



Evaluating the strength of the land–atmosphere moisture feedback in earth system models using satellite observations

Paul A. Levine¹, James T. Randerson¹, Sean C. Swenson², and David M. Lawrence²

¹Department of Earth System Science, University of California, Irvine, CA 92697, USA

5 ²Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80305, USA

Correspondence to: Paul A. Levine (plevine@uci.edu)

Abstract. The relationship between terrestrial water storage (TWS) and atmospheric processes has important implications for predictability of climatic extremes and projection of future climate change. In places where moisture availability limits evapotranspiration (ET), variability in TWS has the potential to influence surface energy fluxes and atmospheric conditions. 10 Where atmospheric conditions, in turn, influence moisture availability, a full feedback loop exists. Here we developed a novel approach for measuring the strength of both components of this feedback loop, i.e., the forcing of the atmosphere by variability in TWS and the response of TWS to atmospheric variability, using satellite observations of TWS, precipitation, solar radiation, and vapor pressure deficit during 2002–2015. Our approach defines metrics to quantify the relationship between TWS anomalies and climate globally on a seasonal to interannual time scale. Metrics derived from the satellite data 15 were used to evaluate the strength of the feedback loop in 38 members of the Community Earth System Model (CESM) Large Ensemble (LENS) and in six models that contributed simulations to Phase 5 of the Coupled Model Intercomparison Project (CMIP5). We found that both forcing and response limbs of the feedback loop in LENS were stronger than in the satellite observations in tropical and temperate regions. Feedbacks in the selected CMIP5 models were not as strong as those found in LENS, but were still generally stronger than those estimated from the satellite measurements. Consistent with 20 previous studies conducted across different spatial and temporal scales, our analysis suggests that models may overestimate the strength of the feedbacks between the land surface and the atmosphere. We describe several possible mechanisms that may contribute to this bias, and discuss pathways through which models may overestimate ET or overestimate the sensitivity of ET to TWS.

1 Introduction

25 Land–atmosphere feedbacks can result from the coupling of the terrestrial moisture state with temperature, precipitation, or radiation (Betts et al., 2014; Findell and Eltahir, 1997; Guillod et al., 2015; Koster et al., 2004). Land–atmosphere coupling occurs when terrestrial moisture anomalies influence the partitioning of surface energy between latent and sensible heat fluxes that, in turn, influence the development of the planetary boundary layer (PBL) (Seneviratne et al., 2010). Temperature coupling generally leads to a positive feedback, with wetter soil contributing to a higher evaporative fraction (EF, the ratio of



the latent heat flux to the sum of the sensible and latent heat fluxes), a lower surface temperature, and decreased evaporative demand (Hirschi et al., 2011; Miralles et al., 2012). Precipitation coupling can lead to both positive and negative feedbacks, as the influence of EF on the development of the PBL can serve to either enhance or suppress cloud formation and precipitation (Findell and Eltahir, 2003; Guillod et al., 2015). Cloud radiative coupling can likewise lead to positive or negative feedbacks (Betts, 2009; Cheruy et al., 2014). Temperature, precipitation, and radiation feedbacks each stem from coupling between terrestrial moisture and evapotranspiration (ET), which occurs most strongly in conditions of intermediate moisture availability (Seneviratne et al., 2010).

Evidence of these feedbacks has been observed in both in situ and remotely sensed data (Eltahir, 1998; Findell and Eltahir, 1997; Guillod et al., 2014, 2015). Some observational analyses have found land–atmosphere feedback strength to be relatively weak compared to the influence of large-scale atmospheric forcing (Alfieri et al., 2008; Phillips and Klein, 2014). Other observational studies have highlighted the role of these feedback mechanisms in the initiation and exacerbation of climatic extremes such as droughts and heat waves (Hirschi et al., 2011; Miralles et al., 2014; Whan et al., 2015).

Large-scale land–atmosphere coupling in general circulation models has been demonstrated by a series of experiments from the Global Land Atmosphere Coupling Experiment (GLACE) project (Guo et al., 2006; Koster et al., 2004, 2006). The GLACE efforts demonstrate that coupled climate models differ greatly in the extent to which soil moisture variations affect precipitation and surface air temperature, but models generally agree on the spatial distribution of relative coupling strength, with “hotspots” of strong coupling during boreal summer found in the central United States, northern Amazonia, the Sahel, western Eurasia, and northern India. These hotspots are found in regions of intermediate soil wetness, which is consistent with the understanding that strong land–atmosphere coupling occurs under conditions in which terrestrial moisture availability limits ET (Seneviratne et al., 2010). GLACE efforts have also demonstrated that correct soil moisture initialization improves the skill of models in seasonal predictability of temperature and, to a lesser extent, precipitation, particularly in hotspots of strong coupling (Koster et al., 2010, 2011).

Additional studies have considered land–atmosphere feedbacks in the coupled earth system models (ESMs) used by the Intergovernmental Panel on Climate Change (IPCC) (Dirmeyer et al., 2013; Notaro, 2008; Seneviratne et al., 2006, 2013). Notaro (2008) was able to confirm the boreal summer GLACE hotspots, as well as identify several additional austral summer hotspots, in the models used for the IPCC Fourth Assessment Report (AR4). Analysis of long-term projections from the Fifth Phase of the Coupled Model Intercomparison Project (CMIP5) indicates an increased control of land surface moisture on boundary layer conditions with climate change (Dirmeyer et al., 2013). The GLACE Coupled Model Intercomparison Project (GLACE–CMIP5) experiment found that modeled coupling strength plays an important role in simulated response to global warming, with greater warming evident in more strongly coupled models due to interactions between soil moisture, temperature, and precipitation (Berg et al., 2015; May et al., 2015; Seneviratne et al., 2013).



Despite the importance of land–atmosphere coupling in both short-term predictability of climatic extremes and long-term uncertainty in climate change, validation efforts have suggested that climate models may not be correctly representing the strength, and in some cases even the sign, of these feedbacks (Ferguson et al., 2012; Hirschi et al., 2014). Efforts to compare
5 GLACE results with observational data suggest that models may overestimate land–atmosphere coupling strength (Dirmeyer et al., 2006). Zeng et al. (2010) found that version 3 of the Community Climate System Model (CCSM3) showed a higher coupling strength than reanalysis or observational data. Mei and Wang (2012) found that coupling strength was reduced when the Community Atmosphere Model (the land surface component of CCSM3) was updated from version 3 (CAM3) to version 4 (CAM4), though the coupling strength of the updated version was still stronger than observations and reanalysis.

10

The Local Land–Atmosphere Coupling (LoCo) Project has focused on developing a suite of metrics for diagnosing land–atmosphere coupling strength in observations and models. LoCo metrics consider both the influence of soil moisture on EF, and the influence of EF on diurnal-scale boundary layer development (Santanello et al., 2009). Ferguson et al. (2012) used the LoCo approach to compare global remote sensing data sets of soil moisture, evaporative fraction, and lifting
15 condensation level with several land surface models and reanalyses. They found that even though the models were able to simulate the correct spatial pattern of stronger coupling in moist–arid transitional regions, the models tended to simulate a stronger influence of soil moisture on surface turbulent fluxes than what was observed in the satellite data. Guillod et al. (2014) used a combination of flux tower, remote sensing, and reanalysis data sets to demonstrate that measured strength of coupling between EF and precipitation depends greatly on the data source and scale, and that strong coupling apparent in a
20 previous modeling analysis (Findell et al., 2011) was not consistent with the observations.

While many of the previously mentioned studies have confirmed the long-standing suspicion that models may overestimate coupling strength relative to observations, more recent work has indicated that observations and models may not even agree on the sign of the precipitation feedback. Taylor et al. (2012) performed a spatial analysis of the relationship between soil
25 moisture and afternoon precipitation using data from remote sensing, reanalysis, and coupled models. They found evidence of a negative feedback in the remote sensing observations, with afternoon rain being more likely over regions of drier soil, as opposed to the positive feedback that was apparent in the models. Guillod et al. (2015) addressed these findings by replicating the spatial analysis and complementing it with a temporal analysis. They found a negative spatial feedback, consistent with the one found by Taylor et al. (2012), but a positive temporal feedback, with afternoon precipitation at a
30 given location being more likely after mornings of relatively moist soil.

These studies highlight the need for continued efforts toward evaluating the coupling strength of models relative to observations using a wide array of data sources at a range of spatiotemporal scales. Apparent coupling strength depends greatly on the spatial scales of analysis (Hohenegger et al., 2009), indicating that observations at the scale of flux towers



should not be expected to yield the same coupling strength as those at the scale of global climate models (Guilod et al., 2014). Consistency between the spatial scale of observations and models is greatly assisted by the fact that earth observation satellites have been continuously monitoring several relevant land surface and atmospheric variables over multiple years (Teixeira et al., 2014). Measured or modeled coupling strength will also depend on the time scales in question (Guilod et al., 5 2015), and while the LoCo efforts have improved the understanding of synoptic and diurnal scale mechanisms, there is an additional need to examine these processes on seasonal to interannual time periods.

Here we introduce a set of metrics for measuring the strength of land–atmosphere interactions on seasonal timescales by combining satellite remote sensing datasets of terrestrial water storage, precipitation, shortwave radiation, and surface 10 atmospheric temperature and water vapor during 2002–2015. These new metrics complement previous studies and are unique in several ways. In particular, while LoCo metrics focus on day-to-day variability across one or more seasons, the present study considers interannual variability of entire seasons. By focusing on this temporal scale and resolution, our metrics are designed to complement LoCo metrics. Land–atmosphere coupling on seasonal timescales has been shown to be 15 enabling seasonal forecasts of fire risk (Chen et al., 2013, 2016).

Until recently, studies using remote sensing data to look for evidence of land–atmosphere coupling relied on data products that provide information about surface soil moisture. The inability to consider root-zone soil moisture has been suggested as an explanation for the relatively weak coupling observed using remote sensing data (Hirschi et al., 2014). In order to 20 consider all of the sources of moisture available across entire seasons, the present study uses the entire terrestrial water storage (TWS) column, which includes soil moisture at all layers along with surface, canopy, snow/ice, and aquifer storage, as each of these components represents a potential source of moisture for fulfilling evaporative demand. The metrics introduced in this study were specifically designed to take advantage of the monthly TWS anomaly (TWSA) anomaly product from the Gravity Recovery and Climate Experiment (GRACE) mission (Landerer and Swenson, 2012; Wahr et al., 25 2004).

Previous studies have largely focused on land surface moisture availability as a forcing mechanism on the atmosphere, as this relationship has important implications for seasonal predictability as well as the projection of the frequency and severity of climatic extremes. However, it is also critical to assess the response of land surface moisture to atmospheric conditions, as 30 an accurate representation of these processes is essential for generating the correct terrestrial moisture variability that will go on to influence the atmosphere. As far as we can tell, this response limb of the land surface feedback loop has not been systematically integrated with existing analyses of land–atmosphere coupling strength.



