

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2021-2-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.



Comment on hess-2021-2

Mirko Mälicke (Referee)

Referee comment on "The International Soil Moisture Network: serving Earth system science for over a decade" by Wouter Dorigo et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-2-RC2>, 2021

The authors present the International Soil Moisture Network (ISMN), a global database and collection of in situ soil moisture measurements. Beyond data storage, general issues about representativeness, data harmonization, and data quality are raised and answered. Finally, the extensive description of the ISMN state is accompanied by a review of scientific studies that used ISMN.

I think the manuscript is overall of high quality, great importance, clear language and I have fewer minor comments than the manuscript has authors.

With Best regards,

Mirko Mälicke

My main comments are:

- Section 4.2.1 This is the first section reviewing the usage of ISMN in scientific publications. The authors present a huge amount of studies, that use ISMN data to evaluate satellite products. The whole section is a collection of references and very brief descriptions, which satellite products were involved. I am wondering if there is any general conclusion that can be drawn from this section. Are there studies that found flaws in satellite products, which would not have been found without ISMN data? Did ISMN data and especially the fact that data from various networks is available, foster or even enable the development of novel satellite-based products, that would not have been possible?

How often is more than one network involved? Or to put it the other way around: Is the multitude of available networks driving satellite product evaluation or is ISMN a collection of isolated data that is used for evaluation anyway? There are so many great studies mentioned in this section, that it would be a waste to just list them without gaining some new insights.

- Section 5.1.2 (P. 25): The authors describe challenges concerning the spatial representativeness of ISMN data and the resulting issues for validating (spatially) coarser soil moisture data. From my point of view, similar issues can result from temporal inconsistencies. While the ISMN has made a lot of effort to harmonize the

temporal resolution of ISMN data, this does not yet solve all representativeness issues. The authors should consider adding a short paragraph on temporal representativeness issues here, as well.

- As a general comment, I would like to point out, that the first almost 20 pages of the manuscript are summarizing and discussing the data in ISMN as well as ISMN itself. While I do think that all of the presented figures and sections are important and helpful, the more classic review of scientific studies using ISMN is noticeably shorter. I think in some parts, like section 4.2.1, the authors might consider extending section 4. Some more insights on how the reviewed studies are fostered or are even made possible in the first place, due to ISMN could be helpful. ISMN is not just a collection of data files, it's a reference database, that has the power to set standards and produce unified quality controls and metadata information for soil moisture measurements. I would love to see a section elaborating how ISMN might already do this today, or what a possible path might look like for the future. Reviewing all the work that is based on or involving ISMN, the authors might see a pattern here.

These are my minor comments in order of appearance:

- P.2 L.10: Which proportion of the total datasets are still updated? A number could help the reader to set `\textit{"many"}` into a context. The number is named on P. 4 L. 75 and the authors could consider moving this up. From my point of view, having roughly 70% of the networks still active is a feature and more than I would have expected.
- P.3 L. 36: At this point I was already so excited about the `\textit{advanced quality control methods}`, that the authors could consider naming at least one of them as an example, already.
- P.4 L. 50 - 51: I personally like the idea of involving citizen scientists in the collection of soil moisture data and have already involved students (who are not 'citizens' in this case) in some short measurement campaigns. I personally doubt, that citizen scientists can really contribute data that hold metadata and quality control requirements of most scientific applications. Therefore, from a personal point of view, I would like to see at least one reference of a successful application here.
- P. 4 L. 69: The authors might consider adding some numbers about the per-continent distribution of networks here, to give the reader a quick overview.
- P. 4 L. 72: I am wondering if it would be feasible to add a graph about the landscape type distribution, here.
- Figure 1 (P. 5) Due to the station density, I suggest putting a sub-figure of the USA into the map. Right now, I can only see that there are a **lot** of different stations of different colors. Another approach could be to rather indicate the geometric centroids of the networks in the USA and keep the stations in a transparent grey?
- Figure 2 (P. 6) I am a little bit confused by figure 2. I can see many networks that are marked as active but did not contribute any data for years. How is this possible? What exactly is an *active network*?
- Table 1 (P. 7) I think this table could really be enhanced by adding another column containing the total count of sensors / timeseries / stations, whatever technically appropriate, of the respective variable. I.e. for soil suction, it will be of importance for many readers, how often soil moisture data is accompanied by suction data.
- P.9 L. 151: The authors state, that the data is formatted according to CEOP **or** a slightly modified version. Does this imply that the data can come in different formats, or is it always the modified format (which was derived from CEOP)?
- P. 10 L. 159 - 160: I was just wondering if the authors think, that this user feedback (on quality issues) might also be a valuable public resource (similar to Github issues)? Having a 'conversation' about quality issues publicly, might help other users to handle these issues for their specific application.

- P. 10 L. 161 - 162: How exactly can a registered download option prevent data misuse? Isn't it the distributed license that regulates the terms and conditions of how the data may be used? Comparable to this manuscript itself, which can also be retrieved without registration and it's the CC BY 4.0 that prevents me from misusing it.
- P. 11 L. 193 - 195: I was a bit confused by this paragraph. Does it mean that an additional flag for the other variables soil temperature, air temperature, and precipitation was added, or that a spurious observation in one of these variables will also flag the soil moisture measurement as spurious?
- P. 15 L. 239: What is ubRSMD?
- P. 16 L. 248: A short clause describing what the "*triple collocation approach*" is would help me to understand this section even better.
- P 20 L. 325: I guess the question mark should not be there.
- Section 4.2.2 (P.20 - 21) This section gives an overview of operational services relying on ISMN data. Unfortunately, the authors only refer to the '*ISMN data*'. I think it could be more helpful for readers, that might happen to be users of the services as well if the specific networks involved for each of these services are listed. Otherwise, I am under the impression, that all data is involved all the time.
- Section 5.1.3 For me it was not clear from this section, if ISMN nowadays already includes a considerable amount of low-cost sensor networks and if the raised issues could be addressed using ISMN. The authors should consider clarifying this.
- Section 5.1.4 This is connected to the point above: Is GROW part of ISMN? I guess it is, as it is listed in the appendix. Further, this section points at some challenges connected to citizen science, but it remains kind of open whether citizen science data should be contributed or not. The authors should make this more clear. A scientific user of ISMN should be aware of the context of data collection and I am not sure if the metadata provided is suitable enough to transport this important information if it is even contained.

Finally, I would like to make two personal comments, that are not part of the review in a strict sense:

Concerning 5.2.3: Working on a web-based data portal for a couple of years myself, I would suggest adding a new section to the data overview Bubbles on the map, that contains a suggestion on how to cite a specific dataset. The same could be implemented for networks and on download. I think this will already significantly increase the visibility and citations, as they are usually not missed out on bad will.

Secondly, I personally don't want to finish my comment, without mentioning ESSD. The authors have described the state of the ISMN database in great detail and will convince every reader why the database and the data itself are unique and a real asset for environmental science. While this manuscript is labeled as a review paper, a considerable amount of it is also halfway a dataset description paper. From my personal point of view, it could be really worth it to prepare an ESSD publication for ISMN itself. That can help to further increase the visibility of ISMN. At the end of the day, I found myself a couple of times hitting 'soil moisture' into the search bar of ESSD, and not finding ISMN is a pity.

With Best regards,

Mirko Mälicke