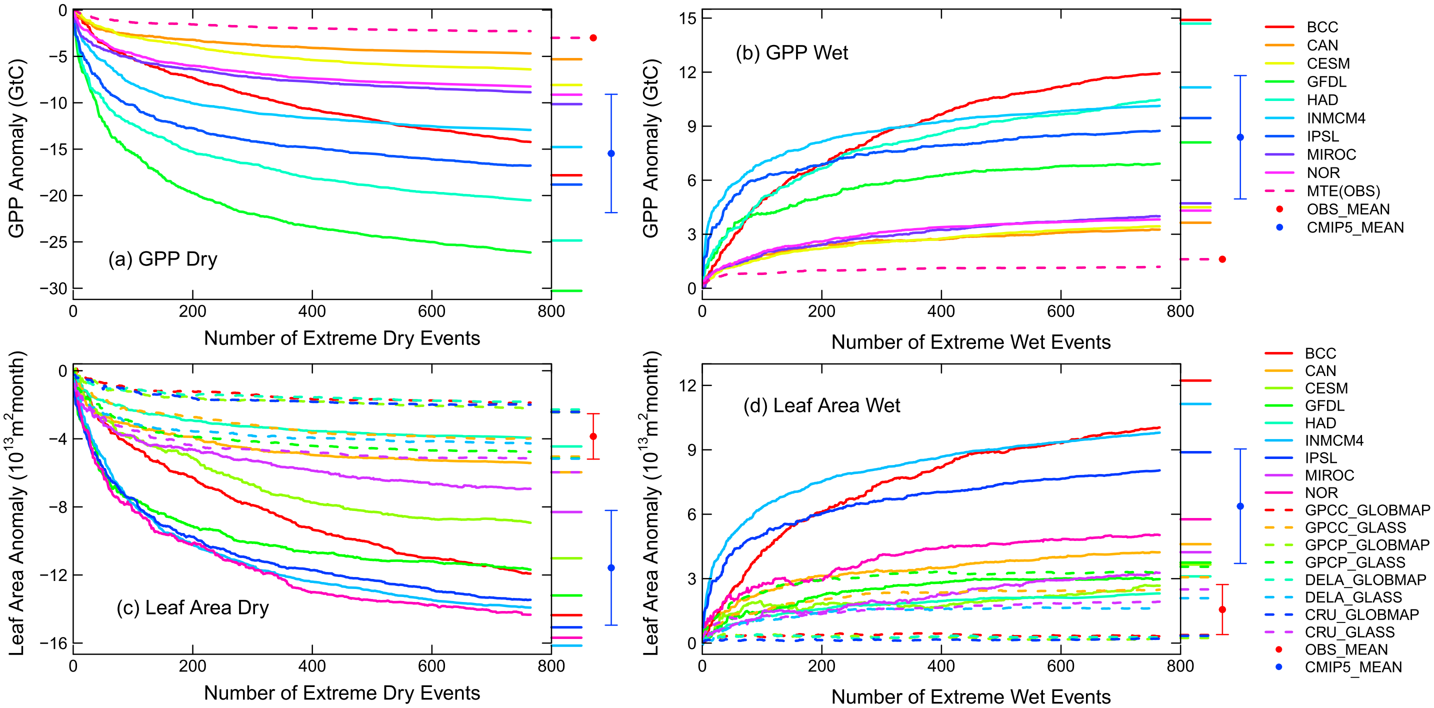

Figure 6. Cumulative GPP (or leaf area) reductions and increases during extreme dry and wet events, respectively. Extreme dry events are defined as three-dimensional (latitude × longitude × time) clusters with SPI < -2. Likewise, extreme wet events are three-dimensional wet (SPI > 2) clusters. For each panel, the in-box lines show the cumulative/sum of GPP (or LAI) anomalies spanned by the largest 760 extreme events (sorted by event-size) and the outside horizontal lines indicate the total from all of extreme events. The points and error bars are means and 95% confidence intervals among CMIP5 models (blue) or observations (red) over all extreme events. The solid lines represent the CMIP5 models, and the dashed lines are from observation-based data sets. Different colors correspond to different CMIP5 models or observations.

shows a much stronger sensitivity to wet extremes than GLOBMAP (0.002 m^2 m^{-2}) despite the volume of extreme wet events being close (9.37 versus 9.10 × 10^8 km^2 month from GLOBMAP and GLASS, respectively). The stronger sensitivity for GLASS compared to GLOBMAP stems largely from differences in the variances in these two data sets, since the sensitivity is comparable between GLASS and GLOBMAP when leaf area anomalies are expressed in normalized anomalies (Figure S13). Normalization of leaf area anomalies does not eliminate inconsistent responses (e.g., with both increases and reductions of leaf area in dry conditions) among CMIP5 models to extreme soil moisture events (Figures S12 and S13).

4. Discussion

Understanding and modeling the response of vegetation to drought are challenging due to the multifaceted nature of drought and markedly varied sensitivities to drought across land biomes and time scales. Here we evaluated the drought response of CMIP5 models from various perspectives: the overall correlations between global annual GPP (or LAI) anomalies versus the drought index SPI across time scales from 1 month to 2 years, the maximum correlation and time scale of maximum response from each grid cell and aggregated regionally or globally, during three-dimensional (longitude × latitude × time) extreme events, and based on soil moisture drought.