Our globally applicable approach uses the annual cycle of TWS drawdown and recharge to isolate the months of the year during which the land surface loses moisture, which we refer to as the drawdown interval (Figure 1a). In our analysis, separate metrics were calculated to consider the influence of TWS at the onset of the drawdown interval on atmospheric conditions in subsequent months, and simultaneously, the influence of atmospheric conditions during the drawdown interval on terrestrial water storage at the end of the season. We refer to these two relationships as the forcing and response limbs, respectively, of the full land surface–atmospheric feedback loop (Figure 1b). We estimated the strength of these feedbacks during 2002–2015 using GRACE and other satellite remote sensing data (Table 1). We then used the satellite observations to evaluate the strength of these feedbacks in the Community Earth System Model (CESM) Large Ensemble (LENS) (Kay et al., 2014) and in several models that contributed simulations to CMIP5 (Table 2).

10

2 Methods

2.1 Remote sensing data

We obtained Level-3 TWSA data from the Gravity Recovery and Climate Experiment (GRACE) platform using the University of Texas at Austin Center for Space Research (CSR) spherical harmonic solutions (Swenson, 2012). GRACE data are available globally over land at a monthly, 1° resolution from September, 2002 through September, 2015, and were scaled using the coefficients provided by Landerer and Swenson (2012), with temporal gaps filled using linear interpolation. At each grid cell, the TWSA time series was decomposed into linear trend, seasonal cycle, and interannual variability components by using ordinary least squares regression. This decomposition allowed us to estimate a mean annual cycle at each grid cell with minimal influence of any long-term trend. While previous studies have focused on the influence of surface soil moisture on boundary layer dynamics, the use of TWS data allows us to include surface storage, canopy storage, snow/ice, deep soil moisture, and groundwater, all of which may be sources of moisture that are potentially limiting factors for ET.

Level-3 near-surface temperature and relative humidity were obtained globally at a monthly, 1° resolution from the ascending (daytime) orbit of the Atmospheric Infrared Sounder (AIRS) platform (Susskind et al., 2014). Vapor pressure deficit (VPD) was calculated from the AIRS data using the August–Roche–Magnus approximation to the Clausius–Clapeyron relation (Lawrence, 2005). Precipitation (PPT) data were obtained from the Global Precipitation Climatology Project (GPCP), a merged satellite and gauge-based data set (Huffman et al., 2009) at a daily, 1° resolution and then integrated monthly. Downwelling shortwave radiation ($SW\downarrow$) was obtained globally at a monthly, 1° resolution from the Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced And Filled (EBAF) Surface product (Loeb et al., 2009). More information describing the remote sensing and reanalysis data products used in this analysis is summarized in Table 1.



To assess how our analysis may be influenced by uncertainty in the satellite data, we substituted data from the European Centre For Medium-range Weather Forecasting (ECMWF) Interim Reanalysis (ERA-Interim) (Dee et al., 2011) in place of AIRS, GPCP, and CERES. We only used atmospheric reanalysis data for this sensitivity analysis, as these data benefit from
5 assimilation of observations, while we continued to use GRACE for TWSA. Comparing results from this GRACE–reanalysis hybrid to those using only satellite data provided an indication of how sensitive our coupling metrics were to the data source.

2.2 Coupling metrics

10 While the existing literature generally defines “land–atmosphere coupling” as the extent to which atmospheric conditions are forced by the land surface state, here we consider the response of the land surface to atmospheric drivers as a coupled relationship that is also important to the full land–atmosphere feedback loop. The land surface response to the atmosphere is governed by many of the same processes through which the land surface forces atmospheric conditions, and therefore we recognize this relationship as a part of land–atmosphere coupling. In the current study, we use the term “coupling” to refer to
15 both the forcing and the response limbs of the full feedback loop.

To quantify the relationship between TWS and the atmosphere on a year-to-year basis, we developed a set of metrics using available remote sensing data. As a first step, we defined a monthly TWS drawdown interval at each 1° land grid cell as the period during which losses from ET and runoff exceeded inputs from precipitation (Figure 1). We selected this interval
20 because past work has shown that the land surface’s influence on the atmosphere is most prevalent during summer in the northern hemisphere (Cheruy et al., 2014; Phillips and Klein, 2014) and during the dry season in tropical forests (Harper et al., 2013; Lorenz and Pitman, 2014). This approach allowed us to investigate land surface coupling at a global scale, and to extend metrics developed in previous work (e.g., Guo and Dirmeyer, 2013; Koster et al., 2006) for pre-defined monthly intervals corresponding with boreal summer to be applicable at any location regardless of seasonality.

25

The drawdown interval was empirically determined as the time period beginning with the month of maximum TWS and ending in the month of minimum TWS, with the maximum and minimum months determined by the climatology of all years for which TWS data were available to ensure that the drawdown interval was consistent from year to year (Figure 2). Northern hemisphere middle and high latitudes exhibit a drawdown interval beginning in the spring (MAM) and ending in
30 the late summer or fall (ASO), reflecting the timing of the boreal summer growing season. At lower latitudes, the North American, African, and Asian monsoons are evident, with Mexico, India, and the Sahel showing a drawdown interval that begins in September, after the monsoonal precipitation has peaked, and ends the following spring after the winter dry season. The onset of the drawdown season reverses abruptly at the equator in Africa and Asia, and with the drawdown interval



reflecting a winter dry season in the austral low latitudes transitioning to a summer growing season in the austral midlatitudes.

With the drawdown interval defined, we then developed a set of coupling metrics for forcing and response limbs of the land surface–atmosphere feedback (Figure 1b). To quantify the land surface forcing, we assessed the influence of the land surface state at the onset of the drawdown interval ($TWSA_{max}$) on surface atmospheric variables during the drawdown interval (ATM_{di}). In our analysis, we selected 3 variables to represent the atmospheric state: VPD, SW_{\downarrow} , and PPT. These atmospheric variables were averaged during the drawdown interval, including during the months of climatological maximum and minimum TWSA. We chose these variables because they represent various aspects of evaporative supply (PPT) and demand (VPD and SW_{\downarrow}). Monthly GRACE, AIRS, GPCP, and CERES observations from September 2002 through September 2015 provided 12 or 13 complete drawdown intervals at each grid cell, depending on the months of $TWSA_{max}$ and $TWSA_{min}$. In regions of strong land surface forcing of the atmosphere, higher than average TWS would typically be followed by lower than average VPD, lower than average SW_{\downarrow} , and higher than average PPT. Thus we would expect strongly coupled regions to have negative correlation between $TWSA_{max}$ and VPD_{di} , negative correlation between $TWSA_{max}$ and $SW_{\downarrow,di}$, and positive correlations between $TWSA_{max}$ and PPT_{di} .

Similarly, to assess the impact of the atmosphere on the land surface state (the response limb), we examined the relationship between atmospheric variables during the drawdown interval (ATM_{di}) and the land surface state at the end of the drawdown interval ($TWSA_{min}$). Strong land surface response to the atmospheric conditions yielded negative correlation between VPD_{di} and $TWSA_{min}$, negative correlation between $SW_{\downarrow,di}$ and $TWSA_{min}$, and positive correlation between PPT_{di} and $TWSA_{min}$. Although most previous diagnoses of land–atmosphere coupling has focused on the forcing limb, we argue the response limb is equally important as a metric for model evaluation. Specifically, if variability in the balance between evaporative supply and demand does not lead to the correct TWS variability, then the incorrect TWS response will feed back into subsequent forcing on the atmosphere.

25

The Pearson product-moment correlation coefficient between $TWSA_{max}$ and ATM_{di} is referred to as the forcing metric, while the correlation coefficient between ATM_{di} and $TWSA_{min}$ is referred to as the response metric. In both metrics, the sign of the correlation indicating strong coupling depends on the atmospheric variable. In the case of PPT, more positive correlation coefficients indicate stronger coupling, while in the cases of VPD and SW_{\downarrow} , more negative correlation coefficients indicate stronger coupling.

30

2.3 Earth system model evaluation

We used the metrics described above to evaluate coupling strength in the Community Earth System Model (CESM) Large Ensemble (LENS). LENS comprises an ensemble of 38 fully coupled runs in which air temperature initial conditions are



perturbed slightly (by an amount less than round-off error) to reveal the internal variability inherent within the interactive climate system. LENS has demonstrated that the uncertainty in climate projections due to internal climate variability inherent in CESM is comparable to the ranges of output within the entire CMIP5 experiment (Kay et al., 2014). LENS uses version 1 of CESM (CESM1) with version 5 of the Community Atmosphere Model (CAM5) and version 4 of the
5 Community Land Model (CLM4) at a horizontal resolution of 1°. The ensemble run follows protocols from the CMIP5 experiment, with historical radiative forcing for the 20th century and representative concentration pathway 8.5 (RCP8.5) forcing for the 21st century.

The LENS data were chosen as a starting point for feedback evaluation for two reasons. First, the availability of a TWS
10 variable enables direct comparison with metrics derived using data from GRACE. The TWS field in CLM4 includes water from surface and canopy storage, snow and ice, soil moisture, and a dynamic aquifer, in addition to river water storage terms from the coupled River Transport Module (RTM). The coupling of CLM4 with RTM is important for simulating both the annual cycle and interannual variability of TWS in comparison with GRACE (Kim et al., 2009).

15 Second, the ensembles allow us to test the internal variability on diagnosed feedback strength. Because the complete satellite record is relatively short (containing no more than 13 drawdown intervals at any location), comparison with an equivalent single time series of model output could reflect the model's internal variability (Kay et al., 2014) more than its ability to reproduce the observed relationships. Analyzing the full ensemble from LENS enables the satellite record to be compared with the range of CESM's ability to reproduce the observed relationships. The equivalent months of the satellite record
20 (September, 2002 – April 2015) were retrieved, with data prior to December, 2005 coming from the historical runs, and data from January, 2006 onward coming from the RCP8.5 simulations.

To extend our analysis to models that do not output an explicit TWS field, the accumulated residuals of precipitation, evapotranspiration, and total runoff (surface and subsurface) were compared with TWS in LENS (supplementary figures).
25 After we determined that the accumulated residuals of the water balance represented much of the variability in the explicit TWS variable, we calculated equivalent metrics for several model simulations in the CMIP5 archive (Table 2). We selected the CMIP5 models that were similar to LENS (CESM1-CAM5 and CESM1-BGC) as well as the models that participated in the GLACE–CMIP5 experiment (Seneviratne et al., 2013) for which all necessary output fields were available (CCSM4, GFDL-ESM2M, GFDL-ESM2G, IPSL-CM5A-MR, and IPSL-CM5A-LR).

30



3 Results

3.1 Drawdown interval and interannual variability

A comparison of the months of maximum and minimum terrestrial water storage as determined by climatologies of GRACE and the LENS ensemble mean indicated that the model largely reproduces the timing of TWSA seasonality evident in the satellite observations (Figure 2). Geographic patterns of seasonality were consistent between the model and observations, though a phase shift in the drawdown interval is apparent in eastern Canada and central Eurasia where LENS had a one-month early phase bias for both the maximum and minimum TWSA, and in southeast North America where the modeled drawdown interval was slightly later than the observations. However, despite capturing generally correct timing, the model exhibited higher interannual variability of $TWSA_{max}$ and $TWSA_{min}$ across the 12–13 drawdown intervals compared with the satellite data (Figure 3) particularly in the southern United States, southern South America, central and eastern Africa, southern Asia, and eastern Australia. One possible explanation for this is the presence of multi-year trends in aquifer storage in CLM that are not consistent with GRACE (Swenson and Lawrence, 2015).

