

Editor

Dear authors,

While the major concerns raised in the first review round have been cleared, there is a number of minor points to be addressed. I invite you to respond to these points and submit a revised manuscript version.

We thank the editor and both reviewers for their useful suggestions, please see below a detailed response (in green) and the modified parts of the manuscript (in blue) to all the provided comments.

Two editorial comments:

- I recommend using "elevation" instead of "altitude" throughout in text and figures, see McVicar TR & Körner C, 2013. On the use of elevation, altitude, and height in the ecological and climatological literature. *Oecologia*, 171(2), 335–337. <http://dx.doi.org/10.1007/s00442-012-2416-7>

We thank the editor for this useful advice, the related parts were rephrased.

- The figures (and text in them) are still slightly blurred, make sure to upload high-resolution images for the final article.

We still think that it is related to pasting it to the word-file, but we will double-check it during the typesetting process.

Please also address the notifications to the authors from review file validation.

Kind regards,
Daniel Viviroli

Editorial team:

1. With the next revision, please add the section "Correspondence to:" into the title page of the *.pdf manuscript. Please see more: <https://www.hydrology-and-earth-system-sciences.net/submission.html#templates> / Technical instructions for MS Word and compatible formats /.

We have added the information as demanded.

2. Please ensure that the colour schemes used in your maps and charts allow readers with colour vision deficiencies to correctly interpret your findings. Please check your figures using the Coblis – Color Blindness Simulator (<https://www.color-blindness.com/coblis-color-blindness-simulator/>) and revise the colour schemes accordingly.

We are thankful for this advice. The color ramps of two figures have been adjusted to increase readability.

3. Please make bold the captions of all figures for more clarity with the next revision.

Thanks for the advice, all captions should be bold now.

Reviewer #3

I have read the revised manuscript as well as the authors' responses to the two referee reports. Overall, the manuscript is suitable for publication in Hydrology and Earth System Sciences, and the major concerns mentioned by the two other referees seem to be cleared out.

We'd like to thank reviewer #3 for this positive evaluation of our study. Detailed response to specific comments is provided below.

However, the following smaller issues and questions should be addressed before publication:

- In your response to point 4 by referee #2 ("Why is the ARSO-d data considered as the benchmark?"), you mention that you also wanted to implement a discussion in the revised manuscript to more clearly address questions in this direction. I may have missed it, but can you indicate where that has been added, or complete if it hasn't been done yet?

We'd like to thank reviewer #3 for pinpointing the matter. The argument for selecting this product for precipitation comparison as the benchmark is provided in the following excerpts:

L170-173: For the validation of the precipitation reanalysis products (PRP), a regionalized daily precipitation data set from the Slovenian Environment Agency (ARSO) is used, from now on referred to as ARSO-d. ARSO-d has a spatial resolution of 1 km raster width length and is available from 01.01.1981 to 31.12.2010. It is based on the regionalization and upscaling of station-based precipitation measurement into spatially and temporally consistent dataset.

L211-212: Hence, for the validation of the PRP the ARSO-d data is used, because it is the only available data set with a sufficient spatial coverage to enable more robust comparisons with the PRP.

L215-217: The PRP are aggregated to daily values to enable the comparison with ARSO-d, since no hourly spatial rainfall product for Slovenia is provided by ARSO.

- Conclusions: Can you make a statement on the extent to which your findings can be applied to other regions, i.e., to what extent can they be generalized, and are limits to that?

We thank the reviewer #3 for this comment. Generalisation is always tricky for rainfall due to its local, regional and hydro-climatological characteristics. We added the following comment in the Conclusions section:

A generalisation of these conclusions is limited due to the regional differences of rainfall processes and the ability of the reanalysis models to represent them. However, the more similar areas of interest are in terms of hydro-climatology, the more likely similar findings can be expected.

- L53: Amend incomplete sentence "... as shown for e.g. by e.g. Gebremichael ..."

Thanks, we have completed the sentence now as suggested.

- L61: Do you refer to a commercial context for a specific reason? Otherwise I would rephrase more generally towards applications.

Thanks, we have removed the 'commercial'-part.

- L229: Clarify or replace the term "catchment-intern analyses", it is used only here.

Thanks, we have rephrased it to 'catchment-specific', which is more appropriate here.

- L228: Clarify the "11 years (Jan-Aug) and 10 years (Sep-Dec)" in parentheses: Why these months?

The period of data availability is 01.01.2009-31.08.2019. So there are 11 years available for months Jan-Aug, but only 10 for Sep-Dec. We agree that this information can be misleading and we removed it. The period of data availability remains in the text and we think it will be sufficient for the reader. Thanks.

- L273: You write that "both calibration and validation periods represent similar climate, and hence catchment conditions. " I would not put it that way, as there are other catchment conditions beyond climate.

We thank reviewer #3 for the recommendation and modified this specific part of the manuscript as

It is very well established that split-sampling is recommended if both calibration and validation periods represent similar climate, soil properties and land cover conditions, i.e. consistent catchment conditions over time.

