

Interactive comment on “Evaluating the strength of the land–atmosphere moisture feedback in earth system models using satellite observation”

by P. A. Levine et al.

P. A. Levine et al.

plevine@uci.edu

Received and published: 10 August 2016

We thank the Referee 3 for their thorough and thoughtful review. The reviewer was very critical of several aspects of our paper, and identified issues that they believe warrant several important revisions. We agree with many of the points the reviewer makes, and we plan to make significant revisions to address these points, which we detail below in our response to the reviewer's comments. However, we disagree with the reviewer's perspective that these represent "fundamental issues" with the manuscript and the research presented in it. We believe that the research represents a meaningful and novel contribution to the field, as indicated by the other two referees who reviewed our manuscript. We believe that the major revisions that we plan to make, detailed

C1

below, will clarify our position and improve our acknowledgement of certain limitations and caveats. We hope these clarifications and revisions will satisfy the reviewer, but we would be happy to engage in an ongoing dialogue with this reviewer as well as the community at large if there are any unresolved issues.

Below, we address the reviewer's general and specific comments by quoting each comment in *italicized* font, providing our response in roman font, and quoting our proposed revisions as indented roman font.

Main comments:

1) My first main comment has to do with the metric used. First, the brevity of the time period of analysis is obviously an issue. The metrics are essentially interannual correlations over 13 years, i.e., correlations over 13 data points. That's very short. I am not sure I have seen many studies looking at interannual variability over 13 years only, in particular in terms of land-atmosphere studies. Accordingly, results for the feedback metrics appear very noisy. First, I believe a discussion of field significance is warranted here: patches of apparently significant values may still be random in that context (e.g., see Livezey and Chen (1983)).

Second, note that recent research underscores the need for long-record datasets to establish land-atmosphere coupling, that coupling metrics require more data than single-variable simple statistics (e.g., mean and variance) to be robustly estimated, and finally that, unlike single-variable statistics, coupling metrics are actually degraded by observational uncertainty (Findell et al. 2015). The latter point, in particular, is in my view a much likelier explanation for the weaker correlations found here in observations – between uncertain observation datasets that are independent of each other – compared to correlations computed with model outputs, which are by definition perfectly consistent with each other. The authors touch on the issue of observational uncertainty by computing correlations with the ERA reanalysis, but I don't think enough is made of that.

C2

So, given the brevity and uncertainty of observations, even without consideration of any other issues (but, see below), I am really uncomfortable with the approach proposed by this paper, which is to consider the observational estimate as a benchmark for model evaluation. Personally, I think an approach where observations and model results are used together to try to infer the “real” coupling would make more sense here. But, this would lead to a very different paper. Overall, if the authors are going to go on with their approach, I would recommend much more caution in how things are presented, including in the title of the study and the conclusions.

The reviewer makes several excellent points, and we are appreciative of the thoughtful perspective. We acknowledge that several of these points warrant revisions and additions to the text (detailed below) in order to clarify the goals of our approach and more appropriately emphasize its limitations. However, we believe that despite these limitations, our approach still represents a valuable and novel system for evaluating model performance using observational constraints.

We recognize that the relatively short time series available from the satellite record warrants caution while interpreting these results. We agree with the reviewer that the findings of Findell et al. (2015) emphasize this limitation and suggest that observational uncertainty would be expected to decrease correlations. As such, we plan to discuss this in Section 4.3 with the following addition:

One important factor contributing toward stronger feedback metrics in models relative to observations is the effect of observational uncertainty combined with a relatively short time series. Adding error to one or more variables in a correlation analysis will reduce the correlation coefficient, and this degradation has been shown to be sensitive to the length of data sets used to establish metrics of land–atmosphere interactions (Findell et al., 2015). Given the relatively short time series available for the current analysis, the correlation coefficients from remote sensing data may be reduced due to

C3

observational uncertainty, unlike those derived from internally-consistent models. We obtained a qualitative estimate of the influence of observational uncertainty on derived feedback metrics by replacing the atmospheric remote sensing data with reanalysis data from ERA-Interim. We found that both sets of observationally based metrics were weaker than those from LENS and several other models, suggesting that some of the overestimated feedback metrics in models may not be fully explained by observational uncertainty.

This acknowledgement further supports our argument that the utility of these metrics will increase as the time series of global satellite data grows longer with continuations of current missions and initiation of new missions (i.e., GRACE follow on) as mentioned in Section 4.5:

Furthermore, we acknowledge that observational error over an insufficiently long time series could reduce the apparent strength of correlations (Findell et al., 2015). Therefore, 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.