4.1. Response to Meteorological Drought

GPP in CMIP5 models is generally more responsive (or less buffered) to water anomalies than the observation-based GPP. This over-response of modeled GPP compared to observation is in line with the findings of *Piao et al.* [2013], which showed a stronger GPP-precipitation relationship in models. We further confirmed the over-response through the correlation with meteorological drought index SPI and through the analysis of GPP in extreme dry and wet events (events outside 2 standard deviations from the mean). Previous studies revealed that CMIP5 models tend to overestimate global mean GPP and LAI, and possibly

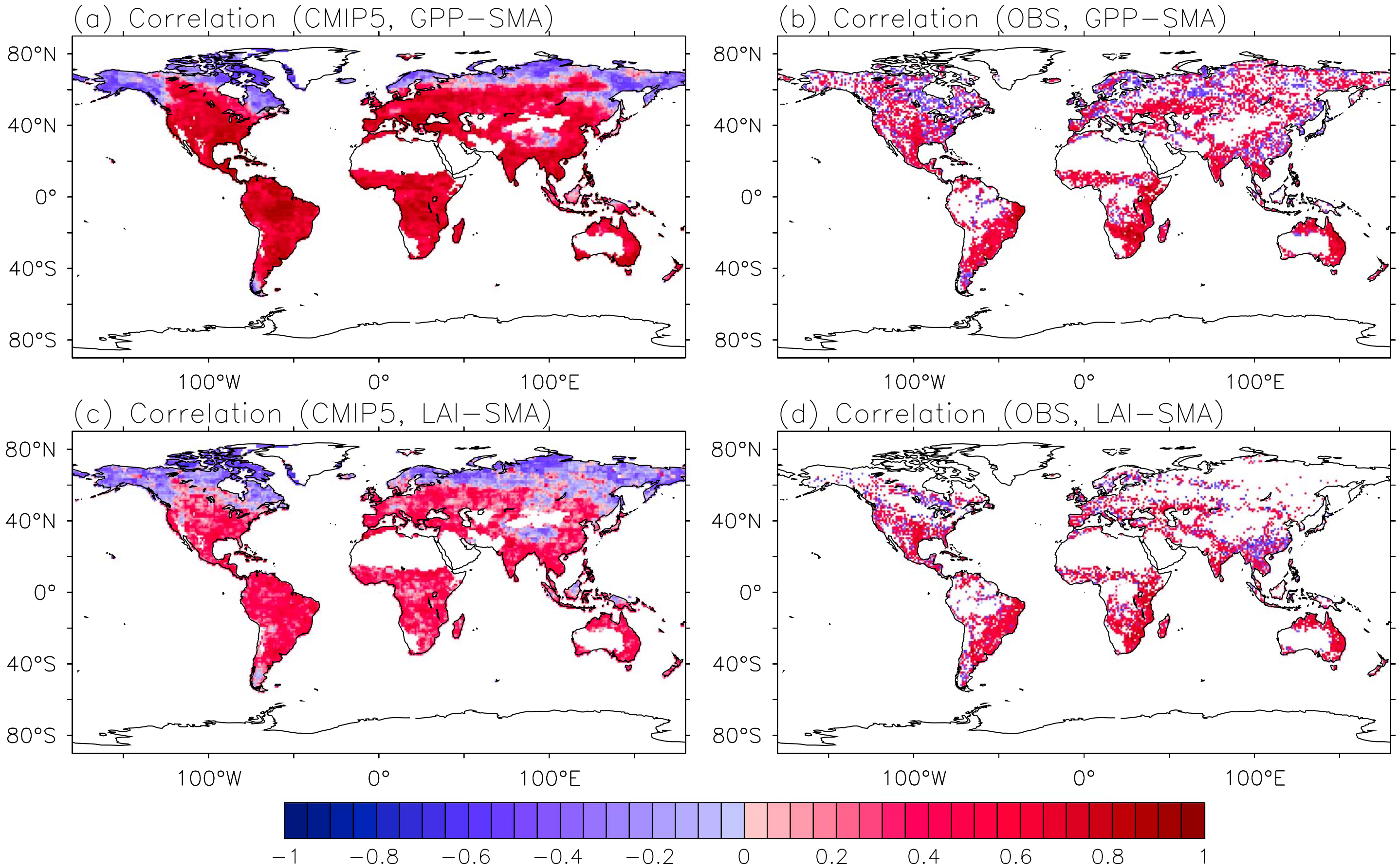

Figure 7. Spatial distribution of the maximum magnitude Pearson correlations between GPP (or LAI) anomalies and surface soil moisture anomalies (SMA) from the 12 possible significant ($P < 0.05$) correlations (one for each month of the year). Averages are shown for (a and c) CMIP5 models and (d) observation-based (OBS) LAI. A single analysis was performed for (b) observation-based GPP. The blank areas represent either polar frost or arid desert according to the Köppen-Geiger climate classification or locations with no significant correlations ($P > 0.05$). The analysis is restricted to the growing season for each grid cell, defined as months where mean monthly GPP or LAI is at least 25% of its maximum.

the magnitude of absolute GPP (or LAI) anomalies [*Shao et al.*, 2013]. It is possible that the variability of observation-based GPP is underestimated [*Jung et al.*, 2011; *Piao et al.*, 2013]. However, the higher GPP reduction in CMIP5 models compared to observation-based data persisted even if GPP anomalies were standardized. This suggests that extreme dry/wet events are more likely to produce GPP excursions in CMIP5 models compared to observation-based data. In addition, the spread of GPP reductions and increases in extreme dry and wet extreme events is large among CMIP5 models, with a 5.72 times difference in GPP reductions and 4.13 times in GPP increases, respectively. This large spread points to large uncertainties in capturing meteorological drought responses among CMIP5 models.

