Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-72-RC1, 2022

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 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.

Author response: We thank the reviewer for commendation and acknowledgement.

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.

Author Response: We thank the reviewer for their suggestions. We are happy to accommodate the reviewer's suggestions:

- *We will elaborate the introduction section of the manuscript to give greater context to different hydroclimatic regions of Australia (lines 66-85) as suggested below, and introduce a new figure in Section 3.1 and elaborate it. We will also review existing literature and include this in the introduction and discussion sections.*
- *We will elaborate the discussion section and will give greater insight to what is unique and novel about this research and why it is suitable for Australia.*

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.

Autor Response: Thanks for the acknowledgement, we highly appreciate it.

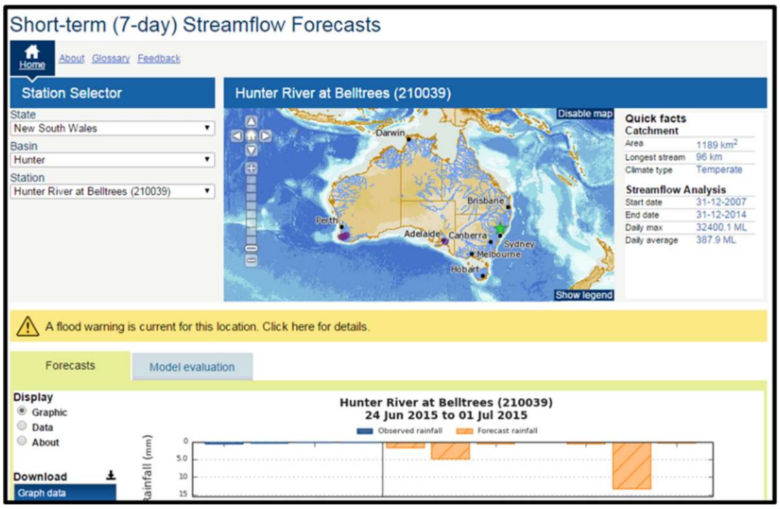
From an open data perspective there is no indication of the source of quality of the observed rainfall and flow data.

Author response: We collected rainfall and discharge data from the Australian Bureau of Meteorology databases. This data is partially quality checked.
The collected data goes through a comprehensive quality checking using a semi-automated workflow by visualising streamflow and nearby rainfall station data side by side. This allows the modeller to identify the connection between rainfall and streamflow (i.e. there should be a high rainfall event for high discharge). This approach assists the modeller to confidently make necessary corrections to the observed data. Then the quality checked data are visually checked (plots) for further quality assurance. The corrections/modifications made to the original data are recorded (a data file) so that all the users of this dataset are aware of them. If necessary, we can elaborate on this process in Section 3.

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.

Author Response: We very much appreciate the reviewer's concerns regarding extreme events, and its performance evaluation using metrics presented in Section 2.5. We certainly agree to acknowledge this limitation and identify it as a future research area.
The Bureau provides more specific and specialised services for flood and drought forecasts. Therefore, if a flood is current for a location, the users of this service are redirected to Bureau's

operational flood forecasting and warning website (please see a screenshot shown below – yellow banner). Therefore, we deliberately avoided discussing extreme events in this paper. However, we appreciate reviewer's comments and agree to acknowledge this limitation and 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?

Author Response: Thanks for the suggestions. We will elaborate the particular relevance to Australia.

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.

Author Response: There are not many operational systems that we are aware of. The HEFS system (Demargne et al. 2014 - https://dx.doi.org/10.1175/bams-d-12-00081.1) is a similar one, but it applies a calibration to rainfall, and an error model. But in practice it uses 'in-the-loop' flood forecasters to do data assimilation manually (or at least, that was the case last we checked), which we imagine would mean it would struggle to produce reliable ensembles in operational setting. The EFAS (EFAS Post-processing - Copernicus Services - ECMWF Confluence Wiki) system uses dynamical models only, as does GloFAS system.
We will elaborate this and present the uniqueness of the system in the Australian context.

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 anther 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.