We do not believe the issue of field significance is relevant in the current context. Our metrics compare a single time series of TWS anomalies with a single time series of atmospheric data in the same region. We are not calculating correlations between a single explanatory variable and a geographically distributed field of dependent variables. Therefore, we are not engaging in the type of hypothesis testing that would warrant consideration of field significance.

C4

We agree with the reviewer that “an approach where observations and model results are used together to try to infer the “real” coupling” would be valuable, and represents a research priority for the community. Our intention was not to present the observationally derived forcing metric as representing the “real” land–atmosphere coupling strength. Instead, it represents the combined effects of land–atmosphere coupling (the “real” coupling strength) along with the remote effects of SST forcing on both the atmosphere and land surface. We believe that despite the relatively short time series, these metrics provide a useful constraint on models’ ability to represent this combined set of processes. The reviewer’s recommendation of greater caution in our presentation is appreciated, and we plan to add the following clarification to section 2.2 when introducing the metrics:

We note that while these metrics provide information about land–atmosphere coupling as a forcing mechanism on the atmosphere and the response of the land surface to the atmosphere, they are potentially influenced by atmospheric and soil moisture persistence, as well as remote forcing from sea surface temperatures (SST) (Orlowsky and Seneviratne, 2010; Mei and Wang, 2011). Nevertheless, these metrics still provide useful benchmarks against which to evaluate the ability of ESMs to reproduce the proper relationships based on the combination of these factors.

We plan to further clarify this in a major revision to the first paragraph of Section 4.1

The metrics developed here from satellite observations provide a means for evaluating land–atmosphere feedback strength on seasonal to interannual timescales in coupled ESMs. The use of correlation coefficients in this study does not enable a direct assessment of whether the relationships are directly causal, as correlation between atmospheric and terrestrial conditions could result from atmospheric persistence and remote forcing from

C5

SST (Orlowsky and Seneviratne, 2010; Mei and Wang, 2011). Nonetheless, the satellite-derived metrics provide a meaningful constraint against which coupled models can be benchmarked, as these models need to correctly represent the combined effects of persistence, remote SST forcing, and land–atmosphere coupling.

We also plan to emphasize the importance of disentangling the influence of land–atmosphere coupling from that of atmospheric persistence and remote SST forcing with the following addition to Section 4.5:

Finally, the issue of causality and the possibility that correlations result primarily from atmospheric persistence and remote forcing from SST rather than land–atmosphere interactions may be addressed using sensitivity experiments similar to those of GLACE and GLACE-CMIP. While the previous experiments have tested the importance of soil moisture interaction with the atmosphere, additional experiments could expand upon these methods by treating SST variability similar to terrestrial soil moisture availability. Such experiments could determine the relative importance of remote SST, including the effect of atmospheric persistence, and local land–atmosphere coupling in explaining correlations between TWS and atmospheric conditions.

We believe that despite the limitations of a relatively short time series and the inability to attribute the sources of covariability, our approach is still valuable. We believe that the revisions described above emphasize our goal of conceptually illustrating an approach towards model benchmarking that will become increasingly useful with longer time series from remote sensing. At this point, we would prefer to retain our title, which we believe is succinct and accurately conveys the content of our paper.

C6

2) Another issue with the metrics involves the definition of the (feedback) metric. The way it is defined, it is looking at the impact of TWS at the end or peak of the rainy season on climate in the subsequent months. The authors indicate as much, and say they want to consider, in the Tropics, the impact of late-rainy season TWS on dry-season climate. I see two issues with that. First, in my view, while that may be useful in the deep Tropics where the dry season is short, this approach is problematic in monsoon regions, or regions of the Tropics that have a well defined rainy season (i.e., outside of the deep Tropics): basically, after the rainy season, there is not much rain to look at any more. For instance, over the Sahel, what the authors are computing is the impact of September TWS on precipitation over September-May. But it doesn't rain much over that time period.

We believe that focusing on the drawdown interval is an important part of our approach. Our algorithm is novel in allowing a global-scale analysis across ecosystems. In mid-latitudes, the drawdown interval contains the summer season that has been the focus of research in land-atmosphere coupling. In tropical latitudes, the drawdown interval contains the dry season, during which precipitation recycling is important for maintaining ecosystems, allowing forests to persist in the absence of circulation-driven precipitation [Keys et al., 2016]. In the example of the Sahel, our algorithm is working as intended, by measuring how variations in TWS at the onset of the dry season are related to atmospheric conditions during the dry season.

