Replies to reviewer #2

“Benchmarking convection-permitting climate simulations for hydrological applications: A comparative study of WRF-SAAG and observation-based products”

Sofía Segovia, Pablo A. Mendoza, Miguel Lagos-Zúñiga, Lucía Scaff, and Andreas Prein

We thank the reviewer for his/her time, revision and suggestions to our paper. We provide responses to each individual point below, and how we will address the main comments of the reviewer. For clarity, comments are given in black italics, and our responses are given in plain blue text.

This manuscript evaluates precipitation and temperature from WRF-SAAG, CR2MET, and RF-MEP against in-situ stations across continental Chile. The topic is important and relevant. Below are my major and minor comments.

Major Comment:

1. Since WRF-SAAG is not expected to reproduce specific events, I am unclear about the motivation for this evaluation. Is WRF-SAAG typically used as forcing for hydrological models to obtain hydrological simulations? If so, is this recommended practice? If not, the value of evaluating individual events is limited. In that case, comparing climatological characteristics across datasets might be more meaningful.

The motivation for our evaluation is not to assess event-scale accuracy, but rather to examine the skill of the WRF-SAAG dataset in representing hydroclimatic variability at the daily time scale. This assessment constitutes a necessary step to determine whether WRF-SAAG provides sufficiently realistic precipitation and temperature fields to be considered for subsequent hydrological modeling experiments. In the revised version of the manuscript, we will make this motivation explicit in the Introduction section, clarifying that the purpose is to assess the general suitability of WRF-SAAG as a meteorological dataset and its potential for hydrological modeling.

Additionally, we will highlight that WRF-SAAG simulations are forced with ERA5 reanalysis, so it is expected to reproduce large-scale precipitation events, and explicitly resolve mesoscale processes such as convective precipitation. WRF simulations are often used in studies involving hydrological modeling (e.g., Wagner et al., 2016) and case study analysis (e.g., Naabil et al., 2017), among others. However, the WRF-SAAG simulations have not been exhaustively evaluated in South America, and their skill as meteorological input forcing for hydrological applications (e.g., Xie et al., 2025) remains unexplored. We will explicitly declare these objectives and opportunities in the introduction.

2. As this is an evaluation study, the manuscript would benefit from clearer take-home messages. For example: Which dataset should be preferred under certain conditions or in specific regions? Are there areas where none of the datasets are recommended?

We agree that the manuscript would benefit from clearer take-home message. In the revised version, we will restructure the Conclusions section to provide a more explicit synthesis of the main findings. This includes identifying the strengths and limitations of each dataset under different hydroclimatic and regional conditions.

These clarifications will be supported by the new analyses included in the revision, such as the intercomparison of gridded precipitation datasets and the evaluation of hydrological signatures, which together allow for more robust and regionally relevant conclusions regarding the suitability of each dataset for hydrometeorological characterizations.

3. Figures 3–5: Consider presenting KGE values as boxplots, since the authors mention medians of the KGE frequently. The spatial maps make it difficult to assess the median.

In the revised version of the manuscript, we will provide additional summary metrics, figures and/or tables to help interpreting the results. Additionally, the figures will be provided in high resolution, so readers can zoom-in to specific regions.

4. Line 285: When precipitation events > 1 mm/day are considered, CR2MET performs best overall, but for thresholds > 5, > 10, and > 20 mm/day, RF-MEP performs best. Could this pattern be influenced by the precipitation distribution? For example, are events between 1–5 mm/day dominant when precipitation events > 1 mm/day? If so, it might cause RF-MEP's superior performance at higher intensities to be diluted when all events > 1 mm/day are included.

We agree that the apparent change in relative performance between precipitation products across precipitation thresholds may be influenced by the underlying distribution of precipitation intensities.

To address this, in the revised version of the manuscript we will add and evaluate the number of observed precipitation events at the station level and in each gridded product, to complement the threshold-based analysis. This will help clarify how event frequency and intensity distributions influence the comparative performance among datasets.

4. Line 262: Please specify the formula used to estimate PET and list the variables involved.

PET was calculated using the formulation proposed by Oudin et al. (2005), as implemented in the R package airGR (Coron et al., 2023), considering the latitude of the centroid of each elevation band. In the revised manuscript, the exact equation and all input variables have been included to ensure reproducibility.

Minor Comments:

6. *Abstract: WRF-SAAG data are available until 2021. Why was the evaluation restricted to 2001–2018?*

The analysis period (2001–2018) was limited by the availability of the RF-MEP dataset, which covers the period 1983–2018 (Baez-Villanueva et al., 2021). We explicitly justify the choice of analysis period at the end of section 4.1:

"Importantly, all the assessments presented in this study were conducted for the period April/2001-March/2018, which corresponds to the common temporal coverage of the three forcing datasets."

7. Line 120: Consider adding polygons to the maps to delineate the four geographical units, which would help international readers.

In the current version of the figure, the four geographical units (macrozones) are delineated by horizontal red lines, and the name of each macrozone is indicated on the left side of the panel. We have ensured that this is more clearly described in the figure caption to guide readers' interpretation.

