Suitability of 17 rainfall and temperature gridded datasets for largescale hydrological modelling in West Africa

Dembélé et al. (2020)

Review (hess-2020-68)

In their paper, Dembélé et al. explore the suitability of combining time series of rainfall and temperature from different climate products as inputs into a hydrological model. The manuscript is well structured and written, methods are robust and results are presented adequately. The research question of the possibility of combining gridded climate variables from different sources to simulate various hydrological components is relevant and timely, and I believe will be of interest for the readers of HESS. Nevertheless, I have a few comments and suggestions for the authors to consider before I can recommend the paper for publication.

Sincerely,

Nadav Peleg

Major comments

1. I found one-step in the methodology (i.e. as presented in Figure 1) to be missing. I think it will be meaningful to know how the climate variables (rainfall, temperature) from each climate products are ranked in comparison to observed data (i.e. from ground stations) before ranking the 102 input combinations based on various hydrological components. I think this step is critical to understand the presented results. For example, JRA-55 and ERA-5 yield poor correlation with Ea (Figure 8), but isn't this because they are poorly reproducing the rainfall statistics over the VRB? GSMaP-std V6 reproduces well the streamflow (Figure 3), St (Figure 4), Su (Figure 5) and Ea (Figure 8) – will this product be ranked #1 when compared to ground stations? I assume there will be a high correlation between the ranks emerging from the comparison to ground stations and hydrological outputs from the model. If this case, wouldn't it be sufficient to evaluate the best products to use in hydrological simulations simply by comparing them to the few climate stations that are available in the catchment of interest or a nearby area? This is a point for discussion.
2. The modelling experiment includes 6 years for model calibration and 4 years for model evaluation. These are very short periods, not necessarily representing well the natural climatic and hydrological variability and not necessarily guarantying a successful calibration of the hydrological model parameters. First, I suggest demonstrating with a simple graph (can be presented as SI) that the natural variability is somehow represented in your 10-year data. Second, consider adding a short discussion regarding the sensitivity (quantified) of the hydrological model parameters to the short period that is used for the model training.

Minor comments

1. Usually, when considering using gridded climate variables from climate re-analysis/other products as inputs into hydrological models the following steps are taken: (i) computing the skills (i.e. temporal dynamics, magnitude, and occurrence) of the climate variables in comparison to observed data; (ii) choosing the (individual) climate product with the best skill to use; and (iii) performing a bias correction to the climate variables to improve the fit to the observed data. I am missing a paragraph in the introduction/discussion explaining why not simply following this practice which should improve the hydrological outputs from the model.
2. Results (Figure 3, for example). 22 values are used to represent the combined performance for the calibration and evaluation periods. This is not clear to me. Why not using a single Ekg value for the entire simulation period (merging the calibration and validation periods to a single period) for each gauge, i.e. 11 values in total per combination of temperature and precipitation product? What is the logic in separating the Ekg values to calibration and validation periods?
3. Table 1. I suggest adding in the table additional column indicating if the product refers to rainfall, temperature or both. Also, please double-check the space-time resolutions reporter in the table. I think that the CMORPH-CRT product, for example, has a resolution of 8-km and 30-min.
4. The use of second-order CV is interesting, I do not recall seeing it in the context of hydrological statistics. Why use it and not simply using Pearson's CV skill? A sentence explaining the motivation is needed.
5. Figures 7 and 10. Too many box-plots are presented. Perhaps present only the median (avoid using box-plots) to compare between products and climatic zones. This will considerably reduce the size and information plotted.
6. Generalization of the results. In lines 437-438 you mentioned that: “The results can be considered valid for West Africa and regions with similar hydroclimatic and physical features. A wider generalization of the findings should be done with caution and after repeating similar evaluation studies in other places”. I do not think that you can generalize the results - they are likely to differ between locations as the quality of climate variables from different climate products differ between locations. In my view, the key message of your paper is that for each large catchment you should consider multiple sources of climate data to find the climate variables combination that is suitable for your region. The VRB is simply a case study used to demonstrate this point.