Toward a better integration of biological data from precipitation manipulation experiments into Earth system models

Nicholas G. Smith1, Vikki L. Rodgers2, Edward R. Brzostek2, Andrew Kulmatiski3, Meghan L. Avolio3, David L. Hoover4, Sally E. Koerner5, Kerstin Grant6, Anke Jentsch7, Simone Fatichi8, and Dev Niyogi9,10,11

1Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA, 2Division of Math and Science, Babson College, Babson Park, Massachusetts, USA, 3Department of Biology, Indiana University, Bloomington, Indiana, USA, 4Department of Wildland Resources and Ecology Center, Utah State University, Logan, Utah, USA, 5Department of Biology, University of Utah, Salt Lake City, Utah, USA, 6Biology Department, Colorado State University, Fort Collins, Colorado, USA, 7Biological Department, University of New Mexico, Albuquerque, New Mexico, USA, 8Department of Disturbance Ecology, University of Bayreuth, Bayreuth, Germany, 9Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland, 10Earth, Atmospheric, Planetary Sciences, Purdue University, West Lafayette, Indiana, USA, 11Department of Agronomy, Purdue University, West Lafayette, Indiana, USA

Abstract The biological responses to precipitation within the terrestrial components of Earth system models, or land surface models (LSMs), are mechanistically simple and poorly constrained, leaving projections of terrestrial ecosystem functioning and feedbacks to climate change uncertain. A number of field experiments have been conducted or are underway to test how changing precipitation will affect terrestrial ecosystems. Results from these experiments have the potential to vastly improve modeled processes. However, the transformation of experimental results into model improvements still represents a grand challenge. Here we review the current state of precipitation manipulation experiments and the precipitation responses of biological processes in LSMs to explore how these experiments can help improve model realism. First, we discuss contemporary precipitation projections and then review the structure and function of current-generation LSMs. We then examine different experimental designs and discuss basic variables that, if measured, would increase a field experiment’s usefulness in a modeling context. Next, we compare biological processes commonly measured in the field with their model analogs and find that, in many cases, the way these processes are measured in the field is not compatible with the way they are represented in LSMs, an effect that hinders model development. We then discuss the challenge of scaling from the plot to the globe. Finally, we provide a series of recommendations aimed to improve the connectivity between experiments and LSMs and conclude that studies designed from the perspective of researchers in both communities will provide the greatest benefit to the broader global change community.

1. Introduction

Current and projected shifts in precipitation have the potential to impact vital terrestrial ecosystem functions including worldwide food production, carbon storage, and patterns of biodiversity loss [Weltzin et al., 2003]. Precipitation is a primary driver of cellular- [e.g., Rodgers et al., 2012], individual- [e.g., Hanson et al., 2001], population- [e.g., Avolio et al., 2012], community- [e.g., Kulmatiski and Beard, 2013], and ecosystem-scale [e.g., Suseela and Dukes, 2012] processes across a variety of temporal scales (Figure 1). These processes are affected by both total precipitation [Wu et al., 2011] and variation in the timing of precipitation, including intraannual [Grant et al., 2014; Peñuelas et al., 2004; Reichstein et al., 2013; Reyer et al., 2012] and interannual [Faticchi and Ivanov, 2014; Hsu et al., 2012; Knapp and Smith, 2001; Peñuelas et al., 2004] variabilities. In light of the projected changes in global precipitation patterns [Intergovernmental Panel on Climate Change (IPCC), 2012, 2013; Sillmann et al., 2013], there is a critical need to understand how changes in precipitation can impact terrestrial ecosystems.

One way to examine these processes in the field is through the use of precipitation manipulation experiments. The design and implementation [Beier et al., 2012; Hanson, 2000; Miranda et al., 2011] as well as results [Reyer et al., 2012; Wu et al., 2011] from these experiments have been recently reviewed. However, there is a need for a continued effort to integrate the results of field experiments into the land surface models.
(LSMs) that simulate the response of the terrestrial biosphere to climate forcing in the context of Earth system models (ESMs) [e.g., Collins et al., 2011] used to project rates and impacts of future climate change.

Currently, LSM representation of many biological responses to precipitation is mechanistically simple and/or poorly constrained due to lack of appropriate parameterizations for many ecosystems [Powell et al., 2013; Todd-Brown et al., 2013]. This reflects both the limited empirical data necessary to improve models and a lack of comparisons between model simulations and observational data. In this review, we outline ways that precipitation manipulation experiments can be designed to help inform models and, ultimately, improve model realism for global change studies. Through collaboration with the modeling community, the applicability of plot-scale results could be greatly enhanced. Although this idea has been discussed before [Beier et al., 2012; Classen and Langley, 2005; Dietze et al., 2013], the transformation of experimental results into model improvements is still fragmentary. This review is designed to provide a framework to bridge the gap between these two communities and to reignite the efforts necessary for interdisciplinary collaborations.

Here we first review the most recent projections for future terrestrial precipitation (i.e., Coupled Model Intercomparison Project phase 5; section 2) and follow this with a brief review of the structure and functioning of current-generation LSMs (section 3). In section 4, we explore the different approaches taken to evaluate the response of terrestrial ecosystems to precipitation in the field. We not only focus on the efficacy of different experimental designs for informing LSMs but also allude to natural experiments that are useful for informing LSMs. In section 5, we highlight routine measurements, including hydrological and meteorological measurements that, if taken, would aid the incorporation of more sophisticated data into models. We then investigate the representation of some precipitation responses in LSMs and examine how similar responses are measured in the field using four commonly measured biological processes as examples: (1) carbon assimilation and productivity, (2) phenology, (3) soil organic matter (SOM) decomposition, and (4) plant community dynamics (section 6). We provide recommendations for how measurements of these processes could be made in a way that is useful for model development. Finally, as this review considers a connection between plot-scale data and global-scale models, we describe ways to improve model-data connectivity (i.e., the ability of experimental data to evaluate and improve models) through the use of targeted observational data (section 7). We conclude that, by following a few simple recommendations, the applicability of field experiments and realism of LSMs could be greatly enhanced (section 8).

2. Future Terrestrial Precipitation Change

Increasing global surface temperatures are expected to enhance rates of evaporation and precipitation, over both land and the ocean [Allan et al., 2013; Held and Soden, 2006; Schneider et al., 2010]. Contemporary Coupled Model Intercomparison Project (CMIP) analyses suggest that the global mean wet-day precipitation (total precipitation on days where precipitation is >1 mm) will increase 3.5–9%, with projected totals increasing under more intense radiative forcing scenarios. However, regional changes in mean precipitation are projected to differ, with amounts increasing in some locations (e.g., high northern latitudes and in Eastern Africa, South and Southeast Asia, and Antarctica) and decreasing in others (e.g., Central America, South Africa, and the Mediterranean) [Sillmann et al., 2013]. These regional effects are expected not only as a result of climate change but also as a result of rapid land use and land cover change underway globally [Pielke et al., 2011].
In concert with altered mean precipitation, models also project an increase in the frequency and intensity of extreme precipitation events [O’Gorman, 2012; O’Gorman and Schneider, 2009; Sillmann et al., 2013; Tebaldi et al., 2006] and an increase in precipitation variability with more frequent droughts and floods [Easterling et al., 2000; IPCC, 2012, 2013]. Observations suggest that these changes are already occurring [Min et al., 2011], with North America showing a strong increase in extreme precipitation events [Alexander et al., 2006]. An increase in these events is projected to occur globally, even in areas of the world where total precipitation is projected to remain unchanged or decrease [Sillmann et al., 2013]. However, while projections of extreme precipitation events over large areas (i.e., the continental scale) are robust, more local projections may be masked by internal climate variability [Fischer et al., 2013].

