Response to Editor Christa Kelleher's comments

We thank the Editor Christa Kelleher for summarising all Reviewers comments and providing constructive comments that help improve the quality of this manuscript.

Editor comment:

The article has received two additional reviews, both of which indicate that the manuscript is strong and has improved in the recent round of iteration.

Authors reply:

Thanks for the positive comments.

Editor comment:

However, both reviewers still had concerns. Reviewer 1 notes that their concerns were not adequately addressed, and encourages some additional text be added to the manuscript. Reviewer 2 would like additional information, particularly with regards to the threshold for intermittence (though other concerns are noted in the review report and should also be addressed) and recommends an additional figure that shows spatial variability in intermittence.

In particular, I agree with Reviewer 2 that additional details on the regression approach (including approach and outcomes) should be added, even though these methods have already been published in another article. All readers need enough information to ensure they fully understand the methods used in this particular study. As this is a key step for determining intermittency, more information is needed.

Authors reply:

We have responded in detail to each of the Reviewers comments and provide a brief summary here.

To address Reviewer 1's further comments, we have now added further details on the representativeness of stream gauges (Section 3.1) and acknowledged that the distribution of selected stream gauges and other factors may be a source of uncertainty for the application of the regression model (Section 3.4).

We have also added a new sentence in the revised Discussion (lines 405-408) to acknowledge that Geofabric may not be able to resolve the smallest streams in the study area, and thus our analyses may have under-estimated the proportion of intermittent streams.

Regarding Reviewer 2's comments, we have now added more details in Section 3.4 describing the linear regression models, including the model fits, methods for model evaluation, assessment of spatial autocorrelation and sources of uncertainty.

We have also added to Figure 9 two new inset plots showing frequency distributions of total stream length by flow intermittency classes identified from AWRA-L daily and WaterDyn monthly streamflow simulations, respectively. These revisions to Figure 9 now amply demonstrate the variations in flow intermittency across the study region.

Editor comment:

Both reviewers also note that the implications of this work, and limitations of this approach, should be further discussed.

Authors reply:

As indicated in our reply to the first comment, we have now acknowledged potential limitations and sources of uncertainty in our study.

We have also restructured a paragraph in the discussion Section 5.3 to make one of our main findings more apparent (new text highlighted in bold, with the original text provided for context).

"**Our results suggest that the temporal resolution of analysis should be dictated by the resolution of input streamflow data. More specifically**, the daily AWRA-L flow showed promise for estimating flow intermittency at a daily time step, while the monthly WaterDyn model was better than the monthly 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, our 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 one day or 30 days as a threshold and, respectively, either substantially overestimated or underestimated flow intermittency."

Editor comment:

I encourage the authors to incorporate these recommendations for adding some clarity to the manuscript.

Authors reply:

We have incorporated all these recommendations in the revised manuscript. Please also see our detailed responses to each of the Reviewer comments.

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 further comments that help improve the quality of this manuscript.

Reviewer comment:

While the authors have addressed almost all of my concerns with the initial version of the manuscript, I view that two of the major issues that I brought up have still not been adequately addressed in the text:

The first concern I raised previously was that I thought the placement of the stream gauges might be biased and not representative of the true hydrological conditions of the entire stream networks. In response, the authors state that the stream gauges are distributed across a range of streams types and therefore are a representative sample. However, they do not provide any evidence that the gauges are actually representative. Instead of asserting that the stream gauges are representative, I would suggest just acknowledging in the text that they are assuming that the gauges are representative and are a source of uncertainty in the analysis.

Authors reply:

We have now added further details on the representativeness of stream gauges and acknowledged that the distribution of selected stream gauges and other factors may be a source of uncertainty for the application of the regression model. The new text (highlighted in bold) reads:

"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). **However, the set of stream gauges used in our analyses under-represented the frequency of small low-order streams in both regions.** Therefore, we regard the selected gauges to be representative of the range of 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." (lines 118-123)

"The unexplained variation in the predictive model **may be due to the limited number of environmental attribute covariates used in the model and hence ability to adequately represent the range of environmental processes that influence streamflow intermittency. Additional uncertainty in model predictions may arise because the distribution of stream gauges used for model calibration under-represented the frequency of extremely small catchments that likely had higher cease-to-flow occurrence**." (lines 227-230)

Reviewer comment:

The second concern is that they largely sidestep my concern that Geofabric doesn't contain data for the smallest streams in the study basin. Yes, I understand that Geofabric is probably the best dataset that is available to conduct this analysis but it still is not able to resolve the finest streams (e.g., see Shanafield et al., 2020, JGR-ES, https://doi.org/10.1029/2019JF005330 or Allen et al., 2018, Nature Communications, https://doi.org/10.1038/s41467-018-02991-w). I suggest that the authors

acknowledge in the text that Geofrabric is not able to resolve the smallest streams in the study area, and therefore the study's analysis likely is biased toward perennial streams when predicting the proportion of streams that are predicted to be intermittent.

In my view, adding these caveats and uncertainties to the text will only strengthen the manuscript and the study will remain interesting and useful to the scientific community.

Authors reply:

We apologise that the reviewer felt we "largely sidestepped" their concern about the spatial resolution of the Geofabric – we attempted to directly respond to their concern and further elaborate here.

We believe that the spatial resolution of the smallest streams identified in the Geofabric stream network (version 2.1.1) is appropriate to achieve the aims of our study (i.e. 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), considering the grain size of the streamline mapping data relative to the large spatial extent of our study areas. We agree that stream mapping at a finer grain size would no doubt reveal the existence of smaller streams, however, the key question is whether spatial patterns of flow intermittency would differ substantially depending on the spatial resolution of the data used to define these patterns. While such an assessment is beyond the scope of this study, the key issues relevant to our paper are the requirement that fine-grained streamline mapping data is actually available, is spatially consistent and is spatially comparable to evaluate spatio-temporal dynamics of streamflow intermittency (the primary objective of this paper). The Geofabric stream network dataset meets all of these requirements.

Nonetheless, we have now added a new sentence in the revised Discussion to acknowledge that Geofabric may not be able to resolve the smallest streams in the study area, and thus our analyses may have under-estimated the proportion of intermittent streams (Lines 405-408):

"However, given the limited spatial resolution of the Geofabric stream network data (9 arc-second longitude-latitude resolution, with the minimum upstream drainage area of 1.5 km^2) and hence ability to resolve the smallest streams, and that small streams are more likely to be intermittent, the proportion of predicted intermittent streams in SEQ may be under-estimated in our study."

Responses to Anonymous Referee #4'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 Anonymous Referee #4 for providing these constructive comments that help improve the quality of this manuscript.

Reviewer comment:

Below, I provide a short overview of the paper, my recommendation, and then more specific comments.

Broad overview. This paper addresses a critical problem for water resource management, regulatory, and research communities – how do we quantify the spatial variability in intermittence across broad spatial scales? Authors address this by applying a water balance model across two watersheds in Australia: Southeast Queensland (temperate to subtropical climates) and Tamar River of Tasmania (temperate climate). The authors modeling approach can be divided into three steps: (i) developing the AWRA-L runoff simulations, (ii) comparing AWRA-L runoff simulations with and without routing, and (iii) quantifying flow intermittency using daily and monthly outputs from the AWRA-L runoff model and comparing those outputs to the WaterDyn model. The authors highlight their model fit using flow regime metrics across multiple gages in each region, highlight that [atleast in this exercise] a routing model is was not necessary to simulate streamflow, and finally demonstrate spatial and temporal trends in intermittency across the two regions.

Recommendation. My recommendation is to accept with minor revisions. I am reviewing this paper for the first time, but it is in its second round of revision. This paper provides a unique solution to estimating large scale spatial and temporal patterns in intermittency. Further, authors should be commended for quantifying the fit [or in some cases lack thereof] between the simulated and observed streamflow regime. My one major concern is the method used to define the threshold for intermittence. I hope the authors will consider providing more information to their readers. Below, I provide several comments in an attempt to help authors improve an already good manuscript.

Authors reply:

Thanks for the positive comments. To address the major concern, we have now added more information for the method used to define the threshold for streamflow intermittency. For details, see our responses to the following two comments.

Reviewer comment:

Specific comments.

-At L205, authors outline their process of defining 'thresholds' for intermittence in the simulated flow data. I encourage the authors to provide information about the distributions of the flow quantiles of these thresholds (either in text, or better yet, a figure!).

Authors reply:

We have now provided the ranges and the median value of these thresholds for streamflow intermittency in the revised manuscript (lines 225-226): "The adopted thresholds ranged from 0 – 1.668 m³/s, with a median value of 0.002 m³/s."

