Response to Editor Christa Kelleher's comments

We thank the Editor Christa Kelleher for inviting us to revise the manuscript. We also thank her for summarising all Reviewers comments and providing constructive comments that help improve the quality of this manuscript.

Editor comment:

The manuscript has received three constructive reviews. All reviewers note that the topic of this manuscript is both important and timely and note that the study will be of interest to a broad readership.

Though the reviewers had several recommendations that I believe will be useful to the authors when revising the manuscript, there were two common themes throughout:

Authors reply:

We have addressed all of the editor and reviewer comments as detailed below.

Editor comment:

The reviewers all had several questions and comments regarding the methods section of this manuscript, and sought additional clarity. I encourage the authors to expand these sections with both text and equations, as they have outlined in their response, including a discussion of the model framework, the selected flow metrics (and why these were selected), and the choice of a metric of model accuracy (why KGE and not an explicitly low flow metric like LRMSE?). The reviewers also had questions regarding some of the decisions made by the authors in the methods section, and encouraged the authors to justify their decisions. In their response, the authors have developed a good plan to address these comments, and I'd encourage them to acknowledge not only their decisions and why they made them, but also to include a discussion of these limitations in the context of how to move the science of predicting intermittent streams forward.

Authors reply:

In response to reviewers comments related to the methods section, we have expanded the methodology section substantially to improve clarity. The major changes are:

1) We added details on the individual models (lines 121-125 for the AWRA-L model, and lines 220-224 for the WaterDyn (initially called "AWAP") model) and observational data (lines 114-118).
2) We provided more explanation of the selected flow metrics and the reason for selecting them (lines 150-154; 166-168). In addition, we also added abbreviations of flow metrics shown in Figure 4 and 6 to Table 1 to make the figures and Table better connected.
3) We included the KGE equation and associated descriptions. Regarding the choice of metric to evaluate model accuracy, we have provided further justification as follows: "The use of KGE provides an overall assessment of AWRA-L model performance, and the flow metrics in Table 1 are used to comprehensively evaluate the model accuracy for various components of flow regimes, including the flow metrics related to low flows" (lines 168-170). We therefore believe that the use of other low flow-focused metrics (e.g. LRMSE) is redundant as we have

already provided a detailed assessment of those related to low flows, including the magnitude, timing, duration, frequency of low flow spells (Table 1).

Regarding the method to classify a month as zero-flowing using daily streamflow, we have now added further rationale for the method we used in the original manuscript, namely that all days in a month had to have zero flow for the flows for that month to be zero. We have now included an additional method that considered a month as non-flowing when at least one day in the month had zero flow (lines 213-219). 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". Consideration of results from both methods should provide readers with both the upper and lower bounds of comparing daily and monthly models in estimating flow intermittency. In addition, we discussed the implication of the two different methods in the revised manuscript (lines 377-381).

Editor comment:

The other common string of comments from the reviewers centered on some of the limitations of the model in the context of the physical hydrology of intermittent streams. For instance, as pointed out by the reviewers, the choice of flow metrics that best describe hydrologic regimes may differ across systems, and from a physical hydrologic perspective, even streams in close proximity can exhibit different behavior. I encourage the authors to acknowledge these limitations, and the implications of these limitations for their work, and for the work of the broader community working to simulate and study intermittent streams. Though the authors responded to these reviewer comments, I request that they revise their manuscript to provide greater discussion of these limitations in their revision.

Authors reply:

We agree that there are limitations of the model in runoff generation, particularly for low flow periods, and this has important implications for estimating spatio-temporal patterns in streamflow intermittency using modelled runoff data. To acknowledge this limitation, we have now included an expanded discussion of this issue (citing the work of Zimmer and colleague) in Section 5.2 of the Discussion as follows (Lines 353-357, new text highlighted in bold, with the original text provided for context):

"This suggests that the AWRA-L model is a generally robust model in predicting average- and high-flows, but still needs some improvement to better simulate low flows. **Runoff generation processes can vary substantially through space and time due to such factors as variations in soil depth, antecedent soil moisture and groundwater connectivity, and this can influence spatio-temporal variations in low flow characteristics, including streamflow intermittency (Zimmer and McGlynn, 2017). However, it is unknown the extent to which this contributed to uncertainty in the simulation of low flows and estimation in streamflow intermittency in this study**. The uncertainty of AWRA-L in low flow simulations can be linked to its over-responsiveness to rainfall, partly caused by overestimation of "in situ" gains and underestimation of transmission losses to low flow discharge, as shown in SEQ. Previous studies found that lateral flow exchange between grid cells of land surface models (e.g. AWRA-L) plays a significant role in redistributing soil water (Kim and Mohanty, 2016), and thus may improve "in situ" surface/subsurface runoff simulations (Lee and Choi, 2017). On the other hand, hydrological process involved in transmission losses have been extensively discussed (Jarihani et al., 2015; Konrad, 2006), and studies have developed methods to calculate transmission losses for better flow simulations (Costa et al., 2012; Lange, 2005). Therefore, low flow simulations by AWRA-L can possibly be improved by incorporating lateral flow exchange algorithms and better accounting for hydrological process such as evapotranspiration from riparian vegetation and infiltration into channel beds. This improvement is made more likely as AWRA-L has been released as a Community Modelling System (https://github.com/awracms/awra_cms), which allows co-development by the research community."

Editor comment:

One reviewer also recommended changes to figures, to make them more interpretable, and I encourage the authors to incorporate these recommendations into their revision.

Authors reply:

We have incorporated these recommendations into the revision as detailed below:

1) We have now provided unit for each of the flow metrics in Figure 4 and 6 (Figure 3 and 5 in the original submission).
2) We have scaled the y axes as $\log_{10}$ in Figure 6 to better display the distribution of the data.
3) We have re-drawn Figure 7 as a scatter plot to improve clarity.

Editor comment:

I request that the authors incorporate their proposed revisions into the discussion paper for re-review.

Authors reply:

We have incorporated all of our proposed revisions into the revised manuscript for re-review.

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. It is also worth noting that our study period is the common period covered by streamflow observations across all gauges, which provides a consistent assessment of model performances spatially. We have added this justification in the revised manuscript (lines 181-184).

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 have revised 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 …" (lines 130-132). We also rephrased all vague sentences we could find to make their meaning more explicit. For example, one vague sentence once read:

"Given that we do not have access to the underlying model to directly adjust model parameters, we instead compared the observed and modelled low flow magnitude…"This has now been more specific about "the underlying model". The revised sentence now reads:

"Given that we do not have access to the AWRA-L model to directly adjust model parameters, we instead compared the observed and modelled low flow magnitude…"

Two additional examples are given below:

Original sentence:

"Due to the fact that water balance models often over-predict the magnitude of very low flows, we adopted the same method used in Yu et al (2018)…"

Revised sentence (lines: 198-200):

"Given the fact that water balance models often over-predict the magnitude of very low flows due to the difficulties of quantifying hydrological processes influencing low flow discharge, we adopted the same method used in Yu et al (2018)…"

Original sentence:

"The spatial patterns of flow intermittency derived from the daily and monthly flow simulations aligned well only for the main stems and some coastal streams…"

Revised sentence (lines: 287-288):

"The spatial patterns of flow intermittency derived from the daily AWRA-L and monthly WaterDyn flow simulations aligned well only for the main stems and some coastal streams…"

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:

The metrics used to characterise streamflow are defined in Table 1. These flow metrics have commonly been used to characterise critical components of flow regimes across average, high, and low flow conditions. In the revised manuscript, we have added more explanation of these metrics with the following text (lines 150-152):

"The calculated flow metrics are commonly used to describe the critical components of flow regimes across average, high, and low flow conditions, including flow magnitude and variability, the timing, frequency and duration of high and low flows, and rates of changes in flow events (Olden and Poff, 2003; Poff et al., 1997)."

We have also added abbreviations of flow metrics shown in Figure 4 and 6 to Table 1, to make the figures and table better connected.

To quantify flow intermittency throughout river networks, we calculated the total number of zero-flow days per year based on the spatially contiguous modelled streamflow simulations. We also 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 by comparison with flow intermittency estimates derived from a monthly model (Yu et al., 2018).

We agree with the Reviewer's important point that "intermittent streams that are close in proximity to each other can behave very differently from each other" as this has important implications for estimating spatio-temporal patterns in streamflow intermittency using modelled runoff data. To address this issue, we have now included an expanded discussion of this issue (citing the work of Zimmer and colleague) in Section 5.2 of the Discussion as follows (Lines 353-357, new text highlighted in bold, with the original text provided for context):

"This suggests that the AWRA-L model is a generally robust model in predicting average- and high-flows, but still needs some improvement to better simulate low flows. **Runoff generation processes can vary substantially through space and time due to such factors as variations in soil depth, antecedent soil moisture and groundwater connectivity, and this can influence spatio-temporal variations in low flow characteristics, including streamflow intermittency (Zimmer and McGlynn, 2017). However, it is unknown the extent to which this contributed to uncertainty in the simulation of low flows and estimation in streamflow intermittency in this study**. The uncertainty of AWRA-L in low flow simulations can be linked to its over-responsiveness to rainfall, partly caused by overestimation of "in situ" gains and underestimation of transmission losses to low flow discharge, as shown in SEQ. Previous studies found that lateral flow exchange between grid cells of land surface models (e.g. AWRA-L) plays a significant role in redistributing soil water (Kim and

Mohanty, 2016), and thus may improve "in situ" surface/subsurface runoff simulations (Lee and Choi, 2017). On the other hand, hydrological process involved in transmission losses have been extensively discussed (Jarihani et al., 2015; Konrad, 2006), and studies have developed methods to calculate transmission losses for better flow simulations (Costa et al., 2012; Lange, 2005). Therefore, low flow simulations by AWRA-L can possibly be improved by incorporating lateral flow exchange algorithms and better accounting for hydrological process such as evapotranspiration from riparian vegetation and infiltration into channel beds. This improvement is made more likely as AWRA-L has been released as a Community Modelling System (https://github.com/awracms/awra_cms), which allows co-development by the research community."

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 aspect of our assessment of model performance.

The assessment of "fair to good overall alignment" related to the AWRA-L model performance in streamflow simulation. 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 248-250). 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 contrast, the R^2 value of 0.56 relates to the assessment of concordance between the modelled and observed flow *intermittency* (Figure 8), which was calculated from streamflow data as the number of days/months with zero flow. In the revised manuscript, we have reworded aspects of the Methodology and Results sections to ensure that the evaluation procedures are explained more clearly. In addition, in the Discussion (Section 5.2), we now also provide an expanded discussion of the potential sources of uncertainty in the ability of the AWRA-L model to accurately simulate low flows (as detailed in the response above).

We evaluated model performance in two hydro-climatically distinctive regions of eastern Australia and found no significant difference in performance between regions (lines 250-252). 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. We have added this information to the Discussion regarding the applicability of our proposed approach (lines 393-400):

The approach developed here to generate spatially continuous estimates of streamflow characteristics (including flow intermittency) throughout stream networks has potential applicability to other regions of Australia and globally. All the data used in this study are available for the Australian national scale, and similar datasets also exist in other countries. For example, similar to the Geofabric data (Stein et al., 2014) used here, the National Hydrography Dataset Plus (NHDPlus) and HydroATLAS (Linke et al., 2019) provide hydrographic datasets and hydro-environmental attributes for national (USA) and global scales, respectively. In addition, similar to the daily flow model AWRA-L used in this study, other global and national-scale hydrologic models are also available, such as the global WaterGAP model (Döll et al., 2003), the community Noah land surface model (Noah-MP) (Niu et al., 2011) in the US and the HYPE model (Lindström et al., 2010) in Sweden.

References:

Costa, A.C., Bronstert, A., de Araújo, J.C., 2012. A channel transmission losses model for different dryland rivers. Hydrology and Earth System Sciences, 16(4): 1111-1135. DOI:10.5194/hess-16-1111-2012

Döll, P., Kaspar, F., Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. Journal of Hydrology, 270(1): 105-134. DOI:https://doi.org/10.1016/S0022-1694(02)00283-4

Jarihani, A.A., Larsen, J.R., Callow, J.N., McVicar, T.R., Johansen, K., 2015. Where does all the water go? Partitioning water transmission losses in a data-sparse, multi-channel and low-gradient dryland river system using modelling and remote sensing. Journal of Hydrology, 529: 1511-1529. DOI:10.1016/j.jhydrol.2015.08.030

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

Kim, J., Mohanty, B.P., 2016. Influence of lateral subsurface flow and connectivity on soil water storage in land surface modeling. Journal of Geophysical Research: Atmospheres, 121(2): 704-721. DOI:10.1002/2015jd024067

Konrad, C.P., 2006. Longitudinal hydraulic analysis of river-aquifer exchanges. Water Resources Research, 42(8): W08425. DOI:10.1029/2005wr004197

Lange, J., 2005. Dynamics of transmission losses in a large arid stream channel. Journal of Hydrology, 306(1-4): 112-126. DOI:10.1016/j.jhydrol.2004.09.016

Lee, J., Choi, H., 2017. Improvements to Runoff Predictions from a Land Surface Model with a Lateral Flow Scheme Using Remote Sensing and In Situ Observations. Water, 9(2). DOI:10.3390/w9020148

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrology Research, 41(3-4): 295-319. DOI:10.2166/nh.2010.007

Linke, S. et al., 2019. Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data, 6(1): 283. DOI:10.1038/s41597-019-0300-6

Niu, G.Y. et al., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. Journal of Geophysical Research: Atmospheres, 116(D12).

Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications, 19(2): 101-121. DOI:10.1002/rra.700

Poff, N.L. et al., 1997. The Natural Flow Regime. BioScience, 47(11): 769-784. DOI:10.2307/1313099

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.