Precipitation is an important forcing that governs hydrology, which in turn affects not only energy and water cycles but also biological processes. As such, models need to simulate the response of these processes to a range of precipitation magnitudes and intensities. A large body of research has been devoted to understanding and properly simulating these responses [Cao and Woodward, 1998; Katul et al., 2007; Knapp et al., 2008; Weltzin et al., 2003]. However, a greater understanding of the mechanisms driving biological responses to precipitation changes and consequent feedbacks is needed to ensure better and more reliable future projections.

### 3. Current-Generation LSMS

In the simplest form, a LSM provides boundary conditions for computing momentum, energy, and mass (water, carbon) fluxes at the interface between the land surface and the atmospheric boundary layer. We use the term LSM generically, including LSM sensu stricto and LSMS embedded in ecohydrological, biogeochemical, and dynamic vegetation models, but we mostly focus on large-scale applications in the context of ESMs and long-term studies. LSMS have evolved greatly [see Sellers et al., 1997; Pitman, 2003], beginning as simple single soil layer, implicit vegetation models in the late 1960s [Manabe, 1969], and expanding to include multiple soil layers and explicit vegetation in the late 1970s and 1980s [Deardorff, 1978; Dickinson et al., 1993, 1986; Entekhabi and Eagleson, 1989; Noilhan and Planton, 1989; Sellers et al., 1986; Verseghy, 1991], carbon assimilation in the early 1990s [Bonan, 1995; Cox et al., 1998; Sellers et al., 1992, 1996], and finally, most currently, different plant types [Bonan et al., 2002] and dynamic changes in carbon pools and vegetation properties [Bonan et al., 2003; Clark et al., 2011; Dickinson et al., 1998; Krinner et al., 2005] as well as dynamic nitrogen pools [Dickinson et al., 2002; Thornton et al., 2009; Zaehle and Friend, 2010].

LSMs have been developed and applied at different scales, and while there is no clear guidance for the scale at which a particular LSM should be applied, the tendency is to go toward finer spatial scales and to validate LSMS at the scale of flux tower footprints (e.g., 1000–10,000 m$^2$) [e.g., Blyth et al., 2010]. However, components that constitute the land surface schemes of climate models and ESMs [e.g., Best et al., 2011; Clark et al., 2011; Krinner et al., 2005; Lawrence et al., 2011; Medvigy et al., 2009; Niu et al., 2011; Noilhan and Mahfouf, 1996; Oleson et al., 2010; Viterbo and Beljaars, 1995] typically operate at larger spatial scales from tens to thousands of kilometers and temporal scales from minutes to days and beyond. These models are continually adding potentially relevant biological responses to precipitation and hydrological changes, often as a function of soil moisture (section 5). Nonetheless, the functions used are often empirical, relying on generalized responses and omit the driving biological processes. Although these functions may simulate historical data well [Kleidon and Heimann, 1998; Porporato et al., 2002], the omission of driving mechanisms decreases the reliability of future projections. Unfortunately, the reliance on empirical models is a necessity due to limited understanding of responses of biological processes in the field [Arthern et al., 2010].

In the sections below, we consider how precipitation manipulation experiments, which are fundamentally designed to characterize these processes, can help decrease the uncertainty associated with precipitation responses in LSMS. As stated above, we focus primarily on LSMS that constitute, or are designed for, the land surface schemes of climate models and ESMs [e.g., Best et al., 2011; Clark et al., 2011; Krinner et al., 2005; Lawrence et al., 2011; Oleson et al., 2010; Raddatz et al., 2007; Shevliakova et al., 2009; Sitch et al., 2003; Zaehle and Friend, 2010]. However, the discussion is also relevant for mechanistic ecohydrological models [e.g., Fatichi et al., 2012b; Ivanov et al., 2008b], which are typically applied at smaller spatial scales and could provide a bridge for improving large-scale models.
4. Field Studies of Biological Responses to Precipitation Changes

Field studies have been underway that manipulate precipitation over small areas (i.e., plots) in order to examine terrestrial responses to future precipitation change (Figure 2). Plot sizes vary greatly between experiments, ranging from tens [e.g., Beier et al., 2004] to hundreds [e.g., Lamersdorf et al., 1998] to thousands [e.g., Hanson et al., 2003; Pangle et al., 2012] of square meters but are typically smaller than the area simulated by common LSM applications (section 3). Plots are typically equipped with instrumentation to monitor soil moisture and meteorological variables (section 5), although the density of instrumentation varies by experiment with some replicating sensors only at the treatment level and others including measurements for each replicate plot.

These studies employ different methodologies, including using either natural studies or experimental manipulations and active or passive treatments [Beier et al., 2012; Hanson, 2000]. For these studies, treatment is defined as the type of manipulation imposed over replicate plots. Experiments also differ in the number of treatment levels used, a design consideration that can critically influence an experiment’s usefulness in informing models [Cottingham et al., 2005]. Below, we highlight the strengths and weaknesses of different methodologies for improving and informing how LSMs model biological responses to precipitation change.

4.1. Natural Studies Versus Experimental Manipulations

Research designed to study how land surfaces respond to precipitation will usually follow one of two strategies: (1) utilize existing natural precipitation events or gradients within a single ecosystem or (2) manipulate precipitation through the use of experimental treatments. Natural experiments allow for the evaluation of large-scale responses (e.g., full system gas, energy, and water fluxes). For example, flux tower data at interannual and intraannual scales have been utilized for describing the response of whole-system fluxes to drying and rewetting cycles in Mediterranean regions, finding that these systems are sensitive to spring rain [Aires et al., 2008; Allard et al., 2008; Ma et al., 2007]. Other examples include studying the response of systems to anomalous conditions such as extreme droughts [Baldocchi, 1997; Leuzinger et al., 2005; Reichstein et al., 2007] and rain pulses [Huxman et al., 2004a, 2004b; Jarvis et al., 2007; Jenerette et al., 2008; Ma et al., 2012]. Also, eddy flux networks have allowed for cross-system comparisons of precipitation responses and changes in water use efficiency [Keenan et al., 2013; Rambal et al., 2003; Ross et al., 2012].

