Office of the Chief Scientist & Engineer

Independent review of the climate risk method for the NSW Regional Water Strategies Program

Additional advice subsequent to April 2020 Panel report – Southern Inland NSW and Greater Sydney Region

Independent Expert Panel

May 2021

Disclaimer

Third Party Reliance

The NSW Government does not take responsibility arising in any way from reliance placed by a third part on this report. Any reliance placed is that party's sole responsibility. We shall not be liable for any losses, claims, expenses, actions, demands, damages, liability or any other proceedings arising out of reliance by any third party on this report.

Note: The proposals and reports that the Panel reviewed for this report were not public at the time of the review.

The Hon Melinda Pavey MP
Minister for Water, Property and Housing
GPO Box 5341
SYDNEY NSW 2001

Dear Minister

Independent Review of Climate Risk Method for the Regional Water Strategies program

In September 2019, you requested that I chair an independent expert panel to assess the proposed methodology being used by DPIE-Water to incorporate climate risk into water resources planning as part of the development of Regional Water Strategies in NSW.

In April 2020, I provided you with the Panel's report. Subsequent to this, the Department has been undertaking work to implement several of the recommendations made in the Panel report.

The Panel was reconvened in December 2020 at the request of DPIE-Water to provide independent comment on the suitability of the methods to address the recommendations, particularly testing for non-stationarity in the historical record and incorporating it into the stochastic modelling for the Southern Inland NSW and Greater Sydney regions.

Please find enclosed the Panel's advice.

I would like to acknowledge the contribution of Panel members: Professor Bryson Bates, Emeritus Professor George Kuczera, Professor Andy Pitman and Dr Scott Power, as well as the advice and support provided by staff and other stakeholders.

Yours sincerely

Dr Chris Armstrong PSM
Deputy NSW Chief Scientist & Engineer
Chair, Independent Expert Panel
9 May 2021

CONTENTS

Table of Contents

EXECUTIVE SUMMARY AND RECOMMENDATIONS

The water modelling group within the NSW Department of Planning, Industry and Environment (DPIE-Water), working with the University of Adelaide (UoA) and the University of Newcastle (UoN), is assessing the presence of non-stationarity of climate and proposing approaches to incorporate non-stationarity into climate modelling for the Regional Water Strategies (RWS) for Southern Inland NSW and Greater Sydney regions. This work is testing for the presence of signals that the climate has changed over recent decades and if it has, how the change should be incorporated into modelling undertaken to inform climate risk. If the climate has started changing over the recent decades and this is not accounted for in the stochastic modelling, future water security risk may be underestimated.

This issue was raised in the April 2020 report by the Independent Expert Panel (the Panel) assessing the climate risk method for the RWS. The Panel was formed at the request of the Minister for Water, Property and Housing in September 2019 to provide independent evaluation of the methods used to assess and account for climate risk in modelling future water supplied for the twelve RWS being developed across NSW.

Note that this report does not provide full background to the RWS and the issues raised in the April 2020 report. It is advised that this report is read in conjunction with the April 2020 report.

DPIE-Water, in conjunction with UoA and UoN, proposed methodologies to the Panel for addressing recommendations made in the April 2020 report. These included addressing the question of the presence of non-stationarity and how to include it in the stochastic modelling, as well as a recommendation to use NARCliM 1.5 data, as it was becoming available at the time.

The scope of this advice is to review three documents:

1. **Draft report:** Devanand, A., Leonard, M. and S. Westra (2020). *Implications of Non-stationarity for Stochastic Time Series Generation in the Southern Basins*. UoA (Appendix 1)
2. **Proposed methodology:** *Provision of stochastic climatic data for the Southern Basin representing historical, current and future climate conditions*. Proposal by UoA to DPIE-Water, 28th August 2020. (Appendix 2)
3. **Proposed methodology:** *Provide stochastic climatic data for the Greater Sydney Region representing historical, current and future climate conditions.* Proposal by University of Newcastle to DPIE-Water, 27 November 2020. (Appendix 3)

The Panel recognises some of the issues raised here are complex and challenging to resolve. Some of the issues are the subject of cutting edge, ongoing research, and strategies to quantify climate risk will require ongoing attention into the future and likely adjustment as understanding improves. This Panel advice elaborates on the issues and suggestions to address them. The Panel acknowledges that some of these will require a longer timeframe to resolve.

Overall, the Panel recommends proceeding with both the UoA and the UoN proposals.

There was Panel consensus that there is no definitive approach to account for non-stationarity in the observed record or future climate. The goal here, ultimately, is to estimate future catchment yield, which requires a series of steps that inherently involve a cascade of uncertainty.

Addressing the preferred methodology of including non-stationarity in stochastic models is complex and was subject to Panel debate. The Panel highlighted the need to be rigorous in documenting historical non-stationarity as a foundation to future work.

The Panel suggested a precautionary approach is taken in that if there is an error, it is better to over-estimate risk of drying than under-estimate it, and the Panel is supportive of using scenarios to test various possibilities and manage the high uncertainty.

The Panel is also supportive of using a 'storyline' approach, where one thinks about the whole system and how it interacts with the landscape and water users (e.g. urban water supply vs irrigation valley). By developing a conceptual storyline for a catchment, components of the model can be designed and applied to reflect the conceptual understanding of the system.

The following table describes the Panel's recommendations on the main process steps outlined in the UoA and UoN proposals to incorporate non-stationarity into the stochastic models. Additional recommendations to address information gaps related to this topic, but that go beyond the immediate proposals from UoA and UoN, and are on a longer timeframe, are also included. More minor suggested improvements to the draft UoA report by Devanand et al are within the body text of this report (Section 2.1.3) and are not reflected in Recommendation 2 in the table.

FINDINGS AND RECOMMENDATIONS

No.	Step/topic	Panel recommendation/comment	Priority	Timeframe	Applies to
1	**Overall - proposals for Southern Basin and Greater Sydney**	Panel recommends proceeding with both proposals[1]	High	Short (0-3 months)	UoA; UoN
2	**Test for non-stationarity in Southern Basin**	The Panel's assessment is that the approach used in the UoA draft report is sound, and the outcomes revealed statistically significant non-stationarity in some, but not all, places. This analysis is consistent with expectations from other relevant research.	N/A	N/A	UoA
3	**Testing for non-stationarity in streamflow for Greater Sydney**	The Panel noted that testing for non-stationarity in Greater Sydney is expected to be affirmative for temperature and unclear for rainfall. The Panel notes that testing for non-stationarity of events that drive rain like East Coast Lows (ECLs) would be useful information. However, the statistics of checking for non-stationarity on something like ECLs may be problematic, given that they are relatively infrequent events. a) The Panel recommends that testing for non-stationarity of temperature and rainfall is not undertaken as proposed at this time, as the Panel recommends that other activities identified as having a higher priority are conducted first, and that this additional exercise is then only undertaken if remaining time and resources allow. b) However, the Greater Sydney region has long streamflow records dating back to 1910. Given there is published evidence of change in these records in the last 30 years and given the timing of the current Greater Sydney water plan, it is recommended that non-stationarity in the streamflow record be investigated with the goal of understanding the causes. This will allow immediate assessment of current climate risk on Greater Sydney Region's water supply. Focussing on streamflow avoids the need to deal with non-stationarity in rainfall, PET/ET and rainfall-runoff. The UoN stochastic model works on seasonal timescales and disaggregates to shorter time scales using a non-parametric method. It is well suited for modelling streamflow.	a) Low (if undertaken) b) High	a) N/A b) Short	a) UoN (if undertaken) b) UoN for stochastic modelling and hydrology If time doesn't permit, potential external parties (e.g. UNSW) for investigation of causes and processes.
4	**Options (5 total) presented by UoA to account for non-stationarity into stochastic model generation**	The Panel consensus is that there is no definitive approach to account for non-stationarity in the observed record or future climate. Hence the Panel supports a multi-pronged approach, particularly where approaches have different conceptual foundations.[2]	a) High b) High[3]	a) Short b) Short[3]	a) UoA; UoN b) UoN

[1] The Panel noted the following recent paper may be useful when considering the recommendations. Fowler, Keirnan; Acharya, Suwash Chandra; Addor, Nans; Chou, Chihchung; Peel, Murray (2020): CAMELS-AUS v1: Hydrometeorological time series and landscape attributes for 222 catchments in Australia. PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.921850

[2] The Panel does not have a single endorsed approach to address these difficult problems and multiple lines of evidence are preferred. The Panel also noted it was too early to settle on a particular approach. Some Panel members were of the view that Recommendation 4a (Option 3- hybrid historical baseline) was the best of the options as it is consistent with a "preparedness rather than prediction" paradigm, which was deemed more likely to lead to a strategy that is more robust to climate variations. Other Panel members expressed concerns about the baseline approach to adequately describe current climate risk. These issues are discussed further in the report.

		a) The Panel recommends Option 3 – hybrid-historical baseline - as a worthwhile option overall for UoA and UoN work. This uses a stationary stochastic model with a representative stationary climate record. There are two possibilities: fit to baseline record (which is assumed to be sufficiently short to be near stationary) or adjust pre-baseline record to match certain statistics in the baseline record (as done by Melbourne Water) and then fit to longer record. b) For UoN, applicable to the Greater Sydney region where the relevant data is available, another worthwhile option is use of a nonstationary stochastic model approach. UoN is already capable of applying such a model to streamflow in Sydney region. It is recommended UoN's remit be extended to applying this approach and comparing it with the baseline approach.[3]			
5	**Adoption of NARCliM to inform climate change**	The Panel recommends using NARCliM 1.5, but undertake due diligence as follows in recommendation 6.	High	Short	UoA; UoN
6	**Validation of NARCliM model outputs to statistics of hydrological importance including extreme or dam filling events**	The NARCliM developers should work with hydrological modelers to identify key metrics in each catchment from a hydrological perspective that could be used to validate how well NARCliM captures statistics of hydrological importance. The NARCliM 1.5/2.0 Technical Reference Group, in which DPIE Water participates, may be the suitable mechanism for this work to be undertaken. It will be important for the metrics to be documented, agreed and peer reviewed.	High	Medium /Long	DPIE NARCliM team with DPIE water resource modelling team, UoA/UoN stochastic modelling expertise
7	**Signal to noise and baseline –** The issue of signal to noise is important and confidence in the approach will largely depend upon the ability to manage this issue. The baseline period will affect the signal to noise ratio. [4]	The Panel recommends that UoA initially (then UoN) investigate the signal to noise ratio to assess options regarding scaling. Choosing a baseline starting point is complex. Victoria DELWP chose 1975, but some Panelists were uncomfortable with using 1975 as a starting point, as there is a significant change signal over that period. Until the trade-offs are better understood, the Panel cannot recommend a specific baseline period. The rolling evaluation approach (Appendix 4) (or something similar) might provide insight into the temporal stability of simulated streamflow and hence changes/trends in hydroclimatic regime. There is no guarantee that any such change will occur simultaneously across the regions or at the same rate.	High	Short	UoA (first) UoN (pending outcomes of UoA)
8	**Estimating upper and lower limiting cases of climate risk on water security[4]**	As it is not complex to undertake, the Panel suggests upper and lower limiting cases for climate risk are estimated as a 'reality' check on other approaches. Note this is the approach used by NARCliM.	Medium	Short	UoA; UoN

[3] The Panel is of the view that Recommendation 4b is feasible for UoN to undertake in a short time frame and high priority. The software is developed, temperature data is readily available, as are long streamflow records, for the Greater Sydney region/catchment which is largely unaffected by land use changes. Running the model to produce diagnostics should take minimal time. The Panel considers this recommendation important because there is published evidence of significant change in streamflow distributions over last 30 years. A stationary streamflow model badly fails split sample tests, and if used, would significantly overestimate current streamflow resources.

[4] The Panel views Recommendations 7 and 8 as producing outcomes that may help better understand the trade-offs and risks with using the baseline approach (Rec 4a).

		For the lower limiting case, calibrating a stationary stochastic model to the full observed record (1910 – 2020) without adjustment ignores the drying change signal and thus will provide an underestimate of current drought risk. The upper limiting case can be estimated in two ways: calibrate to a recent past record (1990 – 2020)[5] or calibrate to the full observed record with constraint on selected statistics.			
9	**Validate against streamflow data –** The Panel expressed discomfort testing the DPIE stochastic model purely in rainfall/evapotranspiration space; important to validate using streamflow data	Stochastic rainfall approach needs to be tested against reliable streamflow records using split sample testing regime or rolling origin approach to model evaluation – run through rainfall/runoff models and compare against long streamflow records. Test the DPIE stochastic model's performance through a forensic look at the rainfall characteristics of dam filling events. Determine metrics that impact water availability and how well the model performs against those. UoA/UoN to provide stochastic rainfall and ET. It is expected DPIE-Water will run forcing through their calibrated models and UoA/UoN will undertake the split-sample and/or rolling origin analysis. There may also be a need to apply post-processing algorithms, which has a precedent through work done by BOM and other parties.[6]	High	Medium	DPIE Water modellers working with UoA; UoN
10	**Weather systems such as ECLs typing**– note there are multiple types of ECLs (Section 4.1.1)	Longer term, consider ECLs by type and mechanism to better assess their impacts, as there may be a different response according to the mechanism. Other weather systems could also be looked at.	Medium/ Low	Long (2-3 years)	External university research (e.g. UNSW, Monash or UoMelb) working in partnership with/contribution from DPIE
11	**PET vs ET (and analysis of changing vegetation effect on flow)**	Do not assume (or prescribe) changes in PET moving forward – keep PET at levels reflecting the last 20 years. While PET will increase in the future, this will likely be an artefact of its formulation and actual ET may not change in ways consistent with PET.[7]	High	Medium/ long	DPIE Water modelling team with external partners (e.g.

[5] This situation assumes current hydroclimate is represented by period 1990-2020. Split sample testing suggests this period is different from pre-1990. The post-1990 record contains two major droughts. As the 1990-2020 stationary scenario assumes this is the new normal, it is considered an upper case. This is similar to the DELWP approach; however, some of the climate change scenarios used by Victoria are more severe than assuming their stationary post-1997 record.

[6] Note: DPIE provided the Panel with information on Quality Control (QC) work undertaken to address this issue. The Panel noted that the QC process used by DPIE does not mention split sample testing, which is arguably the most effective tool to detect non-stationarity. The tests described in the QC document are necessary but not sufficient. The Panel has experience with models that can pass the tests described in the QC document but fail split sample tests. This can occur because the QC tests are focussing on properties of marginal distributions (i.e. properties of the whole record), and they are not designed to look for change.

For the QC tests described by DPIE, the Panel recommends the use of p-values to evaluate the consistency of the observed statistics with model sampling distributions.

The Panel was not aware of DPIEs experience using the QC tests. However, the Panel noted that unless the rainfall-runoff model has a high goodness of fit score, they suspect tests dealing with variability and extremes may not fare well. There are many tests available, but given DPIE's objectives, a judicious selection should lead to a small suite of options. Given there is a distinct possibility of trends in the observed data, the Panel doesn't recommend using the correlation coefficient.

[7]There is ongoing discussion about the shortfalls if using PET in some modelling. This issue will be looked at in the Community of Practice and in longer term research.

		This work needs to examine vegetation’s role in linking climate change with soil moisture and catchment yield.			UNSW, CSIRO CABLE team)
12	**Non-stationarity of climate drivers**	Adding climate drivers to stochastic models increases their complexity, and the challenge is to show evidence that doing this provides statistically significant improvements to the rainfall/streamflow simulations. DPIE should keep abreast of advances in the relevant climate driver/teleconnection research to ensure that approaches used reflect scientific consensus on non-stationarity of climate drivers. This may be addressed through the previously recommended Community of Practice.	Medium	Long	DPIE

1 BACKGROUND AND SCOPE

In September 2019, the Minister for Water, Property and Housing, the Hon Melinda Pavey MP, requested that the Deputy Chief Scientist & Engineer chair an independent expert panel (the Panel) to provide advice on the incorporation of climate risk into the development of Regional Water Strategies (RWS).

The Panel includes expertise in climate drivers, variability and change, extreme weather events, hydrology and statistics and is comprised of:

- Dr Chris Armstrong, Deputy NSW Chief Scientist & Engineer (Chair)
- Professor Bryson Bates, Adjunct Professor, School of Agriculture and Environment, University of Western Australia
- Emeritus Professor George Kuczera, School of Engineering, The University of Newcastle
- Professor Andy Pitman, Director, Australian Research Council Centre of Excellence for Climate Extremes, Climate Change Research Centre, UNSW Sydney
- Dr Scott Power, Adjunct Professor, School of Earth, Atmosphere and Environment, Monash University.

The Panel provided a report, *Independent review of the climate risk method for the NSW Regional Water Strategies Program*, to the Minister in April 2020. Background to the RWS, the Panel's role and specific advice are contained within that report, and this advice should be read in conjunction with the April 2020 report.

Since then, the methodology to incorporate climate risk into the RWS has evolved to address some key recommendations from the April 2020 review, including:

Recommendation 4: DPIE-Water engages external expertise to explore options to improve proxy records by e.g. obtaining and incorporating local proxy records. This may go some way toward improving the situation where current proxy records are derived from distant locations with possibly different climate influences. Efforts to improve both the quality and quantity of proxy records could improve our understanding of climate variability. Moreover, the (possibly interactive) climatic variables that affect the proxies need to be clearly identified by experts. **(Medium priority: Long term)**
Recommendation 6.2: DPIE-Water, in collaboration with experts in climate science and statistics, to explore alternative approaches to generating randomised samples as part of the future research program, particularly for regions where it is not clear what the dominant driver is or where there are multiple dominant drivers. **(Medium/High priority: Short, then Long term)**
Recommendation 7 DPIE-Water engages external expertise to undertake a two-step approach to investigate stationarity to manage the risk that models may underestimate current and future climate risk. The objective is to capture the current climate risk and a baseline that reflects this. First Step **(High priority: Short term)** The first step is to make a careful assessment of non-stationarity in the observed record. Without pre-empting such an assessment two useful methods are described. *Step 1_Method 1* Split-sample testing or hold-out validation splits the observed record into two parts. The stochastic model can be calibrated to each part and the sampling distributions of the parameters compared to determine if there is evidence of significant differences.

Alternatively, the model can be calibrated to one part and its performance in the second part assessed using diagnostics based on predictive distributions. This approach implicitly assumes the record used for calibration is stationary, so care is required in deciding how to split the observed record. It could be split as follows: ● *1990 reference* - the observed record could be split into pre- and post-1990 intervals ● *Drought reference* - the record used for validation and prediction could be chosen with regard to the Millennium Drought and the current drought, with three different calibration periods 1) from start of record (SOR) to the beginning of Millennium Drought; 2) from SOR to end of Millennium Drought; 3) from SOR to beginning of current drought; with the three prediction periods being from the end of the calibration period to today. A variation on this scheme is to constrain the model calibration periods to be the same length, so the earliest parts of the historical record are discarded for Run numbers 2) and 3) Using both of these approaches (*1990 reference* and *drought reference)* could be prudent to confirm the outcomes. *Step 1_Method 2* Calibrate a stochastic model conditioned on an observed covariate associated with a change signal (e.g., temperature). While this approach is still in the research domain it is expected to have more statistical power to identify non-stationarity than the first method (provided a suitable covariate can be found). Second Step **(High priority: Medium/Long term)** If, after application of the first step, it is concluded that the recent past may not be statistically consistent with the full observed record and, more importantly, that differences are hydrologically significant, it will be necessary to revise the stochastic modelling approach described in the Methods Paper. A number of conceptually different approaches are available: *Step 2_Approach 1* Calibrate the stochastic model to the recent past record with the knowledge that parameter uncertainty will be elevated; *Step 2_ Approach 2* Adjust the observed record prior to the recent past so that it reproduces certain statistical features of the recent past and then calibrate the stochastic model to the full adjusted observed record; *Step 2_ Approach 3* Identify suitable change covariates and calibrate a conditional stochastic model. Each of these approaches has strengths and weakness. As there is limited experience with this issue it is expected that methods will evolve in the future.
Recommendation 9.1: DPIE-Water work with the NARCliM developers and together begin the process of planning to incorporate NARCliM 1.5 into calculations. **(High priority: Short term)**

The Panel was reconvened in December 2020 to consider DPIE-Water's proposed approach to implement some of these recommendations for Southern Inland NSW and Greater Sydney regions.

The Southern Inland NSW[8] includes the whole southern catchment of the Murray Darling Basin, incorporating the following water sources:

[8] Note that the UoA draft report (Appendix A) refers to the Southern Basin and the southern region. The Department has also referred to the region as Southern Inland NSW and the Southern Connected Basin.

- Murrumbidgee Regulated River system wholly in NSW
- the Upper Murrumbidgee River in NSW and ACT;
- Murray-Lower Darling Regulated River system in NSW, Victoria and South Australia;
- The Kiewa, Ovens, Broken, Goulburn, Campaspe and Loddon river systems wholly in Victoria
- The Snowy Mountains Scheme in NSW and Victoria

The Panel was provided with three documents for which expert opinion and advice was sought:

- **Draft report:** Devanand, A., Leonard, M. and S. Westra (2020). *Implications of Non-stationarity for Stochastic Time Series Generation in the Southern Basins*. University of Adelaide (draft report). (See overview of method/findings in Section 2.1; full report at Appendix 1)
- **Proposed methodology:** *Provision of stochastic climatic data for the Southern Basin representing historical, current and future climate conditions*. Proposal by UoA to DPIE-Water, 28th August 2020. (Appendix 2 for detail)
- **Proposed methodology:** *Provide stochastic climatic data for the Greater Sydney Region representing historical, current and future climate conditions.* Proposal by University of Newcastle to DPIE-Water, 27 November 2020. (Appendix 3 for detail)

A meeting was held on 4 December to discuss approaches set out in the three documents for review, with two presentations provided to the Panel:

- Professor Seth Westra (UoA) - *Implications of Non-stationarity for Stochastic time series generation in the Southern Basins*
- Associate Professor Anthony Kiem (UoN) - *Stochastic modelling of historical and future hydroclimatic risk for the Greater Sydney Water Strategy (GSWS)*

A follow up discussion occurred on 19 January 2021 that didn't include the UoA and UoN teams. Both meetings included the water and climate modellers from DPIE.