Similar to GPP, modeled LAI is on average more responsive to SPI compared to observation-based data, indicated by higher mean global annual correlations, higher mean maximum correlations over a large area of the global land, and stronger mean excursions and sensitivities to extreme dry or wet events. LAI depends on GPP and the allocation of GPP to leaves, while GPP is regulated by leaf area through the amount of leaf that performs photosynthesis. The strong link between GPP and LAI explains their large similarities in response to water anomalies. The analysis of LAI-SPI relationship provides complementary evidence to our findings based on GPP, since sources of observation-based LAI data sets are different from GPP.

The concept of drought time scale or legacy effect has been widely applied in drought studies [*Vicente-Serrano et al.*, 2013; *Anderegg et al.*, 2015; *Frank et al.*, 2015]. Response time scales are inconsistent between CMIP5 models and observation-based data sets or within individual CMIP5 models. CMIP5 models revealed overall shorter response time scales to meteorological drought compared to observation for GPP. The time lag of observed GPP (MTE-GPP) is partly driven by the forcing FAPAR that was used to produce the global

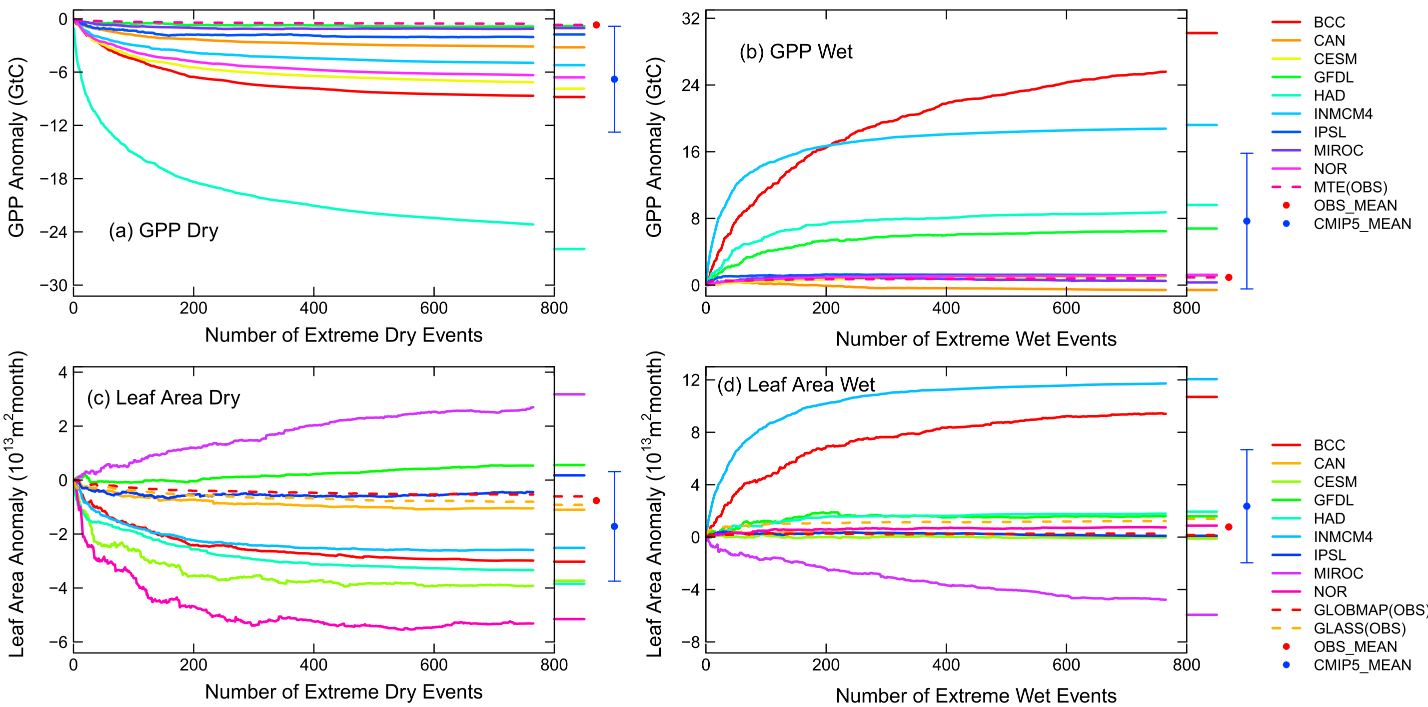

Figure 8. (a–d) Cumulative reductions or increases of GPP and LAI in response to surface soil moisture anomalies (SMA). Symbols and extreme events are defined as in Figure 6 except that dry extremes are defined as surface soil moisture anomalies (SMA) smaller than −2 and wet extremes as SMA greater than 2. The solid lines are from models, and the dashed lines are from observations. The points and error bars are means and 95% confidence intervals among CMIP5 models (blue) and observation-based data sets (red). Different colors correspond to different CMIP5 models or observations.

GPP product. As mentioned in *Jung et al.* [2011], MTE-GPP may not capture the relevant mechanisms that produce a true time lag, and time lag imprinted in MTE-GPP is probably underestimated. In this case, the true gap in time scales between CMIP5 models and the real GPP may be even larger. However, CMIP5 models have longer mean response time scales compared to observation for LAI. This discrepancy indicates potential insufficiency in CMIP5 models in capturing drought responses in processes that other than directly regulate GPP (e.g., leaf longevity and carbon allocation).