Figure 2. Examples of different experimental designs employed in precipitation manipulation experiments. (a) Rainfall exclusion at Konza prairie in Kansas, USA. Note the use of pipes for funneling runoff away from plots. (b) Setup of rainfall exclusion structures at Kruger Park in South Africa. (c) PRICLE rainfall variability experiment in Indiana, USA. Note the “control” structures with fine mesh netting instead of rain excluding slats to control for shading caused by the slats in the treatment plots. (d) Rainfall addition application at the EVENT experiment in Germany with rainfall exclusion shelter in the background.
types of studies are useful for examining how processes defined and modeled at smaller scales translate to larger scales [Baldocchi, 1997; Rambal et al., 2003; Vargas et al., 2013], as is described in section 7. Natural experiments are also better for evaluating responses over long time scales, as most precipitation manipulation experiments last 1–3 years and those lasting greater than 10 years are rare [Beier et al., 2012]. Although natural studies are well suited for analyzing precipitation responses at large scales, in terms of improving LSMs, they suffer from the fact that the responses observed could be the result of numerous smaller-scale processes, which cannot be easily disentangled. For example, at a single site and/or between sites, seasonal changes in abiotic factors such as temperature or photoperiod or biotic factors such as vegetation species distribution or soil type could influence perceived precipitation responses. Therefore, natural studies are best suited for formulating predictions that can be subsequently tested using more controlled experiments and/or examining the results of small-scale experiments at larger scales (section 7). Here we will primarily focus on smaller-scale manipulation studies but include a discussion on how these and larger-scale data can be used in conjunction to evaluate model performance.

As opposed to natural studies, precipitation manipulation experiments evaluate the response of the land surface to changes in precipitation through direct alteration of the amount and/or timing of precipitation (Figure 2). Most often these experiments manipulate rainfall (i.e., warm-season precipitation) rather than full-season precipitation, but for the purpose of this review, we will use the term precipitation. These manipulative experiments will often include control plots to use for comparison to the manipulated, or experimental, plots. These control plots may experience ambient conditions defined by the precipitation pattern at the site or prescribe conditions often representing mean precipitation patterns at the site. Depending on the plot size, a buffer zone and/or trenching will be used to minimize edge effects, separate experimental, plots. These control plots may experience ambient conditions defined by the precipitation pattern at the site or prescribe conditions often representing mean precipitation patterns at the site.

4.2. Active Versus Passive Manipulations

Precipitation manipulation experiments commonly employ either active or passive treatments or both. In active manipulation studies, precipitation is artificially added to plots using sprinklers, hoses, or watering cans to supplement or replace natural precipitation or to alter precipitation chemistry. Alternatively, passive manipulations augment or remove ambient precipitation, typically through rainfall interception using throughfall or overstory shelters [Hanson, 2000]. In areas where vegetation height is low, such as grasslands, overstory shelters are typically employed [e.g., Hoepner and Dukes, 2012; Koerner and Collins, 2014; Yahdjian and Sala, 2002], whereas throughfall shelters are typically used in systems with tall vegetation, such as forests [e.g., Borken et al., 2006; Hanson et al., 1998; Pangle et al., 2012] (but see Misson et al. [2010]). Due to logistical issues, these experiments are often done at small (tens of meters) scales [Beier et al., 2012]. However, experiments at larger scales have been performed [e.g., Hanson et al., 1998; Misson et al., 2010; Nepstad et al., 2002; Pangle et al., 2012].

The benefit of active manipulations is that environmental conditions other than precipitation are minimally altered as a result of the manipulations. With passive manipulations, researchers must be cautious of unintended changes in radiation, temperature, or vapor pressure deficit that might result from treatment structures. In cases where these effects may be confounding, control plots are typically adjusted to create similar conditions to the experimental plots (e.g., by using netting to block out radiation similar to that removed by an overstory shelter). Also, the conditions created by passive manipulations are reliant on ambient conditions. Therefore, interannual differences in precipitation responses may be larger than responses to the manipulation due to year-to-year variation in precipitation.

4.3. Number of Treatment Levels

A major dilemma in the design of precipitation manipulation experiments involves determining the number of treatment levels and number of replicates of each treatment. As precipitation is an environmental driver that occurs at a range of values, multiple experimental treatment levels are desirable [Cottingham et al., 2005]. However, due to logistical constraints, an increase in treatment levels often comes at a cost to replication. Experimental designs that employ only a single treatment level typically have higher replication, often of a precipitation regime expected for a given region [e.g., Jentsch et al., 2007]. These analysis of variance-type designs may increase confidence in evaluating the response to a particular scenario but do not allow for responses to be evaluated across a range of precipitation and/or soil moisture values.
In most LSMs, the biological responses to changes in precipitation are manifested as a function of rainfall interception and soil moisture/soil water potential levels. Thus, experimental data that can be regressed across multiple treatment levels can be integrated easier into LSMs [Cottingham et al., 2005]. However, because of the trade-off between number of treatments and number of replicates in precipitation manipulation experiments, the uncertainty of a model formulation informed from data generated using multiple treatment levels likely increases as the number, range, or replication of levels decreases. Another source of uncertainty arises when the treatment levels fail to encompass soil moisture levels (and variability) expected under future scenarios. This may be particularly true for heavy rainfall or drought events. In fact, experiments that push the system to or beyond the most extreme multimodel projections for a given area can provide the unique ability to identify thresholds in precipitation that severely limit ecosystem functioning [Smith, 2011]. Therefore, ideally, experiments should be designed to include multiple treatments leading to a range of soil moisture values that capture and extend beyond the range of past observations and mean projections.

Typical output from these experiments includes soil moisture, aboveground net primary productivity, leaf onset and offset dates, soil respiration, and species composition, among others (Table 1). As we discuss in the following sections, with the addition and adjustment of a few measurements or products, the applicability of these measurements to models, and thus progresses in the broader global change community, could be greatly improved.

5. Data Needs for Integration of Experiments With Models

To integrate experimental responses into LSMs, high-resolution monitoring of environmental variables is needed (Table 1). One obvious response is the change in water flow through the system as a result of

---

**Table 1.** Data Needs to Incorporate Commonly Measured Field Responses Into Current-Generation Models

<table>
<thead>
<tr>
<th>Process</th>
<th>Common Responses Measured and Reported by the Experimental Community</th>
<th>Data Needed by the Modeling Community&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>All processes</td>
<td>- Precipitation amount and timing&lt;br&gt;- Soil moisture&lt;br&gt;- Air temperature</td>
<td>- Soil moisture, preferably at high temporal and spatial resolution (actual and relative to saturation)&lt;br&gt;- Soil water retention curves&lt;br&gt;- Precipitation amount and timing&lt;br&gt;- Energy, water, and carbon fluxes&lt;br&gt;- Micrometeorological data (e.g., air temperature, humidity, radiation, and wind speed), preferably at high temporal and spatial resolution&lt;br&gt;- Soil temperature&lt;br&gt;- Detailed site characteristics (e.g., plant functional and soil types and fractions)</td>
</tr>
<tr>
<td>Carbon assimilation and productivity</td>
<td>- Aboveground net primary productivity&lt;br&gt;- Net photosynthesis</td>
<td>- Photosynthesis and respiration, preferably at high temporal resolution and in multiple canopy layers&lt;br&gt;- Stomatal conductance&lt;br&gt;- Carbon allocation to different plant tissues including changes in carbon stocks over time</td>
</tr>
<tr>
<td>Phenology</td>
<td>- Date of bud burst/emergence&lt;br&gt;- Leaf drop/senescence</td>
<td>- Estimates of moisture availability and demand at similar time points before, during, and after emergence/senescence&lt;br&gt;- Comparisons across multiple seasons to explore thresholds&lt;br&gt;- LAI, preferably at high temporal and spatial resolution</td>
</tr>
<tr>
<td>Soil organic matter decomposition</td>
<td>- Soil respiration</td>
<td>- Soil respiration, preferably at high temporal and spatial resolution&lt;br&gt;- Microbial abundances at multiple depths&lt;br&gt;- Litter decomposition rates</td>
</tr>
<tr>
<td>Plant community dynamics</td>
<td>- Species richness&lt;br&gt;- Species diversity</td>
<td>- Species abundances grouped by functional type&lt;br&gt;- Functional type shifts through time&lt;br&gt;- Rates of reproduction and mortality</td>
</tr>
</tbody>
</table>

<sup>a</sup>Measurements are grouped by order of importance.
precipitation change. In general, precipitation within LSMs can take three possible routes: (1) fall on the vegetation where it is intercepted and subsequently evaporated or fall on the ground as drip, (2) fall on the soil surface where it can infiltrate the soil and then evaporate, be taken up by plant roots, or percolate down to add to deeper layers including groundwater supplies, or (3) run off from the surface. SL = soil layer; GW = ground water.

