Earth Syst. Sci. Data Discuss., author comment AC1
https://doi.org/10.5194/essd-2021-82-AC1, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Joaquín Muñoz-Sabater et al.

Author comment on "ERA5-Land: a state-of-the-art global reanalysis dataset for land applications" by Joaquín Muñoz-Sabater et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-82-AC1, 2021

We would like to thank this reviewer for his useful and relevant comments that has helped us to improve the manuscript. Please, find here below a point-by-point response to all your comments:

1) In line 6 of the abstract you mention that ERA5-Land enables the characterisation of trends. Yet, the manuscript does not include any analysis of trends. Why were trends not assessed, given that likely the product will be intensively used for trend analyses?

Thanks for your comment. We don't intend to give the impression that this paper also includes analysis trends. The latter is out of the scope of the paper and we believe that trends' analyses deserve a dedicated study, whereas this paper is more focused on the presentation of the ERA5-Land dataset.

Therefore, in order to make it clearer we have replaced in the abstract:

"ERA5-Land describes the evolution of the water and energy cycles over land in a consistent manner over the production period, enabling the characterisation of trends and anomalies" by

"ERA5-Land describes the evolution of the water and energy cycles over land in a consistent manner over the production period, which among others could be used to analyse trends and anomalies".

Also, to align with the comments of the second reviewer, we also have added a sentence in the "discussion and conclusions" section in lines 658-660 of the revised manuscript:

"The reduced impact of discontinuities by using longer spin-up periods, could also be a crucial factor to obtain accurate trends over multi-decadal periods for variables slowly changing in time, for instance the root-zone soil moisture (see bottom panel of Fig.3)".

2) The benefits of ERA5-Land wrt ERA5 are not ubiquitous. Clearly, the resolution

is higher, but this doesn't always lead to better skill. To convince the user of using a ~10fold larger dataset and exponentially increasing subsequent computation times for further use, could you summarise at the end of the manuscript where, for which variables, and for what applications ERA5-Land shall be favoured over ERA5?

Thanks for your comment. As a matter of fact, ERA5-Land is in size much smaller than ERA5 (TB vs PB) since is limited to the land surface layer. However, it is still true that for those fields which are common to ERA5 and ERA5-Land, the latter will be, due to the higher resolution, larger in volume, despite that the data over oceans in ERA5-Land has been masked out and therefore does not inflate the total volume of the field.

The main objective of this paper is the presentation of ERA5-Land while providing the reader sufficient evaluation elements (through the evaluation of several key variables of the water and energy cycles) to guarantee the good quality of the ERA5-Land dataset. At this respect and in wider terms, due to its specific characteristics (hourly frequency, multi-decadal, consistency, higher resolution) we believe ERA5-Land is ideal for all type of land applications, but the choice of an ERA5 or ERA5-Land common field will depend very much on the specific application. Relevant questions are: are data volume and data handling limiting factors? Are resolution and accuracy the most crucial factors? Is the analysis applied at global, regional or local scale? What is the computational power for the required application? In the 'conclusions and discussion' section, we have provided specific comments for each of the variables evaluated in this paper (and to make it clearer which variables are we discussing/concluding, we have replaced the a), b), c)… by the relevant variable name), but it is however impossible to produce a recommendation list for each application. Following the reviewer comment, we have added the following sentence to last paragraph of the 'conclusions and discussion' section, Lines 665-667 of the revised manuscript:

"While, in wider terms, we recommend the use of ERA5-Land fields over ERA5 for all type of land applications, one should factor in their choice elements such as available computer and data handling resources, importance of spatial resolution versus data volume, area of application, temporal consistency, etc."

3) I appreciate that the new product can be comprehensively assessed for all fields, but can you at least justify why these fields were evaluated and others not?

We assume the reviewer means "cannot be comprehensively assessed for all fields". Our choice was based on four factors: a) key variables of the water cycle, b) key variables of the energy cycle, c) availability and accessibility of evaluation data at the time of the study, d) inclusion of global scale evaluation by using reference satellite datasets. Based on all the four previous factors, we came up with the set of selected variables and reference datasets. We agree that the list of variables and reference datasets selected are not exhaustive, for instance 2 m temperature evaluation and soil temperature data from the ISMN as reference dataset could also have been added to this study. However, the length of the paper was a constrain and we had to prioritise some choices (see also our response to the next point). Our hope is that indeed, as stated in lines 647-651 of the original manuscript, the wider scientific community engages in more in-depth analyses of other variables, as it has been the case with the ERA5 atmospheric reanalysis.