Reviewer comment:

-Related to the last comment, authors need to provide more information about the linear models. What were the model fits like, were a split sample design employed, and did the authors consider spatial autocorrelation? The reader currently cannot adequately evaluate this portion of the analysis, which is a critical step when defining intermittency.

Authors reply:

We have now added more details in Section 3.4 describing the linear regression models (lines 205-231), including the model fits, methods for model evaluation, assessment of spatial autocorrelation and sources of uncertainty. The section now reads:

"We used linear regression to model the mean annual zero flow duration (daily time step) at each gauge as a function of catchment environment variables. This regression analysis was only conducted in SEQ as most gauges in the Tamar catchment had perennial flow. 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). Regression models were developed using all possible predictor variable combinations and we selected the "best" model for predicting zero flow duration based on corrected Akaike's Information Criterion (AICc) (Hurvich and Tsai, 1989). To estimate the prediction error of the selected model, we applied leave-one-out cross validation on the selected 43 gauges and reported prediction error (R^2) to estimate the model prediction performance. Regression model development and cross-validation were conducted with the MuMIn and boot packages in R (R Development, 2012). Regression analyses were performed on all combinations of predictor variables and the best model with the lowest AICc (-54.2) retained five covariates, including annual daily maximum temperature, catchment area, slope, average elevation, and average saturated hydraulic conductivity. The developed predictive model showed a good model fit with an adjusted R^2 of 0.71, and the leave-one-out cross validation on the regression model showed relatively good model performance with an average R^2 of 0.64. We checked for spatial autocorrelation of the regression model residuals (as recommended by Dormann et al., 2007) and found they were not significantly autocorrelated (Moran's I = -0.06, p = 0.69). Examination of spatial residual maps further supported this conclusion, with no spatial trends in model residuals apparent.

Next, we used the predictive models to extrapolate estimates of overall flow intermittency (in terms of the proportion of days with zero flow) to each segment throughout the river network. Finally, for each segment, the time-series of daily runoff was truncated (flows below the threshold were set to "0") by adopting an appropriate threshold of "zero flow" that preserved the proportion of days with flow as estimated in the previous step. The adopted thresholds ranged from 0 – 1.668 m³/s, with a median value of 0.002 m³/s. We recognise several sources of uncertainty in our approach to estimating the zero flow thresholds. The unexplained variation in the predictive model may be due to the limited number of environmental attribute covariates used in the model and hence ability to adequately represent the range of environmental processes that influence streamflow intermittency. Additional uncertainty in model predictions may arise because the distribution of

stream gauges used for model calibration under-represented the frequency of extremely small catchments that likely had higher cease-to-flow occurrence."

Reviewer comment:

-It's unclear why authors use parametric comparative statistics (i.e., t-test). I would suggest switching to a nonparametric analogs (i.e., Wilcoxon test) or showing normality to prove t-test is appropriate.

Authors reply:

Although the majority of variables show normality, measured by the Shapiro-Wilk Normality Test, we agree that a nonparametric Wilcoxon test is more appropriate. We have now replaced Student's *t*-test with Wilcoxon rank sum test in the revised manuscript (see new Figure 4), and the new results only change for one variable (i.e. the duration of low flow spells), whose *p* value changes from 0.08 to 0.03. These analyses do not change the main finding that the routing algorithm had negligible effects on flow simulations in our study areas.

Reviewer comment:

- After my first read-through, I was unclear why the WaterDyn model was incorporated into the manuscript. While this is reported in L220, I would strongly recommend authors make this more apparent [i.e., not hidden at the bottom of a paragraph.]

Authors reply:

We have improved clarity of this paragraph by bringing upfront the reason to incorporate the WaterDyn model to make it more apparent. The new added sentence reads: "To evaluate the effect of time step (daily vs. monthly) on the relative performance of the AWRA-L model in replicating observed patterns of cease-to-flow periods, we compared model outputs with those derived from a monthly water balance model – the WaterDyn model (Fig. 2). Monthly flow intermittency estimated from WaterDyn was thus used to benchmark results from the monthly AWRA-L." (lines 233-236)

Reviewer comment:

-Related to the comment above, there is a lot of important information packed in the paragraph starting at 370. [Infact, I would go so far as to say this is one of the main findings of the paper!] My suggestion is to rewrite this paragraph; lead with the argument that "our results suggest that the temporal resolution of analysis should be dictated by the resolution of input streamflow data;" and potentially further develop this discussion point. This is a useful result – one that will help others build on this studies results! Authors might even consider adding this result to the abstract.

Authors reply:

We agree that this finding should be more strongly emphasised. We have now brought the main point from the bottom in the paragraph to the beginning. The paragraph now starts with "Our results suggest that the temporal resolution of analysis should be dictated by the resolution of input

streamflow data". We also made minor text changes in the paragraph, while the majority remained unchanged, as they well fit with the new structure.

In addition, this key finding is highlighted in the abstract (lines 28-29): "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".

Reviewer comment:

-Finally, I was disappointed the authors did not do more to explore the spatial variation of intermittence in their data. I encourage them to add an additional figure to highlight spatial variability. Note, I'm not suggesting they add an additional map. More appropriately, I would encourage authors to add a figure that quantifies spatial variability in intermittence.

Authors reply:

We agree that adding an additional figure to demonstrate spatial variability in flow intermittency is necessary and important. We have now added to Figure 9 two new inset plots showing frequency distributions of total stream length by flow intermittency classes identified from AWRA-L daily and WaterDyn monthly streamflow simulations, respectively. These revisions to Figure 9 now amply demonstrate the variations in flow intermittency across the study region.

We also revisited the text to better describe these newly added inset plots: "Compared to the monthly WaterDyn model, fewer streams were predicted by the daily AWRA-L to experience extremely long dry events as well as less than one month of zero flows (Fig. 9c, d)". (lines 314-315)

The new Figure 9 is shown as below:
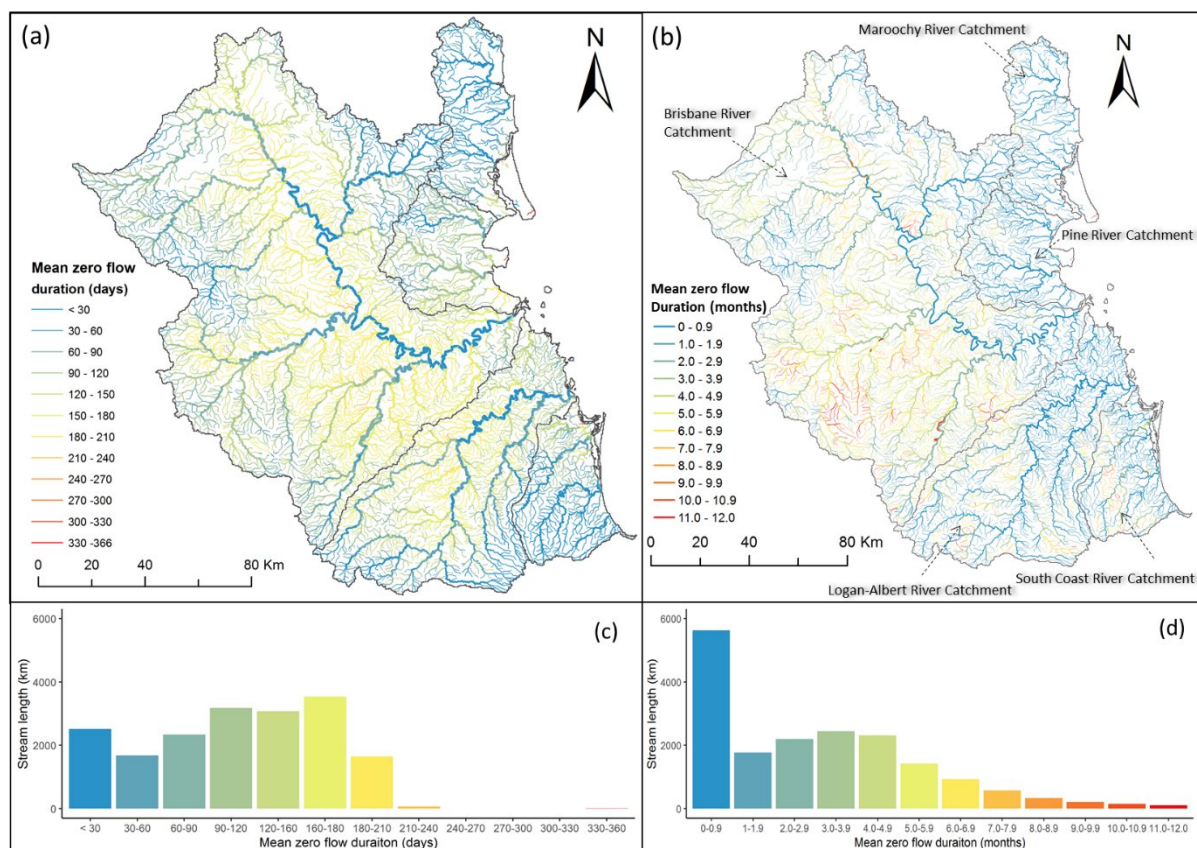
Figure 1. Comparison of the spatial pattern of average annual flow intermittency in SEQ derived from (a) daily flow simulations from the AWRA-L model and (b) monthly flow simulations from the WaterDyn model. Stream segments in both figures are coloured using the same frame but different units. Line thicknesses show the stream orders. Frequency distributions of variations in the total stream length for each of twelve flow intermittency classes are also shown for (c) the AWRA-L model and (d) the WaterDyn model.

