Responses to George Allen's 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 George Allen for providing these constructive comments that help improve the quality of this manuscript.

Reviewer comment:

The manuscript, "Evaluating a landscape-scale daily water balance model to support spatially continuous representation of flow intermittency throughout stream networks" by Yu et al. presents a study that quantifies the ability of two water balance models to simulate streamflow in two basins in Australia, with a particular focus on intermittent streamflow. The authors focus on comparing different water balance models with different timesteps and comparing a flow routing streamflow model to a simple lumped model. Perhaps the most novel component of the analysis involves the characterization of so-called cease-to-flow conditions by applying a gauge derived threshold to the streamflow simulations.

The manuscript is well written and of interest a variety of scientific communities including hydrologists, ecologists, and potentially biogeochemists. While there are a number of improvements that should be made to the manuscript (see below), I don't think there are any issues that warrant a major revision to this manuscript. The three most important issues that should be addressed are as follows:

Authors reply:

Thanks for the positive comments.

Reviewer comment:

1) The authors need to address the uncertainty of, and assumptions involved with, developing linear relationships at a limited number of gauges and then extrapolating these relationships across basins. For example, are the locations of the gauges a representative sample of the population of streams within the two basins, or are they biased towards large, perennial rivers segments? Another example: are the gauges used to calibrate the water-balance and streamflow models used in this study the same gauges used to estimate crease-to-flow occurrence? If so, please include how this fact may impact the results, particularly in terms of uncertainty.

Authors reply:

We agree that the spatial distribution of gauged streams as a representative sample of the population of streams is an important consideration when calibrating a regression model and using it to extrapolate more widely. We considered our sampled gauge locations to be representative of the population of streams and included a statement to this effect in the manuscript lines 113-114: "The gauges were dispersed throughout each study area and encompassed a range of stream sizes and flow regime types". We will add more details to this description in the revision.

We here describe in more detail the sets of streamflow gauges used in the various steps in our analyses and illustrate this in the following Figure R1. The water balance model (AWRA-L) was both calibrated and validated by the developers from the Australian Bureau of Meteorology and CSIRO at the national scale (Viney et al., 2015), with 301 gauges used for calibration and a different set of 304 gauges used for validation (Zhang et al., 2013). Our study converted the AWRA-L water balance model predictions to streamflow estimates and these were validated for different components of the flow regime (high-, average- and low flows), using 25 and 15 gauges in two hydro-climatically distinctive regions, respectively (SEQ and Tamar). Only 6 of the 25 gauges in SEQ and 3 of the 15 gauges in Tamar were the same as those used to calibrate the AWRA-L water balance model. This small overlap between the AWRA-L calibration gauge set (n=301) and the streamflow model validation gauge set (n=25 in SEQ and 15 in Tamar) means that potential overestimation of streamflow model performance is likely to be minimal but we will include a discussion of this potential source of uncertainty in the revised manuscript. A larger set of 43 gauges in SEQ (including 21 of the 25 gauges used by us for streamflow validation) was used to estimate the zero-flow threshold for this region. However, because the validation of the streamflow model applied to the raw discharge simulations, rather than the corrected discharge simulations with zero-flow thresholds, we do not regard this choice of streamflow gauges to be an issue for model validation in this study.