Stein, J.L., Hutchinson, M.F., Stein, J.A., 2014. A new stream and nested catchment framework for Australia. Hydrology and Earth System Sciences, 18(5): 1917-1933. DOI:10.5194/hess-18-1917-2014

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

Yu, S., Bond, N.R., Bunn, S.E., Xu, Z., Kennard, M.J., 2018. Quantifying spatial and temporal patterns of flow intermittency using spatially contiguous runoff data. Journal of Hydrology, 559: 861-872. DOI:10.1016/j.jhydrol.2018.03.009

Zimmer, M.A., McGlynn, B.L., 2017. Ephemeral and intermittent runoff generation processes in a low relief, highly weathered catchment. Water Resources Research, 53(8): 7055-7077. DOI:10.1002/2016wr019742

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, which we believe can be readily 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 85-94):

"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 192-195 (lines 163-166 in the original manuscript) 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 have now provided more details on the individual models (lines 121-125 for AWRA-L and lines 220-224 for WaterDyn (initially called "AWAP") and observational data (lines 114-118) 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 WaterDyn in terms of the ability to estimate flow intermittency. As the monthly flow model WaterDyn outputs monthly average flow, the zero value of monthly flow means all days in the month have zero flows. Additionally, we have now evaluated an alternative method to aggregate the modelled daily flow intermittency into monthly flow intermittency (lines 213-219). We assessed the effect of considering a month as non-flowing when at least one day in the month had 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 bounds of comparing daily and monthly models in estimating flow intermittency. This is expanded upon in response to a detailed Reviewer comment on this issue below.

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 have now added more details on the individual models (including the AWRA-L model to simulate daily flows and the WaterDyn (initially called "AWAP") model used to benchmark flow intermittency based on monthly flow simulations) to the manuscript (lines 121-125 for AWRA-L and lines 220-224 for WaterDyn).

We are not surprised that the AWRA-L model over-estimated 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 7). We also estimated appropriate zero-flow thresholds for each stream segment to mitigate this over-estimation of low flows. In the Discussion (Section 5.2), we also provide an expanded discussion of the potential sources of uncertainty in the ability of the AWRA-L model to accurately simulate low flows (as per the response to this issue raised by Reviewer #1)

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 WaterDyn in terms of their ability to estimate flow intermittency. We have added this rationale in the revised manuscript. Additionally, we also added 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 113-114: “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 10 (Figure 9 in the original submission); instead, the comparison between modelled and observed streamflow data is presented in Figure 5.

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 6. 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 10). We have provided more details about the calibration and validation of the AWRA-L during its development in the revised manuscript (lines 124-125, 170-173).

We agree with the Reviewer’s comment regarding Figure 6. We have re-drawn this figure as a scatter plot to improve clarity (see Figure 7).

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 have thoroughly checked the manuscript and improved the writing and clarity where necessary.

The sentence on line 59-63 has been reworded as follows:

"These kinds 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 rate of change in high or low flow events". (lines 60-63)

Other examples of improvements to writing and clarity are provided in our responses to comments by Reviewer #1.

Regarding organization of the manuscript, we feel the paper is logically arranged to address our main objectives (as explained in an earlier response). To improve clarity of the various model configurations used in the paper, we have now included a new figure (instead of a table as the Reviewer suggested) of all model configurations and other relevant information in the revised manuscript (see Figure 2).

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.

Regarding the comment about our assessment of the ability to accurately model low flows, we addressed this issue in response to comments by Reviewer # 1.

The assessment of "fair to good overall alignment" related to the AWRA-L model performance in streamflow simulation. 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 248-250). 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 contrast, the R^2 value of 0.56 relates to the assessment of concordance between the modelled and observed flow *intermittency* (Figure 8), which was calculated from streamflow data as the number of days/months with zero flow.

In the revised manuscript, we have reworded aspects of the Methodology and Results sections to ensure that the evaluation procedures are explained more clearly. In addition, in the Discussion (Section 5.2), we now also provide an expanded discussion of the potential sources of uncertainty in the ability of the AWRA-L model to accurately simulate low flows (see lines 353-357).

References:

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.

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 an updated statement to this effect in the revised manuscript lines 114-118: "The gauges were widely dispersed throughout each study area and encompassed a range of stream sizes, catchment areas (22-3,881 km^2 in SEQ; 33-3,294 km^2 in Tamar) and flow regime types, ranging from highly intermittent to perennial streams (see results). Therefore, we regard the selected gauges to be representative of the environmental and hydrological conditions in the regions, except for extremely small catchments with an area < 22 km^2 that likely have higher cease-to-flow occurrence."

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. We have now included the following new text in the revised manuscript (lines 170-173) to reflect this:

"Only six of the 25 gauges in SEQ and three 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."

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.

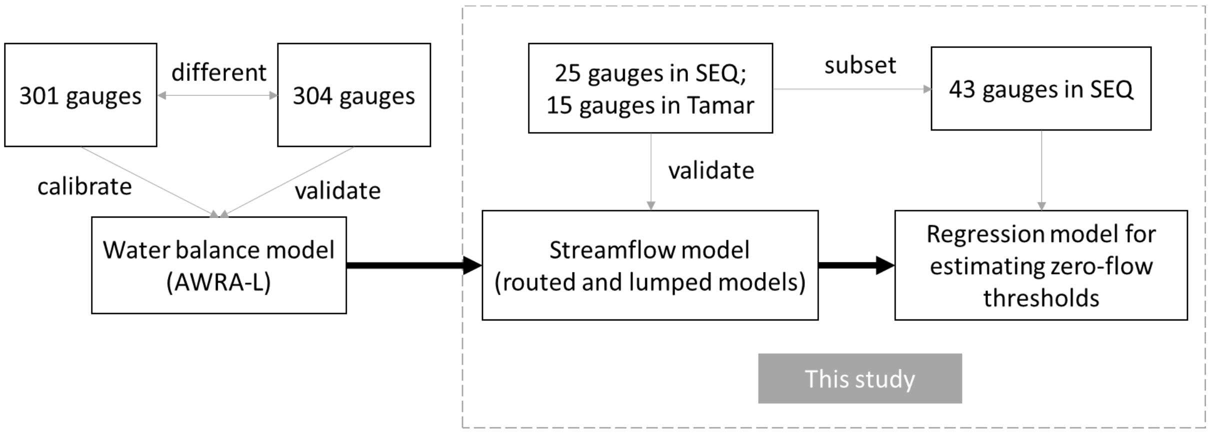


Figure R1. Schematic illustration of the streamflow gauge sets used in the different modelling processes described in this paper.

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:

As detailed in our responses to similar comments by Reviewer # 2, 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 WaterDyn (initially called "AWAP") in terms of the ability to estimate flow intermittency. As the monthly flow model WaterDyn outputs monthly average flow, the zero value of monthly flow means all days in the month have zero flows. Additionally, we have now evaluated an alternative method to aggregate the modelled daily flow intermittency into monthly flow intermittency (lines 213-219). We assessed the effect of considering a month as non-flowing when at least one day in the month had 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 bounds 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 have been 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 114-118): "The gauges were widely dispersed throughout each study area and encompassed a range of stream sizes, catchment areas (22-3,881 km^2 in SEQ; 33-3,294 km^2 in Tamar) and flow regime types, ranging from highly intermittent to perennial streams (see results). Therefore, we regard the selected gauges to be representative of the environmental and hydrological conditions in the regions, except for extremely small catchments with an area < 22 km^2 that likely have higher cease-to-flow occurrence."

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 the 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 have revised 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 3.

Specific comment:

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?

Authors reply:

This thresholds of cease-to-flow duration were chosen to make the flow intermittency estimates from daily and monthly models comparable, assuming 1 month = 30 days.

Specific comment:

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

Authors reply:

We have now provided unit for each of the flow metrics in Figure 4 and 6 (Figure 3 and 5 in the original submission).

References:

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.

Evaluating a landscape-scale daily water balance model to support spatially continuous representation of flow intermittency throughout stream networks

Songyan Yu[1], Hong Xuan Do[2,3,4], Albert I. J. M. van Dijk[5], Nick R. Bond[6], Peirong Lin[7], Mark J. Kennard[1]

[1]Australian Rivers Institute and School of Environment and Science, Griffith University, Nathan, Queensland, Australia
[2]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, Australia
[3]Faculty of Environment and Natural Resources, Nong Lam University, Ho Chi Minh City, Vietnam
[4]School for Environment and Sustainability, University of Michigan, Ann Arbor, Michigan, USA
[5]Fenner School of Environment & Society, The Australian National University, Canberra, Australia
[6]Centre for Freshwater Ecosystems, La Trobe University, Wodonga, Victoria, Australia
[7]Department of Civil and Environmental Engineering, Princeton University, New Jersey, USA

Correspondence to: Songyan Yu (sunny.yu@griffith.edu.au)

Abstract

There is a growing interest globally in the spatial distribution of intermittently flowing streams and rivers, and how their spatial extent varies in relation to climatic factors. However, deriving consistent information on the extent of flow intermittency within river networks is hampered by the fact that streamflow gauges are often sparsely distributed and more often be located within the most perennial parts of the river network. Here, we developed an approach to quantify catchment-wide streamflow intermittency over long timeframes and in a spatially explicit manner, using readily accessible and spatially contiguous daily runoff data from a national-scale water balance model. We examined the ability of the water balance model to simulate streamflow in two hydro-climatically distinctive (subtropical and temperate) regions in Australia, with a particular focus on low flow simulations. We also evaluated the effect of model time step (daily vs. monthly) on flow intermittency estimation to inform future model selection. The water balance model showed better performance in the temperate region characterised by steady baseflow than in the subtropical region with flashy hydrographs and frequent cease-to-flow periods. The model tended to overestimate low flow magnitude mainly due to ~~both~~ overestimation of gains (e.g. groundwater release to baseflow) ~~and underestimation of losses (e.g. transmission losses)~~ during low-flow periods. Modelled patterns of flow intermittency revealed highly dynamic behaviour in space and time, with ~~intermittent flows~~ cease-to-flow events affecting between 29 % and 80% of the river network over the period of 1911-2016, using a daily streamflow model. The daily flow model did not perform better than the monthly flow model in quantifying flow intermittency at a monthly time step, and model selection should depend on the intended application of the model outputs. Our general approach to quantifying spatio-temporal patterns of flow intermittency is transferable to other parts of the world, and can inform hydro-ecological understanding and management of intermittent streams where limited gauging data are available.

Keywords: AWRA-L model, flow regime, river routing, Australia, temporary streams

1 Introduction

Intermittent streams that cease to flow for some period of most years are prevalent within river networks globally (Acuña et al., 2014; Datry et al., 2014). Their spatial extent is projected to increase in regions experiencing drying trends related to climate change and water extraction for human uses (Larned et al., 2010). Intermittent streams have seen increasing research interest over the past decade (e.g. Costigan et al., 2016; Fritz et al., 2013; Gallart et al., 2017; Leigh et al., 2016), and there is a growing interest in conserving these unique ecosystems. The scarcity of spatially-explicit information on flow intermittency has been identified as one of the key issues confronting intermittent stream management (Acuña et al., 2017). Flow intermittency exerts primary control on the transfer of energy, materials and organisms by surface water through river networks (Jaeger et al., 2019) and is a key driver of riverine ecosystems (Datry et al., 2017; Poff et al., 1997; Stanley et al., 1997). Therefore, i~~I~~mproved understanding of temporal and spatial patterns in flow intermittency is fundamentally important for effective river management. ~~Flow intermittency exerts primary control on the transfer of energy, materials and organisms by surface water through river networks (Jaeger et al., 2019) and is a key driver of riverine ecosystems .~~

Previous studies have predominantly relied on the use of gauged streamflow data to make inferences about the distribution of intermittent streams in many regions, including France (Snelder et al., 2013), Australia (Bond and Kennard, 2017; Kennard et al., 2010b), Spain and North America (de Vries et al., 2015). However, spatial biases in the distribution of stream gauges used in such studies may give misleading impressions of spatial patterns and extent of streamflow intermittency (Snelder et al., 2013). Alternative methods for quantifying the extent of intermittent flow include citizen-observation networks supported by regular reports from trained volunteers (Datry et al., 2016; Turner and Richter, 2011), the use of electrical arrays by measuring the electrical conductivity of the streambed (Jaeger and Olden, 2012), development of predictive models for intermittent streams (González-Ferreras and Barquín, 2017), and deployment of unmanned aerial systems (Spence and Mengistu, 2016). These alternatives are generally appropriate over small spatial extents and short time frames~~,~~ but are difficult to scale up to larger areas to quantify flow intermittency in space and time. Satellite remote sensing-based quantification of flow intermittency (Hou et al., 2019) can cover larger spatial extents, but for now, remains applicable only to relatively large rivers (> 30 m in the case of Landsat imagery) and can be affected by factors such as vegetation and cloud obstruction.

Spatially contiguous runoff data derived from water balance models provide another potential alternative to quantify spatio-temporal variations in flow intermittency. For example, Yu et al. (2018) used runoff simulations obtained from a water balance model WaterDyn (Raupach et al., 2009) to generate spatially explicit and catchment-wide estimates of streamflow intermittency, but only at a relatively coarse monthly time step. Depending on the application, flow simulations at a finer temporal scale (e.g. daily) may be necessary to capture the dynamic aspects of hydrological processes. These kinds of simulations are important to better understand the causes of flow intermittency at multiple spatial scales~~ better~~, ~~potentially ~~and enable~~ing~~ ~~more~~ ecologically-relevant characterisation of ~~hydrology~~streamflow properties~~,~~ such as the magnitude, frequency, duration, and ~~change~~ rate of ~~rate~~ change ~~of ecologically important~~in high or low flow events. However, there are few examples of studies quantifying spatial and temporal variation in flow intermittency across river networks using spatially contiguous daily flow data. That is partly because streamflow simulation is more challenging at a daily versus monthly time step due to higher uncertainties in input data at this finer temporal scale (Wang et al., 2011).

Water balance models at a daily time step have been increasingly developed around the world (Bierkens et al., 2015; Lin et al., 2019). One prominent regional example is the Australian Water Resource Assessment Landscape (AWRA-L) model (van Dijk, 2010). The AWRA-L model has been developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Australian Bureau of Meteorology (BoM) to simulate the terrestrial water balance across Australia at a daily time step (Frost et al., 2016; van Dijk, 2010). The model yields spatially contiguous daily water availability values gridded at a spatial resolution of 0.05 arc-degree spatial resolution (approximately 5 × 5 km) (Frost et al., 2016). The development of such water balance models in Australia and other parts of the world provides the potential to quantify spatial and temporal variation in runoff, and hence flow intermittency, at a daily time step. However, this requires an effective and efficient conversion process to translate gridded runoff estimates to accumulated streamflow estimates down the river network. This is especially challenging for large study areas due to lags in runoff, which can influence the timing of flow peaks and rates of recession. Additionally, many national-scale water balance models, including AWRA-L, were calibrated on a large domain that covers multiple climate conditions (Viney et al., 2015), providing a best "~~“~~average"~~”~~ response but potentially inconsistent accuracy of runoff simulations within particular climate domains. As the predictive performance for ungauged basins strongly depends on climate settings, this compromise raises the question as to whether such models can be used to quantify flow intermittency over multiple climate conditions. Although substantial efforts have been made in evaluating hydrological models in different climate conditions (Do et al., 2019; Gudmundsson et al., 2012; Lin et al., 2019; Zaherpour et al., 2018), a limited number of such studies have focused particularly on model performance during low flow conditions, which is particularly important for flow intermittency quantification.