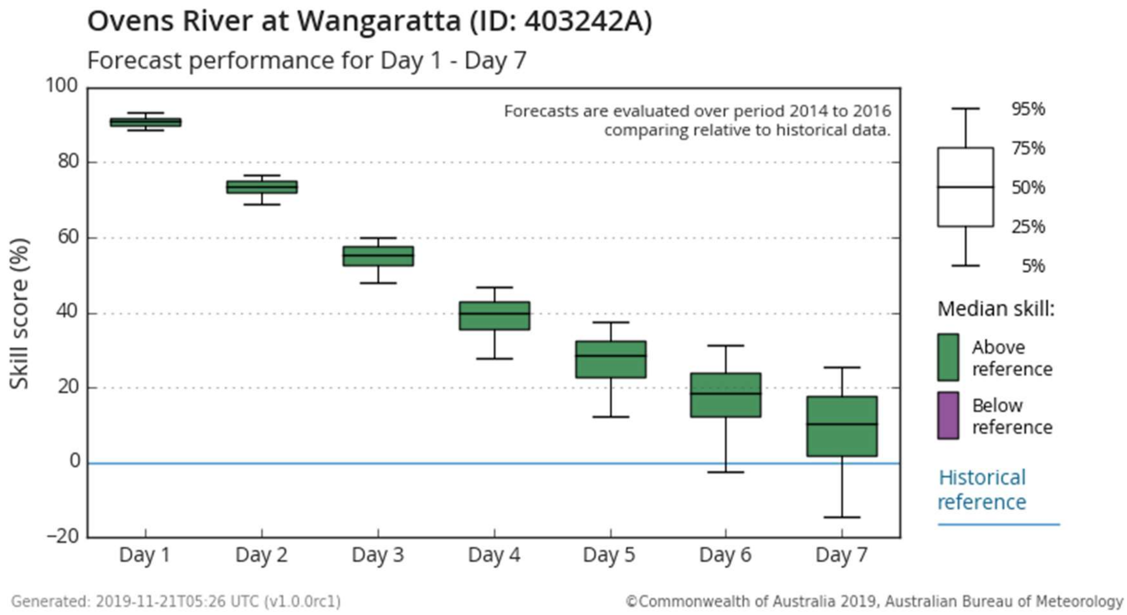
Author Response: We acknowledge the reviewer's suggestions. We will introduce a new spatial map depicting annual rainfall distribution and forecast skill of the forecast locations. We will also discuss any connection between forecast skill and catchment characteristics as the reviewer has suggested.

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

Author Response: We acknowledge the error and will fix it.

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?

Author Response: We adopted NSE of 0.6 from Chew and McMahon (1993) in consultation with the stakeholders. We will elaborate this section and will include reference. We agree that sites where forecasts are not considered ''scientifically acceptable" using this metric may still bring benefit to the user communities. The forecast skill criteria are one set of measures for selecting a forecasting location for the service. We consulted our stakeholders and identified forecast locations critical for their decision making and added to the service. Sometimes, the forecast locations with poor skill are only available to registered users. This is to reduce possible miscommunication with the public and to keep the reputation of the service. To address reviewer's question, we will elaborate Sections 4.5 and 5.4. Please find below a sample plot of forecast skill for a forecast location (with good skill) that is available for users via the service website. A description of how to read this plot is also given on the website.
(http://www.bom.gov.au/water/7daystreamflow/#panel=advanced).



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.

Author Response: We acknowledge the reviewer's suggestion, and we will include another figure (map) showing forecast skills for different locations across Australia.

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.

Author Response: We acknowledge the reviewer's suggestions. We will elaborate Section 5 to integrate more closely with the results and Australian relevance as presented in previous sections. We will present Section 5 after Section 6.

Reference

Chiew, F.H.S., McMahon, T.A., 1993. Assessing the adequacy of catchment streamflow yield estimates. Australian Journal of Soil Research 31, 665–680.

Demargne, J., Limin, W., Regonda, S. K., Brown, J. D., Lee, H., Minxue, H., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., Schaake, J., and Zhu, Y. (2014). The science of NOAA's operational hydrologic ensemble forecast service. *American Meteorological Society*, 95(1), 79–98. https://doi.org/10.1175/BAMS-D-12-00081.1