Impact of infilling method on streamflow derivation

D.C.H. Nguyen [a], **M. Leonard** [a], **S. Westra** [a] **and D. Dutta** [b]

[a] *School of Architecture and Civil Engineering, The University of Adelaide, South Australia*
[b] *NSW Government Department of Planning and Environment*
Email: duc.nguyen@adelaide.edu.au

Abstract: Managing environmental and water resources requires decision-making, with hydrological models playing an important role in translating hydroclimatic inputs into streamflow impacts. To this end, it is essential to consider the integrity of foundational observation series of rainfall and evapotranspiration since the quality of any outputs is strongly dependent on the quality of observations used as inputs.

In Australia, SILO is widely used as an important source for hydrological input timeseries (https://www.longpaddock.qld.gov.au/silo/). In a recent project by the New South Wales Department of Planning and Environment (DPE), 10,000-year time series of stochastic rainfall and evapotranspiration data for a large number of sites have been generated, with input timeseries taken from SILO covering the period 1890 – 2018. Notably, the SILO database is periodically updated, and the online release notes since 2018 state potentially 'large differences' are associated with the removal of erroneous values used within the interpolation. This raises a critical question: To what extent does the interpolation of the historical observation record impact the derivation of streamflow?

This study aims to evaluate the impact of interpolation in terms of streamflow by comparing different interpolations of rainfall and evapotranspiration. The study does not seek to deconstruct or provide an alternative interpolation to the SILO method to identify uncertainties associated with the fundamental interpolation configuration. Instead, this study seeks to identify the accumulated impact of any changes within published SILO data in terms of the magnitude of streamflow. The specific use case is to compare an original set of streamflow outputs from a model calibration in the Lachlan catchment for the period 1890-2018 with a newer dataset spanning 1890-2021 that also has a new underlying interpolation. The reason for the comparison was prompted by a desire to extend the record to account for the potential influence of recent dry years, but with unknown impact of the new interpolation.

Six comparison cases were set up and then applied to 17 sub-catchments of the Lachlan River, having 30 rainfall sites and 13 'Morton Wet' evapotranspiration sites. Cases 1, 2 and 3 respectively identify the impact of the new interpolation algorithm on the two input variables (rainfall and evapotranspiration) and the output flow variable. In all instances the 1890 to 2018 period is used as the baseline so that the effect of the +3 years (2019-2021) does not affect the outcome. Case 4 is a comparison of the impact of the +3 years of data on the flow relative to the impact of the additional data and the change of interpolation method. Cases 5 and 6 change only one of the hydrological inputs to the new interpolation method to determine whether changes to either the rainfall or the evapotranspiration had a bigger impact on streamflow.

The study showed that the new interpolation caused discrepancies in rainfall from –2.5% to +0.7% (Case 1) and –3.5% to +4.1% for evapotranspiration (Case 2) which are modest but nonetheless resulted in streamflow discrepancies from –17.3% to 7.2% (Case 3). The impact of the +3 years of record on flow was minimal, from –0.5% to 1.3% (Case 4) compared to the impact of the interpolation algorithm (Case 3). In addition, the impact of the evapotranspiration on flow is shown to be larger (Case 6) than the rainfall (Case 5). In other words, the change in the interpolated data, especially the evapotranspiration, is more substantial than the influence of +3 drier years at the end of the record. This suggests that care is needed when using interpolated data to appreciate the uncertainty associated with the interpolation and that outputs should be interpreted against a background of uncertainty rather than interpret streamflow estimates to a false level of precision.

Keywords: *SILO, interpolated data, streamflow*

1. INTRODUCTION

Managing environmental and water resources requires decisions to be made, especially accounting for risks to planning, disaster preparedness and infrastructure (Loucks and Van Beek 2017). To achieve this, hydrological models play an important role to translate climatological inputs into streamflow impacts (Refsgaard and Abbott 1996). The quality of any outputs is strongly dependent on the quality of inputs (garbage in garbage out), thus it is essential to consider the integrity of foundational observation series of rainfall and evapotranspiration.

