

Joint Global Change Research Institute (JGCRI)

Pacific Northwest National Laboratory (PNNL)

5825 University Research Ct,

College Park, MD 20740

August 18, 2022

**Response to reviewer comments for:**

“Global monthly sectoral water use for 2010-2100 at 0.5° resolution across alternative futures” (Manuscript ID: SDATA-22-00583), submitted to Nature Scientific Data

**Reviewer Comments Received: 01 August 2022**

Dear Guy Jones,

We would like to thank you and the reviewers for your invaluable feedback and comments. Please find a list of all editor and reviewers’ comments we received followed by an explanation of where and how each comment has been addressed in the revised manuscript. We think that with the suggested changes the paper has become much stronger.

Adasasdas

Sadasdq

Thank you for the guidance,

Sincerely,

Zarrar Khan

Contents

[List of All Reviewer Comments 3](#_Toc110504305)

[E1 Comment 1 6](#_Toc110504306)

[E1 Comment 2 6](#_Toc110504307)

[E1 Comment 3 6](#_Toc110504308)

[E1 Comment 4 6](#_Toc110504309)

[E1 Comment 5 7](#_Toc110504310)

[R1 Comment 1 7](#_Toc110504311)

[R1 Comment 2 7](#_Toc110504312)

[R1 Comment 3 8](#_Toc110504313)

[R2 Comment 1 8](#_Toc110504314)

[R2 Comment 2 9](#_Toc110504315)

[R2 Comment 3 9](#_Toc110504316)

[R2 Comment 4 9](#_Toc110504317)

[R2 Comment 5 9](#_Toc110504318)

[R2 Comment 6 9](#_Toc110504319)

[R2 Comment 7 10](#_Toc110504320)

[R2 Comment 8 10](#_Toc110504321)

[R2 Comment 9 10](#_Toc110504322)

[R2 Comment 10 10](#_Toc110504323)

[R2 Comment 11 10](#_Toc110504324)

[R2 Comment 12 11](#_Toc110504325)

# List of All Reviewer Comments

The following is a list of all reviewer comments received on **31 May 2022**. Each comment has been addressed in the text as described in the corresponding section. Comments are numbered by referee as R1 for Referee 1, and E1 for Editor.

E1 Comment 1: Please ensure that your references conform fully to the Nature style. See the examples at the link below: <http://www.nature.com/sdata/publish/submission-guidelines>#refs

E1 Comment 2: Note that "meta-repository" is not a widely understood term at this journal. If I understand correctly, this provides wider explanations of the methods used and external links to other deposited resources (e.g. Tethys v1.3.1 at Zenodo). Please note that:

* All experimental methods relating to this work should be available in the manuscript itself. I understand that a lot of this content may be duplicated in the paper, but if not, I would encourage you to share any additional workflow, validations, Basin examples, etc within the paper. A single supplementary document can be used if there are large amounts to add. Data visualizations are fine to share soley on this site as long as the underlying data has been been formally deposited
* All external resources should be cited in the manuscript - use the reference list for any citable objects with formal metadata (i.e. DOIs, author names, etc) and simply embed the ULRs in the text in ()s for websites without these. From quick inspection it looked like all the links were in the former category (i.e DOI'd objects that can be formally cited). Again, it looks like a number of these are already included in the paper (e.g. the Code Availability statement), but please check and confirm there is nothing exclusively shared at the met repository.

E1 Comment 3: Please change the Custom licence terms for the Dataverse dataset to a recognizable open licence (CC0 and CC-BY are usually required for this journal).

E1 Comment 4: Please add a data citation for the Dataverse dataset to the reference list using these instructions (https://www.nature.com/sdata/publish/submission-guidelines#data\_citations - note that a DOI URL should be used). Please add the reference number to wherever the dataset is mentioned in the text - the main position should be the first part of the Data Record in a sentence describing where the data has been deposited.

E1 Comment 5: Please also formally cite the items listed in the table in the Code Availability sections in the reference list. Please also note that all tables need a table label and a legend.

R1 Comment 1: Please provide a more detailed description of basin boundaries and selection principles in the manuscript.