The insufficiency in modeled drought response time scale is further confirmed by the difference in arid versus humid regions. In CMIP5 models, arid regions on average need a slightly longer (or similar) lasting dry condition than humid regions to exhibit a maximum response in GPP, while humid regions exhibit a longer response time scale in observation-based data set compared to arid regions (Figure 4c). Our classification of arid and humid regions is broad and did not differentiate between arid and semiarid or humid versus semihumid which might obscure the characteristic time scale of drought response across different biomes. *Vicente-Serrano et al.* [2013] pointed to the biome-dependent time scale of drought responses; however, direct biome by biome comparison is not possible due to differences in the numbers and types of PFTs simulated in CMIP5 models. Nevertheless, our broad classification of arid versus humid biomes revealed potential gaps in models in capturing time scale of drought impacts. Even within CMIP5 models, we found markedly different response time scale pattern for LAI, and further model developments that address lagged response may be critical toward predicting carbon flux anomalies during droughts.

4.2. Response to Soil Moisture Drought

GPP and LAI generally respond stronger to meteorological drought in CMIP5 models compared to observation-based data sets. The stronger response to hydrological anomalies in models is less obvious when soil moisture anomalies are used instead of SPI. Drought responses of modeled GPP stem largely from the water stress functions and should also reflect plant water availability. The large spread of the modeled sensitivity of GPP to extreme dry events indicates either differences in water stress functions or a large range of modeled moisture availability that regulate plant activity, or both. Water stress functions in models vary in their formulations and parameterizations. Some of them down-regulate the leaf level maximum carboxylation capacity of Rubisco, while some scale the potential photosynthetic rate; some formulations of down-regulation are based on soil water conditions alone, while there are models that also explicitly take into account the leaf water demand; some of them are parameterized based on soil matrix potential, while some are based on volume soil water content (see Text S1 for detailed information). These various forms of water stress functions have the potential to create differences in the response of GPP when water is limiting. Modeled LAI responds to SPI coherently with reductions in droughts, but SMA droughts are associated with both reductions and enhancements in different models. Divergent responses of modeled LAI to SMA detected in our study may stem from processes other than water stress function and plant-available water, such as difference in the representation of GPP allocation to leaf biomass, leaf phenology, and functions to derive leaf area.

One of the impediments toward mechanistic understanding of vegetation's drought response is the limitation of knowledge and empirical data on plant-available water, which is the actual pool of water that plants can access to support transpiration. While we found an overestimation of the response of vegetation to meteorological drought, *Rebel et al.* [2012] suggested a possible underestimation of the response of vegetation to drought in Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) land surface model based on the slower decrease of root zone soil moisture after rain compared to surface soil moisture derived from remote-sensed AMSR-E (Land Surface Parameter Model), while the true plant-available moisture remained unknown. Although the strong link between surface and root-zone soil moisture is frequently documented from in situ measurements [e.g., *Albergel et al.*, 2008; *Rebel et al.*, 2012; *Ford et al.*, 2014], it is unclear to what extent soil water anomalies experienced by plants are captured by surface soil moisture anomalies since the variability and availability of soil moisture change with depth [*Hirschi et al.*, 2014; *Cheng et al.*, 2016]. The stronger response of vegetation to SPI compared to SMA detected in our study may be caused by lack of representation of plant-available water anomaly by surface SMA ($<$10 cm). And the response pattern might be different if the whole plant-available water (or deep soil water) is used in evaluation instead of using only the surface layer. Root zone water availability is a better proxy of the dry/wet stress experienced by plant; however, observation-based root zone water is not available in large scale. SPI is widely used as a drought indicator, and short-term SPI has been shown to have good correlations with soil moisture availability/anomaly [*Sims et al.*, 2002; *Scaini et al.*, 2015], but only reflect one side of the multiple processes (e.g., interception and drainage are not captured in SPI) regulating soil water dynamics. Remotely sensed soil moisture anomalies have been shown to be in line with SPI at specific locations and have a high response to precipitation [*Scaini et al.*, 2015]. Nevertheless, the difference between evaluation of the response to SPI and SMA in this study calls for improvement in understanding of plant-available water. The over-response of vegetation to meteorological drought in this study may incorporate uncertainties in the treatment of precipitation that ultimately ends up as water that is available for plant uptake.

4.3. Uncertainties

Quantifying responses to SPI based on observation is afflicted with uncertainties if errors in drought and observation-based data sets are taken into account. We explored the impact of possible errors in SPI on the model-data comparison. Regression slope of LAI to drought is enhanced when errors in the explanatory variable SPI are included. The enhancement alters the result of model-observation comparison in some cases. Ignoring errors in explanatory variable can lead to biased result especially when driver errors are large [*Lichstein et al.*, 2014]. Our illustration through the response of LAI indicates the necessity of taking into account of uncertainties in explanatory variables in future model-data comparison. Further, results for tropical forests derived from observation-based GPP are less reliable: observation-based GPP is obtained from training the FLUXNET flux tower observations with a large number of sites in temperate ecosystems and few located in the tropics. The satellite FAPAR, the variable based on which flux tower measurements are extrapolated, is subject to contamination by cloud especially in tropical forests [*Jung et al.*, 2011]. We use LAI as a complementary variable to evaluate drought response. LAI errors cannot be excluded, particularly in tropical forest with the satellite based LAI [*Fang et al.*, 2012]. Compared to CMIP5 models and observation-based GPP (upscaled to the globe with deterministic algorithm), the noise in LAI products is likely to be larger, which could render the observation-based LAI correlations lower than CMIP5 models. The global and regional average of the highest correlations and the SPI time scales at which the highest correlations are obtained are based on areas with statistically significant ($P < 0.05$) correlations, which may bias the global or regional means with some land areas excluded from the calculation.

In addition to uncertainties mentioned above, the observation-based soil moisture (i.e., ECV_SM) is also error-prone under dense vegetation especially in tropical forests [*Dorigo et al.*, 2010; *Liu et al.*, 2011]. ECV_SM has more data gaps compared to CMIP5 models and other observation-based data sets. We have not assessed how drought responses are affected by data gaps. Other sources, such as model implementations of disturbance regimes (e.g., land use change and fire), model resolution and regridding method, may also contribute to uncertainties in drought response and need to be treated with caution.