In my view, it would be much more interesting to look at the impact of end-of-dry season TWS on the subsequent rainy season to see if, in these regions, available land moisture feeds back on precipitation during the rainy season. Second, in the same example over West Africa, whatever rainfall there is over Sept-May is actually probably the end of the monsoon, Sept-Nov. Because TWS in September is likely to be influenced by precip in September, and Sept. precip is likely to represent large part of the 'response' variable, the causation is muddled a little bit: a clearer temporal offset would be needed in such a case. But more importantly, even precip in the months following September (Oct,

C7

Nov) is likely to be correlated with precip in the previous months – for instance, a year with a strong monsoon that has more rain in Jun-Sept may well tend to also have more rain in Sept-Nov. Because September TWS will largely reflect JJAS rainfall, the TWS-based metric will then show a strong feedback - but the inferred causation would be a misinterpretation.

We believe that ET in the wet season tropics would be energy limited, and therefore would not reflect the influence of land surface moisture availability on the atmosphere. We acknowledge the issue with persistence, which we expand upon below.

This brings me to a more general point: the authors do not discuss how autocorrelation, here at the seasonal time scale, of climate variables, may impact their estimate of land-atmosphere coupling. This is a major issue affecting all empirical studies of land-atmosphere coupling – see, for instance, Wei et al. (2008) and Orlowsky and Seneviratne (2010). The authors do cite the latter study, but, it seems, simply to say that if models underestimate SSTs influence on land climate, they will then appear to overestimate local L-A coupling. They somehow miss the point of that paper in how it should apply to their own results. I just gave one example above (the Sahel) of how that might be the case. Other monsoon regions (e.g., India) might similarly be affected. Interannual variability in the coupled ocean-atmosphere (e.g., ENSO) might also be the source of confounding influence at the time scales investigated here. So, overall, I recommend these caveats be considered and discussed by the authors in their interpretation of their results. Personally, I would need to see some further analysis to be more convinced of the physical reality of the land-atmosphere feedbacks the authors claim to show (e.g., some sensitivity test to the months and time lags considered, some investigation of atmospheric variability and persistence, etc.).

The reviewer's point is well taken, and has already been partially addressed above in the additions to Section 2.2 and Section 4.5 (quoted above). In addition, we plan to discuss these issues more explicitly in a revision of Section 4.3 to include the following:

C8

Another possible explanation stems from the fact that our feedback metrics include the influence of both direct interactions between the land-surface and the atmosphere as well as indirect covariability due to atmospheric persistence and remote forcing by SST (Orlowsky and Seneviratne, 2010; Mei and Wang, 2011). For this reason, we caution that overestimates of feedback metrics do not imply that the land-atmosphere feedback is necessarily stronger, but could be due to an overestimate of SST-driven correlations between the land surface and the atmosphere. Wei et al. (2008) demonstrated that negative correlations between soil moisture and subsequent precipitation can be explained by precipitation persistence combined with negative temporal autocorrelation of precipitation associated with sub-seasonal modes such as the Madden-Julian Oscillation (MJO). Poor representation of the MJO period in CMIP5 models leads to unrealistic patterns of precipitation persistence (Hung et al, 2013). If models are failing to capture MJO-driven negative correlations, this could lead to overly strong positive correlations relative to observations. However, this would depend on the length of the drawdown season relative to persistence time and the period of intra-seasonal modes.

This supports our planned addition to Section 4.5 (quoted above), discussing the importance of modeling experiments to determine relative importance of remote SST forcing, including the effect of atmospheric persistence, and local land-atmosphere coupling in explaining correlations between TWS and atmospheric conditions.

3) Another main comment has to do with the discussion section. The authors discuss why models might exhibit stronger feedback (and forcing) metrics than observations. As mentioned above, I think uncertainty in observations should be mentioned as a primary reason.

We now more explicitly address observational uncertainty as well as uncertainty due

C9

to the short time series in Section 4.5 quoted above.

