



Joint Global Change Research Institute (JGCRI)
Pacific Northwest National Laboratory (PNNL)
5825 University Research Ct,
College Park, MD 20740

27 February 2023

Review 2: 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: 21 December 2022

Dear Chief Editor Guy Jones,

We would like to thank you and the reviewers for your invaluable feedback and comments in this second round of reviews. 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.

Thank you for the guidance,

Sincerely,

Zarrar Khan

Contents

List of All Reviewer Comments	3
R2 Comment 1	4
R2 Comment 2	4
R2 Comment 3	5
R2 Comment 4	5
R2 Comment 5	5
R2 Comment 6	6
R2 Comment 7	6
R2 Comment 8	6
R2 Comment 9	7

List of All Reviewer Comments

The following is a list of all reviewer comments received. Each comment has been addressed in the text as described in the corresponding section. Comments are numbered by referee as R1 for Referee 1, R2 for Referee 2, and E1 for Editor.

In this 2nd round review only comments from Referee 2 (R2) were received.

R2 Comment 1: According to the Data DOI (GCAM, Water use data used as inputs for Tethys, v4.3.chen, <https://data.pnnl.gov/dataset/1322442>) shown in Table 3, I did not find the "Region/Basin Scale" data mentioned in Figure 3. So I cannot see the original input data. The format of the GCAM is all .nc, and their dimensions are 3600*7200. They can be considered that their spatial resolution is 0.05°. Why does the author want to downscale to 0.5°? So, the author should specifically introduce the source of the input data and data pre-processing in the manuscript so readers can understand it better.

R2 Comment 2: This brings me to my second point. Assuming that the authors can explain the region/basin data source. However, the spatial or temporal downscaling method is similar to calculating the values of various water withdrawals based on the share of each grid cell, although this method has been verified in other studies. According to the author's description in the Technical Validation section, I understand that this is like, in terms of space, water withdrawal and consumption in a region/basin are assigned in proportion (e.g., according to the area ratio), and then adding up the assigned values and comparing them to the original data in the validation process. The result is undoubtedly satisfactory, but do the assigned values have practical meaning? The author emphasizes in the manuscript and review that this manuscript involves a data set, but I have doubts about the Technical Validation section (results shown in Figure 4). It is recommended that the authors carefully consider the details of the Technical Validation section so that more researchers can measure future global water scarcity based on this dataset.

R2 Comment 3: L53 Define WRI.

R2 Comment 4: L242 There is no unit after "18".

R2 Comment 5: L248-262 According to the authors' description, "The formula is modified for cells with low annual HDD or CDD as described in Huang et al., 2018". But only HDD is mentioned in the conditions of Formula 19-20, and there is no CDD (only CDD is mentioned in the conditions of Formula 21-22, and there is no HDD), so are the conditions to CDD (HDD) considered? Is it the same as Huang et al., 2018? The authors are requested to indicate.

R2 Comment 6: L263 What is meant by "their respective thresholds".

R2 Comment 7: L369 Please provide the reference for "Graham et al. 2020".

R2 Comment 8: L557 Please update your references.

R2 Comment 9: L569 Please correct "van" to "Van".

R2 Comment 1

R2 Comment 1: According to the Data DOI (GCAM, Water use data used as inputs for Tethys, v4.3.chen, <https://data.pnnl.gov/dataset/1322442>) shown in Table 3, I did not find the "Region/Basin Scale" data mentioned in Figure 3. So I cannot see the original input data. The format of the GCAM is all .nc, and their dimensions are 3600*7200. They can be considered that their spatial resolution is 0.05°. Why does the author want to downscale to 0.5°? So, the author should specifically introduce the source of the input data and data pre-processing in the manuscript so readers can understand it better.

Response: Thank you for pointing this out. Upon checking with the authors of the GCAM paper we see that the link provided in that paper was out of date. We worked with the authors to re-post the data so it is up-to-date and accessible. We fixed the link to point to the correct databases. We also provide a script with details on how to extract the water demand data from the original data source so users can explore it themselves if desired.

This data set is by GCAM region/basin for every year in 5 year intervals from 2010 to 2100 so it needs to be downscaled to 0.5° resolution to match the underlying population data and other proxies used for the downscaling. With Tethys you can directly connect to these GCAM databases to downscale water demands to the gridded resolution.