This advice reflects discussion at those meetings plus additional Panel email correspondence and discussions.

2 COMMENT ON EVIDENCE OF NON-STATIONARITY FOR SOUTHERN BASIN AND GREATER SYDNEY REGIONS

2.1 COMMENTS ON NON-STATIONARITY IN SOUTHERN BASIN

The Panel was asked to comment on the draft report by Anjana Devanand, Michael Leonard and Seth Westra from UoA, *Implications of Non-stationarity for Stochastic Time Series Generation in the Southern Basins*.

The authors tested for presence of a non-stationarity signal of the historical climate data (rainfall, temperature, evapotranspiration) in the Southern Basin by a) assessing non-stationarity of the historical record using selected sites in the Southern Basin to detect trends and b) split-sample testing of the stochastic model.[9] This was contextualised through a literature review to understand past changes and future projections.

Key findings, as articulated by the authors in the presentation to the Panel, include:

- *Temperature is non stationary*
- *Cool season and annual rainfall totals are non-stationary*
- *There are trends in other attributes of rainfall in the pilot analysis, some of which (e.g. short-term spring decline; extremes) are less consistent with general literature*
- *Non-stationarity in evaporation is less well-established*
- *The Millennium drought is a 'high leverage' event; the period used for calibration influences the statistics of the simulated stochastic data.*[10]

The authors noted caveats for the study results which are discussed in Chapter 5 of Appendix 1.

2.1.1 Drying signal in observed record

The stochastic models used in water resource assessment typically assume stationarity. While there are different definitions of stationarity, key is that joint probability distributions of variables at different points in time do not depend on any reference point in time.

When a stationary model is calibrated to an observed record, the implicit assumption is that the observed record is a realisation of a stationary process. If a stationary model is calibrated to an observed record that is not stationary, the sequences generated by the model will not be fully representative of the observed record. Detecting such mis-specification requires the use of diagnostics that compare different periods of time. Diagnostics based on the whole record (such as moment statistics, marginal distributions, autocorrelations) average out time variation and are unlikely to detect the presence of non-stationarity. Suitable approaches include trend analysis over different epochs and split-sample analysis. Failure to properly represent non-stationarity can have major implications for assessment of current and future drought risk.

The UoA analysis in the draft report demonstrated that selected observed rainfall records in the Southern Basin experienced statistically significant change over time. Two techniques were used. One was based on trend analysis over different epochs. The other used split-

[9] Split sample testing methodology used the following time periods:
a) 1990 reference:
1889 calibration to 1990 then validation to 2018
1950 calibration to 1990 then validation to 2018
b) drought reference:
1889 calibration to 2010 then validation to 2018
1950 calibration to 2010 then validation to 2018

[10] Westra, S, Devanand, A and M Leonard. *Implications of Non-stationarity for Stochastic Time Series Generation in the Southern Basins*. UoA. Presentation to Panel, 4 December 2020.

sample analysis which involved calibrating a stationary stochastic model to data prior to 1990 and then testing the model using data independent of the calibration, from 1990 onwards. Both the trend and split-sample tests suggest there has been a drying signal associated with seasonal rainfall.

2.1.2 Prior understanding of stationarity in Southeast Australia

Panel members' prior understanding of stationarity in Southeast Australia is that temperature is now non-stationary, rainfall in the southern half of the Murray-Darling Basin is non-stationary and rainfall in the northern part of the basin cannot be shown to be non-stationary. The UoA draft report (Appendix 1) and the presentation by UoA to the Panel point to similar results. This analysis is broadly consistent with expectations from other relevant research.

In short, the Panel's assessment is that the approach used is sound and the outcomes revealed statistically significant non-stationarity in some, but not all, places. This analysis is consistent with expectations from other relevant research.

2.1.3 Minor suggested improvements to draft report

The following relatively minor suggestions are provided to improve the report.

2.1.3.1 Use of updated BOM temperature and rainfall trend data

The BOM data used in the analysis by Devanand, Leonard and Westra for temperature and rainfall has since been updated. It is important to augment analysis of historical trends of temperature and rainfall data in Southeast Australia with analyses/trends using latest BOM data (see their website), as warming trends have been markedly revised upward.

The Panel acknowledges this won't change the major outcomes or conclusions around the non-stationarity signal. However, for completeness and rigour, and if the report may be used in an external context, the authors should address this.

2.1.3.2 Statistical significance

Stating that something is "statistically significant" or "not statistically significant" can be helpful, but it is typically better to also provide a number giving the likelihood of obtaining a trend as large or larger than observed, under the null hypothesis (e.g. provide estimated p-values). See, for example, Table 1 in Power and Callaghan (2016).[11] This also applies for field significance.

2.1.3.3 Additional suggested reference for literature review

The panel suggests the authors add a brief discussion of Rauniyar and Power (2020)[12] in relation to observed trend and attribution of recent drying (in Victoria only). Rauniyar and Power (2020) estimates the contribution of anthropogenic forcing to drying in recent decades. It also makes use of information from climate models for coming decades. Estimates that ~80% of observed cool season drying since beginning of Millennium Drought in Victoria was due to natural, internal variability, albeit with high uncertainty.

[11] Power, S.B., and J. Callaghan (2016). The frequency of major flooding in coastal southeast Australia has significantly increased since the late 19th century. *J. South. Hemis. Earth Sys. Sci.*, **66**, 2-11. 10.1071/ES16002.
[12] Rauniyar, Surendra, and Scott B. Power (2020). The Impact of Anthropogenic Forcing and Natural Processes on Past, Present, and Future Rainfall over Victoria, Australia. *J. Climate*, 33, 807-8106, 10.1175/JCLI-D-19-0759.1

2.2 GREATER SYDNEY REGION – TEST FOR AND EVIDENCE OF NON-STATIONARITY

2.2.1 Proposed approach to testing for non-stationarity of rainfall, temperature and evapotranspiration

There was no analysis presented by UoN on evidence of non-stationarity for the Greater Sydney catchments, as the project is at the proposal stage. The proposed approach put forward by UoN would follow that for the UoA Southern Basin analysis for testing for non-stationarity in Greater Sydney locations: (i) trend analysis, break point analysis and (ii) split sample calibration/validation of stochastic models.

As scales get smaller (e.g. the Sydney Basin) assessing stationarity becomes harder and is relatively more influenced by phenomenon such as East Coast Lows. The Panel noted that testing for non-stationarity in Greater Sydney is expected to be affirmative for temperature and unclear for rainfall. The Panel notes that testing for non-stationarity of events that drive rain such as East Coast Lows (ECLs) would be useful information if it could be achieved; however, the statistics of checking for non-stationarity on relatively infrequent events such as ECLs may be problematic, given that they are relatively infrequent events.

The Panel recommends that testing for non-stationarity of temperature and rainfall is not undertaken as proposed at this time, as the Panel recommends that other activities identified as having a higher priority are conducted first, and that this additional exercise is then only undertaken if remaining time and resources allow.

2.2.2 Evidence of non-stationarity in streamflow in Sydney catchment

Recent work by Kiem et al (2020)[13] points to non-stationarity of streamflow in the Sydney catchment. Split sample testing offers evidence of change since 1990 in streamflow in the Sydney catchment.[13] This observed change in streamflow may be driven more by warming (causing longer growing seasons and increased ET) than rainfall decline (see 2.4.2). These results support the conclusion that investigating non-stationarity in streamflow records deserves as much if not more priority than investigating rainfall records.

Kiem et al. (2020) show that the observed annual streamflow records for Warragamba, Shoalhaven and Avon exhibit non-stationarity. Figure 1, taken from Kiem et al. (2020), shows a split-sample analysis for 1 to 4-year overlapping sum frequency distributions for Warragamba streamflow. The black circles represent observations from 1990 to 2018. The blue line presents the median sum for a 29-year period simulated by a stationary model calibrated to data from 1910 to 1989. The dashed red lines represent 5 and 95 sum percentiles. The p values shown on each plot are the probability that the stochastic model produces an overlapping cumulative streamflow total that is less than the observed total. The p values are all less than or equal to 0.001 which means the stationary stochastic model would have to generate, on average, one thousand 29-year sequences before a generated 29-year sequence produced a total flow less than observed between 1990 and 2018. This is consistent with the findings for the southern basin, where the evidence suggests that natural variability (as simulated by a stationary stochastic model) alone cannot account for the annual (and seasonal) rainfall behaviour observed between 1990 and 2018.

As above, the Panel is recommending that testing for non-stationarity for rainfall or temperature in Sydney is not undertaken at this time. However, the Greater Sydney region has long streamflow records dating back to 1910. Given there is published evidence of change in these records in the last 30 years and given the timing of the current Greater Sydney RWS, it is recommended that non-stationarity in the streamflow record be

---

[13] Kiem, A. S., Kuczera, G., Kozarovski, P., Zhang, L., & Willgoose, G. (2020). Stochastic generation of future hydroclimate using temperature as a climate change covariate. *Water Resources Research*. 57, 10.1029/2020WR027331

investigated with an additional goal of understanding the causes. This will allow immediate assessment of current climate risk on Greater Sydney Region's water supply. Focussing on streamflow avoids the need to deal with non-stationarity in rainfall, PET/ET and rainfall-runoff.

The UoN stochastic model works on seasonal timescales and disaggregates to shorter time scales using a non-parametric method. It is well suited for modelling streamflow.

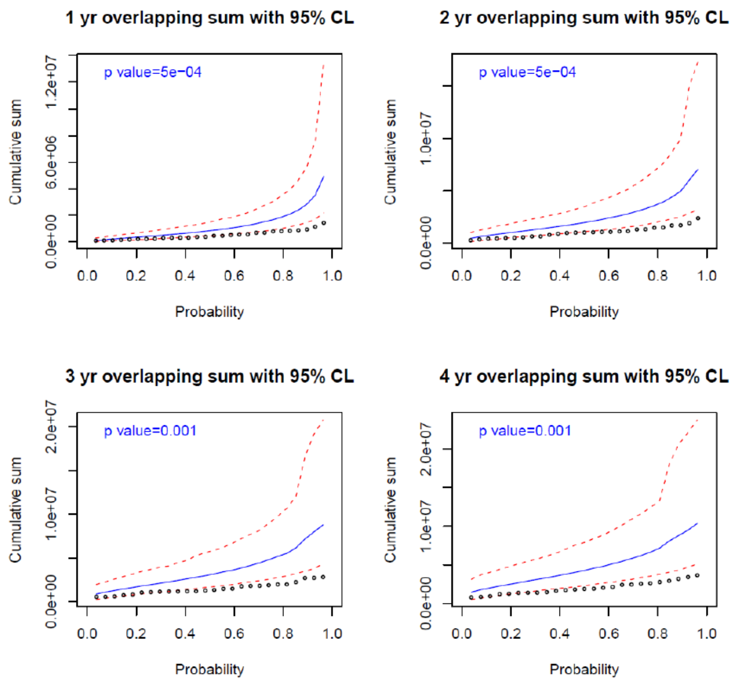

Figure 1 (taken from Kiem et al. (2020)).

Figure 1: Results of split-sample analysis for 1 to 4-year overlapping sum frequency distributions for Warragamba streamflow

2.3 ATTRIBUTION OF NON-STATIONARITY SIGNAL

Given the results of the UoA study on the Southern Basin and the Warragamba study (Kiem et al 2020), as well as other observations in the literature, it is reasonable to conclude there has been a change signal in rainfall in the southern basin (and Greater Sydney region) in the last 30 years. The question then arises: will some uncertain fraction of this drying signal persist or intensify in response to further global warming? If the answer is yes, this will affect current (and long-term) assessment of drought security (which presently assumes a stationary climate) and, in the case of urban water supply, current drought contingency plans.

The change signal in rainfall may be due to some manifestation of (stationary) multi-decadal variability that is yet to be understood or it may be a consequence of anthropogenic-induced climate change or some combination. Statistical models describe associations between random variables; causal attribution is not possible. Unravelling the causes of the change detected by statistical analysis will have to rely on climate science.

Climate science has concluded that global heating is due to anthropogenic emissions and that this causes changes to large-scale atmospheric (and ocean) circulations. For example, the climate in the southern to mid-latitude regions of Australia is susceptible to, for example, the strengthening of Hadley cells and consequent shifts in the mid-latitude high-pressure ridge. Furthermore, climate models consistently project a drying trend in rainfall during the winter-spring period in the southern basin over the remainder of the 21st century (CSIRO and

BOM, 2015)[14] and CMIP5 and CMIP6 models show drying over the southern basin via longer drought duration, and higher drought intensity (Ukkola et al., 2020).[15]

It is worth noting there are other natural and anthropogenic causes of non-stationarity in water availability. Climatic conditions can impact on demand for water.[16] For example, during hot weather, irrigators may pump more water for irrigation or likewise for other water users. When conditions are dry, additional dam release may be required. Land use can cause non-stationarity of flows. Non-stationarity of water availability can also be due to vegetation effects.

Research aimed at improving our understanding of the relative importance of anthropogenic and natural processes in driving observed changes, and the implications this has for future rainfall, is encouraged.

2.4 NON-STATIONARITY AND ITS IMPACT ON PET, ET, CLIMATE DRIVERS

2.4.1 Potential evapotranspiration vs evapotranspiration

The assessment of non-stationarity in potential evapotranspiration (PET) appears to have not been undertaken in this set of work. The Panel agrees with the challenge of assessing stationarity for PET. While it can and has been done, assessing non-stationarity in PET may not be a useful effort moving forward.

The Panel notes that catchment yield won't be well predicted unless it is recognised that PET is not going to capture how catchments respond to climate. Some panel members felt that PET should not be used in future work and may lead to fundamentally wrong future assessments of water security.

The Panel in the April 2020 report argued in support of the need to move to actual evapotranspiration (ET) which is physically represented in future scenarios derived using climate models but is not always used in hydrological modelling (Recommendation 3.2 – April 2020 report). This leads to a fundamental separation of the physically consistent climate model projections with the uncoupled hydrological projections. To further explain, actual ET depends on climate forcing, plant response and availability of water from the soil. For example, in an aerodynamically rough environment (e.g. forest canopy) with adequate soil water, Penman-Monteith ET is approximated by product of canopy conductance (GC), leaf area index (LAI) and vapour pressure deficit (VPD).

The Panel understands the rainfall runoff models used by DPIE Water do not incorporate vegetation responses as expressed by GC and LAI, which respond to climate forcing and environmental stress (such as soil water deficit). This could be a problem with the rainfall runoff models.

The Panel recommends that DPIE not assume (or prescribe) changes in PET moving forward, but rather keep PET at levels reflecting the last 20 years. While PET will increase in the future, this will likely be an artifact of its formulation and *actual* ET will not change in ways consistent with PET.

---

[14] CSIRO and Bureau of Meteorology (2015), Climate Change in Australia Information for Australia's Natural Resource Management Regions: Technical Report, CSIRO and Bureau of Meteorology, Australia

[15] Ukkola, A. M., De Kauwe, M. G, Roderick, M. L., Abramowitz, G., & Pitman, A. J. (2020). Robust future changes in meteorological drought in CMIP6 projections despite uncertainty in precipitation. *Geophysical Research Letters*, 46, e2020GL087820. 10.1029/2020GL087820.

[16] For an analysis for Sydney on this topic see Barker, A., A.J. Pitman, J.E. Evans, F. Spaninks, L. Uthayakumaran, (2020), Drivers of future water demand in Sydney, Australia: examining the contribution from population and climate change, *Journal of Water and Climate Change*, accepted 22 May 2020, doi: 10.2166/wcc.2020.230. While the impact of climate change was relatively small compared with the impact due to population growth, the climate change impact wasn't zero.

This work needs to examine vegetation's role in linking climate change with soil moisture and catchment yield.

There is ongoing discussion about the shortfalls if using PET in some modelling. This issue will be looked at in the Community of Practice and in longer term research.

2.4.2 Analysis of changing vegetation's effect on flow

The issue of longer-term projections was discussed by the Panel in the context of work being undertaken by the BOM on national hydrological projections and streamflow modelling. There is a fundamental change occurring across NSW catchments - the probability that a heavy rainfall event falls on a drier catchment is increasing. This is because the warmer atmosphere and warmer soils allow vegetation to grow for more days a year and "grow" means more transpiration, which dries the soil profile. One can see this in the greening of the satellite record - and every increase in leaf area index reflects embedded carbon resulting in higher net evaporation.

Therefore, irrespective of stationarity in the meteorology, whether or not rainfall is changing, the ability of a catchment to absorb its water is increasing and thereby reducing catchment yields. This may not have any effect on water storages that are filled by 1 in 3 year or 1 in 5-year events, but those dependent on smaller events, this could have a large impact.

Improving the modelling to account for these effects will require a longer-term focus and funding, which has not occurred to date in Australia. There may be an opportunity for NSW to take the lead on this issue.

2.4.3 Non-stationarity of climate drivers

While there may be a non-stationarity signal in the current climate, an associated issue is potential non-stationarity in the climate drivers, and the teleconnections between these drivers and Australian temperature and rainfall, which may already be changing and may change further in the future (e.g. Power and Delage 2018[17]; Delage and Power 2020[18]). The Panel discussed this in section 2.2.4 in the April 2020 report. This section is reproduced below:

> *However all regions in NSW will have contributions from multiple drivers and systems, so full explanations for climate's impacts on rainfall and ET won't be found in any single driver and it is inevitable that multiple interacting drivers could be involved (e,g, ENSO-IPO) in rainfall events and the replenishment of surface water bodies.*
>
> *In addition, there is uncertainty about whether and how climate change will influence climate drivers and their interactions. No matter how these major drivers may have acted historically, they may not act consistently in the future with a changing climate. Indeed, we do not know if they are stationary in the past record, as there is only a relatively small number (approximately 8) of IPO epochs in the measured records and more in the paleo-record.*
>
> *Therefore, considering both the historical record and future changes, there is a need to look at the effects of individual and combined (joint) drivers on rainfall. In relation to the RWS methodology, this could be particularly important for regions where multiple drivers could play important role (e.g. in cases where both IPO and ECL may play a major role in rainfall).*

---

[17] Power, S. B., & Delage, F. P. D. (2018). El Niño–Southern Oscillation and Associated Climatic Conditions around the World during the Latter Half of the Twenty-First Century. *Journal of Climate*,*31*(15), 6189-6207. https://journals.ametsoc.org/view/journals/clim/31/15/jcli-d-18-0138.1.xml

[18] Delage, François & Power, Scott. (2020). The impact of global warming and the El Niño-Southern Oscillation on seasonal precipitation extremes in Australia. *Climate Dynamics*. 54. 10.1007/s00382-020-05235-0.

The Panel believes work to improve stochastic series will be needed, including, but not restricted to, methods that use individual drivers, particularly for regions that may have a combination of dominant drivers. Indeed, for regions with multiple dominant drivers it may be better to explore methods that don't rely on specification of individual drivers, rather than add additional drivers into an existing method, with or without their possible nonlinear interactions. The exploration might find that using drivers is best way to go in future, but it might not.

Overall, the Panel agrees that work should be undertaken in the near term on what approach to use to generate stochastic data sets where multiple drivers dominate, and for future iterations of the RWS, given that climate change may alter the impact of drivers (non-stationarity), we suggest future research put effort into other ways to generate the stochastic data sets.

In the longer term, work with colleagues, such as the community of practice group to examine whether behaviours of climate drivers (e.g. IOD, IPO, ENSO, and SAM) into the future will remain statistically consistent with the past. Active reviews of the latest literature every few years would be prudent.

This issue will need to be considered in the model set up. However, if climate drivers are added to a stochastic model, it increases the complexity. This is because a probability model of the climate drivers is required and a description of the interdependence with the variables of interest (e.g. rainfall, streamflow). The challenge is to show evidence that doing this provides statistically significant improvements to the rainfall/streamflow simulations.

For example, DPIE-Water is using a rainfall/IPO stochastic model. The IPO model is treated crudely as an alternating renewal process (e.g. palaeo data showing long periods (70%) of IPO positive, 15% negative, 15% transition (uncertainty not known here)). Calibrating this model is problematic, as there is only a handful of IPO epochs in the instrumental record. Paleo data can increase this sample provided you trust the paleo reconstructions. The bigger problem is the small sample available to condition rainfall on IPO. Establishing that these distributions are different and stationary is a challenge.

The Panel notes that Victorian DELWP touches on this issue in DELWP (2020).[19]

The Climate Community of Practice, recommended in the April 2020 report, should address this topic. There is lack of consensus in the research community on the importance of climate drivers, with some seeing them as the pathway to predicting future climate and rainfall while others downplay their potential importance. The same is true for use of palaeo data, with some supportive that they provide robust reconstructions, while others are less supportive, in spite of journal articles to the contrary. The topics are complex and subject to ongoing research. But the issue of whether past correlations between rainfall and climate modes are stationary and whether there is any correlation between rainfall events that matter to water security and climate modes are vitally important topics and worthy of future study and consideration.

The Panel recommends that DPIE should keep abreast of advances in the relevant climate driver/teleconnection research to ensure that approaches used reflect scientific consensus on non-stationarity of climate drivers. This may be addressed through the previously recommended Community of Practice.

[19] DELWP. Guidelines for Assessing the Impact of Climate Change on Water Availability in Victoria. Final, November 2020, Department of Environment, Land, Water and Planning, Victoria.

3 COMMENT ON APPLICATION OF NON-STATIONARITY TO STOCHASTIC MODELS

Overall, the Panel supports proceeding with the two proposals for Southern Basin by UoA and Greater Sydney Region by UoN. The Panel recognises some of the issues raised here are complex and challenging to resolve. Some of the key issues are the subject of cutting edge, developing research, and strategies to quantify climate risk will require ongoing attention into the future and likely adjustment as understanding improves. This advice elaborates on the issues and suggestions to address them. The Panel acknowledges that some of these will require a longer timeframe to resolve.