A comparison of the interannual variability of atmospheric variables across multiple drawdown intervals between the model and satellite data showed various degrees of consistency (Figure 4). The magnitude and geographic pattern of VPD_{di} was generally consistent, though LENS showed greater interannual variability in central and western North America, South America, northern and southern Africa, and southern Asia compared with AIRS. In the case of PPT_{di} , LENS showed less interannual variability than GPCP in Southeast North America and much of South America, but the two were largely consistent elsewhere. $SW\downarrow_{di}$ was the least consistent between the model and satellite data, as LENS showed greater interannual variability in southern North America, northern Eurasia, most of Africa, and most of Australasia compared with CERES.

3.2 Evaluating feedbacks for a single model simulation

The forcing metric for VPD derived from GRACE and AIRS showed regions of strong coupling, in which $TWSA_{max}$ was negatively correlated with VPD_{di} , in the northern Great Plains, northern South America, southern Africa, southern and western India, north central Eurasia, and northern Australia (Figure 5a). Regions with strong positive correlation were much less common, and were largely confined to areas of very low GRACE-derived $TWSA_{max}$ variability (Figure 3a). In comparison with the satellite data, the VPD forcing metrics from the first ensemble member of LENS (Figure 5c) showed much stronger coupling in the southern Amazon, and weaker coupling elsewhere, with a few regions of negative coupling (positive correlation).

30

The response metrics for VPD showed much stronger coupling than the forcing metric in both the model and satellite data (Figure 5b,d). Satellite data yielded negative correlation coefficients nearly everywhere, with positive correlations found



only in arid regions of low TWS variability. Particularly strong response metrics were found in eastern North America, northern South America, western Eurasia, the Sahel, India, and eastern Australia. The first ensemble member from LENS showed widespread negative correlations, and did not show the positive correlations found in the satellite data. Response coupling in LENS was much more spatially homogeneous than in the satellite data, though northern South America and western Eurasia still showed stronger coupling than elsewhere.

Many of the areas that showed a strong forcing metric for VPD also showed a relatively strong forcing metric for PPT, though the PPT forcing metric was overall weaker than that for VPD (Figure 6a). The response metric for PPT was generally positive, indicating that for much of the globe, higher precipitation rates were associated with larger $TWSA_{min}$ (Figure 6b). Both the forcing and response metrics were somewhat stronger in the LENS member relative to those evident in the satellite data (Figure 6c,d).

The forcing metrics for $SW\downarrow$ showed a mixture of positive and negative correlations, indicating that higher $TWSA_{max}$ could be either positively or negatively coupled with shortwave radiation (Figure 7a). The response metrics for $SW\downarrow$ were generally negative, indicating that greater seasonal shortwave radiation was associated with more negative $TWSA_{min}$ (stronger coupling), with West Africa being a notable exception (Figure 7b). The LENS member showed generally stronger coupling in both the forcing and response metrics for $SW\downarrow$ (Figure 7c,d).

3.3 Evaluating the CESM Large Ensemble

The internal variability in LENS yielded a distribution of forcing (Figure 8) and response (Figure 9) metrics with a spread on the same order of magnitude as the difference between modeled and satellite-derived zonal averages. In temperate and tropical regions, forcing metrics were generally stronger in LENS (more positive for PPT, more negative for VPD and $SW\downarrow$) than in the satellite and reanalysis data, indicating a stronger land surface forcing of the surface atmospheric state in the model than in the observations. In boreal regions, forcing metrics were much weaker (closer to zero) than at lower latitudes in both the satellite data and in LENS, indicating very little relationship between $TWSA_{max}$ and ATM_{di} . This is consistent with high levels of climate variability in many high latitude regions driven by the Arctic Oscillation, the NAO, and other dynamical modes (Cohen and Barlow, 2005).

Response metrics were also generally higher in LENS than in both the satellite and reanalysis data. Noticeable exceptions were the VPD and PPT response metrics in the tropics, which were close to the satellite observations, and the boreal $SW\downarrow$, midlatitude VPD, and tropical PPT response metrics, were close to the reanalysis estimates. Despite the internal variability evident within the model ensemble, and the difference between metrics as measured by the satellite data compared with the reanalysis data, the general pattern indicates that modeled response metrics were higher than those from observations and reanalysis.



Comparison of the forcing and response metrics calculated using the explicit TWS field from LENS with the equivalent metrics calculated using the accumulating residuals of the surface water budget (precipitation – ET – runoff) indicated that the alternative formulation provides an acceptable substitute when output of an explicit TWS field is not available from an
5 ESM (see supplementary figures). This suggests that metrics calculated for CMIP5 output using accumulating residuals can be reasonably and effectively compared with the metrics derived from LENS and the observations.

3.4 CMIP5 models

Comparison of CMIP5 and LENS models indicated a mostly positive relationship between forcing and response metrics in temperate and tropical latitude bands (Figure 10). Models that showed the strongest forcing metrics often had the strongest
10 response metrics for a given variable. In boreal latitudes, the forcing metrics were all very close to zero, with little distinction between the different models, though there were some clear differences in the response metrics between the various models. LENS exhibited generally the strongest forcing and response metrics among the models, with the CMIP5 models that were closely related to LENS, particularly CESM1-CAM5, found close by. This relationship between forcing and response
15 metrics suggests that analysis of the response limb of the feedback loop is important for understanding how conditions are set up for subsequent forcing via land–atmosphere coupling.

The modeled metrics indicated generally stronger coupling than the observations for both the forcing and response limbs. Exceptions include the VPD response metric in the tropics, the boreal PPT and SW↓ forcing metrics, and the midlatitude SW↓ response metrics. The spread between various models was generally greater than the spread within any single model
20 with a multi-member ensemble, though the NCAR models (CCSM/CESM) were all relatively close to one another. The GFDL-ESM2M model is an exception, as both the forcing and response metrics of SW↓ were far from the remainder of the models.

4 Introduction

4.1 Observed feedback metrics

25 The metrics developed here from satellite observations provide a means for evaluating land–atmosphere feedback strength on seasonal to interannual timescales in coupled ESMs. Although the use of correlation coefficients does not indicate whether the relationships are directly causal, the satellite-derived metrics nonetheless provide a meaningful constraint against which coupled models can be benchmarked. Furthermore, the well-understood physical mechanisms allow causality to be inferred even when not directly demonstrated.

30



The forcing metrics, by indicating the relationship between antecedent TWS and subsequent atmospheric characteristics, provide observational constraints to complement previous research in large scale land–atmosphere coupling in global models (e.g., Guo and Dirmeyer, 2013; Koster et al., 2006; Seneviratne et al., 2013). The current results largely agree with GLACE and other experiments that have demonstrated land–atmosphere coupling being strongest in zones of intermediate wetness (Koster et al., 2006; Seneviratne et al., 2010). The observed forcing metrics also indicated regions of strong coupling in the moist tropics that are considered generally more energy-limited when it comes to ET. Recent observational analyses by (Hilker et al., 2014) demonstrate that at least in the Amazon, deep rooting zone water supplies can become seasonally depleted, leading to a stronger land–atmosphere coupling. This is consistent with findings that deep rooted plants vertically redistribute soil water to shallower layers, allowing higher levels of evapotranspiration to be sustained during the dry season (Lee et al., 2005). It is also consistent with recent work demonstrating that TWSAs can be used as predictors for fire season severity in the Amazon (Chen et al., 2013).

The inclusion of the response metrics allows the full feedback loop to be considered by recognizing the two-way dependence between the land surface and the atmosphere. That models and ensemble members with high forcing metrics were also found to have high response metrics (Figure 10) highlights the importance of the land surface response in priming the system for subsequent forcing on the atmosphere, and the need to consider the full feedback loop. Assessment of modeled vs. observed coupling metrics elucidates whether land–atmosphere feedbacks, which are known to significantly impact long-term climate projections (Berg et al., 2015; May et al., 2015), are of a realistic strength and sign.

4.2 Modeled feedback metrics

To properly represent land–atmosphere moisture feedbacks, models need to accurately simulate several related processes for each limb of the feedback loop. (Guo et al., 2006) explain that the requirements for strong land–atmosphere coupling include a sufficiently strong influence of soil moisture on ET, a sufficiently high ET variability, and a sufficiently strong atmospheric response to soil moisture influenced ET variability. The LoCo efforts (Ferguson et al., 2012; Findell et al., 2011; Guillod et al., 2014; Santanello et al., 2009) similarly characterize coupling as the combined influence of moisture availability on EF and of EF on the development of the PBL. Our forcing metric is limited in that it cannot distinguish between the TWS–ET relationship and the relationship between ET and the atmospheric variable of interest. However, it still is useful for determining whether models sufficiently capture the combined effect of these processes.

In addition, models must capture the correct response of TWS to the supply (precipitation) and demand (ET) of the atmosphere. Incomplete or incorrect representations of the surface water budget, including interception, infiltration, and runoff, would lead to incorrect variability of TWS. This would then feed back into the atmosphere by incorrectly driving the forcing mechanisms. Therefore, even if a model correctly captures the processes in which TWS forces the atmosphere, a correct response is still necessary for accurately capturing both the forcing and response metrics.



Finally, if the relative importance of local land surface processes compared with atmospheric circulation is not correctly represented then the modeled forcing metrics will not be realistic. Even if the models correctly represent the relationship between TWS and ET, dynamical systems resulting from sea surface temperature (SST) forcing or internal atmospheric variability (Seager and Hoerling, 2014) could reduce the influence of local land surface processes. The relative importance of land and atmospheric influence also depends on whether the limitations to ET are correctly represented, as moisture limited conditions allow for strong controls of TWS while energy limited conditions depend on atmospheric conditions (Rahman et al., 2014).

