Responses to Reviewer #2 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 #2 for providing these constructive comments that help improve the quality of this manuscript.

Reviewer comment:

The topic of this paper is certainly timely, and evaluation of how runoff routing, temporal resolution of models and climate impacts on spatial/temporal variability of drying streams is important. The biggest challenge I see in this manuscript is that the hypotheses presented in lines 163-166 are not clear and the framing of the problem and results wanders between computationally efficient streamflow routing to the timescale of importance, to the sub-catchment climate variability, to capturing spatial and temporal patterns of intermittency. With sufficient re-organization, additional details on the individual models and observational data, re-evaluation of the time period of no flow that allows results to be compared across daily and monthly modes this work could provide interesting insights into intermittent stream research. Given the extent of revisions needed, I do not suggest it to be accepted at this time.

Authors reply:

We thank the reviewer for pointing out several issues, however, we also believe that these issues can be sufficiently addressed in this revision.

First, we believe the framing of the research questions was clearly outlined in the last paragraph of the Introduction where we stated (lines 84-93):

"In this study, we sought to apply spatially contiguous daily runoff outputs from the AWRA-L water balance model to quantify the spatial extent and temporal patterns of flow intermittency. To assess the accuracy of the AWRA-L model for daily flow simulations, we first developed a simple but effective technique to convert runoff to streamflow for two hydro-climatically distinctive regions. …. We further assessed the uncertainty of the AWRA-L model in capturing patterns of flow intermittency. Lastly, we evaluated the effect of time step (daily vs. monthly) on the relative performance of the model in replicating observed patterns of cease to flow periods at reference gauges."

The hypotheses presented in lines 163-166 simply provided more details in the Methods section relating to the first objective (assessing the accuracy of the AWRA-L model for daily flow simulations).

We feel the structure and organisation of the paper was logically arranged to address our main objectives. Sections 4.1 and 4.2 of the Results evaluate the effects of streamflow routing and regional differences in hydro-climatic variability on streamflow simulations. Section 4.3 of the Results evaluated our ability to accurately estimate spatial and temporal patterns of flow intermittency using simulated spatially contiguous streamflow data. The importance of model time-step (daily vs. monthly) in estimation of streamflow intermittency was also evaluated in Section 4.3 of the Results.

We acknowledge that more details on the individual models and observational data should be provided, and we will do so in the revised manuscript.

Concerning the time period of no flow to enable comparison of daily and monthly models, we classified a month as no-flow only if every day of the month was estimated to be at zero flow. This classification method was aimed to convert daily flow intermittency to monthly flow intermittency, allowing the daily flow model AWRA-L to be comparable to the monthly flow model AWAP in terms of the ability to estimate flow intermittency. As the monthly flow model AWAP outputs monthly average flow, the zero value of monthly flow means all days in the month have zero flows. That's the reason why we chose that classification method. We will add the rationale in the revised manuscript. Additionally, we will also try using a different method to aggregate the modelled daily flow intermittency into monthly flow intermittency. This way, the results would be biased to be more "intermittent" as compared to our original results that may be biased to be more "perennial", and these two together should provide readers with both the upper and lower bounds of comparing daily and monthly models in estimating flow intermittency.

Reviewer comment:

Using out of the box hydrologic models (AWRA-L, AWAP) that over predict baseflow will certainly limited the ability to capture no-flow conditions (Figure 7, lines 309-311). These models are not fully described, even conceptually in the paper, making it challenging for a reader to understand which assumptions lead to this over-prediction. A previous study was used to benchmark flow intermittency, but was not explained in the methods.

Authors reply:

We will add more details on the individual models (including the AWAP model used to benchmark flow intermittency) and observational data as supplemental material to the manuscript.

We are not surprised that the AWRA-L model over-estimates low flows, a common problem with many hydrological models due to the difficulties of quantifying hydrological processes influencing low flow discharge (Smakhtin, 2001; Staudinger et al., 2011). In this study, we further investigated the potential sensitivity of the model to rainfall events by testing two hypothesis: 1) the overestimation of gains to low flow discharge, and 2) underestimation of transmission losses (Section 3.3; Figure 6). We also estimated appropriate zero-flow thresholds for each stream segment to mitigate this over-estimation of low flows.