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 overestimation of gains (e.g. groundwater release to baseflow) during low-flow periods. Modelled patterns of flow intermittency revealed highly dynamic behaviour in space and time, with 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

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, improved understanding of temporal and spatial patterns in flow intermittency is fundamentally important for effective river management.

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 and enable ecologically-relevant characterisation of streamflow properties such as the magnitude, frequency, duration, and rate of change 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

65　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

70　(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

75　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

80　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.

85　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

90　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 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 characterised by a temperate climate condition, of which rainfall is relatively evenly distributed throughout the year and most months receive very similar averages.

[Figure 1 is about here]

3 Data and Methodology

3.1 Streamflow gauge data

Gauged daily streamflow data were sourced from the BoM water data website (http://www.bom.gov.au/waterdata) and were used to assess accuracy of AWRA-L modelled streamflow (Section 3.3) and to estimate an appropriate zero flow threshold of modelled streamflow data for quantifying patterns of streamflow intermittency (Section 3.4). Based on streamflow data availability, a total of 25 gauges in SEQ and 15 gauges in Tamar were selected (Fig. 1b, c) to assess modelled streamflow accuracy. These gauges had less than 0.5 % missing values over the period from 01/01/2005 to 31/12/2017 and had minimal hydrologic modification due to human activities. 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 (see Yu et al., 2018 for details of stream gauges). 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). However, the set of stream gauges used in our analyses under-represented the frequency of small low-order streams in both regions. Therefore, we regard the selected gauges to be representative of the range of 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.

4

3.2 Conversion from spatially contiguous runoff to streamflow

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
130   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 downloaded from BoM (http://www.bom.gov.au/water/landscape). 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
135   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, 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
140   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
145   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
150   grid (~250m resolution), and associated catchment hierarchy at the national scale. The routing model applied in this study was the Routing Application for Parallel computatIon 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).

155   To test the effects of river routing, we first calculated a series of flow metrics (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 flows, and rates of changes in flow events (Olden and Poff, 2003; Poff et al., 1997). Calculation of these streamflow characteristics allows a comprehensive assessment of the effects of river routing on streamflow simulations in the two regions. We then applied ~~Student's *t* test~~Wilcoxon rank sum test for each flow metric to 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) 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, we 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. 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. 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.

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$$ (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
185 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
190 Australian Millennium Drought (2001-2009), and includes a significant low-flow period. A preliminary analysis showed that AWRA-L modelled streamflow was sensitive to rainfall events, relative to the response of observed flow (Fig. 3). This finding indicates that over-responsiveness of AWRA-L to rainfall may potentially contribute to overestimation of low flow. We hypothesised 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
195 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 AWRA-L model 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
200 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 is about here]

3.4 Quantifying flow intermittency using spatially contiguous flow simulations

Given the fact that water balance models often over-predict the magnitude of very low flows due to the difficulties of
205 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 related measured zero~~no~~-flow duration at each gauge to catchment environment variables. ~~This involved three steps.~~

~~First,~~ ~~Ww~~We used linear regression to model the mean annual ~~no~~zero- flow duration (daily time step) at each gauge as a function of catchment environment variables. This regression analysis was only conducted in SEQ as most gauges in the Tamar
210 catchment had perennial flow. 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). Regression models were developed using all possible predictor variable combinations and we selected the "best" model for predicting zero flow duration based on corrected Akaike's Information Criterion (AICc) (Hurvich and Tsai, 1989). To estimate

7

215   the prediction error of the selected model, we applied leave-one-out cross validation on the selected 43 gauges and reported prediction error (R^2) to estimate the model prediction performance. Regression model development and cross-validation were conducted with the *MuMIn* and *boot* packages in R (R Development, 2017~~2~~). Regression analyses were performed on all combinations of predictor variables and the best model ~~returned the least~~with the lowest ~~–~~AICc (-54.2) retained ~~predictive model with~~ five covariates, including annual daily maximum temperature, catchment area, slope, average elevation, and

220   average saturated hydraulic conductivity. The developed predictive model showed a good model fit with an adjusted R^2 of 0.71, and the leave-one-out cross validation on the regression model showed relatively good model performance with an average R^2 of 0.64. We checked for spatial autocorrelation of the regression model residuals (as recommended by Dormann et al., 2007) and found they were not significantly autocorrelated (Moran's I = -0.06, p = 0.69). Examination of spatial residual maps further supported this conclusion, with no spatial trends in model residuals apparent. ~~The adopted thresholds range from~~

225   ~~0 – 1.668 m³/s, with the median value of 0.002 m³/s. The unexplained variations in the predictive model may be because only a subset of environmental attributes were included in the analysis and that the distribution of selected stream gauges was not fully representative of the environmental and hydrological conditions in the region.~~

~~Second,~~Next, ~~ww~~We ~~then~~ used the predictive models to extrapolate estimates of overall flow intermittency~~permanency~~ (in

230   terms of the proportion of days with zero flow) to each segment ~~in the entire~~throughout the river network. ~~Last,~~Finally, ~~ff~~for each segment, the time-series of daily runoff was truncated (flows below the threshold were set to "0") by adopting an appropriate threshold of "zero flow" that preserved the proportion of days with flow as estimated in the previous ~~at S~~step ~~2~~. The adopted thresholds ranged from 0 – 1.668 m³/s, with a median value of 0.002 m³/s. We recognise several sources of uncertainty in our approach to estimating the zero flow thresholds. The unexplained variation in the predictive model may be

235   due to the limited number of environmental attribute covariates used in the model and hence ability to adequately represent the range of environmental processes that influence streamflow intermittency. Additional uncertainty in model predictions may arise because the distribution of stream gauges used for model calibration under-represented the frequency of extremely small catchments that likely had higher cease-to-flow occurrence.

~~Regression analyses were performed on all combinations of predictor variables and returned the least -AICc (-54.2) predictive~~

240   ~~model with five covariates, including annual daily maximum temperature, catchment area, slope, average elevation, and average saturated hydraulic conductivity. The developed predictive model showed a good model fit with an adjusted R² of 0.71, and the leave-one-out cross validation on the regression model showed relatively good model performance with an average R² of 0.64 . The adopted thresholds range from 0 – 1.668 m³/s, with the median value of 0.002 m³/s. The unexplained variations in the predictive model may be because only a subset of environmental attributes were included in the analysis and~~

245   ~~that the distribution of selected stream gauges was not fully representative of the environmental and hydrological conditions in the region.~~

~~This truncation regression analyses wereas 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~~AWRA-L ~~model~~ in replicating observed patterns of cease-to-flow periods, we compared model outputs with those derived from a monthly water balance model – the WaterDyn model (Fig. 2). Monthly flow intermittency estimated from WaterDyn was thus used to benchmark results from the monthly AWRA-L. To do this, we ~~further~~ aggregated daily outputs to a monthly time step (termed "monthly AWRA-L" hereafter, Fig. 2). We tried two different aggregation methods. One 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), and the other considered that all days in a month must have zero flow for that month to be zero (termed "monthly AWRA-L_30" hereafter). 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 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 (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 AWRA-L, and monthly WaterDyn) was also tested against the measured flow intermittency derived respectively from daily and monthly observed streamflow data at gauged locations in SEQ.