![Figure 3. Scheme of the land water balance within a hypothetical LSM with three soil layers. In this example, and similar to the description in the text, precipitation can either (1) fall on the vegetation where it is intercepted and subsequently evaporated or fall on the ground as drip, (2) fall on the soil surface where it can infiltrate the soil and then evaporate, be taken up by plant roots, or percolate down to add to deeper layers including groundwater supplies, or (3) run off from the surface. SL = soil layer; GW = ground water.](image)
Soil moisture can be observed using direct gravimetric sampling or neutron probe measurements but is most commonly monitored using time domain reflectometry (TDR; see Rundel and Jarrell [2000], Hanson [2000], and Robinson et al. [2008] for more detail information about each technique). The benefit of TDR is that it can be implemented in a nondestructive fashion. However, TDR provides an estimate of soil water content, and further data, such as a soil water retention curves and/or soil texture, are needed to translate the data into soil water or matric potential. As such, researchers should be careful about comparing moisture responses between studies reporting only soil water content data, as similar values may imply differing amounts of stress in multiple systems that vary in soil texture [Vicca et al., 2012].

There is still uncertainty in how precipitation influences soil moisture both directly and indirectly (e.g., through vegetation-driven changes in soil infiltration, shading, or rain funnelling [D’Odorico et al., 2007]), and models are not always able to simulate these responses well. High spatial and temporal resolution monitoring within precipitation manipulation experiments could help reduce model uncertainty. In the field, soil moisture measurements often have better temporal, rather than spatial, resolution, particularly with depth [Robinson et al., 2008; Vereecken et al., 2008]. As more models are beginning to include multiple soil layers [e.g., Ameno and Kumar, 2008; Drewry et al., 2010; Oleson et al., 2013; Parton et al., 1993] and topographic heterogeneity [Fatichi et al., 2012a; Ivanov et al., 2008a], soil moisture measurements at a variety of depths and locations are important, as soil moisture can vary within the vertical profile and among sites with similar soil characteristics [e.g., He et al., 2013].

More general, in models soil moisture is a function of climate, but this is mediated by the characteristics of the site, including plant functional types (PFTs), soil type and texture, and rooting density and depth. Therefore, monitoring and reporting of general site characteristics along with moisture values are necessary for proper model parameterization [Ivanov et al., 2012; Liang et al., 2005] and/or data assimilation products [Heathman et al., 2003; Rodell et al., 2004]. In addition, site characteristic monitoring may allow for the inclusion of dynamic responses (i.e., characteristics that change over time) of these variables (e.g., rooting profiles) into models.

Note that all aspects of the precipitation-soil moisture interaction cannot be evaluated within experiments, cloud-soil moisture feedbacks in particular. Previous studies have found a strong coupling between precipitation and soil moisture that suggests that soil moisture may feedback to observed precipitation patterns [D’Odorico and Porporato, 2004; D’Odorico et al., 2007; Koster et al., 2009, 2004; Teuling et al., 2006]. While these dynamics can be accounted for in coupled land surface climate models, they cannot be addressed by small-scale experiments and remain a pure modeling domain.

Because the biological processes within LSMs respond to soil moisture-soil water potential rather than precipitation, some have called for precipitation manipulation experiments to begin using common metrics for evaluating soil moisture stress [Vicca et al., 2012]. This is important for comparing the results of multiple studies and may also be useful for making generalizations for models. Typically, this requires site characteristic data that can be used to calculate a common metric from measured data (e.g., extractable water or stress intensity; see equations in Vicca et al. [2012]). These data can then be used to generate generalized functions that combine both plant physiological and soil processes to describe the response of plant water uptake to soil moisture [e.g., Bartholomeus et al., 2008; Caylor et al., 2009; Ivanov et al., 2008b; Porporato et al., 2001]. These functions, essentially plant responses at different levels of soil water potentials [e.g., Feddes et al., 1976, 2001; Maherali et al., 2004, 2006], allow models to accommodate combinations of plant and soil types with differing responses.

Finally, micrometeorological data such as soil and air temperatures at various depths and heights, respectively, as well as solar radiation, wind speed, and humidity at different points within the canopy are useful for improving and testing model functioning [Katul et al., 2012; LeMone et al., 2007; Seneviratne et al., 2010]. These data can be used to facilitate the incorporation and test the usefulness of more refined processes into models (i.e., processes that respond only in part to changes in moisture), such as those detailed in the next section. See Table 1 for a description of general data necessary to incorporate processes into a model as well as specific data necessary to incorporate the four processes discussed below.

6. Experimental Data Best Suited for Integration With Models

In the following subsections we compare four biological processes commonly measured in the field with their model analogs: (i) carbon assimilation and productivity, (ii) phenology, (iii) soil organic matter
Figure 4. Conceptual example showing values of a typical soil moisture scaling factor for a land surface process (e.g., photosynthesis and SOM decomposition) under different soil moisture conditions ($\theta$). The model depicted here is for photosynthesis, but a similar theory applies for other processes. The function plotted defines the scaling factor to be 1 above a critical value when moisture does not influence the process ($\theta_c$: 0.27 m$^3$/m$^3$ here), 0 below the permanent wilting point ($\theta_{wp}$: 0.136 m$^3$/m$^3$ here), and $((\theta - \theta_{wp})/((\theta - \theta_c))$ elsewhere, where $q$ is a measure of nonlinearity that describes the shape of the function [Egea et al., 2011a; Keenan et al., 2010; Porporato et al., 2001]. Data are plotted for $q$ values of 0.5, 1, 1.5, 2, and 3.

decomposition, and (iv) plant community dynamics. We chose these processes because they range in scale from leaf to ecosystem level and are affected by precipitation over differing time scales ranging from seconds to decades (see Figure 1).

This range allows us to discuss differences in measured and modeled processes over differing spatial and temporal scales. The purpose of these comparisons is to highlight examples of processes that are measured differently than they are modeled and to explore how measurements or formulations could be designed differently to help make experimental results more suitable for model development. Many of the recommendations provided could be applied to other processes beyond those mentioned below.