One of the important input sources for Australia is the SILO (Scientific Information for Land Owners) climate dataset (https://www.longpaddock.qld.gov.au/silo/). SILO is the Queensland Government's free database that provides continuous daily climate data for Australia from 1889 to present (Queenland Goverment 2023). This database is widely used by researchers, government agencies, and other organizations to study the impacts of climate change, support sustainable land use and management, and inform decision-making related to the environment (Jeffrey et al. 2001).

The SILO database has recently been utilized in a project by the New South Wales Department of Planning and Environment (DPE) in developing its Regional Water Strategies (Leonard et al. 2019). This project aims to assess the risks associated with climate variability and change to address water management issues in the state. As a part of this project, 10,000-year time series of stochastic rainfall and evapotranspiration data for a large number of sites were generated by the rainfall model of Bennett et al. (2018) and the evaporation model of Leonard et al. (2019). These simulations were initially conducted using the SILO point data from January 1890 to December 2018. On seeking to extend the modelling to the present date, it was identified that the timeseries downloaded from SILO had changed. The SILO website indicates that the database is 'constantly evolving' and provides release notes corresponding to updates of the database (https://www.longpaddock.qld.gov.au/silo/about/data-updates/). Recent changes have been made in June 2022, July 2020, September 2019 and April 2018. While there is no published guidance on changes to the overall methodology, the announcements indicate that there have been updates to parameter values used for normalisation and there are potentially 'large differences' associated with the removal of erroneous values underlying the interpolation.

Changes to the interpolated SILO rainfall and evapotranspiration timeseries raises a critical question:

> To what extent does the interpolation of the historical observation record impact the derivation of streamflow?

In the literature to date, the rainfall interpolation method from SILO has been compared to the Australian Water Availability Project (AWAP) interpolation produced by the Bureau of Meteorology (Beesley et al. 2009). This study determined that interpolation methods can have numerous artefacts, including smoothing of extremes, shifts in interpolated values due to changes in gauge availability, the potential for discrepancy in rainfall amounts and overestimation of the number of wet days. Despite contrasting methods of interpolation, both SILO and AWAP have had similar error statistics on gridded rainfall datasets. Whereas the Beesley et al. (2009) study was directly focused on the interpolated product, this study has a key point of differentiation: to evaluate the ultimate impact of differences in weather data in terms of streamflow.

This study does not seek to re-interpolate or unpick the SILO method to identify uncertainties associated with the fundamental interpolation configuration (influence of specific gauges, influence of specific missing days, etc.). Instead, the study adopted here is end-of-system focused, meaning that it seeks to identify the accumulated impact of any changes in terms of magnitude of streamflow and for the use-case of the DPE project seeking to extend the study from 1890-2018 to 1890-2021. The motivation for the study was the acknowledgement that the period since 2018 in the Lachlan catchment represented drier years that could influence the calibration of a stochastic weather generator, yet the impact of the new interpolation was unknown. In other words, for an existing calibration and set of stochastic rainfall (1890-2018), is there a remarkably different result extending to 2021 if the underlying SILO data has a new interpolation? It is undesirable to have to repeat extensive calibration and simulation procedures whenever there is a newly announced change to the underlying dataset. More broadly, it is important to appreciate the potential impact uncertainties within inputs that can affect output values.

3. METHOD

Two SILO datasets were generated from the different interpolation methods (prior to and after 2018). This climate data was used as an input to a rainfall-runoff model (e.g., Sacramento model) to calculate streamflow. This process is shown in Figure 1. By using the same models and only varying the interpolation methods and time-period, the study was able to isolate the effect of the interpolation. The impact of the interpolation methods was demonstrated through the discrepancy in the resulting rainfall/evaporation data and streamflow.