Taking 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 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 AWRA-L and by the mean annual number of calendar months for WaterDyn. 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 AWRA-L and WaterDyn, 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; p values were greater than 0.50 for most flow metrics based on Wilcoxon test~~t test~~ results). There were noticeable but not statistically significant differences for ~~three~~two flow metrics related to low flows (the variability in timing, and the frequency ~~and the duration~~ of low

9

flow spells), and only the duration of low flow spells ~~but these differences were~~ was ~~not~~ statistically significant ~~at the 5 %~~ ~~level~~ (*p* = 0.03). These results suggested that the routing algorithm has nearly negligible effects on flow simulations in our

280 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 is about here]

4.2 Accuracy assessment of modelled streamflow in SEQ and Tamar

285 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). 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 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). However, no significant difference

290 was found in the overall model performance between the two hydro-climatically distinctive regions, according to Wilcoxon~~the~~ ~~two-sample Student's *t*~~ test (*w*~~t~~ = 247 ~~1.46~~, *p* = 0.10~~5~~).

[Figure 5 is about here]

Concerning model performance in simulating different components of flow regimes, the 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

295 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). 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). However, flow simulations in Tamar also exhibited slightly earlier estimations for the timing of low flow spells (13 % earlier), overestimations

300 for low flow spell frequency (92 % lower on average) and underestimation for low flow spell duration (58 % lower) (Fig. 6).

[Figure 6 is about here]

Varying degrees of difference in the magnitude of low flow between the observed and modelled were found among the gauges. There appeared to be a tendency toward larger differences with increasing catchment area in SEQ but not in Tamar (Fig. 7). The models appeared to over-estimate "*in situ*" gains to low flow in some reaches in both regions, while under-estimating

305 transmission losses in SEQ, suggesting that over-estimation of "*in situ*" gains in AWRA-L likely contribute to the overall overestimation of low flow in downstream catchments.

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
310   AWRA-L, and monthly WaterDyn). Annual flow intermittency calculated using daily AWRA-L flow (i.e., the average number of cease-to-flow days per year) was tested against annual flow intermittency estimated using observed data (Fig. 8a). The AWRA-L model displayed the potential to be used to estimate flow intermittency at a daily time step, with a 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
315   in relatively dry areas (e.g. ≥ 40 days of flow intermittency per year).

Figure 8b shows annual flow intermittency calculated using monthly AWRA-L flow and monthly WaterDyn flow. In this case, annual flow intermittency was defined as the average number of months characterized with zero flow. The WaterDyn model showed much more accuracy than the two aggregation methods based on the monthly AWRA-L model in estimating flow intermittency ($R^2 = 0.53$, 0.43 and 0.32 respectively for monthly WaterDyn, monthly AWRA-L_01 and monthly AWRA-
320   L_30). More specifically, the WaterDyn 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: monthly AWRA-L_01 overestimated flow intermittency and monthly AWRA-L_30 underestimated flow intermittency at nearly all gauges (Fig. 8b).

325

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. 9a, b). There w~~ereas a~~ noticeable spatial differences between model predictions of streamflow intermittency for low order inland streams~~, especially those lower-order streams~~. ~~More specifically~~For example, in the western Brisbane River catchment and the South
330   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. 9a). Compared to the monthly WaterDyn model, fewer streams were predicted by the daily AWRA-L ~~and~~ to experience extremely long dry events as well as less than one month of zero flows (Fig. 9c, d). ~~But~~ However, more streams on average (60 % vs. 49 % for the AWRA-L and WaterDyn model, respectively) were predicted
335   to flow intermittently (> 30 days or > 1 month) to varying degrees in SEQ, which suggest~~ed~~s that flow intermittency was prevalent in SEQ, irrespective of the water balance model used.

11

[Figure 9 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). 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 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, characterise 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 ≥ -0.41 at all gauges). Interestingly, although streamflow was more accurately simulated in the Tamar than in SEQ (the median values of KGE 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 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 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

Our results suggest that the temporal resolution of analysis should be dictated by the resolution of input streamflow data. 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. More specifically, tThe daily AWRA-L flow showed promise for estimating flow intermittency at a daily time step, while the monthly WaterDyn model was better than the monthly 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, our 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 one day or 30 days as a threshold and, respectively, either substantially overestimated or underestimated flow intermittency. This suggests that the temporal resolution of analysis (e.g. for flow intermittency estimation) should be dictated by the resolution as the streamflow data used in the analysis.

Spatially contiguous runoff data were used in this study 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 in SEQ was shown to vary significantly from year to year, ranging from 29 % to 80 % of total stream length for the AWRA-L model. However, Ggiven the limited spatial resolution of ability of the the current version of Geofabric stream network data (9 arc-second longitude-latitude resolution, with the minimum upstream drainage area of 1.5 km^2) used here and hence ability to resolve the smallest streams in SEQ, and that small streams are more likely to be intermittent, the proportion

14

of predicted intermittent streams in SEQ may be under-estimated in our study. Although there are differences in the temporal patterns of estimated flow intermittency between the AWRA-L and WaterDyn 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 is still scarce.

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.

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 a daily time step. The proposed approach has 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.

15

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

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.

Costigan, K.H., Jaeger, K.L., Goss, C.W., Fritz, K.M., Goebel, P.C., 2016. Understanding controls on flow permanence in intermittent rivers to aid ecological research: integrating meteorology, geology and land cover. Ecohydrology, 9(7): 1141-1153. DOI:10.1002/eco.1712

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

F. Dormann, C. et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30(5): 609-628. DOI:10.1111/j.2007.0906-7590.05171.x

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

Fritz, K.M. et al., 2013. Comparing the Extent and Permanence of Headwater Streams From Two Field Surveys to Values From Hydrographic Databases and Maps. JAWRA Journal of the American Water Resources Association, 49(4): 867-882. DOI:doi:10.1111/jawr.12040

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

Gallart, F. et al., 2017. TREHS: An open-access software tool for investigating and evaluating temporary river regimes as a first step for their ecological status assessment. Sci Total Environ, 607-608: 519-540. DOI:10.1016/j.scitotenv.2017.06.209

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

Hurvich, C.M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. Biometrika: 297-307.

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. et al., 2016. Ecological research and management of intermittent rivers: an historical review and future directions. Freshwater Biology, 61(8): 1181-1199. DOI:10.1111/fwb.12646

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

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.

18

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

610   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.

615   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

620   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.