4.4. Implications for Future Studies

Efforts are underway to better understand and forecast terrestrial ecosystems' response to drought, such as through precipitation manipulation experiments and site-specific model-data intercomparisons [*Beier et al.*, 2012; *Powell et al.*, 2013; *Knapp et al.*, 2015]. Our study focuses on the performance of large scale drought response from CMIP5 models and serves as a baseline to identify current knowledge gaps in models, to guide manipulative drought experiments (such as those in Drought-Net:www.drought-net.org) and for future integration of observations with Earth system models.

Several of our findings are beneficial for future studies. First, multiple perspective evaluation is helpful in providing a thorough picture on the performance of modeled vegetation's response to drought. Drought is a complex phenomenon, and drought impacts are multifaceted. Although modeled GPP is more sensitive compared to observation-based data on short to medium time scale, the oversensitivity does not hold for long-term (e.g., 2 years in this study) meteorological drought. Modeled LAI is more responsive to meteorological drought; however, the over-response is not true if based on drought quantified by soil moisture. Future manipulative drought experiments should be designed to provide model relevant measurements that span different temporal scales, while modelers should consider validating multiple variables and processes, especially model formulations of plant transpiration/GPP relationships and parametrization that affect LAI (including leaf carbon allocation and turnover in response to drought). Second, our results suggest that at short time scales (e.g., 3 months) the GPP response to drought is oversensitive to drought compared to observation. As the world is projected to experience more severe, frequent, or/and widespread droughts and wet extremes under future global warming [*Dai*, 2013; *Orlowsky and Seneviratne*, 2013], GPP swings in models may even more deviating from the realistic variability under future climate. The oversensitivity may be improved through incorporating processes that are currently in lack in most CMIP5 models, such as nitrogen interactions with drought [*Huang and Gerber*, 2016] and realistic plant community dynamics and functional diversity [*Weng et al.*, 2015]. Third, work is required to improve mechanistic understandings of how water stress affects vegetation's activity. Models vary in their water stress functions which may result in difference in model performance. Manipulative drought experiments with focus on how water stress regulates plant activity across terrestrial ecosystems are essential to step forward. Fourth, additional focus needs to be put on LAI dynamics. We found distinct response pattern of LAI among individual models with respect to sensitivity and time scale in response to SPI (Figure S5), with respect to SMA (Figure S10), and in the response to extreme SMA events (Figures 8 and S11). Since LAI is a common parameter used to upscale leaf level productivity to the ecosystem level, accurate simulation of LAI response is beneficial for carbon cycling-drought studies. Finally, empirical global data sets with low uncertainty is essential in benchmarking models and identifying model insufficiency. Observation data sets that are afflicted with uncertainties can mislead the evaluation of model's performance and should be taken into account in future studies. Especially important and urgent is the generation of a global data set of plant-available water perhaps through integration of radar-based data, in situ measurements, and hydrologic modeling. While SPI and SMA are commonly used but indirect indicators of water stress experienced by plants, an empirical based data set of plant-available water may be important to reveal the actual water stress that regulate vegetation's response and would help further to evaluate model performance.

5. Conclusions

We compared drought responses of vegetation activities derived from nine CMIP5 models with observations based on meteorological drought index SPI and surface soil moisture anomalies (SMA). Several lines of evidence suggest an oversensitivity of CMIP5 models to SPI, including higher mean global annual correlation, higher maximum correlations, and higher reduction/increase in extreme events for both GPP and LAI. However, the oversensitivity from CMIP5 models is less apparent based on surface SMA for GPP and even in contradictory when the LAI response is evaluated. Future work should be directed toward a better understanding of plant-available water and plant stress functions that regulate vegetation activities.

Acknowledgments
The authors thank Matthew Cohen and Patrick Inglett from the University of Florida for their valuable suggestions. All data are available through e-mail request to Yuanyuan Huang (yyhuang@ufl.edu).