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. The translation of gridded runoff to aggregated streamflow/discharge on vector river flow lines make AWRA-L outputs more accessible to fluvial geomorphologists and ecologists, who may intend to relate daily hydrologic characteristics of rivers to a broad range of physical and ecological phenomena. 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. A previous study conducted at the monthly time step (Yu et al., 2018) was used to benchmark flow intermittency estimated from the AWRA-L~~daily~~ model.

2 Study areas

This research was conducted in two hydro-climatically distinctive regions: South-east Queensland and the Tamar River catchment in Tasmania (Fig. 1). The South-east Queensland (SEQ) region is located in the eastern part of Australia (Fig. 1a) and comprises five major coastal river basins with a total area of 21,331 km^2 (Fig. 1b) (Australian Bureau of Statistics, 2011). SEQ has 7,229 stream segments and their corresponding sub-catchments according to the Australian Hydrologic Geospatial Fabric (Geofabric), with the minimum upstream drainage area of 1.5 km^2. SEQ is a region of transitional temperate to subtropical climate (Fig. 1a) with substantial inter- and intra- annual variation in rainfall. The majority of rainfall and streamflow usually occur in the summer months of January to March, often followed by a second minor discharge peak between April and June, but high and low flows may occur at any time of year (Kennard et al., 2007). Thus, there are a range of flow regimes with many streams being intermittent to varying degrees. The Tamar River catchment (Tamar) is located in Tasmania, an island state off Australia'~~'~~s south coast (Fig. 1a, c). It drains a catchment area of approximately 11,215 km^2, comprising over one-~~-~~fifth of Tasmania'~~'~~s land mass and is located in north-east and central Tasmania. According to climate data from BoM (http://www.bom.gov.au/climate/data), Tamar is ~~characterized~~ characterised by a temperate climate condition, of which rainfall is relatively evenly distributed throughout the year and most months receive very similar averages~~, according to climate data from BoM (http://www.bom.gov.au/climate/data)~~.

[Figure 1 is about here]

3 Data and Methodology

3.1 Streamflow gauge data

~~Observed~~ Gauged streamflow data were sourced from the BoM water data website (http://www.bom.gov.au/waterdata). Based on streamflow data availability, a total of 25 gauges in SEQ and 15 gauges in Tamar were selected (Fig. 1b, c). All gauges have less than 0.5 % missing values over the period from 01/01/2005 to 31/12/2017. The gauges were widely dispersed throughout each study area and encompassed a range of stream sizes, catchment areas (22-3,881 km^2 in SEQ; 33-3,294 km^2 in Tamar) and flow regime types, ranging from highly intermittent to perennial streams (see results). Therefore, we regard the selected gauges to be representative of the environmental and hydrological conditions in the regions, except for extremely small catchments with an area < 22 km^2 that likely have higher cease-to-flow occurrence.~~.~~ Daily flow data for the study period were used to validate flow simulations and test flow intermittency estimates.

3.2 Conversion from spatially contiguous runoff to streamflow

~~Simulated daily runoff from the AWRA-L model (version 5) were downloaded from BoM (http://www.bom.gov.au/water/landscape).~~ AWRA-L is a daily 0.05° grid-based distributed water balance model that is conceptualised as a small catchment. It simulates the water flow through the landscape from the rainfall entering the grid cell through the vegetation and soil and then out of the grid cell through evapotranspiration, surface water flow or lateral flow of groundwater to the neighbouring grid cells (Viney et al., 2015). AWRA-L was calibrated and validated at the national scale during its development by CSIRO and BoM, with 301 gauges used for calibration and a different set of 304 gauges used for validation (Zhang et al., 2013). Simulated daily runoff from the AWRA-L model (version 5) was~~ere~~ downloaded from BoM (http://www.bom.gov.au/water/landscape~~; last access: May 25, 2020~~). These data are in gridded format and require conversion to streamflow for each sub-catchment by aggregating the gridded runoff data with a hierarchically nested catchment to simulate streamflow throughout river networks. The conversion process may or may not need to use a river routing model to propagate streamflow through river networks, partly depending on the size of the catchment of interest (Robinson et al., 1995). If streamflow simulated with a routing model shows little difference to that without a routing model, ~~can be simulated at an acceptable accuracy without a routing model,~~ then the conversion process is more efficient without a routing model, and the readily available runoff data can be more accessible for potential applications, such as flow characterisation for ungauged stream segments. In addition, a conversion process involving a routing model can be computationally-intensive and usually requires parallel computing to speed up the calculations (David et al., 2011b). Therefore, in this study, we applied two approaches to determine an effective and efficient runoff-streamflow conversion. The first approach coupled a river routing model to the water balance model, and its effects on flow simulations are compared to the model performance of a lumped model, which was operated without any river routing (Fig. 2). As the conversion process was achieved using the "~~“~~*catchstats*"~~”~~ package (https://github.com/nickbond/catchstats) in the R programming language (R Development Core Team, 2017), so the second approach was to speed up the conversion process by incorporating parallel algorithm to exiting functions of that package. The conversion process was run on a Griffith University High-~~-~~Performance Computing node with 12 cores and RAM 12 GB.

[Figure 2 is about here]

The hierarchically nested catchment dataset used in this study was sourced from the Geofabric dataset (Stein et al., 2014), which provides a fully connected and directed stream network derived from the national 9 arc-second DEM and flow direction grid (~250m resolution), and associated catchment hierarchy at the national scale. The routing model applied in this study was the Routing Application for Parallel c~~C~~omputatI~~i~~on of Discharge (RAPID) model (David et al., 2011b). RAPID solves the matrix-based Muskingum equation to route flow through each stream of the river network and performs streamflow computation for every stream segment of a river network, including ungauged streams. Various water balance models have been used in combination with RAPID (Follum et al., 2017; Lawrence et al., 2011; Lin et al., 2019).

To test the effects of river routing, we first calculated ~~summary~~ a series of flow metrics ~~describing the critical components of hydrological variation across average, high- and low-flow conditions~~ (Table 1) for flow simulations from both the lumped and coupled models. The calculated flow metrics are commonly used to describe the critical components of flow regimes across average, high, and low flow conditions, including flow magnitude and variability, the timing, frequency and duration of high and low ~~extreme~~ flows, ~~the frequency and duration of extreme flows,~~ and rates of changes in flow events (Olden and Poff, 2003; Poff et al., 1997). Calculation of these streamflow characteristics allows ~~to~~a comprehensive~~ly~~ assessment ~~understand~~ of the effects of river routing on streamflow simulations in the two regions. ~~Then w~~We then ~~conducted~~ applied Student's *t*-test for each flow metric to ~~identify~~ determine whether the inclusion of river routing can improve model accuracy based on a significance level of 5 %. We used the 10th and 90th percentiles of daily flows to respectively describe low-flow and high-flow thresholds (Gudmundsson et al., 2019; Leigh and Datry, 2016). The calculation process was conducted with the "*hydrostats*" package in the R language (Bond, 2016).

[Table 1 is about here]

3.3 Accuracy assessment of modelled streamflow

To evaluate overall model performance in streamflow simulations, we calculated the modified Kling-Gupta efficiency (KGE; Kling et al, 2012) ~~(Equation 1) (Kling et al., 2012)~~ between the observed and modelled streamflow for all gauges in SEQ and Tamar (Eq. (1)). KGE is an integrated skill metric, which measures the Euclidean distance between a point and the optimal point that has the maximum correlation coefficient, zero variability error and zero bias error between the simulated and observed streamflow (Gupta et al., 2009; Kling et al., 2012). 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 (Knoben et al., 2019). To evaluate model performance in different components of flow regimes, w~~W~~e also calculated each summary flow metric (Table 1~~Table 1~~) for observed and modelled streamflow data at all gauges in SEQ and Tamar and visually compared their frequency distributions.~~ Therefore, t~~The use of KGE ~~is supposed to~~ provides an overall assessment of AWRA-L model performance and the flow metrics in Table 1 are ~~applied~~used to comprehensively evaluate the model accuracy for various components of flow regimes, including the flow metrics related to low flows. Only six of the 25 gauges in SEQ and three 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.

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \qquad \text{(Equation 1)}$$

$$\beta = \frac{\mu_s}{\mu_o};\ \gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}$$

where KGE is the modified KGE-statistic (dimensionless), r is the correlation coefficient between simulated and observed runoff (dimensionless), β is the bias ratio (dimensionless), γ is the variability ratio (dimensionless), μ is the mean runoff in $m^3 s^{-1}$, CV is the coefficient of variation (dimensionless), σ is the standard deviation of runoff in $m^3 s^{-1}$, with subscripts s and o referring to simulated and observed runoff values, respectively.

Furthermore, considering that this study aims to apply flow simulations to quantify flow intermittency, the model accuracy of low flow simulation is particularly important. The study period (01/01/2005-31/12/2017) was considered sufficient to assess low flows. The 13 year study period is close to a discharge record length of 15 years which Kennard et al. (2010a) 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), and includes a significant low-flow period. A preliminary analysis showed that AWRA-L modelled streamflow was ~~very~~ sensitive to rainfall events, relative to the response of observed flow (Fig. 3~~2~~). This finding indicates that over-responsiveness of AWRA-L to rainfall may potentially contribute to overestimation of low flow. We hypothesis~~z~~ed that this over-responsiveness is partly due to overestimation of "~~“~~*in situ*"~~”~~ gains to low flow discharge (e.g. groundwater release to baseflow) as well as underestimation of transmission losses (e.g. depression filling and evapotranspiration) during water movement through various flow paths in the stream network (Davison and van der Kamp, 2008). Given that we do not have access to the ~~the underlying~~ AWRA-L model ~~models~~ to directly adjust model parameters, we instead compared the observed and modelled low flow magnitude at all gauges in the two study areas along the gradient of their catchment areas (22-3,881 km^2 in SEQ; 33-3,294 km^2 in Tamar) to test this hypothesis. We expect that 1) if the difference in low flow magnitude occurs at all gauges, then low flow overestimation can be at least attributed to the overestimation of gains to low flow discharge. Alternatively, 2) if the difference in low flow magnitude occurs towards the downstream of the catchment, then low flow overestimation may be related to underestimation of transmission losses.

[Figure 3~~2~~ is about here]

3.4 Quantifying flow intermittency using spatially contiguous flow simulations

~~Due to~~Given the fact that water balance models often over-predict the magnitude of very low flows due to the difficulties of quantifying hydrological processes influencing low flow discharge (Smakhtin, 2001; Staudinger et al., 2011; Ye et al., 1997), we adopted the same method used in Yu et al. (2018) to estimate a threshold of zero flow from the model that ~~correlated with~~related measured ~~cease to~~ no-flow duration at each gauge to catchment environment variables. This involved three steps.

1) We used linear regression to model the ~~cease to~~ no-flow duration at each gauge as a function of catchment environment variables. The environmental variables were the same as those in Yu et al. (2018), and included variables

related to climate (annual daily maximum temperature), catchment geology topography (catchment area, catchment average slope, and catchment average elevation), and catchment soil properties (catchment average saturated hydraulic conductivity).

2) We then used the predictive models to extrapolate estimates of overall flow permanency (in terms of the proportion of days with flow) to each segment in the entire river network.

3) For each segment, the time-series of daily runoff was truncated (flows below the threshold were set to "~~“~~0"~~”~~) by adopting an appropriate threshold ~~for~~ of '~~‘~~zero flow~~s~~'~~’~~ that preserved the proportion of days with flow as estimated at Step 2.

This truncation was only conducted in SEQ as most gauges in the Tamar catchment had perennial flow. Based on the modelled daily streamflow from AWRA-L, we calculated annual flow intermittency as the number of zero-flow days per year over the period of 2005-2016~~.~~. To evaluate the effect of time step (daily vs. monthly) on the relative performance of the model in replicating observed patterns of cease-to-flow periods~~In examining the model outputs~~, we further ~~also compared the patterns of cease to flow from the daily AWRA-L model, with those derived by~~ aggregated~~ing~~ daily outputs to a monthly time step (termed "~~“monthly-aggregated~~ monthly AWRA-L"~~”~~ hereafter, Fig. 2)~~, as well as results from the AWAP water balance model, which operates only at a monthly time step. For the monthly-aggregated AWRA-L outputs~~.~~,~~ We tried two different aggregation methods. One ~~is~~considered that the flows for a month were zero when at least one day in that month had zero flow (termed "monthly AWRA-L_01" hereafter)~~all days in a month had to have zero flow for the flows for that month to be zero~~, and the other considered ~~is~~ that all days in a month ~~had~~must ~~to~~ have zero flow for ~~the flows for~~ that month to be zero (termed "monthly AWRA-L_30" hereafter)~~ as the zero value of monthly average flow means all days in the month have zero flows~~. These two methods together should provide both upper and lower bounds of comparing daily and monthly models in estimating flow intermittency. The monthly AWRA-L outputs were compared with results from the WaterDyn water balance model, which operates only at a monthly time step (Fig. 2). The ~~AWAP~~ WaterDyn model was developed to provide monthly spatially contiguous water balance data at the Australian continental scale by CSIRO and BoM with a similar model structure to AWRA-L (Raupach et al., 2018), and has been used to quantify the spatial and temporal patterns of flow intermittency in SEQ following similar methods to this study ~~in our previous research~~ (Yu et al., 2018). Monthly flow intermittency estimated from WaterDyn was thus used to benchmark results from the monthly AWRA-L. Modelled flow intermittency from all three sources (i.e. daily and monthly~~-aggregated~~ AWRA-L, and monthly WaterDyn~~AWAP~~) was also tested against the measured flow intermittency derived respectively from daily and monthly observed streamflow data at gauged locations in SEQ.