625   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

630   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

635   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

640
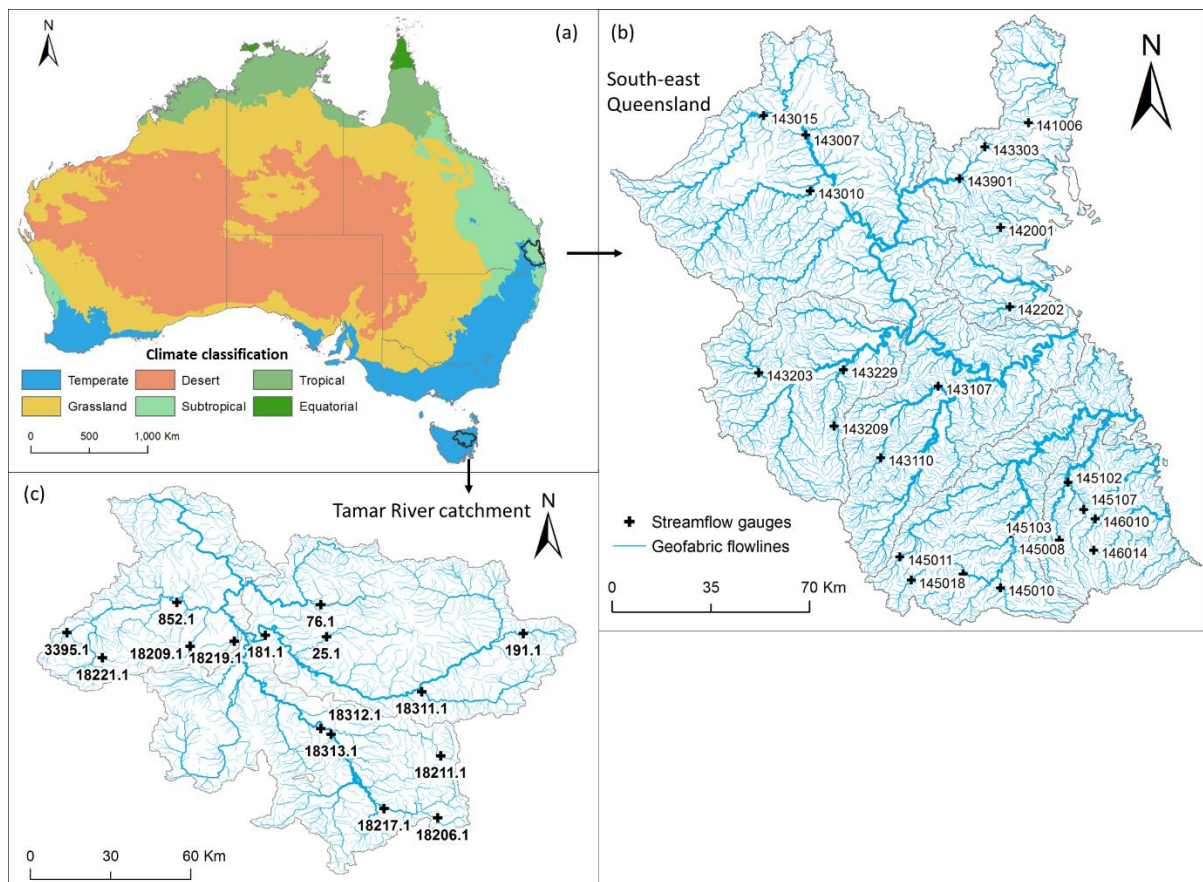
Figure 1. Locations of the two climatically and hydrologically distinctive regions in Australia (a): South-east Queensland (SEQ) (b) and the Tamar River catchment (Tamar) (c) with Geofabric river networks and selected stream gauges (25 and 15 gauges for SEQ and Tamar, respectively). The climate classification in (a) is based on the Köppen classification system (Australian Bureau of Meteorology, 2014).
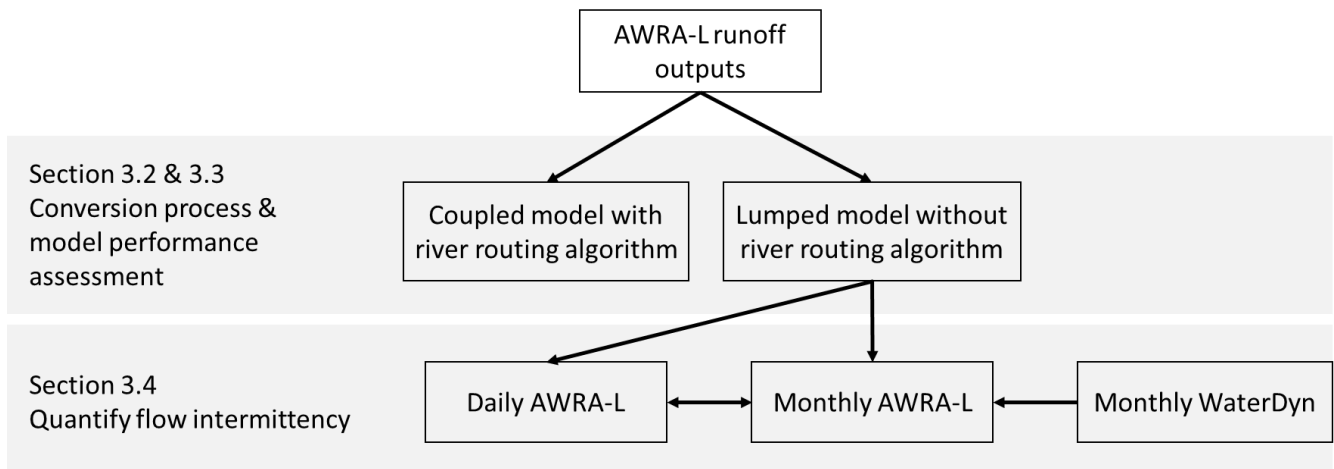
Figure 2. Model configurations and their applications in this study. AWRA-L runoff outputs are translated to accumulated streamflow estimates with river routing algorithm (coupled model) and without (lumped model). These two model configurations are applied to test the effect of river routing on streamflow 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.

650

655

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 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.

22
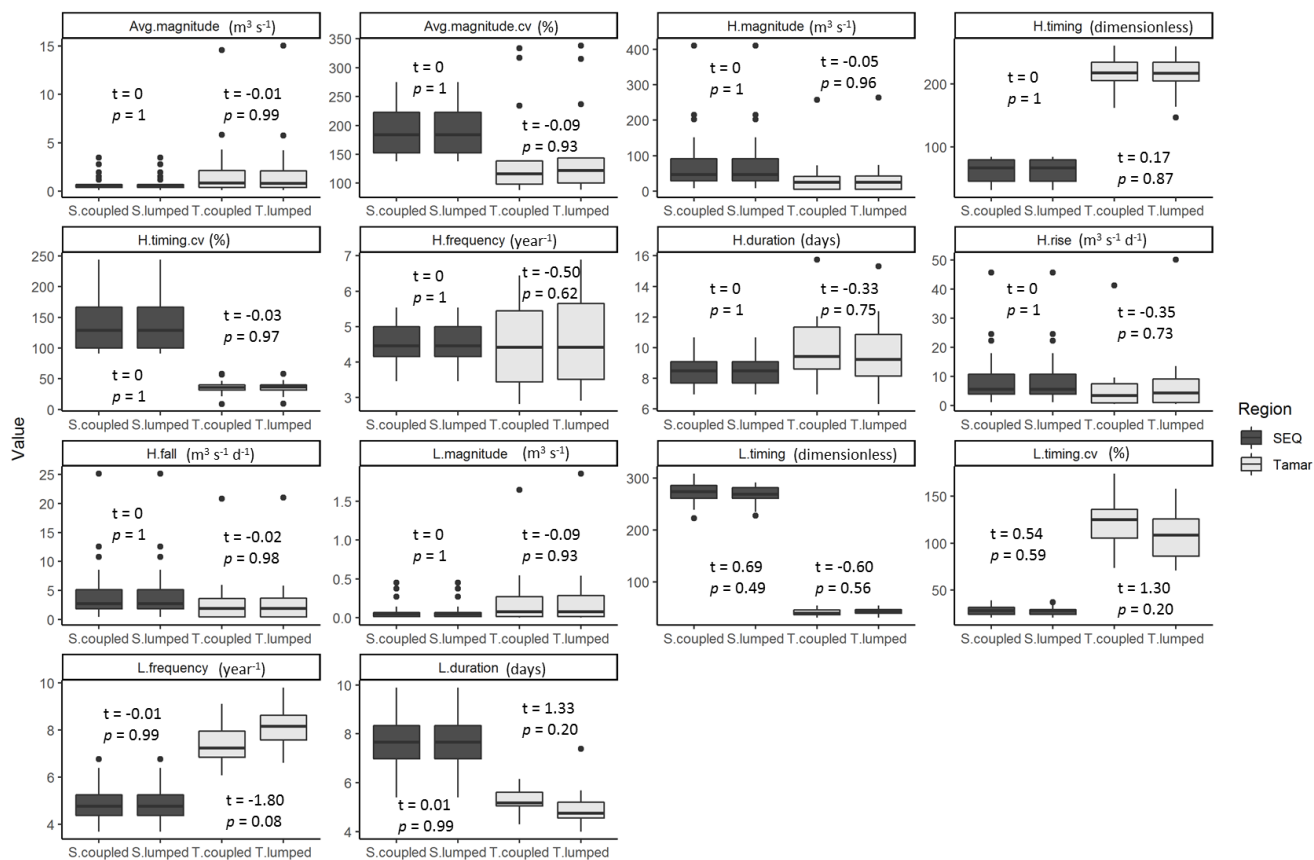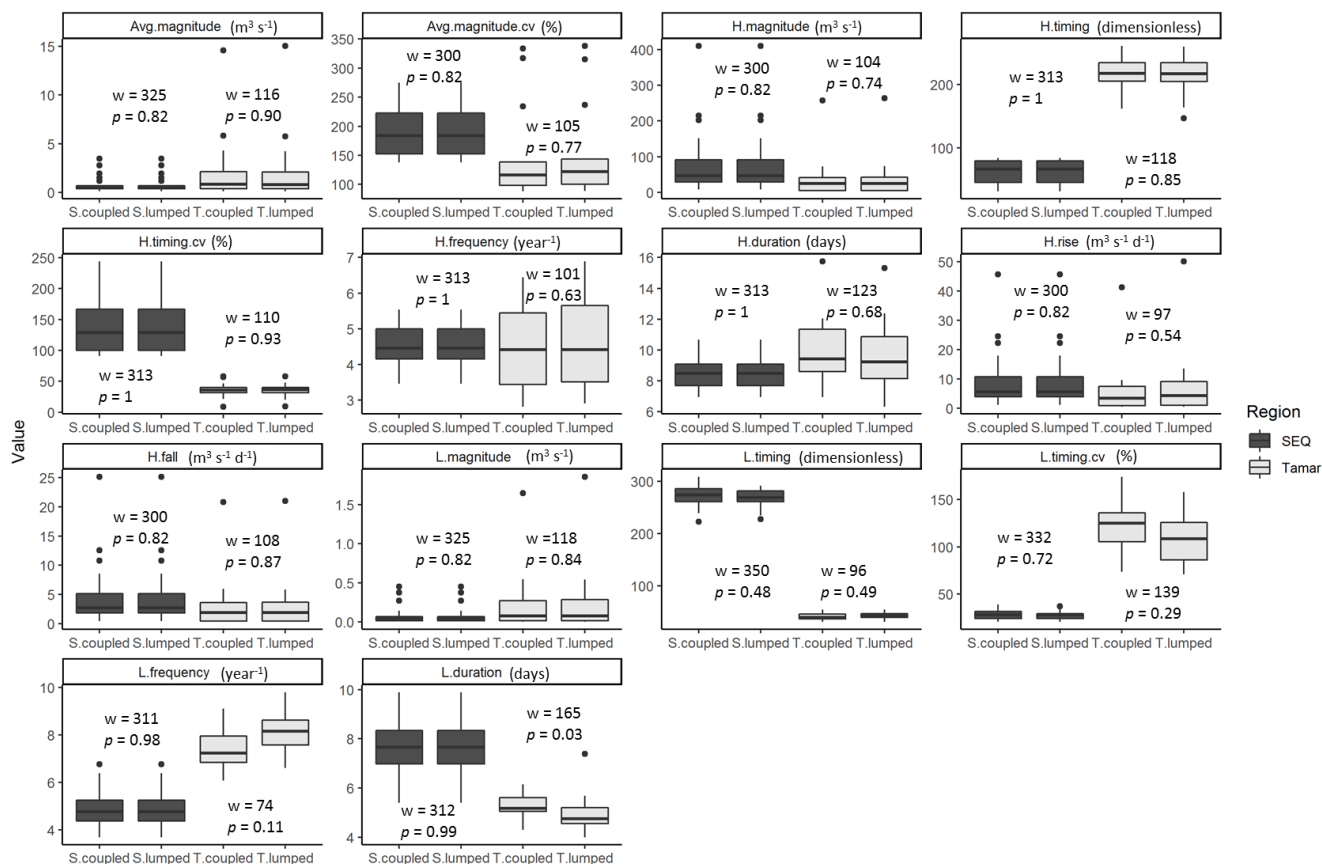
Figure 4. Comparison of hydrological characteristics between the lumped and coupled models in SEQ and Tamar. Refer to Table 1 for measurement description for each flow metric. Metrics are grouped according to average (Avg), high (H) and low (L) flow conditions. The values of _wt_ statistic and associated _p_ values are also shown to indicate whether there is any significant difference between the coupled and lumped simulations.