* Discovery of the problem: Irrigation water accounts for a higher proportion of the total water consumption in the region, and the combination of region and basin can indeed be better constrained. But I am confused about the basin boundary in Figure 2 and how to select the basin for constraints. For example, the boundary of the Yangtze River Basin in Asia is not correct ([1], etc.), and the Yukon River Basin in North America [2] mentioned in the background that there is irrigation in the Basin, but is not isolated in Figure 2.
* References:

[1] Zhang Q, Xu C, Becker S, et al. Sediment and runoff changes in the Yangtze River basin during past 50 years. Journal of hydrology, 2006, 331(3-4): 511-523.

[2] Yang D, Zhao Y, Armstrong R, et al. Yukon River streamflow response to seasonal snow cover changes. Hydrological Processes: an International Journal, 2009, 23(1): 109-121.

R1 Comment 2: Please give a reasonable explanation for the degree to which the historical data results deviate from the 45° curve, and add the verification of the historical water data in time and space.

* Discovery of the problem: The dataset is to serve users, and the first consideration of users should be more about the quality of the dataset. When verifying the reliability of downscaling, the downscaled data must be compared to the original input data. However, the quality of the dataset also includes the accuracy of the water withdrawal/consumption data in time and space, this needs to be verified and should be reflected in the manuscript [3]. (1) Since neither the manuscript nor the meta-repository has found the result data used for the verification of this paper, the data verification has not been carried out. It is recommended to add the result data of the historical period of this article in the data set to facilitate the verification of the historical period for users and reviewers. (2) Through the comparison of two similar datasets in the meta-repository (Huang et al. 2018, Mekonnen, M.M and Hoekstra, A.Y. 2011), the discrete differences between regions can be explained,but the degree of deviation from the 45° curve doesn't seem to be well answered. (3) Therefore, it is further suggested to supplement the manuscript validation section with the available tabular data or integration results of government departments (such as USGS water withdrawal, water resource bulletins in China, etc.) or international organizations (FAO-AQUASTAT data). Comparison with the regional statistical results in this manuscript [4].
* References:

[3] Chiarelli D D, Passera C, Rosa L, et al. The green and blue crop water requirement WATNEEDS model and its global gridded outputs. Scientific data, 2020, 7(1): 1-9.

[4] Zhang, K., Li, X., Zheng, D., Zhang, L., Zhu, G. (2021). Satellite-based Global Irrigation Water Use data set (2011-2018). National Tibetan Plateau Data Center, DOI: 10.11888/Hydro.tpdc.271220. CSTR: 18406.11.Hydro.tpdc.271220.

R1 Comment 3: Please add the verification of future forecast data in the manuscript.

* Discovery of the problem: There may not be a similar combination method and cannot be accurately compared for the future sector's forecast data. But the similarities, differences ,trends or ranges of its predictions can still be compared in some aspects [5-6]. Future forecast data is another key content in the manuscript, and the availability needs to be further analyzed.
* References:

[5] Wada Y, Bierkens M F P. Sustainability of global water use: past reconstruction and future projections[J]. Environmental Research Letters, 2014, 9(10): 104003.

[6] Fujimori S, Hanasaki N, Masui T. Projections of industrial water withdrawal under shared socioeconomic pathways and climate mitigation scenarios. Sustainability Science, 2017, 12(2): 275-292.

R2 Comment 1: First, the methods of temporal downscaling are similar to Huang et al., 2018. The authors were required to demonstrate the manuscript's novelty.

R2 Comment 2: Second, there are no detailed description that can be applied to future water withdrawals and consumption in the Methods chapter.

R2 Comment 3: Third, the authors do not provide rigorous verification. The content of the technical validation section does not guarantee the accuracy and availability of this dataset.

R2 Comment 4: The manuscript describes the importance of water withdrawals and consumption, but the research progress on spatial and temporal downscaling methods is insufficiently described. Authors should explain the limitations of existing spatial and temporal downscaling methods and clearly explain how they have addressed them in this manuscript.

R2 Comment 5: L30-63 This paragraph is about the models and datasets used in this manuscript, which is more like what is in the Methods chapter. The authors should have organized the logical structure of the Background & Summary chapter.