The authors propose that ET may be consistently overestimated in climate models, and a large part of the discussion then consists in speculation as to why that may be the case. First, while I appreciate the effort to discuss things further and not just show results, I found this whole section a bit too speculative. IF the models overestimate ET, then IF stomatal conductance, IF convection, IF bare soil, etc. . . Can the authors actually point to any evidence that ET is consistently overestimated in climate models, in the first place (or at least in CESM)?

These points are well taken, and while this section is speculative by its very nature, we agree that it warrants revision. We plan to modify the discussion so that it does not center on models overestimating ET, but instead focuses on ways in which models could make moisture too readily available for ET. We plan to clarify the basis of our argument with the following revision to Section 4.3:

A set of possible explanations involves models overestimating the amount of water available for ET during the drawdown interval. The land surface influence on the atmosphere requires water to be a limiting factor to ET but not limiting enough to prevent it altogether. Under more moisture-limited conditions, a drawdown interval may experience multiple shorter time periods during which ET is inhibited due to insufficient water, and the terrestrial moisture state exerts no control over flux partitioning. These periods of insufficient moisture would tend to reduce the overall feedback strength integrated across the duration of the drawdown interval. Model shortcomings that make water too readily available for ET could reduce the amount of time spent in periods of insufficient moisture during the drawdown interval, thereby unrealistically strengthening the longer-term feedback. We note that the opposite could take place under near-saturated conditions if a model overestimates the amount of time in which ET is energy-limited, but

C10

we would not expect these conditions to be as prevalent during the drawdown interval that was the time period of focus in our analysis.

We also plan to add further discussion to Section 4.5 citing evidence of models overestimating ET:

CMIP5 models are known to have a high ET bias (Mueller and Seneviratne, 2014), which could be due in part to the explanations proposed as possible reasons for overestimated feedback metrics in models.

Second, if soil water is too readily available in models, and ET is overestimated, wouldn't that actually mean that feedbacks should be underestimated in models? Indeed, ET would then be less water-limited and more energy-limited, with less potential for soil moisture-atmosphere feedbacks.

We designed our metrics around the drawdown interval in order to specifically consider the time of year during which ET would be water-limited. The issues we discuss with insufficient representation of bare soil processes and big leaf parameterizations would, during this time of year, unrealistically make too much water available for ET. This would allow ET to take place in the model when in reality that water would have run off or is unavailable for transpiration. Under these conditions, the atmosphere in the model would be influenced by moisture availability when in reality no moisture is available. These points will be clarified with the revision to Section 4.3 quoted above.

Surface climate variability would then be influenced by the atmosphere to a greater extent. Along the same lines, the authors claim that their results, showing an overestimation of land-atmosphere feedback by models, are consistent with prior studies, and have implications for projected warming (e.g., Cheruy et al., 2014). However, these previous studies, it seems to me, point to ET being underestimated in these models,

C11

and models getting to easily “locked” in a dry and warm surface mode. So, in effect, while the authors agree with prior studies that land-atmosphere feedbacks are overestimated in models, they provide opposite reasons for that (overestimated versus underestimated of ET). I would like to see the authors clarify that apparent contradiction.

We plan to modify our discussion, described above, that clarifies our point so as not to rely on whether models overestimate or underestimate ET.

4) Finally, the author interpret the relationship they find between the strength of the feedback and forcing metrics in CMIP5 models as showing that: “the response limb of the feedback loop is important for understanding how conditions are set up for subsequent forcing via land-atmosphere coupling”. They claim that it highlights “the importance of the land surface response in priming the system for subsequent forcing on the atmosphere”. I am not convinced by this interpretation, which sounds a bit hand-wavy to me. I don’t see a strong physical reason why a model where, for instance, TWS responds strongly to precipitation, should have a strong feedback of TWS onto precipitation.

Conceptually, we disagree with this perspective. The strength of land-atmosphere coupling depends on moisture availability enabling some ET while still limiting it. Models must therefore simulate the correct moisture availability in order to simulate the proper amount of land-atmosphere coupling. The response metrics reflect whether models simulate the right moisture availability based on precipitation and evaporative demand, and whether this is the right amount to set up subsequent land-atmosphere coupling. We plan to clarify our reasoning with the following modification to Section 4.1:

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. The generally higher correlation coefficients in observed response metrics indicates the importance of the land surface re-

C12

sponse in priming the system for subsequent forcing on the atmosphere. For example, if TWS response too strongly coupled to atmospheric forcing, a small change in atmospheric conditions could yield an unrealistically large change in TWS. The unrealistically large TWS anomaly would have the potential to impart a larger land surface forcing of the atmosphere in subsequent time steps. That models and ensemble members with high forcing metrics were also generally found to have high response metrics (Figure 10) highlights the need to consider this.