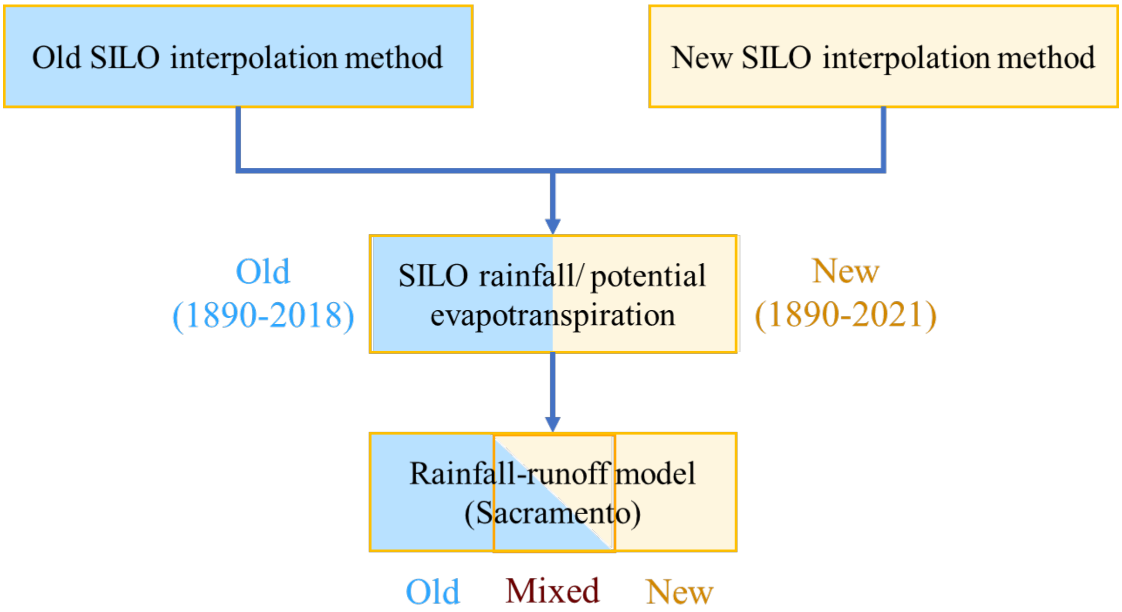

Figure 1. A process to compare the impacts of two interpolation methods for the SILO database (blue represents old data (rainfall, evaporation, and flow), yellow represents new data, and a combination of both colors represents mixed data, e.g. streamflow generated from ‘old’ rainfall and ‘new’ evapotranspiration)

3.1. Discrepancy calculation

Table 1 shows a list of comparisons made in this study. In this table, ‘Old’ and ‘New’ refers to data from the prior and newer interpolation algorithms, respectively. ‘Mixed’ for flow refers to a method of having one input timeseries as ‘Old’ (either rainfall or evapotranspiration) while having the other as ‘New’. Case 1 and 2 focus on analysis of the inputs. The other cases consider streamflow outputs from the Sacramento model.

Case 1 and 2 and 3 compare the impact of the interpolation algorithm on the two input variables (i.e., rainfall and evapotranspiration) and the output flow variable respectively. In all instances the 1890 to 2018 period is used as the baseline so that the effect of the additional 3 years (2019-2021) does not affect the outcome.

Case 4 is a comparison of the impact of the additional 3 years of data on the flow, i.e., sensitivity to added record. In both instances the new interpolated data was used. This case allows for a comparison of the relative effect on streamflow of additional data compared to the effect of changing interpolation method.

Case 5 and 6 use the 1890 to 2018 baseline and change only one of the hydrological inputs to the new interpolation method to determine whether either the rainfall or the evapotranspiration has a bigger impact on streamflow.