670
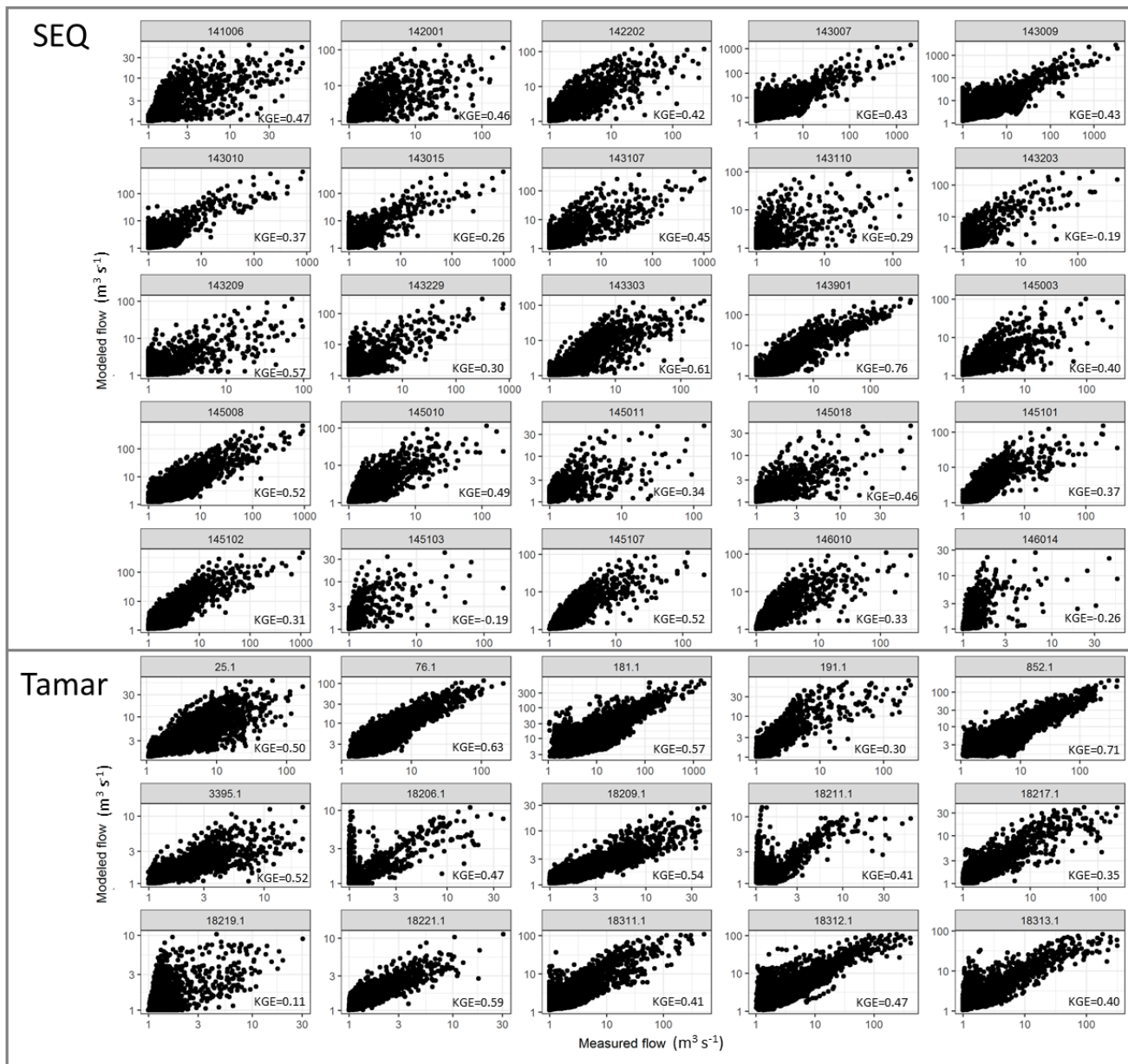
Figure 5: Scatter plots of the measured and modelled (lumped) streamflow for each gauge station in SEQ and Tamar. The modified Kling-Gupta efficiency (KGE) is presented in each panel. The x and y axes are log-transformed ($\log_{10}(x+1)$) to aid interpretation.