4.3 Enhanced feedback strength in models

- 10 The feedback metrics as calculated from the output of the ESMs analyzed in the current study indicated generally stronger coupling compared with those derived from the satellite observations. There are exceptions to this pattern, but it holds generally true, particularly across middle and lower latitudes, and particularly in the LENS data. This is consistent with previous studies conducted at finer temporal resolutions (Ferguson et al., 2012) and across more limited spatial domains (Hirschi et al., 2011). As described below, there are several possible explanations as to why models may simulate a stronger
- 15 feedback than is observed in the satellite record. Each of these explanations represents a potential direction for future research to test these mechanisms in individual models. Possible explanations include overestimation of ET, overly simplistic convective and stomatal conductance parameterizations, and underestimation of and remote SST forcing of atmospheric dynamics over land.
- 20 One set of explanations involves models overestimating the amount of water available for ET, either by improperly partitioning precipitation between runoff and storage or incorrectly moving water between terrestrial storage reservoirs (e.g., canopy, soil, and surface storage). If incoming precipitation is not sufficiently routed out of the model grid cell through runoff, water will be too readily available for ET, which would yield an unrealistically strong connection between TWS and the atmosphere. Such a model would then overestimate both the forcing and response metrics, and would also exhibit a high
- 25 bias for ET. Similarly, any model shortcomings that cause water to move between storage reservoirs in a way that makes water too readily available for ET will cause the model to simulate too strong of a feedback. Notable examples of these issues, described below, include incorrect simulation of precipitation intensity and misrepresentation of bare soil fraction and processes.
- 30 ESMs are known to simulate unrealistically homogeneous rainfall intensity, with overestimates of drizzle and underestimates of large infrequent events (Dai, 2006). Infrequent high-intensity rainfall events would yield much more runoff from saturated soil, which would lead to a weaker connection between the land and atmosphere than frequent low-intensity drizzle. If a model simulates too much drizzle, precipitation could lead to too much storage, which would cause a model to overestimate



the response metrics. Too much storage also could allow water to be too readily available for ET, causing an overestimate of the forcing metrics. Contributions from drizzle could be offset if insufficient rainfall intensity does not allow high enough throughfall or soil moisture recharge. The issue of rainfall intensity is related to issues of convective parameterization (described below), and may be addressed in future versions of ESMs through atmospheric superparameterization, in which a model's convective parameterization is replaced with embedded cloud resolving models (Kooperman et al., 2016).

A misrepresentation of either the amount of bare soil or of bare soil processes also could lead to overestimates of ET and coupling strength. Current land surface schemes of ESMs are based on the “big leaf” model paradigm, which could lead to overestimates of ET if recharge and runoff are underestimates as a consequence of too small of a bare soil fraction. In addition, even if bare soil fraction is correct, overestimates of ET due to an incomplete representation of surface resistance of bare soil, as found in CLM by Swenson and Lawrence (2014), would amplify positive feedbacks.

Additional explanations for why models may overestimate feedback strength include the parameterization of convection in the PBL or stomatal conductance responses to soil moisture. Previous work using a regional climate model (RCM) with a higher spatial resolution have determined that convective parameterizations are as important as spatial resolution in the simulation of precipitation coupling (Hohenegger et al., 2009). Taylor et al. (2013) similarly found parameterized convection in an RCM yielding a positive coupling in contrast to both observations and model runs with explicitly simulated convection, both of which yielded negative coupling. It is therefore possible that the convective parameterizations of ESMs are leading to an overestimate of feedback metrics. Similarly, the diversity of stomatal conductance parameterizations in CMIP5 ESMs is relatively low (Medlyn et al., 2011), and if stomatal apertures close too rapidly in response to an initial deficit in terrestrial water storage, transpiration–humidity feedbacks may be artificially intensified.

Finally, if the model underestimates the strength of remote forcing from SST anomalies or dynamic atmospheric processes, the strength of the coupling metrics will be overestimated. Orlowsky and Seneviratne (2010) urge caution when inferring land–atmosphere coupling relationships based on statistical similarities between land surface moisture and atmospheric conditions due in part to the confounding influence of SST forcing. Mei and Wang (2011) argue that remote SST forcing and local land surface conditions are both important factors for determining summertime precipitation in a midlatitude location, with soil moisture being most important during years of low precipitation. These findings indicate the importance of remote atmospheric forcing, and suggest that insufficient representation of these factors could lead to an overestimation of modeled feedback strength.

4.4 Implications for climate modeling

The ability of coupled models to correctly represent observed feedbacks, and knowledge of which models provide better or worse representations, has important implications for projections of future climate change. If models are representing



feedbacks with too great a strength, trajectories of future warming and drying may be unrealistically extreme. Midlatitude warm biases in CMIP5 models are known to be greatest in regions of strong coupling, and were determined by Cheruy et al. (2014) to result from both the underestimation of evaporative fraction and an overestimation of shortwave radiative forcing, both of which yield positive feedbacks. Cheruy et al. (2014) found that while the radiative feedback was more widespread, the evaporative fraction feedback is limited to regions of strong coupling, and both mechanisms are responsible for the present day temperature bias as well as the broad uncertainty in future projections.

Similarly, the GLACE–CMIP5 experiment (Seneviratne et al., 2013) found that the models exhibiting the strongest coupling strength also showed the strongest warming trends through the 21st century. Berg et al. (2015) investigated this experiment further, and found that negative correlations between precipitation and temperature resulted from both a direct atmospheric effect, in which higher air temperatures can hold more precipitable water, as well as an indirect land surface effect, in which both temperature and precipitation are linked through land surface interactions. Berg et al. (2015) determined that if models overestimate the contribution of the indirect terrestrial pathway, then they will yield overestimates of warming. This has particularly important consequences for predictions of climatic extremes such as droughts and heat waves (Hirschi et al., 2011; Miralles et al., 2014).

4.5 Uncertainties and future applications

The current study demonstrates the utility of the feedback metrics presented here, but conclusions are limited by the time span of the satellite record. While LENS enables the internal variability of these relationships to be investigated within the model, it is unclear how much natural climate variability affects these relationships in reality on timescales longer than the 13-year satellite record. The utility of the feedback metrics will increase alongside the length of the time series available from remote sensing platforms. This emphasizes the importance of the GRACE follow-on mission (Flechtner et al., 2014) and the need for continuity in the record between missions.

Furthermore, incorporating additional remote sensing products can reduce uncertainties inherent in the satellite-derived data sets. We presented metrics derived using ERA-Interim in place of AIRS, GPCP, and CERES in order to qualitatively illustrate this uncertainty. We found a non-negligible amount of uncertainty in both forcing and response metrics due to inconsistencies between the remote sensing and reanalysis products. Future work will address these uncertainties by incorporating additional observations and observationally constrained data sets such as those from the Global Energy and Water Experiment, the Global Precipitation Climatology Centre, and the Global Land Data Assimilation System. In addition, as increasingly long time series of data become available from the Soil Moisture–Ocean Salinity (Mecklenburg et al., 2012) and Soil Moisture Active Passive (Panciera et al., 2014) missions, the metrics developed here can be applied to those data sets as well, which will elucidate the importance of surface soil moisture relative to TWS in these interactions.



5 Conclusion

We have developed a new approach for measuring the strength of the two-way feedback relationships between TWS and the atmosphere. This approach was designed specifically to take advantage of the 13 years of TWSA data from the GRACE mission, along with concurrently collected remote sensing and reanalysis data sets of atmospheric variables, in a manner that could then be applied to coupled models of the land and atmosphere. Feedback metrics quantify both the relationship between antecedent TWS and subsequent atmospheric conditions, as well as antecedent atmospheric conditions and subsequent TWS.

Regions of strong forcing, in which the TWSA at the beginning of the drawdown interval was related to the subsequent atmospheric state, coincided with the semi-arid zones previously found to be hot spots of land atmosphere coupling, as well as some new tropical zones that may have moisture limited ET regimes. Regions of strong coupling metrics, in which the TWSA at the end of the drawdown interval is related to the atmosphere, are much more widespread. Modeled feedback metrics are generally found to be stronger than those observed in the satellite record, which suggests that some of these models may have difficulty properly predicting warming trends and climatic extremes.

The results of this study are consistent with previous studies at smaller temporal scales indicating land–atmosphere coupling strength may be stronger in models than in observations. There are several possible mechanisms that may contribute to the overestimation of land–atmosphere coupling in models, and future studies may incorporate the metrics introduced here to assess the role of these mechanisms. These metrics will become increasingly useful as the temporal coverage of the remote sensing record grows longer and additional missions come online.

Acknowledgements. We received funding support from the United States Department of Energy Office of Science Biogeochemistry Feedbacks Scientific Focus Area and the Climate Change Prediction Program, Cooperative Agreement (DE-FC03-97ER62402/A010). The CESM Large Ensemble Community Project was supported by the National Science Foundation (NSF) with supercomputing resources provided by the Climate Simulation Laboratory at NCAR’s Computational and Information Systems Laboratory (CISL). We also acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 2 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.



References

- Alfieri, L., Claps, P., D'Odorico, P., Laio, F. and Over, T. M.: An analysis of the soil moisture feedback on convective and stratiform precipitation, *J. Hydrometeorol.*, 9(2), 280–291, doi:10.1175/2007JHM863.1, 2008.
- Berg, A., Lintner, B. R., Findell, K., Seneviratne, S. I., van den Hurk, B., Ducharne, A., Chérut, F., Hagemann, S., Lawrence, D. M., Malyshev, S., Meier, A. and Gentile, P.: Interannual Coupling between Summertime Surface Temperature and Precipitation over Land: Processes and Implications for Climate Change, *J. Clim.*, 28(3), 1308–1328, doi:10.1175/JCLI-D-14-00324.1, 2015.
- Betts, A. K.: Land-surface–atmosphere coupling in observations and models, *J. Adv. Model. Earth Syst.*, 1(3), doi:10.3894/JAMES.2009.1.4, 2009.
- 10 Betts, A. K., Desjardins, R., Worth, D. and Beckage, B.: Climate coupling between temperature, humidity, precipitation, and cloud cover over the Canadian Prairies, *J. Geophys. Res. Atmos.*, 119(23), 13,305–13,326, doi:10.1002/2014JD022511, 2014.
- Chen, Y., Velicogna, I., Famiglietti, J. S. and Randerson, J. T.: Satellite observations of terrestrial water storage provide early warning information about drought and fire season severity in the Amazon, *J. Geophys. Res. Biogeosciences*, 118(2), 15 495–504, doi:10.1002/jgrg.20046, 2013.
- Chen, Y., Morton, D. C., Andela, N., Giglio, L. and Randerson, J. T.: How much global burned area can be forecast on seasonal time scales using sea surface temperatures?, *Environ. Res. Lett.*, 11(4), 45001, doi:10.1088/1748-9326/11/4/045001, 2016.
- Cherut, F., Campoy, A., Dupont, J.-C., Ducharne, A., Hourdin, F., Haefelin, M., Chiriaco, M. and Idelkadi, A.: Combined 20 influence of atmospheric physics and soil hydrology on the simulated meteorology at the SIRTa atmospheric observatory, *Clim. Dyn.*, 40(9), 2251–2269, doi:10.1007/s00382-012-1469-y, 2013.
- Cherut, F., Dufresne, J. L., Hourdin, F. and Ducharne, A.: Role of clouds and land–atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations, *Geophys. Res. Lett.*, 41(18), 6493–6500, doi:10.1002/2014GL061145, 2014.
- 25 Cohen, J. and Barlow, M.: The NAO, the AO, and global warming: how closely related?, *J. Clim.*, 18(21), 4498–4513, doi:10.1175/JCLI3530.1, 2005.
- Dai, A.: Precipitation characteristics in eighteen coupled climate models, *J. Clim.*, 19(18), 4605–4630, doi:10.1175/JCLI3884.1, 2006.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, 30 G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N.