There was consensus that there is no definitive approach to account for non-stationarity in the observed record or future climate. The goal here, ultimately, is to estimate catchment yield which requires a series of steps which inherently involve a cascade of uncertainty. So, overall the Panel cannot assess whether the approaches being proposed are “right”, but rather, whether they are reasonable.

Panel discussions were focused around the options for addressing issues, including non-stationarity, in stochastic modelling, both benefits and shortcomings, and arriving at a strategy that makes best use of what is known and communicates the uncertainty.

3.1 CHOOSING THE APPROACH/MODELS

3.1.1 Options to deal with non-stationarity in stationary stochastic modelling

Stationary stochastic models are designed to simulate stationary climates. There are numerous ways to deal with non-stationarity in stochastic modelling – one approach is to amend a stationary stochastic model or otherwise use a non-stationary stochastic model (Section 3.1.3). The approach put forward here in DPIE-Water material is the former, applying NARCliM 1.5 scaling factors.

The UoA proposal provided five potential approaches to deal with non-stationarity in stochastic model generation (Figure 2). The UoN proposal stated the approach will aim for consistency with the UoA approach noting the outcomes of the pending Panel advice.

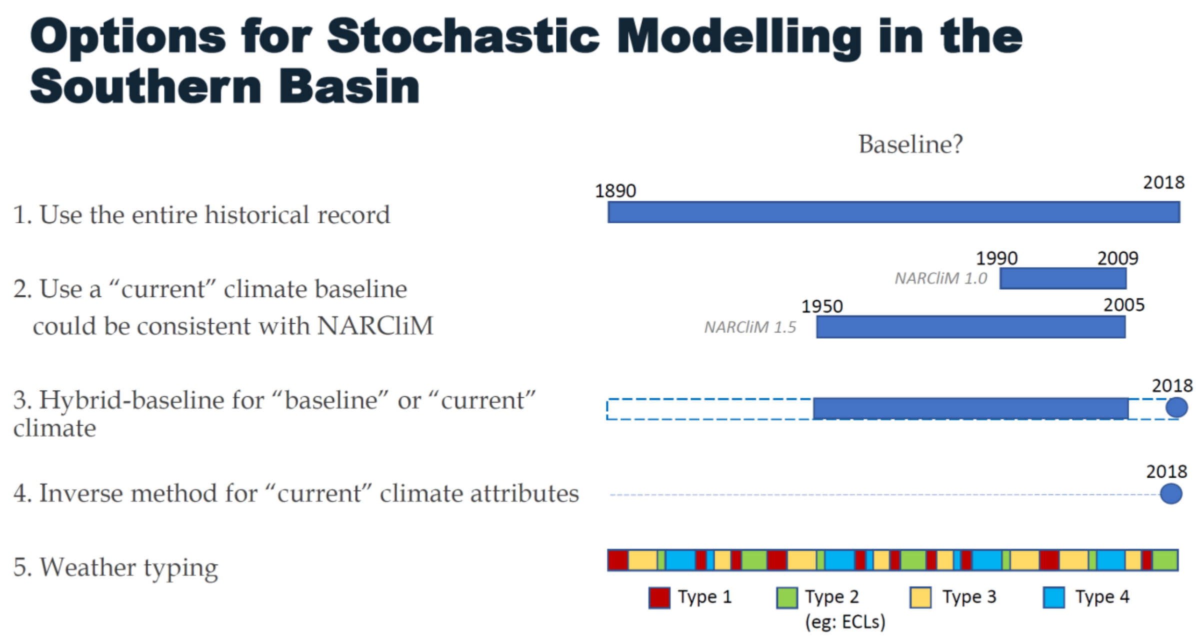

Figure 2: Outline of options

(Source: Westra presentation to panel 19 January)

The approach to date used by DPIE has assumed climate stationarity. This approach, as well as the rationale and complexities for modifying to incorporate non-stationarity, are described by UoA.

> *The stochastic method that has been used for the northern basins is based on the historical record and simulated to be conditional to the Interdecadal Pacific Oscillations (IPOs), with the implicit assumption that the climate is stationary over this period. These simulations are intended to capture variability in the baseline historical record, which was composed of (SILO infilled) climatic measurements from the period 1890 to 2018 as well as IPO distribution dwell times informed by paleoclimatic records. DPIE developed an additional methodology to amalgamate the variability from the stochastic replicates with climatic signals informed by RCMs. This method applies NARCliM based scaling factors to the stochastic model outputs.*
>
> *There are a number of trade-offs associated with modifying this methodology to incorporate climate non-stationarity, balancing the complexity of the climate signal, the strength of various non-stationarity elements as compared to the underlying variability, and the variety of methods that may be available to explicitly accommodate known trends. Here we detail a range of options available to account for the presence of non-stationarity in stochastics time series generation. The options vary in complexity.*[20]

The trades-offs to each option presented by UoA are noted in Table 2.

[20] Devanand, A., Leonard, M. and S. Westra (2020). *Implications of Non-stationarity for Stochastic Time Series Generation in the Southern Basins*. University of Adelaide (draft report), pg. 37.

Table 1: Trade-offs and application of five options outlined by UoA to account for non-stationarity
(Source: Westra presentation to panel 19 January)

Option	Advantages	Disadvantages	Application
1. Entire record	• Long record, enables increased precision of parameter estimation	• Neither reflective of current climate nor potential climatological baselines	Stationarity reflection of long term historical climate
2. Historical baseline	• Matches NARCliM baseline and thus can be used as the basis for 'current' and 'future' climate assessments	• Need to choose which NARCliM, noting that NARCliM 1.0 is too short relative to WMO & VicCI recommendations • Not reflective of current climate and thus requires some level of subsequent processing • Simple seasonal scaling would not reflect changes in wet days and other statistics, although there may be alternative (e.g. quantile-based) methods that could address this issue	NARCliM-derived 'current' and 'future' climate applications
3. Hybrid historical baseline	• Uses full historical record • Possible extra precision in some attributes compared to baseline only	• Advantage of extra data before baseline is unclear • There are multiple complicating factors associated with implementation of this method, including complex variations with key attributes, and the need to separate the natural and anthropogenic components of any historical trends prior to adjustments	Adjustments would enable stochastic data to reflect baseline or 'current' climate
4. Inverse method	• Can be implemented to match current and future climate • Allows changes in all key attributes	• Currently designed to represent 'point' rainfall and thus would require augmentation to account for spatial dependence. • Method development and testing required	Could be designed to reflect 'current' and 'future' climate
5. Weather typing	• Allows for specific mechanisms (e.g. ECLs)	• Additional complication introduced by multiple types • Scaling is complicated by types, and there exists a significant possibility of 'double counting'	Specific allowance for future changes to weather types rather than simple scaling

3.1.2 Comment on Option 3 – historical hybrid baseline

Overall, the Panel suggested a precautionary approach is taken in that if there is an error, it is better to over-estimate risk of drying than under-estimate it. The Panel is also supportive of using multiple 'storylines' to test different scenarios.

Panel consensus is that there is no definitive approach to account for non-stationarity in the observed record or future climate.

The Panel recommends Option 3 – hybrid-historical baseline – as a worthwhile option for both the UoA and UoN work. This uses a stationary stochastic model with a representative stationary climate record. There are two possibilities: fit to baseline record (which is assumed to be sufficiently short to be near stationary) or adjust pre-baseline record to match certain statistics in the baseline record (as done by Melbourne Water) and then fit to longer record.

Some Panel members have concerns about the ability of the baseline approach to adequately describe current climate risk. Hence the Panel supports a multi-pronged approach, particularly where approaches have different conceptual foundations.

The Panel noted that split sample tests have confirmed recent non-stationarity in the Southern Basin rainfall as well as Sydney streamflow records. To represent this recent non-stationarity the baseline record needs to be short, otherwise if too long, the change will be averaged out. However, a short baseline comes at the price of sampling uncertainty or noise regarding what is the actual climate representing the present. As a result, there is the risk that the baseline scaling is unrepresentative of the current climate.

The other four options from 3.1.1. were deliberated, with the Panel noting:

- Option 1 (entire record) risks providing an overly optimistic assessment of current climate risk and was not a viable option if there is a conclusion of non-stationarity for a particular region. As the Panel was comfortable with the UoA conclusions around non-stationarity in the Southern Basin, this option was excluded for that region. It is also expected that the Sydney region has experienced a non-stationary temperature signal and also for streamflow.
- Option 2 (historical baseline) was not the preferred option. The question arose, if NARCliM 1.5 is applied, what would this mean?

- Option 4 (inverse method) was not viable in the available timeline.
- Option 5 (weather typing) was also not viable in the available timeline, although the Panel noted that it has merits for the Greater Sydney catchment (discussed in section 4.1.1).

3.1.3 Another option - non-stationary stochastic model

A sixth option was noted at the meeting on 4 December 2020, which is to condition the model parameters on external covariates, such as time or temperature. This option was not discussed in detail in the Panel meetings. However, recent work on this topic has been published by Kiem et al (2020)[21] and is described briefly here.

A non-stationary stochastic model can be implemented by conditioning its parameters on an exogenous variable which is associated with the change signal, both past and future – the exogeneous variable could be time itself or some climate-related variable. Such a model avoids the need to adjust observed data and the definition of reference periods. However, it requires identification of a climate variable that is meaningfully associated with the climate change signal, has long and reliable records and is estimated with reasonable skill by climate models.

There is limited literature using this approach and little practical experience. However, the recent work of Kiem et al. (2020) is briefly mentioned because it has been applied to the Sydney catchments with satisfactory results – for instance, the split sample results shown in Figure 1 for a stationary model are significantly improved with p values increasing from <0.001 to ~0.13. One immediate implication is that the model can represent current climate risk.

The Kiem et al. (2020) approach uses a simple generalisation of a long-established stationary stochastic model that is applied at seasonal or annual time scales. It uses maximum temperature as the exogenous variable. Although there are significant homogeneity issues with long-term maximum temperature records, stable and useful associations can be found in some cases. In cases where the association survives diagnostic scrutiny (in the form of multi-block split-sample analysis), it offers an additional line of evidence to the approach based on stationary models.

In a non-stationary stochastic model the same principles can be used as with the stationary stochastic models to generate different probability distributions that describe the uncertainty in the way the non-stationary process evolves over time. In that regard the intent to describe uncertainty in climate is similar to the baseline approach. The big difference is that the uncertainty is data-driven, which is both a strength and liability. This uncertainty is straightforward to evaluate and is illustrated in Kiem et al (2020).

For UoN's proposed approach in the Greater Sydney region, use of a nonstationary stochastic model approach is a worthwhile option. UoN is already capable of applying such a model to streamflow in Sydney region. It is recommended UoN's remit be extended to applying this approach and comparing it with the baseline approach.

3.1.4 Choice of NARCliM 1.5

As described above, there is a range of approaches that could be used to incorporate non-stationarity into a stochastic data model. One set of approaches uses climate models to scale historical stochastic data and future stochastic generated data for a location or region. In some cases, global climate models (GCMs) have been used directly (e.g. the DELWP modelling for Victoria) to reflect current and future climate. The Panel members had mixed views on using GCMs directly for local modelling.

---

[21] Kiem, A. S., Kuczera, G., Kozarovski, P., Zhang, L., & Willgoose, G. (2020). Stochastic generation of future hydroclimate using temperature as a climate change covariate. *Water Resources Research*,57, 10.1029/2020WR027331

In other cases, where regionally downscaled climate modelling is available these can be used to reflect recent or future climate change, in NSW's case, by applying NARCliM-derived scaling factors to stochastic simulations.

In NSW, with the availability of NARCliM that covers the whole state and the Murray Darling Basin, a library of downscaled models is available. NARCliM is not a single model, but rather an ensemble that includes alternative representations of some physical processes, as described further in the Panel's April 2020 report. NARCliM is currently going through a set of updates, with NARCliM 1.5 now available and NARCliM 2.0 underway (planned completion 2022).

Certain characteristics of NARCliM 1.5 mean that it is preferable to use in the current project on non-stationarity in stochastic models compared with its predecessor, referred to as NARCliM 1.0. Using NARCliM 1.0 in this project is possible, but it has a shorter baseline simulation. However, there may be useful information in NARCliM 1.0 if using a pooled approach (i.e. combining all modelled years from NARCliM 1.0 and NARCliM 1.5), adding additional information to the more recent NARCliM 1.5. A comparison of the two versions is in Table 3.

The attributes of NARCliM 1.5 that are of particular relevance for this project include the following:

- NARCliM 1.5 provides 150-year simulation of physically plausible climate (1950-2100).
- NARCliM 1.5 reflects large scale drivers of rainfall in the Southern Basin (including expansion of tropics, shift of storm tracks). It gives a better view of the low rainfall/drought risk over NARCliM 1.0, as the CMIP5 models perform better than CMIP3 models, and NARCliM 1.5 samples a 'hotter' future change space that NARCliM 1.0 doesn't.
- In setting up NARCliM, climate model drivers were intentionally chosen to be relatively wet, relatively dry and medium; an approach that lead the downscaling exercise to 'bookend' the Coupled Model Intercomparison Project (CMIP).
- For NARCliM 1.5, two configurations of the Weather Research and Forecasting Models (WRF) (version 3.6.0.5) regional climate model (RCM) were used to downscale projections from three CMIP5 global climate models (GCMs), with two future emissions scenarios (RCP4.5 and RCP8.5). In effect, for each 20 years of time generated using NARCliM1.5, there is 240 years' worth of data for physically-plausible future climate. For the 20-year periods also modelled for NARCliM 1.0 (2020-2040, 2060-2080), NARCliM now provides 480 years of physically-plausible future climate.
- NARCliM isn't one model, but rather an ensemble of models. Pooling all NARCliM models for one 20-year time slice equates to 480 years of model data representative of the 20-year period. Considering the NARCliM ensemble as a whole, NARCliM has dynamically downscaled approximately 1,500 years of future climate data. The question is how best to mine and use that data to align with the stochastic model. If focusing on a 20-year period from 2060 – 2080, there are 24 physically plausible replications of that 20-year period – plus 100 years of observational data.
- Using WRF as the only downscaling model could introduce shared biases, so to overcome this NARCliM developers used three different convection and three different planetary boundary layer (PBL) parameterisations to examine dependencies but using the same dynamical core.
- For NARCliM 1.5 the chosen models for downscaling were selected on the basis of being independent using the Bishop and Abramowitz method.
- For historical climate data using NARCliM, if the whole period from 1950 – 2009 is considered, then there are three models (based on the CMIP5 historical simulations), two for the period from 1980 – 2005 (based on ERA-Interim reanalysis), and three for

the period 1950-2009 (based on NCEP/NCAR reanalysis), providing eight historical simulations in total.

- All data provided by NARCliM 1.5 is stored as unadjusted, but for some parameters bias corrected data is also available. Note that while bias correction can improve model performance compared to current observed data, the bias correction process can also introduce its own uncertainties.
- NARCliM1.5 is uploaded and available to be accessed by UoA and UoN on the National Computational Infrastructure (NCI).

Table 2: Comparison of features of NARCliM 1.0 and 1.5

	NARCliM 1.0	NARCliM 1.5
Models	12	6
Model timeframes	20 years 1990-2009, 2020-2039, 2060-2079 (& 1950-2009 NCEP-forced simulations)	150 years 1950-2100 (& 1981-2010 ERA-Interim forced simulations)
Grid	50 km (CORDEX-Australasia) & 10 km (NARCliM)	50 km (CORDEX-Australasia) & 10 km (NARCliM)
Timestep	3-hourly (all vars), hourly (precip, 2m temp, RH, wind)	3-hourly (all vars)
Variables	100+ (20 post-processed)	120+ (25 post-processed: 1-hr, 3-hr, 6-hr, daily, monthly, season)
GCMs	4 CMIP3	3 CMIP5
Regional models	3 regional models per GCM (WRF3.3)	2 per GCM (WRF3.6.0.5)
Future emissions scenario(s)	SRES A2	RCP4.5 & RCP8.5
Example uses	Regional climate snapshots, near versus far future climate analyses for temperature, heat, snow, fire, rainfall, etc. Strategic planning (e.g. State Infrastructure Strategy, Sydney Region Plan). Transport for NSW Asset Management Authority – guidelines for rolling stock. Asset risk management (XDi). NSW Common Planning Assumptions. NSW NPW adaptation strategy, etc.	**In addition to previous iteration:** climate extremes, thresholds for impacts, compare with non-climate datasets.

3.1.5 Overview of Victorian Government work on climate modelling for water planning, including non-stationarity timeframe post-1997

In November 2020, the Victorian Department of Environment, Land, Water and Planning (DELWP) released guidelines to assist Victorian Water Corporations and agencies to incorporate climate change into planning for water supply sustainability in Victorian urban and rural water. The Guidelines set out an approach for applying climate change scenarios for temperature, PET, rainfall, runoff and groundwater (GW) recharge and include changes to climate variability associated with climate change.

A brief summary list of the key DELWP Guideline approaches is as follows:

1. A post-1975 historical climate baseline period is used, as well as a post-1997 historical reference period, from which to generate a step climate change scenario, to capture non-stationarity.
2. Extending the post-1975 and post-1997 periods to incorporate a broader range of natural climate variability using either scaling techniques or stochastic data generation.
3. Use of additional scenarios to represent low, medium and high current and future projected climate and streamflow (in addition to the post-1997 scenario). Scenarios are derived using average annual climate change projections applied to the post-1975 period for temperature, PET, rainfall and runoff under the RCP8.5 scenario for the years 2040 and 2065. The low, medium and high represent the 10th, 50th and 90th percentile from the 42 available GCMs.
4. Linear interpolation of low, medium and high projections between 1995, 2040 and 2065, and extrapolation of GCM projections up to 2075.
5. Post-1997 step change climate scenario (in addition to low, medium and high) to represent current and future climate and streamflow – projecting the post-1997 reference period independent of climate modelling. The scenario assumes the dry conditions post-1997 are a permanent step-change in climate compared to pre-1997.
6. Adjustments to peak daily and sub-daily rainfall and run-off and changes to average annual rainfall and run-off to project any projected changes in rainfall intensity under global warming.
7. Applications of climate change scenarios to estimate change in water use demand, GW recharge, additional changes to run-off due to changes in snow cover, changes in water supply availability, such as due to changes in yield, reliability etc.
8. Identification of near-term risks.
9. Optional application of additional information and techniques including using also RCP4.5, optional use of regional downscaled projections (from the Victorian Climate Projections 2019), and optional use of additional climate change impact assessment techniques such as sensitivity testing. [22]

The Victorian DELWP guidelines will be updated after the CMIP6 climate modelling is published in 2022.

Lessons for NSW from the Victorian approach include the following observations:

- *Detection of non-stationarity* – Given the overall comfort of the Panel that the work of the UoA team in the Southern Basin indicates that there is evidence of non-stationarity in the data, the Panel then considered whether it is feasible to determine the relative importance of anthropogenic contributors to non-stationarity versus natural contributors. It is noted that in the Victorian DELWP work, the observed "externally-forced" component (which includes the important anthropogenic component) was approximately 20% of the observed change since the Millennium Drought, but with very high uncertainty (for example, interquartile range is 40% to -4%). As there is a large degree of uncertainty with this estimate, an approach is required to manage uncertainty. The approach used by DELWP was to use a range of projections – 10%, mean (i.e. 50%), 90%, derived from the GCM runs.
- *Model set-up and scaling factors* – The DELWP approach used as a baseline from 1975 onward to near-present. It used stochastic modelling of that period and applied scaling factors based on global, coupled GCM outputs to stochastic variability so the stochastic data was run once, but three sets of scaling factors were then applied (10%, median (i.e. 50%), 90%) over all the state at the catchment level, giving three scenarios. A fourth scenario was obtained by using data from 1997 onwards only, independent of climate model output. The Panel thought that reflecting uncertainty through the selection of low,

[22]DELWP. *Guidelines for Assessing the Impact of Climate Change on Water Availability in Victoria.* November 2020, Department of Environment, Land, Water and Planning, Victoria.

medium and high climate change scenarios was prudent, and noted this was similar to the approach used in NARCliM. Having a baseline for Southern Inland NSW that aligns with that for Victoria could have benefits for the ability to compare the two regions.

- *Model performance and extremes* – The approach of fitting the model to the entire record and then fitting to a post-change period (for Victoria it was 1997) is a good one. It provides an idea of how climate has changed – seen in model parameters – allowing comparisons of frequency and length of dry spells. However, the issue of dam filling events requires a forensic look at the records. Some parts of NSW rely on these events, so if NSW had that insight it might give a better idea of model performance. If more metrics are brought in, then it will be possible to determine whether models were performing well or not. Having the metrics that impact water availability would give a clearer picture of how well the model performs for water security and this exercise would help identify the relative performance of different models. The DELWP work dealt with extreme drought and rainfall by undertaking extreme event sampling (see DELWP report). Refer to section 3.3 for more on this topic.
- *Regional downscaled models versus GCMs* – There are attributes from this DELWP approach that work well, but there are some caveats for NSW. First, the GCM models used have very coarse resolution and do not resolve what happens at the catchment level. Furthermore, the evaluation of extreme events over NSW is limited.
    - There are benefits to adopting a systematic downscaling approach, which is what NARCliM has done. NARCliM 1.5 can be used for drivers of variability. How well downscaling products reflect rainfall or drought events is an important question for water security. They will be part of the answer but not a silver bullet.
    - There is concern by some Panel members of the use of the 10th, 50th and 90th percentiles from the 42 available climate models. If these 42 models were independent then it might be ok, but they are not. This means that the percentiles are merely the statistical artefacts of a biased sample. There is no "reason" why these 42 models exist - it is an "ensemble of opportunity" and not an ensemble designed to sample uncertainty.[23] So, NSW should not follow this method without first undertaking a formal assessment of independence and only using those GCMs that are independent.
    - For NSW, NARCliM has worked through these issues. NARCliM tested for model independence when selecting those to downscale. NARCliM also picked, in effect, a wet, dry and intermediate climate model to downscale. Unless it is possible to demonstrate *a priori* that the NARCliM data are biased it is strongly recommended not expanding the ensemble without very great thought – adding more models would increase the range of projections, but not necessarily usefully.
    - It should be noted, however, that model dependence might be less of an issue for rainfall projections (e.g., Power et al. 2012) and that some, but not all, Panel members saw value in augmenting projection information (e.g., scaling factors) from NARCliM, with information from the global GCMs, using a similar approach to that adopted by DELWP.[24]

---

[23] Model co-dependence is an issue with GCMs. Power et al. (2012) introduced the concept of "multi-family mean change" (see e.g. Fig. 11 in the 2012 paper) as a way of addressing model dependence. As there aren't many families, in order to reflect model uncertainty in projections, one might have to resort to something ad hoc like e.g. using largest and smallest, or 2nd highest and 2nd smallest.

