RC1

Received and published: 11 March 2020

The paper investigates the skill of the land surface energy partitioning in the ERA5 reanalysis. For reference, the ERA5 skill is compared to that of its predecessor, ERA Interim (ERA-I). Skill is determined in several different ways: (1) directly vs flux tower measurements, (2) vs fluxes from water and energy balance estimates, (3) by driving the GLEAM land surface scheme with reanalysis data and validating the output vs. flux tower measurements, and, finally, (4) by driving the CLASS4GL boundary layer model with reanalysis data and validating the output vs. balloon observations. The authors find that ERA5 land surface energy partitioning is generally improved over that of ERA-I. In particular, the overestimation of the latent heat flux in ERA-I is reduced (but not eliminated) in ERA5.

The paper is of interest to GMD readers and makes an important contribution by documenting the quality of ERA5 land surface estimates. By and large, the writing and graphics are clear and concise, and the conclusions drawn from the study are supported by the results. I recommend eventual publication of the paper in GMD provided the authors address the comments below. It would be particularly helpful to include other reanalysis data in the comparison.

We thank the referee for reviewing the paper and providing us with useful comments, feedback, and corrections. Below, we give a point-to-point reply to the comments posted by the reviewer and list the changes that will be implemented in the revised version of the manuscript.

Major comments (no particular order):

1. The title and introduction do not make it sufficiently clear that the turbulent fluxes investigated here are for the land only. I suggest changing the title to: "Evaluating the *land* surface energy partitioning in ERA5" and occasionally replacing the expression "surface [..] fluxes" with "land surface [..] fluxes", e.g., P3/L15, P4/L3, and probably a few more places.
   The authors agree with the referee and will update the title.

   <u>Changes in manuscript:</u> the title of the manuscript will be updated per suggestion of the referee.

2. P4/L1: Aren't there advances in land data assimilation from ERA-I to ERA5? And do these not matter for the quality of the land surface turbulent flux estimates? Land data assimilation in ERA-I and ERA5 and its impact on land surface estimates should be included briefly in the Introduction and further discussed in the Results and Discussion section.
   The authors agree with the referee that changes in the data assimilation system were not well-emphasized in the manuscript.

   <u>Changes in manuscript:</u> the changes in the data assimilation system will be more highlighted in the revised version of the manuscript.

3. Section 2 is a somewhat odd mix of "methods" and "data". E.g., the FLUXNET 2015 section 2.2 includes discussion of how the climatology is derived for the computation of the standardised anomalies, but this would also apply to the other datasets (incl. ERA-I and ERA5). The entire section needs to be reorganized and be more clearly separated into "Data" and "Methods".
   The authors agree that this would improve the readability of the manuscript.

   <u>Changes in the manuscript:</u> the authors will re-organise Section 2 and introduce a section where methods (such as metrics calculation and anomaly calculation) are described.

4. The violin plots are great visual tools, but I assume their construction involves is some fitting of the distributions. These details should be in the "Methods" section.
   The method used to construct the violin plots is mentioned in the caption of the figures and involves a kernel density estimation approach.

   Changes in the manuscript: the authors will briefly mention this in the method section of the revised manuscript.

5. There are gaps in the literature discussion. E.g., Draper et al (2018) is a highly relevant assessment of reanalysis estimates of land surface energy flux estimates, incl ERA-I. Draper, C. S., R. H. Reichle, and R. D. Koster (2018), Assessment of MERRA-2 Land Surface Energy Flux Estimates, Journal of Climate, 31, 671-691, doi:10.1175/JCLI-D-17-0121.1.
   The authors thank the reviewer for this suggestion.

   Changes in the manuscript: the authors will cite the paper in the manuscript, and reconsider whether other references may be of relevance.

6. Related to (5), the present paper only investigates ERA-I and ERA5. The paper would be considerably more relevant to readers if these ECMWF products were assessed along with at least one or two other, major reanalysis products (such as MERRA-2).
   The authors thank the referee for this suggestion and fully agree that this would be an interesting analysis. However, the focus of the paper is to make a first assessment of the quality of the turbulent fluxes in ERA5 against different observation-based data sets. Including other reanalyses – next to ERA-I, which is there to show improvements of ERA5 upon its predecessor – would deviate the focus from ERA5, and turn this manuscript into a reanalysis comparison paper.

   Changes in the manuscript: no changes.

