Responses to Reviewer #1 comments on “Evaluating a landscape-scale daily water balance model to support spatially continuous representation of flow intermittency throughout stream networks” [hess-2020-10]

We thank Reviewer #1 for providing these constructive comments that help improve the quality of this manuscript.

Reviewer comment:

Sunny Yu and coauthor present work modelling intermittent streamflow in two Australian catchments. Overall the paper is well-written and covers a topic that is of huge interest right now. I have some concerns about how the work is presented and will describe them below.

-Authors do not discuss how much data is required to assess low-flows. I fully acknowledge that it can be very difficult in getting long-term data sets. The data that they use here covers a period from January 1, 2005 through December 31, 2017, so 13 years. I do not think that this is a long enough period of time to quantify, characterize, or model low-flows.

Authors reply:

Thanks for the positive comments. The 13 year study period (01/01/2005-31/12/2017) is close to a discharge record length of 15 years which Kennard et al. (2010) concluded is sufficient to enable accurate estimation of low flow metrics. In addition, our study period begins in the middle of the Australian Millennium Drought (2001-2009), meaning that the low-flow assessment in this study included a significant low-flow period. Therefore, we believe that our study period is appropriate to assess low flows. We will add this justification in the revised manuscript.

Reviewer comment:

-Related to this, I am left wondering why the authors would choose a model that they know overestimates streamflow following a rainfall even when they are specifically interested in the lower-end of the streamflow continuum in systems that are dominated by streamflow that is mostly from rainfall.

Authors reply:

First, the AWRA-L model is able to provide spatially contiguous runoff outputs that are necessary to characterise flow intermittency throughout river networks. Equally important, the runoff outputs from the model are readily available and covers all of Australia from 1900 to the current year, enabling the potential application of the method proposed in this study to other areas.

In addition, due to the fundamental difference in runoff drivers for mean and peak flows and low flows, and that many hydrological models are calibrated to the average condition, low flows are usually overestimated (Smakhtin, 2001; Staudinger et al., 2011). To better apply the modelled low flows, we hypothesised and tested the potential causes of the AWRA-L model to overestimate low flows. Furthermore, to mitigate the overestimation of low flows by the model, we estimated segment-specific zero-flow thresholds by developing regression relationships relating gauged zero-flow duration to surrounding environmental variables.

It is also worth mentioning that the AWRA-L model used in this study has already been calibrated and validated for a range of hydrological conditions when it was being developed by CSIRO and the Australian Bureau of Meteorology (Viney et al., 2015). In this study, we applied more rigorous validations to the AWRA-L model, including evaluating its ability to estimate flow intermittency. To our knowledge, there was no other models that have undergone comparable calibration and validation processes for the Australian setting, making AWRA-L the most appropriate option for this study.

Reviewer comment:

-Some of the language is very vague. For instance on MS line 121-122 "If streamflow can be simulated at an acceptable accuracy...." but do not provide any guidelines for what is an acceptable accuracy.

Authors reply:

Thanks for pointing that sentence out. We will revise it to “If streamflow simulated with a routing model shows little difference to that without a routing model, then the conversion process can be more efficient …”. We will also read through the manuscript thoroughly and revise all vague sentences to make their meaning more explicit.

Reviewer comment:

-There is absolutely no explanation of the metrics that the authors used to characterize streamflow. Olden and Poff 2003 describe that it is really hard to characterize intermittent streams with metrics because the metrics that are used to describe one type of intermittent system are not the best to describe other types of intermittent systems. We also know that intermittent streams that are close in proximity to each other can behave very differently from each other from Margaret Zimmer, Adam Ward, and Katie Costigan’s work. There’s no discussion of metrics or how even close intermittent streams can behave differently in the manuscript. I also thought having to flip between a table and a figure to figure out what they were displaying could be improved.

Authors reply:

We are assuming that the reviewer was referring to the flow metrics in Table 1. Those flow metrics have been commonly used to characterise flow regimes, but we will add more explanation of those metrics to the main text. However, we did not aim to use those metrics to characterise intermittent streams, but only to evaluate the ability of the AWRA-L model in simulating streamflow for different components (high-, average- and low-flows).

Based on the model discharge simulations, we further calculated the number of zero-flow days/months to characterise flow intermittency throughout river networks (described in Section 3.4), which is the main focus of this study. We agree with the reviewer that finding the most suitable metrics to characterise different types of intermittent streams is itself an important topic, however, it is beyond the scope of this research.

In our revision, we will revisit section 3.3 to ensure that the purpose of these flow metrics are presented explicitly. We will also add abbreviations of flow metrics shown in Figure 3 and 5 to Table 1, to make the figures and table better connected.

Reviewer comment:

-Related to this, the discussion seems to be over emphasizing the implications of the results. Yes, this is an important first step to modelling intermittent streamflow. However, their results only show that the modeled and observed streamflows have a r2 of less than 0.56! I would not say that this is "fair to good overall alignment" like the authors say in line 299. There is also no discussion, as I mentioned above, about how difficult it is to transfer metrics and results from this coastal Australian sites to other phyiographic areas. They admit that the two catchments used here are similar but the models preformed very differently for them.

Authors reply:

We feel that the reviewer may have misunderstood this part of model performance.

The R^2 value of 0.56 is not about the AWRA-L model performance in streamflow simulation, but about the modelled and observed flow *intermittency* (Figure 7), that latter being calculated from streamflow data as the number of days/months with zero flow. The metric used in this study to evaluate model performance in estimating daily streamflow is the Kling-Gupta efficiency (KGE; results presented in Section 4.2). KGE takes values from -1 to 1: KGE = 1 indicates perfect agreement between simulations and observations, and KGE < -0.41 indicates that the mean of observations provides better estimates than simulations. Our results showed that KGE values were more than -0.41 for all gauges, ranging from -0.19 to 0.76 in SEQ and 0.11 to 0.71 in the Tamar (lines 214-216). These results justify our interpretation that daily streamflow estimates showed a fair to good overall alignment with the observed flows in our study regions.

We evaluated model performance in two hydro-climatically distinctive regions of eastern Australia and found no significant difference in performance between regions (lines 216-218; 302). We also believe the proposed approach has a potential applicability to other regions of Australia and globally. All the data we used in this study to characterise flow intermittency are available for the Australian national scale, and similar datasets also exist in other countries. For example, similar to Geofabric used here, the National Hydrography Dataset Plus (NHDplus) and HydroSHEDS provide hydrographic dataset and hydro-environmental attributes for the national and global scale, respectively. In addition, similar to the daily flow model AWRA-L used in this study, other national-scale hydrologic models are also available around the world, such as the community Noah land surface model (Noah-MP) in the US and the HYPE model in Sweden. We will add more details to the Discussion regarding the applicability of our proposed approach.

References:

Kennard, M.J., Mackay, S.J., Pusey, B.J., Olden, J.D., Marsh, N., 2010. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. River Research and Applications, 26(2): 137-156. DOI:10.1002/rra.1249

Smakhtin, V.U., 2001. Low flow hydrology: a review. Journal of Hydrology, 240(3): 147-186. DOI:https://doi.org/10.1016/S0022-1694(00)00340-1

Staudinger, M., Stahl, K., Seibert, J., Clark, M., Tallaksen, L., 2011. Comparison of hydrological model structures based on recession and low flow simulations. Hydrology and Earth System Sciences, 15(11): 3447-3459.

Viney, N. et al., 2015. AWRA-L v5.0: Technical description of model algorithms and inputs. CSIRO, Australia.