Using e.g. 10th percentile of CMIP5 ensemble is likely a biased or at least inaccurate estimate of 10th percentile of multi-family ensemble. It would be interesting to see what scaling factors are for each family and would seem to be better to use this distribution to estimate uncertainty in scaling factors. But others might argue that it might not be a major issue (especially for rainfall projections – see Power et al. 2012) and the approach taken is adequate, given other uncertainties. It is in fact the approach taken by IPCC (2013).

[24] Note there are other ways of testing for and dealing with model dependence. See, for example:

3.1.6 Storyline approach

The Panel was supportive of a conservative approach using ‘multiple lines of evidence’, as well as a ‘storyline’ approach, where consideration is given to the whole system and how it interacts (e.g. urban water supply vs irrigation valley). By developing a conceptual storyline for that catchment, components of the model can be designed and applied to reflect the conceptual understanding of the system.

A storyline would be informed by the characteristics of valleys such as their location (coastal vs inland), dependent water users (e.g. urban vs irrigation; large or small dependent), regulated vs unregulated waterways, connectivity and role of groundwater resources, agricultural, forestry or irrigation types, etc.

The complexity of choosing the “best” design highlights the value of the storyline approach, informed by NARCliM, and informed by CMIP6 models, as appropriate. Among those storylines should, without question, be a strong drying trend in the Southern Basin slowly expanding northward into the Northern Basin. For the Sydney Basin, the ‘best’ storyline is to explore sensitivity to a range of East Coast Lows as proposed in earlier analyses and proposed by Kiem based on earlier work.

This approach links to the discussion in section 3.3.2 about what generates streamflow and dam recharge. Metrics for the model should be defined around ‘what matters’.

3.2 MODEL SET UP: VALIDATION, SIGNAL TO NOISE, BASELINE, LIMITING CASES, SCALING

3.2.1 Validation of NARCliM model outputs to statistics of hydrological importance

From a stochastic modelling perspective, the baseline approach remains problematic, as an issue with the proposed option 3 is how representative of past records are NARCliM baseline simulations (unforced by historical observations)? Before this approach is undertaken, it needs to be ascertained how well NARCliM 1.5 historical runs capture the statistics of variability in rainfall, of relevance to streamflow, over the regions of interest in the Southern Basin and Greater Sydney. This should be checked.

The degree to which NARCliM 1.5 can be used will be partly dependent on how well the simulations capture the historical period in a statistical sense. They will not likely match observations in terms of the timing of variability, but they might capture the statistics of variability and any change due to increased CO_2.

This is important because, as the Panel was advised, the change factors will be derived from NARCliM simulations of historic baseline and future epochs.

To understand this question, it would be prudent for the UoA team to analyse the NARCliM time series (using trend analysis/split sample analysis) to compare against the historical record (rainfall/streamflow), comparing key measurement statistics, trends and key variables. This could be, for example, an assessment of the probability of the Millennium Drought under alternative stochastic replicates (calculated using different baseline periods).

- Abramowitz, G. and Bishop, C.H. (2014). Climate Model Dependence and the Ensemble Dependence Transformation of CMIP Projections, J. Climate, 28, 2332-2348, doi: 10.1175/JCLI-D-14-00364.1
- Bishop, C.H. and Abramowitz, G. (2013). Climate model dependence and the replicate Earth Paradigm, *Climate Dynamics*, 41, 885-900, doi:10.1007/s00382-012-1610-y
- Power, S.B., F. Delage, R. Colman, and A. Moise (2012). Consensus of 21st century rainfall projections in climate models more widespread than previously thought. *J. Climate*, 25, 3792-3809, DOI: 10.1175/JCLI-D-11-00354.1.

The selection of the specific regions and the key measurements selected for comparison can be informed through process of ‘storyline’ development.

Of importance is whether the simulations are consistent with observed statistics and trends that affect storage behaviour. Persistence in time and space would be an important attribute. It is not “consecutive dry/wet days” or “RX1” [the wettest day of the year] in all catchments. Sometimes it will be the event that occurs once every three years, or it will be some sequence of wet days that are not consecutive. Two aims of this are to identify how reliable and useful NARCliM is, as a due diligence assessment, and also to define benchmarks for NARCliM in the future – informing the value of NARCliM for specific purposes.

Key to this, and more generally, is for the streamflow modellers to inform the climate modellers of the characteristics of events that matter to them. This could be different for different catchments, depending on their location (coastal vs inland), dependent water users (e.g. urban vs irrigation), regulated vs unregulated.

Testing the NARCliM outputs to the statistics of specific catchments provides an indicator for the confidence that the climate variability signal has been captured and can help recommend the distribution of possibilities for scaling factors from NARCliM, providing clarity on what parameters would be suitable for modelling such as rainfall, temperature, soil moisture, etc.

However, there may be a question about the ability of climate models to capture more extreme events which are important for water security. Regional Climate Models (RCMs) are useful but may not be a silver bullet for this issue. A particular question remains about whether another or an additional method would need to be considered for drought and extreme rainfall, including East Coast Low dam-filling events?

In relation to extreme events and validation of the NARCliM model to these events, the Panel recommends that the developers of NARCliM work with hydrology and hydraulic modelers to identify some key metrics that could be used to validate NARCliM against and to give confidence and clarity on this question. NARCliM developers can then build these metrics into the NARCliM data flow, to build in some aspect of hydrological assessment and validation into the project’s post data processing. Relevant metrics may be regionally-specific, or an array of metrics may be developed to choose from, with the methodology to apply these techniques documented. This task would be challenging, but worthwhile. Informing the NARCliM user community about which metrics to refer to for a particular catchment and what the scale and range of the metrics should be, for the data output to be useful (e.g. could be variability in rainfall), would be interesting to know what that is for a set of catchments and judge accordingly.

The NARCliM 1.5/2.0 Technical Reference Group, in which DPIE Water participates, may be the suitable mechanism for this work to be undertaken. It will be important for the metrics to be documented, agreed and peer reviewed.

The Panel notes that our ability to accurately know long-term basic statistics of crucial hydrological variables is very limited due to the brevity of historical records.

The Panel acknowledges that the recommended work requires a longer-term effort and could take 12-18 months to determine and develop the methodology. There is likely to be a suite of candidate metrics that work could commence on. The approach to choosing / evaluating metrics would require guidance from experts and stakeholders in the hydrological community. It would be trivial to add a dozen or fewer metrics to the coding, as coding for a couple of hundred metrics occurs now. The Panel is advised that DPIE Water has started some work on this, and an evolution of that would be required.

3.2.2 Signal to noise ratio and sampling uncertainty

A stationary model is fundamentally incapable of quantifying the relative contribution of natural variability and climate change to an observed change. A climate model (such as NARCliM), on the other hand, is capable in principle of such quantification.

The ideal way to accomplish such a separation would be to run many ensembles (or replicates) of the climate model over the same historical period with both natural and anthropogenic forcing, and then conduct two additional sets of runs in which only anthropogenic and then only natural forcing is applied. The historical record could then be detrended and used to calibrate a reference stationary stochastic model.

The proposed approach using NARCliM relies on identifying a reference (or baseline) period and then providing some mechanism for comparing future climate periods against the baseline. The stochastic model could be calibrated to the baseline historical period. Hydroclimate sequences generated for the baseline period could then be adjusted in a manner that captures the changes in key statistics between NARCliM simulations of the reference and future periods.

The issue of signal to noise ratio is very important to this and confidence in this approach will largely depend on the ability to manage the signal-noise ratio. The baseline period chosen (next section) will affect the signal to noise ratio.

If the rainfall response to anthropogenic forcing is regarded as the signal and internally generated rainfall variability is regarded as the noise, then the ratio is smaller for coming twenty years than it is later in 21st century under e.g. RCP8.5 (as the climate change signal increases). Clearly advice for the 21st century needs to factor in what natural climate variability might do about these projected changes. This is evident in Rauniyar and Power (2020).[25]

Changes in key statistics between NARCliM simulations of the reference and future periods will be affected by both the underlying change signal and sampling error. How big does the change signal have to be in order to "break through" the sampling noise? To put this question in perspective, it is worth noting that the annual average inflow into Warragamba reservoir is only known within ±20% with 95% confidence even though there are over 100 years of records. So how much useful information about change is conveyed by NARCliM simulations of the reference and future periods?

Given high natural variability, any statistic used to derive change factors to scale stochastically generated data may be subject to considerable sampling uncertainty (which decreases proportionately to the square root of record length). The problem is that the climate change signal implicit in the change factors may be masked by sampling variability. It is prudent to make an effort to quantify this and use it to inform decision making and selection of the baseline period.

3.2.3 Choosing the baseline length and starting year

Because the approach is to calibrate stationary stochastic models, the data used needs to be adjusted so it represents a stationary snapshot in time. Choosing a baseline starting point is complex.

One of the many problems is landing on an appropriate baseline when the baseline may already be changing. The 'change' is more likely to be a trend of some kind rather than one or more sizeable break points at specific instances in time. As already noted, the 'change'

---

[25] Rauniyar, Surendra, and Scott B. Power (2020). The Impact of Anthropogenic Forcing and Natural Processes on Past, Present, and Future Rainfall over Victoria, Australia. *J. Climate*, 33, 807-8106, 10.1175/JCLI-D-19-0759.1

may be small relative to natural variability, but the direct and indirect impacts may be more noticeable.

The chosen baseline can either be short for calibration or a longer adjusted record (adjusting pre-baseline data to be consistent with baseline data, as done by Melbourne Water).[26] There are trade-offs with each of these options.

The longer the period, the more precise the calibration and the less sampling error in calibration and in statistics used to compute change factors. However, the longer the period, the stronger the change signal; and the historical baseline may have significant non-stationarity.

The Panel deliberated changing the baseline. One option was a longer baseline of 1950 – 2005, which would err on the wet side and likely result in overestimation of rainfall. Another option was to choose a shorter baseline to post 1997, which may be too conservative (dry) to somewhere in between, noting it's best to err on the dry side. The assessment of post 1997 as too dry was debated by Panel members, as there is no straightforward way to assess whether post-1997 is too dry, too wet or representative.

In the southern basin there is limited evidence to suggest that using post 1997 is "overly pessimistic". Current ability to quantify the relative importance of anthropogenic and natural processes to observed drying records or recent multidecadal changes in Australia is limited, which further complicates the ability of current science to estimate future rainfall change. There is evidence the drying signal is related to large-scale drivers including the expansion of the tropics associated with global warming. It nevertheless seems prudently cautious to assume that some fraction of the observed drying will persist and possibly intensify into the future and is therefore not 'overly pessimistic'. The northern basin is not drying to the same degree and given the lack of evidence for non-stationarity one could choose a longer baseline.

Victorian and Western Australian water corporations adjusted their pre-1997 flow data to reflect key statistics of post-1997 using a 'pessimistic view' (Melbourne Water) as the Millennium Drought forms part of that signal, which in part reflects natural variability, so would calibrate to a baseline that likely reflects both a change signal and some element of natural variability.

Victoria DELWP chose 1975 as baseline starting point, with some Panelists noting merit in NSW being consistent with the DELWP approach, particularly for the Southern Basin enabling ease of comparison across the border. On the other hand, some Panelists were uncomfortable with using 1975 as a starting point, as there is a significant change signal over that period. Until the trade-offs are better understood, the Panel cannot recommend a specific baseline period. The Panel overall suggested a precautionary approach in that if we make an error it is better to over-estimate risk of drying than under-estimate it.

The rolling evaluation approach (Appendix 4) (or something similar) might provide insight into the temporal stability of simulated streamflow and hence changes/trends in hydroclimatic regime. There is no guarantee that any such change will occur simultaneously across the regions or at the same rate.

3.2.4 Estimating upper and lower limiting cases of climate risk on water security

The current approach develops one stochastic single realisation (single-best method) to enable probabilities to be estimated. Probabilities are conditional on assumptions. A scenario-based approach would provide a different decision-making route – providing upper

[26] DELWP. Guidelines for Assessing the Impact of Climate Change on Water Availability in Victoria. Final, November 2020, Department of Environment, Land, Water and Planning, Victoria.

and lower cases to stress test the system. The simplest approach to calibrate to a climate change signal relies on bounding.

As it is not complex to undertake, the Panel suggests upper and lower cases are estimated as a ‘reality’ check on other approaches. Note this is the approach used by NARCliM.

Note this approach will apply to temperature and rainfall, but not drought or extreme dam filling events (e.g. ECLs).

Lower limiting case

Calibrating a stationary stochastic model to the full observed record (1910 – 2020) without adjustment ignores the drying change signal and thus will provide an underestimate of current drought risk.

Upper limiting case

The upper bound on current drought risk can be estimated in two ways:

1) Calibrating to the recent past record, from 1990 - 2020. While this will capture the statistics of the last 30 years, the stochastic model will be sensitive to sampling variability. A caveat is that this approach may not apply to all regions and will depend upon the relative importance (and sign) of natural and anthropogenic processes and what is expected to happen. What happens, for example, in a region where rainfall increased since 1990 but projected change is drying?
2) Calibrating to the full observed record with constraint on selected statistics. Melbourne Water have adopted this approach adjusting pre-1997 records to match post-1997 quantiles. UoA’s proposed inverse method is more flexible in that it gives more freedom in specifying the constraint; however, there is insufficient time to develop and evaluate this method and identify what are the most appropriate constraints.

If emphasis is placed on reproducing post-1990 statistics, which are affected by the Millennium Drought, it is likely that the overestimation of current drought risk may be quite severe.

While the upper and lower case approach is intuitive, it could be of limited practical value because the range it produces on simulated climate risk may be too wide. However, as the approach is straightforward and uncontroversial, it should be undertaken if only to provide a check on other approaches.

3.2.5 Use NARCliM to scale the stochastic data

In addition to estimating the bounds described above, the panel suggests using downscaled data available to NSW from NARCliM 1.5 to estimate scaling factors.

The teams will need to look at NARCliM 1.5 first before deciding on seasonal or quantile scaling.

Sampling the NARCliM 1.5 data runs using different 20-30 year blocks will provide a range of scaling factors which can then be applied to the stochastic data.

The risk with the approach of relying on short blocks of data, real or simulated, is that the change signal may be dominated by sampling error noise. Investigating the signal to noise ratio should be done prior to assessing options regarding scaling.

The available climate data could be used to understand the statistics, developing a set of scaling factors (distribution of factors), using a range of possibilities. In the case of Victoria DELWP, they used scenarios and apply scaling factors from GCMs (e.g. 10%, median, 90% for RCP8.5 and RCP4.5).

The scenarios also touch on how to handle extreme events. While this approach may not work out the ‘best’ answer, it will reflect the range of uncertainty.

Chapter 4 of the Victorian DELWP report may be a useful reference for the methods being described in this section.

3.2.6 Choosing projected output timescale

If assessments are needed for years in between future multidecadal blocks (over short time periods), consider using linear interpolation as per the DELWP approach.

The DELWP approach also uses extrapolations to expand the time frame out beyond 2070, although the NARCliM 1.5 data set goes out to 2100, so extrapolation would not be required until after 2100.

3.3 TEST/VALIDATE THE DPIE STOCHASTIC MODEL USING STREAMFLOW DATA

3.3.1 Issues in testing models with rainfall/ET

The Panel expressed discomfort with testing DPIE stochastic models purely in the rainfall/evapotranspiration space, noting that the end-game for the use of the datasets is surface water management, therefore testing in the context of streamflow is also important.

The presumption is that routing stochastically generated rainfall and ET through a calibrated rainfall-runoff model will produce streamflow series whose statistics are consistent with observed streamflow records.

This remains an untested assumption. Given that the input required in river basin simulation is streamflow, it is critical to the integrity of river basin modelling process that this assumption be tested.

Therefore, the Panel recommends the use of a stochastic rainfall and evapotranspiration stochastic model, running it through rainfall/runoff models, and comparing it to long streamflow records (not regulated streams) to test final outputs. If this isn’t done, there is a risk that the issue is only partially considered.

The need for testing is all the more important when one considers the inherent risks in the stochastic rainfall approach. These include:

- It is likely that despite best efforts, the DPIE stochastic model of rainfall and PET will have deficiencies which may be amplified by the rainfall-runoff transformation.
- Rainfall-runoff models are conceptual models that require calibration to observed data. At best they reproduce satisfactorily all aspects of the observed data, more likely they provide a compromise, estimating some aspects of the data well and others poorly. Moreover, they may not perform adequately under climate forcing that is qualitatively different to the calibration period such as during extended drought. The work of Saft and her colleagues[27] has identified this as a significant issue for models calibrated to pre-Millennium drought data. A major weakness of conceptual rainfall-runoff models is that they do not account for vegetation dynamics (including response to prolonged water stress and carbon fertilisation).
- The rainfall-runoff model does not account for all the variability in observed streamflow. Ignoring this residual variability, which can be quite large, can further undermine the integrity of the process

[27] Saft, M., M. C. Peel, A. W. Western, J.-M. Perraud, and L. Zhang (2016), Bias in streamflow projections due to climate induced shifts in catchment response, *Geophys. Res. Lett.,* 43, 1574–1581, doi:10.1002/2015GL067326.

3.3.2 Split sample testing/validation using streamflow data

The Panel notes the importance of validating the rainfall approach using streamflow data within the current timeline. At the very minimum, the stochastic rainfall approach needs to be tested against reliable streamflow records using a split-sample testing regime. It is expected the outcome of such testing will not be comforting so it is important that any shortcomings be taken into account when making decisions in the near future. It may end up being better to stochastically simulate streamflow directly.[28]

Emphasis should be given not only to the frequency and duration characteristics of dry spells but also the rainfall characteristics of dam-filling (or near-filling) events (single or otherwise). The latter exercise would require a forensic look at the records to give a better idea of model performance, that is, determine metrics that impact water availability and how well the model performs against those.

Having knowledge about local runoff generation processes can help here too. In Southwest Western Australia, part of the recent streamflow decline is due to the detachment of perched water tables from stream channels and associated variable source areas. It now takes more rainfall to produce runoff, and the increase in the interarrival time between moderate and larger rainfall events exacerbates the problem.

DPIE provided the Panel with information on Quality Control (QC) work undertaken to address this issue. The Panel noted that the QC process used by DPIE does not mention split sample testing, which is arguably the most effective tool to detect non-stationarity. The tests described in the QC document are necessary but not sufficient. The Panel has experience with models that can pass the tests described in the QC document but fail split sample tests. This can occur because the QC tests are focussing on properties of marginal distributions (i.e. properties of the whole record), and they are not designed to look for change.

For the QC tests described by DPIE, the Panel recommends the use of p-values to evaluate the consistency of the observed statistics with model sampling distributions.

The Panel was not aware of DPIEs experience using the QC tests. However, the Panel noted that unless the rainfall-runoff model has a high goodness of fit score, they suspect tests dealing with variability and extremes may not fare well. There are many tests available, but given DPIE’s objectives, a judicious selection should lead to a small suite of options. Given there is a distinct possibility of trends in the observed data, the Panel doesn’t recommend using the correlation coefficient.

3.3.2.1 Rolling approach to model evaluation

A potentially useful approach to model evaluation involves the use of a rolling approach and drawing traces of the resulting parameter estimates (for rainfall and hydrological models) to gauge the impact of additional observations as the analysis moves through the historical record. The traces would capture any changes in the parameter estimates (size and sign) over time and reveal the extent of any systematic change. It would also facilitate a performance comparison between 'frozen' calibrated models and models in which the parameter estimates are updated over time. The performance metrics should concentrate on water availability. This is described in further detail in Appendix 4.

The rolling approach corresponds to a moving window through the observed time series. By calibrating parameters for each window, it can be thought of as a non-parametric nonstationary model with no assumptions made about the nature of the non-stationarity

[28] It was noted that some agencies like Hunter Water generate stochastic models directly for streamflow and that there could be merit to test this approach - to generate stochastic models directly in the streamflow space and compare to the approach of generating stochastic models for rainfall/ET as input into the rainfall/runoff models. There are pros and cons to both approaches, but the stochastic generation of flow could address some of the PET issues and save time; however, there may also be need for consistency of approaches across the state.

other than window length. This contrasts with the parametric non-stationary models which use an exogenous covariate in a parametric setting.

Both of these approaches are data-driven and are capable of representing present climate risk and possibly near future risk. They definitely offer an alternative (possibly tidier) approach to the baseline approach using a stationary model.

The Panel recommends that the stochastic rainfall approach needs to be tested against reliable streamflow records using split sample testing regime or rolling origin approach to model evaluation. This will involve running the DPIE stochastic model through rainfall/runoff models and comparing against long streamflow records, as well as testing the DPIE stochastic model's performance through a forensic look at the rainfall characteristics of dam filling events. Metrics that impact water should be part of the evaluation process.

This will involve UoA/UoN making available stochastic rainfall and ET to DPIE-Water, and for DPIE-Water to run forcing through their calibrated models and UoA/UoN would undertake the split-sample and/or rolling origin analysis. There may also be a need to apply post-processing algorithms, which has a precedent through work done by BOM and other parties.

4 FURTHER COMMENTS - PROPOSAL FOR GREATER SYDNEY REGION

The Panel agrees that it is worth proceeding with the UoN (Kiem) proposal for Greater Sydney region (Appendix 3). The Panel's comments reflected below acknowledge the challenges to this work. Kiem in his proposal has emphasised these challenges and the risk of outcomes that may not end up being necessarily useful. Conversely, if this work did provide insight it would be very valuable; the proposal should proceed with the understanding that expectations need to be managed.