R2 Comment 6: The Region/Basin Scale (Electricity and Irrigation) in figure 3 shows a very large differences in water withdrawals between adjacent regions, so how did authors deal with making the spatial distribution smooth in the Gridded Scale in Figure 3, because there are some regions with clear boundary in the northern Africa and in the northern North America. Meanwhile, many methods of spatial downscaling are average methods, how did authors ensure the accuracy of results between adjacent regions. In addition, how did authors eliminate the impact of some cross-regional water transfer projects and water facilities on the spatial downscaling of water withdrawals and consumption.

R2 Comment 7: L361-370 Authors did not validate the accuracy of the water withdrawals and consumption. Although the spatial and temporal downscaling methods were validated, the results of this validation verification are obvious according to the methods presented earlier by authors. It is recommended that authors increase the comparison results with statistics published by government agencies to prove the accuracy of the historical part of this dataset.

R2 Comment 8: L375-402 Authors explains some differences between this dataset and the other two datasets, but do not explain in detail whether these differences will affect the accuracy of the results. Furthermore, although the authors list the scatterplots of the validation with other datasets on a meta-repository, they do not give detailed explanations. The correctness of the validation results is not known only from Figure 5 and 6. Using the data of only one year (2010) is not very convincing. It is recommended that the author extend the period (e.g., 2005-2010) of validation results of this dataset and other datasets in detail to prove the stability of the prediction data.

R2 Comment 9: There are some regions with clear boundary in the Gridded Scale of Figure 3, such as in Sahara in Africa and in northern North America. What is the reason for this situation?

R2 Comment 10: In the formula 15 and 16, two identical “tempmax” appear in the denominator.

R2 Comment 11: There is no label in the first figure of “Temporal” of Figure 4b.

R2 Comment 12: L238 What is the meaning of “pρ values”?

# E1 Comment 1

**E1 Comment 1: Please ensure that your references conform fully to the Nature style. See the examples at the link below:** [**http://www.nature.com/sdata/publish/submission-guidelines**](http://www.nature.com/sdata/publish/submission-guidelines)**#refs**

**Response:**

**Edits Made:**

# E1 Comment 2

**E1 Comment 2: Note that "meta-repository" is not a widely understood term at this journal. If I understand correctly, this provides wider explanations of the methods used and external links to other deposited resources (e.g. Tethys v1.3.1 at Zenodo). Please note that:**

* **All experimental methods relating to this work should be available in the manuscript itself. I understand that a lot of this content may be duplicated in the paper, but if not, I would encourage you to share any additional workflow, validations, Basin examples, etc within the paper. A single supplementary document can be used if there are large amounts to add. Data visualizations are fine to share soley on this site as long as the underlying data has been been formally deposited**
* **All external resources should be cited in the manuscript - use the reference list for any citable objects with formal metadata (i.e. DOIs, author names, etc) and simply embed the ULRs in the text in ()s for websites without these. From quick inspection it looked like all the links were in the former category (i.e DOI'd objects that can be formally cited). Again, it looks like a number of these are already included in the paper (e.g. the Code Availability statement), but please check and confirm there is nothing exclusively shared at the met repository.**

**Response:**

**Edits Made:**

# E1 Comment 3

**E1 Comment 3: Please change the Custom licence terms for the Dataverse dataset to a recognizable open licence (CC0 and CC-BY are usually required for this journal).**

**Response:**

**Edits Made:**

# E1 Comment 4

**E1 Comment 4: Please add a data citation for the Dataverse dataset to the reference list using these instructions (https://www.nature.com/sdata/publish/submission-guidelines#data\_citations - note that a DOI URL should be used). Please add the reference number to wherever the dataset is mentioned in the text - the main position should be the first part of the Data Record in a sentence describing where the data has been deposited.**

**Response:**

**Edits Made:**

# E1 Comment 5

**E1 Comment 5: Please also formally cite the items listed in the table in the Code Availability sections in the reference list. Please also note that all tables need a table label and a legend.**

**Response:**

**Edits Made:**

# R1 Comment 1

**R1 Comment 1: Please provide a more detailed description of basin boundaries and selection principles in the manuscript.**