~~-~~Taking ~~the~~ advantage of the modelled long-term runoff data from AWRA-L over the period of 1911-2016, we further quantified spatial and temporal dynamics of flow intermittency for every stream segment within SEQ, and compared the results with those from the WaterDyn~~AWAP~~ model over the same period (Yu et al., 2018). The spatial pattern of flow intermittency was represented by the mean annual number of zero flow days across the period of 1911-2016 for ~~the~~ AWRA-L and by the mean annual number of calendar months for ~~the~~ WaterDyn~~AWAP~~. The temporal pattern of flow intermittency was expressed as the proportion of streams with flow intermittency > 30 days or 1 month (termed "~~“~~intermittent streams"~~”~~ hereafter) for ~~the~~ AWRA-L and WaterDyn~~AWAP models~~, respectively.

4 Results

4.1 Negligible effects of river routing on daily flow simulations

The lumped and coupled (i.e. with routing) models using AWRA-L simulated runoff were run in both SEQ and Tamar, and produced similar values for various flow metrics between the lumped and coupled in both regions (Fig. 4~~3~~; *p* values were greater than 0.50 for most flow metrics based on *t*-test results). There were noticeable differences for three flow metrics related to low flows (the variability in timing, the frequency and the duration of low flow spells), but these differences were not statistically significant at the 5 % level. These results suggested that the routing algorithm has negligible effects on flow simulations in our study areas, which is reasonable because of the small size of the two watersheds. Therefore, in the subsequent analysis, we only used the results from the AWRA-L lumped model as it is relatively less computationally intensive and was able to maintain a comparable model performance to that of the coupled model taking into account the routing effect.

[Figure 4~~3~~ is about here]

4.2 Accuracy assessment of modelled streamflow in SEQ and Tamar

The overall accuracy of streamflow estimated by AWRA-L lumped model (referred to as "~~“~~modelled streamflow"~~”~~ in this section) was evaluated for 25 gauges in SEQ and 15 gauges in Tamar. Results suggested a fair to good explanatory value across all gauges (Fig. 5~~4~~). The KGE values varied across the 25 gauges in SEQ, ranging from -0.19 (Gauge No. 145103) to 0.76 (143901), with a median value of 0.42, while the model generally performed ~~generally~~ better in Tamar and the KGE values ranged from 0.11 (18219.1) to 0.71 (852.1) across 15 gauges, with a median value of 0.47 (Fig. 5~~4~~). However, no significant difference was found in the overall model performance between the two hydro-climatically distinctive regions, according to the two-sample Student'~~’~~s *t*-test (t = -1.46, p = 0.15).

[Figure 5~~4~~ is about here]

~~When it comes to~~Concerning model performance in simulating different components of flow regimes, t~~T~~he modelled streamflow in SEQ revealed a generally good match with the observed streamflow across all high-flow metrics and the magnitude of average flow, but the model tended to overestimate the variation in the magnitude of average flow (almost two times higher on average), report earlier timing of low flows, overestimate the frequency (48 % higher), and underestimate the duration (74 % lower) of low flows (Fig. 6~~5~~). Compared to the model performance in SEQ, the flow simulations in Tamar showed slightly better performance, predicting well not only for the high-flow metrics but also for the metrics related to average flows (Fig. 6~~5~~). However, flow simulations in Tamar also exhibited slightly earlier estimations for the timing of low flow spells (13 % earlier), overestimations for low flow spell frequency (92 % lower on average) and underestimation for low flow spell duration (58 % lower) (Fig. 6~~5~~).

[Figure ~~5~~6 is about here]

Varying degrees of difference in the magnitude of low flow between the observed and modelled were found ~~at nearly all~~among the gauges. ~~At the same time,~~ T~~t~~here appeared to be a tendency toward larger differences with increasing catchment area in ~~both~~ SEQ ~~and~~ but not in Tamar (Fig. 7~~6~~). The models appeared to ~~both~~ over-estimate "~~“~~*in situ*"~~”~~ gains to low flow in some reaches in both regions, while ~~also~~ under-estimating transmission losses in SEQ, suggesting that~~;~~ over-estimation of "*in situ*" gains in AWRA-L likely ~~these both~~ contribute to the overall overestimation of low flow in downstream catchments.

[Figure 7~~6~~ is about here]

4.3 Quantifying flow intermittency using flow simulations

We calculated annual flow intermittency at gauged locations in SEQ using three sources of modelled flow (daily and monthly~~-aggregated~~ AWRA-L, and monthly WaterDyn ~~AWAP~~). Annual flow intermittency calculated using daily AWRA-L flow (i.e., the average number of cease-to-flow days per year) was tested against ~~the~~ annual flow intermittency estimated using observed data (Fig. 8~~7~~a~~)~~). The AWRA-L model displayed the potential to be used to estimate flow intermittency at a daily time step, with a ~~good~~ fair match with the observed flow intermittency (R^2= 0.56) in SEQ. Nonetheless, the model tended to overestimate flow intermittency for gauges located in relatively wet areas (e.g. ≤ 40 days of flow intermittency per year) while underestimating for gauges located in relatively dry areas (e.g. ≥ ~~more than~~ 40 days of flow intermittency per year~~Fig.~~8~~7a~~).

Fig~~.~~ure 8b shows ~~Annual~~ annual flow intermittency calculated using monthly~~-aggregated~~ AWRA-L flow and monthly WaterDyn~~AWAP~~ flow. In this case, annual flow intermittency was ~~-~~defined as the average number of months characterized with ~~intermittent~~zero flow~~were also compared with the observed (Fig.87b)~~. The WaterDyn~~AWAP~~ model showed much more accuracy than the two aggregation methods based on the monthly~~-aggregated~~ AWRA-L model in estimating flow intermittency (R^2 = 0.53~~, and~~ 0.43 and 0.32~~,~~ respectively for ~~the two models~~monthly WaterDyn, monthly AWRA-L_01 and monthly AWRA-L_30). More specifically, the WaterDyn~~AWAP~~ model displayed a similar estimation pattern as the daily AWRA-L model: overestimation in relatively wet areas while underestimation in relatively dry areas. By contrast, not surprisingly, the two aggregation methods showed the upper and lower bounds of flow intermittency estimates from the monthly AWRA-L model: ~~the~~ monthly AWRA-L_01 overestimated flow intermittency and monthly~~-aggregated~~ AWRA-L_30 ~~model~~ underestimated flow intermittency at nearly all gauges (Fig. 8~~7~~b).

[Figure 8~~7~~ is about here]

The spatial patterns of flow intermittency derived from the daily AWRA-L and monthly WaterDyn flow simulations aligned well only for the main stems and some coastal streams, which were predicted to flow for most of the time (Fig. 9~~8~~). There was a noticeable difference for inland streams, especially those lower~~-~~order streams. More specifically, in the western Brisbane River catchment and the South Coast River catchment, most inland streams were predicted by the daily model to flow for longer period than by the monthly model; while in the Pine River catchment and the Logan-Albert River catchment, many inland streams were predicted by the daily model to flow for a shorter period (Fig. 9~~8~~a). Compared to the WaterDyn~~AWAP~~ model, fewer streams were predicted to experience extremely long dry events. But more streams on average (60 % vs. 49 % for the AWRA-L and WaterDyn~~AWAP~~ model, respectively) were predicted to flow intermittently (> 30 days or > 1 month) to varying degrees in SEQ, which suggested that flow intermittency was prevalent in SEQ.

[Figure 9~~8~~ is about here]

Temporally, the daily model estimated that the proportion of intermittent streams in SEQ varied from 29 % to 80 % over the study period (1911-2016), while the monthly model estimated the range to be from 3 % to 80 % estimated during the same time span (Fig. 10~~9~~). The two temporal patterns were temporally correlated (r = 0.71) and similar predictions with higher proportions of intermittent streams were estimated for the dry years by both models. Compared to dry years, the two models exhibited greater differences in predictions for the wet years, where the daily model tended to predict more proportion of intermittent streams. Overall, the daily model suggested a drier history in SEQ in terms of flow intermittency than the monthly model. The models successfully identified the extensive drying associated with severe drought periods. Notably, the Widespread drought (1914-1920), WWII drought (1939-1946) and Millennium drought (2001-2009) were all visible in both two sets of model predictions.

[Figure 10~~9~~ is about here]

5 Discussion

The scarcity of information on the spatial and temporal extent of flow intermittency has been identified as a major barrier for ecologists and managers to understand and protect intermittent stream ecosystems (Acuña et al., 2017). This barrier has been partly overcome in previous studies by using statistical models relating flow intermittency to surrounding environmental variables (Bond and Kennard, 2017; González-Ferreras and Barquín, 2017; Jaeger et al., 2019; Snelder et al., 2013), but most of these studies focused on only the spatial variations in flow intermittency, except for Jaeger et al. (2019), overlooking its temporal aspects. This issue becomes particularly urgent in the time when flow regimes of streams are changing worldwide, mainly in response to climate change and water extraction for human uses (Chiu et al., 2017; Jaeger et al., 2014). Monthly runoff data have been recently used to quantify flow intermittency for entire river networks (Yu et al., 2018), and the current study takes one step further to use daily runoff data in flow intermittency estimation, which is especially needed for studies aimed at quantifying ecological responses to short term flow events (e.g. frequent zero flow events within a month). In this study, we comprehensively examined the ability of a daily water balance model to simulate streamflow, with a particular focus on low flow simulations. We also investigated how to better choose water balance models to estimate flow intermittency by answering the question that whether daily flow models outperform monthly flow models at both daily and monthly scales. Our study can not only inform the estimation of the spatial distribution of intermittent flow~~,~~ but also reveal the temporal dynamics of intermittent streams over long timeframes.

5.1 Efficient runoff-streamflow conversion for eco-hydrological research

Effects of river routing on daily flow simulations were found negligible in SEQ and Tamar, most probably due to the relatively small size of the two catchments and the relatively short length of even the longest streams (Cunha et al., 2012). This can be verified with the concept of time of concentration, which is commonly used to measure the time needed for water to flow from the most remote point in a catchment to the catchment outlet. By following the formula for calculating the time of concentration proposed by Pilgrim and McDermott (1982) that has been widely used in Australia, we found the time of concentration in SEQ is around 33 hours, only slightly more than a daily time step (24 hours). This illustrates why it is difficult for a daily time-step routing model to effectively capture routing lags in our study domain. Negligible effect of river routing on flow simulations was also observed in previous studies (David et al., 2011a). Robinson et al. (1995) found that catchment size is a primary factor to determine which process, the hillslope or the channel network transport component, characteris~~z~~e lags in catchment runoff down the river network. In areas such as SEQ and Tamar that have a relatively small catchment size, the inclusion of channel network transport contributes little to the improvement of flow simulations. The negligible effect of river routing in SEQ and Tamar allowed us to simplify the simulation of daily flows without coupling with a river routing model. Hence we were able to use existing runoff outputs from the daily AWRA-L model. Arguably, similar opportunities exist in other small catchments.

5.2 Accuracy assessment of modelled daily streamflow in two hydro-climatically distinctive regions

Daily streamflow estimates showed a fair to good overall alignment with the observed flows in both SEQ and Tamar, with all gauges showing that flow simulations were better estimates than the mean of observations (KGE~~EG~~ ≥ -0.41 at all gauges). Interestingly, although streamflow was more accurately simulated in the Tamar than in SEQ (the median values of K~~E~~GE were 0.47 and 0.42, respectively), the differences between the two hydro-climatically distinctive regions were relatively small. Despite ongoing efforts to calibrate AWRA-L against a set of reference scales distributed across the continent (Viney et al., 2015), this finding was reassuring given the much higher variability in rainfall and soil moisture in SEQ, factors that typically can lead to a more nonlinear streamflow response to rainfall (Poncelet et al., 2017), which possibly undermines the ability of water balance models to reliably predict runoff (Sheng et al., 2017). These results hence bode well for the application of AWRA-L outputs across diverse hydroclimatic regions.

When looking into the model performance for specific components of the flow regime, average- and high-flow metrics were both modelled well in Tamar, while only high-flow metrics were modelled well in SEQ. However, in both regions, the AWRA-L model showed poor performance in low flow metrics: overestimating the frequency and underestimating the duration of low flows, consistent with previous studies (Costelloe et al., 2005; Ivkovic et al., 2014; Staudinger et al., 2011; Ye et al., 1997). This suggests that the AWRA-L model is a generally robust model in predicting average- and high-flows, but still needs some improvement to better simulate low flows. Runoff generation processes can vary substantially through space and time due to such factors as variations in soil depth, antecedent soil moisture and groundwater connectivity, and this can influence spatio-temporal variations in low flow characteristics, including streamflow intermittency (Zimmer and McGlynn, 2017). However, it is unknown the extent to which this contributed to uncertainty in the simulation of low flows and estimation in streamflow intermittency in this study. The uncertainty of AWRA-L in low flow simulations can be linked to its over-responsiveness to rainfall, partly caused by ~~both~~ overestimation of "~~“~~*in situ*"~~”~~ gains and underestimation of transmission losses to low flow discharge, as shown in SEQ. Previous studies found that lateral flow exchange between grid cells of land surface models (e.g. AWRA-L) plays a significant role in redistributing soil water (Kim and Mohanty, 2016), and thus may improve "~~“~~*in situ*"~~”~~ surface/subsurface runoff simulations (Lee and Choi, 2017). On the other hand, hydrological processes involved in transmission losses have been extensively discussed (Jarihani et al., 2015; Konrad, 2006), and ~~recently many~~ studies have developed methods to calculate transmission losses for better flow simulations (Costa et al., 2012; Lange, 2005). Therefore, low flow simulations by AWRA-L can possibly be improved by incorporating lateral flow exchange algorithms and better accounting for hydrological process such as evapotranspiration from riparian vegetation and infiltration into channel beds. This improvement is made more likely as AWRA-L has been released as a Community Modelling System (https://github.com/awracms/awra_cms), which allows co-development by the research community.

