M. Zaidman (Referee)

maxine.zaidman@jbaconsulting.com

Received and published: 9 July 2018

Overall comments:

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 his thoughtful thinking 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 374-383, the text says "*The modelling experiments and findings from this study could have important implications to other parts of the world. First, to develop appropriate gap-filling modelling experiments, it is necessary to evaluate 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 (Equation 1). This distribution is very useful for the similar missing patterns in other regions. Second, hydrological modelling is a very good tool for filling gaps since it can fully take the advantage for 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 the 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 384 to 390, the text now says *"It should 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. In the future, this operational method can further be comprehensively evaluated against other approaches, such as interpolation or correlations with nearby gauging sites".*

Specific comments:
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.

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 93-95):

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 324 to 334, the text now says "Streamflow data gap could only occur in high flow or low flow condition in the extreme case though majority of missing data for the Australian catchments are more or less evenly distributed through the year. We further tested the impact of filling streamflow data in high flow or low flow condition. In that case, 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 of is less than 50%. In contrast, the high flow gap-filled shows a noticeable change in annual streamflow trend when the missing rate is 5% (or at 95th percentile). This is understandable since high flow is usually several orders of magnitude higher than low flow, and a certain error in filling high flow could have large impact on annual flow and its trends".*

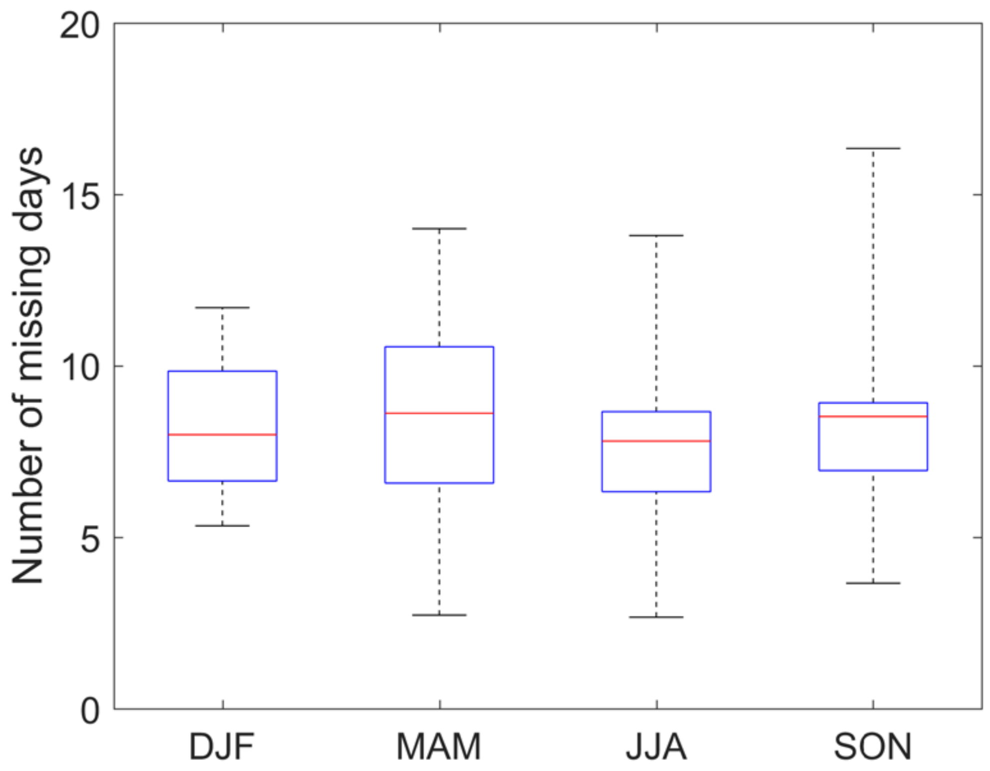

Fig. 4. Distribution of number of missing days across different seasons, summarised from 39 catchments with a missing rate ranging from 8% to 12% (i.e. 10% missing data group).

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%.

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.

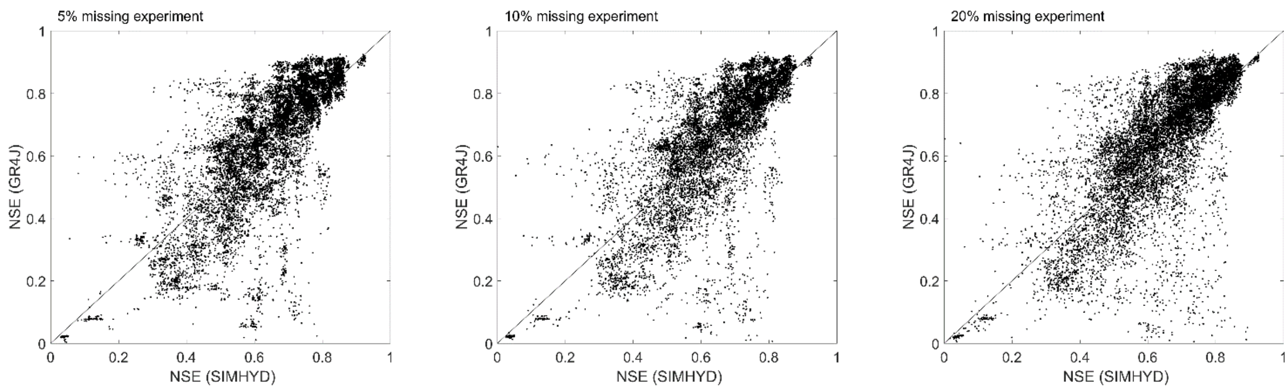

Fig. 5. Comparisons between calibrated GR4J and calibrated SIMHYD for 44 catchments of the 5% missing experiment, 39 catchments of the 10% missing experiment, and 22 catchments of the 20% missing experiment. In each catchment, there were100 replicates carried out.

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).