In summary, the proposal from UoN offers the potential to gain the kind of understanding that was obtained for Newcastle catchments in previous RWS work. The value of that could be considerable. There is also some valuable scaffolding from the UoA work, as well as foundation elements achieved through UoN's published work.

The Panel also notes that while it is good that methods used across the state are as consistent as possible, differences in specific regions may warrant augmenting approaches. Importantly, the methods used have to be justified.

Comments on some of the steps in the proposal follow.

4.1.1 Identify weather systems

The Panel discussed that the Greater Sydney water supply comes from a large region covering both coastal and inland areas that might be subject to different weather patterns or multiple drivers in different parts of the catchment.

Approaches to identify key weather systems are informative. Weather typing has been undertaken a number of times before, such as PhD work by Acacia Pepler at UNSW Sydney on East Coast Lows. Repeating this exercise may not lead to new knowledge, although doing it with BARRA or ERA-5 may create new knowledge. As the data exist at National Computing Initiative (NCI) and the codes to do the weather typing exist, it would not be a major challenge and something new might be discovered.[29]

Weather typing is important, but practitioners need to think beyond simple correlation analysis and consider things beyond drivers such as ENSO, IPO, etc. Eastern Australia seasonal rainfall (east of the Great Dividing Range) doesn't correlate well with ENSO. Previous work has tried to identify what causes interannual to decadal variability in Sydney's rain/flow.

It is not known what controls the variability in ECL frequency, in part because early work lumped ECLs into one type, but newer work recognises that there are at least three types[30] and perhaps five types[31], each associated with different mechanisms. This means

[29] Previous work on weather typing can be used as a starting point (e.g. ESCCI work by Acacia Pepler, Stuart Browning and Ian Goodwin on large scale drivers of East Coast Lows (10.22499/3.6602.004), Danielle Udy and colleagues work on Tasmania and Antarctica, as well as weather typing work by Pook and colleagues and Kiem and colleagues on Victoria.

[30] Dowdy AJ, Pepler A, Di Luca A, Cavicchia L, Mills G, Evans JP, Louis S, McInnes KL, Walsh K. (2019) Review of Australian east coast low pressure systems and associated extremes. *Climate Dynamics*, 53(7): 4887–4910.https://doi.org/10.1007/s00382-019-04836-8.

[31] ECL types: 1) Easterly trough lows (ETL): events that track mostly east of the Great Dividing Range and in a southerly direction; 2) Southern secondary lows (SSL): events that track mostly over the ocean and in a northerly direction;3) Inland troughs (IT): events that evolve mostly over land, west of the Great Dividing Range and north of 30°S; 4) Continental lows (CL): events that evolve mostly over land, west of the Great Dividing Range and south of 30°S; 5) Extratropical cyclones (XTC): differentiation between ETLs and ECLs that evolve from tropical cyclones based on storm track trajectories is problematic, as both storm types develop in a similar region and follow similar tracks. However, the BOM maintains a database of all tropical cyclone occurrences and tracks from

correlating ECLs with any driver is hugely challenging, as they need to be separated by type and mechanism and one might get different responses according to the mechanism. The Panel recommends this work as a worthwhile long-term research project.

As referenced in Section 2.4.3, it is unknown whether the teleconnection patterns between climate drivers and temperature and rainfall will remain stationary into the future and whether they have been stationary in the past. Given work to date coming from the UoN, the Panel is also confident that this issue is well understood by the UoN team, and any analysis would be undertaken with the appropriate caution. As noted earlier, Kiem is clearly acutely aware of these issues and "merely hopes" to tease something useful from a combination of approaches. The Panel thinks this is the right way to view this, given that if something is identified it might be extremely valuable.

A recently completed PhD from ANU presented back-trajectory modelling of all rainfall events in the Murray Darling, separating the northern from the southern basin.

- Holgate, C.M., J.P. Evans, A.I.J.M. van Dijk, A.J. Pitman, 2020, Local and Remote Drivers of Southeast Australian Drought, *Geophysical Research Letters*, 47, e2020GL090238, doi: 10.1029/2020GL090238.
- Holgate, C.M., J.P. Evans, A.I.J.M. van Dijk, A.J. Pitman and G. Di Virgilio, 2020, Australian precipitation recycling and evaporative source regions, *J. Climate*, 33, 8721–8735, doi: 10.1175/JCLI-D-19-0926.1

That data set might be worth interrogating to see where the rain actually came from, noting this is over the reanalysis period, not over multiple centuries. Perhaps one could take the back trajectories and then ask about the synoptic types coincident with the rain.

4.1.2 Use of palaeo data

The Panel commented that using palaeo data is a good idea, but with the awareness that it is just information that provides data to infer other plausible scenarios or gain insights into previous examples of non-stationarity. The palaeo data shouldn't be overstated. It can be used to get an idea of the baseline and of how representative the instrumental record is for things like length of IPO phases etc.

The Panel was less supportive of using the Vance ice core but was strongly supportive of using the Wombeyan record. The latter was seen to have genuine potential value to inform Sydney dam inflows.

4.1.3 Climate modelling approach

The Panel was open about the approach going forward, noting that the preceding steps will dictate the method, which should look for consistency. It was expected this approach would likely be a combination of weather systems and climate drivers.

If variability can be linked to one (or a few) weather types, then the weather typing approach used for the South Coast region (linked to ECLs) might be feasible. However, it is expected that Greater Sydney will be more complicated and an approach similar to that used by UoA elsewhere (e.g. Hunter) may need to be used.

1900 to present (http://www.bom.gov.au/cyclone/history/). XTCs are therefore defined as events that evolved from storms. (from NSW Office of Environment & Heritage AdaptNSW, 2016, *Eastern Seaboard Climate Change Initiative – East Coast Lows Research Program Synthesis for NRM Stakeholders.* Accessed at https://climatechange.environment.nsw.gov.au/Impacts-of-climate-change/East-Coast-Lows.)

APPENDIX 1: DRAFT UOA REPORT

Implications of Non-Stationarity for Stochastic Time Series Generation in the Southern Basins

Anjana Devanand, Michael Leonard, Seth Westra

THE UNIVERSITY of ADELAIDE

DOCUMENT STATUS RECORD

Project Title: Implications of Non Stationarity for Stochastic Time Series Generation in the Southern Basin

Client: New South Wales Department of Planning, Industry and Environment

Document Number: 2020/07

File Name: PilotStudy_NonStationarity_SouthernBasin.docx

Issue No.	Date of Issue	Description	Signatures	
			Authors	Approved
1	10/07/2020	Draft report	AD, ML, SW	SW

Disclaimer:

1. The Intelligent Water Decisions research group has taken all reasonable steps to ensure that the information contained within this report is accurate at the time of production. In some cases, we may have relied upon the information supplied by the client.
2. This report has been prepared in accordance with good professional practice. No other warranty expressed or implied is made as to the professional advice given in this report.
3. The Intelligent Water Decisions maintains no responsibility for the misrepresentation of results due to incorrect usage of the information contained within this report.
4. This report should remain together and be read as a whole.
5. This report has been prepared solely for the benefit of the client listed above. No liability is accepted by the Intelligent Water Decisions with respect to the use of this report by third parties without prior written approval.

Contact Details:

Intelligent Water Decisions research group

School of Civil, Environmental and Mining Engineering

University of Adelaide

Physical:

Engineering North Building

North Terrace Campus
University of Adelaide, Adelaide.

Postal:

University of Adelaide, Adelaide, 5005

Telephone: (08) 8313 5451

Fax: (08) 8313 4359

Contents

List of Figures

List of Tables

Executive Summary

The Department of Planning, Infrastructure, Environment (DPIE) has undertaken a risk-based methodology to account for climate variability and change in developing its Regional Water Strategies (RWS), in which the risk assessment is informed by the use of stochastically generated long-term sequences that reflect climate variability beyond that contained within the instrumental record. An independent expert panel review of this climate risk method commended many aspects of the methodology, but recommended further review of:

1. **The potential non-stationarity of historical climate data used to inform the stochastic models,** to determine whether changes in climate in recent decades affect estimates of present-day climate risk compared with climate risk based on the whole observed record, by (a) assessing non-stationarity of the historical record; and (b) split-sample testing of the stochastic model.
2. **Investigation into the effects of multiple climate drivers,** to develop the methodology for stochastic generation in regions where multiple climate drivers influence the regional hydro-climate.

These recommendations are addressed in this report in the context of rainfall, temperature and evapotranspiration for the Southern Basin, using a ‘multiple lines of evidence’ approach to determine the presence and potential causes of any non-stationarity. Specifically, this report documents a review of literature on the physical mechanisms influencing the regional climate of southeast Australia, reported trends in hydro-climatic variables, and future projections in this region. This review is complemented by a pilot study using data from Ovens, Upper Murray, and Snowy catchments to assess trends in historical record and implications for stochastic simulations generated using the data. The pilot analysis used 49 rainfall time series and 30 evapotranspiration time series from these catchments, and three temperature time series at nearby locations from a homogenized dataset. The trends in seasonal and annual evapotranspiration, temperature and attributes of rainfall are studied using Mann-Kendall test. Split sample stochastic simulations are performed on the rainfall and evapotranspiration time series.

The main findings are summarised as follows.

1. Stationarity in temperature, rainfall and evapotranspiration

- **Temperature:** Literature documents temperature increases in this region, especially post 1960. The ‘Climate Change in Australia’ initiative reports an increase of 0.8°C in mean annual temperature in the Murray basin cluster (which contains the pilot catchments) over the 1910-2013 time period assuming a linear trend, with higher trends for temperature minima than for maxima; this is broadly consistent with a mean temperature increase over Australia by just over 1°C during the slightly longer period from 1910 to 2018. Moreover, climate projections indicate that increases of 0.6 to 1.3 degrees are expected in the Murray Basin cluster in the near-term (2020 to 2039) with respect to a 1986-2005 baseline. Trends in homogenized temperature sites near the pilot catchments are broadly consistent with these findings, with increases of 0.8-1.5°C in maximum temperatures and 1.9-2.6°C in minimum temperatures during the period 1913-2018. These multiple lines of evidence are in agreement that there is non-stationarity in the temperature record of the southern system.
- **Rainfall:** Literature documents decreasing trends in cool season (April to October) rainfall by 10-20% in southeast Australia since mid-1990s, predominantly in autumn and early winter. There are accompanying decreasing trends in the number of wet days during the cool season. Literature documented that “*the decline in rainfall across south-eastern Australia was at least partly attributable to climate change*” (CSIRO, 2012) and “*drying across southern Australia cannot be explained by natural variability alone*” (CSIRO & BOM, 2015). In the Ovens Murray catchment, Victoria Climate Projections 2019 (VCP19) project median annual rainfall decreases during the 2020 to 2039 period amounting to about 6% (11%) under the medium (high) emission scenarios (uncertainty range from -3% to -18%), with respect to a 1986-2005 baseline. The results of the trend assessment using data from the pilot sites are in general agreement with literature, showing decreasing trends in cool season rainfall and the number of wet days. The median total autumn

rainfall trends amount to -5.5%/decade at the pilot sites over the period from 1950 to 2018. Thus multiple lines of evidence indicate that the historical rainfall record in this region is non-stationary due to the changes in the cool season. The pilot sites also shows a short term decline in spring (SON) rainfall and increase in intensity of extreme rainfall intensity. These trends are not in agreement with all literature, but consistent with regional studies in nearby catchments.

- **Evapotranspiration:** Literature documents decreasing trends in pan evaporation over the period 1975-2002, whereas studies using more recent data (up to 2016/2018) report insignificant increasing trends. There are long term increasing trends in annual Morton wet evapotranspiration at the pilot sites, whereas short-term trends are insignificant. The pilot study results are not directly comparable with available literature based on pan evaporation data given differences in processes that drive pan evaporation and Morton wet evapotranspiration. Thus, non-stationarity in the historical record of evapotranspiration remains highly uncertain. VCP19 projects 8 to 10% increases in pan evaporation by 2030s (2020-2039).

2. Influence of climate drivers on meteorological variables in the Southern Basin

- The documented influence of climate drivers on key meteorological variables in the Southern Basin is summarised in this report. The declining cool season rainfall is associated with an expansion of the tropics, increasing intensity of the sub-tropical ridge over the continent and positive trends in the Southern Annular Mode (SAM). Literature indicates that these changes in large scale patterns during the cool season are at least partly attributable to climate change. Other climate drivers, notably El Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD), influence the inter-annual variability of regional rainfall, primarily affecting rainfall in winter and during the warm season (CSIRO, 2012; CSIRO & BOM, 2015; Hope et al., 2017).

3. The importance of stochastic generation calibration period on key statistics

- Consistent with the trend assessment, the split-sample tests demonstrate that the period used for calibration can significantly influence the statistics of the simulated data. In particular, the split-sample tests show that the millennium drought is a 'high-leverage' event, in the sense that statistics of the simulated rainfall can vary significantly depending on whether the drought is included in the calibration period or validation period. When the drought is included in the calibration period, the extent of similarity between the simulated rainfall and observed data following the drought is dependent upon the statistic and season under consideration. For example, inclusion of the drought in the calibration period brings the autumn rainfall in the simulated data close to recent (2010-2018) observations (reduction of biases in simulations from +24% to -4%) but can cause a larger deviations in simulated summer rainfall (increase in biases from +16% to +29%) from recent observations.
- The implications of non-stationarity is that the calibration period needs to be carefully considered in the context of any Southern Basin stochastic analysis, with the appropriate approach likely to depend on whether the objectives are to represent risk over 'historical', 'current' or 'future' periods.

4. Recommendations for stochastic data generation

TO BE COMPLETED

1 Introduction

Department of Planning, Infrastructure, Environment (DPIE) has undertaken a risk-based methodology to account for climate variability and change in developing the Regional Water Strategies (RWS). The method involves the use of stochastically generated long term sequences of climate data to characterize the "current" climate, and application of scaling factors to the stochastic data to generate future climate projections. The stochastic modelling uses historical (observed/reconstructed) records of daily rainfall, evapotranspiration and temperature to generate synthetic data for 10,000 years that reflect variability over the instrumental record from 1889 to 2018. The stochastic sequences provide insights into natural climate variability beyond the available observations. The scaling factors for a future climate are derived based on projections from the NSW and ACT Regional Climate Modelling (NARCliM) project, and applied to the stochastic data to characterise future climate risk.

The stochastic data generation methodology has been applied to multiple basins across the New South Wales using a multi-site stochastic data generator conditioned on the Inter-decadal Pacific Oscillation (IPO) documented in (Leonard & Westra, 2020). A similar method is planned to be applied for data generation in the southern basin of New South Wales, consisting of the New South Wales Murray and Lower Darling, Murrumbidgee, and parts of Victorian Murray water resource plan areas, shown in Figure 1. An independent expert panel review of the DPIE climate risk method recommended that ongoing improvement of the stochastic generation methodology be given high priority. The panel recommended further work to understand the implications of (1) the influence of multiple climate drivers, and (2) the existence of non-stationarity in the instrumental record of the southern New South Wales region on stochastic time series generation. The work presented in this report addresses these questions.

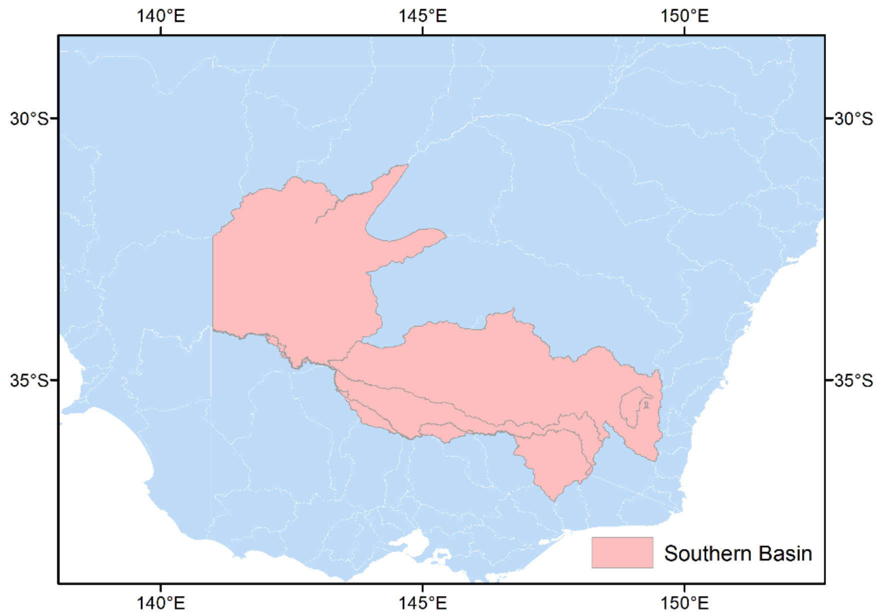

Figure 3 Southern basin of New South Wales

The regional climate of southeast Australia is highly variable, and there is a strong *a priori* reason to suspect that key meteorological variables are likely to already be experiencing change in the southern system. Literature notes warming signals in mean temperatures that are distinguishable from the background inter-annual and low frequency variability in this region (Ukkola et al., 2019). These temperature changes are consistent with expected changes due to climate warming (Jones, 2012; Karoly & Braganza, 2005). The regional rainfall patterns exhibit substantial variability (CSIRO, 2012; Hope et al., 2017), and is potentially affected by both natural and anthropogenic influences on the climate system. These influences on the changing patterns of rainfall in this region have been the subject of much focussed research, especially since the Millennium drought.

Section 2 of the report briefly reviews literature on the influences of regional and large scale climate drivers on the precipitation patterns in southeast Australia and the historical and expected changes in these influences, to assist in interpreting non-stationarity results in subsequent parts of this report. Section 3 summarises literature on the historical trends in climatic variables in this region and section 4 summarizes the future climate projections. Subsequently, a pilot study using data from representative catchments in the southern New South Wales and Victoria are undertaken to assess the observational record. The assessment involves an analysis of trends in the observed data from the pilot sites and split-sample calibration and validation of stochastic replicates as recommended by the expert review panel. The results of this pilot study is presented in Section 5 of this report. Based on the review of literature and the results of the pilot study, we provide options and recommendations for stochastic data generation to characterise "historical" and "current" climates in southeast Australia in Section 7.

2 Large-scale patterns of variability and change relevant to southeast Australia

The precipitation in southeast Australia exhibits substantial inter-annual and intra-seasonal variability, influenced by large scale patterns of global ocean-atmosphere variability. Different moisture systems contribute to precipitation in this region. These include low pressure systems that bring in moisture from the Southern ocean, north western cloud bands that originate in the Indian Ocean, eastern coastal troughs from northern Australia, and east coast low pressure systems (CSIRO, 2012; Dowdy et al., 2019; Hope et al., 2017). The region exhibits a cool season precipitation regime which is dominated by moisture contributions from Southern westerlies. The circulation patterns in the Southern Oceans affect the location of sub-tropical ridge in the Australian mid-latitudes which in turn influence the penetration of low pressure systems from the south into the continent during the cool season.

Thus, the regional precipitation in this region is influenced by the oceanic and atmospheric patterns in the surrounding Pacific, Indian and Southern Oceans, which collectively 'modulate' weather patterns by imparting long-term (generally interannual) persistence, and also mediate some (particularly circulation-related) aspects of the anthropogenic climate change signal on local weather patterns. The indicators of these patterns, typically referred to as climate 'drivers', are the El Niño Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), Southern Annular Mode (SAM), and the Inter-decadal Pacific Oscillation (IPO). These drivers influence the regional rainfall in SEA in different seasons and at different scales (CSIRO, 2012; Hope et al. 2017). We briefly summarize the current understanding of these influences in the following sub-sections. The summary is based on the technical reports by the southeast Australia Climate Initiative (SEACI) (CSIRO, 2012), Climate Change in Australia (CSIRO & BOM, 2015; Timbal et al., 2015), Victoria Climate Initiative (VicCI) (DELWP, 2016; Hope et al., 2017) and Victoria Climate Projections 2019 (VCP19) (Clarke et al., 2019a).

2.1 Influence of large scale drivers: ENSO, IOD, SAM, and IPO

ENSO and IOD are dipole modes of ocean-atmosphere variability in the tropical Pacific and Indian Oceans. Positive values of IOD refer to higher sea surface temperatures (SSTs) in the western equatorial Indian Ocean, and El Niño is the ENSO mode associated with higher SSTs in eastern equatorial Pacific Ocean. Positive IOD and El Niño are thus associated with higher SSTs and tropical convective centres that are located farther away from the Australian continent, leading to lower rainfall in Australia. ENSO and IOD are understood to influence the rainfall in SEA primarily during winter and spring (CSIRO, 2012; Hope et al., 2017).

The SAM is a mode of mid- and high-latitude climate variability associated with north-south shift of the atmospheric mass between the polar region and the mid latitudes. A positive value of SAM indicates a shift of mid-latitude storm tracks towards the South Pole. SAM influences the rainfall over SEA differently in different seasons. During winter, a positive SAM is associated with a shift of mid-latitude storm tracks towards the poles, reducing rainfall over SEA. During the summer, a positive value of SAM is associated with increased onshore transport of tropical moisture in eastern Australia and a subsequent increase in the warm season rainfall.

Thus, in general, positive SAM, positive IOD and El Niño events result in lower cool season rainfall in SEA. The relationships are further complicated due to interactions between these modes. On decadal time scales, a pattern of Pacific climate variability, the IPO, affects the inter-annual variations associated with ENSO and IOD. When the IPO is in the negative phase (cold phase), variability of ENSO and IOD are weakened, and the coupling between them is also weakened. During this phase, the impacts of ENSO and IOD on the eastern Australian rainfall is strengthened. The variability in SAM is also related to the ENSO, mainly during the warm season (Hope et al., 2017).