5.3 Choose appropriate water balance models to quantify spatio-temporal dynamics of flow intermittency

To mitigate the overestimation of low flow simulations, we identified segment-specific zero-flow thresholds and used the corrected runoff estimates to quantify flow intermittency. The daily AWRA-L flow showed promise for estimating flow intermittency at a daily time step, while the monthly WaterDyn~~AWAP~~ model was better than the monthly~~-aggregated~~ AWRA-L model in flow intermittency estimation at a monthly time step. This suggests that monthly flow models can sometimes outperform daily flow models in quantifying flow intermittency, depending on the intended temporal resolution of the analysis. For example, daily flow models may be appropriate for studies aimed at quantifying ecological responses to short term flow events, while monthly flow models are more suitable for research requiring the average degree of flow intermittency at a large spatial or temporal scale, such as examining the effect of flow intermittency on aquatic/streamside vegetation or species distributions (Stromberg et al., 2005). In addition, ou~~t~~r study also suggested that the suitability of a monthly model (WaterDyn) for monthly resolution of analysis was not challenged by a daily model (AWRA-L) simply through aggregating daily streamflow simulations to a monthly time step. The aggregation methods used here applied~~ying~~ one day or 30 days as a threshold and, respectively, either substantially overestimated or underestimated flow intermittency.~~ Other thresholds between 1 and 30 days are hard to be selected due to lack of appropriate reasons. Therefore, until a better aggregation method is found,~~ This suggests that~~ we suggest that a certain~~ the temporal resolution of analysis (e.g. for flow intermittency estimation) ~~is~~should be dictated by the ~~better done with the same~~ resolution as the~~of~~ streamflow data used in the analysis.

Spatially contiguous runoff data were used in this study ~~as an alternative method~~ to quantify spatial and temporal dynamics of flow intermittency, shedding light on the temporal aspect of flow intermittency that has been often overlooked in previous studies. Annual flow intermittency was shown to vary significantly from year to year, ranging from 29 % to 80 % for the AWRA-L model. Although there ~~is difference~~are differences in the temporal patterns of estimated flow intermittency between the AWRA-L and WaterDyn~~AWAP~~ models, neither model estimated intermittency to have a clear trend over the past century. However, there is still the concern about the potential shift of some perennial streams to intermittent streams due to climate change and intense human activities, as it has been evident in several regions where the number of low-flow and/or no-flow days is increasing (King et al., 2015; Ruhí et al., 2016; Sabo, 2014). Jaeger et al. (2014) investigated the effect of climate change on flow intermittency patterns and found that annual zero-flow days frequency were projected to increase by 27 % by mid-century in the Lower Colorado River Basin of United States. Research looking into projected changes in regional climate regimes can provide insights into future scenarios people may face, but such research ~~of similar types~~ is still scarce.

The ~~proposed~~ approach ~~described~~developed here to generate spatially continuous estimates of streamflow characteristics (including flow intermittency) throughout stream networks 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 the Geofabric data (Stein et al., 2014) used here, the National Hydrography Dataset Plus (NHDP~~p~~lus) (McKay et al., 2012) and HydroATLAS~~SHEDS~~ (Linke et al., 2019) provide hydrographic datasets and hydro-environmental attributes for ~~the~~ national (USA) and global scales, respectively. In addition, similar to the daily flow model AWRA-L used in this study, other global and national-scale hydrologic models are also available~~ around the world~~, such as the global WaterGAP model (Döll et al., 2003), the community Noah land surface model (Noah-MP) (Niu et al., 2011) in the US and the HYPE model (Lindström et al., 2010) in Sweden.

6 Conclusions

In this study, we presented an approach to quantifying spatially explicit and catchment-wide flow intermittency over long timeframes based on spatially contiguous daily runoff data from a readily accessible water balance simulation. This research builds upon previous studies using monthly runoff data, and paves the way for ecological research looking for metrics of flow intermittency at a daily time step. By testing this approach in eastern Australia, we not only confirmed our previous finding that intermittent flow conditions prevailed in the majority of streams, but also provided more detailed information on their spatio-temporal variability at ~~short time frames~~a daily time step. The proposed approach has ~~the~~ potential applicability to other catchments globally, but our results also highlighted some complexities that future research should address to help improve the reliability of model outputs.

Data availability

The data used in this study are available publicly online and the access websites have been listed in the main text where they were first mentioned.

Competing interests

The authors declare that they have no conflict of interests.

Author contribution

AvD, HXD, MK and SY designed the research, and SY and HXD carried it out. SY wrote the original draft, and HXD, AvD, PL, NB and MK contributed to writing of subsequent drafts.

Acknowledgements

The project was supported by the Australian Climate and Water Summer Institute organised by the Australian Energy and Water Exchange Research Initiative (OzEWEX) and partners. This research was undertaken on the NCI National Facility in Canberra, Australia, which is supported by the Australian Commonwealth Government. We also gratefully acknowledge the support of the Griffith University eResearch Services Team and the use of the High Performance Computing Cluster "Gowonda" to complete this research. SY was financially supported by the China Scholarship Council and Griffith University. HXD is currently funded by School for Environment and Sustainability, University of Michigan through grant number U064474.

References

Acuña, V. et al., 2014. Why should we care about temporary waterways? Science, 343(6175): 1080-1081.

Acuña, V., Hunter, M., Ruhí, A., 2017. Managing temporary streams and rivers as unique rather than second-class ecosystems. Biological Conservation, 211(part B): 12-19.

Australian Bureau of Meteorology, 2014. Key Climate groups of the objective classification of Australian Climates using Koppen's scheme, Bioregional Assessment Source Dataset, http://data.bioregionalassessments.gov.au/dataset/01089f5d-33ef-4fc4-b044-6e1e1f72d698.

Bierkens, M.F.P. et al., 2015. Hyper-resolution global hydrological modelling: what is next? Hydrological Processes, 29(2): 310-320. DOI:10.1002/hyp.10391

Bond, N., 2016. Package 'hydrostats'. The Comprehensive R Archive Network (CRAN).

Bond, N.R., Kennard, M.J., 2017. Prediction of Hydrologic Characteristics for Ungauged Catchments to Support Hydroecological Modeling. Water Resources Research, 53(11): 8781-8794. DOI:10.1002/2017wr021119

Chiu, M.-C., Leigh, C., Mazor, R., Cid, N., Resh, V., 2017. Chapter 5.1 - Anthropogenic Threats to Intermittent Rivers and Ephemeral Streams. In: Datry, T., Bonada, N., Boulton, A. (Eds.), Intermittent Rivers and Ephemeral Streams. Academic Press, pp. 433-454. DOI:https://doi.org/10.1016/B978-0-12-803835-2.00017-6

Costa, A.C., Bronstert, A., de Araújo, J.C., 2012. A channel transmission losses model for different dryland rivers. Hydrology and Earth System Sciences, 16(4): 1111-1135. DOI:10.5194/hess-16-1111-2012

Costelloe, J.F., Grayson, R.B., McMahon, T.A., 2005. Modelling stream flow for use in ecological studies in a large, arid zone river, central Australia. Hydrological processes, 19(6): 1165-1183.

Cunha, L.K., Mandapaka, P.V., Krajewski, W.F., Mantilla, R., Bradley, A.A., 2012. Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model. Water Resources Research, 48(10). DOI:10.1029/2012wr012138

Datry, T., Bonada, N., Boulton, A.J., 2017. Chapter 1 - General Introduction. In: Datry, T., Bonada, N., Boulton, A. (Eds.), Intermittent Rivers and Ephemeral Streams. Academic Press, pp. 1-20. DOI:https://doi.org/10.1016/B978-0-12-803835-2.00001-2

Datry, T., Larned, S.T., Tockner, K., 2014. Intermittent Rivers: A Challenge for Freshwater Ecology. BioScience, 64(3): 229-235. DOI:10.1093/biosci/bit027

Datry, T., Pella, H., Leigh, C., Bonada, N., Hugueny, B., 2016. A landscape approach to advance intermittent river ecology. Freshwater Biology, 61(8): 1200-1213. DOI:10.1111/fwb.12645

David, C.H., Habets, F., Maidment, D.R., Yang, Z.-L., 2011a. RAPID applied to the SIM-France model. Hydrological Processes, 25(22): 3412-3425. DOI:10.1002/hyp.8070

David, C.H. et al., 2011b. River Network Routing on the NHDPlus Dataset. Journal of Hydrometeorology, 12(5): 913-934. DOI:10.1175/2011jhm1345.1

Davison, B., van der Kamp, G., 2008. Low-flows in deterministic modelling: a brief review. Canadian Water Resources Journal, 33: 181-193.

de Vries, S.B. et al., 2015. Characterizing spatial and temporal patterns of intermittent rivers, EGU General Assembly Conference Abstracts.

Do, H.X. et al., 2019. Historical and future changes in global flood magnitude – evidence from a model-observation investigation. Hydrology and Earth System Sciences Discussions: 1-31. DOI:10.5194/hess-2019-388

Döll, P., Kaspar, F., Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. Journal of Hydrology, 270(1): 105-134. DOI:https://doi.org/10.1016/S0022-1694(02)00283-4

Follum, M.L., Tavakoly, A.A., Niemann, J.D., Snow, A.D., 2017. AutoRAPID: A Model for Prompt Streamflow Estimation and Flood Inundation Mapping over Regional to Continental Extents. JAWRA Journal of the American Water Resources Association, 53(2): 280-299. DOI:10.1111/1752-1688.12476

Frost, A.J., Ramchurn, A., Smith, A., 2016. The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Repor, Melbourne.

González-Ferreras, A.M., Barquín, J., 2017. Mapping the temporary and perennial character of whole river networks. Water Resources Research, 53(8): 6709-6724. DOI:10.1002/2017wr020390

Gudmundsson, L., Leonard, M., Do, H.X., Westra, S., Seneviratne, S.I., 2019. Observed Trends in Global Indicators of Mean and Extreme Streamflow. Geophysical Research Letters, 46(2): 756-766. DOI:10.1029/2018gl079725

Gudmundsson, L. et al., 2012. Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe. Journal of Hydrometeorology, 13(2): 604-620. DOI:10.1175/jhm-d-11-083.1

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1-2): 80-91. DOI:10.1016/j.jhydrol.2009.08.003

Hou, J., van Dijk, A.I.J.M., Renzullo, L.J., Vertessy, R.A., Mueller, N., 2019. Hydromorphological attributes for all Australian river reaches derived from Landsat dynamic inundation remote sensing. Earth Syst. Sci. Data, 11(3): 1003-1015. DOI:10.5194/essd-11-1003-2019

Ivkovic, K., Croke, B., Kelly, R., 2014. Overcoming the challenges of using a rainfall–runoff model to estimate the impacts of groundwater extraction on low flows in an ephemeral stream. Hydrology Research, 45(1): 58-72.

Jaeger, K.L., Olden, J.D., 2012. Electrical Resistance Sensor Arrays as a Means to Quantify Longitudinal Connectivity of Rivers. River Research and Applications, 28(10): 1843-1852. DOI:10.1002/rra.1554

Jaeger, K.L., Olden, J.D., Pelland, N.A., 2014. Climate change poised to threaten hydrologic connectivity and endemic fishes in dryland streams. Proceedings of the National Academy of Sciences, 111(38): 13894-13899.

Jaeger, K.L. et al., 2019. Probability of Streamflow Permanence Model (PROSPER): A spatially continuous model of annual streamflow permanence throughout the Pacific Northwest. Journal of Hydrology X, 2: 100005. DOI:10.1016/j.hydroa.2018.100005

Jarihani, A.A., Larsen, J.R., Callow, J.N., McVicar, T.R., Johansen, K., 2015. Where does all the water go? Partitioning water transmission losses in a data-sparse, multi-channel and low-gradient dryland river system using modelling and remote sensing. Journal of Hydrology, 529: 1511-1529. DOI:10.1016/j.jhydrol.2015.08.030

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

Kennard, M.J., Olden, J.D., Arthington, A.H., Pusey, B.J., Poff, N.L., 2007. Multiscale effects of flow regime and habitat and their interaction on fish assemblage structure in eastern Australia. Canadian Journal of Fisheries and Aquatic Sciences, 64(10): 1346-1359. DOI:10.1139/f07-108

Kennard, M.J. et al., 2010b. 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

Kim, J., Mohanty, B.P., 2016. Influence of lateral subsurface flow and connectivity on soil water storage in land surface modeling. Journal of Geophysical Research: Atmospheres, 121(2): 704-721. DOI:10.1002/2015jd024067

King, A.J., Townsend, S.A., Douglas, M.M., Kennard, M.J., 2015. Implications of water extraction on the low-flow hydrology and ecology of tropical savannah rivers: an appraisal for northern Australia. Freshwater Science, 34(2): 741-758.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology, 424-425: 264-277. DOI:10.1016/j.jhydrol.2012.01.011

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrology and Earth System Sciences Discussions: 1-7. DOI:10.5194/hess-2019-327

Konrad, C.P., 2006. Longitudinal hydraulic analysis of river-aquifer exchanges. Water Resources Research, 42(8): W08425. DOI:10.1029/2005wr004197

Lange, J., 2005. Dynamics of transmission losses in a large arid stream channel. Journal of Hydrology, 306(1-4): 112-126. DOI:10.1016/j.jhydrol.2004.09.016

Larned, S.T., Datry, T., Arscott, D.B., Tockner, K., 2010. Emerging concepts in temporary-river ecology. Freshwater Biology, 55(4): 717-738. DOI:10.1111/j.1365-2427.2009.02322.x

Lawrence, D.M. et al., 2011. Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. Journal of Advances in Modeling Earth Systems, 3(1): M03001. DOI:doi:10.1029/2011MS00045

Lee, J., Choi, H., 2017. Improvements to Runoff Predictions from a Land Surface Model with a Lateral Flow Scheme Using Remote Sensing and In Situ Observations. Water, 9(2). DOI:10.3390/w9020148

Leigh, C., Datry, T., 2016. Drying as a primary hydrological determinant of biodiversity in river systems: a broad-scale analysis. Ecography, 40(4): 487-499. DOI:10.1111/ecog.02230

Lin, P. et al., 2019. Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. Water Resources Research. DOI:10.1029/2019wr025287

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrology Research, 41(3-4): 295-319. DOI:10.2166/nh.2010.007

Linke, S. et al., 2019. Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data, 6(1): 283. DOI:10.1038/s41597-019-0300-6

McKay, L. et al., 2012. NHDPlus Version 2: user guide. US Environmental Protection Agency, https://www.epa.gov/waterdata/basic-information.