6.1. Carbon Assimilation and Productivity

Within LSMS, net primary productivity is the difference between simulated carbon assimilation and autotrophic respiration. Carbon assimilation is simulated through leaf gas exchange processes that respond to changes in environmental conditions, including soil moisture. However, the structure of these soil moisture responses varies greatly among models [De Kauwe et al., 2013; Egea et al., 2011a]. Some models (e.g., Orchidee-CN (O-CN) [Zaeihle et al., 2010]) use a modifier to alter the relationship between photosynthesis and conductance within coupled photosynthesis-stomatal conductance schemes, which are based on empirical relationships between stomatal conductance and assimilation [e.g., ball et al., 1987; Jacobs et al., 1996; Leuning, 1995]. These modifiers are based on soil water content [e.g., Fatichi et al., 2012b; Wang and Leuning, 1998], soil water potential [e.g., Laurent, 2004], or leaf water potential [e.g., Vico and Porporato, 2008]. See Figure 4 for an example of such a function. Alternatively, other models (e.g., Community Land Model (CLM) [Oleson et al., 2010, 2013]) include a similar type of modifier that alters the biochemical capacity of the photosynthetic system based upon the soil water available to plant roots. For example, CLM uses a scaling factor, $\beta_t$, which scales down the maximum rate of carboxylation of photosynthesis ($V_{\text{cmax}}$). The $\beta_t$ value ranges from 0 to 1 depending on the amount of roots and soil moisture in each soil layer (i.e., $V_{\text{cmax}}$ will be decreased less if a greater amount of water is available to a larger fraction of roots). This modified $V_{\text{cmax}}$ value is then used to calculate photosynthetic rates [Collatz et al., 1991; Farquhar et al., 1980]. The carbon that is taken up through photosynthesis is then allocated to different processes, including growth [Oleson et al., 2013].

As opposed to models that simulate photosynthetic responses to moisture and infer productivity responses, field experiments often directly measure productivity and infer that photosynthesis is a main factor driving the response [e.g., Fay et al., 2003]. This is because the positive relationship between precipitation, moisture, and aboveground production is well documented and the explanation for this pattern is well understood: when soil water is abundant, a plant can leave its stomates open, allowing for greater CO$_2$ diffusion into leaves [Chaves et al., 2009, 2003, 2002; Niyogi and Xue, 2006; Pinheiro and Chaves, 2011; Potter et al., 1993; Shaw et al., 2002]. However, other evidence suggests that the link between assimilation and growth is more complex because of the lag between assimilation and allocation of carbon [Sala et al., 2012]. This link may indeed be mediated by plant carbon storage and hydraulic controls acting directly on meristematic activity (tissue growth) rather than photosynthesis [Fatichi et al., 2014; Körner, 2013]. In addition, in dryic sites, water is not limiting and additional water can decrease soil oxygen concentrations and nutrient cycling, resulting in nutrient leaching and increased aboveground productivity [Schuur, 2003]. Unfortunately, many studies do not fully link the flow of carbon (from photosynthesis to growth), limiting their ability to develop model
parameterizations because models rely on a structure that links changes in growth in response to moisture availability through carbon uptake and allocation. As such, there remains a need to better elucidate these responses in the field (Table 1).

An additional area of uncertainty in the field and in models is the interactive effect of elevated CO$_2$ and water stress on carbon assimilation and productivity. Water use efficiency (i.e., the amount of carbon gained through assimilation per water lost through transpiration) has been shown to increase with experimental and historical increases in CO$_2$ [Battipaglia et al., 2013; Keenan et al., 2013; Morgan et al., 2011], but the ability of these potential water savings to mitigate drought stress, increase runoff, and stimulate productivity remains debated [Donohue et al., 2013; Fatichi and Leuzinger, 2013; Huntington, 2008; Warren et al., 2011; Zaehle et al., 2014]. Direct experimental tests of this response on plant productivity are rare and suggest little interactive effect [Dukes et al., 2005]. Observational studies suggest that enhanced water use efficiency does not always translate into productivity gains [Peñuelas et al., 2011] or mitigate the impacts of soil moisture stress on plant growth [Brezosek et al., 2014; Grünzweig and Körner, 2001; Morgan et al., 2004]. As models vary greatly in how they simulate photosynthesis and conductance responses to CO$_2$ and soil moisture [De Kauwe et al., 2013], there is the potential for an important synergy between field researchers and modelers to address this knowledge gap.

Currently, model simulations of terrestrial productivity responses to changes in precipitation are challenging, likely due to improper parameterization of photosynthetic responses and/or the connection between assimilation and productivity [Powell et al., 2013]. In fact, model-data comparisons have found that models that only include $V_{cmax}$ responses to soil moisture are not able to reproduce observations as well as models that include more physiologically relevant mechanisms [Egea et al., 2011a] (but see Keenan et al. [2010]). As such, there has been a recent push to include more physiologically relevant responses into models (e.g., mesophyll conductance [Egea et al., 2011a]). Although these studies are a good start, systematic model-data comparisons examining photosynthetic and/or growth responses to changes in soil water content are lacking.

Although poor parameterization of stomatal conductance, assimilation, and growth responses to soil moisture is a major limitation of current models, this limitation provides an opportunity for future field experiments to help understand and quantify these responses. To help improve model performance, field researchers should measure the photosynthetic and conductance responses [e.g., Egea et al., 2011b; Grassi and Magnani, 2005; Rodgers et al., 2012] and link them to the allocation processes that lead to changes in growth [Franklin et al., 2012]. Figure 5 illustrates how in the absence of this link, field-derived productivity data, while useful for evaluating the overall model performance, may not aid in improving the parameterization of underlying processes. It also is worth noting that model simulation of these responses is difficult because of the need for current models to organize species into broad categories, or plant functional types (PFTs), and to be able to account for microclimate (e.g., temperature, radiation, humidity, and wind speed) variability within the plant canopy.

### 6.2. Phenology

Shifts in phenology can have important feedbacks on ecosystem processes, particularly carbon uptake, biotic interactions, and energy-water linkages [Dragoni et al., 2011; Richardson et al., 2009, 2012]. Phenological responses to soil moisture in LSMS are often only simulated in “raingreen” deciduous species (i.e., species that shed their leaves in response to soil moisture stress). This is implemented by simulating leaf senescence in these trees as a function of the ratio of soil moisture and canopy conductance [e.g., Sitch et al., 2003]. Phenology in other deciduous plant functional types (i.e., “sumergreen” species) responds to temperature rather than soil moisture. In some models, phenology in deciduous species may be determined by either temperature, day length, or moisture depending on which factor reaches a predetermined threshold first [e.g., Fatichi et al., 2012b; Shevliakova et al., 2009]. Similar alternative models have also been proposed to determine phenology in deciduous species based on a cost-benefit structure, where leaves are only present when environmental conditions, including temperature, precipitation, and photoperiod (an important driving factor [Körner and Bosler, 2010]), result in a net carbon gain [e.g., Arora and Boer, 2005].

Unfortunately, manipulation experiments in areas where phenology is most likely driven by soil moisture (e.g., monsoon regions and tropical dry forests [Eamus, 1999; van Schaik et al., 1993]) are scarce (but see
As such, it remains unclear how to properly parameterize phenological responses in these biomes. Phenological studies examining precipitation responses in temperate systems typically find that temperature drives phenology [Bloor et al., 2010; Cleland et al., 2006]. However, flux tower studies in Mediterranean grassland [Xu and Baldocchi, 2004; Xu et al., 2004] and precipitation manipulation experiments in temperate grassland [Jentsch et al., 2009], Mediterranean shrubland [Llorens and Peruelas, 2005], and Mediterranean forest [Misson et al., 2011] have shown that moisture can have a strong effect on phenological responses in extratropical systems, typically delaying spring phenology under drier conditions. For example, Misson et al. [2011] found that the number of Mediterranean trees producing functionally mature leaves was decreased by 50% under heavy (87% removal) spring drought.