Edits Made:

- Link to raw GCAM databases updated in Table 3
- Added script to meta-repo so users can extract data if desired to view the raw data from the original databases with a note in the text:

* Note: For users wanting to explore the water consumption and withdrawal data directly from the original GCAM databases we provide a short R script at:

https://github.com/JGCR/khan-etal_2022_tethysSSPRCP/blob/v1-pre-publish/scripts/extract_water_data.R (<https://doi.org/10.5281/zenodo.7636762>)

R2 Comment 2

R2 Comment 2: This brings me to my second point. Assuming that the authors can explain the region/basin data source. However, the spatial or temporal downscaling method is similar to calculating the values of various water withdrawals based on the share of each grid cell, although this method has been verified in other studies. According to the author's description in the Technical Validation section, I understand that this is like, in terms of space, water withdrawal and consumption in a region/basin are assigned in proportion (e.g., according to the area ratio), and then adding up the assigned values and comparing them to the original data in the validation process. The result is undoubtedly satisfactory, but do the assigned values have practical meaning? The author emphasizes in the manuscript and review that this manuscript involves a data set, but I have doubts about the Technical Validation section (results shown in Figure 4). It is recommended that the authors carefully consider the details of the Technical Validation section so that more researchers can measure future global water scarcity based on this dataset.

Response: Thank you for your comments on the validation. As described in the text (lines 102 to 109 and line 370) the underlying data at the regional/basin scale comes from Graham et al. 2020 in which it

has been validated at that scale. In this study we show two kinds of validations. The first shown in Figure 4 is to check if mass is conserved during the downscaling as we go from regions to grids and from years to months. The second shown in Figures 5 and 6 show that the spatial and temporal distributions match those from other studies in the past. As explained in the paper (lines 413 to 415) we are not trying to match the results from these other papers exactly since they use different methods, sectoral categorizations, and proxies. We show that overall patterns for spatial and temporal distributions reasonably match. Additionally, both those papers have highlighted several of their own limitations, which would make cell-by-cell differences reflect methodological differences more so than differences from actual water usage on the ground.

Having said that we agree that a detailed validation of the gridded spatial and temporal distributions against observed data is very important but beyond the scope of the current study. We are in the process of conducting this validation in a series of follow-up papers that will focus on detailed validation for particular regions at a time. The first of these compares Tethys downscaled outputs reaggregated to the USGS county scale in the U.S. since this is the scale at which we were able to obtain reliable observational data. We add a note about this in the section discussing planned developments.

Edits Made: L475: Text added to planned developments-

“Comparing gridded outputs against observational data for individual sectors and regions where data is available.”

R2 Comment 3

R2 Comment 3: L53 Define WRI.

Response: Updated to expand Acronym.

Edits Made: L53 – Text updated to expand Acronym.

R2 Comment 4

R2 Comment 4: L242 There is no unit after "18".

Response: Updated to add units.

Edits Made: L242 – updated to added units for “°C” after 18.

R2 Comment 5

R2 Comment 5: L248-262 According to the authors' description, "The formula is modified for cells with low annual HDD or CDD as described in Huang et al., 2018" . But only HDD is mentioned in the conditions of Formula 19-20, and there is no CDD (only CDD is mentioned in the conditions of

Formula 21-22, and there is no HDD), so are the conditions to CDD (HDD) considered? Is it the same as Huang et al., 2018? The authors are requested to indicate.

Response: We modify the text slightly to make this clearer. As indicated the modifications are made when HDD or CDD are less than certain thresholds.

- In Formula 19-20 when HDD is less than 650 the HDD term is removed leaving only CDD (which is what you see in 19-20).
- In Formula 21-22 when CDD is less than 450 the CDD term is removed leaving only HDD (which is what you see in 21-22).
- Yes, this is the same method as used in Huang et al., as mentioned on line 247: “The formula is modified for cells with low annual HDD or CDD as described in Huna et al. 2018”

Edits Made:

- L251: Text in red added to clarify: “the HDD term is removed (leaving only CDD)”
- L257: Text in red added to clarify: “the CDD term is removed (leaving only HDD)”

R2 Comment 6

R2 Comment 6: L263 What is meant by "their respective thresholds".

Response: This refers to the thresholds mentioned in before the formulas 19-22 above. We have expanded the text to explicitly describe these thresholds.

Edits Made:

- L265-266: Text in red added to clarify: “When annual HDD and CDD are both below their respective thresholds (<650 for HDD and <450 for CDD)”

R2 Comment 7

R2 Comment 7: L369 Please provide the reference for "Graham et al. 2020".

Response: Reference added.

R2 Comment 8

R2 Comment 8: L557 Please update your references.

Response: Updated to version 1.2

R2 Comment 9

R2 Comment 9: L569 Please correct "van" to "Van".

Response: All mentions of “van” updated and corrected in text and references.