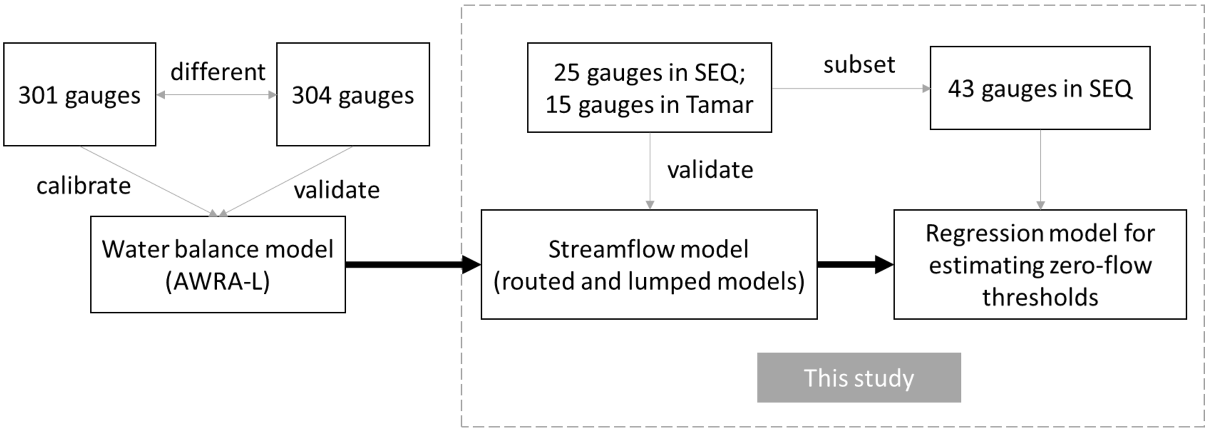

Figure R1. Schematic illustration of the streamflow gauge sets used in the different modelling processes described in this paper.

Reviewer comment:

2) When comparing monthly and daily models, the authors classify a month as no-flow only if every day of the month is estimated to be at zero flow. Wouldn't this approach bias the results to be more perennial? Is this why the daily model doesn't perform as well as the monthly model at the monthly timestep? Please provide some rationale on this decision for the monthly classification.

Authors reply:

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. We propose to try regarding a month as non-flowing when as least one day in the month have zero flow. 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 bound of comparing daily and monthly models in estimating flow intermittency.

Reviewer comment:

3) Finally, I suspect that Geofabric is missing some of the smallest streams (see Benstead & Leigh (2012) An expanded role for river networks, Nature Geoscience). If so, this error will control the proportion of rivers that are predicted to be intermittent, a primary finding of this study.

Authors reply:

We agree that small streams comprise a large proportion of river networks, may be more frequently intermittent than larger streams, and their prevalence may be underestimated using readily available spatial datasets such as used in our study. We do not, however, regard these issues as compromising the main objectives of our study.

We believe that the spatial resolution of the smallest streams identified in the Geofabric stream network (version 2.1.1) is appropriate considering the relatively large spatial extent of our study areas. The Geofabric stream network is a fully connected and directed stream network derived from the national 9 arc-second DEM and flow direction grid (~250m resolution). Streams of seven Strahler orders were delineated in Geofabric for the study river networks, with the minimum upstream drainage area of 1.5 km^2, while the two study areas (SEQ and Tamar) are 21,331 km^2 and 11,215 km^2, respectively. In addition, the Geofabric is the finest resolution national stream network layer with supporting environmental attributes available for Australia and is of much finer resolution than similar products such as HydroSHEDS (15 arc-second (~500 m) resolution).

Moreover, an updated version of Geofabric (version 3) is now being developed (http://www.bom.gov.au/water/geofabric/about.shtml). The new version is based on a finer scale digital elevation model (~30m resolution) and aims to provide continent-wide river networks with eight Strahler stream orders. Our proposed approach to characterising flow intermittency can also be built upon this new version of Geofabric.

Specific comments:

Abstract:

L27: replace "intermittent flows" with "cease-to-flow events"

L28: add "at a monthly timestep" after "intermittency"

L29: add ", using a daily streamflow model" after "1911-2016". The monthly model produced a different estimate.

Main text:

L92: As mentioned above, add an acknowledgement about how the location of these reference gauges is likely biased towards particular river types (e.g. large perennial rivers) and river forms (e.g.

narrow, single threaded rivers located near bridges), and how this bias might influence the extrapolation of the cease-to-flow threshold to all Geofabric stream segments.

L99: Add a little more information about Geofabric. What is the spatial resolution? Does it contain all of the smallest streams in the basins? If there is a channelization threshold, it will control the proportion of rivers that are estimated to be intermittent.