References

Adler, R. F., et al. (2003), The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, *4*(6), 1147–1167.
Albergel, C., C. Ruediger, T. Pellarin, J. C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, B. Piguet, and E. Martin (2008), From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations, *Hydrol. Earth Syst. Sci.*, *12*(6), 1323–1337.
Anav, A., P. Friedlingstein, M. Kidston, L. Bopp, P. Ciais, P. Cox, C. Jones, M. Jung, R. Myneni, and Z. Zhu (2013), Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth system models, *J. Clim.*, *26*(18), 6801–6843.
Anderegg, W. R. L., L. Plavcova, L. D. L. Anderegg, U. G. Hacke, J. A. Berry, and C. B. Field (2013), Drought's legacy: Multiyear hydraulic deterioration underlies widespread aspen forest die-off and portends increased future risk, *Global Change Biol.*, *19*(4), 1188–1196.
Anderegg, W. R. L., et al. (2015), Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models, *Science*, *349*(6247), 528–532.
Arora, V. K., and G. J. Boer (2010), Uncertainties in the 20th century carbon budget associated with land use change, *Global Change Biol.*, *16*(12), 3327–3348.
Ball, J. T., I. E. Woodrow, and J. A. Berry (1987), A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in *Progress in Photosynthesis Research*, edited by I. Biggins, pp. 221–224, Springer, Netherlands.
Baret, F., et al. (2007), LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION—Part 1: Principles of the algorithm, *Remote Sens. Environ.*, *110*(3), 275–286.
Beier, C., et al. (2012), Precipitation manipulation experiments—Challenges and recommendations for the future, *Ecol. Lett.*, *15*(8), 899–911.
Bigler, C., D. G. Gavin, C. Gunning, and T. T. Veblen (2007), Drought induces lagged tree mortality in a subalpine forest in the Rocky Mountains, *Oikos*, *116*(12), 1983–1994.
Bonan, G. B. (1996), A Land Surface Model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide, *NCAR Tech. Note NCAR/TN-417+STR*.
Capehart, W. J., and T. N. Carlson (1997), Decoupling of surface and near-surface soil water content: A remote sensing perspective, *Water Resour. Res.*, *33*, 1383–1395, doi:10.1029/97WR00617.
Chen, T., G. R. van der Werf, R. A. M. de Jeu, G. Wang, and A. J. Dolman (2013), A global analysis of the impact of drought on net primary productivity, *Hydrol. Earth Syst. Sci.*, *17*(10), 3885–3894.
Cheng, L. Y., M. Hoerling, A. AghaKouchak, B. Livneh, X. W. Quan, and J. Eischeid (2016), How has human-induced climate change affected california drought risk? *J. Clim.*, *29*(1), 111–120.
Ciais, P., et al. (2005), Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, *437*(7058), 529–533.
Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*(2–4), 107–136.
Collatz, G. J., M. Ribas-Carbo, and J. A. Berry (1992), Coupled photosynthesis-stomatal conductance model for leaves of C4 plants, *Aust. J. Plant Physiol.*, *19*(5), 519–538.
Cox, P. M. (2001), Description of the "TRIFFID" Dynamic Global Vegetation Model, Rep., Hadley Cent. Tech. Note 24, Hadley Cent., Met Off., U. K.
Dai, A. (2013), Increasing drought under global warming in observations and models, *Nat. Clim. Change*, *3*(1), 52–58.
Dorigo, W., R. de Jeu, D. Chung, R. Parinussa, Y. Liu, W. Wagner, and D. Fernandez-Prieto (2012), Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture, *Geophys. Res. Lett.*, *39*, L18405, doi:10.1029/2012GL052988.
Dorigo, W. A., K. Scipal, R. M. Parinussa, Y. Y. Liu, W. Wagner, R. A. M. de Jeu, and V. Naeimi (2010), Error characterisation of global active and passive microwave soil moisture datasets, *Hydrol. Earth Syst. Sci.*, *14*(12), 2605–2616.
Dorigo, W. A., et al. (2014), Evaluation of the ESA CCI soil moisture product using ground-based observations.
Dufresne, J. L., et al. (2013), Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5, *Clim. Dyn.*, *40*(9-10), 2123–2165.
Dunne, J. P., et al. (2012), GFDL's ESM2 global coupled climate-carbon Earth system models. Part I: Physical formulation and baseline simulation characteristics, *J. Clim.*, *25*(19), 6646–6665.
Fang, H., S. Wei, C. Jiang, and K. Scipal (2012), Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method, *Remote Sens. Environ.*, *124*, 610–621.
Farooq, M., A. Wahid, N. Kobayashi, D. Fujita, and S. M. A. Basra (2009), Plant drought stress: Effects, mechanisms and management, *Agron. Sustainable Dev.*, *29*(1), 185–212.
Farquhar, G. D., S. V. Caemmerer, and J. A. Berry (1980), A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, *Planta*, *149*(1), 78–90.
Ford, T. W., E. Harris, and S. M. Quiring (2014), Estimating root zone soil moisture using near-surface observations from SMOS, *Hydrol. Earth Syst. Sci.*, *18*(1), 139–154.
Frank, D. A., et al. (2015), Effects of climate extremes on the terrestrial carbon cycle: Concepts, processes and potential future impacts, *Global Change Biol.*, *21*(8), 2861–2880.
Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: Algorithms and early results, *Remote Sens. Environ.*, *83*(1–2), 287–302.
Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti (2014), Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, *J. Clim.*, *27*(2), 511–526.
Fuller, W. A. (1987), *Measurement Error Models*, John Wiley, New Nork.
Gent, P. R., et al. (2011), The Community Climate System Model version 4, *J. Clim.*, *24*(19), 4973–4991.
Hirschi, M., B. Mueller, W. Dorigo, and S. I. Seneviratne (2014), Using remotely sensed soil moisture for land-atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability, *Remote Sens. Environ.*, *154*, 246–252.
Huang, Y., and S. Gerber (2016), Nitrogen restrictions buffer modeled interactions of water with the carbon cycle, *J. Geophys. Res. Biogeosci.*, *121*, 218–232, doi:10.1002/2015JG003148.
Ji, J., M. Huang, and K. Li (2008), Prediction of carbon exchanges between China terrestrial ecosystem and atmosphere in 21st century, *Sci. China, Ser. D: Earth Sci.*, *51*(6), 885–898.

Ji, L., and A. J. Peters (2003), Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices, *Remote Sens. Environ.*, *87*(1), 85–98.

Jones, C. D., et al. (2011), The HadGEM2-ES implementation of CMIP5 centennial simulations, *Geosci. Model Dev.*, *4*(3), 543–570.

Jones, P. D., and I. Harris (2013), CRU TS3.21: Climatic Research Unit (CRU) Time-Series (TS) version 3.21 of high resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2012), doi:10.5285/D0E1585D-3417-485F-87AE-4FCECF10A992.

Jung, M., et al. (2011), Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.*, *116*, G00J07, doi:10.1029/2010JG001566.

