CSIRO Land and Water
Clunies Ross Street, Black Mountain, Canberra ACT
GPO Box 1700, Canberra ACT 2601 Australia
Telephone: +61 2 6246 5700 Facsimile: +61 2 6246 5800
www.csiro.au
ABN 41 687 119 230

07 August 2018

Dr Louise Slater

Editor for Hydrology and Earth System Sciences

Dear Louise,

First, we would like thank you and the two reviewers for the quite favourable reviews on our manuscript "How good are hydrological models for gap-filling streamflow data?" (hess-2018-250). We appreciate that all of you acknowledge that this is a concise but very interesting paper. It is really encouraging. Although the reviewers provided favourable comments and acknowledge the research value of this paper, they also gave insightful comments to clarify several important points, i.e. seasonality of missing data, model comparison, broad implications. All of these comments have been carefully considered and all comments have been adopted and incorporated to the improved revised version. The follows are key improvements:

a. New analysis on seasonality of missing data;
b. New analysis on model comparisons;
c. More literature review; and
d. More discussion on the broad implications, comparison for various approaches and future directions.

In the following sections, we provide point-to-point response to the comments, followed by the track changed version. Please let us know if there are any questions. Thanks again for you and the reviewers for your time, suggestions and comments.

Yours sincerely,

Yongqiang Zhang (on behalf of all co-authors)

Principal Research Scientist

CSIRO Land and Water

Clunies Ross St. Canberra ACT 2601, Australia

Dear Yongqiang Zhang,

We are pleased to inform you that the Editor report for the following manuscript is now available:

Journal: HESS

Title: How good are hydrological models for gap-filling streamflow data?

Author(s): Yongqiang Zhang and David Post MS No.: hess-2018-250 MS Type: Research article

Iteration: Minor Revision

The Editor has decided that minor revisions are necessary before the manuscript can be accepted. Please find the Editor Report at https://editor.copernicus.org/HESS/ms_records/hess-2018-250.

We kindly ask you to revise your manuscript accordingly and to upload the revised files, a point-by-point reply to the comments, and a marked-up manuscript version showing the changes made in your File Manager no later than 14 Aug 2018: https://editor.copernicus.org/HESS/file_manager/hess-2018-250. Please find all information on manuscript submission under https://www.hydrology-and-earth-system-sciences.net/for_authors/submit_your_manuscript.html.

Your revised manuscript will be reviewed by the Editor and you will be informed about the outcome by separate email.

Besides adjustments requested by the Editor or Referees, please check your manuscript carefully for typos, missing co-authors and their affiliations, terminology, updates of data in tables, or updates of variables in equations. All these have to be clarified with the Editor and therefore have to be included before you submit your revised manuscript. Should your manuscript be finally accepted it will not be possible to include such rather substantial changes anymore when your manuscript is in final production (proofreading).

To log in, please use your Copernicus Office user ID 217293.

Please note that all Referee and Editor reports, the author's response, as well as the different manuscript versions of the peer-review completion (post-discussion review of revised submission) will be published if your paper will be accepted for final publication in HESS.

You are invited to monitor the processing of your manuscript via your MS Overview: https://editor.copernicus.org/HESS/my_manuscript_overview

In case any questions arise, please contact me. Thank you very much for your cooperation.

Kind regards,

Natascha Töpfer

Copernicus Publications

Editorial Support

editorial@copernicus.org

on behalf of the HESS Editorial Board

Reviewer comments are in black and our responses are provided in this blue colour. We also use codes R1C1 to mean Reviewer 1 Comment 1, to allow for cross-referencing in the response letter and to aid navigation.

COMMENTS FROM EDITORS AND REVIEWERS:

Dear Authors,

Thank you for your responses to the two referees' reports. Based on my own reading of the manuscript, I find this is a concise but interesting paper that fits the scope of HESS well and will be of interest to the community.

EC1): The two reviews are both quite favourable, but they also make some important points about comparing the effect of gap-filling in different types of sites (e.g. in different regions and hydrological regimes), providing contextualisation and implications of the work (its wider significance and transferability to other contexts), and comparing the models. Additionally, some of the discussion and assertions could be better supported by references, such as the statement that 'it is well recognised that hydrological modelling is the best option' (for gap-filling). See for instance papers describing the utility/efficacy of different gap-filling approaches (e.g. those discussed in https://doi.org/10.1002/joc.4954).

I would therefore like to invite you to upload a revised manuscript, incorporating the proposed changes and additions, and making any other modifications where you see fit.

I look forward to receiving the revised manuscript.

With best regards,

Louise Slater

Response:

Thanks for you quite favourable and constructive comments. As stated in the above letter to Editor, we made following revisions based on the comments from Editor and two reviewers Maxine Zaidman and Juraj Parajka:

a. New analysis on seasonality of missing data;

b. New analysis on model comparisons;
c. More literature review; and
d. More discussion on the broad implications, comparison for various approaches and future directions.

We add more literatures on various approaches for data gap-filling. In lines 60-63, the clear text now says "*There are many methods used for gap-filling the missing data, including interpolation from nearby gauges (Hannaford and Buy, 2012; Lavers et al, 2010; Lopes et al., 2016), statistical methods (Gedney et al., 2006b), hydrological modelling (Dai et al., 2009; Sanderson et al., 2012), and multiple infilling methods (Harvey et al., 2012).*". We add more references to support the use of hydrological modelling approaches. In lines 63-67, the text now says "*Among them, the hydrological modelling method is widely used since it fully considers the spatial heterogeneity and temporal variability of climate forcing data, and can achieve sufficient simulations when it is calibrated against a small number of observations (Peña-Arancibia et al. 2014; Rojas-Serna et al., 2016; Seibert and Beven, 2009; Liu and Zhang, 2017)*".

We also add discussion on high-flow gap filling impacts. In lines 334-337, the clear text now says "*In contrast, the high flow gap-filled data shows a noticeable change in annual streamflow trend when the missing rate is 5% This is understandable since high flow is usually several orders of magnitude higher than low flow, and errors in filling high flow could have large impacts on annual flow and its trends (Slater and Villarini, 2017).*".

We also discuss more comparisons between hydrological modelling and other gap filling approaches. In lines 388-394, the text now says "*It would also be interesting to compare hydrological modelling to other approaches for filling streamflow data gaps. Hydrological modelling is a most useful method used in Australia for predicting daily streamflow in ungauged catchments (Chiew et al., 2009; Li and Zhang, 2017; Zhang and Chiew, 2009; Viney et al., 2009). It has been used operationally by the Australian Bureau of Meteorology for filling daily streamflow data gap for many years. In the future, this operational method could further be comprehensively evaluated against other approaches, such as interpolation or correlations with nearby gauging sites.*".

New References:

Hannaford, J., and Buys, G.: Trends in seasonal river flow regimes in the UK, Journal of Hydrology, 475, 158-174, 10.1016/j.jhydrol.2012.09.044, 2012.

Harvey, C. L., Dixon, H., and Hannaford, J.: An appraisal of the performance of data-infilling methods for application to daily mean river flow records in the UK, Hydrology Research, 43, 618-636, 10.2166/nh.2012.110, 2012.

Lavers, D., Prudhomme, C., and Hannah, D. M.: Large-scale climate, precipitation and British river flows Identifying hydroclimatological connections and dynamics, Journal of Hydrology, 395, 242-255, 10.1016/j.jhydrol.2010.10.036, 2010.

Sanderson, M. G., Wiltshire, A. J., and Betts, R. A.: Projected changes in water availability in the United Kingdom, Water Resources Research, 48, 10.1029/2012wr011881, 2012.

Slater, L., and Villarini, G.: On the impact of gaps on trend detection in extreme streamflow time series, International Journal of Climatology, 37, 3976-3983, 10.1002/joc.4954, 2017.

Referee #1

M. Zaidman (Referee)
maxine.zaidman@jbaconsulting.com
Received and published: 9 July 2018

Overall comments:
R1C1): Generally a clear well-written paper. The underlying science appears to have been undertaken robustly, methodically and consistently. My main thoughts, having digested the submission, were to the wider scientific significance of the work presented. Has this been suitably explored within the context of the work? Currently the paper has a colloquial emphasis (Australia) and as a reader in the UK, I would like the authors to make a comment on whether the results are transferable elsewhere and also on how much dependency there is on the type of model used for infilling and patterns of missing data. Even for a more direct audience (e.g. users of Australian streamflow data / those wising to understand the reliability of trend detection analysis in an Australian context) the benefits/implications of the outcomes of the work could be drawn out in the paper a little more. At the very end of the paper the authors tantalise the reader by hinting at other patterns within the dataset, beyond the scope of the study to explore at this point. Ideally for me, this paper would give more value if it took a stance of saying, having established that gap filling does not impact on trend analysis, what the trend analysis on the gap-filled data shows and whether this changes our perception on the strength and direction of trend for either individual sites or regionally. Finally, what a shame the paper does not address the potential payback of infilling with modelled data compared with other methods (like interpolation or correlations with nearby sites for example). Would there be less confidence in the trend analysis results if modelling had not been used as the gap filling method. Having said the above, I would not object to publication of this paper in its current form (no suggested corrections to the text). It is a self-contained work that no doubt many hydrologists will find useful.

Response: We do appreciate the favourable comments from Maxine Zaidman. Maxine highlights the underlying science appears to have been undertaken robustly, methodically and consistently.

We are grateful for her thoughtful thinking on how to transfer the results obtained from Australia to other parts of the work. It is indeed it is important to discuss the implication. To this end, we add one paragraph in Discussion section. In lines 377-387, the text says "*The modelling experiments and findings from this study could have important implications for* *other parts of the world as well as Australia. First, to develop appropriate gap-filling modelling experiments, it is necessary to evaluate the distribution of consecutive missing data pattern. The probability distribution of consecutive missing data is skewed toward the low end, which can be nicely simulated using the Gamma distribution (Eq.1). This distribution should be very useful for similar missing patterns in other regions. Second, hydrological modelling is a very good tool for filling gaps since it can fully take the advantage of climate forcing and non-gap streamflow data, and obtain the best possible daily simulations. Third, the threshold of 10% identified in this study should be applicable to regions/catchments with similar missing patterns. However, if the data gaps continue for seasons or years, the threshold may not hold."*

In terms of comparisons between modelling and other methods (like interpolation or correlations with nearby sites for example), it is well recognised that hydrological modelling in Australia is the best option since it fully takes advantage for climate forcing and non-gap streamflow data. We add one paragraph for discussing the comparison. In lines 388 to 394, the text now says "*It would also be interesting to compare hydrological modelling to other approaches for filling streamflow data gaps. Hydrological modelling is a most useful method used in Australia for predicting daily streamflow in ungauged catchments (Chiew et al., 2009; Li and Zhang, 2017; Zhang and Chiew, 2009; Viney et al., 2009). It has been used operationally by the Australian Bureau of Meteorology for filling daily streamflow data gap for many years. In the future, this operational method could further be comprehensively evaluated against other approaches, such as interpolation or correlations with nearby gauging sites".*

Specific comments:
R1C2) Abstract: The point that springs to my mind is that if gap filling has so little impact, then why bother to undertake it in the first place? Presumably the gap filling is being undertaken de rigour/as part of data QA for reasons of consistency / completeness and the purpose here is to show this does not have negative impact on key hydrological analyses (of which trend analysis might be just one?). The abstract also states there is a lack of quantitative analysis of gap filled data. Is this really true, across the entirety of the international body of scientific literature.

Response: In our knowledge, it is indeed that there is lack of quantitative evaluation of the gap-filled data accuracy in most hydrological studies. The scientists basically use a threshold, based on some kind of gut feeling. This study can fill knowledge gap. This study provides two key findings: (1) when the missing rate is less than 10%, the gap-filled streamflow data obtained using calibrated hydrological models perform almost as same as the benchmark data (less than 1% missing) for estimating annual trends for 217 unregulated catchments widely spread in Australia; (2) the relative streamflow trend bias caused by the gap-filling is not very large in very dry catchments where the hydrological model calibration is normally poor.
In terms of why it is undertaken in the first place, it is generally done by the collecting agency (in Australia, the Bureau of Meteorology), as end users often require streamflow data with no gaps.

