

Water Availability Review 2 Responses

We would like to thank the reviewers once again for their time taken to evaluate the revised manuscript. Our responses to their individual comments are included below.

Reviewer 1

1. In the Responses to Reviewers document, the authors clearly describe the motivation for the paper. Two very helpful sentences state the aim as "...to enable decision makers to benefit from the additional information they might obtain from using large ensembles of coarse resolution ESMs", clarifying that decision makers "wish to understand how the local [downscaled?] data they employ for decisions are situated within a cascade of uncertainty stemming from model and scenario selections."

Those clear phrases should be captured a little better in the manuscript. The last sentence of the abstract points toward this as a conclusion, but not as a driving motivation for the study, so that could be added to the abstract.

The third sentence of the abstract has been revised to read "However, the models are seldom evaluated for their ability to reproduce metrics that are important for practitioners, such that practitioners can situate higher-resolution model outputs within a cascade of uncertainty stemming from different models and scenarios."

Also, line 98 includes a third bullet (as a motivation for the study) as looking at projected changes, which the revised paper no longer includes. That can be replaced by something like one of the quoted phrases above.

The bullet has been removed and replaced with "the range of CESM2 structural uncertainty and internal variability for these metrics"

2. Lines 377-381 is a residual paragraph that had been part of a section that was deleted from the paper (prior section 5.2 on climate projections). This should be removed.

Thank you for noting this, the paragraph has been removed.

Reviewer 2

My main point is that instead of showing results for all basins, you have chosen some examples – which in general is fine. However, some parts of the results are based on data that is not presented in the manuscript (see specific in-line comments below). I suggest to include that data (e.g. in the Appendix if you don't want to blow up the Figures) and refer to those Figures in the Results.

Thank you for this suggestion. We have now included the following Supplementary figures and pointed to them within the main text

- Average number of wet days per month (Fig 3a) remaining 16 basins - S1
- Variability in mean annual rainfall on wet days (Fig 3b) remaining 16 basins - S2
- Number of heavy rain days and Proportion of Annual total from heavy rain days (Fig 5) for 4 additional basins - S3
- Annual Runoff (Fig 7a) for 8 additional basins - S4
- Cumulative Runoff Hydrographs (Fig. 7b) for 8 additional basins - S5

p.1 line 16: Explaining the term/job “water (resource) manager” and what the respective responsibilities are might help for readers from different fields.

To minimize confusion, we have removed the first instance of “water resource manager”, and similarly changed “water resource metrics” to “water management metrics” in the abstract to make the text more consistent with the article title.

p.3 line 98: I suggest to remove this motivation point (“how such metrics are projected to change”), since apparently you removed that part of the analysis during the revision.

This has been removed, and an additional motivation point added as suggested by Reviewer 1.

p.5 line 150: You have added some text to accommodate a reviewer question, however this addition needs to be explained, in order to not confuse readers. → I suggest adding sth like “Pierce et al. (2021) provide an updated version of the Livneh dataset, by ... ” to explain this.

This now reads: “Pierce et al. (2021) provided an update to the Livneh data set to address time adjustments that result in an underestimation of the most extreme daily precipitation totals and resultant runoff and flood potential (Pierce et al. 2021). However, as we are also interested other measures of in precipitation and in runoff minimat, we did not employ the updated gridded observations.”

p.6 line 187: “expected once per decade” – is this still the case with the current rate of climate change – or has it changed already?

This is a good point. We have added a sentence “Stationarity was assumed over the climatological period for the purposes of these analyses, acknowledging that changes may have already occurred in the frequency of these events.”

p.7 line 200: Should say "Appendix C".

Changed.

p.7 line 207-210: Why is there no figure for regions where there are larger differences? Would be interesting to compare. If the discrepancies are especially large in months with snow fall and or melt, I would assume that the missing processes related to snow accumulation and melt are playing a large role here.

This is a good point, we have added Supplemental Figure S1 for the other regions as noted above.

p.9 line 229-233: Again, here you are describing a results that you have not shown.

Supplemental Figure S2 has been added as noted above.

p.10 line 252: They are "defined" the same (both via the Q95 and Q99) I assume, but calculated separately and compared, correct?

This is correct. The sentence has been amended.

p.10 line 268-269: You write that the agreement is good (qualitative subjective judgement I guess), however the P95Tot and its variability is consistently overestimated by CESM2, compared to Livneh. You might want to mention that.

"Subjective" added to the sentence regarding good agreement. Supplemental Figure S3 has been added to illustrate the differences in other basins. However, we disagree that P95Tot and its variability is consistently overestimated - some basins are overestimated, others are well captured and this is already reported in the text.

p.12 line 287: Should say "Appendix D".

Changed.

Figure 3: Interannual spread for Livneh (pink area) is missing in the legend. You might want to use different shadings for the CESM2 range for the top (interannual + ensemble) and bottom panels (only ensemble), because they are not the same – and quickly mention this in the text to guide the reader.

Shading altered and legend updated.

Figure 5: Which single ensemble member do you plot and why? Why not plot the ensemble mean/median here?

The ensemble member was selected at random. This better shows the interannual variability, which is smoothed out and meaningless when the ensemble mean/median is used. This was in the text relating to Figure 3.

Table A1: “Metrics in bold are presented in this article.” - I cannot find any bold text except for the headings.

The formatting has been amended to highlight the metrics used in the article in bold.

Reviewer 3

The authors present a work to evaluate Earth System Models by comparing its precipitation and runoff with that from gridded daily observations and VIC simulations. The motivation of this study is interesting to me, and can be helpful for the ESM users in hydrology. My major concerns come from the depth of the comparison. As we know, all the models may generate some results that can be comparable to but may be greatly different from the truth. A simple comparison cannot well address our concerns on their performance, especially for a water manager perspective, who may already has some information from remote sensing, in situ monitoring or other modeling.

This is a good point, we have added to the conclusions:

“The present evaluation is also only the first step in evaluating ESM performance. Additional research is needed to support water managers placing these results and their uncertainty in the context of additional observational data (such as remote sensing) that may already be available to them. “

(1) Only P and R are investigated, and how about other water balance compartments, such as soil moisture, evapotranspiration.

The reviewer makes a good point. At the beginning of this project, many water balance components were evaluated. In the interests of keeping this manuscript to a readable size we opted to focus on the most interesting results. As the available soil moisture and evaporation data were also from the derived Livneh VIC simulation, these results were not considered to present additional information for the benefit of the reader.

(2) The comparison between ESM and Livensh are displayed in a statistical way for some selected regions. How about direct validation at grid scale or in situ scale? Such information (or may be some cases) can better show the performance of ESM.

Direct validation at grid scale isn't appropriate for such a coarse scale, this is explained within the text and demonstrated in the figures in Appendix C seasonal mean precipitation and Appendix D Maximum daily runoff.

(3) The simulated runoff is used for comparison. Can in situ observations at river flow gauges be available for comparison? If so, the findings will be more reliable.

Again, comparing in situ observations to the coarse scale model results is not appropriate as described in the text. Furthermore, the results are of catchment runoff, rather than river flow and so a direct comparison cannot be made.

(4) Figure 3 shows the large deviations among different solutions in ESM. This large uncertainty should be highlighted for its use in practice.

Thank you for this suggestion, we have added a sentence at the end of the paragraph describing Figure 3:

The figures also highlight the scale of model (structural and internal variability) uncertainty present in the ensemble. As noted in previous sections, water management decision-makers are aware of the potential scale of uncertainty and expressed a desire for the full ensemble range to be presented to them instead of ensemble means.