8. The first part of the abstract reads like that WRF-SAAG was evaluated against CR2MET and RF-MEP, but in fact all three datasets were compared against in-situ observations. Please revise for clarity.

We have revised the abstract to clarify that all three datasets—WRF-SAAG, CR2MET, and RF-MEP—were evaluated against in-situ station observations:

"In this paper, we evaluate the quality of WRF-SAAG daily precipitation and daily maximum and minimum temperature simulations using observations from meteorological stations over continental Chile for the period 2001–2018. The results are compared with the performance of two widely used gridded meteorological products – CR2MET and RF-MEP – which combine reanalysis data with in-situ measurements."

9. Line 135: What do "daily observations" refer to? Are they discharge (Q) and precipitation (P)? Does this also include temperature?

We refer to daily discharge observations. We have modified the text to clarify this:

"The selected catchments fulfill the following criteria: (i) at least 80% coverage of daily discharge observations during the period April/2001 – March/2018…"

10. Line 210: It seems the reference should be "1b (1c)" instead of "1c (1d)."

Thanks for catching this! We have corrected the references in the revised manuscript.

11. Line 255: The DEM data should be introduced in the Data section.

Although it is used to derive elevation bands in the analysis, it is not a hydrometeorological dataset and therefore does not fit within the "Hydrometeorological datasets" section.

12. Equation 8: Please clarify what Q and 1/Q represent in the KGE' formulation.

Q is daily discharge, and 1/Q is the reciprocal of daily discharge. We have modified the text as follows to explain this:

"where Q is daily discharge, 1/Q is the reciprocal of daily discharge, and KGE' is the modified Kling-Gupta efficiency (KGE', Kling et al., 2012)".

We have also added the following lines of text in section 4.3:

"Finally, KGE'(Q) (KGE'(1/Q)) denotes the KGE obtained from comparing daily time series of simulated and observed Q (1/Q)".

13. Line 220: When referring to precipitation events > 1 mm/day, do these include events exceeding 5, 10, and 20 mm/day, or are they limited to events between 1–5 mm/day? Please clarify.

In our analysis, we are including all events with magnitude larger than the specified threshold, without limiting them to the next lower limit. We have clarified this in the revised version to avoid ambiguity:

"We used metrics formulated from contingency tables to assess the ability of the datasets to replicate historically observed daily precipitation events exceeding thresholds of 1, 5, 10, and 20 mm/d (i.e., events >1, >5, >10, and >20 mm/d)."

14. Line 297: This section appears to discuss only summer results. Why are "all seasons" mentioned here?

The intention was to highlight that, although Figure 3 focuses on summer (DJF), the Southern macrozone consistently shows high KGE_T median values (≥ 0.65) not only in summer but also across the other seasons.

We have modified the text to clarify this:

"Figure 3 shows the spatial distribution of KGE_T (and its components) for summer (DJF) daily precipitation estimates from WRF-SAAG, CR2MET, and RF-MEP. The highest KGE_T values in summer are obtained in the Southern macrozone, where median KGE_T values reach ≥ 0.65, and remain at or above this level throughout all seasons"

References:

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Mendoza, P. A., McNamara, I., Beck, H. E., Thurner, J., Nauditt, A., Ribbe, L., & Thinh, N. X. (2021). On the selection of precipitation products for the regionalisation of hydrological model parameters. *Hydrology and Earth System Sciences*, *25*(11), 5805–5837. https://doi.org/10.5194/hess-25-5805-2021

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., Michel, C., Andréassian, V., Bourgin, F., Brigode, P., Le Moine, N., Mathevet, T., Mouelhi, S., Oudin, L., Pushpalatha, R., & Valéry, A. (2023). *Suite of GR Hydrological Models for Precipitation-Runoff Modelling, R Package version 1.7.4*. https://cran.r-project.org/web/packages/airGR/

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424–425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Naabil, E., Lamptey, B. L., Arnault, J., Olufayo, A., & Kunstmann, H. (2017). Water resources management using the WRF-Hydro modelling system: Case-study of the Tono dam in West Africa. *Journal of Hydrology: Regional Studies*, *12*, 196–209. https://doi.org/https://doi.org/10.1016/j.ejrh.2017.05.010

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, *303*(1–4), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

Wagner, S., Fersch, B., Yuan, F., Yu, Z., & Kunstmann, H. (2016). Fully coupled atmospheric-hydrological modeling at regional and long-term scales: Development, application, and analysis of WRF-HMS. *Water Resources Research*, *52*(4), 3187–3211. https://doi.org/https://doi.org/10.1002/2015WR018185

Xie, K., Li, L., Chen, H., & Xu, C.-Y. (2025). Assessing the performance of convection-permitting climate model in reproducing basin-scale hydrological extremes: A western Norway case study. *Journal of Hydrology*, *656*, 132989. https://doi.org/https://doi.org/10.1016/j.jhydrol.2025.132989