To make this point more clear, we have additionally added the following sentence to lines 183-184 of the revised manuscript:

"Note that the list of evaluated variables and reference datasets is not exhaustive and was based on factors such as availability of data at the time of the evaluation".

4) Similarly, one needs to make choices regarding the reference data to be used. The choice seems slightly arbitrary and hence, the conclusions drawn can be inconsistent between fields. Especially, as ERA5-Land is embedded in the C3S Climate Data Store I would have expected that the assessment would draw on the various (satellite) products provided by this service. Vice versa, for LST you used a satellite product but you could also have used soil temperature data from the ISMN.

Although ERA5-Land is available from the C3S Climate Data Store, the evaluation exercise started much before that some of the current satellite datasets in the CDS were available. Evaluation against satellite data is not trivial, and requires dedicated efforts for each dataset considered. Our primary choice has been using in-situ ground data as 'ground truth', but we didn't want to limit ourselves to in situ only.

For skin temperature, in situ soil temperature measurements from ISMN are representative of soil at different temperatures, while we aimed at validating the joint contribution of soil (top surface) and vegetation into skin temperature, which is more closely related to satellite estimates from thermal sensors. Among the different sensors, we chose MODIS for its spatio-temporal coverage, since it provides four daily global long-term observations combining Terra and Aqua platforms, and daytime and night time overpasses. Note this MODIS ensemble has in turn been validated with in-situ data from tower-based measurements (see Chen et al., 2017).

Since, in this paper, the evaluation exercise cannot be exhaustive for all fields, we have been guided by the aim of being representative and relevant in our choices for selected and independent validation results, while keeping the size of the paper limited. Please, see also the response to the previous comment.

Minor issues

- **Line 227: please also cite the recently published paper on the ISMN: https://doi.org/10.5194/hess-2021-2**

Done (now in line 232 of the revised paper)

- **Did you also apply the quality flags of the ISMN data? https://doi.org/10.2136/vzj2012.0097**

No, although we acknowledge the usefulness of such a flags, the quality flags of the ISMN data were not used in this study.

- **Line 391: From the boxplots one cannot tell that ERA5-Land on average performs slightly better than ERA5, as the results for the different regions are based on a different number of stations. Does this statement still hold if you perform the analysis for all stations in these regions together?**

We have produced the boxplots adding all the stations used of the ISMN, independently of the region. The plots are generated for the top soil layer, as for the root-zone we only had observations for North-America. As the reviewer can see (fig.1 & 2 attached), the boxplots

are very similar to those shown in Fig. 7 (a-d) of the manuscript, as most of the stations are located in North-America. The conclusions are the same, so we prefer to keep the current boxplots to show the break-down per regions, and which also show quite clearly that the impact of ERA5-Land is larger in the root-zone.

Figure 1.- Box plots showing the evaluation metrics of ERA5-Land and ERA5 against in situ measurements of the ISMN network at 5 cm. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th (q25) and 75th (q75) percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers.

Figure 2.- Box plots showing the evaluation metrics of ERA5-Land and ERA5 anomalies against in situ measurements of the ISMN network at 5 cm. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th (q25) and 75th (q75) percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers.

- **URL of the ISMN: change into ismn.earth**

Thanks to point this out. Done.

- **As this is a dataset paper, it would be useful to include a list and short description of the input and output fields**

Following the reviewer recommendation, we have included an appendix to the manuscript with the list of input/output fields, and provided a reference for the description of each field in the caption of the table. The tables are also referenced in the main text, in lines 87 and 126-127 of the revised manuscript.

- **For ET: briefly describe the evaluation metrics, even as the fields sshf, slhf (l492), and bowen ratio**

Thanks for this remark. We have removed the description of the evaluation metrics from the beginning of section 4.5.1, and moved them and slightly modified to the end of section 3.7.1. (lines 361-363 of the revised manuscript), as this is more consistent with the evaluation metrics described for each subsection.

We are not sure what the reviewer means by "even as the fields sshf, slhf (l492) and bowen ratio". If the reviewer means what these abbreviations stands for, they were defined in lines 334 and 335 of the original manuscript, section 3.7.1. However, in the revised version of the paper and also in line with the comments of reviewer-2, we have replaced all "sshf and slhf" entries by "H" and "λρE", respectively.

Please also note the supplement to this comment:
https://essd.copernicus.org/preprints/essd-2021-82/essd-2021-82-AC1-supplement.zip