R1C3. Data and methods: I'm interested to know whether the timing of missing data impacts on the trend analysis outcomes. Presumably the % rates are across the period of record of each site? Could you explain reasons for the gaps in the records, e.g. are all the stations gauged in the same way or are some types of station/ river more vulnerable to gaps than others (e.g. stations on smaller flashy rivers). Was there consideration of data quality outside of periods with gaps. Are stations with more gaps likely to suffer poorer data quality overall.

Response: In most cases, missing data are randomly distributed and different gauges show different missing pattern. This can be seen from missing patterns in Fig. 3 that there is a skewed distribution for consecutive missing days. This means that majority of the consecutive missing days are less than 30 days. The data gaps for Australian streamflow gauges mainly include (see lines 95-97):

1. Non-sensible record
2. Sensor broken
3. No recorded data (Instrumentation removed)
4. No data exists
5. No record or record lost

In terms of timing of missing data and reasons of gaps, we further plot a boxplot plot (Figure 4). Yes, the missing data are more-less evenly distributed through different seasons across all 39 catchments (with missing rate of 8% to 12%) within the 10% missing data group. This indicates that the data gaps were not skewed toward a particular season and it occurred randomly through the year. Having said that, we actually conducted independent modelling experiments (but did not show them in the previous version) to test the consequence if the missing streamflow only occurs in high-flow or low flow seasons in the extreme cases. *In lines 327 to 337, the text now says "It is possible that data gaps may only exist during high flow or low flow conditions, although that is not what we observed here with the majority of missing data being more or less evenly distributed throughout the year (Figure 4). We did however test the impact of filling streamflow data in high flow or low flow conditions (results not shown here). In those cases, the missing patterns were selected using only high flow (>95th percentile) or low flow (less than 50th percentile) data. The results obtained from the low flow gap-filling indicates that there is only a negligible influence on annual streamflow trend estimates when the missing rate is less than 50%. In contrast, the high flow gap-filled data shows a noticeable change in annual streamflow trend when the missing rate is 5% This is understandable since high flow is usually several orders of magnitude higher than low flow, and errors in filling high flow could have large impacts on annual flow and its trends (Slater and Villarini, 2017).".*

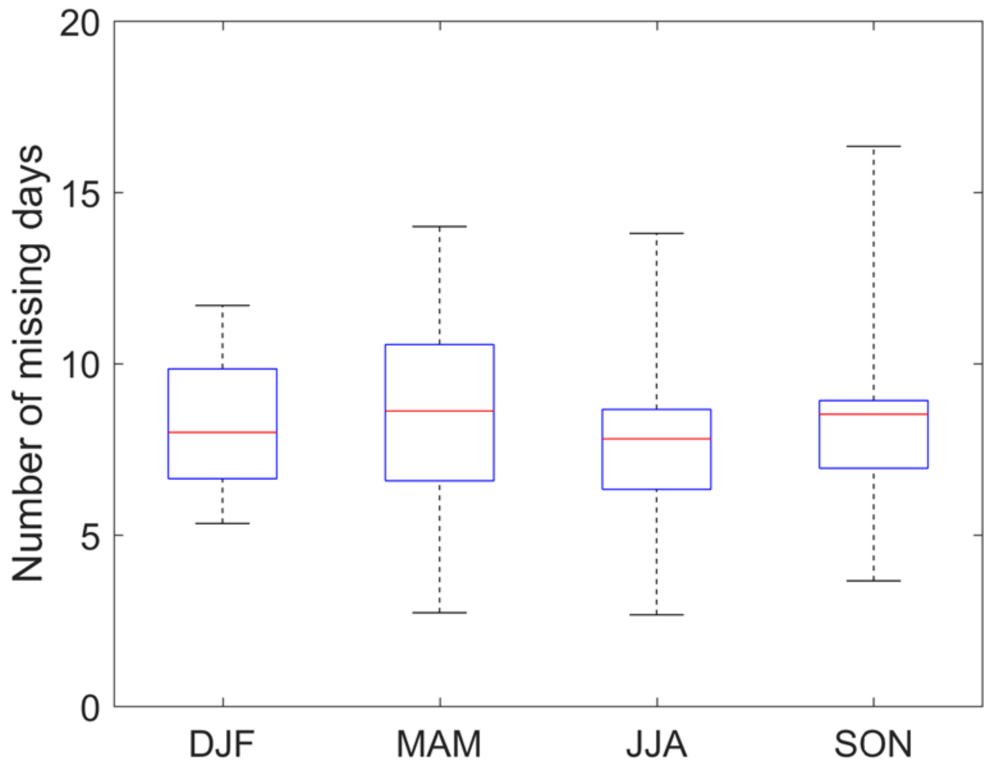


Fig. 4. Distribution of number of missing days across different seasons, summarised from 39 catchments with a missing rate ranging from 8% to 12% (i.e. 10% missing data group).

We did not consider data quality outside of periods with gaps.

In term of "Are stations with more gaps likely to suffer poorer data quality overall", we do not sure what the poor data quality refer to. If the review talked about poorer simulation quality, we compared the trends between the three gap-filling experiments (Fig. 7). It is clear that the trend biases between 5% and 10% missing experiments are similar. For GR4J, both have the trend bias varying from -1 to 1 mm/year/year; For SIMHYD, the trend bias between the two is similar when it varies from -0.5 to 1 mm/year/year, and the trend bias for 5% missing experiment is even larger than that for 10% missing experiment. The trend bias for 20% missing experiment is noticeably larger than that for 10% and 5% missing experiments for both models, and the underperformance is more noticeable from SIMHYD gap-filled than that from GR4J gap-filled. This result suggests that the trend bias is reasonable when the missing rate is less than 10%, and can be large for small number of catchments when the missing rate is to 20%.

R1C4. Results: It is stated that the model performance is not as good for high flows, but the analysis considers annual trends (annual average flows?). Was any analysis of trends in high flow patterns attempted and if so was there a different outcome. I'd also like to see more exploration and explanation of differences seen between the SIMHYD and GR4J results. Does one model theoretically outperform the other? Are the differences between the infilled trend analysis for the two models the same order of magnitude as between trend from filled and unfilled series etc. I just wonder if we need more discussion in this section to draw out some useful implications or provisos. Should one model be preferred or give a greater payback (i.e. Gap filling is just as good but the model is more practicable to use/more straight forward to parameterise).

Response: The two model are overall good for high flow simulations as demonstrated by high NSE and low bias. It only slightly underestimates very high flow (i.e. floods).

We have not had analysis of trends in high flow patterns.

We include a comparison between SIMHYD and GR4J models. Figure 5 summarises the Comparisons between calibrated GR4J and calibrated SIMHYD for 44 catchments of the 5% missing experiment, 39 catchments of the 10% missing experiment, and 22 catchments of the 20% missing experiment. It is in general that there is no systematic difference between the two. In lines 232-237, the text now says "Overall, the two models perform well and GR4J does not systematically outperform SIMHYD (Figure 5). For the three groups of gap-filling experiments, these two models performs similarly (i.e. the difference of NSE of daily runoff between two is less than 0.02) in 18-19% catchments; SIMHYD model outperforms GR4J model (NSE difference between two is larger than 0.02) in 30-31% catchments; GR4J model outperforms SIMHYD model in 50-51% catchments". In 18-19% catchments, these two models performs similarly (i.e. the difference of NSE of daily runoff between two is less than 0.02); in 30-31% catchments SIMHYD model outperforms GR4J model (NSE difference between two is larger than 0.02); in 50-51% catchments, GR4J model outperforms SIMHYD model.

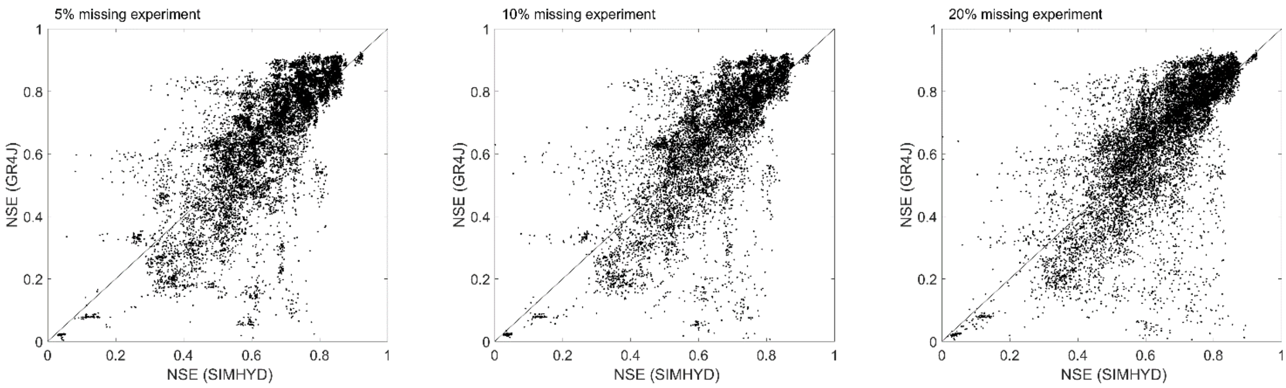
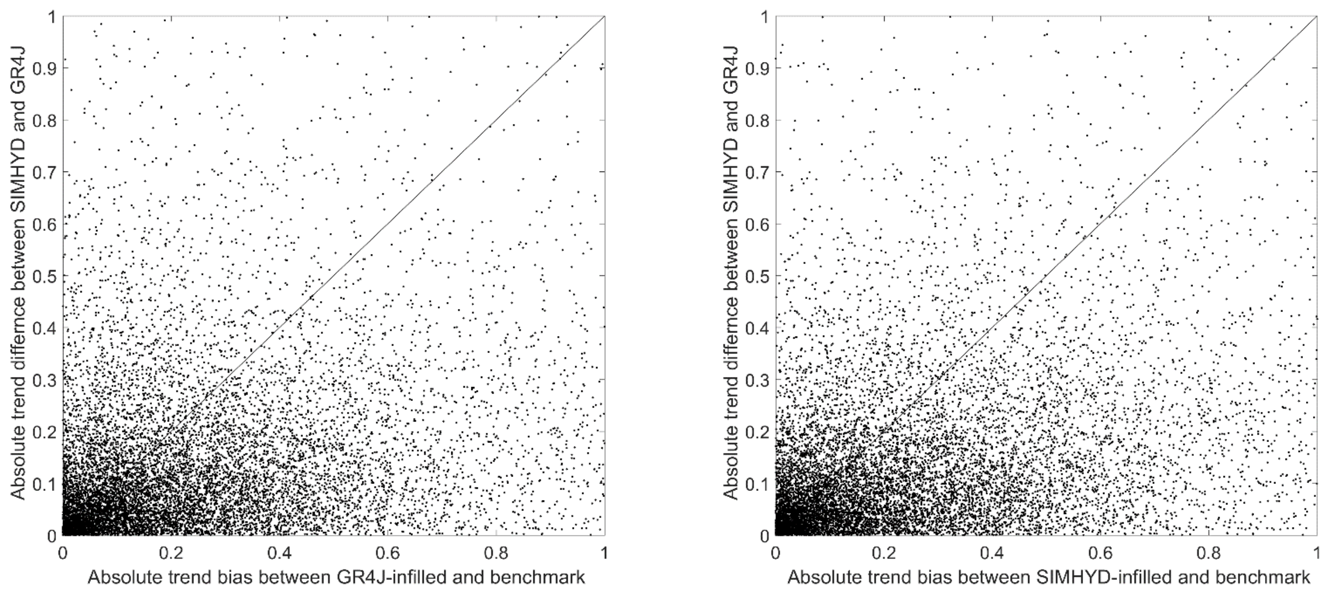


Fig. 5. Comparisons between calibrated GR4J and calibrated SIMHYD for 44 catchments of the 5% missing experiment, 39 catchments of the 10% missing experiment, and 22 catchments of the 20% missing experiment. In each catchment, there were100 replicates carried out.

We also compare the difference between the infilled trends for the two models to the difference the infilled and infilled trends. As shown in the following figure, they are with the similar order (but this figure is not shown in the main text).