2.2 Influence of east coast lows on the southeast coast

In addition to the large scale drivers outlined above, the precipitation in the southeast coast is influenced by low pressure systems known as East Coast Lows (ECLs) (Dowdy et al., 2019). ECLs are cyclonic systems that occur near the south-east coast due to both mid-latitude and tropical influences. These systems can occur during any time of the year, but are more common and intense during the cooler months (Dowdy et al., 2013). ECL-related rainfall events exhibit a spatial contrast since the rainfall from these events primarily occur on the eastern coastal regions rather than inland areas, due to presence of the Great Dividing Range. These systems may also cause snowfall in the mountainous regions of southeast Australia (Fiddes et al., 2015). ECLs are associated with large rainfall events and multiple climate hazards in the southeast coast. These impacts have resulted in focussed research in recent years to study the characteristics of ECLs, the influence of climate drivers on ECLs, and their expected changes into the future.

Studies have characterized ECLs in observations using multiple climatic features and report a large inter-annual variability in the number of ECL systems in the historical record (Di Luca et al., 2015; Pepler et al., 2015). The relationship of ECLs to large scale climate drivers (ENSO, IOD and SAM) are reported to be generally weak, but some studies report mixed findings that indicate that some types of ECLs may potentially be related to the SAM (Dowdy et al., 2013; Pepler et al., 2015). A warming climate is expected to result in fewer ECL-related rainfall events during the cooler months, while there are large uncertainties in the expected ECL changes during the warmer months. The historical record also shows a decline in the number of ECLs, but the trend is not significant (Dowdy et al., 2013; Pepler et al., 2015).

Studies have examined the skill of regional climate models in simulating ECLs in historical and future scenarios. Regional simulations reproduce the climatology of ECLs (Di Luca et al., 2016), and future regional climate projections from NARCliM exhibit a decline in ECL frequency during the cooler months. The decreasing signal is reported to be robust in the different ensemble members and is consistent with the decreasing trend in the historical record and expected changes into the future (Pepler et al., 2016).

3 Climatic trends in the historical record of southeast Australia

In this section, we review publications documenting the historical trends in precipitation, evaporation and temperature in southeast Australia (SEA). This region broadly covers the mainland Australian region south of 33°S and east of 135°E and encompasses most of the southern basin that is the focus of this report. It is to be noted that there are some differences in the region referred to as "southeast" in various studies in literature, however these latitude–longitude ranges are broadly consistent.

3.1 Rainfall

There is a well-documented decreasing trend in mean rainfall during the cool season, primarily during autumn, in SEA. This trend is reported in studies since the mid-2000s; the reduction in cool season rainfall post mid-1990s influenced the long-term and medium-term mean precipitation trends in this region (Gallant et al., 2007). The reported trends in the attributes of precipitation are summarized in Table 1.

Further research explored the climatic features that lead to the reduction in precipitation. The cool season decline was a notable feature of the millennium drought that this region experienced from 1996 to 2009; the signal has persisted post the drought. Figure 2 shows the cool season (April to October) rainfall anomaly in southeast Australia during 1900 to 2018 from the 2018 State of the Climate report (BOM and CSIRO, 2018). The report documents that the southern half of Australia (south of 26°S) had below-average Apr-Oct rainfall in 17 of the last 20 cool seasons, since 1999. The recent years with above-average rainfall (2010, 2016) were generally associated with drivers of higher than usual rainfall across Australia (strong negative IOD in 2016; La Niña in 2010) (BOM & CSIRO, 2019)

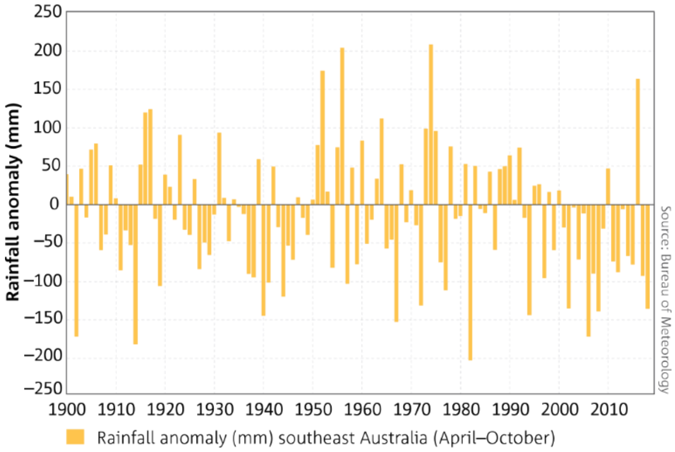

Figure 4 Anomalies of April to October rainfall for southeastern (south of 33° S, east of 135° E inclusive) Australia. Anomalies are calculated with respect to 1961 to 1990 averages (Source: BOM and CSIRO, 2019)

The current understanding indicates that changes in large-scale atmospheric features influence the observed trends. ENSO and IOD primarily influence the rainfall in SEA during winter and spring and they do not have a major impact on the autumn rainfall. Hence these climate drivers are not thought to be the primary cause of the declining autumn rainfall (Timbal & Hendon, 2011). The general consensus from SEACI and VicCI (CSIRO, 2012; CSIRO & BOM, 2015; Hope et al., 2017) suggests that the decline in cool season rainfall is associated with the expansion of the tropics, and increasing intensity of the sub-tropical high located over the continent. Tropical expansion causes the mid-latitude storm tracks responsible for most of the cool season rainfall in south-east Australia to move further south. The tropical expansion and changes in the intensity of sub-tropical high are also associated with positive trends in SAM, which indicate a poleward shift of the mid-latitude westerlies.

These climatic changes and the precipitation trends in the region are at least partly attributable to climate warming (CSIRO, 2012; CSIRO & BOM, 2015; Hope et al., 2017). Experiments using climate models indicate that expanding tropics can only be reproduced when global atmospheric changes in greenhouse gases, aerosols and depletion of stratospheric ozone are incorporated into global model simulations (Nguyen et al. 2015). Literature has studied the regional variation in Hadley circulation by dividing the globe into three "sectors" based on centres of upper level tropical divergence (depicting upward branch of the Hadley circulation) (Nguyen et al., 2018). There is enhanced expansion of the Hadley circulation in the Asia-Pacific (Australian) sector compared to the African and South American sectors. This regional expansion is linked to the negative phase of the IPO, and may reduce when the IPO changes phase (Nguyen et al., 2018). Therefore, both natural variability and climate warming are understood to contribute to the declining rainfall trends in SEA (Hope et al., 2017).

While tropical expansion and trends in SAM appear to be the drivers of the cool season precipitation trend in this region, the influence of other drivers, such as the ENSO and IOD, are relevant with respect to the changes during warm season. A La Niña mode concurrent with a negative phase of the IOD is associated with the rainfall events during spring of 2010/11 that marked the end of the millennium drought.

Table 3 Summary of reported precipitation trends in the historical record of southeast Australia

Authors	Dataset	Period	Region	Index Type	Findings
Gallant et al. (2007)	Stations (95 across Australia)	1910-2005 1950-2005	Six regions in Australia	Total rain Rain days (threshold = 1 mm) Mean rain per rain day	For the southeast region, long-term (1910-2005) results show decreasing trends in autumn in total rainfall and extreme intensity.

				Extreme intensity, frequency, and proportion of the total (95th and 99th percentile thresholds)	In the medium term (1950-2005), autumn rainfall indices show decreasing trends in most characteristics of rainfall, except the extreme proportion indices which showed significant increasing trends
Alexander et al. (2007)	0.25° x 0.25° gridded data	1910-2005 1950-2005	Whole Country	Annual and seasonal precipitation Extremes (Max 1-day, Max 5-day) Number of heavy and very heavy precipitation days Consecutive wet and dry days Annual proportions from extremes (Note: detailed results from these indices are not included in the publication)	Decreasing trends in autumn precipitation in southeast Australia during 1950-2005 Trends in extremes are correlated with the trends in mean, in general, across the whole country.
Taschetto & England, (2009)	Gridded BoM Rainfall data at 0.5° resolution	1970-2005	Whole Country	Annual and Seasonal Total Rainfall Frequency of moderate (up to 1 SD from the mean), heavy (from 1 to 2 SD from mean), very heavy (more than 2 SD from mean) rainfall events	Decreasing trends in rainfall over Victoria, Southern South Australia and Southern New South Wales during summer and autumn (stronger in autumn). Frequency of very heavy rainfall events during MAM also show a decline.
Risbey et al., (2013)	Average rainfall from 8 stations in Malle region in Victoria	1956-2009	Malle region in Victoria	Total rainfall during the cool season (Apr to Oct)	Decreasing trends in cool season rainfall Reported to be primarily associated with a decline in rainfall from cut-off lows
Theobald et al. (2016)	Station data	1958-2012	Snowy Mountains	Annual, cool season (Apr to Oct), and warm season (Nov to Mar) rainfall total and frequency of wet days and heavy precipitation days (P > 1mm) & (P > 10mm) Frequency and intensity of extreme precipitation (P > 90th percentile threshold)	Annual decreasing trends in the frequency P>10mm events, but an increase in the total precipitation the events generate. The increase in precipitation from these events occur during the warm season. Annual intensity of extreme precipitation shows an increasing trend.

Ukkola et al., (2019)	Area average records from BoM for 6 regions across Australia	1900-2018	Whole Country	Annual and Seasonal totals	No significant trends in regional mean rainfall in southeast Australia

3.2 Temperature

Literature reports a warming signal in temperature records in southeast Australia, consistent with the expected changes due to global warming. Karoly and Braganza (2005)studied the trends in minimum (Tmin), maximum (Tmax) and mean daily temperatures in southeast Australia over a 50 year period from 1954 to 2003, after removing the rainfall related component of temperature variations. The study reports a clear anthropogenic warming signal in observed temperature trends in southeast Australia. Jones (2012) note discrete step changes in minimum and maximum temperatures in southeast Australia, and attribute these changes to episodes of anthropogenic regional warming. The study reports that the temperature data of southeast Australia is stationary from 1910 to 1967 followed by a series of step changes: 0.7°C in Tmin in 1968, 0.5°C in 1973 in Tmax and 0.8°C in Tmax in 1997, attributed to anthropogenic warming. We summarize the reported trends in temperature records in the SEA in Table 2, which shows increasing trends in mean, minimum and maximum temperatures over the region.

Thus, there is a well-established warming signal in the temperature records in SEA, especially post 1960. The 'Climate Change in Australia' initiative reports an increase of 0.8°C in the Murray Basin cluster (Timbal et al., 2015) (which consists of most of the southern basin), over the 1910-2013 time period assuming a linear trend. The trends are higher for temperature minima (total increase of 1°C) than for maxima (total increase of 0.7°C) The mean temperature over Australia has increased by just over 1°C during the period 1910-2018, the Victoria mean historical changes over 1910-2018 reported by VCP19 (Clarke et al., 2019a) is also just above 1°C.

Table 4 Summary of reported temperature trends in the historical record of southeast Australia

Authors	Dataset	Period	Region	Index Type	Findings
Ashcroft & Karoly (2012)	Daily data from 103 stations across the country Monthly area-averaged anomalies for states and Northern Territory	1860-1909 1910-1959 1960-2011 1860-2011	Southeast Australia	Annual, DJF, and JJA means of Tmax, Tmin Diurnal Temperature Range (DTR)	Positive trends in Annual, DLF, and JJA Tmax and Tmin. Stronger trends in Tmax (1.12°C) and Tmin (0.93°C) post 1960.
Jones (2012)	Homogenised temperature data on a 0.25° grid	1910-2010	Southeast Australia	Annual means of Tmin and Tmax	Positive step changes in Tmin and Tmax post 1967
Ukkola et al. (2019)	Area average records from BoM for 6 regions across Australia	1910-2018	Whole Country	Annual and seasonal mean temperature	The mean air temperature trends during all seasons are positive and statistically significant in southeast Australia. The strongest trends are in summer (DJF, 0.015 °C/yr) and autumn (MAM, 0.012 °C/yr)

3.3 Pan evaporation

The non-stationarity in historical evapotranspiration records is not well established. The available literature have used pan evaporation records to study trends post year 1975. Earlier studies reported a decreasing trend in pan evaporation over 1975-2002 period across the country potentially due to reduced atmospheric demand associated with decreasing wind speed or radiation (Roderick & Farquhar, 2004). This trend is reduced or reversed (i.e., increasing trends or no trends) in more recent analyses using data up to 2016/2018 due to temperature driven changes in vapour pressure deficit (Stephens et al., 2018; Ukkola et al., 2019). We summarize these reported trends in Table 3.

Table 5 Summary of reported evapotranspiration trends in the historical record of southeast Australia

Authors	Data-set	Period	Region	Index Type	Findings
Roderick and Farquhar (2004)	30 BoM sites across the country	1970-2002	Entire Country	Pan Evaporation	Decreasing trends over Australia primarily due to a decline in atmospheric demand, associated with declines in surface radiation or wind speed
Stephens et al. (2018)	41 BoM sites across the country	1975-2016	Entire Country	Pan Evaporation	The declining trends detected earlier have reduced, or become neutral in southeast Australia. The changes in trends from previous reports are due to increases in temperature driven vapour pressure deficits.
Ukkola et al. (2019)	Area average monthly observations from BoM	1975-2018	Whole Country	Pan Evaporation	No significant trends in southeast Australia

4 Future climate projections for southeast Australia

We include a summary of the future climate projections for this region as an alternate line of evidence to understand climatic non-stationarity. Climate projections for southeast Australia indicate future precipitation and temperature changes consistent with the trends in the current observational record, i.e., warmer and drier in general, especially in the cool season. From climate model projections, positive values of the SAM and an increase in the number of positive IOD events are likely in the future, bringing drier conditions to south-east Australia during the cool season. The changes in ENSO and its interactions with the SAM and IOD into the future are currently unknown (Clarke et al., 2019a; CSIRO, 2012).

Future projections for the Murray Basin cluster are documented as part of the CCIA initiative (Timbal et al. 2015). This cluster includes most of the southern basin region that is the focus of this report. Climate projections indicate that increases of 0.6 to 1.3 degrees are expected in the Murray Basin cluster in the near-term (2020 to 2039) with respect to a baseline of 1986-2005.The report considers the physical understanding of climatic relationships and results from downscaling future projections to conclude that there is high confidence that cool season rainfall will decline in future, but the magnitude of decline is very uncertain (Timbal et al. 2015). In the near term (2030), it is reported with high confidence that natural climate variability will remain the major driver of rainfall differences from the climate of 1986–2005.

The latest climate projections from the Victoria Climate Projections 2019 (VCP19) include regional projections for the Ovens and Murray catchments (Clarke et al., 2019b) that are also used in the pilot study documented in Section 5 of this report. The key changes projected for these catchments are:

Daily minimum and maximum temperatures are projected to continue to increase, with an increase in Tmax of 1.0 to 1.9 °C (since 1990) are expected by 2030s

Rainfall is projected to be very variable, but will continue to decline in winter and spring (medium to high confidence) and autumn (low confidence)

Intensification of 1-in-20 year maximum daily extreme rainfall events

The VCP19 report projects changes under medium (RCP 4.5) and high (RCP 8.5) emission scenarios with respect to a baseline climate of 1986 to 2005. The changes are projected for 20 year time slices up to year 2090. The first time slice of projected changes spans 20 years from 2020 to 2039. Thus, the "current" climate (year 2020) marks the beginning of the first period of projected change in this report. The projected rainfall changes in the Ovens and Murray catchments reported by VCP19 are shown in Figure 3.

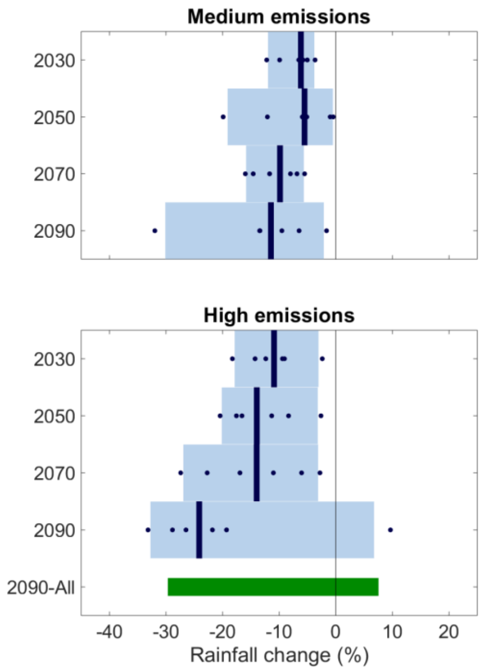

Figure 5 Projected changes (compared to 1986-2005) in annual mean rainfall in the Ovens Murray catchment for medium emissions (top) and high emissions (bottom). Bars show the 10th to 90th percentile range. Blue bars: results from the new downscaled modelling. Dark vertical line: median. Dark blue dots: individual models. Green bar: results from all available modelling (high resolution and GCM) for comparison at 2090 (Source: Clarke et al. 2019b)

The annual mean rainfall during 2020 to 2039 is projected to decrease with the largest declines during spring. The projected median change in annual rainfall is is -6% (range of -12% to -4%) under the medium emissions scenario, and -11% (range of -18% to -3%) under high emissions scenario with further declines anticipated in subsequent decades (Figure 3). The annual maximum temperatures are expected to increase by 1.1°C (1.4°C) under medium (high) emission scenarios during the 2020 to 2039 time slice; the projected changes in minimum temperatures are smaller in magnitude for the Ovens Murray catchments. The projected annual changes in pan evaporation during 2020 to 2039 are positive, with a median change of 8% (10.8%) under medium (high) emission scenarios.

5 Pilot study: assessment of the historical record

The review of literature indicates the presence of non-stationarity in the historical record of southeast Australia, with a number of references attributing key aspects of change to anthropogenic climate change. A pilot study using data from a few representative catchments are undertaken to assess the presence of significant trends in the attributes of climatic variables and the implications for stochastic time series generation in the southern New South Wales region. As part of the assessment, we perform a trend analysis of the historical record of the pilot catchments and split-sample testing of the stochastic model, according to the recommendations of the independent review panel.

The focus of this pilot study was to analyse key time series provided by DPIE that are used as inputs for hydrological modelling. As a result, interpretation of the pilot results needs to consider several caveats:

- The homogeneity of the data has not been reviewed in this analysis. For this reason, the results presented below can be used to assess non-stationarity of the analysed *data*, but cannot be interpreted as a climate change attribution study.
- The analysis assesses a relatively small number of stations relative to other peer reviewed assessments of trends in key weather variables in southeast Australia, and thus is not suitable to assess large-scale drivers of change (e.g. the effects of changes to circulation patterns) that present themselves when evaluating trends over larger geographical areas.

For these reasons, results from this section should be considered in the context of the wider literature as summarised in Sections 3 and 4, rather than viewed in isolation.

5.1 Data

We use data from pilot sites located in the Upper Murray, Ovens and Snowy catchments, shown in Figure 4.

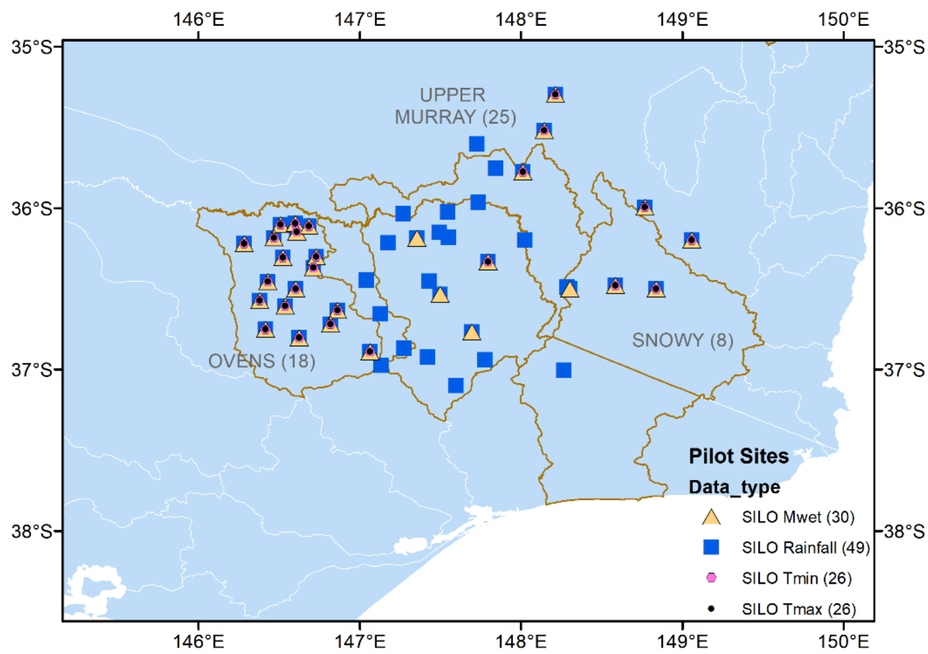

Figure 6 Location of rainfall, evapotranspiration and temperature pilot sites

Table 4 provides a summary of the data from the pilot sites which consists of precipitation, evapotranspiration (Mwet – the tag refers to Morton Wet formulation), and minimum/maximum daily temperature (Tmin, Tmax) time series. All data were sourced from the SILO database. The time-series span 130-years from 1/1/1889 to 31/12/2018 and there were no missing values (owing to pre-determined infilling methods used to construct the data).

Table 6 Summary of number of different observation timeseries by variable type

	Pilot Basins		
Variable Type	**Upper Murray**	**Ovens**	**Snowy**
SILO Rain	25	18	6
SILO Mwet	5	18	7

SILO Tmin/Tmax	0	18	8

As described above, the analysis of trends using raw station observations of Tmax and Tmin are not advisable because signals of changes in instrument, methods of data collection, and station location may exist in the data (Trewin, 2013). Such analysis are generally performed using homogenized temperature datasets. In the process of homogenization, raw temperature data from multiple sites are examined visually and statistically to create homogenised datasets that minimize discrepancies across time (Ashcroft & Karoly, 2012; Trewin, 2013)(Ashcroft et al. 2012; Trewin et al. 2012). The Australian Climate Observations Reference Network – Surface Air Temperature version 2 (ACORN SATv2) (Trewin et al., 2020) data is one such high quality dataset prepared and made available by the Bureau of Meteorology, Australia. This data has been used in literature to study the changes in temperature in Australia (Allen et al., 2019; van Wijngaarden & Mouraviev, 2016), and thus analyses of data provided by DPIE for the pilot sites are supplemented by an analysis of ACORN SATv2 data at key locations in the study area.