L112: As mentioned above, are the gauges used to calibrate the water-balance and streamflow models used in this study the same gauges used to estimate cease-to-flow? If so, please include how this fact may impact the results, particularly in terms of uncertainty.

L114: As mentioned above, please provide more information on the types of rivers and streams that these gauges are located on. This can help the reader understand the uncertainty associated with this analysis.

L122-123: "the readily available runoff data can be more accessible for potential applications" I don't follow this logic. Using a flow propagation model doesn't limit accessibly and should be relatively fast using RAPID, especially at the scale of these two catchments.

L160-161: "given that we do not have access to the underlying models to directly adjust model parameters." RAPID is open source and you can adjust these parameters.

L162: Gauges are on rivers with large upstream drainage areas. There should be an acknowledgement that there are many smaller streams that likely have higher cease-to-flow occurrence and the gauges are likely not representative of these smaller streams.

L187: "all days in a month had to have zero flow for the flows for that month to be zero". Wouldn't this approach bias the results to be more perennial? Is this why the daily model doesn't perform as well as the monthly model at the monthly timestep? Please provide some rationale on this decision.

L197-198: As mentioned above, "The temporal pattern of flow intermittency was expressed as the proportion of streams with flow intermittency > 30 days or 1 month" – is this definition of intermittent streams based off of something or is it just arbitrary?

L239: insert "fair" before "match"

L288: Please explain "time of concentration" for the uninformed reader. Would be best to introduce it earlier on in the manuscript.

L300 and L301: typo: replace "KEG" with "KGE"

L318-319: "and recently many studies have developed methods to calculate transmission losses for better flow simulations (Lange, 2005; Costa et al., 2012)." The citations provided are neither recent nor many. L329: add "temporal" before "resolution"

L337: replace "is difference" with "are differences"

Figures:

Figure 3: Providing a y-axis with units would make it easier to interpret these boxplots

Figure 5: Perhaps considering scaling some of these y-axes as log, outliers make it difficult to compare the distributions and see the distribution of data where most of the data are located.

Authors reply:

All of the above comments will be addressed as suggested, except for the following, for which we provide individual responses below:

Specific comment:

L92: As mentioned above, add an acknowledgement about how the location of these reference gauges is likely biased towards particular river types (e.g. large perennial rivers) and river forms (e.g. narrow, single threaded rivers located near bridges), and how this bias might influence the extrapolation of the cease-to-flow threshold to all Geofabric stream segments.

Authors reply:

As explained in our response to the first major comment (above), we considered our sampled gauge locations to be representative of the population of streams and included a statement about this in the manuscript (lines 113-114): “The gauges were dispersed throughout each study area and encompassed a range of stream sizes and flow regime types”. We will add more details to this description in the revision.

Specific comment:

L122-123: “the readily available runoff data can be more accessible for potential applications” I don’t follow this logic. Using a flow propagation model doesn’t limit accessibly and should be relatively fast using RAPID, especially at the scale of these two catchments.

Authors reply:

This sentence is in a context of whether the conversion process can be more efficient without a routing model. Here we actually mean that if the conversion process does not need a routing model (e.g. RAPID), the users of the AWRA-L runoff data can confidently skip the routing process, which makes the runoff data more accessible for potential applications.

Specific comment:

L160-161: “given that we do not have access to the underlying models to directly adjust model parameters.” RAPID is open source and you can adjust these parameters.

Authors reply:

Here “the underlying models” was meant to be the AWRA-L model. We will revise the sentence as “given that we do not have access to the AWRA-L model to directly adjust model parameters.”

Specific comment:

L187: “all days in a month had to have zero flow for the flows for that month to be zero”. Wouldn’t this approach bias the results to be more perennial? Is this why the daily model doesn’t perform as well as the monthly model at the monthly timestep? Please provide some rationale on this decision.

Authors reply:

This comment is same to the second major comment. Please refer to our response to that major comment on Page 2.

Reference:

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

Zhang, Y. et al., 2013. Collation of Australian modeller's streamflow dataset for 780 unregulated Australian catchments, CSIRO: Water for a Healthy Country National Research Flagship.