Referee #2

J. Parajka (Referee)
parajka@hydro.tuwien.ac.at
Received and published: 21 July 2018

Overall comments:
General comments
R2C1. This study explores the efficiency of gap-filling of streamflow data by using simulations of a hydrologic model. The main objective is to evaluate the annual trends and annual variables obtained from gap-filled streamflow data using two hydrological models (GR4J and SIMHYD) in 217 catchments in Australia. The results show that when the missing rate of streamflow data is less than 10%, the gap-filled streamflow data from hydrological models perform very close to the benchmark data. Interestingly, the relative streamflow trend bias caused by the gap-filling is not very large even in very dry catchments where typically the hydrological model calibration is poor. Authors conclude that the gap filling using hydrological modelling has little impact on the estimation of annual streamflow and its trends in selected catchments in Australia.

Overall, the study is very clearly written, has a good structure and it is within the scope of HESS. The presentation of take home messages is very compact and clear. I have only one question which remained unanswered after reading the manuscript. What is the impact of patterns of missing data in terms of dominant hydrologic regime in the catchments? I expect that the large dataset in Australia covers catchments with different hydrological (seasonal) runoff regime. Are the missing data more-less evenly distributed thorough the year in all catchments or are there some seasonal patterns of gaps? What is the impact if majority of missing data are from the most/least important season (in terms of maximum monthly runoff)? I would expect that if the majority of e.g. 10% missing data are from seasons with minimum monthly runoff then the impact
on annual mean or trend will be smaller and vice versa. Are there some differences between catchments with different seasonal regime? Some more discussion around it will be interesting.
Finally I would like to congratulate the authors for a very nice analysis. I enjoyed reading it.

Response: We do appreciate the favourable comments from Juraj Parajka. Juraj highlights the science quality of this study and quality presentation.

To address the question Juraj raised regarding seasonal pattern of number of the missing days, we have included a new boxplot in the paper (Figure 4). Yes, the missing data are more-less evenly distributed through different seasons across all 39 catchments (with missing rate of 8% to 12%) within the 10% missing data group. This basically suggests the streamflow is missing randomly through the year. Having said that, we actually conducted independent modelling experiments (but did not show them in the previous version) to test the consequence if the missing streamflow only occurs in high-flow or low flow seasons in the extreme cases. *In lines 327 to 337 the text now says "It is possible that data gaps may only exist during high flow or low flow conditions, although that is not what we observed here with the majority of missing data being more or less evenly distributed throughout the year (Figure 4). We did however test the impact of filling streamflow data in high flow or low flow conditions (results not shown here). In those cases, the missing patterns were selected using only high flow (>95th percentile) or low flow (less than 50th percentile) data. The results obtained from the low flow gap-filling indicates that there is only a negligible influence on annual streamflow trend estimates when the missing rate is less than 50%. In contrast, the high flow gap-filled data shows a noticeable change in annual streamflow trend when the missing rate is 5% This is understandable since high flow is usually several orders of magnitude higher than low flow, and errors in filling high flow could have large impacts on annual flow and its trends (Slater and Villarini, 2017)."*

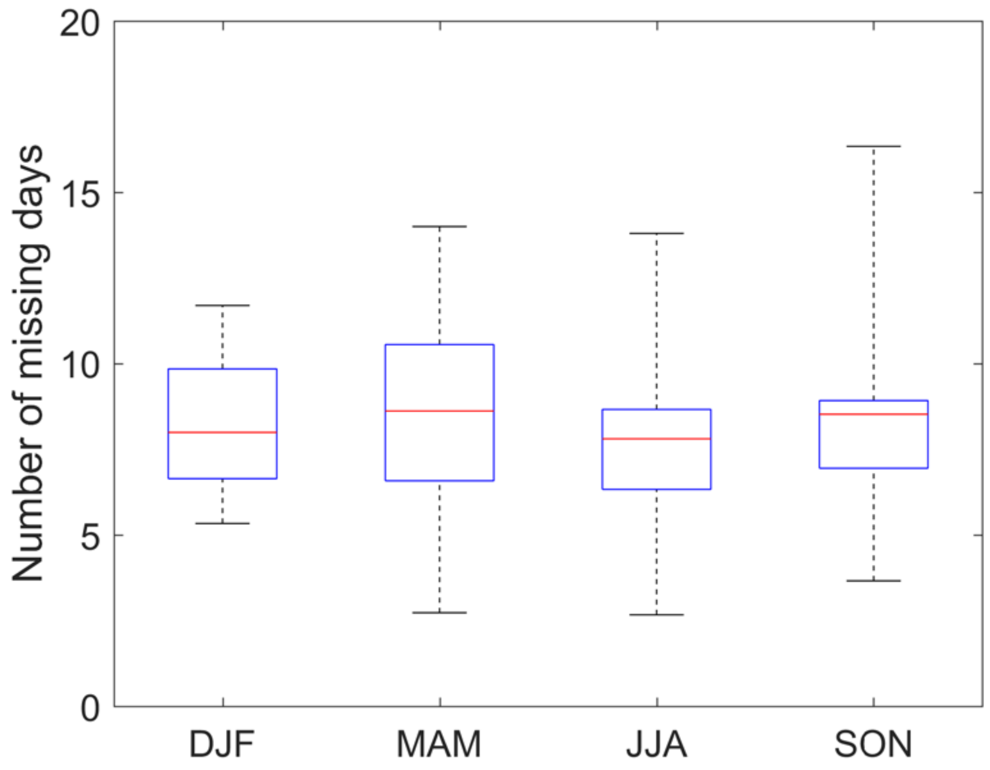


Fig. 4. Distribution of number of missing days across different seasons, summarised from 39 catchments with a missing rate ranging from 8% to 12% (i.e. 10% missing data group)

How good are hydrological models for gap-filling streamflow data?

Yongqiang Zhang[1*], David Post[1]

[1] CSIRO Land and Water, GPO Box 1700, ACTON 2601, Canberra, Australia

Submission to: ***Hydrology and Earth System Sciences***

Submission date: May 2018

[*]Corresponding author: Yongqiang Zhang

Email: yongqiang.zhang@csiro.au; yongqiang.zhang2014@gmail.com

Address: CSIRO Land and Water, Clunies Ross Street, Canberra 2601, Australia

Tel.: +61 2 6246 5761

Key Points:

- Gap-filling of streamflow data performs well when the missing rate is less than 10%
- Small number of catchments showing large trend bias when the missing rate is up to 20%
- Poor gap-filling occurring in some wet catchments even with reasonable model calibration

Abstract. Gap-filling streamflow data is a critical step for most hydrological studies, such as streamflow trend, flood and drought analysis and hydrological response variable estimates and predictions. However, there is lack of quantitative evaluation of the gap-filled data accuracy in most hydrological studies. Here we show that when the missing rate is less than 10%, the gap-filled streamflow data obtained using calibrated hydrological models perform almost as same as the benchmark data (less than 1% missing) for estimating annual trends for 217 unregulated catchments widely spread in Australia. Furthermore, the relative streamflow trend bias caused by the gap-filling is not very large in very dry catchments where the hydrological model calibration is normally poor. Our results clearly demonstrate that the gap-filling using hydrological modelling has little impact on the estimation of annual streamflow and its trends.

Keywords: streamflow, data, gap-filled, hydrological model, trend

1 Introduction

Streamflow is channel runoff, i.e. the flow of water in streams and rivers and accumulated from surface runoff from land surface and groundwater recharge. It is one of the major water balance components in a catchment where precipitation is partially stored in surface water, soil and groundwater stores, and the rest is partitioned into two fluxes: evapotranspiration and streamflow. It is almost impossible to measure evapotranspiration dynamics at a catchment scale. In contrast, streamflow time series can be easily measured at a catchment outlet. Therefore, streamflow data becomes a fundamental dataset underpinning hydrological studies. Without such a dataset, it is hard to understand catchment hydrological processes under climate change and non-stationarity（Dai et al., 2009; Gedney et al., 2006a; Ukkola et al., 2015; Zhang et al., 2016b).

Unfortunately, streamflow data are not always continuously available and most gauges suffer from streamflow data missing issues（Dai et al., 2009). Often, the missing rate is important when selecting streamflow gauges, especially when the data is used for annual trend analysis. To choose qualified catchments, researchers often set up a threshold for the missing ratio, for instance 1%（Petrone et al., 2010), 5%（Ukkola et al., 2015), 10%（Déry et al., 2009), 15%（Liu and Zhang, 2017), and 20%（Lopes et al., 2016). Only those gauges with missing rate less than a particular threshold are selected, and the rest are excluded for further analysis because of high missing rates.

There are many methods used for gap-filling the missing data, including interpolation from nearby gauges（Hannaford and Buy, 2012; Lavers et al, 2010; Lopes et al., 2016), statistical methods（Gedney et al., 2006b), ~~and~~ hydrological modelling（Dai et al., 2009; Sanderson et al., 2012), and multiple infilling methods (Harvey et al., 2012). Among them, the hydrological modelling method is widely used since it fully considers the spatial heterogeneity and temporal variability of climate forcing data, and can achieve sufficient simulations when it is calibrated against a small number of observations (Peña-Arancibia et al. 2014; Rojas-Serna et al., 2016; Seibert and Beven, 2009; Liu and Zhang, 2017). This is particularly important in Australia where hydrological modelling is a major tool for simulating continuous streamflow at a catchment scale. More recently, the Australian Bureau of Meteorology used a hydrological model –GR4J– to infill missing daily streamflow data for 222 Hydrologic Reference Stations (http://www.bom.gov.au/water/hrs/about.shtml). The gap-filled streamflow data are then used for trend analysis and providing hydrological information to all users.

One major concern for the hydrology community is to understand how reliable the gap-filled data is. Unfortunately there are no studies in the literature to comprehensively evaluate the reliability and accuracy of the gap-filled data that are influenced by different thresholds and by data missing patterns. Our study aims to provide a framework to evaluate the annual trends and annual variables obtained from gap-filled streamflow data using two hydrological models (GR4J and SIMHYD) together with a large streamflow dataset available across the Australian Continent（Zhang et al., 2013). This can guide researchers to more sensibly define a threshold for catchment selection and hydrological analysis.

2 Data and Methods

2.1 Data

We obtained daily streamflow data set from 780 unregulated catchments widely spread across Australia（Zhang et al., 2013). The dataset has undergone strict quality assurance and quality control, including quality codes check and spike (i.e. outlier points) control, and covered the period from 1975 to 2012. This dataset has been used by modellers for various hydrological modelling and extreme-event studies（Li and Zhang, 2017; Liu and Zhang, 2017; Ukkola et al., 2016; Yang et al., 2017). The missing rate for the pre-1980 and post-2010 periods were high. To meet our study requirement, we selected 217 catchments with a data missing rate less than 1% for the period 1981-2010 and the streamflow data for the 217 catchments are regarded as 'benchmark' data (Figure 1). Out of the 780 catchments there are 146, 91, and 61 with the missing rate of 1-5%, 5-10%, and 10-20% during 1981-2010, respectively (Figure 1), and these catchments account for 38% of total available catchments. Table 1 summarises major catchment attributes for the 217 selected catchments. The data gaps for Australian streamflow gauges mainly include: i) non-sensible record; ii) sensor broken; iii) no recorded data (instrumentation removed); iv) no data existed; and v) no record or record lost.

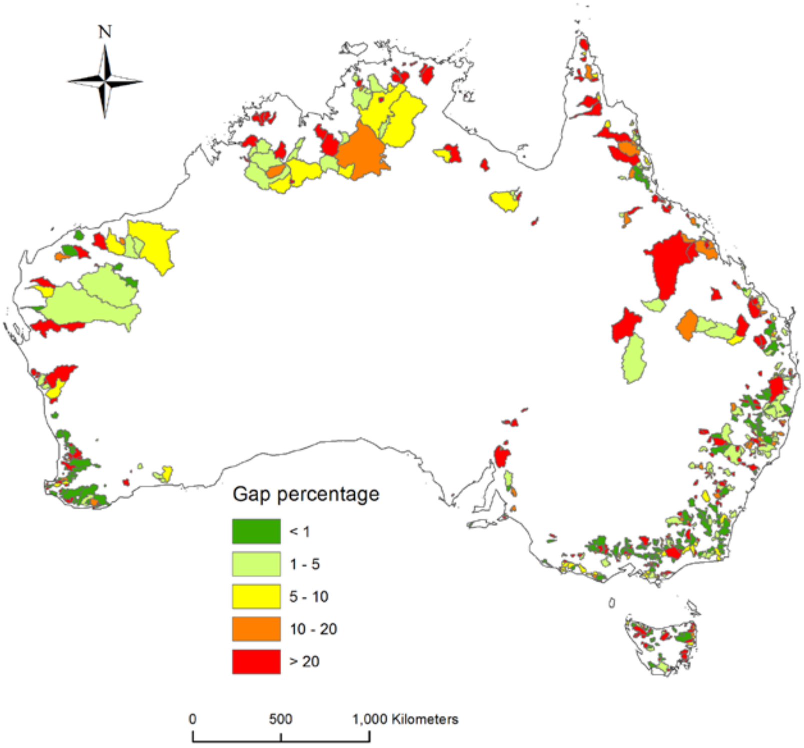


Fig. 1. The 780 unregulated catchments grouped by different streamflow data gaps for the period of 1981-2010.