The temperature stations from the pilot sites are located in the Ovens and Snowy catchments. Three ACORN SATv2 stations are located in the vicinity of these sites as shown in Figure 5.

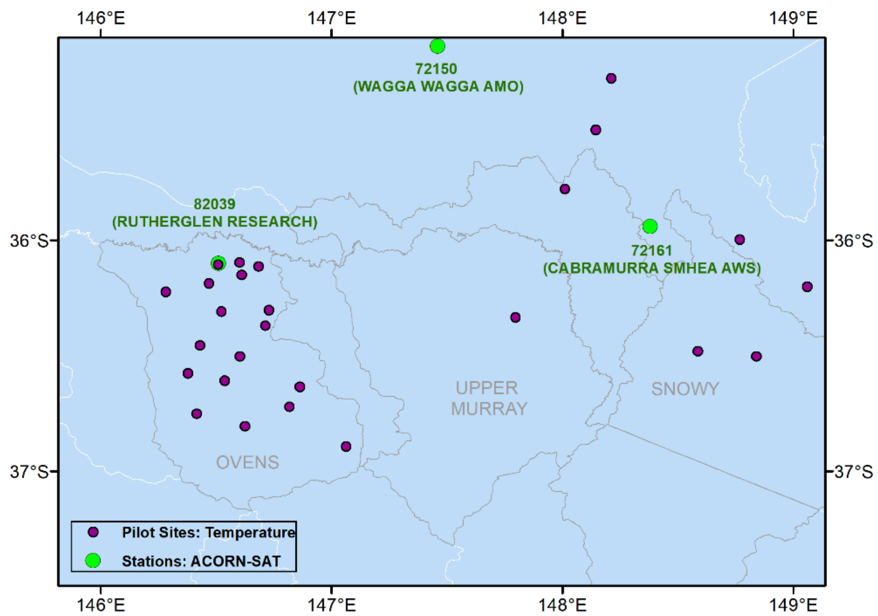

Figure 7 Location of pilot temperature sites and three ACORN SAT v2 stations in the vicinity

Homogenized data at pilot site 82039 (Rutherglen Research) is available from the ACORN SAT v2 dataset, so that the differences between the data using the two methods can be compared. Figure 6 shows the differences in long-term trends calculated from the raw station data and the homogenized ACORN SATv2 data at the same site using data for years 1913 to 2018 (101 years). There are major differences in the trends estimated from the two data sources. Such differences also exist in the short term trends estimated from the two data sources at this site (Appendix A). These differences indicate that the site level temperature observations from the pilot sites are unsuitable for the analysis of trends. While there is some debate on the homogenization of temperature data (Marohasy & Abbot, 2016), the use of homogenized data for assessment of trends is the existing globally accepted standard of analysis (Hewaarachchi et al., 2017; Squintu et al., 2019; Vincent et al., 2020). Therefore, we perform trend analysis using the homogenized data from the three ACORN SATv2 stations located close to the pilot region (shown in Figure 5) as part of this study. The length of homogenised data records at these sites are listed in Table 5.

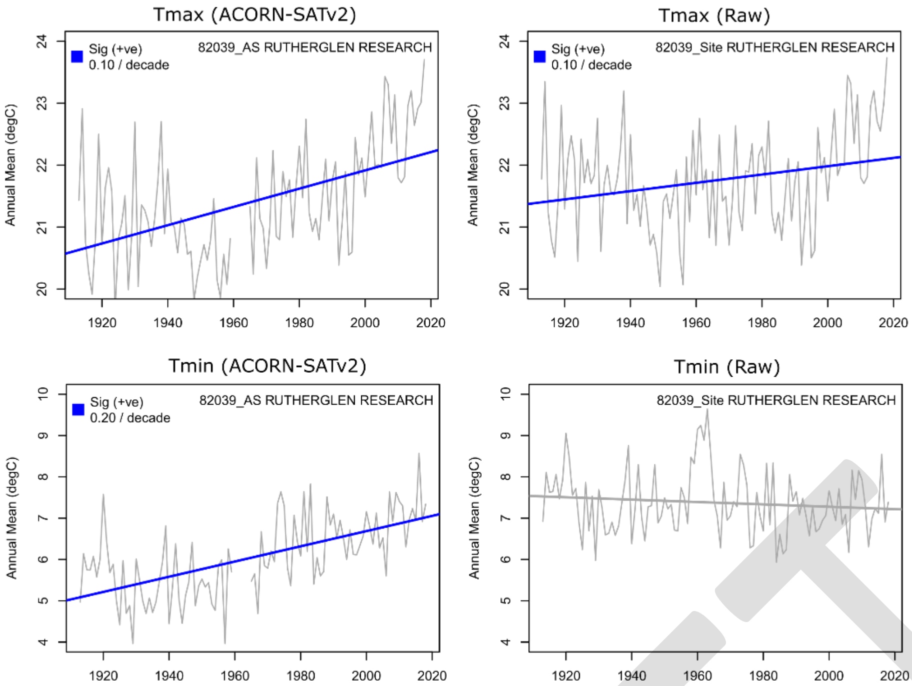

Figure 8 Annual mean Tmax and Tmin at site 82039 (Rutherglen Research) from raw station data and homogenized ACORN SATv2 data with linear trend lines. Blue trend lines indicate the presence of significant trends (at 5% level) using Mann-Kendall trend test. The data from 1960 to 1964 is missing at this site in the ACORN-SATv2 data.

Table 7 ACORN SATv2 stations used for analysis

Variable Type	Station Number	Period of Record
	82039	08/11/1912 to 31/05/2019
Tmax, Tmin	72150	01/01/1910 to 31/05/2019
	72161	01/01/1962 to 31/05/2019

5.2 Methodology

5.2.1 Trend detection

The statistical significance of temporal trends in various attributes of the time series are assessed using non-parametric Mann Kendall test at two sided 5% significance level. The Mann Kendall test is a well-established technique employed in studies for assessment of hydro-climatic time series (Lavender & Abbs, 2013; Theobald et al., 2016; Ukkola et al., 2019). The magnitude of significant trends is quantified using least square regression, similar to the analysis performed by Theobald et al. (2016).

While this approach can assess whether individual sites exhibit statistically significant trends, in multi-site analyses there is often a non-negligible probability of detecting one or more individual sites with significant trends even under the null hypothesis of no trends (for example at the 5% significance level one would expect an average of five out of every 100 sites to experience statistically significant trends under the null hypothesis that there is no trend). As such, a ‘field significance’ test is used to determine whether the number of stations experiencing statistically significant trends are more than would expected under the null hypothesis. The field significance of the trends are assessed in this study using a bootstrap resampling procedure (e.g. Do et al. (2017)). The bootstrap procedure uses resampled data to obtain an estimate of the 95th percentile value of the percentage of significant sites that may occur due to chance. If the percentage of sites exhibiting significant trends in the historical record exceeds this estimate, then the trends are considered field significant. The methodology used for the analysis of trends consists of the following steps:

The significance of site level trends are estimated using the Mann Kendall test. The proportion of sites that exhibit significant positive and negative trends in the historical record are calculated.

The entire dataset is randomly resampled in time while preserving the spatial structure. The new resampled data therefore contains a new sequence of years (eg: {1967, 1954, 2003, 1895, 1920…}). The site level significant trends in the resampled dataset are estimated using Mann Kendall test and the proportion of sites that exhibit positive and negative trends are calculated as done in step A.

A bootstrapping procedure is used to repeat step B 1000 times. The samples are used to create a distribution of percentage of significant sites that may occur in the region due to chance. If the proportion of significant sites in the historical data (step A) is higher than the 95th percentile value of the proportion of significant sites that may occur due to chance, the historical trend is considered to be field significant.

The attributes of rainfall, evaporation and temperature used for the trend analyses are listed in Table 6. These attributes are selected to comprise 'hydrologically relevant' features of the respective variables, which may have a bearing on hydrological response of the respective catchments. The trend analysis is performed using the entire dataset of 130 years (1889 to 2018) as well as a recent subset of the dataset, to enable comparison with existing literature.

Table 8. Attributes of hydro-climatic variables and the time periods used for trend analyses

Variable	Attribute	Definition	Analysis Period
Rainfall	Total	Total annual and seasonal (DJF, MAM, JJA & SON) rainfall (mm)	1889 to 2018 1950 to 2018
	Wet Day Rainfall	Mean annual and seasonal wet day (P >= 1 mm) rainfall (mm/day)	
	Number of Wet Days	Annual and seasonal number of wet days (P >= 1mm) (days)	
	Heavy Day Rainfall	Annual and seasonal heavy day (P >= 10 mm) rainfall (mm/day)	
	Number of Heavy Rainfall Days	Mean annual and seasonal heavy day (P >= 10 mm) rainfall (mm/day)	
	Mean Dry Spell Duration	Annual mean number of consecutive days with rainfall less than 1 mm (days)	
	Maximum Dry Spell Duration	Annual maximum number of consecutive days with rainfall less than 1 mm (days)	
	Extreme Intensity	Annual mean rainfall during days with rainfall greater than the 95^{th} percentile (mm/day)	
	Extreme Frequency	Annual mean number of days with rainfall greater than the 95^{th} percentile (days)	
Evapotranspiration	Total Evapotranspiration	Total annual and seasonal (DJF, MAM, JJA & SON) evapotranspiration (mm)	1889 to 2018 1975 to 2018
Temperature	Mean Temperature	Annual and seasonal (DJF, MAM, JJA & SON) mean daily minimum temperature Annual and seasonal (DJF, MAM, JJA & SON) mean daily maximum temperature	1913 to 2018 1960 to 2018

5.2.2 Split sample stochastic simulations

We assess the potential implications of non-stationarity on the results of stochastic analyses using split sample tests. This is achieved using a split sample methodology, in which the stochastic generation model is calibrated to one part of the time series (usually the earlier part of the record) and then validated on the other part of the record. This split-sample approach provides an analogy to possible issues that could arise by calibrating a stochastic model to the full historical record and assuming it is representative of current or future conditions. We perform split sample stochastic simulations using the precipitation and evapotranspiration (Mwet) data from the pilot sites to assess the ability of a stochastic model calibrated on the earlier part of the record to capture statistics corresponding to the later part of the record. The split sample tests are designed based on the recommendations of the independent review panel, and consists of "1990 Reference" and "Drought Reference" experiments. These are defined as follows:

- **1990 Reference:** The experiment uses data up to year 1990 to calibrate the stochastic model and data after year 1990 to validate stochastic simulations.
- **Drought Reference:** The experiment includes data up to the end of the Millennium drought (year 2009) to calibrate the model, and the remaining period data to validation the simulations.

We perform the calibration-validation tests using the full record (1889 to 2018, 130 years) as well as a shorter recent period of data (1950 to 2018, 69 years) to assess the performance of the stochastic model while calibrated using different record lengths. Note that the year 1950 used here is selected arbitrarily to consider the potential of a using a shorter baseline (eg: corresponding to the NARCliM 1.5 baseline starting in 1950). In total the experiment suite consists of the four experiments shown in Figure 7.

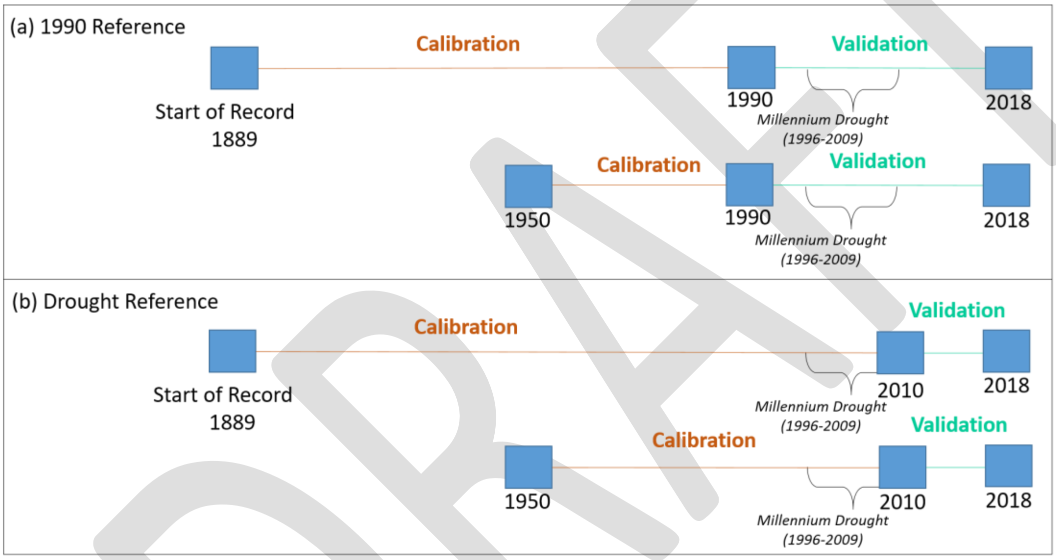

Figure 9 Split-sample experiments

The stochastic model used for the experiments uses is based on the latent-variable model formulation documented in Bennett et al. (2018). The model implemented for this pilot study is not conditioned on the IPO and it is a single site version of the spatial field rainfall model used for stochastic time series generation in the northern New South Wales basins (Leonard & Westra, 2020). The model is calibrated site-wise using observations from the calibration time period and used to generate stochastic time series. Time series are generated for 100 replicates of data length corresponding to the validation time period.

The simulated data is compared to observations during both the calibration and validation time periods for assessment of the split sample experiments. The comparison is based on the attributes of hydro climatic time series that show major trends in the historical record identified using non-parametric trend testing at the pilot sites. The mean values of the attributes from the simulated data are compared to the mean values of observations from the calibration and validation periods at site level. The results are presented using histograms of the differences between the observations and simulations across all the pilot sites.

5.3 Results of non-parametric trend testing

5.3.1 Trends in rainfall attributes

Table 7 summarizes the number of sites that exhibit significant trends in rainfall attributes during the different two time periods of analysis. The trends in the attributes of rainfall that is significant in both long term (1889 to 2018) and short term (1950 to 2018) analyses are:

- Decreasing trend in cool season totals
- Decreasing trend in the number of wet days annually and during the cooler seasons of the year
- Increasing trend in annual extreme rainfall intensity

Table 9 Number of sites that exhibit significant trends (pos = significant positive trend, neg = significant negative trend, none = no significant trends) in annual and seasonal attributes of rainfall. The value in brackets indicate the number of sites as a percentage of the total number of sites (49 sites). Shading indicates trends that are field significant (at 5% level).

	Total Rainfall (mm)									
	1889 to 2018					**1950 to 2018**				
	ANN	**DJF**	**MAM**	**JJA**	**SON**	**ANN**	**DJF**	**MAM**	**JJA**	**SON**
pos	5 (10%)	11 (22%)	2 (4%)	2 (4%)	3 (6%)	1 (2%)	1 (2%)	0 (0%)	1 (2%)	0 (0%)
neg	8 (16%)	0 (0%)	6 (12%)	15 (31%)	1 (2%)	13 (27%)	0 (0%)	13 (27%)	8 (16%)	24 (49%)
none	36 (73%)	38 (78%)	41 (84%)	32 (65%)	45 (92%)	35 (71%)	48 (98%)	36 (73%)	40 (82%)	25 (51%)
	Mean Wet Day (P >= 1 mm) Rainfall (mm/day)									
	1889 to 2018					**1950 to 2018**				
	ANN	**DJF**	**MAM**	**JJA**	**SON**	**ANN**	**DJF**	**MAM**	**JJA**	**SON**
pos	21 (43%)	19 (39%)	5 (10%)	8 (16%)	22 (45%)	5 (10%)	6 (12%)	1 (2%)	2 (4%)	3 (6%)
neg	7 (14%)	0 (0%)	7 (14%)	10 (20%)	2 (4%)	7 (14%)	0 (0%)	4 (8%)	8 (16%)	4 (8%)
none	21 (43%)	30 (61%)	37 (76%)	31 (63%)	25 (51%)	37 (76%)	43 (88%)	44 (90%)	39 (80%)	42 (86%)
	Number of Wet (P >= 1 mm) Days									
	1889 to 2018					**1950 to 2018**				
	ANN	**DJF**	**MAM**	**JJA**	**SON**	**ANN**	**DJF**	**MAM**	**JJA**	**SON**
pos	7 (14%)	10 (20%)	1 (2%)	3 (6%)	3 (6%)	0 (0%)	6 (12%)	0 (0%)	1 (2%)	0 (0%)
neg	19 (39%)	3 (6%)	16 (33%)	19 (39%)	11 (22%)	24 (49%)	0 (0%)	26 (53%)	10 (20%)	29 (59%)
none	23 (47%)	36 (73%)	32 (65%)	27 (55%)	35 (71%)	25 (51%)	43 (88%)	23 (47%)	38 (78%)	20 (41%)
	Mean Heavy Day (P >= 10 mm) Rainfall (mm/day)									
	1889 to 2018					**1950 to 2018**				
	ANN	**DJF**	**MAM**	**JJA**	**SON**	**ANN**	**DJF**	**MAM**	**JJA**	**SON**
pos	17 (35%)	9 (18%)	2 (4%)	7 (14%)	27 (55%)	7 (14%)	4 (8%)	2 (4%)	2 (4%)	5 (10%)
neg	2 (4%)	0 (0%)	0 (0%)	3 (6%)	0 (0%)	1 (2%)	0 (0%)	1 (2%)	1 (2%)	1 (2%)
none	30 (61%)	40 (82%)	47 (96%)	39 (80%)	22 (45%)	41 (84%)	45 (92%)	46 (94%)	46 (94%)	43 (88%)
	Number of Heavy Rainfall (P >= 10 mm) Days									
	1889 to 2018					**1950 to 2018**				
	ANN	**DJF**	**MAM**	**JJA**	**SON**	**ANN**	**DJF**	**MAM**	**JJA**	**SON**
pos	4 (8%)	15 (31%)	1 (2%)	2 (4%)	6 (12%)	0 (0%)	1 (2%)	0 (0%)	1 (2%)	0 (0%)
neg	6 (12%)	0 (0%)	5 (10%)	11 (22%)	1 (2%)	15 (31%)	0 (0%)	7 (14%)	6 (12%)	20 (41%)

none	39 (80%)	34 (69%)	43 (88%)	36 (73%)	42 (86%)	34 (69%)	48 (98%)	42 (86%)	42 (86%)	29 (59%)
	Mean Extreme (P > 95th percentile) Rainfall (mm/day)									
	1889 to 2018					**1950 to 2018**				
	ANN					**ANN**				
pos	15 (31%)					9 (18%)				
neg	1 (2%)					1 (2%)				
none	33 (67%)					39 (80%)				
	Frequency of Extreme (P > 95th percentile) Rainfall Days									
	1889 to 2018					**1950 to 2018**				
	ANN					**ANN**				
pos	13 (27%)					3 (6%)				
neg	3 (6%)					2 (4%)				
none	33 (67%)					44 (90%)				
	Maximum Dry (P < 1 mm) Spell Duration (days)									
	1889 to 2018					**1950 to 2018**				
	ANN					**ANN**				
pos	0 (0%)					13 (27%)				
neg	31 (63%)					1 (2%)				
none	18 (37%)					35 (71%)				
	Average Dry (P < 1 mm) Spell Duration (days)									
	1889 to 2018					**1950 to 2018**				
	ANN					**ANN**				
pos	6 (12%)					0 (0%)				
neg	13 (27%)					2 (4%)				
none	30 (61%)					47 (96%)				

Based on these results, the significant trends in the various rainfall attributes are summarised as follows:

- **Annual and Seasonal Totals:** The long term (1889 to 2018) analysis shows negative trends in the JJA total rainfall and an increase in the DJF total. In the short term (1950 to 2018) analysis, the decreasing trends are significant at a larger number of stations, and the declines occur in MAM, JJA and SON seasons. The highest number of stations show statistically significant declines in SON (49%).
- **Mean Wet Day Rainfall:** In the long term analysis, positive trends in the mean wet day rainfall are field significant at the annual scale. The positive trends occur in DJF and SON; negative trends occur in the cool season (MAM & JJA). In the short term analysis, only the negative trends in JJA are significant.
- **Number of Wet Days:** In both the short term and long term analyses, negative trends are significant at the annual scale. In the long term, positive trends occur in DJF; negative trends during the MAM, JJA, and SON. In the short term analysis, only the negative trends during MAM, JJA and SON are significant. The negative trends during MAM and SON occur at more number of stations.
- **Mean Heavy Day Rainfall days and the Number of Heavy Rainfall Days:** The trends in heavy rainfall intensity and frequency are vary between the analyses performed at long and short time scales. In the long term analysis, the prominent signal is an increase in mean heavy day rainfall

during all seasons except MAM. In the short term analysis, the prominent signal is a decrease in the number of heavy rainfall events in all seasons except DJF.

- **Intensity and Frequency of Extreme Rainfall Days:** Both long term and short term analysis show an increase in the intensity of extreme rainfall days. The increase in frequency of extreme days is only significant in the long term analysis.
- **Dry Spell Durations:** Long term analysis shows declining trends in the mean dry spell duration and maximum dry spell duration. The short term analysis shows an increasing trend in maximum dry spell duration.

It is to be noted that we refrain from comparing the trends at the pilot sites with literature that examined trends in precipitation characteristics using gridded datasets primarily for the purpose of assessing the fidelity of global climate models in capturing these trends (eg: Alexander & Arblaster, 2009, 2017). The spatial scale of these studies are very different from the station level analyses of trends at the pilot sites; hence comparison proves difficult, and potentially misleading. However, in general, the declining trends in cool season rainfall appears to be sufficiently widespread in spatial scale to be apparent in literature documenting analysis at both larger and finer spatial scales (Nicholls, 2010; Theobald et al., 2016).