Better model representation of phenological responses to precipitation change is likely to have a large influence on model performance. Similar to productivity, field experiments should not only measure typical changes in phenology such as timing of bud burst and leaf drop but also carefully measure how these processes vary with changes in moisture availability and demand across seasons as well as within seasons (e.g., time evolution of leaf area index (LAI) at high temporal resolution), carefully noting threshold moisture values where phenological shifts occur. Studies should consider overstory as well as understory phenology, as understory plants have been shown to constitute a large portion of carbon flux in some ecosystems [Baldocchi et al., 1997], an effect that, due to the phenology of these species, has been shown to impair model performance [Kimball et al., 1997]. These types of measurements will allow for a much better incorporation of responses and processes into LSMs.

**6.3. Soil Organic Matter Decomposition**

It is well known that soil organic matter (SOM) decomposition is sensitive to changes in water availability [Ise and Moorcroft, 2006; Manzoni and Porporato, 2009; Moyano et al., 2013; Schimel et al., 1994; Suseela et al.,
Many LSMs simulate SOM decomposition in a similar fashion [Todd-Brown et al., 2013], using a framework similar to the CENTURY model framework [Parton et al., 1987, 1993]. In the CENTURY framework, decomposition is modeled as a first-order kinetic reaction with rate constants that differ as a function of the reactivity of carbon in various soil and litter pools (i.e., passive, slow active [Parton et al., 1987, 1993]). To model climate responses for each pool, this rate constant is modified as a function of soil temperature following Arrhenius kinetics [Lloyd and Taylor, 1994]. Soil moisture influences this response by altering soil temperature. Soil moisture is also included as a direct influence along with the temperature function into these models in the form of an increasing [e.g., Andrén and Paustian, 1987] or peaked [e.g., Colema and Jenkinson, 1999] modifier, similar to the Vmax, scalar, βt, noted above.

In the field, SOM decomposition responses to precipitation and moisture are well studied. It is understood that when soil moisture is low, soil microbes become less active, exhibit stress responses, and have less access to C- and N-bearing substrates [Schimel et al., 2007]. When soils are water saturated, anoxic conditions can impede microbial and enzymatic activity [Freeman et al., 2001]. Thus, shifts in precipitation that lead to dry or wet conditions appear to lead to strong declines in decomposition rates [Liu et al., 2009; Suseela et al., 2012] and these responses are typically included into models [e.g., Porporato et al., 2003].

On the other hand, shifts in precipitation that move soil moisture conditions away from these wet and dry thresholds into more optimal conditions stimulate decomposition often leading to large pulsed releases of carbon and nutrients [Cleveland et al., 2010; Craine and Gelderman, 2011]. For example, in arid and semiarid ecosystems, rainfall events that follow long periods of dry conditions lead to pulsed releases of CO2 and nutrients [Carbone et al., 2011; Huxman et al., 2004a; Sponseller, 2007; Xu et al., 2004]. These pulses in soil respiration and N mineralization can comprise a substantial portion of the annual production of CO2 and plant-available N [Austin et al., 2004; Carbone et al., 2011]. Models could be tested to reproduce pulse respiration events and nonlinear threshold responses [Todd-Brown et al., 2013], but further evaluation of model performance is necessary, as these responses may become less apparent and less important when evaluating carbon cycling over longer time scales. This is only possible with longer manipulation experiments or larger-scale observations (section 7).

In addition, different types of soil organisms have different strategies for acquiring and utilizing soil carbon and nutrients, an effect that may have a large influence on soil carbon stocks [Averill et al., 2014; Orwin et al., 2011]. These organismal differences are basically unaccounted for in LSMs [Manzoni and Porporato, 2009]. Recent research has focused on explicitly representing soil microbial processes into ecosystem models [Allison et al., 2010; Orwin et al., 2011; Treseder et al., 2012]. At the global scale, the integration of a functioning soil microbial community into CLM substantially improved predictions of current soil C stocks [Wieder et al., 2013]. The majority of the parameters and outputs of this CLM microbial model can be measured in the field (i.e., microbial growth efficiency, microbial biomass, and enzyme activity) and should be more often reported as experimental results. Given that enzyme kinetics and microbial growth efficiency are modeled primarily as a function of soil temperature in most microbial models, there is the potential for field researchers to directly inform the next generation of models by investigating how these parameters vary as a function of experimental changes in soil moisture or, better, soil water potential (via changes in precipitation).

### 6.4. Plant Community Dynamics

A few models have been designed to incorporate individual- and community-level processes such as survival and competition into ESMs through the use of dynamic global vegetation models (DGVMs). However, DGVMs typically have a simplified land surface component, and integration of state-of-the-art LSMs with ecosystem demography dynamics for global-scale analyses is still limited [e.g., Arora et al., 2013]. DGVMs typically come in three classes: (1) “area-based” models that simulate the plant functional type (PFT) occupancy of a grid cell based on environmental and climatic variables in a deterministic manner (e.g., CLM [Bonan et al., 2002; Oleson et al., 2010, 2013], Lund-Potsdam-Jena (LPJ) [Sitch et al., 2003], and Top-down Representation of Interactive Foliage and Flora Including Dynamics (TRIFFID) [Cox, 2001]), (2) “individual-based” gap models that simulate competition between individual plants, primarily for light, in a stochastic manner (e.g., LP General Ecosystem Simulator [Smith et al., 2001] and the adaptive dynamic global vegetation model [Scheiter and Higgins, 2009]), and (3) hybrid models that simulate succession and light competition between PFTs in a deterministic, computationally efficient manner (e.g., Ecosystem Dynamics Model [Medvigy et al., 2009; Moocroft et al., 2001]).
Individual responses in individual-based and hybrid DGVMs are typically defined primarily by light competition, while changes in soil moisture affect vegetation at the community level. Community responses are primarily driven by feedbacks that differ between PFTs [e.g., Medvigy et al., 2009]. In essence, these responses are similar to the ones presented above but are parameterized differently for different PFTs or species. For example, models may employ lower moisture stress thresholds for drought-tolerant than drought-intolerant species, which have higher water use efficiency. This allows community dynamics to play out via differences in carbon uptake of different PFTs under different moisture conditions. Additionally, ESMs may simulate plant community changes to precipitation indirectly through changes in fire occurrence [Li et al., 2013]. However, as human disturbances/decisions are critical in driving fire projections, we focus on more direct precipitation responses here.