Niu, G.Y. et al., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. Journal of Geophysical Research: Atmospheres, 116(D12).

Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications, 19(2): 101-121. DOI:10.1002/rra.700

Pilgrim, D., McDermott, G., 1982. Design floods for small rural catchments in eastern New South Wales, First National Local Government Engineering Conference 1981: Reprints of Papers. Institution of Engineers, Australia, pp. 138.

Poff, N.L. et al., 1997. The Natural Flow Regime. BioScience, 47(11): 769-784. DOI:10.2307/1313099

Poncelet, C. et al., 2017. Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. Water Resources Research, 53(8): 7247-7268. DOI:10.1002/2016wr019991

R Development Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raupach, M. et al., 2009. Australian water availability project (AWAP): CSIRO marine and atmospheric research component: final report for phase 3. Centre for Australian weather and climate research (bureau of meteorology and CSIRO). Melbourne, Australia, 67.

Raupach, M. et al., 2018. Australian Water Availability Project, Data Release 26m. CSIRO Oceans and Atmospheric, <http://www.csiro.au/awap>.

Robinson, J.S., Sivapalan, M., Snell, J.D., 1995. On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments. Water Resources Research, 31(12): 3089-3101. DOI:10.1029/95wr01948

Ruhí, A., Olden, J.D., Sabo, J.L., 2016. Declining streamflow induces collapse and replacement of native fish in the American Southwest. Frontiers in Ecology and the Environment, 14(9): 465-472.

Sabo, J.L., 2014. Predicting the river's blue line for fish conservation. Proceedings of the National Academy of Sciences, 111(38): 13686-13687.

Sheng, M., Lei, H., Jiao, Y., Yang, D., 2017. Evaluation of the Runoff and River Routing Schemes in the Community Land Model of the Yellow River Basin. Journal of Advances in Modeling Earth Systems, 9(8): 2993-3018. DOI:10.1002/2017ms001026

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

Snelder, T.H. et al., 2013. Regionalization of patterns of flow intermittence from gauging station records. Hydrology and Earth System Sciences, 17(7): 2685-2699.

Spence, C., Mengistu, S., 2016. Deployment of an unmanned aerial system to assist in mapping an intermittent stream. Hydrological Processes, 30(3): 493-500.

Stanley, E.H., Fisher, S.G., Grimm, N.B., 1997. Ecosystem expansion and contraction in streams. BioScience, 47(7): 427-435.

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.

Stein, J.L., Hutchinson, M.F., Stein, J.A., 2014. A new stream and nested catchment framework for Australia. Hydrology and Earth System Sciences, 18(5): 1917-1933. DOI:10.5194/hess-18-1917-2014

Stromberg, J.C., Bagstad, K.J., Leenhouts, J.M., Lite, S.J., Makings, E., 2005. Effects of stream flow intermittency on riparian vegetation of a semiarid region river (San Pedro River, Arizona). River Research and Applications, 21(8): 925-938. DOI:10.1002/rra.858

Turner, D.S., Richter, H.E., 2011. Wet/dry mapping: using citizen scientists to monitor the extent of perennial surface flow in dryland regions. Environmental management, 47(3): 497-505.

van Dijk, A.I., 2010. The Australian Water Resources Assessment System. Technical Report 3. Landscape Model (version 0.5) Technical Description. CSIRO: Water for a Healthy Country National Research Flagship, CSIRO, Australia.

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

Wang, Q.J. et al., 2011. Monthly versus daily water balance models in simulating monthly runoff. Journal of Hydrology, 404(3-4): 166-175. DOI:10.1016/j.jhydrol.2011.04.027

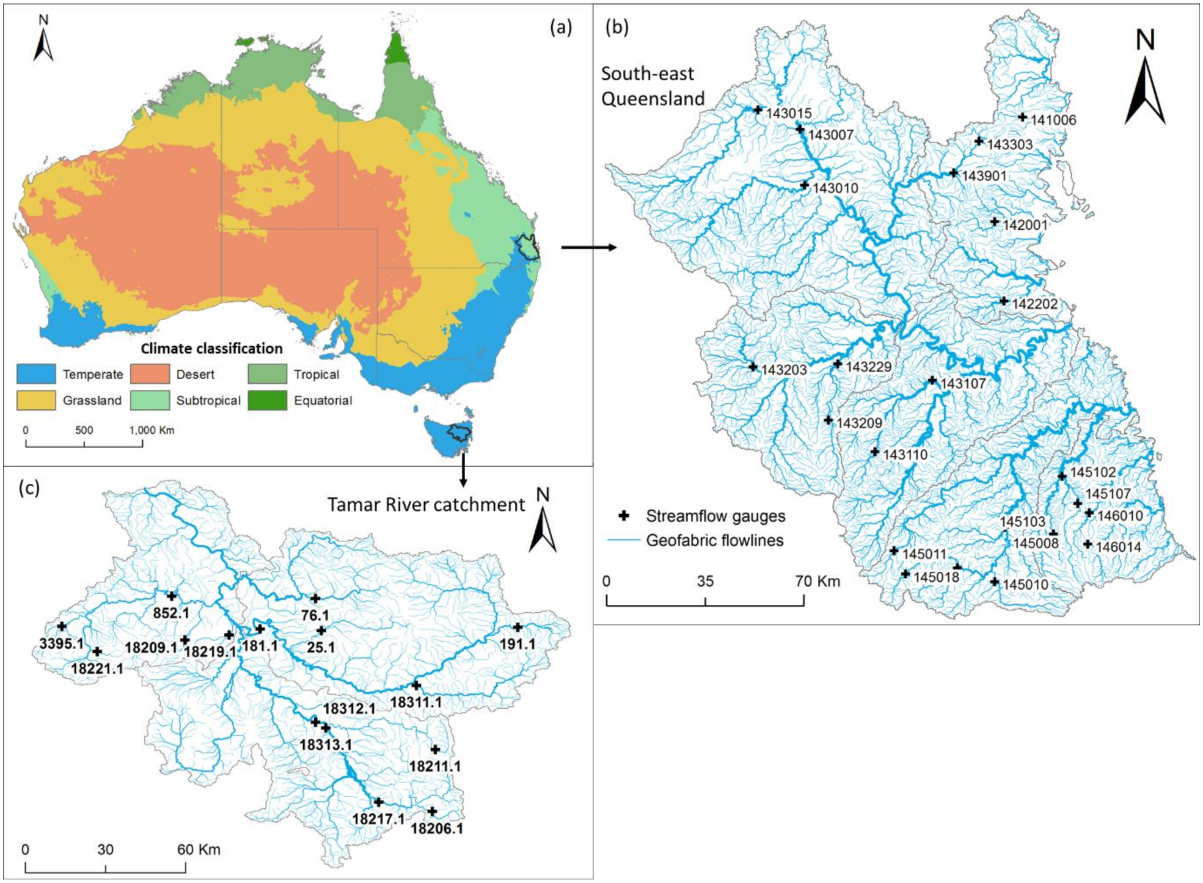
Ye, W., Bates, B.C., Viney, N.R., Sivapalan, M., Jakeman, A.J., 1997. Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. Water Resources Research, 33(1): 153-166. DOI:10.1029/96wr02840

Yu, S., Bond, N.R., Bunn, S.E., Xu, Z., Kennard, M.J., 2018. Quantifying spatial and temporal patterns of flow intermittency using spatially contiguous runoff data. Journal of Hydrology, 559: 861-872. DOI:10.1016/j.jhydrol.2018.03.009

Zaherpour, J. et al., 2018. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environmental Research Letters, 13(6). DOI:10.1088/1748-9326/aac547

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.

Zimmer, M.A., McGlynn, B.L., 2017. Ephemeral and intermittent runoff generation processes in a low relief, highly weathered catchment. Water Resources Research, 53(8): 7055-7077. DOI:10.1002/2016wr019742



605 **Figure 1. Locations of the two climatically and hydrologically distinctive regions in Australia (a): South-east Queensland (SEQ) (b) and the Tamar River catchment (Tamar) (c) with Geofabric river networks and selected stream gauges (25 and 15 gauges for SEQ and Tamar, respectively). The climate classification in (a) is based on the Köppen classification system (Australian Bureau of Meteorology, 2014).**

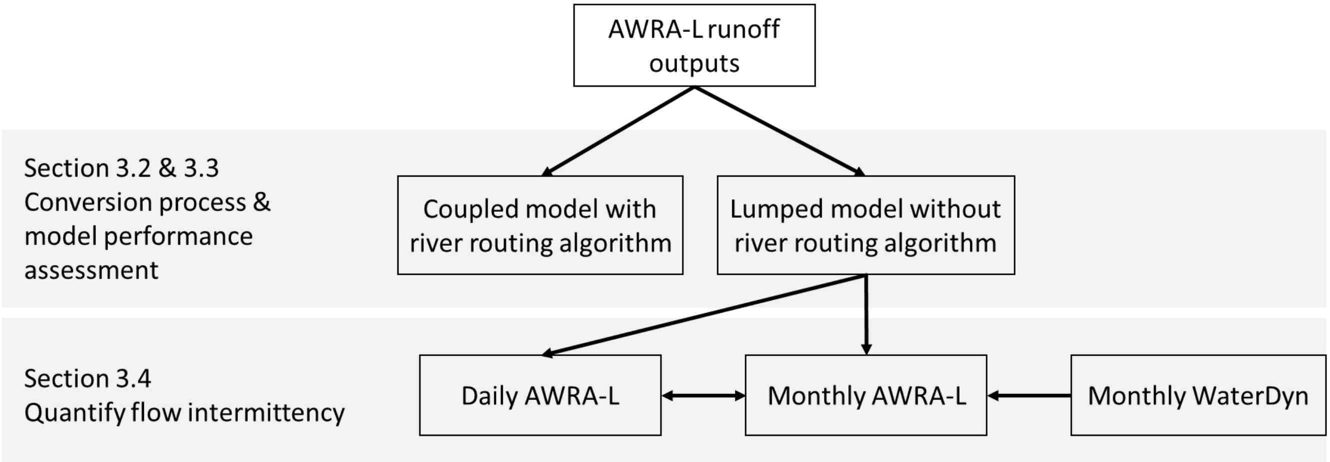


Figure 2. Model configurations and their applications in this study. AWRA-L runoff outputs are translated to accumulated streamflow estimates with river routing algorithm (coupled model) and without (lumped model). These two model configurations are applied to test the effect of river routing on streamflow simulation accuracy. Based on the lumped model, we simulate daily streamflow throughout river networks (Daily AWRA-L) and further convert the daily stimulations to monthly outputs (Monthly AWRA-L). Both simulations are used to quantify streamflow intermittency, while results from a different monthly model (Monthly WaterDyn) are used to benchmark the flow intermittency estimates from Monthly AWRA-L.

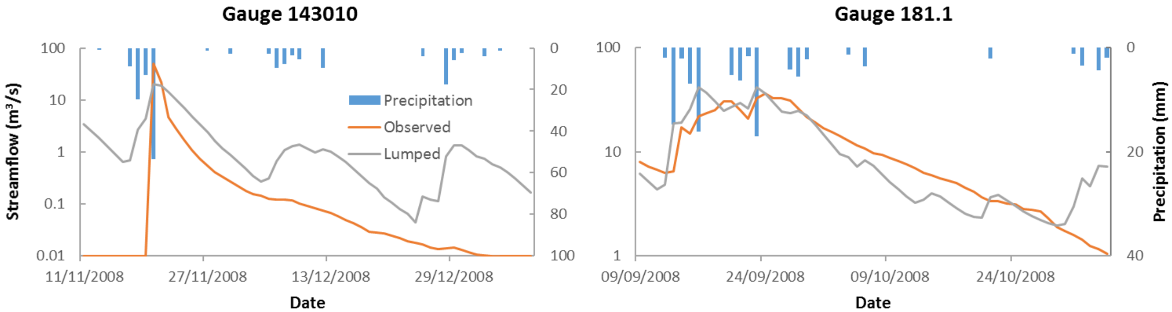


Figure 3. Comparison of the observed and modelled hydrograph with the rainfall time series at gauges 143010 in SEQ and 181.1 in Tamar. The over-responsiveness of the model to rainfall is illustrated in the ~~dramatic~~ noticeable increase in modelled streamflow when a rainfall event occurred, while there is no obvious increase in observed streamflow. Rainfall data were sourced from the AWRA-L input.