* **Discovery of the problem: Irrigation water accounts for a higher proportion of the total water consumption in the region, and the combination of region and basin can indeed be better constrained. But I am confused about the basin boundary in Figure 2 and how to select the basin for constraints. For example, the boundary of the Yangtze River Basin in Asia is not correct ([1], etc.), and the Yukon River Basin in North America [2] mentioned in the background that there is irrigation in the Basin, but is not isolated in Figure 2.**
* **References:**

**[1] Zhang Q, Xu C, Becker S, et al. Sediment and runoff changes in the Yangtze River basin during past 50 years. Journal of hydrology, 2006, 331(3-4): 511-523.**

**[2] Yang D, Zhao Y, Armstrong R, et al. Yukon River streamflow response to seasonal snow cover changes. Hydrological Processes: an International Journal, 2009, 23(1): 109-121.**

**Response:** The boundaries are imposed by the GCAM input data, as these areas correspond to its representation of regions and basins, so we did not select them as part of this study. We will refer readers to Moirai, the land data system used by GCAM, for more details on the selection principles. The basin boundaries are aggregated from watersheds, and named for river basins and coasts, so some area of a river basin may be part of the coast and included in that boundary instead. Additionally, Figure 2b shows the intersection of GCAM regions and basins, which will cause basins that are shared between regions to be divided even further.

Di Vittorio, A.V., Vernon, C.R. and Shu, S., 2020. Moirai Version 3: A Data Processing System to Generate Recent Historical Land Inputs for Global Modeling Applications at Various Scales. *Journal of Open Research Software*, 8(1), p.1. DOI: <http://doi.org/10.5334/jors.266>

Narayan, Kanishka, Di Vittorio, Alan, & Vernon, Chris. (2021). GCAM boundary spatial products from moirai v3.1 (1.1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4688451>

**Edits Made:** Methods section, added "(these boundaries are determined by Moirai , the land data system used by GCAM)”

# R1 Comment 2

**R1 Comment 2: Please give a reasonable explanation for the degree to which the historical data results deviate from the 45° curve, and add the verification of the historical water data in time and space.**

* **Discovery of the problem: The dataset is to serve users, and the first consideration of users should be more about the quality of the dataset. When verifying the reliability of downscaling, the downscaled data must be compared to the original input data. However, the quality of the dataset also includes the accuracy of the water withdrawal/consumption data in time and space, this needs to be verified and should be reflected in the manuscript [3]. (1) Since neither the manuscript nor the meta-repository has found the result data used for the verification of this paper, the data verification has not been carried out. It is recommended to add the result data of the historical period of this article in the data set to facilitate the verification of the historical period for users and reviewers. (2) Through the comparison of two similar datasets in the meta-repository (Huang et al. 2018,** **Mekonnen, M.M and Hoekstra, A.Y. 2011), the discrete differences between regions can be explained,but the degree of deviation from the 45° curve doesn't seem to be well answered. (3) Therefore, it is further suggested to supplement the manuscript validation section with the available tabular data or integration results of government departments (such as USGS water withdrawal, water resource bulletins in China, etc.) or international organizations (FAO-AQUASTAT data). Comparison with the regional statistical results in this manuscript [4].**
* **References:**

**[3] Chiarelli D D, Passera C, Rosa L, et al. The green and blue crop water requirement WATNEEDS model and its global gridded outputs. Scientific data, 2020, 7(1): 1-9.**

**[4] Zhang, K., Li, X., Zheng, D., Zhang, L., Zhu, G. (2021). Satellite-based Global Irrigation Water Use data set (2011-2018). National Tibetan Plateau Data Center, DOI: 10.11888/Hydro.tpdc.271220. CSTR: 18406.11.Hydro.tpdc.271220.**

**Response:** For the comparison with Huang et al. 2018 on the metarepo, the strong lines are a result of the same population proxy being used to downscale country scale data, rather than region scale data, as some countries will have a higher or lower per capita demand than the region they are part of (reflected in the slope of the line). The difference in the irrigation sector is because a different irrigation proxy was used altogether (Demeter downscaled crop maps were not used in that study). Some difference with Mekonnen, M.M and Hoekstra, A.Y. 2011 can be attributed to a difference in data year, as that study is meant to characterize the years 1996 through 2005, which is close but not identical to 2010. Additionally, proxy quantities likely don’t replicate all details of actual water demand, especially when compared at a grid cell by grid cell basis, but the overall spatial distribution is similar.