Although DGVMs are typically combined with LSMs to simulate plant community responses, experimental results suggest that precipitation responses of plant communities are more complex than those represented in models [e.g., Balvanera et al., 2006; Barger et al., 2011; Knapp, 1993; Knapp et al., 2012; Pérez-Ramos et al., 2010; Walter et al., 2012; Yachi and Loreau, 1999]. For example, increases in rainfall intensity may specifically alter the distribution of PFTs by increasing shrub encroachment [Barger et al., 2011; Kulmatiski and Beard, 2013], which could, in turn, influence the microclimate of the system [He et al., 2011]. Changes in rainfall amount may affect seed production [Pérez-Ramos et al., 2010], and changes in rainfall timing may also affect seed germination [Chou et al., 2008; Rivas-Arancibia et al., 2006], an important effect in annual species. Finally, evidence suggests that biodiversity increases functional resilience to environmental fluctuations [Balvanera et al., 2006; Yachi and Loreau, 1999], with other experimental evidence suggesting that more functionally diverse communities experience less tissue dieback under extreme drought [Walter et al., 2012].

These community responses can potentially be included into models using a PFT approach (e.g., by allowing for shifts in PFT composition under different soil moisture levels or stresses); however, experimental data are still insufficient for providing thresholds necessary for parameterization of these responses in different biomes and likely a very large within biome variability has to be expected. More data on reproduction and mortality of different species are anyhow warranted. More likely these responses could be incorporated using a trait-, rather than PFT-, based framework for modeling species heterogeneity. A trait-based framework involves employing known relationships between plant traits that arise a consequence of evolutionary trade-offs [Osnas et al., 2013; Reich, 2014; Wright et al., 2004] to describe species distributions in time and space [Van Bodegom et al., 2012]. This work is promising in that it overcomes many of the problems associated with using PFTs (e.g., few PFT levels and many individual parameterizations); however, this work is still in its infancy [Douma et al., 2012; Pavlick et al., 2013; Scheiter et al., 2013].

6.5. Scales of Responses

Long-term (>10 years) precipitation manipulation experiments are rare [Beier et al., 2012]. Short-term experiments may be able to adequately explore the response of stomatal functioning or fortuitously capture extremes. However, responses such as plant and/or microbial species change may not be able to be observed over the short time scales utilized in most experiments (Figure 1) [Smith et al., 2009]. This is also important for model development, as the processes that respond to longer-term changes are typically higher-level processes (e.g., phenology on the scale of a season and plant community dynamics on the scale of years to decades). These higher-level processes ultimately define lower level processes and, thus, can be critical for model functioning. For instance, it is rather different for a model to reproduce the plant response to a few days-long drought or to several years of below average precipitation.

To highlight the importance of this issue, we developed a simple model sensitivity experiment. We evaluated model sensitivity of the carbon assimilation, SOM decomposition, phenology, and community composition processes above using the land component of the NOAA/GFDL ESM (LM3) [Shevliakova et al., 2009]. We ran simulations at a grid cell corresponding to a temperate deciduous broadleaf forest flux tower site in Indiana, USA [Dragoni et al., 2011], for 7 years (1999–2005) following a 299 year spin-up. The model structures used included a base version of the model [Shevliakova et al., 2009], as well as versions that (1) increased conductance limitation under drought stress by a factor of 2 (carbon assimilation), (2) increased the moisture level at which leaves drop by a factor of 1.5 (phenology), (3) changed the SOM decomposition-soil moisture function from a peaked to constantly increasing function (SOM decomposition), and (4) altered the PFT
Our results confirmed the idea that models are most sensitive to higher-level processes, particularly shifts in vegetation cover, as changes to the PFT distribution had the largest effect on the modeled NEE, increasing NEE by 36%, compared to a <5% change for alterations to leaf gas exchange, phenology, or SOM decomposition functions (Figure 6). This may be surprising considering that leaf gas exchange, phenology, and SOM decomposition are more closely linked to the carbon flux of the system (i.e., NEE). However, within the model, changes in PFT structure outweigh relatively subtle changes in leaf gas exchange and SOM decomposition. PFT structure ultimately describes the magnitude of change that occurs as a result of changes in these lower level processes (e.g., the way in which CO₂ is taken up is determined by the growth and survival strategy and photosynthetic pathway of the plants within each PFT). While the case study shown here is only for one model formulation under a subjectively chosen set of scenarios, it highlights the critical need for studies to improve the formulations and/or parameterizations for higher-level functions. It is also supported by a recent study which showed that sensitivity of ET and vegetation productivity to changes in annual precipitation increases when an ecosystem undergoes reorganization (e.g., successional and invasion processes, shifts in composition [Fatichi and Ivanov, 2014]). In addition, these results highlight the importance of site-level conditions in modeling studies. As such, it is critical that experiments report site conditions such as species distribution and soil type and texture in order to help integrate results with models.

7. Connecting Plot-Scale Data With Models Using Intermediate-Scale Observations

Above, we have outlined examples of ways in which plot-scale precipitation manipulation experiments can help to improve LSMs. However, each example often assumes that plot-scale responses will be representative of larger spatial scales (e.g., ecosystem, regional, or global scale). Also, many processes are measured in the field at different temporal scales than their model analogs. For example, leaf gas exchange responses occurring over short time periods (e.g., seconds) in the field must be assumed to be similar when incorporated into or tested against a model utilizing a larger time step. Also, there may be small-scale (i.e., subgrid from a model perspective) variability that can influence the data (e.g., soil moisture differences due to topographic effects or heterogeneity in soils). These spatial and temporal mismatches are likely to invalidate direct comparisons between large-scale models and data from precipitation manipulation experiments. Therefore, ad hoc solutions are often used for making these comparisons [LeMone et al., 2008].

One possible solution is to use and develop independent, intermediate-scale data (i.e., data at a scale between the plot level and the level that the model is functioning) to test models and formulations designed using data from precipitation manipulation experiments, similar to model benchmarking techniques that have been previously proposed [Henderson-Sellers et al., 1993; Luo et al., 2012; Randerson et al., 2009] and carried out [e.g., Egea et al., 2011a; Keenan et al., 2010; Kleidon and Heimann, 1998; Powell et al., 2013; Richardson et al., 2012; Schaefer et al., 2012; Todd-Brown et al., 2013; Vargas et al., 2013]. One such effort, the
Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) [Henderson-Sellers et al., 1993, 1995], was designed to improve LSMs through comparison and evaluation of different models. This led to a broader understanding of the processes that result in differences between models. For instance, the PILPS project found that soil moisture estimation contributed greatly to differences between models [Henderson-Sellers et al., 1995]. Model intercomparisons help to determine the formulation differences between models that lead to uncertainty but do not necessarily pinpoint the particular process or parameterization that needs to be better represented, as indirect and multiple effects may influence these differences. In fact, problems with model functioning may balance out under typical conditions and not manifest until extreme conditions are seen [Niyogi et al., 1999], conditions that may not occur until long into the future.

As a solution, parameterization-level (i.e., equation), rather than model-level, studies could be performed to test the influence of the addition and/or change of a particular parameterization within a single LSM, comparing versions of a single model with differing structures in an ensemble mode. For example, one could run two simulations with inputs and forcings from an observational site (e.g., flux tower site). One simulation would use the original formulation or parameterization of a process, while the other would use a formulation or parameterization from a precipitation manipulation experiment. Following the simulations, comparisons to observational data could then be made (e.g., using Taylor scores [Taylor, 2001]) to see if the new model outperformed the old model. This would provide an appropriate test for the scalability of the new formulations or parameterizations.