- and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137(656), 553–597, doi:10.1002/qj.828, 2011.
- Dirmeyer, P. A., Koster, R. D. and Guo, Z.: Do global models properly represent the feedback between land and atmosphere?, *J. Hydrometeorol.*, 7(6), 1177–1198, doi:10.1175/JHM532.1, 2006.
- 5 Dirmeyer, P. A., Jin, Y., Singh, B. and Yan, X.: Trends in Land–Atmosphere Interactions from CMIP5 Simulations, *J. Hydrometeorol.*, 14(3), 829–849, doi:10.1175/JHM-D-12-0107.1, 2013.
- Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L.,
- 10 Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N. and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, *Clim. Dyn.*, 40(9), 2123–2165, doi:10.1007/s00382-012-1636-1, 2013.
- 15 Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T. and Zadeh, N.: GFDL’s ESM2 global coupled climate–carbon earth system models. Part I: Physical formulation and baseline simulation characteristics, *J. Clim.*, 25(19), 6646–6665, doi:10.1175/JCLI-D-11-00560.1, 2012.
- 20 Eltahir, E. A. B.: A soil moisture–rainfall feedback mechanism: 1. Theory and observations, *Water Resour. Res.*, 34(4), 765–776, doi:10.1029/97WR03499, 1998.
- Ferguson, C. R., Wood, E. F. and Vinukollu, R. K.: A global intercomparison of modeled and observed land–atmosphere coupling, *J. Hydrometeorol.*, 13(3), 749–784, doi:10.1175/JHM-D-11-0119.1, 2012.
- Findell, K. L. and Eltahir, E. A. B.: An analysis of the soil moisture–rainfall feedback, based on direct observations from
- 25 Illinois, *Water Resour. Res.*, 33(4), 725–735, doi:10.1029/96WR03756, 1997.
- Findell, K. L. and Eltahir, E. A. B.: Atmospheric controls on soil moisture–boundary layer interactions. Part I: Framework development, *J. Hydrometeorol.*, 4(3), 552–569, doi:10.1175/1525-7541(2003)004<0552:ACOSML>2.0.CO;2, 2003.
- Findell, K. L., Gentine, P., Lintner, B. R. and Kerr, C.: Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation, *Nat. Geosci.*, 4(7), 434–439, doi:10.1038/ngeo1174c, 2011.
- 30 Flechtner, F., Morton, P., Watkins, M. and Webb, F.: Status of the GRACE Follow-On Mission, in *Gravity, Geoid and Height Systems: Proceedings of the IAG Symposium*, edited by U. Marti, pp. 117–121, Springer International Publishing, 2014.
- Guillod, B. P., Orlowsky, B., Miralles, D., Teuling, A. J., Blanken, P. D., Buchmann, N., Ciais, P., Ek, M., Findell, K. L., Gentine, P., Lintner, B. R., Scott, R. L., den Hurk, B. and I. Seneviratne, S.: Land-surface controls on afternoon precipitation



- diagnosed from observational data: uncertainties and confounding factors, *Atmos. Chem. Phys.*, 14(16), 8343–8367, doi:10.5194/acp-14-8343-2014, 2014.
- Guilod, B. P., Orlowsky, B., Miralles, D. G., Teuling, A. J. and Seneviratne, S. I.: Reconciling spatial and temporal soil moisture effects on afternoon rainfall, *Nat Commun*, 6, doi:10.1038/ncomms7443, 2015.
- 5 Guo, Z. and Dirmeyer, P. A.: Interannual variability of land–atmosphere coupling strength, *J. Hydrometeorol.*, 14(5), 1636–1646, doi:10.1175/JHM-D-12-0171.1, 2013.
- Guo, Z., Dirmeyer, P. A., Koster, R. D., Sud, Y. C., Bonan, G., Oleson, K. W., Chan, E., Versegny, D., Cox, P., Gordon, C. T., McGregor, J. L., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Mitchell, K., Malyshev, S., McAvaney, B., Oki, T., Yamada, T., Pitman, A., Taylor, C. M., Vasic, R. and Xue, Y.: GLACE: The Global Land–
- 10 Atmosphere Coupling Experiment. Part II: Analysis, *J. Hydrometeorol.*, 7(4), 611–625, doi:10.1175/JHM511.1, 2006.
- Harper, A., Baker, I. T., Denning, A. S., Randall, D. A., Dazlich, D. and Branson, M.: Impact of evapotranspiration on dry season climate in the Amazon forest, *J. Clim.*, 27(2), 574–591, doi:10.1175/JCLI-D-13-00074.1, 2013.
- Hilker, T., Lyapustin, A. I., Tucker, C. J., Hall, F. G., Myneni, R. B., Wang, Y., Bi, J., de Moura, Y. and Sellers, P. J.: Vegetation dynamics and rainfall sensitivity of the Amazon, *Proc. Natl. Acad. Sci.*, doi:10.1073/pnas.1404870111, 2014.
- 15 Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B. and Stepanek, P.: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, *Nat. Geosci.*, 4(1), 17–21, doi:10.1038/ngeo1032, 2011.
- Hirschi, M., Mueller, B., Dorigo, W. and Seneviratne, S. I.: Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability, *Remote Sens. Environ.*, 154(0), 246–252, doi:10.1016/j.rse.2014.08.030, 2014.
- 20 Hohenegger, C., Brockhaus, P., Bretherton, C. S. and Schär, C.: The soil moisture–precipitation feedback in simulations with explicit and parameterized convection, *J. Clim.*, 22(19), 5003–5020, doi:10.1175/2009JCLI2604.1, 2009.
- Hourdin, F., Foujols, M.-A., Codron, F., Guemas, V., Dufresne, J.-L., Bony, S., Denvil, S., Guez, L., Lott, F., Ghattas, J., Braconnot, P., Marti, O., Meurdesoif, Y. and Bopp, L.: Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model, *Clim. Dyn.*, 40(9), 2167–2192, doi:10.1007/s00382-012-1411-3, 2013.
- 25 Huffman, G. J., Adler, R. F., Bolvin, D. T. and Gu, G.: Improving the global precipitation record: GPCP Version 2.1, *Geophys. Res. Lett.*, 36(17), doi:10.1029/2009GL040000, 2009.
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K.,
- 30 Polvani, L. and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability, *Bull. Am. Meteorol. Soc.*, 96(8), 1333–1349, doi:10.1175/BAMS-D-13-00255.1, 2014.
- Kim, H., Yeh, P. J.-F., Oki, T. and Kanae, S.: Role of rivers in the seasonal variations of terrestrial water storage over global basins, *Geophys. Res. Lett.*, 36(17), doi:10.1029/2009GL039006, 2009.



- Kooperman, G. J., Pritchard, M. S., Burt, M. A., Branson, M. D. and Randall, D. A.: Robust effects of cloud superparameterization on simulated daily rainfall intensity statistics across multiple versions of the Community Earth System Model, *J. Adv. Model. Earth Syst.*, doi:10.1002/2015MS000574, 2016.
- Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence,
5 D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y. and Yamada, T.: Regions of strong coupling between soil moisture and precipitation, *Science*, 305(5687), 1138–1140, doi:10.1126/science.1100217, 2004.
- Koster, R. D., Sud, Y. C., Guo, Z., Dirmeyer, P. A., Bonan, G., Oleson, K. W., Chan, E., Verseghy, D., Cox, P., Davies, H., Kowalczyk, E., Gordon, C. T., Kanae, S., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Mitchell, K., Malyshev, S.,
10 McAvaney, B., Oki, T., Yamada, T., Pitman, A., Taylor, C. M., Vasic, R. and Xue, Y.: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview, *J. Hydrometeorol.*, 7(4), 590–610, doi:10.1175/JHM510.1, 2006.
- Koster, R. D., Mahanama, S. P. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., Dirmeyer, P. A., Doblas-Reyes, F. J., Drewitt, G., Gordon, C. T., Guo, Z., Jeong, J.-H., Lawrence, D. M., Lee, W.-S., Li, Z., Luo, L., Malyshev, S., Merryfield, W. J., Seneviratne, S. I., Stanelle, T., van den Hurk, B. J. J. M., Vitart, F. and Wood, E. F.: Contribution of land
15 surface initialization to subseasonal forecast skill: First results from a multi-model experiment, *Geophys. Res. Lett.*, 37(2), L02402, doi:10.1029/2009GL041677, 2010.
- Koster, R. D., Mahanama, S. P. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., Dirmeyer, P. A., Doblas-Reyes, F. J., Drewitt, G., Gordon, C. T., Guo, Z., Jeong, J.-H., Lee, W.-S., Li, Z., Luo, L., Malyshev, S., Merryfield, W. J., Seneviratne, S. I., Stanelle, T., van den Hurk, B. J. J. M., Vitart, F. and Wood, E. F.: The second phase of the Global Land–
20 Atmosphere Coupling Experiment: Soil moisture contributions to subseasonal forecast skill, *J. Hydrometeorol.*, 12(5), 805–822, doi:10.1175/2011JHM1365.1, 2011.
- Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, 48(4), W04531, doi:10.1029/2011WR011453, 2012.
- Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L.,
25 Levis, S., Sakaguchi, K., Bonan, G. B. and Slater, A. G.: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, *J. Adv. Model. Earth Syst.*, 3(1), n/a–n/a, doi:10.1029/2011MS00045, 2011.
- Lawrence, M. G.: The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications, *Bull. Am. Meteorol. Soc.*, 86(2), 225–233, doi:10.1175/BAMS-86-2-225, 2005.
- 30 Lee, J. E., Oliveira, R. S., Dawson, T. E. and Fung, I.: Root functioning modifies seasonal climate, *Proc. Natl. Acad. Sci. U. S. A.*, 102(49), 17576–17581, doi:10.1073/pnas.0508785102, 2005.
- Lindsay, K., Bonan, G. B., Doney, S. C., Hoffman, F. M., Lawrence, D. M., Long, M. C., Mahowald, N. M., Keith Moore, J., Randerson, J. T. and Thornton, P. E.: Preindustrial-control and twentieth-century carbon cycle experiments with the earth system model CESM1(BGC), *J. Clim.*, 27(24), 8981–9005, doi:10.1175/JCLI-D-12-00565.1, 2014.



- Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., Manalo-Smith, N. and Wong, T.: Toward optimal closure of the Earth's top-of-atmosphere radiation budget, *J. Clim.*, 22(3), 748–766, doi:10.1175/2008JCLI2637.1, 2009.
- Lorenz, R. and Pitman, A. J.: Effect of land–atmosphere coupling strength on impacts from Amazonian deforestation, *Geophys. Res. Lett.*, 41(16), 5987–5995, doi:10.1002/2014GL061017, 2014.
- 5 May, W., Meier, A., Rummukainen, M., Berg, A., Chéruy, F. and Hagemann, S.: Contributions of soil moisture interactions to climate change in the tropics in the GLACE–CMIP5 experiment, *Clim. Dyn.*, 45(11), 1–23, doi:10.1007/s00382-015-2538-9, 2015.
- Mecklenburg, S., Drusch, M., Kerr, Y. H., Font, J., Martin-Neira, M., Delwart, S., Buenadicha, G., Reul, N., Daganzo-
10 Eusebio, E., Oliva, R. and Crapolicchio, R.: ESA's Soil Moisture and Ocean Salinity Mission: Mission performance and operations, *Geosci. Remote Sensing, IEEE Trans.*, 50(5), 1354–1366, doi:10.1109/TGRS.2012.2187666, 2012.
- Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M., Crous, K. Y., De Angelis, P., Freeman, M. and Wingate, L.: Reconciling the optimal and empirical approaches to modelling stomatal conductance, *Glob. Chang. Biol.*, 17(6), 2134–2144, doi:10.1111/j.1365-2486.2010.02375.x, 2011.
- 15 Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., Gettelman, A., Lawrence, D. M., Sanderson, B. M. and Strand, W. G.: Climate change projections in CESM1(CAM5) compared to CCSM4, *J. Clim.*, 26(17), 6287–6308, doi:10.1175/JCLI-D-12-00572.1, 2013.
- Mei, R. and Wang, G.: Impact of sea surface temperature and soil moisture on summer precipitation in the United States based on observational data, *J. Hydrometeorol.*, 12(5), 1086–1099, doi:10.1175/2011JHM1312.1, 2011.
- 20 Mei, R. and Wang, G.: Summer land–atmosphere coupling strength in the United States: Comparison among observations, reanalysis data, and numerical models, *J. Hydrometeorol.*, 13(3), 1010–1022, doi:10.1175/JHM-D-11-075.1, 2012.
- Miralles, D. G., van den Berg, M. J., Teuling, A. J. and de Jeu, R. A. M.: Soil moisture–temperature coupling: A multiscale observational analysis, *Geophys. Res. Lett.*, 39(21), L21707, doi:10.1029/2012GL053703, 2012.
- Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C. and de Arellano, J.: Mega-heatwave temperatures due to combined
25 soil desiccation and atmospheric heat accumulation, *Nat. Geosci.*, 7(5), 345–349, doi:10.1038/ngeo2141, 2014.
- Neale, R. B., Richter, J., Park, S., Lauritzen, P. H., Vavrus, S. J., Rasch, P. J. and Zhang, M.: The mean climate of the Community Atmosphere Model (CAM4) in forced SST and fully coupled experiments, *J. Clim.*, 26(14), 5150–5168, doi:10.1175/JCLI-D-12-00236.1, 2013.
- Notaro, M.: Statistical identification of global hot spots in soil moisture feedbacks among IPCC AR4 models, *J. Geophys. Res. Atmos.*, 113(D9), doi:10.1029/2007JD009199, 2008.
- 30 Orłowsky, B. and Seneviratne, S. I.: Statistical analyses of land–atmosphere feedbacks and their possible pitfalls, *J. Clim.*, 23(14), 3918–3932, doi:10.1175/2010JCLI3366.1, 2010.



- Panciera, R., Walker, J. P., Jackson, T. J., Gray, D. A., Tanase, M. A., Ryu, D., Moneris, A., Yardley, H., Rudiger, C., Wu, X., Gao, Y. and Hacker, J. M.: The Soil Moisture Active Passive Experiments (SMAPEX): Toward soil moisture retrieval from the SMAP mission, *Geosci. Remote Sensing, IEEE Trans.*, 52(1), 490–507, doi:10.1109/TGRS.2013.2241774, 2014.
- Phillips, T. J. and Klein, S. A.: Land–atmosphere coupling manifested in warm-season observations on the U.S. southern great plains, *J. Geophys. Res. Atmos.*, 119(2), 509–528, doi:10.1002/2013JD020492, 2014.
- Rahman, M., Sulis, M. and Kollet, S. J.: The concept of dual-boundary forcing in land surface–subsurface interactions of the terrestrial hydrologic and energy cycles, *Water Resour. Res.*, 50(11), 8531–8548, doi:10.1002/2014WR015738, 2014.
- Santanello, J. A., Peters-Lidard, C. D., Kumar, S. V., Alonge, C. and Tao, W.-K.: A modeling and observational framework for diagnosing local land–atmosphere coupling on diurnal time scales, *J. Hydrometeorol.*, 10(3), 577–599, doi:10.1175/2009JHM1066.1, 2009.
- Seager, R. and Hoerling, M.: Atmosphere and ocean origins of North American droughts, *J. Clim.*, 27(12), 4581–4606, doi:10.1175/JCLI-D-13-00329.1, 2014.
- Seneviratne, S. I., Luthi, D., Litschi, M. and Schar, C.: Land–atmosphere coupling and climate change in Europe, *Nature*, 443(7108), 205–209, doi:10.1038/nature05095, 2006.
- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B. and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, *Earth-Science Rev.*, 99(3–4), 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010.
- Seneviratne, S. I., Wilhelm, M., Stanelle, T., van den Hurk, B., Hagemann, S., Berg, A., Cheruy, F., Higgins, M. E., Meier, A., Brovkin, V., Claussen, M., Ducharne, A., Dufresne, J.-L., Findell, K. L., Ghattas, J., Lawrence, D. M., Malyshev, S., Rummukainen, M. and Smith, B.: Impact of soil moisture–climate feedbacks on CMIP5 projections: First results from the GLACE–CMIP5 experiment, *Geophys. Res. Lett.*, 40(19), 5212–5217, doi:10.1002/grl.50956, 2013.
- Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P. C. D., Caspersen, J. P., Sentman, L. T., Fisk, J. P., Wirth, C. and Crevoisier, C.: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, *Global Biogeochem. Cycles*, 23(2), doi:10.1029/2007GB003176, 2009.
- Susskind, J., Blaisdell, J. M. and Iredell, L.: Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm, *J. Appl. Remote Sens.*, 8(1), 84994, doi:10.1117/1.JRS.8.084994, 2014.
- Swenson, S. C.: GRACE MONTHLY LAND WATER MASS GRIDS NETCDF RELEASE 5.0. Ver. 5.0, Datasets, doi:10.5067/TELND-NC005, 2012.
- Swenson, S. C. and Lawrence, D. M.: Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data, *J. Geophys. Res. Atmos.*, 119(17), 10,210–299,312, doi:10.1002/2014JD022314, 2014.
- Swenson, S. C. and Lawrence, D. M.: A GRACE-based assessment of interannual groundwater dynamics in the Community Land Model, *Water Resour. Res.*, 51(11), 8817–8833, doi:10.1002/2015WR017582, 2015.



- Taylor, C. M., de Jeu, R. A. M., Guichard, F., Harris, P. P. and Dorigo, W. A.: Afternoon rain more likely over drier soils, *Nature*, 489(7416), 423–426, doi:10.1038/nature11377, 2012.
- Taylor, C. M., Birch, C. E., Parker, D. J., Dixon, N., Guichard, F., Nikulin, G. and Lister, G. M. S.: Modeling soil moisture–precipitation feedback in the Sahel: Importance of spatial scale versus convective parameterization, *Geophys. Res. Lett.*, 5 40(23), 6213–6218, doi:10.1002/2013GL058511, 2013.
- Teixeira, J., Waliser, D., Ferraro, R., Gleckler, P., Lee, T. and Potter, G.: Satellite observations for CMIP5: The genesis of Obs4MIPs, *Bull. Am. Meteorol. Soc.*, 95(9), 1329–1334, doi:10.1175/BAMS-D-12-00204.1, 2014.
- Wahr, J., Swenson, S., Zlotnicki, V. and Velicogna, I.: Time-variable gravity from GRACE: First results, *Geophys. Res. Lett.*, 31(11), doi:10.1029/2004GL019779, 2004.
- 10 Whan, K., Zscheischler, J., Orth, R., Shongwe, M., Rahimi, M., Asare, E. O. and Seneviratne, S. I.: Impact of soil moisture on extreme maximum temperatures in Europe, *Weather Clim. Extrem.*, 9, 57–67, doi:10.1016/j.wace.2015.05.001, 2015.
- Zeng, X., Barlage, M., Castro, C. and Fling, K.: Comparison of land–precipitation coupling strength using observations and models, *J. Hydrometeorol.*, 11(4), 979–994, doi:10.1175/2010JHM1226.1, 2010.



Table 1: Remote sensing and reanalysis products used for analysis.

Variable	Abbr.	Data product	Spatial resolution	Temporal resolution	Reference
Terrestrial water storage	TWS	GRACE Tellus RL05.1	1°	monthly	Landerer and Swenson (2012)
Vapor pressure deficit	VPD	AIRS AIR3XSTM v6	1°	monthly	Susskind et al. (2014)
Precipitation	PPT	GPCP v2.1	1°	daily	Huffman et al. (2009)
Downwelling shortwave radiation	SW↓	CERES EBAF Ed2.8	1°	monthly	Loeb et al. (2009)

Table 2: CMIP5 models used for analysis.

Model acronym	Atmospheric model	Land surface model	Horizontal resolution	Ens. size	Reference(s)
CCSM4	National Center for Atmospheric Research (NCAR) Community Atmospheric Model version 4 (CAM4)	Community Land Model (CLM4)	288 x 192	6	Lawrence et al. (2011); Neale et al. (2013)
CESM1-CAM5	NCAR Community Atmosphere Model version 5 (CAM5)	CLM4	288 x 192	3	Lawrence et al. (2011); Meehl et al. (2013)
CESM1-BGC	NCAR CAM4 with biogeochemistry	CLM4	288 x 192	1	Lawrence et al. (2011); Lindsay et al. (2014); Neale et al. (2013)
IPSL-CM5A-LR	Laboratoire de Météorologie Dynamique atmospheric model (LMDZ5A)	Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE)	96 x 96	3	Cheruy et al. (2013); Dufresne et al. (2013); Hourdin et al. (2013)
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory (GFDL) Earth System Model 2 (ESM2)	Land Model 3.0 (LM3.0)	144 x 90	1	Dunne et al. (2012); Shevliakova et al. (2009)
GFDL-ESM2G	GFDL ESM2	LM3.0	144 x 90	1	Dunne et al. (2012); Shevliakova et al. (2009)