In all instances the discrepancy was calculated (with reference to V1, V2 naming in Table 1) as

$$\text{discrepancy} = \frac{Variable2 - Variable1}{Variable1} \quad (1)$$

In short, an increase is related to either the impact of the ‘New’ interpolation method (Case 1,2,3,5,6) or the impact of an additional 3 years (Case 4).

Table 1. A list of comparisons made between hydrological variables

		Input variables 1890 2018 2021		System output 1890 2018 2021
Case 1: Input only analysis. This case gives % discrepancy in rainfall over a common period.	Old Rain	V1		
	New Rain	V2		
Case 2: Input only analysis. This case gives % discrepancy in evapotranspiration over a common period.	Old Evap	V1		
	New Evap	V2		
Case 3: Output analysis. This case gives % discrepancy in streamflow over a common period.	Old Rain		→ Old Flow	V1
	Old Evap			
	New Rain		→ New Flow	V2
	New Evap			
Case 4: Output analysis. This case gives % discrepancy in streamflow due to extending the dataset, i.e. sensitivity to added record.	New Rain		→ New Flow	V1
	New Evap			
	New Rain		→ New Flow	V2
	New Evap			
Case 5: Output analysis. This case gives % discrepancy in streamflow due to only changing rainfall but not evapotranspiration.	Old Rain		→ Old Flow	V1
	Old Evap			
	New Rain		→ Mixed Flow	V2
	Old Evap			
Case 6: Output analysis. This case gives % discrepancy in streamflow due to only changing evapotranspiration but not rainfall.	Old Rain		→ Old Flow	V1
	Old Evap			
	Old Rain		→ Mixed Flow	V2
	New Evap			

3.2. Application

The Lachlan Catchment was used as the case study (Figure 2). It covers an area of approximately 84,700 km^2 and is characterized by a diverse landscape, ranging from the high country of the Great Dividing Range to the flat plains. The water resources of the Lachlan Catchment are crucial for irrigation, urban water supply, and environmental needs, making it a key focus for sustainable water management practices. This study uses 17 sub-catchments, that have 30 rainfall sites and 13 'Morton Wet' evapotranspiration sites in total.