Couldn't the relationship on Figure 10 be due to intermodel differences in what TWS (or its estimate, here) encompasses in each model? For instance, different soil depths? A deeper soil would lead to weaker links between TWS and climate both in terms of response and feedback to the atmosphere.

No, because we are using the total terrestrial water storage column. In the case of LENS, this is an explicitly output field that includes this entire column. In the case of the CMIP5 output, we used the accumulated residuals of the surface water balance (i.e., the integral of precipitation minus evaporation and runoff) to approximate this quantity.

In any case, I found Figure 10 to be insufficiently explained and encourage the authors to discuss this further.

We plan to add the following to the figure caption in order to clarify how TWS was determined for each model.

For CMIP5 models, we estimated TWSA using the accumulated residuals of the surface water balance. For LENS, TWSA values were internally calculated from water masses in soils and other terrestrial reservoirs

Specific comments:

C13

- p.2 line 4: "cloud radiative coupling": please explain and clarify.

We plan to clarify the text as follows:

Cloud radiative coupling can likewise lead to positive or negative feedbacks, as enhanced (suppressed) cloud formation decreases (increases) insolation and evaporative demand (Betts, 2009; Cheruy et al., 2014).

- P.2 line 24: actually, no: a surprising result of GLACE II was that predictive skill was not enhanced over the Great Plains "hot spot" from GLACE I, but rather to the North of it (see Koster et al. 2011). Consider rephrasing.

We thank the reviewer for pointing out this discrepancy, which plan to correct by removing the reference to GLACE II

- P.3 line 5: the text should make it clear that GLACE-like metrics cannot be directly compared to observations, and that other more simple metrics, not strictly equivalent, have to be used, like SM-ET correlations, etc.

We agree that this warrants clarification, and we plan to modify the text as follows:

GLACE metrics are based on model experiments with no direct observational equivalents. However, correlation based metrics that do enable direct comparison with observations suggest that models may overestimate land-atmosphere coupling strength (Dirmeyer et al., 2006).

- P.3 line 20: Findell et al. 2011 is actually based on reanalysis data, not "modeling". Also, Findell et al. 2015 should be included in this discussion, to highlight the issue, discussed above, of data length requirements to estimate land-atmosphere coupling.

We reference Findell et al. (2011) in the context of Guillod et al. (2014), which emphasizes that the surface state and fluxes are still model based, even if the atmosphere

C14

is constrained by some observations. However, to avoid confusion, we now omit the word “modeling” from the description of Findell et al. (2011). Furthermore, as discussed above, we now include Findell et al. (2015) in the discussion in Section 4.3 (quoted above).

- P.5 line 15: is that version of the GRACE data downscaled in any way, and if so, how? I thought the basic GRACE data was at coarse resolution (e.g., 500km).

The GRACE gridded land product that we use is provided at a 1-degree resolution. We plan to clarify this in the methods section by rewording the beginning of Section 2.1 Remote sensing data as follows:

We obtained Level-3 TWSA data from GRACE using the University of Texas at Austin Center for Space Research (CSR) spherical harmonic solutions (Swenson, 2012). Global land data at a 1° resolution were scaled using the coefficients provided by Landerer and Swenson (2012).

- P.5 line 17: “the TWS time series”. I read that GRACE data are actually anomalies compared to the mean over 2004-2009. How is that accounted for in the computation of the metrics? Are the other variables centered on the same years? Does that affect results in any way? What about model outputs?

In the context of our metrics, the baseline against which GRACE Anomalies are compared is arbitrary. In our calculations, the baseline only affects the intercepts of the linear correlations; it does not affect the correlation coefficients that comprise our metrics.

- P.6 line 32: see main comment above: I am not sure this is the most relevant time of year to investigate, and they are issues of rainfall autocorrelations.

We have addressed this in our response to the reviewer's main comment above, in which we explain why we chose to focus on this time of year.

C15

- P.7 lines 12-15: that is, if the feedback is actually a positive moisture feedback. In other words, the authors adopt the a priori view that they are looking at a positive, moisture recycling feedback. This should be stated more explicitly, and perhaps earlier in the manuscript.