**Edits Made:**

# R1 Comment 3

**R1 Comment 3: Please add the verification of future forecast data in the manuscript.**

* **Discovery of the problem: There may not be a similar combination method and cannot be accurately compared for the future sector's forecast data. But the similarities, differences ,trends or ranges of its predictions can still be compared in some aspects [5-6]. Future forecast data is another key content in the manuscript, and the availability needs to be further analyzed.**
* **References:**

**[5] Wada Y, Bierkens M F P. Sustainability of global water use: past reconstruction and future projections[J]. Environmental Research Letters, 2014, 9(10): 104003.**

**[6] Fujimori S, Hanasaki N, Masui T. Projections of industrial water withdrawal under shared socioeconomic pathways and climate mitigation scenarios. Sustainability Science, 2017, 12(2): 275-292.**

**Response:** Verification of the region-scale forecasts from GCAM was carried out in <https://doi.org/10.1029/2018WR023452> (see the supporting information), which we have now made reference to. This accounts for the plausibility of the projections at region scale. The work done for this publication is the downscaling of those projections to gridded scale, and we include the comparison with Mekonnen and Hoekstra (2011) to show that these downscaling methods reasonably reproduce the spatial and temporal distributions.

**Edits Made:**

* In Technical Validation, added “Sectoral comparison between GCAM’s future water demand projections and other studies is carried out in the supporting information of Graham et al. 2018.”

# R2 Comment 1

**R2 Comment 1: First, the methods of temporal downscaling are similar to Huang et al., 2018. The authors were required to demonstrate the manuscript's novelty.**

**Response:** We do employ the same techniques as Huang et al., 2018 (which themselves were developed in earlier papers) for temporal downscaling. The novelty in this dataset comes from applying these techniques to future years across 75 combinations of SSP, RCP, and GCM, with spatial downscaling of irrigation demand using a land model also forced by these scenarios.

**Edits Made:**

# R2 Comment 2

**R2 Comment 2: Second, there are no detailed description that can be applied to future water withdrawals and consumption in the Methods chapter.**

**Response:** Future withdrawals/consumption at the region/basin level come from GCAM, and the downscaling methods are applied to these coarse-scale data to produce the finer future withdrawals and consumption. Details of these GCAM runs are in Graham et al. 2018.

**Edits Made:**

* In Technical Validation, added “Sectoral comparison between GCAM’s future water demand projections and other studies is carried out in the supporting information of Graham et al. 2018.”

# R2 Comment 3

**R2 Comment 3: Third, the authors do not provide rigorous verification. The content of the technical validation section does not guarantee the accuracy and availability of this dataset.**

**Response:** We have added a reference to the validation of region-scale water demand inputs from Graham et al. 2018., which compares those values to other projections to assess their reasonableness, and we verify that these conditions are satisfied in this data set. The methods used for spatial and temporal downscaling are verified on historical data in their respective papers, so the quality of the spatial and temporal detail in this data set is dependent on how suitable these techniques continue to be in the future, as well as uncertainty in the proxies. We included an example comparison with data for the year 2010 as a way of illustrating the degree to which these methods reproduce patterns of demand.

**Edits Made:**

# R2 Comment 4

**R2 Comment 4: The manuscript describes the importance of water withdrawals and consumption, but the research progress on spatial and temporal downscaling methods is insufficiently described. Authors should explain the limitations of existing spatial and temporal downscaling methods and clearly explain how they have addressed them in this manuscript.**

**Response:** This dataset utilizes existing downscaling algorithms, but applies them across a suite of future scenarios, utilizing a land use downscaling model (Demeter) to provide irrigation proxies forced by each scenario. Since this manuscript primarily concerns the dataset, not the development of new downscaling techniques, we have included an overview of similar datasets.

**Edits Made:**

# R2 Comment 5

**R2 Comment 5: L30-63 This paragraph is about the models and datasets used in this manuscript, which is more like what is in the Methods chapter. The authors should have organized the logical structure of the Background & Summary chapter.**

**Response:** We have reorganized this section for increased emphasis on background, keeping discussion on models and datasets broad to provide a summary of the research, and saving details for the methods section.