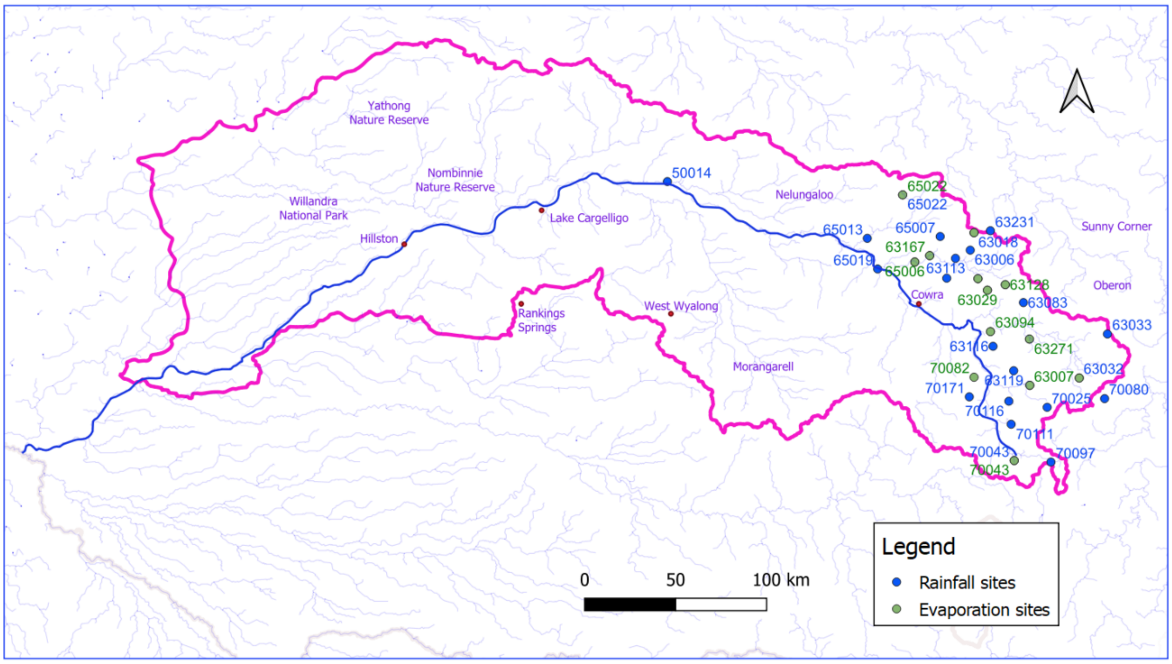

Figure 2. Rainfall and evaporation sites in the Lachlan catchment, Australia

4. RESULTS

As an example, detailed results will be shown for a selected catchment, **412080**, prior to summarising across all catchments. Figure 3 (top row) shows that the new interpolation method causes considerable differences to both the rainfall input (left) and potential evapotranspiration (right). For example, on one of the days, the old rainfall interpolation was 90 mm and under the new interpolation it is 0 mm. The bottom-left panel shows the impact of these inputs aggregated to the annual scale, and while the rainfall has some mild scatter with some positive and negative discrepancies (overall –2.2% lower) the PET shows a positive discrepancy (+4.1%). Transforming these inputs via the Sacramento model generates a -17% difference in streamflow (the bottom-right panel). In other words, the drier conditions from the rainfall and evapotranspiration have a large impact on the streamflow for this catchment.