This point is well taken, and we plan to modify the Methods section to more explicitly state this assumption by removing lines 12–15 and 19–20 on page 7, and replacing them with the following:

Here we note that our evaluation of both the forcing and response metrics will follow a nomenclature that considers strong coupling as acting in the direction of an overall positive feedback loop. In regions with a strong positive feedback, higher than average TWS would be followed by lower than average VPD, as more available water is able to fulfill evaporative demand. Therefore, strong TWS forcing on VPD would be associated with a negative correlation coefficient. Higher VPD during the drawdown interval would increase evaporative demand, potentially leading to a lower TWS anomaly, therefore a strong response of the land surface to VPD would also be associated with a negative correlation coefficient.

Because the partitioning of surface fluxes can, depending on the spatiotemporal scale, cause a change of either sign to both cloudiness and precipitation (Taylor et al., 2012; Guillod et al., 2015), correlation coefficients of either sign could indicate strong land surface forcing on PPT and SW_{\downarrow} . However, the response metrics would be expected to show greater consistency. Higher PPT during the drawdown interval would be expected to increase TWS (positive correlation), while higher SW_{\downarrow} would increase evaporative demand, thereby decreasing TWS (negative correlation). Therefore, to maintain consistent nomenclature based on evaluating the strength of a positive moisture feedback, we consider strong coupling in both the forc-

C16

ing and response metrics to be associated with a positive correlation in the case of PPT and a negative correlation in the case of SW \downarrow .

- *P.8 line 18: what about AMIP simulations?*

Correlations in our forcing metric come from both land-atmosphere coupling and the effects of remote SST forcing. AMIP simulations could reduce internal variability, but will not capture ocean-atmosphere interactions. We are interested in evaluating fully coupled models that are used for 21st century projections.

- *P.8 line 28: It's unclear to me why the authors restrict themselves to the GLACE-CMIP5 models. There is no further comparison in the manuscript, on a model-by-model basis, with results from that experiment. So why not use the whole CMIP5 ensemble?*

The purpose of this manuscript is not to evaluate the entire CMIP5 ensemble, but rather to introduce a new approach toward model benchmarking using a small ensemble of models as an example.

- *P.9 line 9: so what? What is made of that? What are the implications for the correlation-based metrics? This comment applies to the whole sub-section, actually, including the result about climate variability. If anything, higher variability in model outputs would point to lower correlations, if the covariance between TWS and climate is similar.*

This point is well taken, and we address this with an addition to this section that clarifies the implications of this analysis:

Comparing both the timing of TWS dynamics and the interannual variability of TWS and the atmospheric variables between the observations and model output provides context for interpreting the correlation-based metrics we present next. Although there are some inconsistencies, as noted

C17

above, the model largely reproduces the same patterns evident in the remote sensing data. In most regions, interannual variability in model output is within an order of magnitude of the observed variability, indicating that CESM can reasonably simulate the baseline properties (timing and variability) that influence the feedback dynamics.

- *P.9 line 11: aren't trends removed from this data? Please clarify.*

Trends are removed only for the purpose of generating the annual climatology, as indicated in the methods section, are retained in the correlation analysis in order to capture decadal-scale variability that would represent a trend in the relatively short time series.

- *P.9 line 26: still, why would the covariance be positive?*

We plan to address this question by adding the following text to this section:

Positive correlations are unlikely to reflect direct land-atmosphere interactions. Instead, they demonstrate how remote SST forcing can, depending on lag times, lead to apparent negative relationships such as those demonstrated by Wei et al. (2008).

- *P.9 whole section 3.2: this whole subsection feels very descriptive. On the other hand, there is not much description of the processes themselves. This might feel obvious to the authors, but some further discussion of what the correlations mean physically, when describing the figures, may be welcome.*

The purpose of this section is to help the reader understand how the metrics work for a single ensemble member before we present the aggregated results for the whole ensemble. We feel this is a critical step in explaining how our metrics work and justifying their use in comparing models with observations.

C18

- P.10 line 13: the link with cloud cover and precipitation should be explicitly mentioned here.

The link is now explicitly mentioned:

This is consistent with coupling between cloud cover and terrestrial moisture being both positive and negative on shorter time scales, which sometimes yields negative coupling over shorter time scales (Taylor et al., 2012; Guillod et al., 2015).

- P. 11 lines 14-15: see main comment 4 above.

We have addressed this in the response to the reviewer's comment 4 above.

- P.11 line 23: "Discussion".

Thank you for pointing out this typographical error, which has been corrected.

