

Dear Dr. Hanqin Tian and reviewers,

We appreciate for the opportunity to improve our manuscript.

We would like to thank you for the time and effort spent on reviewing. In the resubmitted manuscript, we explicitly address all of the comments. Language was edited thoroughly in preparing the resubmission.

Additionally, as we were concerned about the modest manuscript score on data quality criterion, we improved our data quality after a short communication with the editor (Dr. Hanqin Tian). Specifically, we improved the accessibility and usability of the dataset by providing an alternative data format including selected data summaries, and we extended insights in dataset validity by performing additional comparisons with FAOSTAT structural data which was firstly released two months ago. FAOSTAT structural data includes and significantly extends one of the datasets that we compared in the original manuscript, thus, we reorganized the respective content in the resubmission.

Below you can find our revisions, structured as: **[Comment]** from reviewers, **[Response]** from authors, and **[Change]** made in the manuscript (clean mode).

Best regards,

Han Su, on behalf of all co-authors

PhD Candidate

Multidisciplinary Water Management group, University of Twente

Report #1

[Comment] The manuscript has addressed the concerns raised during the previous round of review. Editorial suggestions are listed below, but it is up to the author to decide whether to adopt them. This is a nice work, and I acknowledge the authors' efforts to develop the dataset.

[Response] Thank you for your acknowledgment. We appreciate your comments, the time and effort spent on reviewing. All the suggestions are agreed on and implemented.

Besides what was suggested, we were eager to improve our data quality once receiving your review report, due to its only modest rating.

Already in the original manuscript, following the EESD guidelines, we have made the best use of state-of-art publicly available data sources to develop and validate this dataset. The uncertainty issue is explicitly addressed in the current manuscript. Although there obviously is uncertainty, as in any dataset, we do not think current uncertainty will limit the potential applications of our dataset in global studies, particularly since there are no comparable other datasets available. Dataset validity information may guide interpretation in the applications.

Still, we see potentials for data quality improvement regarding data accessibility and comparison with additional dataset. After communicating with the editor, we decided to provide the dataset in an additional data format (netCDF) with selected data summaries. We hope our dataset is more accessible for some potential users with the new data format. We also grasped the opportunity to add an extra comparison with FAOSTAT structural data on farm size which was firstly released just two months ago. The FAOSTAT structural data covers one of the datasets that we compared before, thus, we replaced one of the original comparisons with this new one.

[Change] The description of FAOSTAT structural data and how it covers Lowder et al. (2016), that we compared in the original manuscript, were added in line 255-262. The new comparison with FAOSTAT structural data can be found in line 357-376, Fig. 7 and Fig. A3.

Technical corrections:

[Comment] 1. L108-109. Although the two crop maps appear line 118, I would suggest mentioning the name of datasets (GAEZv4 and SPAM2010) in lines 108-109 for clarity. Since many datasets are used in this study, I was confused which datasets the authors mention.

[Response] We agree.

[Change] The names of datasets (GAEZv4 and SPAM2010) were added in line 107.

[Comment] 2. For GAEZv4 and SPAM2010, the indicator is labelled "crop map" in Fig. 1 but it appears to be "harvested area" in Table 1. It would be more readable if a note is added to the figure and table to state these are interchangeable here.

[Response] We agree.

[Change] Notes were added to Fig. 1 and Table 1 showing the “crop map” and “harvested area” are interchangeable.

[Comment] 3. L435. “SDG2 (Zero Hunger)” is true. But more precisely, it is Target 2.3 of SDG2 (see <https://unstats.un.org/sdgs/metadata/?Text=&Goal=2&Target=2.3>).

[Response] We agree.

[Change] The specification “Target 2.3” was added in line 497.

Report #2

[Comment] The authors tried to develop a gridded map of farm-size dataset by harmonizing different sources of data. I agree this is an interesting study and the data developed has potential to improve global/regional simulations. However, the current version is not in good shape.

[Response] Thank you for your comments which enable us to improve the manuscript. We appreciate the time and effort you spent on reviewing. All the suggestions are agreed on, and implemented.

Besides what was suggested, we were eager to improve our data quality once receiving your review report, due to its only modest rating.

Already in the original manuscript, following the EESD guidelines, we have made the best use of state-of-art publicly available data sources to develop and validate this dataset. The uncertainty issue is explicitly addressed in the current manuscript. Although there obviously is uncertainty, as in any dataset, we do not think current uncertainty will limit the potential applications of our dataset in global studies, particularly since there are no comparable other datasets available. Dataset validity information may guide interpretation in the applications.