Reviewer comment:

Some of the methods of examining the low-flows themselves seem questionable, namely, that all days in a month had to have zero flow for the flows in that month to be zero from the AWRA-L outputs (line 187). There is work being done that suggests that a stream that goes dry for 15 days in a year is considered intermittent, so using a consistently dry 30 day window could be an exceptionally high threshold, either way, description of why a given threshold is used is necessary.

Authors reply:

As explained in our response above, we classified a month as no-flow only if every day of the month was estimated to be at zero flow. This classification method was aimed to convert daily flow intermittency to monthly flow intermittency, allowing the daily flow model AWRA-L to be comparable to the monthly flow model AWAP in terms of their ability to estimate flow intermittency. We will add this rationale in the revised manuscript. Additionally, we will also try using a different method to aggregate the modelled daily flow intermittency into monthly flow intermittency (in line with our first response).

Reviewer comment:

The timeframe of observational data included is not clear, and it was not presented with the comparison between model output in Figure 9. The modeled flow from 1911-2016 is included in the paper, with no reference to how well that model actually did at capturing low flows in the calibration time period. Figure 6 is slightly misleading because the dashed line is not a continuous variable and the catchment areas do not increase linearly.

Authors reply:

The timeframe of observational streamflow data included was clearly described in lines 111-112: "All gauges have less than 0.5 % missing values over the period from 01/01/2005 to 31/12/2017". This observational streamflow data was not presented in Figure 9; instead, the comparison between modelled and observed streamflow data is presented in Figure 4.

The AWRA-L model has already been calibrated and validated by its developers from the Australian Bureau of Meteorology and CSIRO (Viney et al., 2015). In our study, we further evaluated the model accuracy in streamflow simulations over the period of 2005-2017, with a particular focus on low flows. The model performance at capturing low flows was clearly illustrated in Figure 5. Based on the accuracy assessment, we applied a longer period (1911-2016) of the model outputs to estimate the temporal dynamics of flow intermittency in SEQ (Figure 9). We will provide more details about the calibration and validation of the AWRA-L during its development as supplemental material to the revised manuscript.

We agree with the Reviewer's comment regarding Figure 6. We will redraw this figure to remove the dashed line and only retain the value dots.

Reviewer comment:

The writing and organization of the manuscript could be improved throughout (e.g. 59-63). References to the multiple model configurations throughout is particularly confusing (e.g. a table that has the 4 model configurations and associated details with acronyms would be useful). One important caveat relevant to modelling intermittent streams at a daily-time step using contiguous data is not referenced (e.g. stream gaging locations are generally put where there is usually surface water flow). There are several quantitative results (r2) that are presented, yet the discussion poses that the models "showed fair to good overall alignment" which seems to overstate the ability to capture the low flows given how low the r2 was.

Authors reply:

We will thoroughly check the manuscript and improve the writing and clarity where necessary. However, as explained in an earlier response, we feel the organisation of the paper is logically arranged to address our main objectives. We will revise the sentence in lines 59-63 to: "These kind of simulations are important to better understand the causes of flow intermittency at multiple spatial scales, and enable ecologically-relevant characterisation of streamflow properties such as the magnitude, frequency, duration, and rates of change in high or low flow events".

We will also include a table of all model configurations and other relevant information in the revised manuscript to improve clarity.

We disagree with the Reviewers' assertion that "stream gaging locations are generally put where there is usually surface water flow". In Australia this is not the case with around 70% of 830 streamflow gauges found to be located on streams with varying degrees of flow intermittency (Kennard et al., 2010). Therefore, the Reviewers' suggested caveat does not apply to the Australian situation.

The R^2 value of 0.56 does not relate to the AWRA-L model performance in streamflow simulation; instead, it relates to the concordance between modelled and observed flow *intermittency* (see Figure 7), which was calculated from streamflow data as the number of days/months with zero flow. AWRA-L model performance was instead assessed using the Kling-Gupta efficiency (KGE) metric compare the alignment between modelled and observed streamflows. 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.

In our revision, we will revisit the Methodology and Results sections to ensure that the evaluation procedures are explained more explicitly.

Reference:

Kennard, M.J. et al., 2010. Classification of natural flow regimes in Australia to support environmental flow management. Freshwater Biology, 55(1): 171-193. DOI:10.1111/j.1365-2427.2009.02307.x

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.