- P.11 line 28: as mentioned above, these "well understood mechanisms" are actually never really explained.

The planned revision to Section 4.1, described above in the response to the first general comment, removes this phrase and more clearly describes what is being shown.

- P.12 lines 3-4: that's exaggerated. Feedback results on Figures 5-7 are very noisy, and even from a simply qualitative perspective, it is a stretch to say that they agree with results from GLACE 1. One could just as well point out all the regions on Figures 5-7 that do NOT show up in GLACE 1 and say results are completely different. Besides, I find it a bizarre impulse (or maybe, a testament to the strength of the GLACE 1 study) that every land-atmosphere study seemingly feels the need to point out some level of agreement with GLACE results, even when, as is the case here, the match is very weak at best, and more importantly, when different data (observations versus models), processes and spatio-temporal scales are considered. Consider removing

C19

that comparison.

The reviewer's point is well taken, and we do not believe the comparison with GLACE 1 is a necessary component of our discussion. As part of the major revision of the discussion section, as described above, we plan to remove this reference to GLACE.

- P.12 line 16: see main comment 4 above.

We have addressed this in the response to main comment 4 above.

- P.12 line 26: the authors could still look at this in models results, though. In fact, showing the link between TWS and ET, for instance, would reinforce their results and the physical interpretation that they propose.

This is limited by the lack of a global remote sensing ET data set spanning the study period.

- P. 13, first paragraph: this is unclear. Do the author mean that that the models underestimate remote influences of SSTs, for instance, and thus appear to have too strong a coupling?

Along side the other major revisions to our discussion section, we plan to remove this passage from the text and replace it with a clearer explanation, which we have quoted above in response to main comment 2.

- P.13 lines 16-18: see main comment 3 above.

This is addressed in major revisions to this section, quoted above in the response to main comment 3.

- P. 14 line 18: but here observations show positive coupling, too! Please clarify.

The observational metrics in this study include the effects of remote SST forcing as well as land-atmosphere interactions integrated across seasonal time scales. The negative soil moisture–precipitation coupling mechanism found from observations by Taylor

et al. (2013) would tend to reduce the overall positive correlation. If parameterized convection prevents models from correctly capturing this mechanism, then the correlations may be overestimated. We plan to clarify this with the following addition to the discussion:

If negative coupling mechanisms are present in reality but absent from models, this could contribute to an overestimate of feedback metrics and under-representation of negative feedbacks in models.

- *P. 14 line 21: but reduced stomatal opening would be associated with reduced ET, too. Please clarify.*

We plan to remove the discussion of stomatal opening, both for the sake of brevity and to prevent any confusion.

- *P. 14 lines 18-30: See main comment 2. There is a fundamental issue with the manuscript here.*

We plan to make major revisions to this section, as addressed above in the response to main comments 1 and 2. As we explained above, we do not feel that this represents a fundamental issue with our approach or our manuscript.

- *P.15 line 3: see main comment 3.*

We have addressed this above in our response to comment 3.

- *P.15 lines 8-9. not really: Seneviratne et al. (2013) show that long-term soil moisture change leads to more warming, differently across models in the GLACE-CMIP5. That, in and of itself, could be considered an estimate of (long-term) soil moisture-atmosphere coupling in these models; but, in any case, there is no comparison to estimates of present-day coupling.*

We agree that the linkage between our metrics and the results of GLACE-CMIP5 are

C21

too speculative. For this reason and for the sake of brevity, we plan to remove this portion of the discussion section.

- *P.15 lines 11: No. Warmer air “holding” more water vapor and leading to more precipitation would lead to positive temperature-precipitation correlations – not negative.*

We thank the reviewer for pointing out this erroneous characterization of the results of Berg et al. (2015). The phrase “in which higher air temperatures can hold more precipitable water” was intended to read “cloud cover variability drives precipitation and temperature in opposite directions” However, as mentioned in the response to the previous comment, we intend to remove this portion of the text from our discussion.

- *P.15 line 13: “determined”: not really. What Berg et al. (2015) show is that because of land-atmosphere interactions, the interannual negative temperature-precipitation relationship that they identify in present-day climate holds on longer time scales, including in the case of climate change. This may be interpreted as suggesting, as the authors say here, that models with too strong a coupling will then overestimate future warming; however, it is not directly shown by that study. Consider rephrasing.*

As mentioned in the response to the previous two comments, we plan to remove this portion of the text from our discussion.