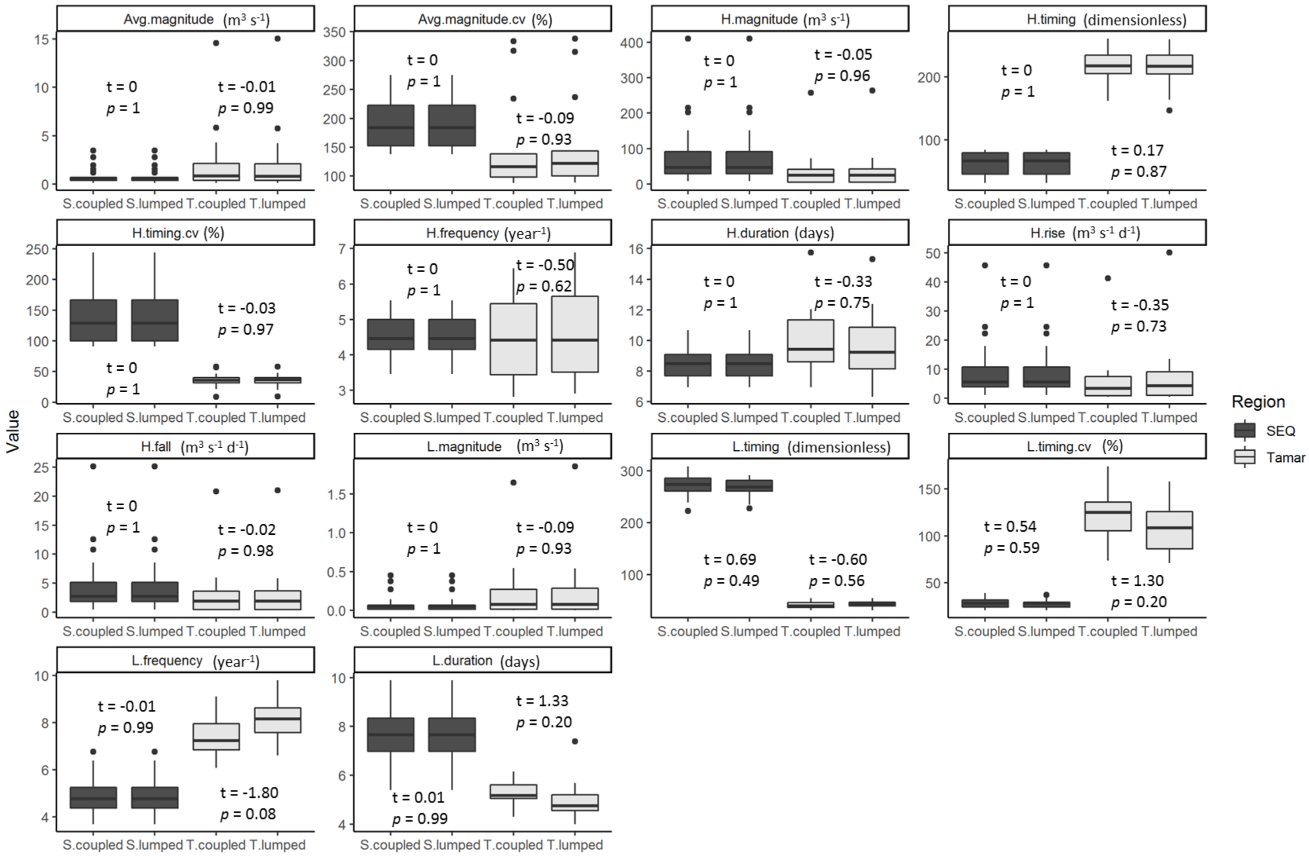


Figure 4. Comparison of hydrological characteristics between the lumped and coupled models in SEQ and Tamar. Refer to Table 1 for measurement description ~~and units of measurement~~ for each flow metric. Metrics are grouped according to average (Avg), high (H) and low (L) flow conditions. The values of *t* statistic and associated *p* values are also shown to indicate whether there is any significant difference between the coupled and lumped simulations.

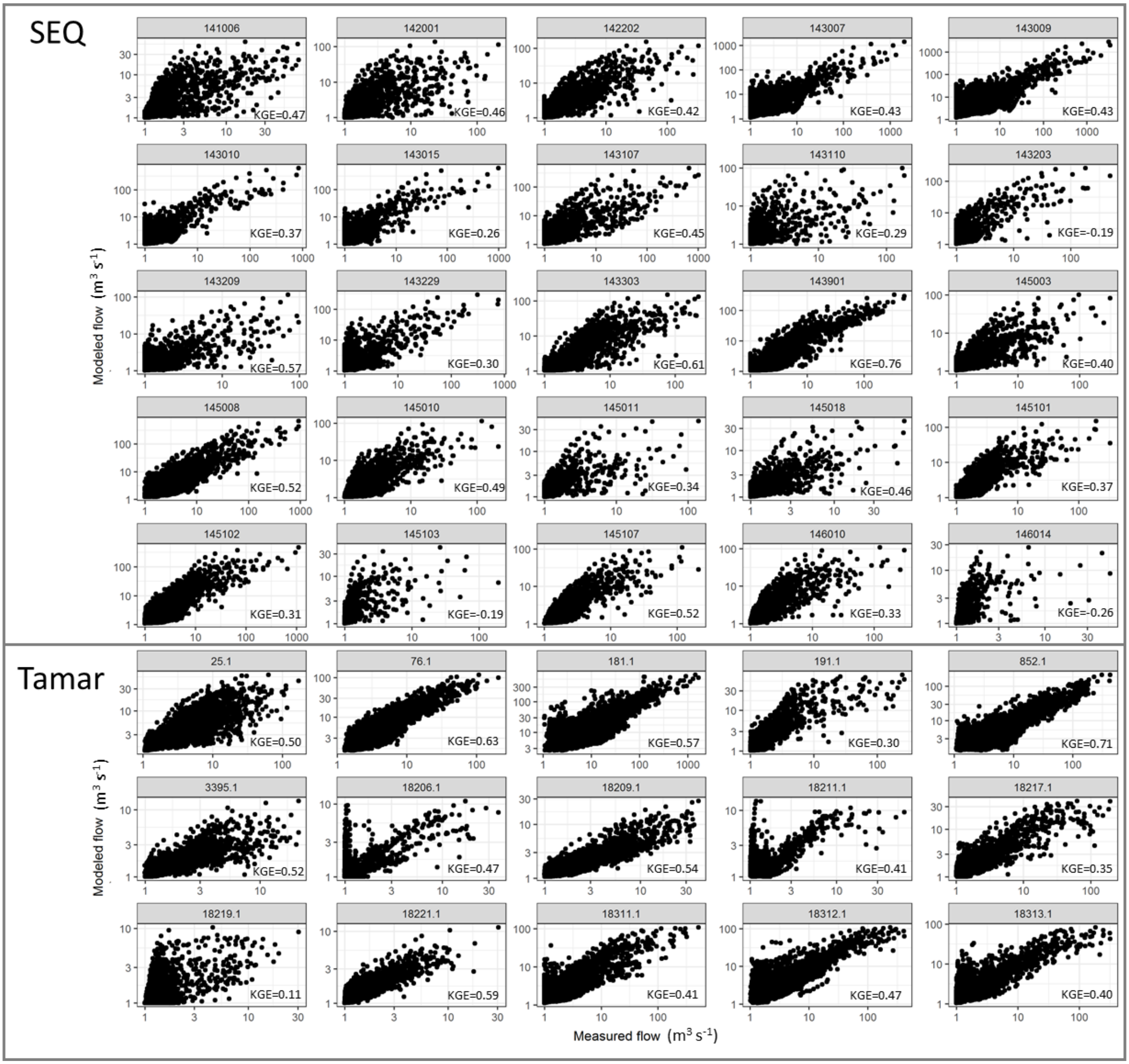


Figure 5: Scatter plots of the measured and modelled (lumped) streamflow for each gauge station in SEQ and Tamar. The modified Kling-Gupta efficiency (KGE) is presented in each panel. The x and y axes are log-transformed ($\log_{10}(x+1)$) to aid interpretation.

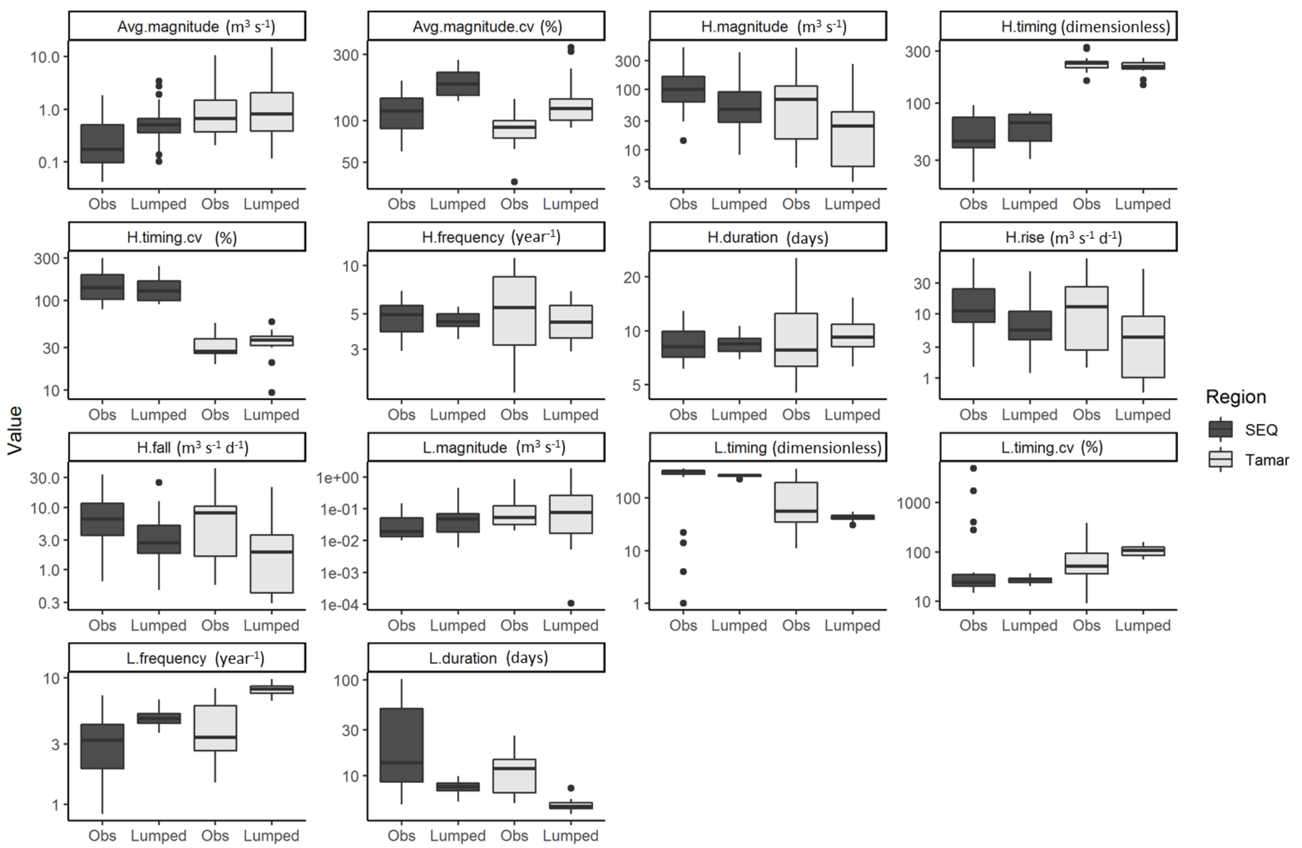
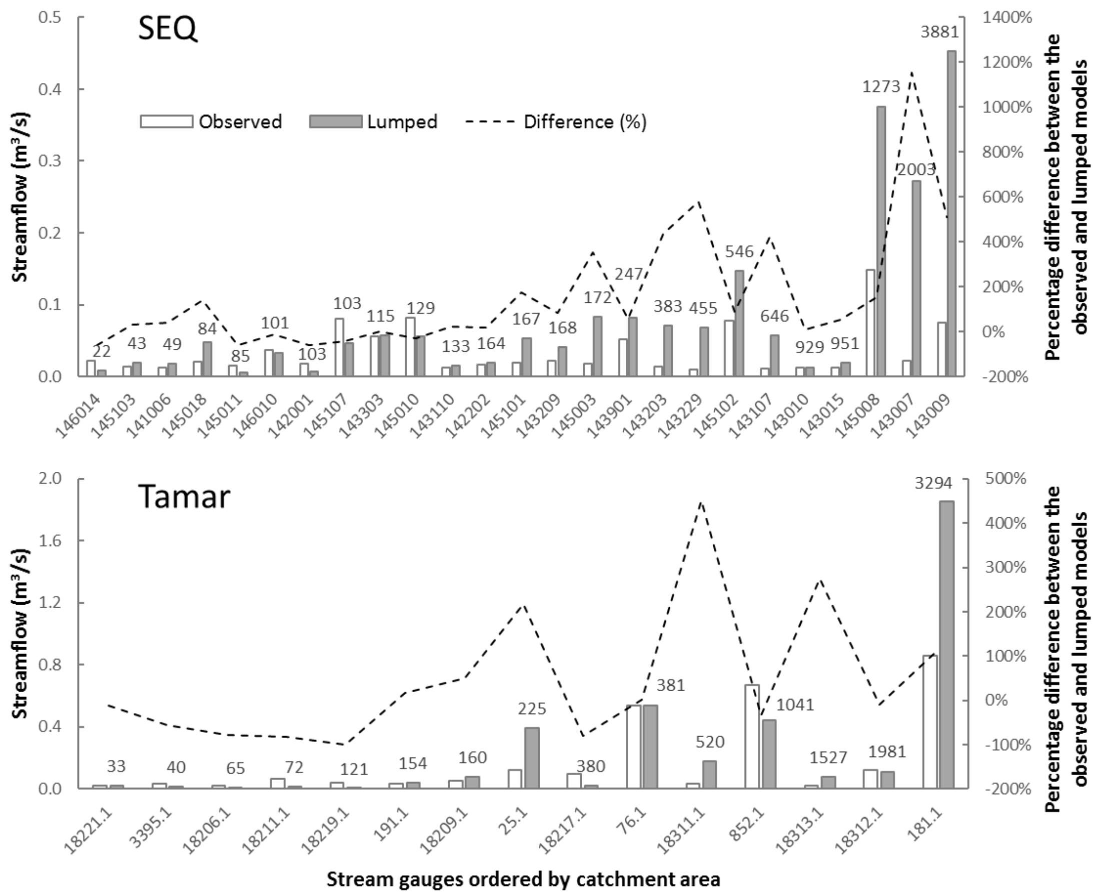


Figure 6. Variation in observed and modelled (lumped) hydrologic characteristics in SEQ and Tamar (n= 25 and 15 gauge locations, respectively). Refer to Table 1~~Table 1~~ for measurement description ~~and units of measurement~~ for each flow metric. Metrics are grouped according to average (Avg), high (H) and low (L) flow conditions. The y-axis is on a log scale for better interpretation.