Knapp, A. K., C. J. W. Carroll, E. M. Denton, K. J. La Pierre, S. L. Collins, and M. D. Smith (2015), Differential sensitivity to regional-scale drought in six central US grasslands, *Oecologia*, *177*(4), 949–957.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006), World map of the Koppen-Geiger climate classification updated, *Meteorol. Z.*, *15*(3), 259–263.

Krinner, G., N. Viovy, N. de Noblet-Ducoudre, J. Ogee, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.

Lei, T., J. Wu, X. Li, G. Geng, C. Shao, H. Zhou, Q. Wang, and L. Liu (2015), A new framework for evaluating the impacts of drought on net primary productivity of grassland, *Sci. Total Environ.*, *536*, 161–172.

Le Quere, C., et al. (2009), Trends in the sources and sinks of carbon dioxide, *Nat. Geosci.*, *2*(12), 831–836.

Leuning, R. (1995), A critical appraisal of a combined stomatal-photosynthesis model for C3 plants, *Plant, Cell Environ.*, *18*(4), 339–355.

Liang, S., and Z. Xiao (2012), Global land surface products: Leaf area index product data collection(1985–2010), Beijing Normal Univ.

Lichstein, J. W., N.-Z. Golaz, S. Malyshev, E. Shevliakova, T. Zhang, J. Sheffield, R. A. Birdsey, J. L. Sarmiento, and S. W. Pacala (2014), Confronting terrestrial biosphere models with forest inventory data, *Ecol. Appl.*, *24*(4), 699–715.

Liu, Y., R. Liu, and J. M. Chen (2012), Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data, *J. Geophys. Res.*, *117*, G04003, doi:10.1029/2012JG002084.

Liu, Y., Y. Zhou, W. Ju, S. Wang, X. Wu, M. He, and G. Zhu (2014), Impacts of droughts on carbon sequestration by China's terrestrial ecosystems from 2000 to 2011, *Biogeosciences*, *11*(10), 2583–2599.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. De Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. P. Evans (2011), Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, *Hydrol. Earth Syst. Sci.*, *15*(2), 425–436.

Lloyd-Hughes, B. (2012), A spatio-temporal structure-based approach to drought characterisation, *Int. J. Climatol.*, *32*(3), 406–418.

Lloyd-Hughes, B., and M. A. Saunders (2002), A drought climatology for Europe, *Int. J. Climatol.*, *22*(13), 1571–1592.

Loew, A., T. Stacke, W. Dorigo, R. de Jeu, and S. Hagemann (2013), Potential and limitations of multidecadal satellite soil moisture observations for selected climate model evaluation studies, *Hydrol. Earth Syst. Sci.*, *17*(9), 3523–3542.

Mao, J., P. E. Thornton, X. Shi, M. Zhao, and W. M. Post (2012), Remote sensing evaluation of CLM4 GPP for the period 2000-09, *J. Clim.*, *25*(15), 5327–5342.

McKee, T. B., N. J. Doesken, and J. Kleist (1993), The relationship of drought frequency and duration to time scales, in *Proceedings of the 8th Conference of Applied Climatology*, 17–22 January, Anaheim, Calif., Am. Meterol. Soc., Boston, Mass., pp. 179–184.

Niu, S., Y. Luo, D. Li, S. Cao, J. Xia, J. Li, and M. D. Smith (2014), Plant growth and mortality under climatic extremes: An overview, *Environ. Exp. Bot.*, *98*, 13–19.

Orlowsky, B., and S. I. Seneviratne (2013), Elusive drought: Uncertainty in observed trends and short- and long-term CMIP5 projections, *Hydrol. Earth Syst. Sci.*, *17*(5), 1765–1781.

Palmer, W. C. (1965), Meteorological droughts, Weather Bur. Res. Pap. 45, 58 pp., U.S. Dep. of Comm.

Phillips, O. L., et al. (2009), Drought sensitivity of the Amazon rainforest, *Science*, *323*(5919), 1344–1347.

Phillips, O. L., et al. (2010), Drought-mortality relationships for tropical forests, *New Phytol.*, *187*(3), 631–646.

Piao, S., et al. (2013), Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends, *Global Change Biol.*, *19*(7), 2117–2132.

Porporato, A., E. Daly, and I. Rodriguez-Iturbe (2004), Soil water balance and ecosystem response to climate change, *Am. Nat.*, *164*(5), 625–632.

Potter, C., S. Klooster, C. Hiatt, V. Genovese, and J. C. Castilla-Rubio (2011), Changes in the carbon cycle of Amazon ecosystems during the 2010 drought, *Environ. Res. Lett.*, *6*(3), 034024.

Poulter, B., et al. (2014), Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, *Nature*, *509*(7502), 600–603.

Powell, T. L., et al. (2013), Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought, *New Phytol.*, *200*(2), 350–364.

Pratola, C., B. Barrett, A. Gruber, G. Kiely, and E. Dwyer (2014), Evaluation of a global soil moisture product from finer spatial resolution SAR data and ground measurements at Irish sites, *Remote Sens.*, *6*(9), 8190–8219.

Rebel, K. T., R. A. M. de Jeu, P. Ciais, N. Viovy, S. L. Piao, G. Kiely, and A. J. Dolman (2012), A global analysis of soil moisture derived from satellite observations and a land surface model, *Hydrol. Earth Syst. Sci.*, *16*(3), 833–847.

Reichstein, M., et al. (2013), Climate extremes and the carbon cycle, *Nature*, *500*(7462), 287–295.

Reyer, C. P. O., et al. (2013), A plant's perspective of extremes: Terrestrial plant responses to changing climatic variability, *Global Change Biol.*, *19*(1), 75–89.