Comparisons of model output could be made with any number of available observational data, including flux tower, forest inventory, remote sensing, or aircraft data. In fact, flux tower data sets have been designed, in part, for such studies [Reichstein et al., 2005]. Each data source would come with its own set of positive and negative aspects relating to data availability and global coverage, size of the data footprint, and degree to which the data are truly observational, which reflects the amount of post processing needed and assumptions met to obtain usable data. Therefore, careful consideration should be made when choosing the type of data to compare with and, in most cases, it is best to make comparisons to multiple data sets.

In addition, the initial conditions and equilibrium spun-up conditions likely play a large influence on model-data comparisons at the scale of observations. Therefore, initial conditions, including plant species and SOM quality and amount, among others, should be set to best match the observational site and ensemble simulations should be done to explore the uncertainty related to initial conditions that are unrelated to the model formulation. Following single-model analysis, similar analyses using other models could be performed to test the generality of the response across models. This could be done using a framework similar to that used in PILPS [Henderson-Sellers et al., 1993, 1995].

Data from precipitation manipulation experiments could also be used for model-data comparisons, assuming that data used for comparison are independent of data used for model parameterization. Comparisons with data from manipulation studies have been performed for CO$_2$ enrichment studies [De Kauwe et al., 2013; Fatichi and Leuzinger, 2013; Hickler et al., 2008; Warren et al., 2011; Zaehle et al., 2014] but are rare for precipitation manipulation studies (but see Fisher et al. [2007] and Powell et al. [2013]). The recommendations provided here would help these comparisons to be made more broadly.

8. Conclusions

Current-generation LSMs are becoming more sophisticated, but, as we have shown above, opportunities exist to improve their ability to simulate biological responses to precipitation and soil moisture. Precipitation manipulation experiments provide an excellent structure by which modeled processes can be examined. These field experiments have proved valuable for understanding processes under different hydrological conditions. We propose that if these experiments are conducted with models in mind (Table 1), they can help to improve model realism. This is particularly true if adequate covariates that allow the response of biological processes to be modeled are examined (Figure 5). At least two types of field to model studies could be realized: one using field data to improve mechanistic understanding and reformulate or reparameterize a modeled process, the other testing models using field data as a reference.

Similarly, modeling studies are needed that keep experiments in mind, as they can be helpful for defining the processes and parameters that can be directly improved using field measurements (Figure 6). The objective
of this review is not to suggest the restructuring of all precipitation manipulation experiments but to highlight the potential value of these experiments to models and to reignite the dialog necessary to reconnect the experiments to models and thus scale findings from the plot to the globe.

To summarize, the following are six specific comments and recommendations for helping to improve the connectivity between precipitation manipulation experiments and LSMS:

1. Experimentalists should consider using an experimental design that could help improve models and expand the level of inference of their research by becoming familiar with a model structure and tailoring their experimental design appropriately. This includes using designs with multiple (i.e., $\geq 2$) treatment levels to allow for responses to be regressed across a broad range of precipitation and/or moisture values. The targeted model in this case does not matter, as other models will likely be flexible enough to adjust. If the goal of a particular project is to understand how an ecosystem may respond in the future, the connection with modeling studies is particularly necessary.

2. Modelers should encourage this investment by experimentalists and become involved in experimental design, indicating necessary and important measurements. However, modelers should be mindful of experimental limitations and come up with creative ways to incorporate experimental results and validate those results (e.g., using larger-scale observations).

3. Although precipitation manipulation experiments are becoming more common, the responses of different processes across a high spatial resolution are still unknown due to the poor experimental representation of many biomes. As such, experiments in unrepresented biomes, particularly those occurring at high and low latitudes as well as urban biomes, should be prioritized. Input from the modeling community regarding regions that show the highest uncertainty will be of value.

4. Following from 3, many of the responses discussed above occur over long time scales (Figure 1). Therefore, experiments that run over long time periods (>10 years) or capture extremes are particularly valuable.

5. Decreased model performance or increased model uncertainty resulting from the inclusion and/or reparameterization of processes evaluated in the field should not deter modeling studies from examining, and even including, more mechanistic formulations. Getting the right causality for a given mechanism is important, even at the cost of getting a worse model fit, initially. Representing observed processes increases model realism, which results in a decrease in uncertainty in a future, changing world.

6. Following from 5, there is still a need to improve the way in which models are evaluated. For example, contemporary assessments (such as those mentioned previously) may find that a version of a model that omits certain mechanisms performs better than a more realistic version. The conclusion may be made that the more realistic version is worse, when instead more studies are needed to improve its formulation and/or parameterization. In addition, parameterization evaluations are needed to compliment full model evaluations that have been performed or are currently underway.

Precipitation manipulation experiments are an invaluable tool for helping to improve models, and through increased discussion between the different communities, heeding to the recommendations above, greater progress could be made in understanding how terrestrial ecosystems will respond to future change, under hydrological extremes.

Acknowledgments
We would like to thank Robert Dickinson and one anonymous reviewer for their comments on an earlier version of this review, which strengthened its breadth and focus. This review grew out of an organized oral session on precipitation manipulation experiments at the 96th Annual Meeting of the Ecological Society of America. We would like to thank the organizers and participants of that session for their discussion of topics relevant to this review. We would also like to thank the members of the INTERFACE research coordination network (RCN), particularly Jeff Dukes, for their discussion of topics related to this review. N.G.S. was supported by a Purdue Climate Change Research Center (PCCRC) Fellowship and a NASA Earth and Space Science Fellowship. D.N. acknowledges support through NSF CAREER (AGS 0847872) and USDA/NIFA U2U project. Data from analyses and figures will be archived to the Purdue University Research Repository and can be accessed by contacting the corresponding author (NGS). This is publication number 1414 of the PCCRC.

The Editor on this paper was Gregory Okin. He thanks Simone Fatichi, Robert Dickinson, and two anonymous reviewers for their review assistance on this manuscript.

References


Avolio, M., J. Beaulieu, and M. Smith (2012), Genetic diversity of a dominant C₄ grass is altered with increased precipitation variability, Oecologia, 1–11, 571–81.


Beier, C., et al. (2004), Novel approaches to study climate change effects on terrestrial ecosystems in the field: Drought and passive nighttime warming, Ecosystems, 7(6), 583–597.


Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch (2010), A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale, Geosci. Model Dev. Discuss., 3(4), 1829–1859.


Cox, P. M. (2001), Description of the “TRIFFID” Dynamic Global Vegetation Model, Hadley Center Met Office, London Road, Bracknell, Berks, RG12 2YX, U. K.


Hanson, P. J., D. E. Todd, and J. S. Amthor (2001), A six-year study of sapling and large-tree growth and mortality responses to natural and induced variability in precipitation, and throughfall, Tree Physiol., 21(6), 345–358.


He, L., V. Y. Ivanov, G. Bohrer, J. E. Thomsen, C. S. Vogel, and M. Moghaddam (2013), Temporal dynamics of soil moisture in a northern temperate mixed successional forest after a prescribed intermediate disturbance, Agric. For. Meteorol., 180(0), 22–33.


photosynthesis under water-stressed conditions, 

SMITH ET AL.

Reviews of Geophysics

Xu, L., D. D. Baldocchi, and J. Tang (2004), How soil moisture, rain pulses, and growth alter the response of ecosystem respiration to temperature,