- L275: Please be more specific in this sentence, it sounds rather generic: "In addition, different approaches are available as calibration and model evaluation strategies, and some studies support the split-sampling approach (e.g., Perrin et al., 2003)."

We thank the reviewer #3 for the recommendation and modified this part of the manuscript as:

Nonetheless, in the data used for this study there are minimal fluctuations within the selected periods in terms of very wet or dry periods. In addition, amongst the various methodologies for calibration/validation period selection found in literature, some studies support the split-sampling approach (e.g., Perrin et al., 2003).

-L296: I would mention here in addition that a KGE value of more than -0.41 (i.e., $1-\sqrt{2}$) indicates model efficiency better than the mean flow benchmark according to Knoben et al. (2019).

We thank reviewer #3 for this literature hint, we added the information as follows:

Knoben et al. (2019) point out that mean flow as benchmark is already outperformed with KGE values >-0.41 .

- L426: The figure legend seems to be mixed with the main text.

Yes, we have separated it now. Thanks for this remark.

- L440ff: Clarify what you mean by "ad-hoc calibration".

We thank reviewer #3 for the recommendation and added the following segment on L478

Ad-hoc calibration refers to the implementation of the Mitchel algorithm with each new insertion of reanalysis within the model; that is, instead of keeping the model calibrated using observations, its ingrained parameters are re-optimized in order for the reanalysis to converge closer to discharge observations.

- L454: "No significant trends are observable in relation to latitude." Do you use "significant" in a statistical sense? If yes indicate the test done and the p-value used, otherwise change the wording.

We thank reviewer #3 and replaced significant with substantial.

- L476: Complete "distributed" at the end of sentence "PRP is of little added, value compared to a spatially distr."

We have corrected the flaw, thanks for pointing it out.

- L500: Can you give some indications why there are overestimations > 1000 m a.s.l., and why deviations are smaller for catchments below this elevation?

Thanks for this useful comment. We think that the reanalysis models simulate on a certain elevation and got 'corrected'/adapted to the true elevation afterwards. With increasing elevation this correction becomes less accurate, hence higher deviations occur. However, this is an impression from our findings and others from the literature, but further investigations are needed to better evaluate this behavior.

- L503: "the *following* conclusions"?

Thanks, rephrased as suggested.

- L510: "show*s*"

Thanks, the 's' was added.

- L521: You conclude that bias-correction should be implemented, but on L112ff you mention that an advantage of the present research is the examination of raw data. Can you clarify?

We thank reviewer #3 for pointing that concern: in L531 it is emphasized that an evaluation of the reanalysis products that undergo a bias-correction procedure would be beneficial to further assess their applicability in r-r applications.

- Equation 1: Should it be "RC[PRP, n]" and "RC[Obs, n]" instead of "RC[PRP]" and "RC[Obs]" (both in numerator and denominator)?

Yes, indeed we refer to station-based rainfall characteristics. We added the 'i' for all terms.

- Table 1: Use "discharge regime" instead of "water regime" for consistency?

We have rephrased the header as suggested.

- Figure 2: 1) The rivers do not seem to be in alphabetical order, but rather the maxima seem to increase from left to right. Please clarify and amend. 2) I recommend mentioning in the legend that the values are displayed on a logarithmically scaled y-axis. 3) Using different symbols for min/mean/max would improve readability of the figure.

We thank reviewer #3 for his thoughts on the figure, all three suggestions were incorporated.

- Figure 3: Is there a reason for the larger paragraph spacing between Mislinja, Dravinja and Radovna?

No, there was no reason, spacings are uniform now. Thanks for pointing to this issue.

- Figure 6: The figure legend isn't clear to me: "Deviations of areal rainfall extreme values (5- and 50-

year return periods) between ERA5-Land and COSMO-REA6 over all 16 catchments." ? But isn't ARSO-d the reference?

Yes, the disaggregation of ARSO-d using the closest rain gauge is the reference. The caption was misleading, we have changed it to (also for Fig5):

Deviations of areal rainfall extreme values (5- and 50-year return periods) of ERA5-Land and COSMO-REA6 in comparison to observations over all 16 catchments.

- Figure 9: It doesn't seem correct to show negative values in white. KGE has a value range from inf to 1, so values below 0 do not fall outside of that range. Values <0 should rather have an own class in the color bar, and use a color that shows it's a continuum, e.g., dark red.

We thank reviewer #3 for this suggestion, we did as suggested (also for Fig. 12).

- Figure 11: I recommend adding a 1:1 line and making the figure square for better interpretation.

We thank reviewer #3 for this suggestion, the straight line was added.

- Figure 12: This is for the recalibrated model, isn't it? Mention this in legend for clarity.

We have added this information in the caption. Thanks.

Reference:

W. J. M. Knoben, J. E. Freer, and R. A. Woods, 2019: Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrol. Earth Syst. Sci. 23, 4323–4331. doi:10.5194/hess-23-4323-2019.