

Comment on hess-2022-72

Anonymous Referee #1

Referee comment on "Development of a national 7-day ensemble streamflow forecasting service for Australia" by Hapu Arachchige Prasantha Hapuarachchi et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-72-RC1>, 2022

Overview

This is a detailed descriptive article on the methodology followed to set up an Australian ensemble streamflow service. I commend the authors on the clear description and succinct summary of what I imagine was a very large project. I believe the submission would be of interest to readers of HESS, particularly due to the value of sharing the development of operational systems with the academic community.

The paper is understandable heavily focused on Australia. I have a couple of suggestions which would help make this work relevant to a wider audience. Firstly, I suggest that more context is given to help the reader understand the hydro-climatic context that the model is being validated over for example by including some maps instead of / alongside the box blots and table summaries (further comments on this are detailed below). Secondly, I would like to see more discussion of how the development of this service in Australia builds on, and moves forward, the development of ensemble streamflow services around the world. At present the work is situated in the Australia context and the reader is given limited insight into what is novel or new about this work or why a particular approach is suitable for Australia but may not have been used elsewhere. A wider review of existing literature would help support this.

From a technical perspective the work appears sound, an assessment of the strengths and limitations on the underlying data is made and a series of established verification metrics applied. The methodological steps are clearly documented throughout. From an open data perspective there is no indication of the source or quality of the observed rainfall and flow data. My main technical concerns come from the representation of extremes within the skill assessment. L66-85 sets the context of hydrological extremes in Australia and identifies both floods and droughts as particular water management challenges. The representation of high and low flows in forecast systems leads to different challenges at different parts of the flow regime yet the discussion around model assessment does not address this as you use evaluation metrics across the full flow regime, it is well

documented that it is much easier to model non-extreme flows. Is there also a need to consider the skill of the forecast system in identifying events that cross a high / low threshold as it is during these events that the system will have more operational value and your results may be skewed depending on characteristic of individual catchments. I appreciate the system is already operational and it may not be appropriate to add this to this paper, but it would be helpful to acknowledge this limitation and maybe identify it as a future research area.

Specific comments on the text and figures

L56 – 65 – it is unclear to me what this paragraph on continental and global scale models adds to the paper. Could you integrate this in the context of developing a streamflow model for Australia e.g. what lessons did you learn from the existing global models?

L98 – do you know of other examples of “hybrid dynamical-statistical streamflow forecasting systems” or similar set ups. It would be helpful here to identify if there is anything unique about the Australian system compared to other operational systems in other countries.

Table 3 – for those not familiar with Australian climatology it would be helpful to show some of the info in this table graphically e.g. could you include a map of mean annual rainfall distribution (or another representative variable) across Australia, it's hard to fathom this from the table, especially as the number of catchments in each drainage division are quite varied. Other information that might be interesting is an indication of the catchment response time, are you looking at steep flashy catchments or slowly responding catchments. Later on you mention ephemeral rivers as a reason for lower forecast skill, again is there a particular region where they are more common? This type of characteristics information would help readers compare your approach to approaches taken in other countries and understand potential spatial variations in your model skill.

Fig 7 – the caption and x axis label for fig 7b are inconsistent

Section 4.5 Acceptance Criteria - How did you specify the 0.6 NSE threshold? Was this in conjunction with user requirements or based on existing published thresholds? Do you have any indication of the acceptable forecast skill for users? I find it interesting that there were additional sites when the forecast skill wasn't 'scientifically acceptable' yet users still wanted to receive this information. How have you addressed presenting forecast skill in the user interface? Also see my comments above re: the skill for different parts of the flow regime, did you incorporate this in any way? Section 5.1 goes on to discuss some reasons for variability in forecast skill, could you show the forecast skill spatially on a map and any links to catchment/meteorological forecast characteristics? Again the table display in Table 4 is difficult to interpret due to the number of forecast locations lumped into each jurisdiction.

Section 5 is interesting and raises established challenges of operational streamflow forecasting however it lacks integration with the rest of the paper. Possibly this could be improved with incorporation of wider literature on development of streamflow forecasting systems mentioned above. I also suggest it is moved after section 6 so that it links to the summary and conclusions section.