675
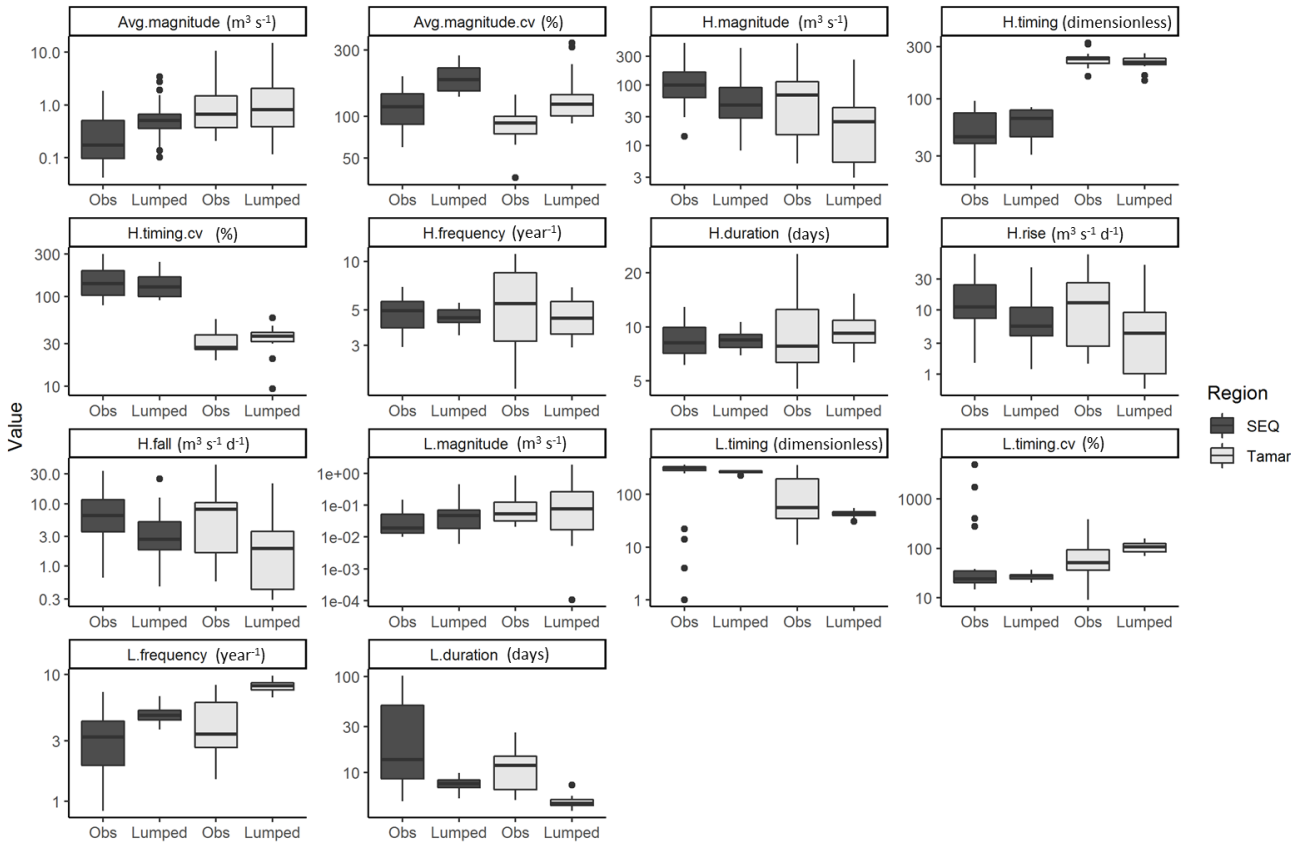
Figure 6. Variation in observed and modelled (lumped) hydrologic characteristics in SEQ and Tamar (n= 25 and 15 gauge locations, respectively). Refer to Table 1~~Table 1~~ for measurement description for each flow metric. Metrics are grouped according to average (Avg), high (H) and low (L) flow conditions. The y-axis is on a log scale for better interpretation.
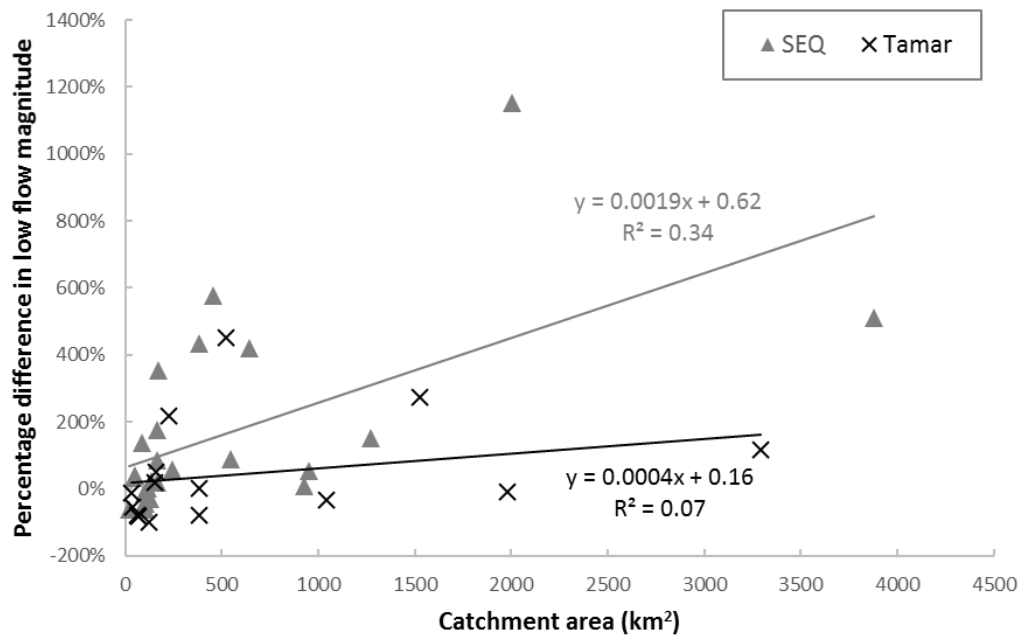
680   **Figure 7. Scatter plot of gauged catchment areas and percentage difference in low flow magnitude between the observation and simulation in SEQ (solid grey triangle) and Tamar (black cross). The regression line for each region is also shown as solid line (grey line for SEQ and black line for Tamar) with the regression function and R² value.**
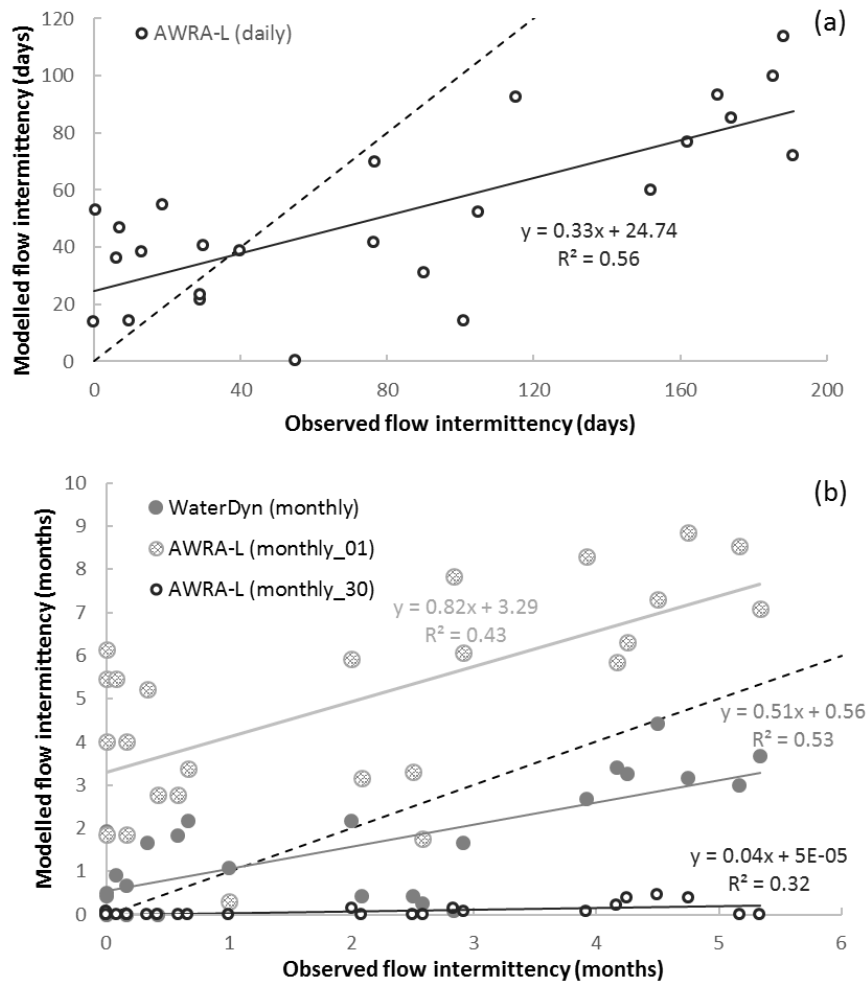
Figure 8. Scatter plots of the observed and modelled flow intermittency by the two models (AWRA-L and WaterDyn model) for SEQ. daily AWRA-L and monthly WaterDyn are derived from the original data from the two models, while AWRA-L monthly_01 and monthly_30 are flow intermittency estimates using the two 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.