Table 1. Major catchment attributes for the 217 catchments

Attribute	Definition	Unit	Min	2.5th	25th	Median	75th	97.5th	Max
Area	Catchment area	km^2	53	70	180	392	844	4562	72902
Elevation	Catchment average elevation above sea level	m	46	100	278	449	753	1194	1351
Slope	Catchment mean slope	Degrees	0.3	0.6	2.0	3.9	7.7	12.0	13.6
P	Mean annual precipitation	mm/year	256	371	703	853	1107	1966	2473
ET_p	Mean annual potential evapotranspiration	mm/year	906	968	1149	1235	1408	1791	1892
AI	Aridity index	-	0.38	0.55	1.11	1.44	1.89	4.75	6.47
Forest ratio	Ratio of forest to all land cover types	-	0.02	0.06	0.39	0.55	0.67	0.83	0.90

Out of the 217 catchments, about half of the catchments showed a significant decreasing trend, 37% showing non-significant decreasing trend, and 13% showing non-significant increasing trend (Figure 2), detected using Mann-Kendall trend analysis (see 2.3). This is because Australia experienced the Millennium drought over the period 2001-2009, which caused a dramatic streamflow reduction in this period（van Dijk et al., 2013). Trend analysis for the 217 catchments is explained in Section 2.3 and trend results are summarised in Section 3.

Out of the 217 catchments, about 46% of catchments have no missing data in 1981-2010, 12% with the missing rate <0.1%, 22% with the missing rate 0.1-0.5% and 20% with the missing rate of 0.5-1% (Figure 2).

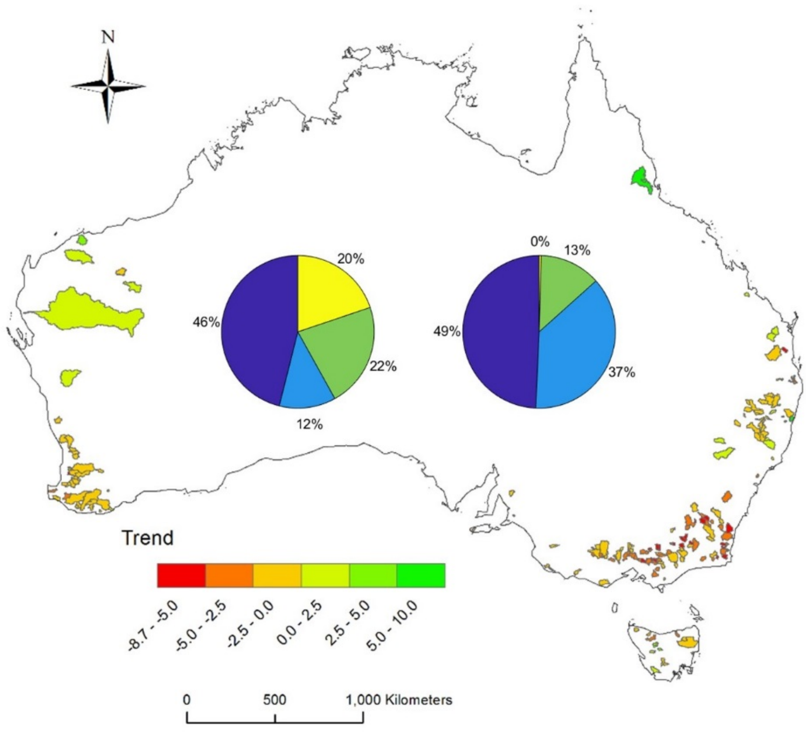


Fig. 2. Trends and streamflow data summary for the 217 catchments used in this study. Trend in annual streamflow is with a unit of mm/year/year. Left pie indicates the catchment percentage with different missing rates (dark blue with missing rate of 0%, navy blue with missing rate of 0-0.1%, green with missing rate of 0.1-0.5%, yellow with missing rate of 0.5-1.0%); right pie indicates the catchment percentage with different trends (dark blue with significant ($p \leq 0.05$) decreasing trend, navy blue with non-significant ($p > 0.05$) decreasing trend, green with non-significant ($p > 0.05$) increasing trend, and yellow with significant ($p \leq 0.05$) increasing trend).

To drive the two hydrological models, we obtained daily meteorological time series (including minimum temperature, maximum temperature, incoming solar radiation, actual vapour pressure and precipitation) from 1975 to 2012 at 0.05° (~5 km) grid resolution from the SILO Data Drill of the Queensland Department of Natural Resources and Water (www.nrw.gov.au/silo). The data quality is reasonably good, indicated by the mean absolute error for maximum daily air temperature, minimum daily air temperature, vapour pressure, and precipitation at 1.0°C, 1.4°C, 0.15 kPa and 0.40 mm/day（Jeffrey et al., 2001).

2.2 Gap-filling experiments

For thoroughly investigating the potential impacts of infilled streamflow data on annual trend accuracy, we conducted three groups of experiments to test how the missing rates at 5%, 10% and 20% impact on streamflow trends. We followed three steps for each missing rate of experiments:

1. Missing patterns were obtained using actual streamflow data. We selected consecutive missing day pattern from actual data from the 780 catchments. For 5% group of missing rate experiments, we selected 44 catchments with missing rates in 4-6%; for 10% group of missing rate experiments, we selected 39 catchment with missing rate in 8-12%; for 20% group of missing rate experiments, we selected 22 catchments with missing rate in 18-22%. Figure 3 shows the probability distribution of consecutive missing days from each group of catchments, which is skewed toward the low end. We therefore used the two-parameter Gamma distribution to simulate probability distribution of consecutive missing days (Figure 3). The Gamma distribution is expressed as

$$X \sim \Gamma\left(k,\theta\right) = Gamma(k,\theta), \tag{1}$$

where X is the consecutive missing days number, k is shape parameter, and θ is scale parameter. The corresponding probability density function in the shape-scale parameterization is

$$f\left(x;k,\theta\right) = \frac{1}{\Gamma\left(k\right)\theta^{k}} x^{k-1} e^{-\frac{x}{\theta}}, \tag{2}$$

where $\Gamma\left(k\right)$ is the gamma function.

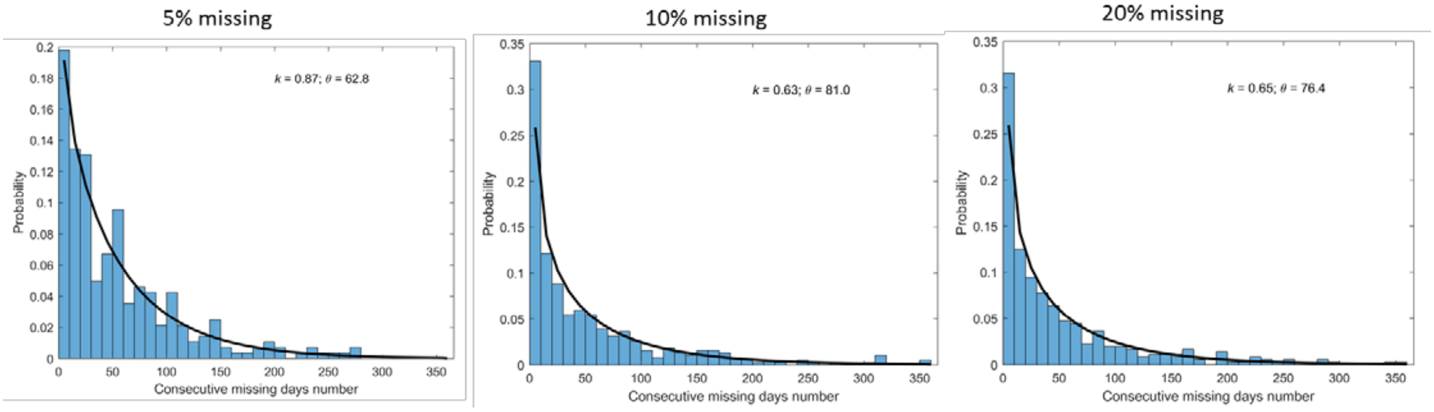


Fig. 3. Missing patterns for three groups of catchments with missing rates 4-6%, 8-12%, 18-22% that represent 5%, 10% and 20% missing rates, respectively.

As seen from Figure 3, the two parameters are stable under the three groups of catchments. The k parameter varies from 0.63 to 0.87 and the θ parameter changes from 62 to 81. It is noted that we removed all times when the number of consecutive missing days was > 365. We did that for a number of reasons. Firstly, gap-filling an entire year of missing data would likely impact annual trends. Secondly, the focus of this paper is on gap-filling short periods of missing data to be able to include more catchments in streamflow analyses. Thirdly, removing all periods of greater than 365 days allowed us to better fit a gamma distribution to the number of missing days.

We also checked the seasonality of missing data to see if one season were more likely to have missing data than another. As seen from Figure 4, the missing data are more or less evenly distributed through different seasons across all the 39 catchments (with missing rate of 8% to 12%) within the 10% missing data group. This indicates that the data gaps were not skewed toward a particular season and it occurred randomly through the year.

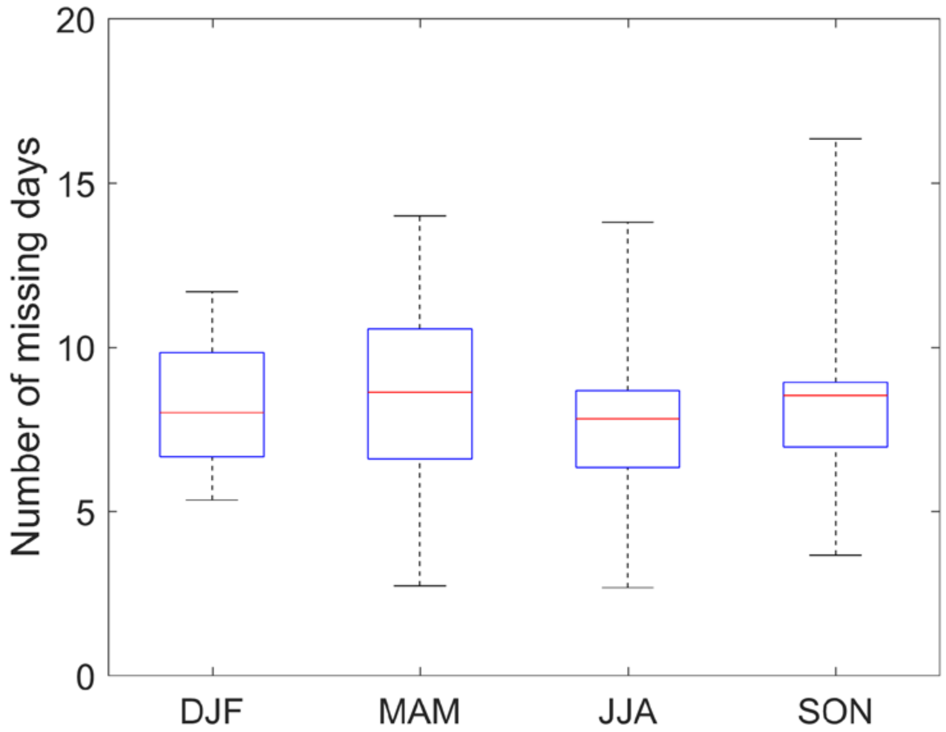


Fig. 4. Distribution of number of missing days across different seasons, summarised from 39 catchments with a missing rate ranging from 8% to 12% (i.e. 10% missing data group).