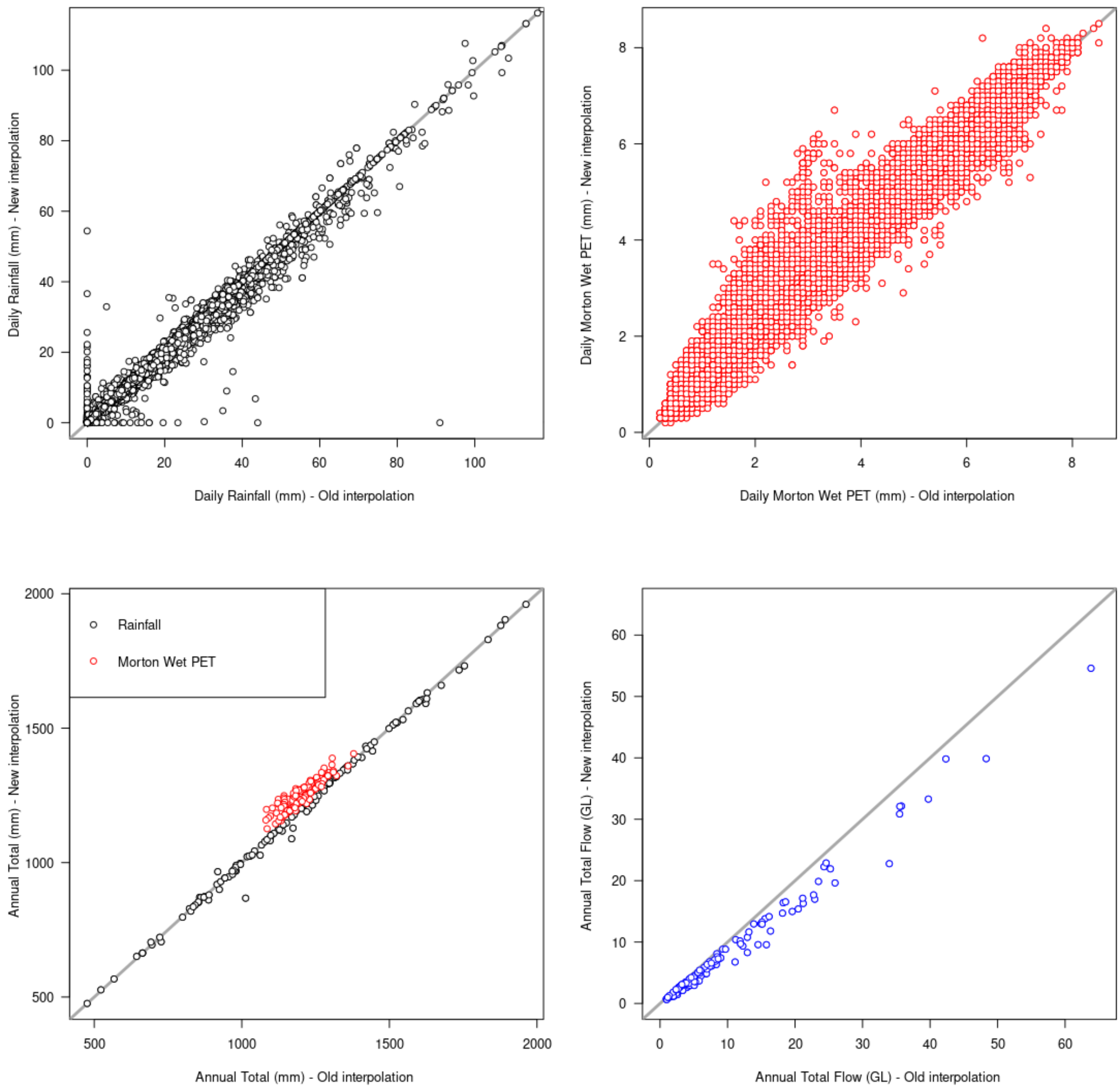

Figure 3. (top row) Change in daily rainfall and Morton Wet PET respectively due to a change in the interpolation method, (bottom left) change in annual totals of rainfall and PET inputs and (bottom right) streamflow output for catchment 421080

Figure 4 dissects the respective contribution of the newly interpolated rainfall and evapotranspiration. The solid dash and blue circle respectively give the streamflow from the old/new interpolation (same values as in Figure 4 bottom right), the green and purple squares show the contribution when only one of the newly interpolated inputs (either rainfall or PET) is used while keeping the other input from the old interpolation. Case 5 uses the newly interpolated rainfall and has a –7.3% difference in streamflow on average and Case 6 uses the newly interpolated PET and has a –10.8% difference on average. While PET has a larger overall impact on streamflow, the change in the input was 4.4% giving a 2.2x change factor. In comparison the streamflow response to rainfall is more sensitive and has a 3.5x change factor. Case 3 is not shown and had negligible impact +0.8%, which is the change due to the additional 3 years of data.