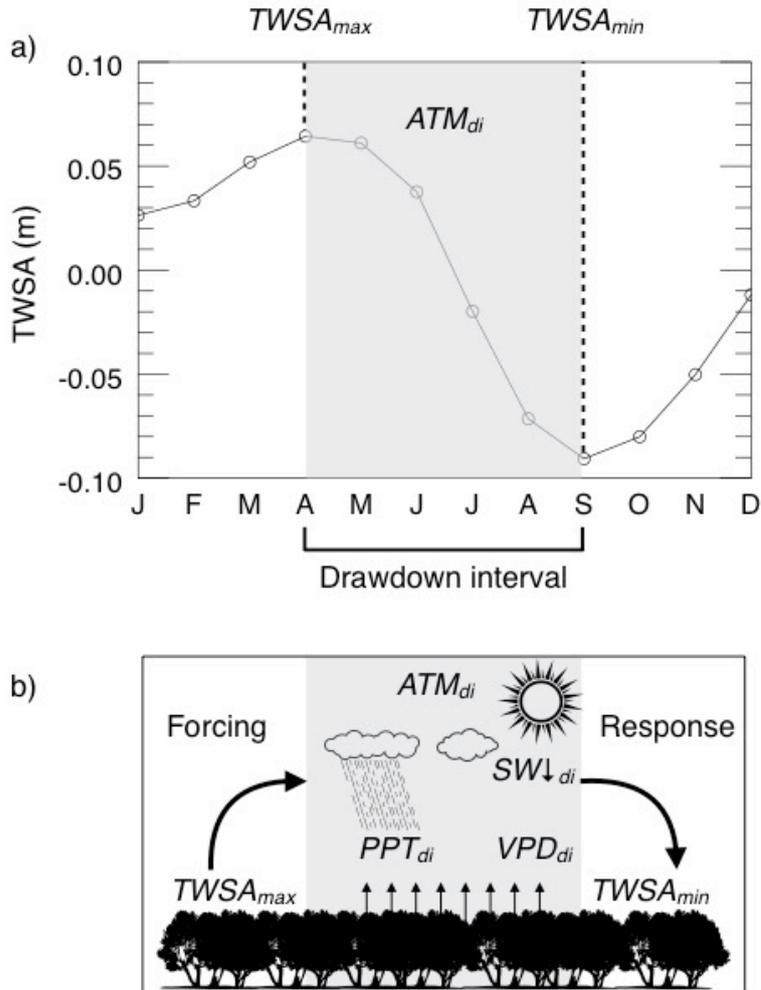


Figure 1: Conceptual description of feedback metrics: a) Example TWS climatology from a typical mid-latitude location in central North America (38° N, 92° W) illustrating the definition of the drawdown interval as the months from the maximum TWSA through the minimum TWSA. $TWSA_{max}$ and $TWSA_{min}$ are the TWSA values (in units of water height) during the maximum and minimum months respectively, and ATM_{di} is the atmospheric variable of interest averaged across the months of the drawdown interval. b) Representation of the interactions between TWS and atmospheric component, demonstrating the forcing limb of the feedback loop, in which $TWSA_{max}$ forces subsequent atmospheric conditions, as well as the response limb, in which $TWSA_{min}$ responds to the atmospheric state during the drawdown interval.

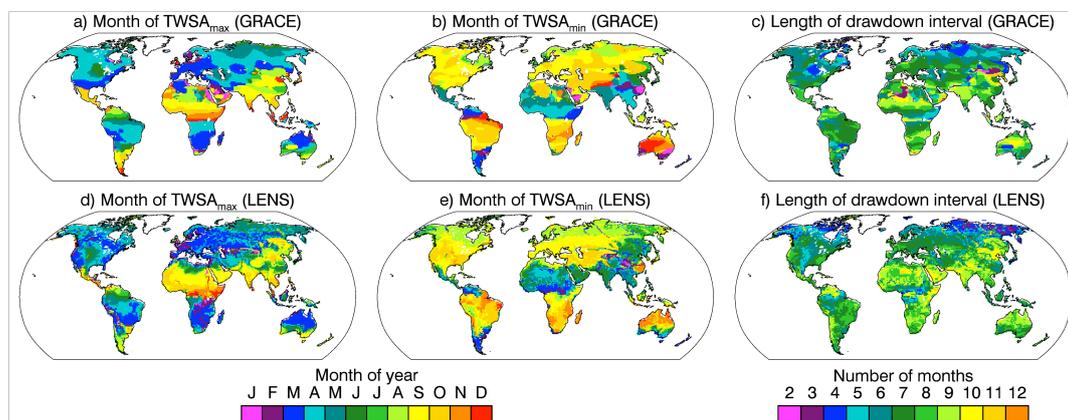


Figure 2: Month of maximum and minimum TWSA and the length of the drawdown interval from GRACE (a–c) and the LENS ensemble mean (d–f). Months of maximum and minimum were based on the climatology of detrended TWSA over the 157 months in the GRACE record.

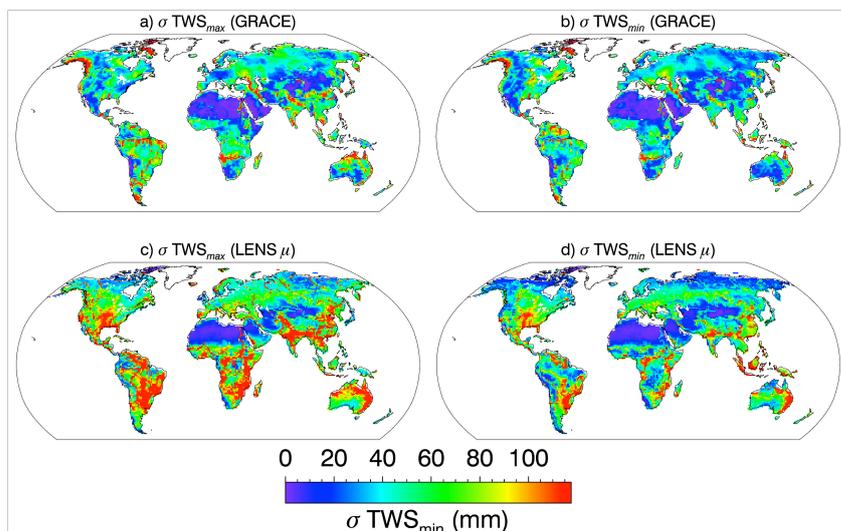


Figure 3: Interannual variability (standard deviation) of $TWSA_{max}$ and $TWSA_{min}$ from GRACE (a and b) and the LENS ensemble mean (c and d).

5

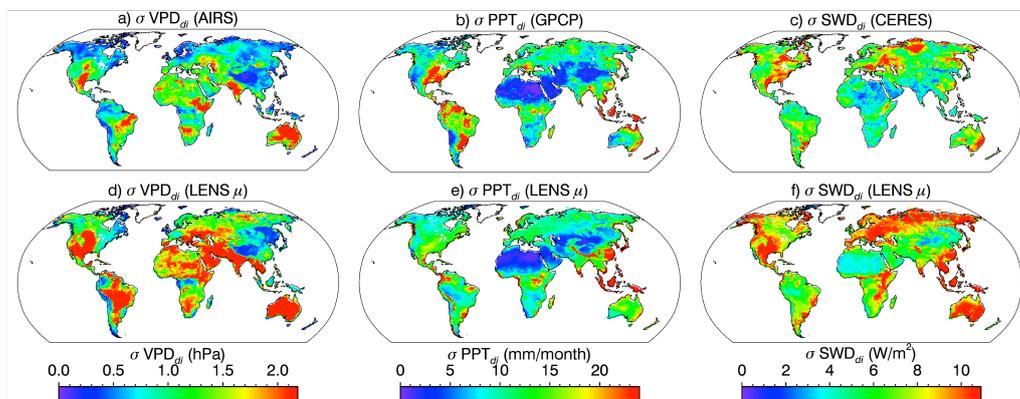
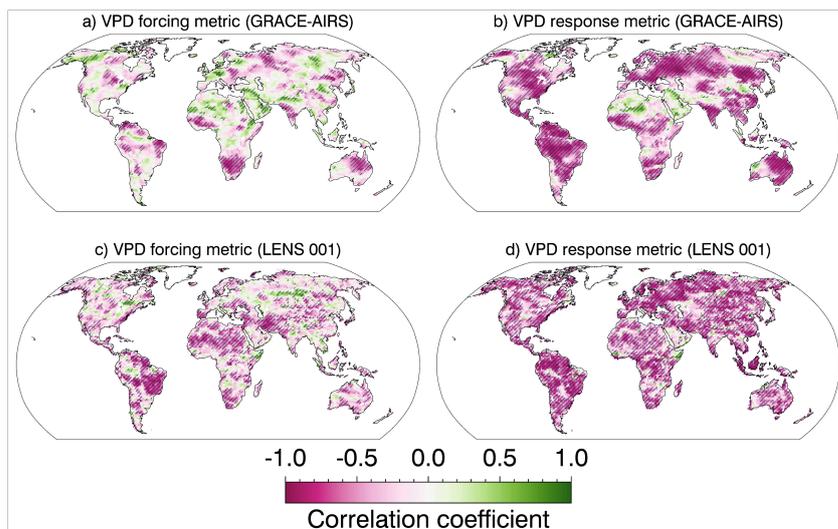


Figure 4: Interannual variability (standard deviation) of VPD_{di} from AIRS (a), PPT_{di} from GPCP (b), SWD_{di} from CERES (c) and the equivalent quantities from the LENS ensemble mean (d–f).



5 Figure 5: Forcing and response metrics for VPD from GRACE/AIRS (a and b) and LENS ensemble member 001 (c and d). Crosshatching indicates a correlation coefficient that is statistically significant at $p \leq 0.05$ (one-tailed Student's t-test).

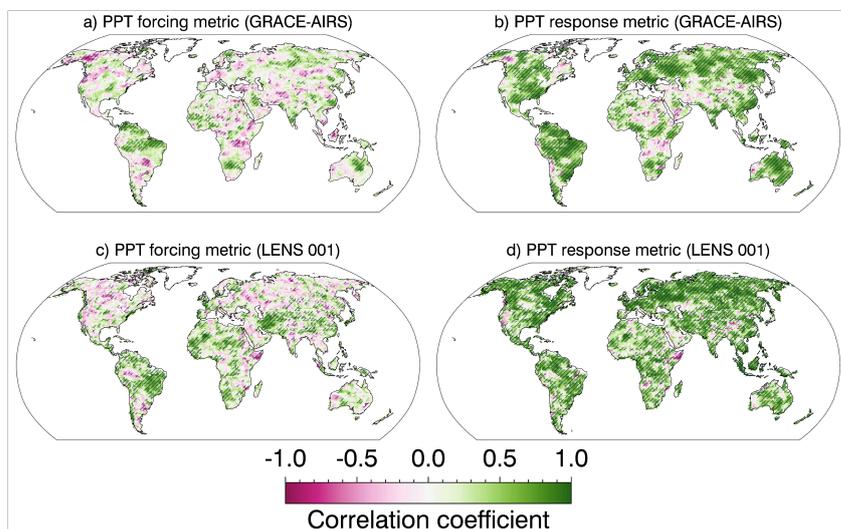
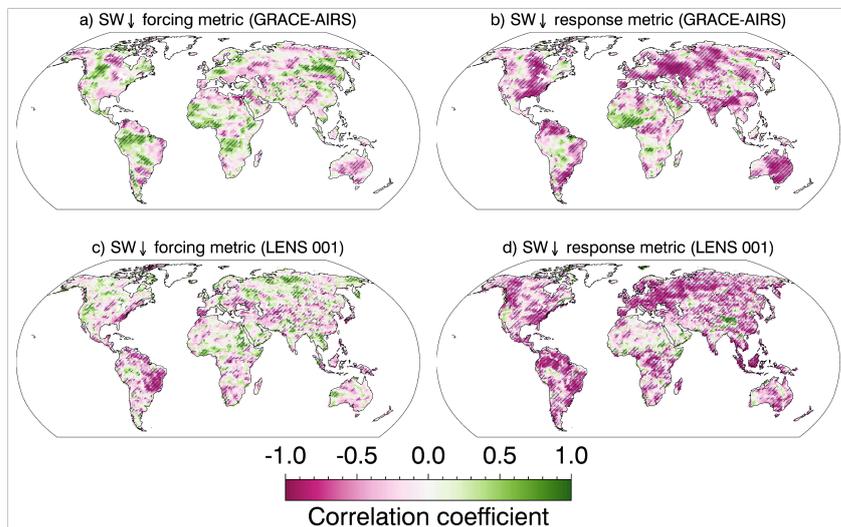


Figure 6: Forcing and response metrics for PPT from GRACE/GPCP (a and b) and LENS ensemble member 001 (c and d). Crosshatching indicates a correlation coefficient that is statistically significant at $p \leq 0.05$ (one-tailed Student's t-test).