2. *Generating random consecutive missing day numbers using random number generator (sampling without replacement) based on the Gamma distribution*. The random number generator was repeated 100 times to ensure the selected samples cover a wide range of streamflow time series.

3. *Gap-filling streamflow data*. The selected days were treated as 'missing' data and the unselected data were used for hydrological model calibration. The 'missing' data were then gap-filled using the simulated streamflow from the calibrated GR4J and SIMHYD models, respectively.

For consistent interpretation thereafter, the benchmark streamflow data is regarded as 'observed' and the experiment ones as 'filled' ones. For each of the three experiments, there are 100 x 217 (21,700) ‘missing’ time series, with 100 representing sample times using the random number generator and 217 representing the number of catchments.

2.3 Trend analysis

We used the Mann–Kendall Tau-b non-parametric test including Sen’s slope method（Burn and Elnur, 2002) for annual streamflow trend analysis and significance testing for all the three groups of experiments and benchmark data.

We used the following equation to quantify the trend bias*:*

$$B_t = T_{filled} - T_{obs}, \tag{3}$$

where B_t is the bias in annual streamflow trend (mm/year/year), T_{filled} is annual trend for gap-filled streamflow (mm/year/year), T_{obs} is annual trend in observed streamflow (mm/year/year). It measures the trend error between the infilled and observed runoff trends with $B_t \approx 0$, which indicates that the trend in observed annual runoff is almost the same as that in the infilled annual runoff.

We also defined relative trend bias (P_{Bt}) as

$$P_{B_t} = \frac{T_{filled} - T_{obs}}{T_{obs}} \times 100 \text{ ,} \tag{4}$$

2.4 Hydrological models

Two widely used hydrological models SIMHYD and GR4J (Chiew et al., 2002; Chiew et al., 2010; Li et al., 2014; Oudin et al., 2008; Perrin et al., 2003; Zhang and Chiew, 2009; Zhang et al., 2016a) were used to infill daily ‘missing’ streamflow. Both models require daily precipitation and daily potential evaporation (Priestley and Taylor, 1972) as model inputs, and model outputs are daily streamflow at each gauge. The daily inputs of the maximum and minimum temperatures, incoming solar radiation, and vapour pressure data were used to calculate the Priestley–Taylor daily potential evaporation.

The two models were calibrated using a global optimiser: genetic algorithm (The MathWorks, 2006) at each catchment, with the first six years (i.e., 1975–1980) for spin up and remainder (1981 to 2010) for modelling experiments. Since this study mainly evaluates the trends obtained using the gap-filled streamflow from hydrological modelling, it is crucial to predict high flow and mean flow as accurate as possible. To this end, the model calibration was to minimize the following objective function (F) (Viney et al., 2009; Zhang et al., 2016b):

$$F = (1 - NSE) + 5\left|\ln(1 + B)\right|^{2.5}, \tag{5}$$

$$B = \frac{\sum_{i=1}^{N} Q_{sim,i} - \sum_{i=1}^{N} Q_{obs,i}}{\sum_{i=1}^{N} Q_{obs,i}}, \tag{6}$$

where NSE is the Nash-Sutcliffe-Efficiency of daily streamflow, B is the model bias, Q_{sim} and Q_{obs} are the simulated and observed daily runoff, i is the ith day, N is the total number of days sampled. The NSE gives higher streamflow more weight, and varies between $-\infty$ to 1 with $NSE > 0.6$ indicating a good agreement (Zhang and Chiew, 2009). The B measures water balance error between the observed and modelled daily streamflow, with $B = 0$ indicating that the average of modelled daily streamflow is the same as the average of observed daily streamflow.

For each catchment, GR4J and SIMHYD were calibrated using benchmark data and 100 time series of streamflow data with ‘missing’ data (see Section 2.2), respectively. For benchmark data without any missing data (46% catchments) there are no gap-filling required; for the benchmark data with missing rate less than 1%, the calibrated continuous streamflow data were used to fill the gaps. For the ‘missing’ experiments, the calibrated continuous streamflow data for each 'missing' replicate were used to infill the artificially-made 'missing' data. Table 2 summarises the model calibrations carried out for benchmark and each experiment. Finally, there were ~~1,302,434~~130,634 model calibrations and ~~1,302,000~~130,200 times of gap-filling carried out. Finally, the trends estimated from benchmark were used to evaluate those obtained from the 'missing' experiments.

Table 2. Summary of model calibration number carried out for benchmark and data 'missing' experiments

Model	Benchmark	5% missing	10% missing	20% missing	Sum
GR4J	217	~~217,000~~21,700	~~217,000~~21,700	~~217,000~~21,700	~~651,217~~65,317
SIMHYD	217	~~217,000~~21,700	~~217,000~~21,700	~~217,000~~21,700	~~651,217~~65,317
Sum	434	~~434,000~~43,400	~~434,000~~43,400	~~434,000~~43,400	~~1,302,434~~130,634

3 Results

The gap-filled data from the two hydrological models were evaluated against the benchmark data. Overall, the two models perform well and neither significantly outperforms the other (Figure 5). For the three groups of gap-filling experiments, these two models perform similarly (i.e. the difference of NSE of daily runoff between two is less than 0.02) in 18-19% catchments; SIMHYD model outperforms GR4J model (NSE difference between two is larger than 0.02) in 30-31% catchments; GR4J model outperforms SIMHYD model in 50-51% catchments.

Figures ~~4~~6 and ~~5~~7 summarise the performance of the gap-filled data for estimating annual trend, annual streamflow, monthly streamflow and daily streamflow, respectively. ~~Overall, the two models perform similarly.~~ The three missing rate experiments (5%, 10%, and 20%) perform almost the same as the benchmark (Figures ~~4~~6 and ~~5~~7). The coefficient of determination (r^2) between the gap-filled trends and observed trends is more than 0.98 for the three experiments and two hydrological models.

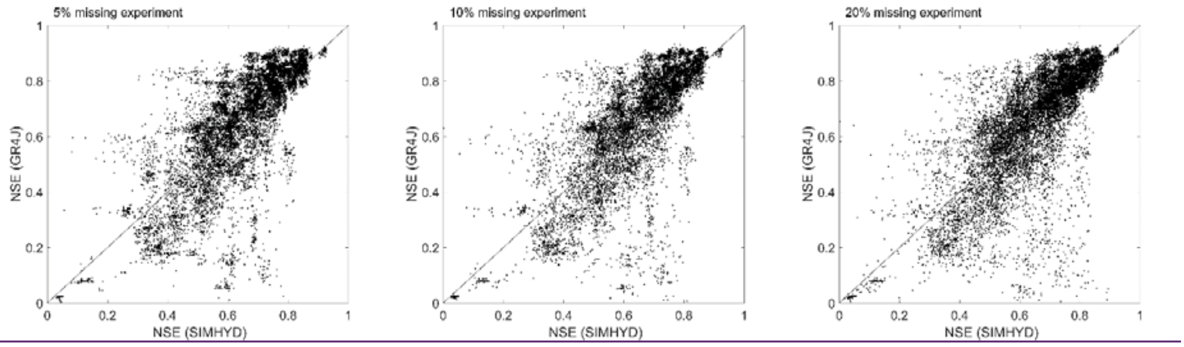


Fig. 5. Comparisons between calibrated GR4J and calibrated SIMHYD for 44 catchments of the 5% missing experiment, 39 catchments of the 10% missing experiment, and 22 catchments of the 20% missing experiment. In each catchment, there were100 replicates carried out.

Since errors in gap-filled trends likely to be different and different time steps when daily infilled streamflow data is used, we further investigate how gap-filled errors are propagated from daily to monthly and to annual scales under the three gap-filling cases (5%, 10%, and 20%) (Figures ~~4~~6 and ~~5~~7). It is expected that daily gap-filled streamflow has a larger standard deviation from the benchmark than monthly and annual streamflow since the streamflow was gap-filled at daily scale. This indicates that the temporal aggregation smooths the gap-filled error strongly, and it generates very reasonable monthly and annual streamflow estimates with less standard deviation. It is interesting to note that both models tend to underestimate very high flows though they are calibrated against the NSE of daily streamflow which puts a larger weight on correctly representing higher flows.

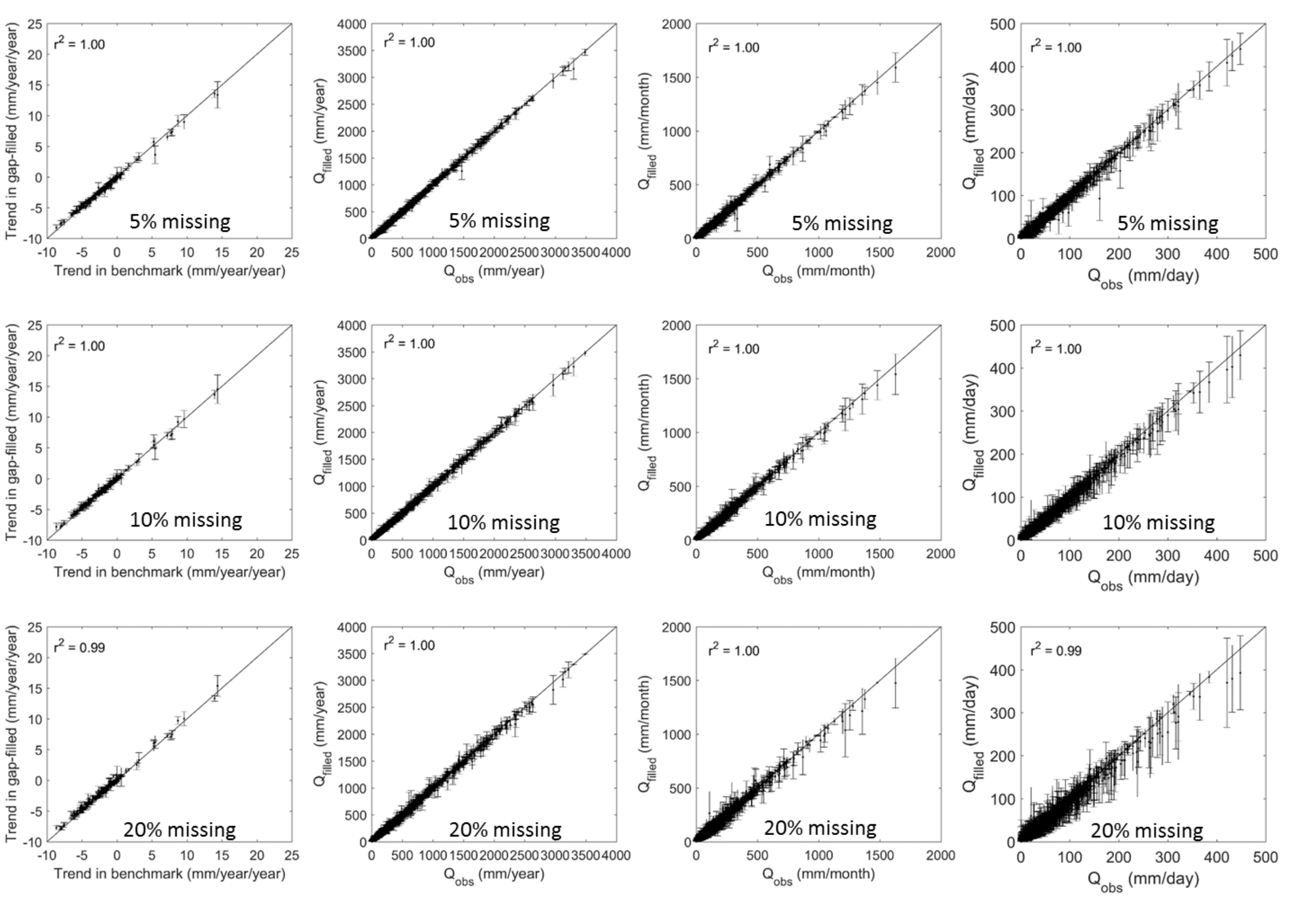


Fig. ~~4~~6. Comparisons between the observed streamflow (x-axis) and gap-filled ones (y-axis) for streamflow trend (mm/year/year, left panels), annual streamflow (mm/year, second left panels), monthly streamflow (mm/month, second right panels) and daily streamflow (mm/day, right panels). The gaps were filled using GR4J. Error bar represents standard deviation of the 100 replicates for each group of ‘missing’ experiments.