7. If I understood this correctly, the GLEAM and CLASS4GL analyses work as follows: (i) use ERA-I and, separately, ERA5 to force GLEAM (or CLASS4GL), then (ii) evaluate the results against tower (or balloon) measurements. This approach depends on the assumption that GLEAM and CLASS4GL are very good models, or at least that they do not have compensating errors. If, say, errors in GLEAM were to compensate for errors in ERA-I, forcing GLEAM with a better reanalysis isn't necessarily going to deliver better GLEAM outputs. Similarly for CLASS4GL. This is a major caveat that needs to be discussed prominently in the paper.
   The authors agree that this type of analysis comes with assumptions, yet it still provides useful insights in the quality of ERA5. Unless errors in ERA5 and GLEAM (or CLASS4GL) are somehow (anti-)correlated, forcing either of the models with better inputs can only lead to better model output as compared against in situ observations when done systematically at a large number of sites. Given the strong differences between the model concepts of ERA5 and GLEAM (or CLASS4GL), the authors believe that both are sufficiently independent to be used in such analysis.

   Changes in the manuscript: this issue will be briefly discussed in the revised version of the manuscript.

8. Figures A.1-A.4 are oddly placed. Either they need to be put into a proper Appendix, with discussion in the appendix as well, or they need to placed in a separate "Supplementary Information" document. As assembled, Figures A.1-A.4 are simply out of order.
The figures will be uploaded as supplementary material, following the guidelines of Copernicus.

Changes in the manuscript: the figures will be uploaded as supplementary material, following the guidelines of Copernicus.

9. Figures 2 and 4 are a confusing mix of dimensional metrics (MD for raw fluxes) and unit-less metrics (MD for raw Bowen ratio, and MAD and R for standardised anomalies). These figures also lack basic information about a dimensional 2nd-order metric such as MAD or RMSE for dimensional (raw or anomaly) variables. The readers are going to want to know typical MAD or RMSE values for fluxes in units of W/m2, incl. and/or excl. the seasonal cycle. I suggest revisiting the assembly of Figures 2, 4, A.1-A.4. The paper would be much easier to follow and more informative if, say, one figure includes only (dimensional) metrics computed from raw time series and another figure includes only (non-dimensional) metrics from standardised anomaly time series.
The primary reason to focus on the evaluation of anomalies is to minimise the effect of the strong seasonal cycle in the turbulent fluxes, which might mask important differences between the quality of ERA5 and ERA-I. Then, we prefer to focus on standardised anomalies to allow direct comparison of metrics from the different turbulent fluxes that typically range in a different order of magnitude. Needless to say that calculating the Mean Difference on anomaly time series is useless as anomalies are mean zero; hence we report the Mean Difference calculated on raw time series.

However, we do agree with the referee that readers might be interested in the statistics calculated on raw time series as well. Therefore, we report these corresponding statistics in Figures A1 and A4.

Changes in the manuscript: The authors believe that the content of the figures is described with sufficient detail in the captions and corresponding text to avoid confusion, thus we foresee no changes.

10. P10/L15-16: "Figure 4b shows that ERA5 is better at capturing..." Doesn't this invalidate the conclusion drawn from Fig 4a, that is, the evaluation of ERA-I and ERA5 through GLEAM? See also comment (7) above.
As replied to comment #7, the authors are convinced that forcing GLEAM with better inputs, can only lead to better validation statistics against in situ, when done over a large number of sites and for long-term periods. Hence, we think the statements are thus correct.

Changes in the manuscript: no changes.

11. P11/L3-8: One aspect that may come into play here is that GLEAM+ERA5 is an off-line (land-only) modeling system that does not permit feedback whereas ERA5 is a coupled land-atmosphere modeling system. This may be related to a finding by Draper et al (2018): "Finally, the SH results for MERRA-Land are troubling. While MERRA-Land did have the desired reduction in the LH biases compared to MERRA (to 1 W/m2 in the global land annual average), it also had a compensating, and much larger, increase in the SH bias (up to 15W m22 in the global land average)" [beginning of p689

of Draper et al. 2018]. See also comment 5) above about the need for a better integration of the results of the present paper into the literature context.
We fully agree with the referee that this is a main difference between both modelling approaches that may strongly affect the simulated fluxes. We also thank the reviewer for the suggested paper that fits well within the current manuscript.

Changes in the manuscript: this finding will be more discussed in detail and be put into better perspective within the existing literature.

12. P11/L26-P12/ L2: There is no discussion of Fig 9! In this paragraph, insert explicit references to Figs 7, 8, and 9 in the relevant place within the paragraph. E.g., reference Fig 7 in P11/L28, reference Fig 8 in P11/L33. This reveals that Fig 9 has not been discussed.
We thank the referee for picking this up and will extend the discussion of Figures 7–9.