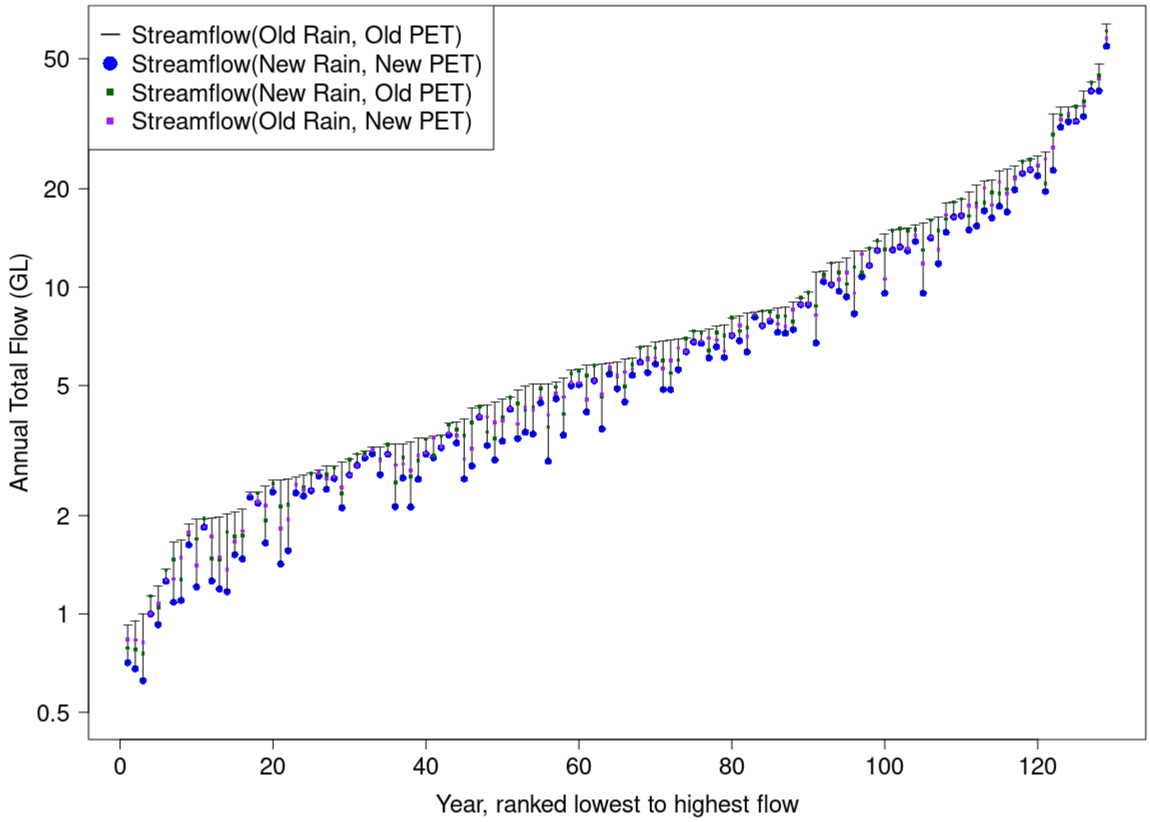

Figure 4. The reduction in streamflow from the old interpolation to the new interpolation is broken into components that show partial reductions in flow when only the rainfall or the streamflow is changed for catchment 421080

Table 2 shows the outcome of the study for simulated annual mean totals across all sub-catchments. Comparing Case 1, Case 2 and Case 3, the discrepancy in the rainfall and evapotranspiration is relatively low (rainfall ranging from –2.5% to 0.7%, evapotranspiration from –3.5% to 4.1%), when compared to the discrepancy those inputs have on streamflow (ranging from –17.3% to 7.2%). This indicates the sensitivity of flow conditions to rainfall and evapotranspiration inputs.

Comparing Case 3 and Case 4, it can be seen that the impact of the additional 3 years of record in terms of streamflow is low (ranging from –0.5% to 1.3%) compared to the impact of the interpolation algorithm. This means that the change in streamflow due to the additional years of data is not important when compared to the change caused by the interpolation method.

Case 5 and Case 6 dissect the relative contribution of rainfall and evapotranspiration to the impact on streamflow identified in Case 3. In most instances the impact of the evapotranspiration on flow (Case 6) is larger than rainfall (Case 5), which is because there were bigger discrepancies in evapotranspiration (Case 2) relative to rainfall (Case 1). The range of the flow discrepancy due to the rainfall change is from –7.3% to 3.2%, while the flow discrepancy due to the evapotranspiration change ranges from –10.8% to 8.9%. In other words, while the streamflow substantially amplifies the discrepancy of the input variables, the proportional impact of the streamflow discrepancy follows the discrepancy of the input variable.