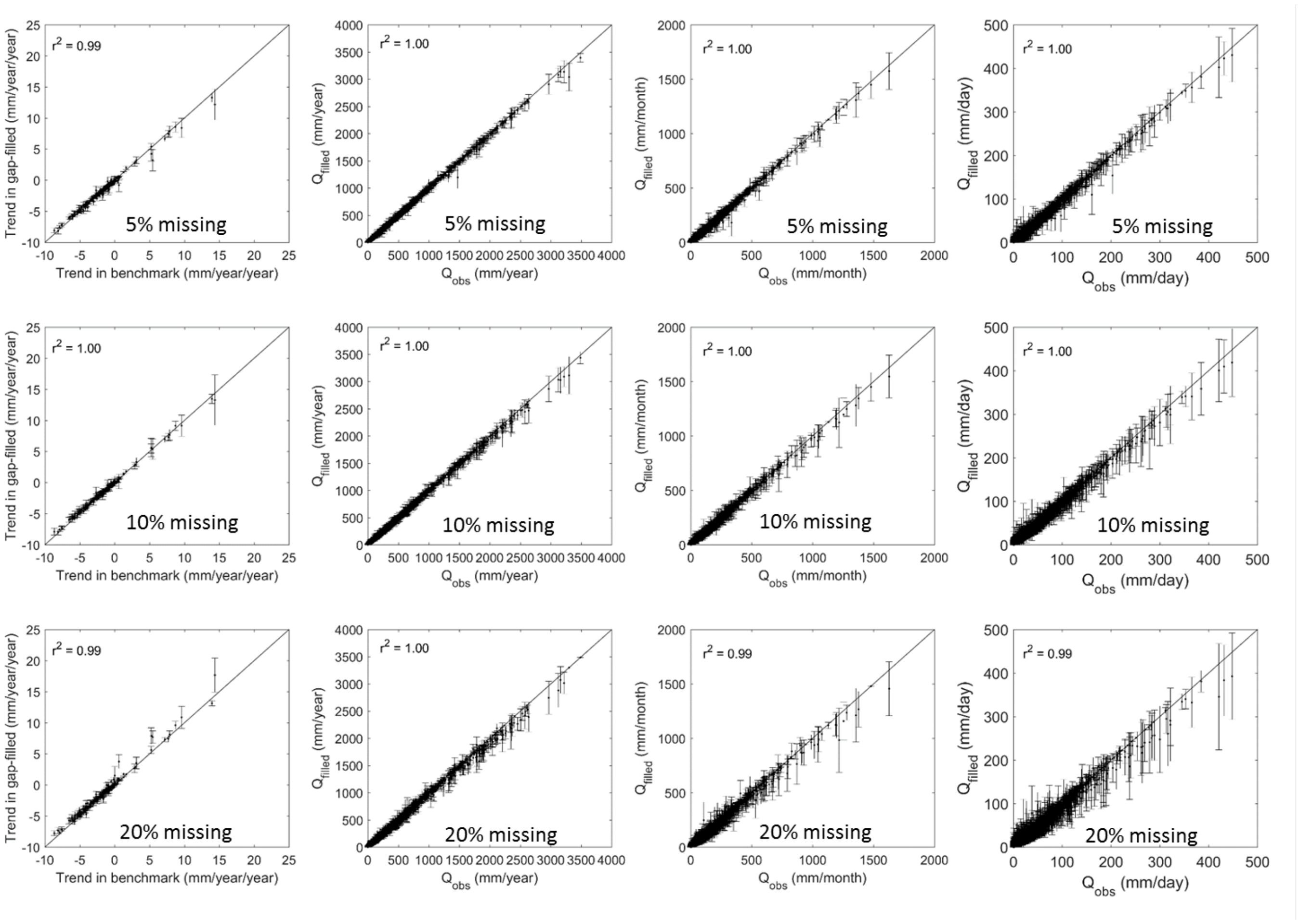


Fig. ~~5~~7. Same as Fig. ~~4~~6 but using SIMHYD.

Figure ~~6~~8 further summarises the catchments with trend direction mismatch between the benchmark and gap-filled data (i.e. change from negative to positive or change from positive to negative). For the experiments with 5% and 10% missing rates and for GR4J, there are less than 8 out of the 217 catchments showing a trend mismatch and almost all of them show non-significant trends ($p > 0.05$). For the experiments with a 20% missing rate for GR4J, there are less than 10 out of the 217 catchments showing trend mismatch and all of them show non-significant trends. SIMHYD results are almost the same as GR4J results. All these indicate that there is very marginal influence on annual streamflow trend directions when the missing rate is less than 20%.

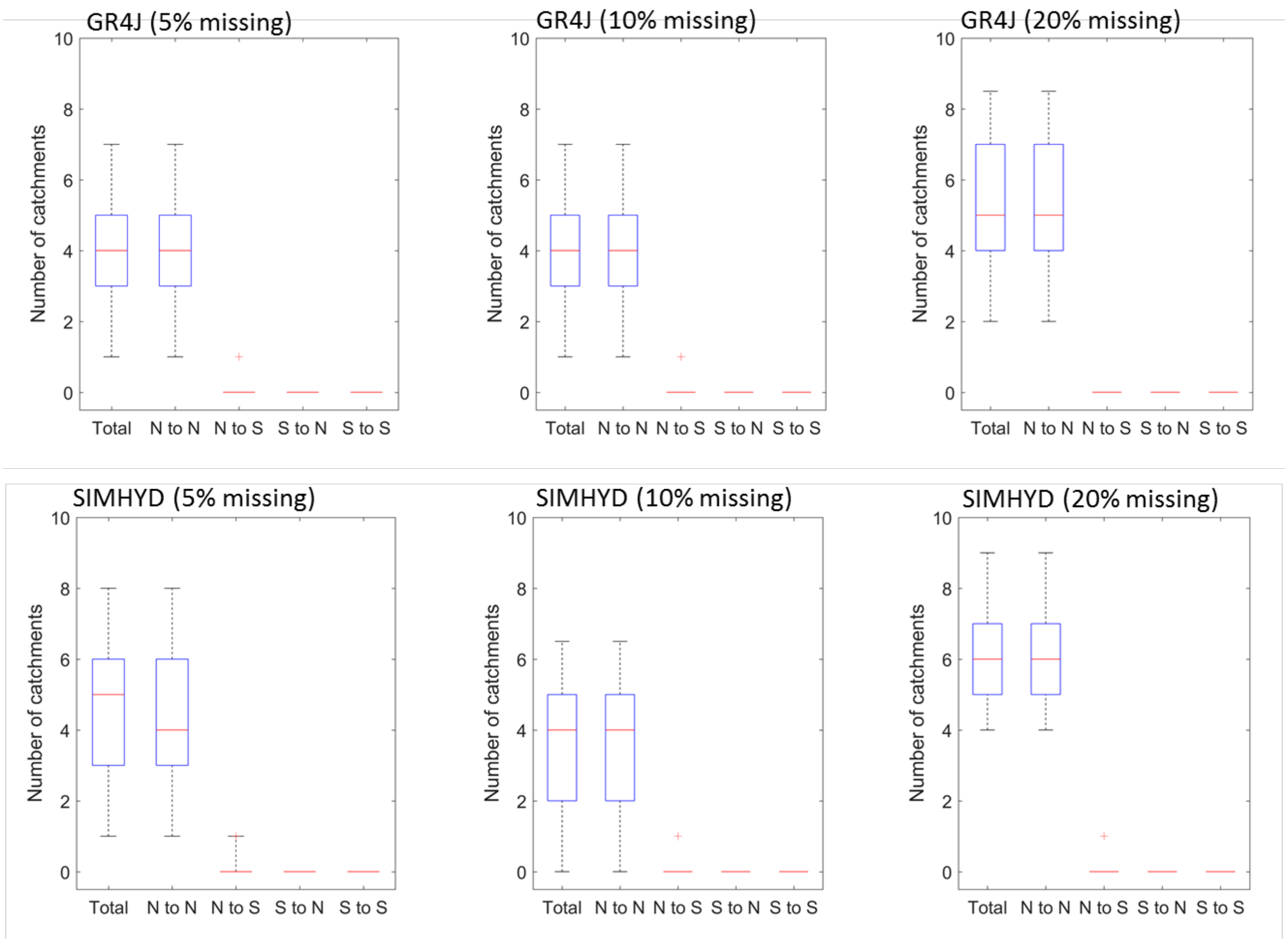


Fig. ~~6~~8. Trend mismatch analysis between the gap-filled and benchmark. Total means all mismatch catchments; 'N' means not significant trends ($p > 0.05$); 'S' means significant trends ($p \leq 0.05$). The bottom, middle and top of each box are the 25th, 50th and 75th percentiles, and the bottom and top whiskers are the 5th and 95th percentiles.

Though the three groups of experiments show small trend direction changes (Figure ~~6~~8), it is not clear how the trend bias (Eq. 3) looks. To this end, Figure ~~7~~9 further compares the trend bias between the experiments. It is clear that the trend biases between 5% and 10% missing experiments are similar. For GR4J, both have the trend bias varying from -1 to 1 mm/year/year; For SIMHYD, the trend bias between the two is similar when it varies from -0.5 to 1 mm/year/year, and the trend bias for 5% missing experiment is even larger than that for 10% missing experiment. The trend bias for 20% missing experiment is noticeably larger than that for 10% and 5% missing experiments for both models, and the underperformance is more noticeable from SIMHYD gap-filled than that from GR4J gap-filled. This result suggests that the trend bias is reasonable when the missing rate is less than 10%, and can be large for small number of catchments when the missing rate is to 20%.

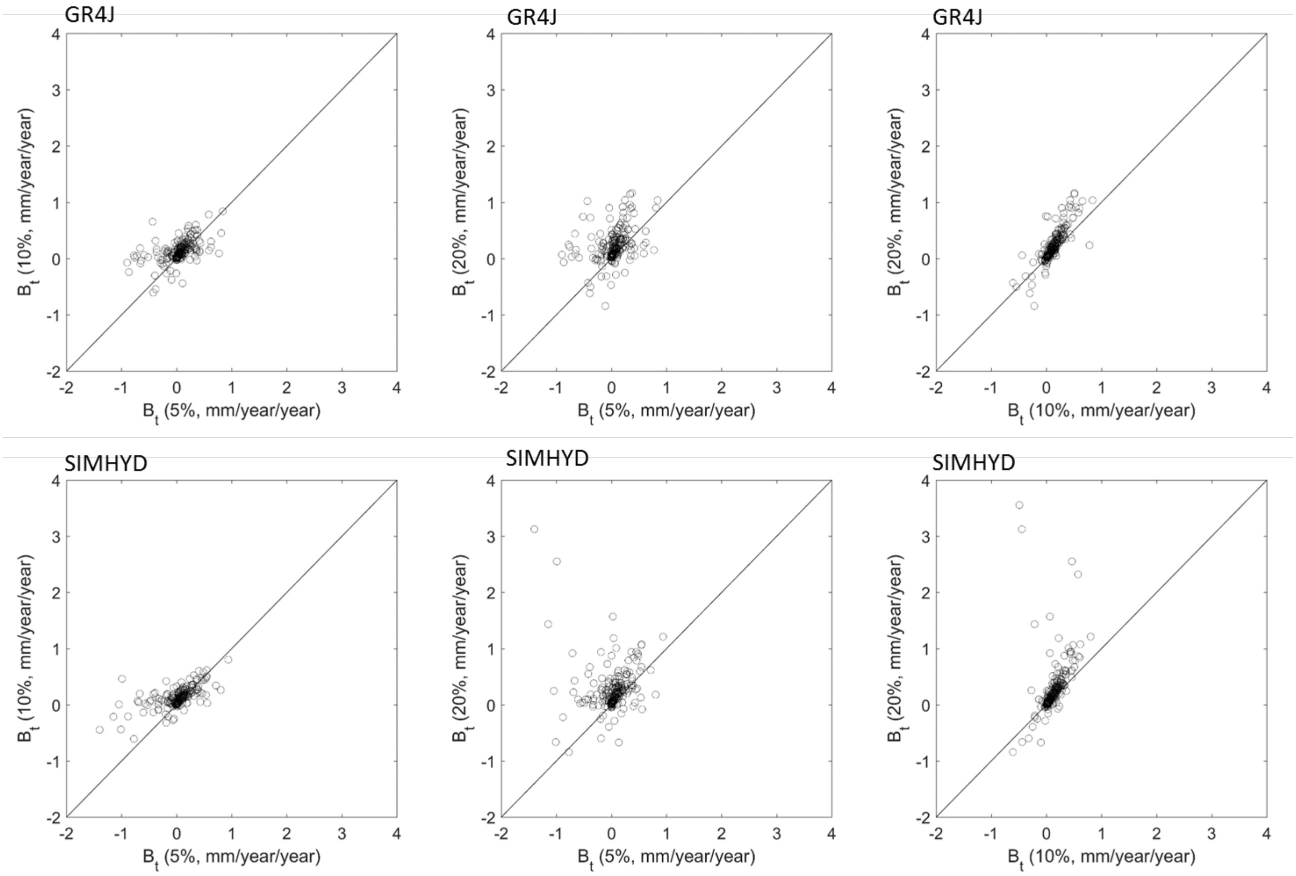


Fig. ~~7~~9. Trend biases comparison between the three groups of gap-filling experiments (5%, 10% and 20%). Top three are for GR4J and bottom three are for SIMHYD.