- *P.16 line 10: see comment above on P.12*

As mentioned in the response to the comment above, we plan to remove the assertion that our observed metrics reflect the patterns found in GLACE. In addition to the revisions to the discussion section, mentioned above, we also plan to revise the conclusion section to remove this portion.

- *P.16 line 11: “regions of strong RESPONSE metrics”, I believe.*

We thank the reviewer for pointing out this typographical error

- *P.16 line 14: the implication is bit too implicit here. Consider being more explicit.*

C22

We plan to make this statement more explicit with the following revision:

Modeled feedback metrics are generally found to be stronger than those observed in the satellite record. If this discrepancy is due to models overestimating the two-way feedback between the land surface and the atmosphere, this could lead to models incorrectly projecting future warming trends and climatic extremes.

Figures:

- *Figure 1: nice figure that helps understand the study. The y-axis on a) refers to anomalies, I presume – see comment on GRACE values above.*

Correct, as we mentioned in response to the comment above. We plan to clarify this by replacing “TWS” with “TWSA” in the caption to Figure 1.

- *As noted above, Figure 3 and 4 are nice, but not much is made of them in the analysis.*

As described in response to the reviewer’s comment above, we plan to add some additional text to expand upon and clarify the purpose of these figures. As quoted above, we feel these figures are important for demonstrating that LENS is able to capture the baseline properties of our analysis (timing and variability) before presenting the correlation coefficients.

- *Figure 5-7: I suggest the authors modify the color legend here. More color shades is not always better. It is actually not easy to see differences in color shades on a continuous bi-color palette like here, and for the reader things essentially end up being two colors, one positive (green) and one negative (red). It would actually be easier to have fewer shades, more clearly separated, and with perhaps several different colors as well.*

After experimenting with several color and shading schemes, we determined that the

C23

spatial variability in Figures 5–7 are best illustrated using the employed color scheme combined with crosshatching to indicate statistically significant correlations.

- *Figure 8: I suggest showing the mean of the CESM distribution as well.*

We considered including this, but determined that it made the figures too busy without adding useful information.

References cited in this response:

Berg, A., Lintner, B. R., Findell, K., Seneviratne, S. I., van den Hurk, B., Ducharne, A., Chéry, F., Hagemann, S., Lawrence, D. M., Malyshev, S., Meier, A. and Gentine, 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., 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.

Cheruy, 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.

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.

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.

Findell, K. L., Gentine, P., Lintner, B. R. and Guillod, B. P.: Data length requirements

C24

for observational estimates of land–atmosphere coupling strength, *J. Hydrometeorol.*, 16(4), 1615–1635, doi: 10.1175/JHM-D-14-0131.1, 2015.

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.

Hung, M.-P., Lin, J.-L., Wang, W., Kim, D., Shinoda, T. and Weaver, S. J.: MJO and convectively coupled equatorial waves simulated by CMIP5 climate models, *J. Clim.*, 26(17), 6185–6214, doi: 10.1175/JCLI-D-12-00541.1, 2013.

Keys, P. W., Wang-Erlundsson, L., and Gordon, L. J.: Revealing invisible water: Moisture recycling as an ecosystem service, *PLoS ONE*, 11(3), 1–16, doi:10.1371/journal.pone.0151993, 2016.

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.

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.

Mueller, B., and Seneviratne, S. I.: Systematic land climate and evapotranspiration biases in CMIP5 simulations, *Geophys. Res. Lett.*, 41(1), 128–134, doi: 10.1002/2013GL058055, 2014.

Orlowsky, B. and Seneviratne, S. I.: Statistical analyses of land–atmosphere feedbacks

C25

and their possible pitfalls, *J. Clim.*, 23(14), 3918–3932, doi:10.1175/2010JCLI3366.1, 2010.

Swenson, S. C.: GRACE MONTHLY LAND WATER MASS GRIDS NETCDF RELEASE 5.0. Ver. 5.0, Datasets, doi:10.5067/TELND-NC005, 2012.

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.*, 40(23), 6213–6218, doi:10.1002/2013GL058511, 2013.

Wei, J., Dickinson, R. E. and Chen, H.: A negative soil moisture–precipitation relationship and its causes, *J. Hydrometeorol.*, 9(6), 1364–1376, doi: 10.1175/2008JHM955.1, 2008.

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, doi:10.5194/hess-2016-206, 2016.