Table 2. Discrepancy in simulated annual mean for 17 catchments (rows) and cases 1 to 6. Orange/Blue corresponds to "drier"/"wetter" conditions, e.g. increased rainfall is blue but increased evapotranspiration is orange. Cases 1 and 2 are only about input/forcing changes, cases 3 and 4 are about streamflow changes, whereas cases 5 and 6 are the impact of inputs on streamflow

Catchment	Case 1 *Rain discrepancy*	Case 2 *Evap discrepancy*	Case 3 *Flow discrepancy*	Case 4 *+3yr Flow discrepancy*	Case 5 *Rain impact on flow discrepancy*	Case 6 *Evap impact on flow discrepancy*
412027	-0.4%	-3.5%	7.2%	1.1%	-1.5%	8.9%
412083	0.7%	-2.1%	6.9%	1.0%	3.2%	3.7%
412010	-0.1%	-2.0%	6.6%	-0.5%	0.0%	6.6%
412195_412033_res_inf	0.1%	-1.5%	6.5%	0.3%	0.0%	6.5%
412056_412195_res_inf	0.5%	-2.1%	5.2%	0.6%	0.3%	4.8%
412073	0.0%	-1.5%	4.5%	0.3%	0.0%	4.5%
412071	-1.2%	-2.1%	2.0%	-0.2%	-1.8%	3.8%
412028	0.1%	0.1%	0.2%	-0.5%	0.6%	-0.4%
412065	-0.5%	-1.0%	-0.7%	1.3%	-2.5%	1.8%
412106_412077_res_inf	-1.1%	-0.3%	-2.2%	0.8%	-1.8%	-0.4%
412050	-0.2%	0.2%	-2.4%	0.9%	-0.8%	-1.5%
412043	0.0%	-1.0%	-3.2%	0.3%	0.0%	-3.2%
412077_412056_res_inf	-0.8%	-0.3%	-3.4%	1.3%	-2.8%	-0.6%
412081	-0.5%	1.3%	-5.9%	0.3%	-1.8%	-4.2%
412092	-1.9%	0.1%	-7.1%	0.0%	-6.9%	-0.3%
412070	-2.5%	4.1%	-15.0%	0.4%	-6.4%	-9.4%
412080	-2.2%	4.1%	-17.3%	0.8%	-7.3%	-10.8%

5. DISCUSSION AND CONCLUSION

Evapotranspiration and rainfall inputs to streamflow models should be checked to ensure that the values used are representative. A recent change in the SILO interpolation algorithm had a modest change to input timeseries (i.e. <5% discrepancy) but streamflow models amplify this discrepancy to significant levels for some sub-catchments (from +7.2% to –17.3%). An additional reflection is that it is important to avoid 'over fitting' of models to specific values of streamflow. Instead, decisions should be based on observed data by interpreting results against a backdrop of uncertainty in the data. This study can be extended to investigate if the differences are due to daily variations or systematic changes in persistence across different time scales. It is also important to explore potential discrepancies in longer-term aggregations, such as the influence of the new interpolation method on autumn rainfall.

ACKNOWLEDGEMENTS

This work is part of the project "Multisite Rainfall and Evapotranspiration Data Generation for the Lachlan Region". It is completed by Water Research Centre research group, The University of Adelaide and funded by New South Wales Department of Planning and Environment (DPE).

REFERENCES

Beesley, C., Frost, A., Zajaczkowski, J., 2009. A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets, 18th world IMACS/MODSIM congress, Cairns, Australia. Citeseer, pp. 13-17.

Bennett, B., Thyer, M., Leonard, M., Lambert, M., Bates, B., 2018. A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model. Journal of Hydrology 556 1123-1138.

Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling & Software 16(4) 309-330.

Leonard, M., Westra, S., Bennett, B., 2019. Multisite Rainfall and Evaporation Data Generation for the Macquarie Water Infrastructure Project, Report 2019/01: The University of Adelaide.

Loucks, D.P., Van Beek, E., 2017. Water resource systems planning and management: An introduction to methods, models, and applications. Springer.

Queenland Goverment, 2023. SILO - Australian climate data from 1889 to yesterday.

Refsgaard, J.C., Abbott, M.B., 1996. The role of distributed hydrological modelling in water resources management. Distributed hydrological modelling 1-16.