5 **Figure 7:** Forcing and response metrics for SW↓ from GRACE/CERES (a and b) and LENS ensemble member 001 (c and d). Crosshatching indicates a correlation coefficient that is statistically significant at $p \leq 0.05$ (one-tailed Student's t-test).

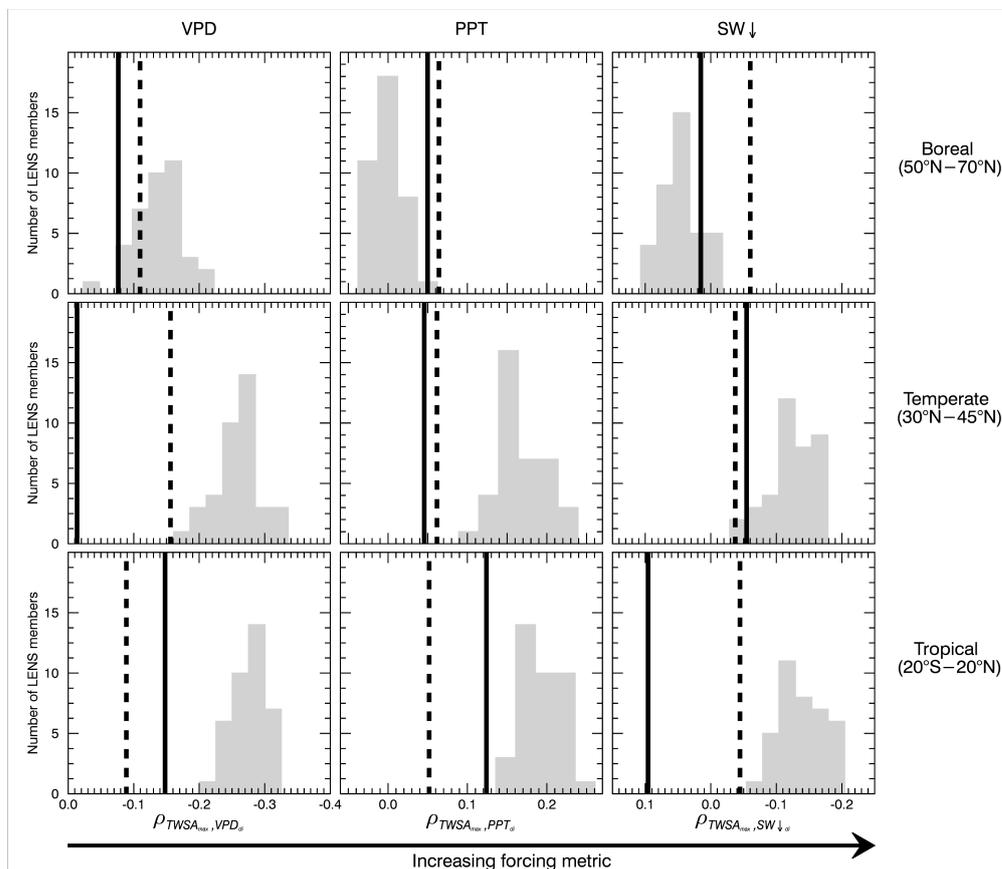


Figure 8: Ensemble histogram of forcing metrics from LENS (grey bars) with the satellite observations from GRACE/AIRS/GPCP/CERES (solid black line) and the alternate LENS and ERA-Interim (dashed black line), averaged across land regions within different latitude bands.

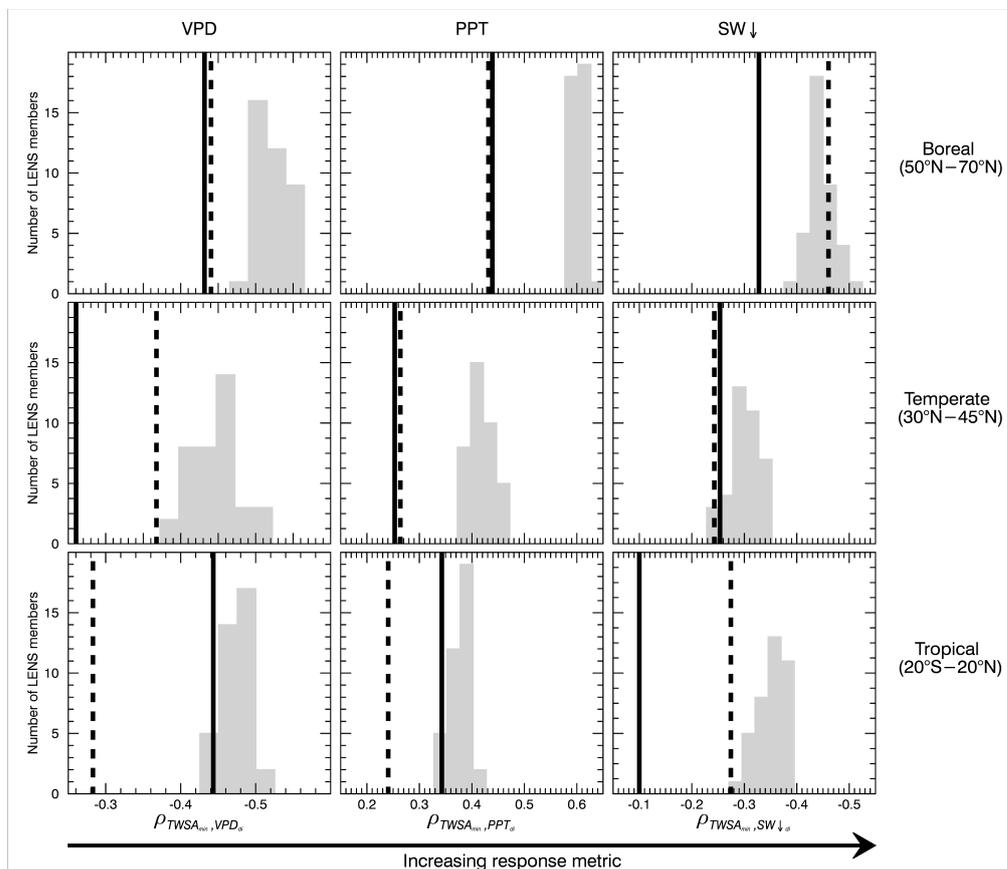


Figure 9: Ensemble histogram of response metrics from LENS (grey bars) with the satellite observations from GRACE/AIRS/GPCP/CERES (solid black line) and the alternate observations from GRACE and ERA-Interim (dashed black line), averaged across land regions within different latitude bands.

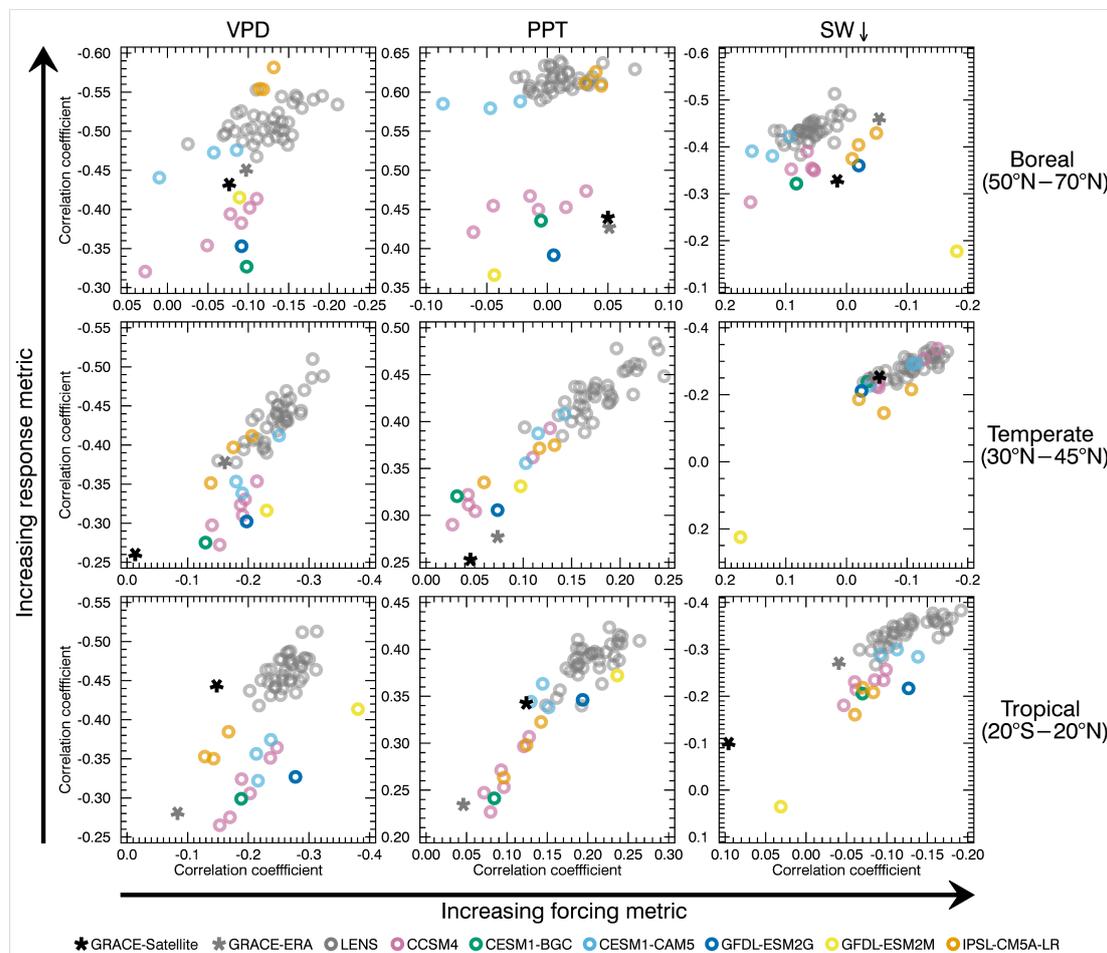


Figure 10: Scatter plots of forcing and response metrics for LENS and CMIP5 models with observations, averaged across land regions within different latitude bands.