**Edits Made:**

* Put importance of dataset first, and summary second.

# R2 Comment 6

**R2 Comment 6: The Region/Basin Scale (Electricity and Irrigation) in figure 3 shows a very large differences in water withdrawals between adjacent regions, so how did authors deal with making the spatial distribution smooth in the Gridded Scale in Figure 3, because there are some regions with clear boundary in the northern Africa and in the northern North America. Meanwhile, many methods of spatial downscaling are average methods, how did authors ensure the accuracy of results between adjacent regions. In addition, how did authors eliminate the impact of some cross-regional water transfer projects and water facilities on the spatial downscaling of water withdrawals and consumption.**

**Response:** While total withdrawals can vary largely between adjacent regions, these are often strongly related to total population or total land area, so withdrawals per grid cell may actually be very similar near the border. Additionally, most of a region’s large withdrawals usually belong to a small number of cells (such as those containing cities), and as a result the distribution at the border appears fairly smooth without additional consideration of boundary effects on our part, although these effects likely do exist in some manner. Cross basin transfer is not modeled as part of our downscaling, as we are looking at location of water demand, not physical water supply.

**Edits Made:**

* In Methods, added “While many adjacent regions differ largely in total water demand, most of this demand is directly related to total population or land area, and often concentrated in a few cells, such as those containing cities. As a result, spatial distributions at the border are smoother than they appear on the region scale map, without additional consideration of the boundaries by Tethys.”

# R2 Comment 7

**R2 Comment 7: L361-370 Authors did not validate the accuracy of the water withdrawals and consumption. Although the spatial and temporal downscaling methods were validated, the results of this validation verification are obvious according to the methods presented earlier by authors. It is recommended that authors increase the comparison results with statistics published by government agencies to prove the accuracy of the historical part of this dataset.**

**Response:**

**Edits Made:**

# R2 Comment 8

**R2 Comment 8: L375-402 Authors explains some differences between this dataset and the other two datasets, but do not explain in detail whether these differences will affect the accuracy of the results. Furthermore, although the authors list the scatterplots of the validation with other datasets on a meta-repository, they do not give detailed explanations. The correctness of the validation results is not known only from Figure 5 and 6. Using the data of only one year (2010) is not very convincing. It is recommended that the author extend the period (e.g., 2005-2010) of validation results of this dataset and other datasets in detail to prove the stability of the prediction data.**

**Response:**

**Edits Made:**

# R2 Comment 9

**R2 Comment 9: There are some regions with clear boundary in the Gridded Scale of Figure 3, such as in Sahara in Africa and in northern North America. What is the reason for this situation?**

**Response:** These clear boundaries area a byproduct of the areal-weighting method used in the proxy population data <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/methods/method1>, meaning population counts were only available at the scale of those visible regions, and population was assumed to be uniformly distributed among cells (in proportion to land area). These clear boundaries can also appear in the irrigation sector when using the crop distribution directly as a proxy would lead to a situation that would imply more irrigated land than total cell area, so the remaining demand is reallocated uniformly across the rest of the region, resulting in clear boundaries.

**Edits Made:**

* In Spatial Downscaling – Non-Agriculture: “Large groups of cells with the same value are a byproduct of the areal-weighting method used in the proxy, where coarse census data are evenly distributed.”

# R2 Comment 10

**R2 Comment 10: In the formula 15 and 16, two identical “tempmax” appear in the denominator.**

**Response:** The corrections to the formulas have been made.

**Edits Made:**

* Replaced with be (temp\_month - temp\_mean)/(temp\_max - temp\_min)

# R2 Comment 11

**R2 Comment 11: There is no label in the first figure of “Temporal” of Figure 4b.**

**Response:** The words have been added.

**Edits Made:**

* TODO: ~~photoshop~~ inkscape the words “By Sector” onto the figure.

# R2 Comment 12

**R2 Comment 12: L238 What is the meaning of “****pρ values”?**

**Response:** This formatted wrong, and has been corrected to “ρ values,” which are proportions for electricity referred to in the formulas earlier.

**Edits Made:**

* Removed the extra p