4 Discussion and conclusions

Researchers are keen to have a comprehensive understanding of rules for excluding catchments with gaps in the streamflow record. Our results indicate that when the streamflow data gaps are up to 10%, the gap-filled data obtained using hydrological modelling are very reasonable for annual trend analysis and annual streamflow estimates. Choosing the threshold of 10% missing rate will allow the use of many more catchments in modelling and data analysis studies. For example, of the 780 unregulated Australian catchments available for modelling studies (Zhang et al., 2013), there are 237 catchments with the missing rate of 1-10% during 1981-2010, accounting for 38% of total available catchments (Figure 1). Of these 237, 67 (~28%) also have gaps lasting more than one year (which we did not consider in this analysis), and therefore these may not be suitable for use. With an increased number of catchments, more reliable large-scale hydrological modelling studies can be carried out (Beck et al., 2016; Parajka et al., 2013; Zhang et al., 2016a).

The 'missing' rate experiments designed in this study are based on the actual data missing patterns obtained from the 780 catchments. In most cases, the consecutive missing days are less than 10, as indicated by Figure 3, indicating brief periods of gauge malfunctions. It is however interesting to note that there are streamflow gaps lasting much longer than this in many catchments, with gaps of many months in some cases, noting that we excluded gaps lasting one year or more. It is highly likely that filling a gap of one year or more will result in biases larger than those presented here.

Furthermore, we also tested the quality of random gap-filled daily streamflow. In that case, the missing patterns were randomly selected using a random number generator. The results obtained from the random gap-filling (not shown) are similar to the results presented here. Thus, it is likely that the length of the gaps (as long as it is less than one year) is unlikely to impact the results of the gap-filling experiment. We would conclude from this that the use of hydrologic modelling for filling the substantially gapped data (up to 10% missing rate) described here for Australia will not impact annual trends of streamflow. Impacts on other streamflow characteristics also need to be examined, as well as seeing if the results obtained in Australia are comparable with those in other parts of the world, where the length of observational gaps may be quite different to those shown in Figure 3.

It is possible that data gaps may only exist during high flow or low flow conditions, although that is not what we observed here with the majority of missing data being more or less evenly distributed throughout the year (Figure 4). We did however test the impact of filling streamflow data in high flow or low flow conditions (results not shown here). In those cases, the missing patterns were selected using only high flow (>95th percentile) or low flow (less than 50th percentile) data. The results obtained from the low flow gap-filling indicates that there is only a negligible influence on annual streamflow trend estimates when the missing rate is less than 50%. In contrast, the high flow gap-filled data shows a noticeable change in annual streamflow trend when the missing rate is 5% This is understandable since high flow is usually several orders of magnitude higher than low flow, and errors in filling high flow could have large impacts on annual flow and its trends (Slater and Villarini, 2017).

To understand if the quality of gap-filled streamflow is related to catchment attributes and calibration accuracy, we conducted further analysis among the trend bias, model calibration efficiency (i.e. *NSE*) and catchment aridity index (mean annual potential evaporation divided by mean annual precipitation) (Figure ~~8~~10). The model calibration results at dry catchments are normally poorer than those at wet catchments. However, the trend bias (mm/year/year) obtained from dry catchments is usually smaller. The large biases are observed from the catchments with aridity index less than 2 and with the calibrated NSE being larger than 0.60. In part, this is to be expected since the streamflow is also lower in more arid catchments, meaning that the trend bias is also likely to be lower.

Figure ~~9~~11 shows the relationship between relative trend bias (%, Eq. 4) and aridity index. It shows that not only is the actual trend bias lower in drier catchments, but so too is the relative (%) trend. This result suggests that the large bias in annual trends as a result of gap-filling is observed in relatively wet catchments where model calibrations are reasonably good. This result seems counter-intuitive and requires further exploration, which is beyond the scope of the current paper.

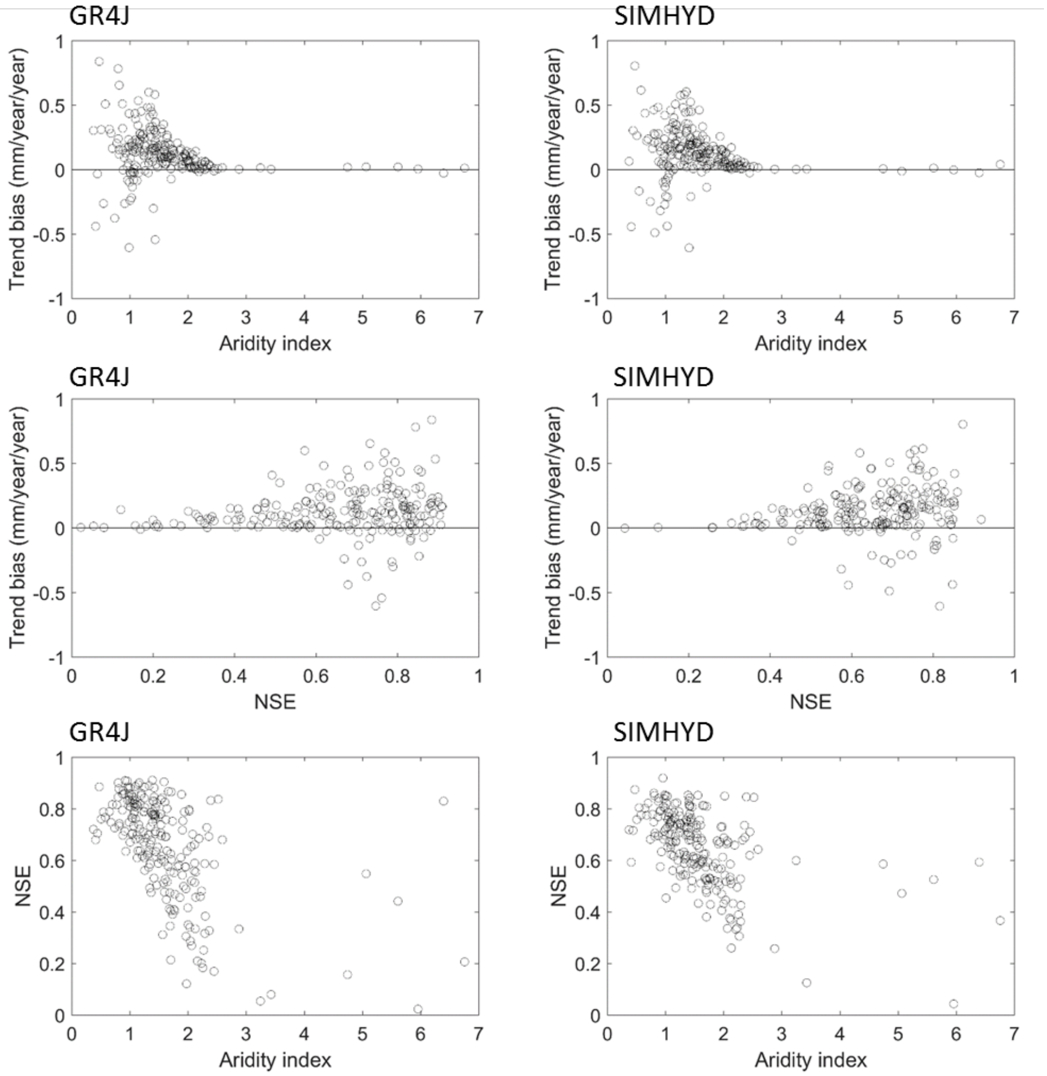


Fig. ~~8~~10. Relationships among trend bias (mm/year/year), model calibration Nash-Sutcliffe Efficiency and aridity index for each catchment and for the experiment of 10% missing rate.

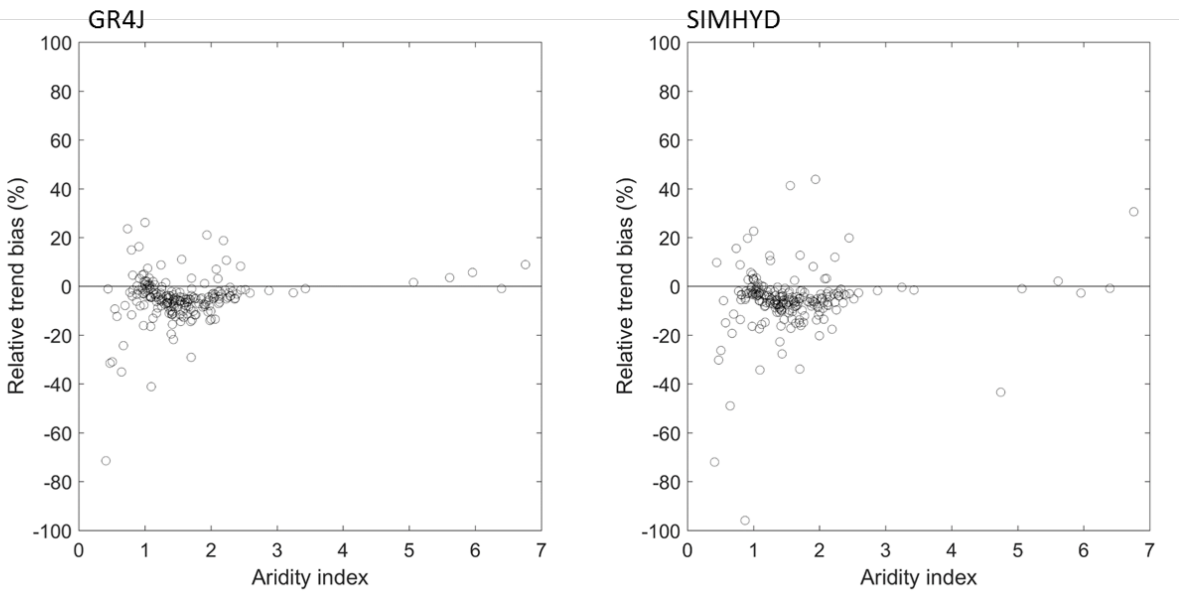


Fig. ~~9~~11. Relationships between relative trend bias (mm/year/year) and aridity index for each catchment and for the experiment of 10% missing rate.

This study focuses on evaluating annual streamflow and its trends. Therefore, we used the Nash-Sutcliffe Efficiency plus model bias (Eqs. 5 and 6) to calibrate the two hydrological models. If other hydrological response variables such as low flow metrics are required, other model calibration schemes should be used since the NSE model calibration scheme gives more weight to reproducing high flows at the expense of low-flows (Zhang et al., 2014). Low flow metrics have important ecological implications (Mackay et al., 2014; Smakhtin, 2001). In general however, it is challenging to use hydrological modelling for low flow simulations and predictions (Pushpalatha et al., 2012; Staudinger et al., 2011). To have credible low flow gap-filling, model calibrations should use an objective function that puts more weights on low flows, such as NSE of daily inverse streamflow and the direct low flow metrics. Another possible method is to combine hydrological modelling with other methods for gap-filling, such as using nearby gauges (Lopes et al., 2016) and statistical methods (Gedney et al., 2006b).