Changes in the manuscript: the discussion on Figures 7–9 will be extended in the revised version of the manuscript.

13. In some cases the results are overstated. E.g., P12/L22: "The improvements are less clear..." suggests that there are some (hard-to-see) improvements, when in fact the results are neutral at best. P14/L3-4: The statement here is not consistent with the results of Fig 4 that show that ERA-I estimates of the sensible heat flux and Bowen ratio are better than those of ERA5.
The authors will have a detailed look at the conclusions again to soften some statements and to align the conclusions better with the results discussed in the remainder of the manuscript.

Changes in the manuscript: The manuscript will be screened for conclusions that might be too strong.

14. The tower validation results should come with some measure of statistical significance or error bars. Are the improvements, that is, the small shifts in the distributions of the metrics as shown in the violin plots, meaningful?
We agree with the reviewer that differences in quality are sometimes marginal, although often consistent. Although relying on assumptions on its own, we agree that a measure of statistical significance would be useful to report.

Changes in the manuscript: a measure of statistical significance will be included in the revised version of the manuscript.

Minor comments:

1. P2/L13: Reichle et al. 2017 is a reference primarily for MERRA-2 land surface estimates. MERRA-2 which is a full atmospheric reanalysis, similar to ERA-I and ERA5. In this place, however, the authors are here referring to land-only reanalyses, such as ERA-Interim/Land and MERRA-Land. For the latter, Reichle et al. (2011) is a better reference. Note that there is *not* a land-only reanalysis associated with MERRA-2.
Reichle, R. H., et al. (2011), Assessment and enhancement of MERRA land surface hydrology estimates, Journal of Climate, 24, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.
We thank the referee for the suggested correction.

Changes in the manuscript: the reference will be updated.

2. P2/L26-30: The text here is about very geographically limited results (southern Antarctic peninsula) or sea-ice, which is not the focus of the present paper. I suggest deleting this text or moving it further down. It confuses the reader by distracting from the focus of the paper on the global land surface turbulent fluxes.
   The main idea of this paragraph is to give a brief overview of studies that have already evaluated the quality of ERA5, irrespective of the scientific field, and to indicate that – although the number of studies is rather limited – results generally show a high quality of ERA5 compared to other datasets. We agree that these studies had a different focus than the current manuscript, but the authors believe that this short discussion is still relevant.

   Changes in the manuscript: no changes.

3. P4/L19: On first reading, I completely missed the term "standardised" here and later got confused about the lack of dimensions/units in the graphics. In many papers, anomalies from the seasonal cycle are examined without standardisation, i.e., they are dimensional anomalies. There is nothing wrong per se with the standardised anomalies, but please make it clearer that you are focusing on dimensionless anomalies.
   Thanks for pointing this out.

   Changes in the manuscript: the use of standardised anomalies and the reason for this choice will be better described in the manuscript.

4. P10/L31: typo: "we should emphasis" –> "we should emphasise"
   We thank the referee for their detailed look at the paper and picking this up.

   Changes in the manuscript: the typo will be corrected.

5. P12/L16+L20: The numbers referenced here contradict the numbers in the graphic.
   We thank the referee for picking this up, the numbers were indeed incorrectly reported.

   Changes in the manuscript: the correct numbers will be reported in the revised version of the manuscript.

6. Caption Fig 2: Replace "For MD, the distribution of beta…" with "The distribution of the MD of beta…"???
   We thank the referee for the suggestion.

   Changes in the manuscript: the captions will be updated per suggestion of the reviewer.

7. Caption Fig 7: "…between the absolute bias in ERA5 and ERA-I;"??? Maybe I'm misunderstanding this, but I think the bottom panel shows abs(bias(ERA-I)) minus abs(bias(ERA5), that is, the sign of the abs bias difference is different from what the caption suggests.
   We thank the reviewer for pointing this out. The caption should indeed read: "The bottom map represents the difference between the absolute bias in ERA-I and ERA5; hence, green colors represent lower bias in ERA5 than in ERA-I."

Changes in the manuscript: the caption will be corrected.

8. Caption Fig 10: "...versus ERA-I (squares)"??? Should this read "...versus ERA-I (triangles)"???
   We thank the referee again for his detailed look at the figures, the symbols were indeed wrongly referenced in the caption.

   Changes in the manuscript: the cation will be updated per suggestion of the reviewer.

9. Figure 11 needs units on the colorbars. I also suggest making the graphic bigger so it can be read more clearly in a hard copy for the next round of reviews.
   The authors agree that the figures were too small and that units need to be included on the colorbar.

   Changes in the manuscript: the size of the figure will be increased and units will be included.