Sala, O. E., L. A. Gherardi, L. Reichmann, E. Jobbagy, and D. Peters (2012), Legacies of precipitation fluctuations on primary production: Theory and data synthesis, *Philos. Trans. R. Soc. B*, *367*(1606), 3135–3144.

Sato, H., A. Itoh, and T. Kohyama (2007), SEIB-DGVM: A new dynamic global vegetation model using a spatially explicit individual-based approach, *Ecol. Modell.*, *200*(3–4), 279–307.

Scaini, A., N. Sanchez, S. M. Vicente-Serrano, and J. Martinez-Fernandez (2015), SMOS-derived soil moisture anomalies and drought indices: A comparative analysis using in situ measurements, *Hydrol. Processes*, *29*(3), 373–383.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese (2011), GPCC full data reanalysis version 6.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data, doi:10.5676/DWD_GPCC/FD_M_V6_050.

Schwalm, C. R., et al. (2010), Assimilation exceeds respiration sensitivity to drought: A FLUXNET synthesis, *Global Change Biol.*, *16*(2), 657–670.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling (2010), Investigating soil moisture-climate interactions in a changing climate: A review, *Earth Sci. Rev.*, *99*(3–4), 125–161.

Shao, P., X. Zeng, K. Sakaguchi, R. K. Monson, and X. Zeng (2013), Terrestrial carbon cycle: Climate relations in eight CMIP5 Earth system models, *J. Clim.*, *26*(22), 8744–8764.

Shevliakova, E., S. W. Pacala, S. Malyshev, G. C. Hurtt, P. C. D. Milly, J. P. Caspersen, L. T. Sentman, J. P. Fisk, C. Wirth, and C. Crevoisier (2009), Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, *Global Biogeochem. Cycles*, *23*, GB2022, doi:10.1029/2007GB003176.

Sims, A. P., D. D. S. Niyogi, and S. Raman (2002), Adopting drought indices for estimating soil moisture: A North Carolina case study, *Geophys. Res. Lett.*, *29*(8), 1183, doi:10.1029/2001GL013343.

Taylor, K. E., R. J. Stouffer, and G. A. Meeh (2011), A summary of the CMIP5 experiment design. [Available at http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf.]

Thornton, P. E., and N. E. Zimmermann (2007), An improved canopy integration scheme for a land surface model with prognostic canopy structure, *J. Clim.*, *20*(15), 3902–3923.

Thornton, P. E., et al. (2002), Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, *Agric. For. Meteorol.*, *113*(1–4), 185–222.

Tjiputra, J. F., C. Roelandt, M. Bentsen, D. M. Lawrence, T. Lorentzen, J. Schwinger, O. Seland, and C. Heinze (2013), Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM), *Geosci. Model Dev.*, *6*(2), 301–325.

Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison (2013), Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, *Biogeosciences*, *10*(3), 1717–1736.

Trenberth, K. E., A. Dai, G. van der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield (2014), Global warming and changes in drought, *Nat. Clim. Change*, *4*(1), 17–22.

Vicente-Serrano, S. M., et al. (2013), Response of vegetation to drought time-scales across global land biomes, *Proc. Natl. Acad. Sci. U.S.A.*, *110*(1), 52–57.

Vicente-Serrano, S. M., J. Julio Camarero, and C. Azorin-Molina (2014), Diverse responses of forest growth to drought time-scales in the Northern Hemisphere, *Global Ecol. Biogeogr.*, *23*(9), 1019–1030.

Volodin, E. M., N. A. Dianskii, and A. V. Gusev (2010), Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations, *Izv. Atmos. Oceanic Phys.*, *46*(4), 414–431.

Watanabe, S., et al. (2011), MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments, *Geosci. Model Dev.*, *4*(4), 845–872.

Weng, E., and Y. Luo (2008), Soil hydrological properties regulate grassland ecosystem responses to multifactor global change: A modeling analysis, *J. Geophys. Res.*, *113*, G03003, doi:10.1029/2007JG000539.

Weng, E. S., S. Malyshev, J. W. Lichstein, C. E. Farrior, R. Dybzinski, T. Zhang, E. Shevliakova, and S. W. Pacala (2015), Scaling from individual trees to forests in an Earth system modeling framework using a mathematically tractable model of height-structured competition, *Biogeosciences*, *12*(9), 2655–2694.

Willmott, C. J., and K. Matsuura (2012), *Terrestrial Precipitation: 1900–2010 Gridded Monthly Time Series (1900–2010)*. [Available at https://www.esrl.noaa.gov/psd/data/gridded/data.UDel_AirT_Precip.html, Accessed date 03-2014.]

WMO (2012), World Meteorological Organization, 2012: Standardized Precipitation Index User Guide (M. Svoboda, M. Hayes and D. Wood). (WMO-No. 1090), Geneva.

Wu, T., et al. (2013), Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century, *J. Geophys. Res. Atmos.*, *118*, 4326–4347, doi:10.1002/jgrd.50320.

Xiao, Z., S. Liang, J. Wang, P. Chen, X. Yin, L. Zhang, and J. Song (2014), Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance, *IEEE Trans. Geosci. Remote Sens.*, *52*(1), 209–223.

Zeng, J., Z. Li, Q. Chen, H. Bi, J. Qui, and P. Zou (2015), Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations.

Zhao, M., and S. W. Running (2010), Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, *Science*, *329*(5994), 940–943.

Zscheischler, J., M. D. Mahecha, S. Harmeling, and M. Reichstein (2013), Detection and attribution of large spatiotemporal extreme events in Earth observation data, *Ecol. Inf.*, *15*, 66–73.

Zscheischler, J., et al. (2014a), A few extreme events dominate global interannual variability in gross primary production, *Environ. Res. Lett.*, *9*(3), 035001.

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