It is noted that the infilled data purely refers to the 'missing' data. All streamflow gauges are only rated to a certain flow. Once the flow exceeds that level during flooding, the results are interpolated using stage-discharge relationships（Peña-Arancibia et al., 2015). These interpolations could be a major source of observation error. However, investigating high flow interpolation and data quality is beyond the scope of this study.

The modelling experiments and findings from this study could have important implications for other parts of the world as well as Australia. First, to develop appropriate gap-filling modelling experiments, it is necessary to evaluate the distribution of consecutive missing data pattern. The probability distribution of consecutive missing data is skewed toward the low end, which can be nicely simulated using the Gamma distribution (Eq.1). This distribution should be very useful for similar missing patterns in other regions. Second, hydrological modelling is a very good tool for filling gaps since it can fully take the advantage of climate forcing and non-gap streamflow data, and obtain the best possible daily simulations. Third, the threshold of 10% identified in this study should be applicable to regions/catchments with similar missing patterns. However, if the data gaps continue for seasons or years, the threshold may not hold.

It would also be interesting to compare hydrological modelling to other approaches for filling streamflow data gaps. Hydrological modelling is a most useful method used in Australia for predicting daily streamflow in ungauged catchments (Chiew et al., 2009; Li and Zhang, 2017; Zhang and Chiew, 2009; Viney et al., 2009). It has been used operationally by the Australian Bureau of Meteorology for filling daily streamflow data gap for many years. In the future, this operational method could further be comprehensively evaluated against other approaches, such as interpolation or correlations with nearby gauging sites.

In summary, our results clearly demonstrate that the gap-filled data is most accurate when examining trends at the annual scale, followed by monthly scale, and with least satisfaction at the daily scale. This gives researchers confidence for annual trend analysis, a hot topic in hydrological and climate sciences. Our results also clearly indicate that the gap-filling of Australian streamflow data using hydrological model is very reasonable when the missing rate is less than 10%, with only a small number of catchments showing a large trend bias when the missing rate is to 20%. The results also indicate that gap-filling drier catchments appears to be more successful than gap-filling wetter catchments.

Acknowledgements. This study was supported by the CSIRO strategic project "Next generation methods and capability for multi-scale cumulative impact assessment". The authors thank constructive suggestions from Editor Louse Slater and two reviewers: Maxine Zaidman and Juraj Parajka. Data used for gap-filling experiments are freely available upon request from the corresponding author (Yongqiang Zhang, email: yongqiang.zhang2014@gmail.com). The authors declare no conflict of interests.

References

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resources Research, 52, 3599-3622, 10.1002/2015wr018247, 2016.

Burn, D. H., and Elnur, M. A. H.: Detection of hydrologic trends and variability, Journal of Hydrology, 255, 107-122, 2002.

Chiew, F. H. S., Kirono, D. G. C., Kent, D. M., Frost, A. J., Charles, S. P., Timbal, B., Nguyen, K. C., and Fu, G.: Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates, Journal of Hydrology, 387, 10-23, 10.1016/j.jhydrol.2010.03.025, 2010.

Chiew, F. H. S., Peel, M. C., and Western, A. W.: Application and testing of the simple rainfall-runoff model SIMHYD, in: Mathematical Models of Small Watershed Hydrology and Applications, edited by: Singh, V. P., and Frevert, D. K., Water resources Publication, Littleton, Colorado, USA, 335-367, 2002.

Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C., and Viney N. R.: Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method, Water Resources Research. 45, W10414, doi:10.1029/2008WR007338, 2009.

Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D.: Changes in Continental Freshwater Discharge from 1948 to 2004, Journal of Climate, 22, 2773-2792, 10.1175/2008jcli2592.1, 2009.

Déry, S. J., Hernández-Henríquez, M. A., Burford, J. E., and Wood, E. F.: Observational evidence of an intensifying hydrological cycle in northern Canada, Geophysical Research Letters, 36, 1-5, 10.1029/2009gl038852, 2009.

Gedney, N., Cox, P. M., Betts, R. A., Boucher, O., Huntingford, C., and Stott, P. A.: Detection of a direct carbon dioxide effect in continental river runoff records, Nature, 439, 835-838, 10.1038/nature04504, 2006a.

Gedney, N., Cox, P. M., Betts, R. A., Boucher, O., Huntingford, C., and Stott, P. A.: Continental runoff - A quality-controlled global runoff data set - Reply, Nature, 444, E14-E15, 10.1038/nature05481, 2006b.

Hannaford, J., and Buys, G.: Trends in seasonal river flow regimes in the UK, Journal of Hydrology, 475, 158-174, 10.1016/j.jhydrol.2012.09.044, 2012.

Harvey, C. L., Dixon, H., and Hannaford, J.: An appraisal of the performance of data-infilling methods for application to daily mean river flow records in the UK, Hydrology Research, 43, 618-636, 10.2166/nh.2012.110, 2012.

Jeffrey, S. J., Carter, J. O., Moodie, K. B., and Beswick, A. R.: Using spatial interpolation to construct a comprehensive archive of Australian climate data, Environmental Modelling & Software, 16, 309-330, 2001.

Lavers, D., Prudhomme, C., and Hannah, D. M.: Large-scale climate, precipitation and British river flows Identifying hydroclimatological connections and dynamics, Journal of Hydrology, 395, 242-255, 10.1016/j.jhydrol.2010.10.036, 2010.

Li, F., Zhang, Y., Xu, Z., Liu, C., Zhou, Y., and Liu, W.: Runoff predictions in ungauged catchments in southeast Tibetan Plateau, Journal of Hydrology, 511, 28-38, 10.1016/j.jhydro1.2014.01.014, 2014.

Li, H., and Zhang, Y.: Regionalising rainfall-runoff modelling for predicting daily runoff: Comparing gridded spatial proximity and gridded integrated similarity approaches against their lumped counterparts, Journal of Hydrology, 550, 279-293, 10.1016/j.jhydrol.2017.05.015, 2017.

Liu, J., and Zhang, Y.: Multi-temporal clustering of continental floods and associated atmospheric circulations, Journal of Hydrology, 555, 744-759, 10.1016/j.jhydrol.2017.10.072, 2017.

Lopes, A. V., Chiang, J. C. H., Thompson, S. A., and Dracup, J. A.: Trend and uncertainty in spatial-temporal patterns of hydrological droughts in the Amazon basin, Geophysical Research Letters, 43, 3307-3316, 10.1002/2016gl067738, 2016.

Mackay, S. J., Arthington, A. H., and James, C. S.: Classification and comparison of natural and altered flow regimes to support an Australian trial of the Ecological Limits of Hydrologic Alteration framework, Ecohydrology, 7, 1485-1507, 10.1002/eco.1473, 2014.

Morton, F. I.: Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology, Journal of Hydrology, 66, 1-76, https://doi.org/10.1016/0022-1694(83)90177-4, 1983.

Oudin, L., Andreassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, Water Resources Research, 44, W03413, doi:10.1029/2007WR006240, 10.1029/2007wr006240, 2008.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, Hydrology and Earth System Sciences, 17, 1783-1795, 10.5194/hess-17-1783-2013, 2013.

Peña-Arancibia, J. L., Zhang, Y., Pagendam, D. E., Viney, N. R., Lerat, J., van Dijk, A. I. J. M., Vaze, J., and Frost, A. J.: Streamflow rating uncertainty: Characterisation and impacts on model calibration and performance, Environmental Modelling & Software, 63, 32-44, 10.1016/j.envsoft.2014.09.011, 2015.

Perrin, C., Michel, C., and Andreassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275-289, 10.1016/s0022-1694(03)00225-7, 2003.

Petrone, K. C., Hughes, J. D., Van Niel, T. G., and Silberstein, R. P.: Streamflow decline in southwestern Australia, 1950-2008, Geophysical Research Letters, 37, 1-7, 10.1029/2010gl043102, 2010.

Priestley, C.H.B., Taylor, R.J.: On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100, 81–92, 1972.

Pushpalatha, R., Perrin, C., Le Moine, N., and Andreassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420, 171-182, 10.1016/j.jhydrol.2011.11.055, 2012.

Rojas-Serna, C., Lebecher el, L., Perrin, C., Andreassian, V., Oudin, L: How should a rainfall-runoff model beparameterized in an almost ungaugedcatchment? A methodology tested on609 catchments, Water Resour. Res., 52,4765–4784, doi:10.1002/2015WR018549., 2016.

Sanderson, M. G., Wiltshire, A. J., and Betts, R. A.: Projected changes in water availability in the United Kingdom, Water Resources Research, 48, 10.1029/2012wr011881, 2012.

Seibert, J., Beven, K. J.: Gauging the ungauged basin: How many discharge measurements are needed?, Hydrol. Earth Syst. Sci.,13(6), 883-892, 2009.

Slater, L., and Villarini, G.: On the impact of gaps on trend detection in extreme streamflow time series, International Journal of Climatology, 37, 3976-3983, 10.1002/joc.4954, 2017.

Smakhtin, V. U.: Low flow hydrology: a review, Journal of Hydrology, 240, 147-186, 10.1016/s0022-1694(00)00340-1, 2001.

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., and Tallaksen, L. M.: Comparison of hydrological model structures based on recession and low flow simulations, Hydrology and Earth System Sciences, 15, 3447-3459, 10.5194/hess-15-3447-2011, 2011.

Ukkola, A. M., Keenan, T. F., Kelley, D. I., and Prentice, I. C.: Vegetation plays an important role in mediating future water resources, Environmental Research Letters, 11, 10.1088/1748-9326/11/9/094022, 2016.

Ukkola, A. M., Prentice, I. C., Keenan, T. F., van Dijk, A. I. J. M., Viney, N. R., Myneni, Ranga B., and Bi, J.: Reduced streamflow in water-stressed climates consistent with CO2 effects on vegetation, Nature Climate Change, 6, 75-78, 10.1038/nclimate2831, 2015.

van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, Water Resources Research, 49, 1040-1057, 10.1002/wrcr.20123, 2013.

Viney, N. R., Perraud, J., Vaze, J., Chiew, F. H. S., Post, D. A., and Yang, A.: The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, 18th World Imacs Congress and Modsim09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, 3421-3427, 2009.

Yang, Y., McVicar, T. R., Donohue, R. J., Zhang, Y., Roderick, M. L., Chiew, F. H. S., Zhang, L., and Zhang, J.: Lags in hydrologic recovery following an extreme drought: Assessing the roles of climate and catchment characteristics, Water Resources Research, 53, 4821-4837, 10.1002/2017wr020683, 2017.

Zhang, Y. Q., Viney, N. R., Frost, A., Oke, A., Brooks, M., Y., C., and N., C.: Collation of Australian modeller's streamflow dataset for 780 unregulated Australian catchments, CSIRO: Water for a Healthy Country National Research Flagship, 117, 2013.

Zhang, Y., and Chiew, F. H. S.: Relative merits of different methods for runoff predictions in ungauged catchments, Water Resources Research, 45, W07412, doi:10.1029/2008WR007504, 10.1029/2008wr007504, 2009.

Zhang, Y., Pena-Arancibia, J. L., McVicar, T. R., Chiew, F. H., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y. Y., Miralles, D. G., and Pan, M.: Multi-decadal trends in global terrestrial evapotranspiration and its components, Sci Rep, 6, 19124, 10.1038/srep19124, 2016a.

Zhang, Y., Vaze J., Chiew, F. H. S., Teng, J., Li, M.: Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall–runoff modelling, Journal of Hydrology, 517, 936-948, 2014.

Zhang, Y., Zheng, H., Chiew, F. H. S., Pena-Arancibia, J., and Zhou, X.: Evaluating Regional and Global Hydrological Models against Streamflow and Evapotranspiration Measurements, Journal of Hydrometeorology, 17, 995-1010, 10.1175/jhm-d-15-0107.1, 2016b.

Zhang, Y.Q., Viney, N.R., Frost, A., Oke, A., Brooks, M., Y., C., N., C: Collation of Australian modeller's streamflow dataset for 780 unregulated Australian catchments, CSIRO: Water for a Healthy Country National Research Flagship, 2013.