Still, we see potentials for data quality improvement regarding data accessibility and comparison with additional dataset. After communicating with the editor, we decided to provide the dataset in an additional data format (netCDF) with selected data summaries. We hope our dataset is more accessible for some potential users with the new data format. We also grasped the opportunity to add an extra comparison with FAOSTAT structural data on farm size which was firstly released just two months ago. The FAOSTAT structural data covers one of the datasets that we compared before, thus, we replaced one of the original comparisons with this new one.

[Change] The description of FAOSTAT structural data and how it covers Lowder et al. (2016), that we compared in the original manuscript, were added in line 255-262. The new comparison with FAOSTAT structural data can be found in line 357-376, Fig. 7 and Fig. A3.

[Comment] Language editing is strongly recommended. The language should be improved, especially the descriptions in the results section.

[Response] We agree.

[Change] Language was edited thoroughly in preparing the resubmission, paying special attention to results section.

[Comment] Another major issue is that a substantial amount of discussion was mixed into the results. These contents should be moved to the discussion section. A typical example is the content in Lines275-280.

[Response] We agree. Considering this manuscript is a data description paper, we limited the results section to present the data statistics, validation of the datasets, and comparison of the dataset, and moved other content to discussion.

[Change] We moved the explanations on farming systems and farm sizes to discussion section 4.1, line 407-426.

[Comment] Besides, there are numerous, minor errors, which indicate that the authors didn't pay much attentions on this work.

[Response] We tried our best to avoid any errors when preparing and revising the manuscript. Unfortunately, there were still some minor errors. We appreciate you pointing them out.

[Change] All the errors found by reviews (as below) and ourselves were fixed.

Other suggestions:

[Comment] 1. Line67, why 2018b is cited before 2018a.

[Response] Citations and references are managed by EndNote using the official style provided in the ESSD submission guideline. According to the style, the reference is sorted alphabetically not by the citation order. Thus, it is possible to cite 2018b first. But, the Ricciardi et al. (2018a) and Ricciardi et al. (2018b) are special because they are the same study: Ricciardi et al. (2018a) is the data brief paper of Ricciardi et al. (2018b).

We agree citing 2018b before 2018a might be confusing. For this specific case, we cite both Ricciardi et al. (2018a) and Ricciardi et al. (2018b) to avoid potential confusion.

[Change] Ricciardi et al. (2018a) and Ricciardi et al. (2018b) were simultaneously cited throughout the manuscript.

[Comment] 2. Why a reference at the end of the Line69?

[Response] The reference was meant as the start of the next sentence.

[Change] We clarified the citation in line 70.

[Comment] 3. ?Line79-80, the references.

[Response] We interpret the comment to mean that references were not cited properly, and agree with that. We improved the whole sentence and citations.

[Change] The citations were clarified, lines 84.

[Comment] 4. Line90, should be "estimate"

[Response] We agree.

[Change] The sentence was rewritten, line 93-95.

[Comment] 5. In section 3.2, you actually only talked about the dash line (overall) in Figure 4a. I would suggest to use thin lines for other cases. The figure is distracting and readers' attentions will be directed to the colored lines.

[Response] We agree. The dark line is firstly described than the colored lines. The figure is not optimal.

[Change] We improved the figure style based on the above suggestion, Fig. 5.

[Comment] 6. Line268, what observation. Should be revised to be more clear to readers. This is a typical example that the descriptions in this section should be improved.

[Response] We agree. The “observation” refers to “small farms irrigate a larger share of their area than large farms”, which is shown by our dataset and previous studies

[Change] We improved the description in this section for better clarity, line 289-291.

[Comment] 7. Lines282-283, what do you mean? Not clear.

[Response] We mean large farms irrigate more when water is scarce.

[Change] We improved the description for better clarity, line 307-309.

[Comment] 8. For Figure 4, “significant” is not clear to readers, although I know it indicates a water scarcity level higher than moderate. Explanation will be needed in the title. For example, low, moderate... indicate ...

[Response] We agree.

[Change] Explanations on the classification of water scarcity levels were added in the figure title, line 317-319, 418-420.

[Comment] 9. Suggest to redraw all scatter plots using more professional software (e.g. R). They are not aesthetically attractive.

[Response] We agree.

[Change] All the scatter plots were redrawn using *seaborn* and *matplotlib.pyplot* package in Python, Fig. 6.

[Comment] 10. Line312, please cite the source of the data. Which table?

[Response] We agree.

[Change] The tables were added as supplementary materials ([S6] and [S7]) and cited in the manuscript, line 326.