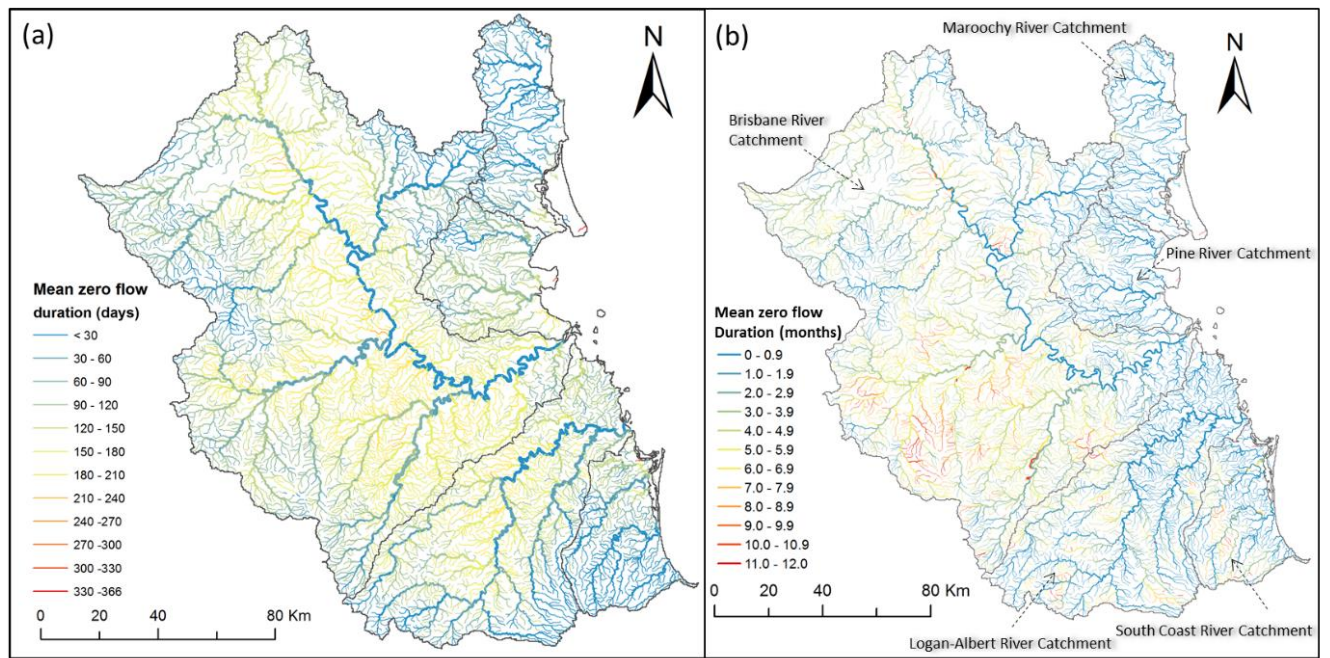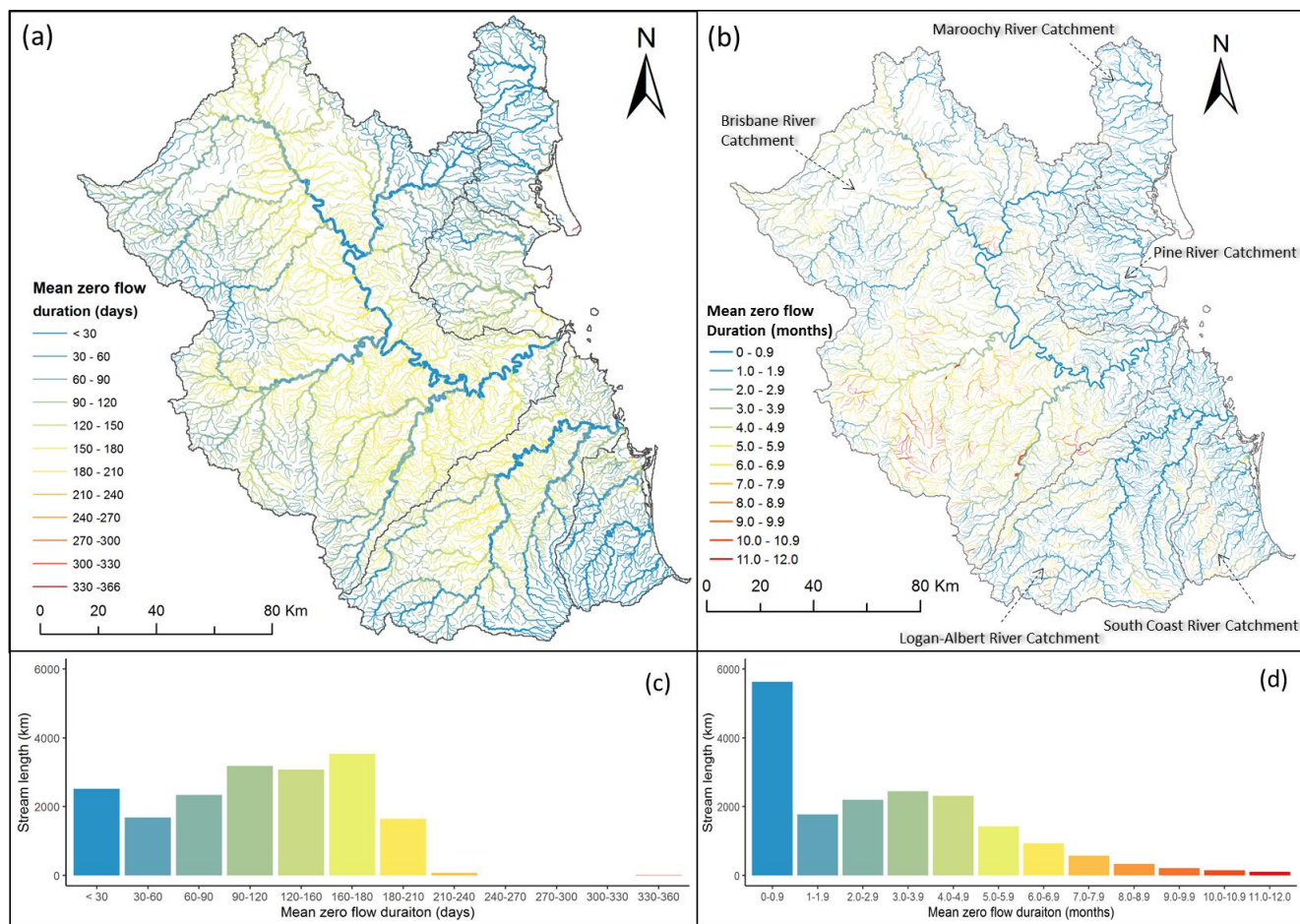
Figure 9. Comparison of the spatial pattern of average annual flow intermittency in SEQ derived from (a) daily flow simulations from the AWRA-L model and (b) monthly flow simulations from the WaterDyn model. Stream segments in both figures are coloured using the same frame but different units. Line thicknesses show the stream orders. Frequency distributions of The variations in the total stream length for each of the twelve flow intermittency groupsclasses are also shown as bar plot for (c) the AWRA-L model and (d) the WaterDyn model.
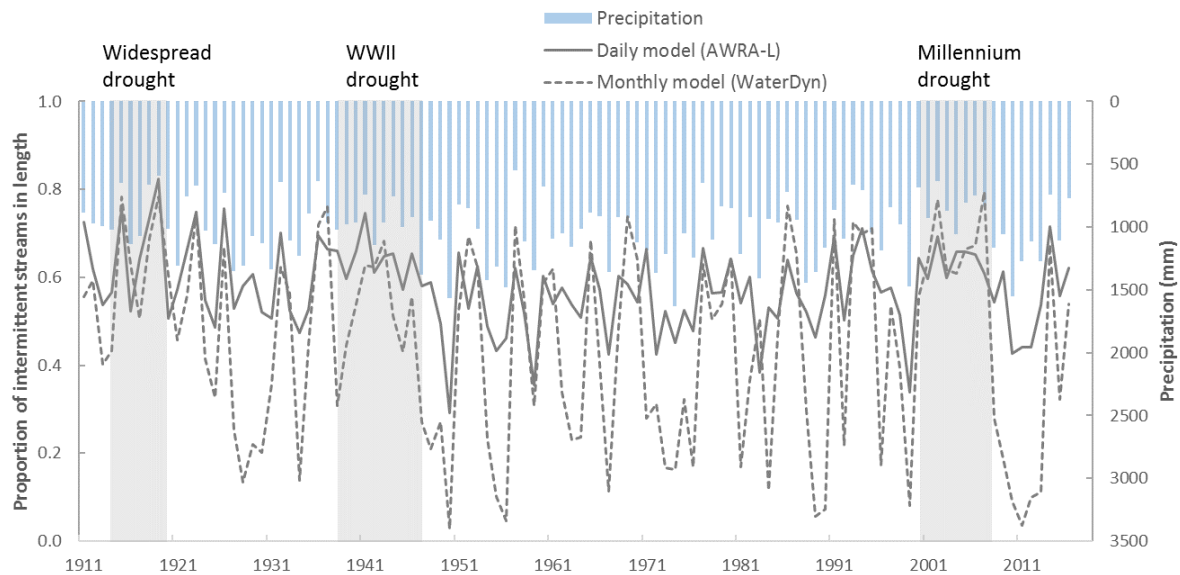
700 **Figure 10. Comparison of intra-annual variation of the proportion of intermittent streams in length from 1911-2016 across SEQ, derived from stream flow simulations from the daily flow model (lumped, solid grey line) and monthly flow model (grey dash line). Three severe droughts in Australia were 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 is shown for reference.**

705 **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