~~**Figure 6. Comparison of the observed and modelled low flow magnitudes (left y-axis) at all gauges in each region, with their percentage difference shown as dashed line (right y-axis). The stream gauges are ordered by catchment area (km2), which is labelled above each column**~~**.**

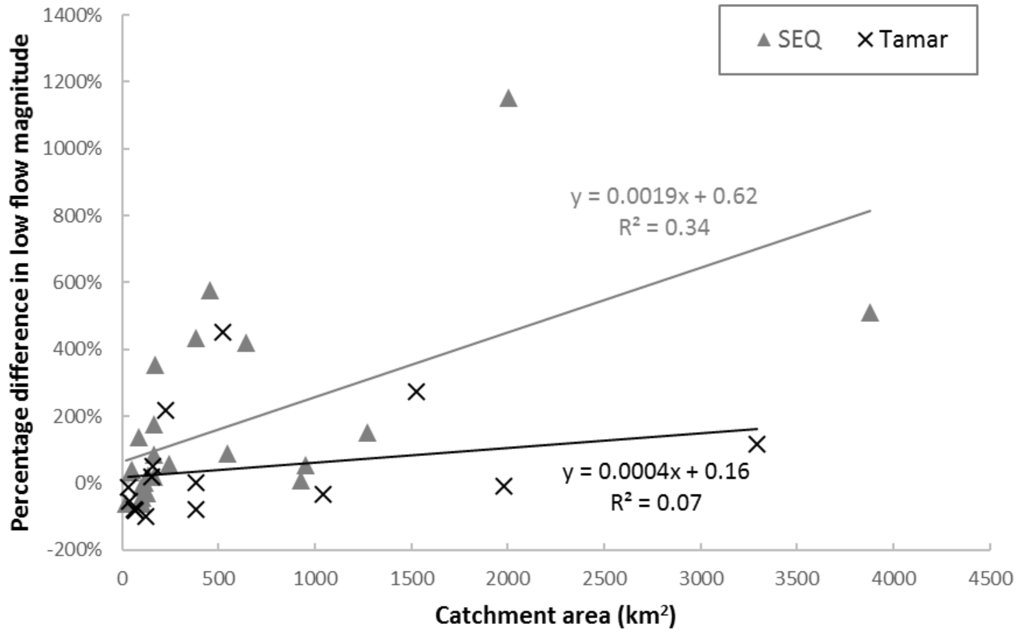
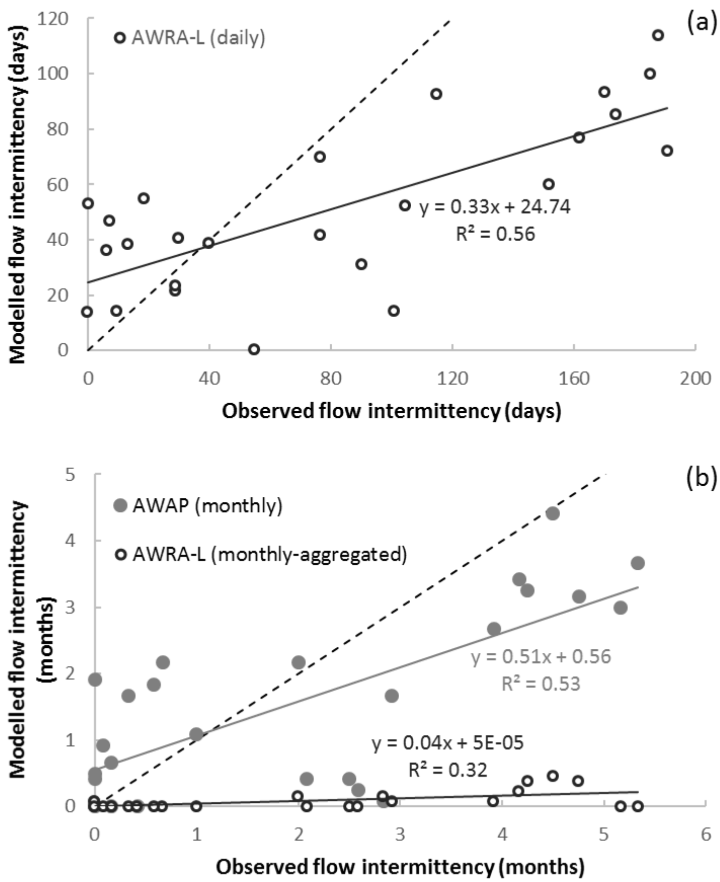


Figure 7. Scatter plot of gauged catchment areas and percentage difference in low flow magnitude between the observation and simulation in SEQ (solid grey triangle) and Tamar (black cross). The regression line for each region is also shown as solid line (grey line for SEQ and black line for Tamar) with the regression function and R^2 value.~~Comparison of the observed and modelled low flow magnitudes (left y-axis) at all gauges in each region, with their percentage difference shown as dashed line (right y-axis). The stream gauges are ordered by catchment area (km²), which is labelled above each column.~~



~~Figure 7. Scatter plots of the observed and modelled flow intermittency by the two models (AWRA-L and AWAP model) for SEQ. The solid line represents the regression line for each model. The 1:1 line (dashed line) is plotted for reference.~~

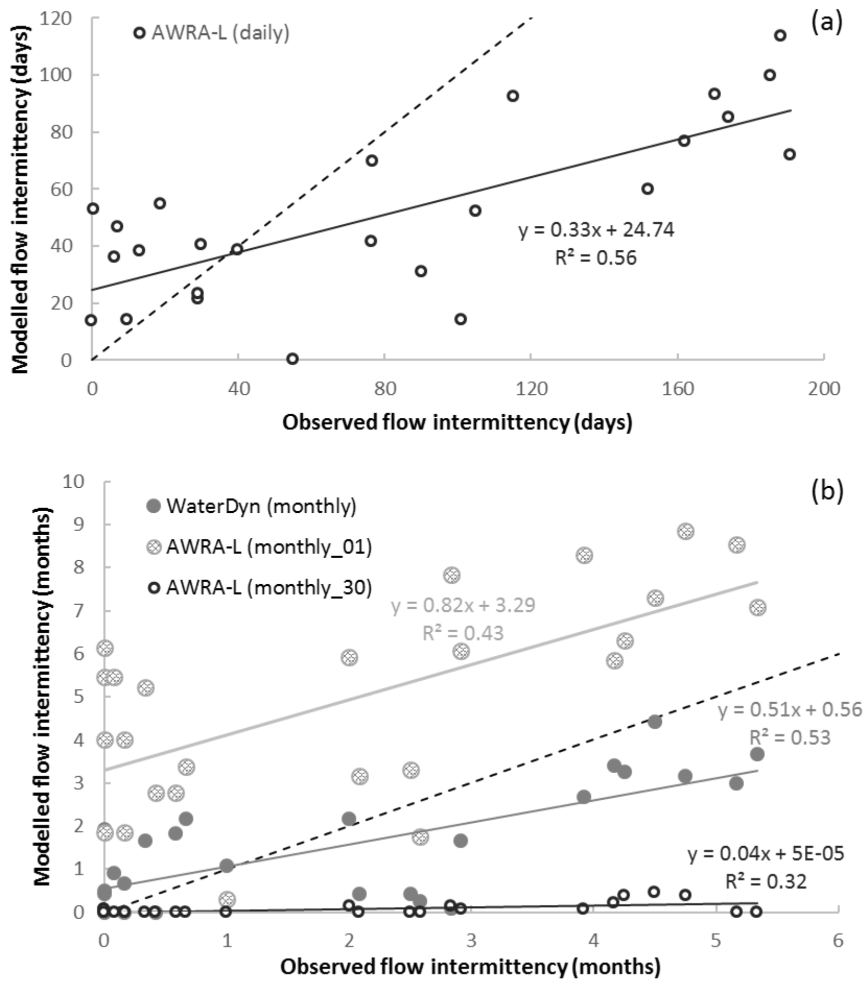


Figure 8. Scatter plots of the observed and modelled flow intermittency by the two models (AWRA-L and WaterDyn model) for SEQ. daily AWRA-L and monthly WaterDyn are derived from the original data from the two models, while AWRA-L monthly_01 and monthly_30 are flow intermittency estimates using the two different aggregation methods with different thresholds (one day vs. 30 days) to classify a month as zero-flowing. The solid line represents the regression line for each model. The 1:1 line (dashed line) is plotted for reference.

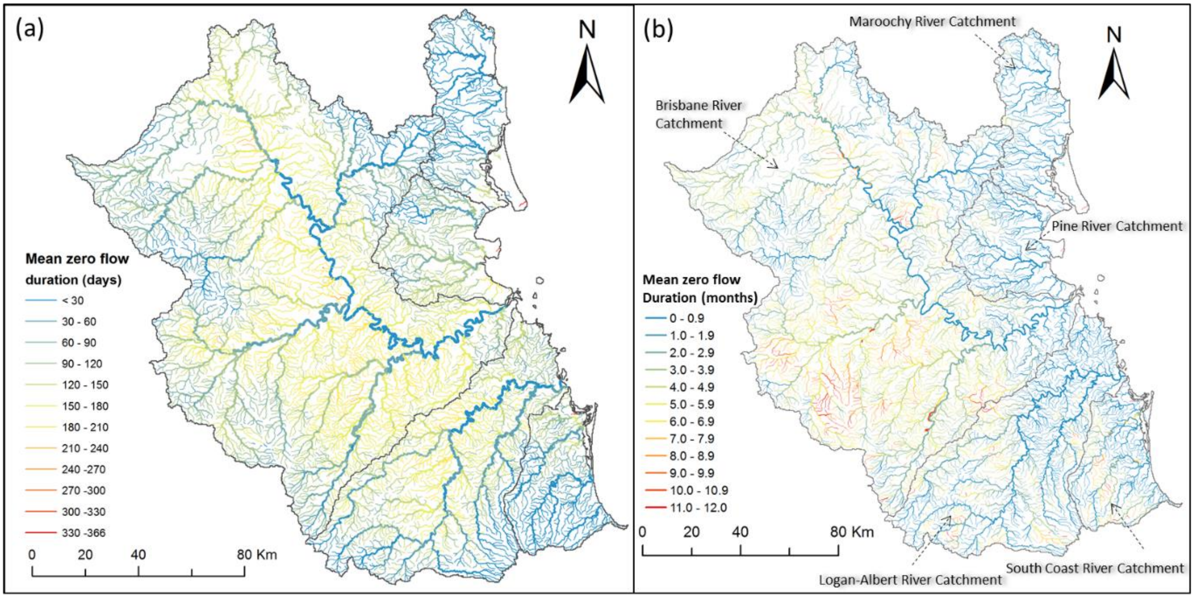


Figure 9. Comparison of the spatial pattern of average annual flow intermittency in SEQ derived from (a) daily flow simulations from the AWRA-L model and (b) monthly flow simulations from the WaterDyn~~AWAP~~ model. Stream segments in both figures are coloured using the same frame but different units. Line thicknesses show the stream orders.

665

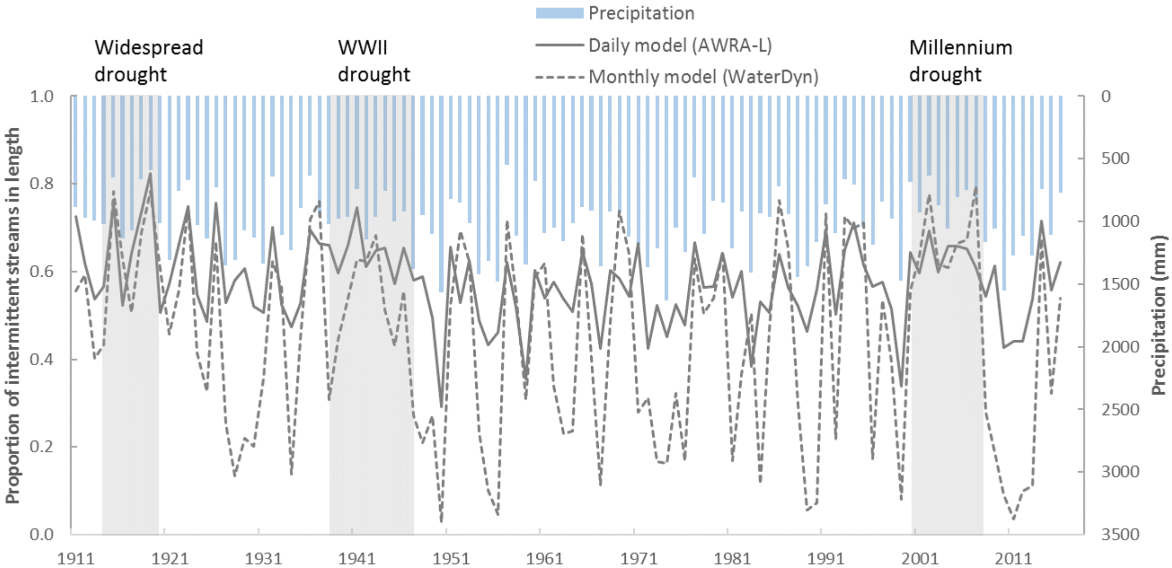


Figure 10. Comparison of intra-annual variation of the proportion of intermittent streams in length from 1911-2016 across SEQ, derived from stream flow simulations from the daily flow model (lumped, solid grey ~~solid~~ line) and monthly flow model (grey dash line). Three severe droughts in Australia were also presented as transparent grey rectangles: Widespread drought (1914-1920), WWII droughts (1939-1946) and Millennium droughts (2001-2009). The time series of annual mean precipitation ~~were~~ is shown ~~presented~~ for reference ~~and they were sourced from the AWAP model (Raupach et al., 2009; Raupach et al., 2018)~~.

Table 1. Flow metrics used to describe average-, high- and low-flow conditions across key components of hydrological variation. Note that a spell independence criteria of 5 days was applied to regard periods between spells of less than 5 days as "“in spell"”.

Conditions	Component	Abbreviation	Definition	Units
Average flow	Magnitude	Avg.magnitude	Mean daily flow for entire period	$m^3\ s^{-1}$
	Variability	Avg.magnitude.cv	Coefficient of variation in mean daily flow	%
High-flow	Magnitude	H.magnitude	The average annual maximum flow	$m^3\ s^{-1}$
	Timing	H.timing	The mean Julian date of annual maximum	unitless
	Variability	H.timing.cv	Coefficient of variation in Julian date of annual maximum flow	%
	Frequency	H.frequency	Mean of annual count of spells above the 90th percentile flow	unitless
	Duration	H.duration	Mean duration of all spells above the 90th percentile flow	days
	Rate of rise	H.rise	Mean rate of positive changes in flow from one day to the next	$m^3\ s^{-2}$
	Rate of fall	H.fall	Mean rate of negative changes in flow from one day to the next	$m^3\ s^{-2}$
Low-flow	Magnitude	L.magnitude	The average annual minimum flow	$m^3\ s^{-1}$
	Timing	L.timing	The mean Julian date of annual minimum	unitless
	Variability	L.timing.cv	Coefficient of variation in Julian date of annual minimum flow	%
	Frequency	L.frequency	Mean of annual count of spells below the 10th percentile flow	unitless
	Duration	L.duration	Mean duration of all spells below the 10th percentile flow	days