The trends in rainfall attributes are compared to the reported trends in literature around the same region as summarised in Section 3 in Table 8. The decreasing trend in cool season precipitation in the pilot sites is consistent with this reported decline in this region. The decreasing trends in the frequency of wet days during the cool season are also broadly consistent with the trends documented literature. The increasing intensity in the annual intensity of extreme precipitation we note in the pilot sites do not appear to be as widespread spatially. The increasing trend in extreme intensity is consistent with literature focussed on near catchments (Theobald et al. 2016), but inconsistent with literature documenting larger spatial scale analyses (Gallant et al. 2007).

Table 10 Trends in rainfall attributes in the pilot sites and the changes reported in literature

Attribute	Study/Publication	Period	Ann	D	J	F	M	A	M	J	J	A	S	O	N	Comment
Total rainfall	Pilot study long	1889-2018	-	↑	↑	↑	-	-	-	↓	↓	↓	-	-	-	T&E (2009): Used gridded data; the comparison is based on approximately locating the pilot region from their figures.
	Gallant et al. (2007)	1910-2005	-	-	-	-	↓	↓	↓	-	-	-	-	-	-	
	Pilot study short	1950-2018	↓	-	-	-	↓	↓	↓	↓	↓	↓	↓	↓	↓	
	Taschetto and England (2009)	1970-2005	↓	-	-	-	↓	↓	↓	-	-	-	-	-	-	
	Risbey et al. (2013)	1956-2009	n/a	n/a	n/a	n/a	n/a	↓	↓	↓	↓	↓	↓	↓	n/a	Gallant (2007): Used a larger southeast Australia region for their analysis.
	Gallant et al. (2007)	1950-2005	-	-	-	-	↓	↓	↓	-	-	-	-	-	-	
Mean wet day rainfall	Pilot study long	1889-2018	↑	↑	↑	↑	↓	↓	↓	↑/↓	↑/↓	↑/↓	↑	↑	↑	-
	Gallant et al. (2007)	1910-2005	-	-	-	-	-	-	-	-	-	-	-	-	-	
	Pilot study short	1950-2010	-	-	-	-	-	-	-	↓	↓	↓	-	-	-	
	Gallant et al. (2007)	1950-2005	-	-	-	-	↓	↓	↓	-	-	-	-	-	-	
Number wet days	Pilot study long	1889-2018	↓	↑	↑	↑	↓	↓	↓	↓	↓	↓	↑	↑	↑	T&E (2009): Analyzed moderate rainfall events as events within one standard deviation of the mean.
	Gallant et al. (2007)	1910-2005	-	-	-	-	-	-	-	-	-	-	-	-	-	
	Pilot study short	1950-2018	↓	-	-	-	↓	↓	↓	↓	↓	↓	↓	↓	↓	
	Gallant et al. (2007)	1950-2005	-	-	-	-	↓	↓	↓	-	-	-	-	-	-	
	Taschetto and England (2009)	1970-2005	↓	-	-	-	↓	↓	↓	-	-	-	-	-	-	Theobald et al. (2016): Analyzed changes in the mean rainfall in the Snowy region.
	Theobald et al. (2016)	1958-2012	-	-	-	-	-	↓	↓	↓	↓	↓	↓	↓	-	
Heavy day rainfall	Pilot study long	1889-2018	↑	↑	↑	↑	-	-	-	↑	↑	↑	↑	↑	↑	-
	Pilot study short	1950-2018	↑	-	-	-	-	-	-	-	-	-	-	-	-	
Number of heavy rainfall days	Pilot study long	1889-2018	-	↑	↑	↑	-	-	-	↓	↓	↓	-	-	-	T&E (2009): Analyzed the number of heavy (1 to 2 SD from mean) and very heavy (more than 2 SD from mean) rainfall events
	Pilot study short	1950-2018	↓	-	-	-	↓	↓	↓	↓	↓	↓	↓	↓	↓	
	Taschetto and England (2009)	1970-2005	-	-	-	-	↓	↓	↓	-	-	-	-	-	-	
	Theobald et al. (2016)	1958-2012	-	-	-	-	-	↓	↓	↓	↓	↓	↓	↓	-	
	Pilot study long	1889-2018	↓	n/a												-

Mean dry spell duration*	Pilot study short	1950-2018	-		
Maximum dry spell duration*	Pilot study long	1889-2018	↓	n/a	-
	Pilot study short	1950-2018	↑		
Extreme rainfall intensity	Pilot study long	1889-2018	↑	n/a	Gallant (2007): defined extreme rainfall as the 95th percentile of rainfall
	Gallant et al. (2007)	1910-2005	-		
	Pilot study short	1950-2018	↑	n/a	
	Theobald et al. (2016)	1958-2012	↑		
	Gallant et al. (2007)	1950-2018	-		
Extreme rainfall frequency	Pilot study long	1889-2018	↑	n/a	
	Gallant et al. (2007)	1910-2005	-		
	Pilot study short	1950-2018	-	n/a	
	Gallant et al. (2007)	1950-2005	-		
	Theobald et al. (2016)	1958-2012	-		

-	No significant Trends
	Trends that are present at more than 25% of the sites
n/a	The analyses for attribute/season not available
*	The colours are reversed since increase in the attribute indicates a drier climate

Having reviewed the statistical significance of trends, we now turn to an examination of the magnitude of trends for cases that are field significant in the pilot catchments and consistent with literature. The magnitude of trends are estimated site-wise using linear least squares regression at all 49 rainfall sites. The median absolute value of the trends and the median percentage of trends at the 49 sites are presented in Table 8.

Table 8. Median magnitude of trends at 49 pilot sites (x marks trends that are not field significant). Absolute trends per decade are presented first, followed by percentage change per decade in parentheses.

Attribute	Time Period	Median Trends				
		ANN	DJF	MAM	JJA	SON
Total Rainfall in mm/decade (%/decade)	1889 to 2018	x	3.4 (2%)	x	-2.4 (-1%)	x
	1950 to 2018	-16.8 (-2.1%)	x	-10.7 (-5.5%)	-3.7 (-1.5%)	-8.1(-3.7%)
Number of wet days in days/decade (%/decade)	1889 to 2018	-0.6 (-0.6%)	0.1 (0.7%)	-0.2 (-0.9%)	-0.3 (-1%)	-0.2 (-0.9%)
	1950 to 2018	-1.9 (-1.8%)	x	-0.9 (-4.4%)	-0.2 (-0.7%)	-0.9 (-3.6%)
Extreme rainfall Intensity in mm/day/decade (%/decade)	1889 to 2018	0.2 (0.5%)	n/a			
	1950 to 2018	0.3 (0.7%)				

The rainfall totals and the number of wet days in MAM and SON exhibit the largest trends, especially in the short term analysis. The spatial pattern of the magnitude of the short term trends in the number of wet days and total seasonal rainfall during MAM, JJA and SON are shown in Figures 8 and 9. The decreasing trends in the number of wet days are present at a number of sites across the study region during MAM and SON. The decreasing trends in seasonal totals are more widespread during SON.

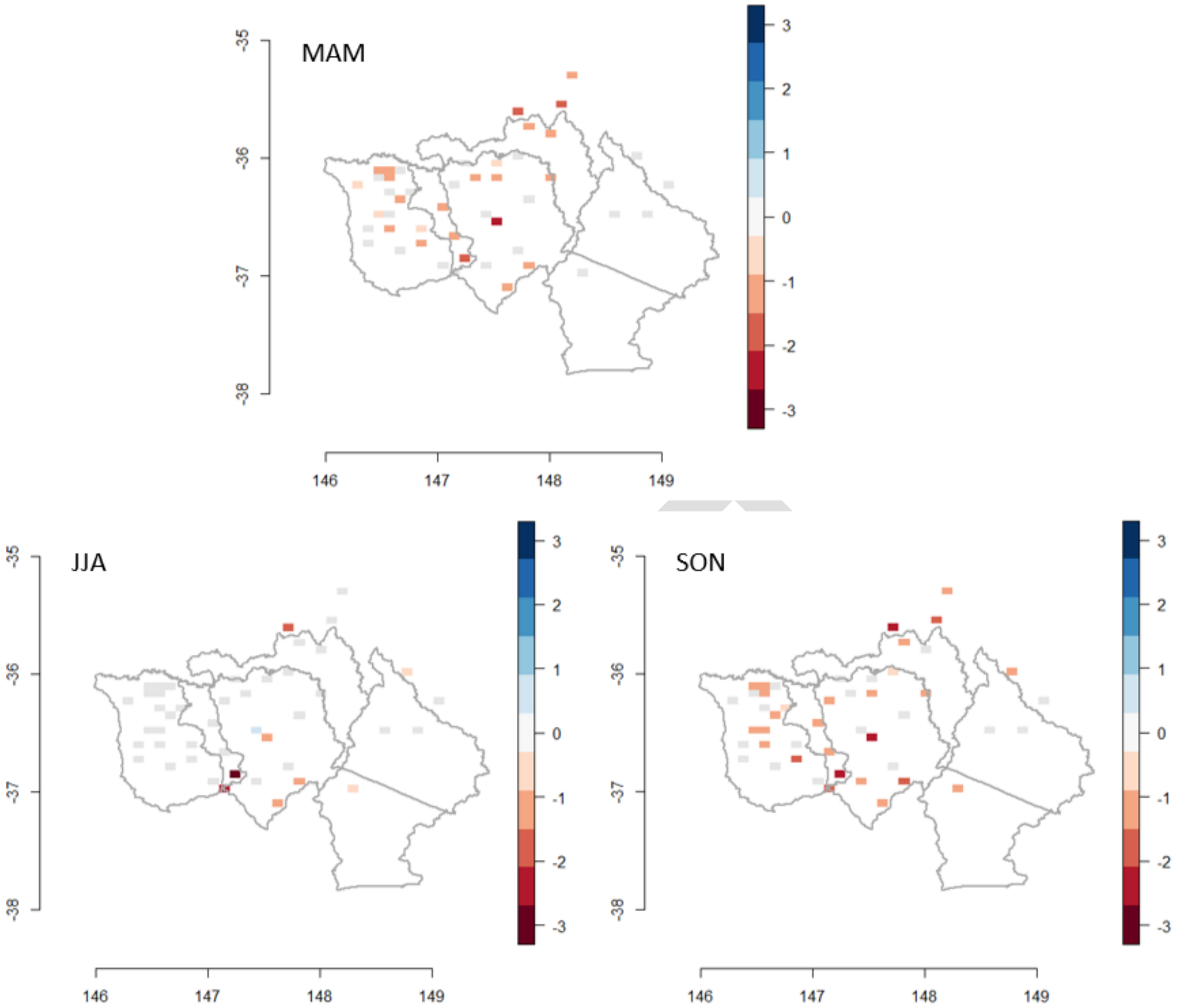

Figure 10 Short term (1950 to 2018) seasonal trends in number of wet days (in days/decade)

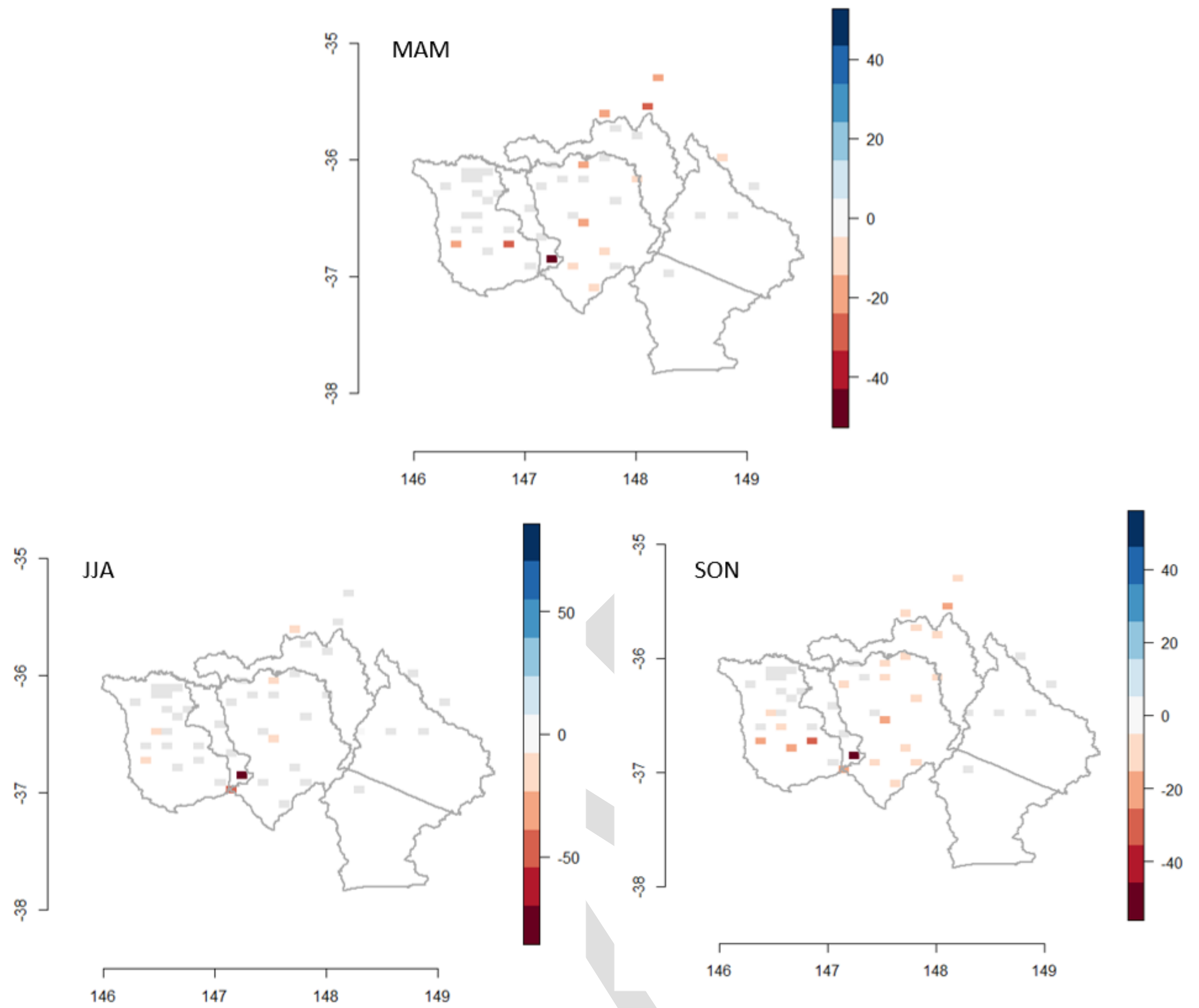

Figure 11 . Short term (1950 to 2018) seasonal trends in total rainfall (in mm/decade)

5.3.2 Trends in evapotranspiration

We examine the trends in evapotranspiration (Mwet) at the pilot sites over the full period of record (1889 to 2018) as well as a more recent short term period (1975 to 2018), consistent with the pan evaporation trends reported in literature. Table 9 shows the number of sites that exhibit significant trends in evapotranspiration during the two time periods of analysis. The long term (1889 to 2018) analysis shows an increasing trend in annual total Mwet. At seasonal scale, the increase occurs during MAM, JJA and SON. The trend in JJA is widespread with 97% of the sites exhibiting an increasing trend. In the short term (1975 to 2018) analysis, the annual and seasonal trends are not field significant.

Table 11 Number of sites that exhibit significant trends in annual and seasonal total evapotranspiration. The value in brackets indicate the number of sites as a percentage of the total number of sites (29 sites). Shading indicates trends that are field significant (at 5%).

Sign of Trend	**Total Mwet (mm)**									
	1889 to 2018					**1975 to 2018**				
	ANN	**DJF**	**MAM**	**JJA**	**SON**	**ANN**	**DJF**	**MAM**	**JJA**	**SON**
pos	8 (27%)	2 (7%)	8 (27%)	29 (97%)	9 (30%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (10%)
neg	1 (3%)	3 (10%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

none	21 (70%)	25 (83%)	22 (73%)	1 (3%)	21 (70%)	30 (100%)	30 (100%)	30 (100%)	30 (100%)	27 (90%)

We compare the evapotranspiration (Mwet) trends from the pilot sites with reported trends in pan evaporation in Table 10. It is to be noted that pan evaporation records are different from the Morton wet evapotranspiration data from the pilot sites; however we include this comparison here in the absence of trend studies in literature using estimated Morton wet evapotranspiration data. Literature reports decreasing trends in pan evaporation during the 1975 to 2002 period, possibly due to decreasing wind speeds or atmospheric demand (Roderick and Farquhar, 2004). The trends are reported to be reversed or insignificant in southeast Australia when recent observations up to year 2016/2018 are used for analyses (Ukkola et al. 2019; Stephens et al. 2018). The insignificant trends in the short term pilot analysis is consistent with the results of Ukkola et al. (2019).

Table 12 Trends in evapotranspiration in the pilot sites and the changes reported in literature

Attribute	Study	Period	Ann	Comment
Total Evaporation	Pilot study long	1889-2018	↑	The literature is uses pan evaporation records which is different from the estimated Morton wet evaporation used in the pilot study
	Pilot study short	1975-2018	-	
	Roderick and Farquhar (2004)	1975-2002	↓	
	Stephens et al. (2018)	1975-2016	↑/-	
	Ukkola et al. (2019)	1975-2018	-	

5.3.3 Trends in temperature

The trends in minimum and maximum temperatures from three ACORN SATv2 stations close to the pilot sites are presented here. The long term (1913 to 2018) trends are calculated using available data from two of the stations; the short term (1960 to 2018) trends are calculated using data from all three stations. Since the analysis uses only three sites we do not perform field significance test of the results, and instead report trends at individual sites.

All the temperature sites show significant increasing trends. Tables 11 and 12 present the magnitude of significant trends in Tmax and Tmin at the site-level data and the magnitude of trends reported in literature in this region. The trends in the ACORN-SATv2 sites near the pilot region shows significant increasing trends in Tmax and Tmin, consistent with literature. The magnitudes of trends are also in general consistent. The short term (1960-2018) increasing trends in Tmax and Tmin are slightly higher at these stations compared to the regional trends reported by Ashcroft et al. (2012).

Table 13 The magnitude of trends in Tmax at the ACORN-SATv2 sites and the trends reported in literature. All trends are in °C/decade

Study/Publication	Period	Pilot Station	ANN	DJF	MAM	JJA	SON
Pilot study long	1913-2018	82039	0.15	0.10	0.22	0.11	0.12
		72150	0.07	0.10	0.12	-	-
Ashcroft et al. (2012)	1860-2011		0.04	0.03	n/a	0.08	n/a
Ukkola et al. (2019) *	1910-2018		0.11	0.15	0.12	0.09	0.10
Jones (2012)	1910-2010		0.07	n/a	n/a	n/a	n/a
Pilot study short	1960-2018	82039	0.35	0.31	0.31	0.26	0.45
		72150	0.29	0.30	0.25	0.17	0.45

		72161	0.37	0.40	0.26	0.33	0.52
Ashcroft et al. (2012)	1960-2011		0.22	0.21	n/a	0.22	n/a

** Studied trends in mean temperatures as the average of Tmax and Tmin*

Table 14 The magnitude of trends in Tmin at the ACORN-SATv2 sites and the trends reported in literature. All trends are in °C/decade

Study/Publication	**Period**	**Pilot Station**	**ANN**	**DJF**	**MAM**	**JJA**	**SON**
Pilot study long	1913-2018	82039	0.18	0.29	0.14	0.12	0.17
		72150	0.25	0.26	0.29	0.19	0.20
Ashcroft et al. (2012)	1860-2011		0.07	0.10	n/a	0.05	n/a
Ukkola et al. (2019) *	1910-2018		0.11	0.15	0.12	0.09	0.10
Jones et al. (2012)	1910-2010		0.11	n/a	n/a	n/a	n/a
Pilot study short	1960-2018	82039	0.22	0.40	-	-	0.19
		72150	0.32	0.42	0.24	0.25	0.36
		72161	0.21	0.30	0.13	0.10	0.32
Ashcroft et al. (2012)	1960-2011		0.18	0.23	n/a	0.18	n/a

** Studied trends in mean temperatures as the average of Tmax and Tmin*

5.4 Results of split-sample testing

Four split sample simulations are performed using the historical data from the pilot sites to assess the ability of the stochastic model to capture the attributes of rainfall and evapotranspiration in recent observations, following the methodology outlined in section 5.2.2. The split sample tests consists of the 1990 reference and drought reference experiments. The mean attributes of the simulated data are compared with the mean attributes of observations during the calibration and validation time periods in the subsections below.

5.4.1 Rainfall

The trend analysis showed significant trends in cool season rainfall totals, annual/cool season number of wet days, and annual extreme rainfall intensity at the pilot sites. Here we examine these attributes in the simulated data with respect to the mean values in observations during the calibration and validation periods. The results of the experiments for some key attributes of rainfall are examined in this section. The figures showing the split sample results for other attributes are included in Appendix A for brevity.

Figure 10 shows histograms of the differences in the mean MAM rainfall totals during the calibration and validation time periods from the MAM seasonal rainfall totals in the simulated time series. The 1990-reference split sample tests are shown in Figure 10a-b. When stochastic models are calibrated using observed data up to year 1989, the simulated data shows substantial differences (42 to 58 mm) when compared to the post-1990 validation period. When the millennium drought is included in the calibration period, the MAM rainfall during the validation period is closer to both the calibration period and the stochastically simulated data (Figure 10c-d). For example, the simulations exhibit biases of 42 mm (+24%) with respect to the validation period in one of the 1990-reference experiments; the biases reduce to 10 mm (+4%) when the drought is included in the calibration period. The histograms of total rainfall during JJA and SON are available in Appendix A. In each of these experiments, the seasonal rainfall totals during the calibration period are close to the simulated data, as expected. If the seasonal rainfall totals during the validation time periods are different from that in the calibration time period, the mean values during the validation period are different from the simulated data as well. The mean monthly rainfall is used for the calibration of the stochastic model, and the simulated time series closely replicate this statistic during the calibration period.

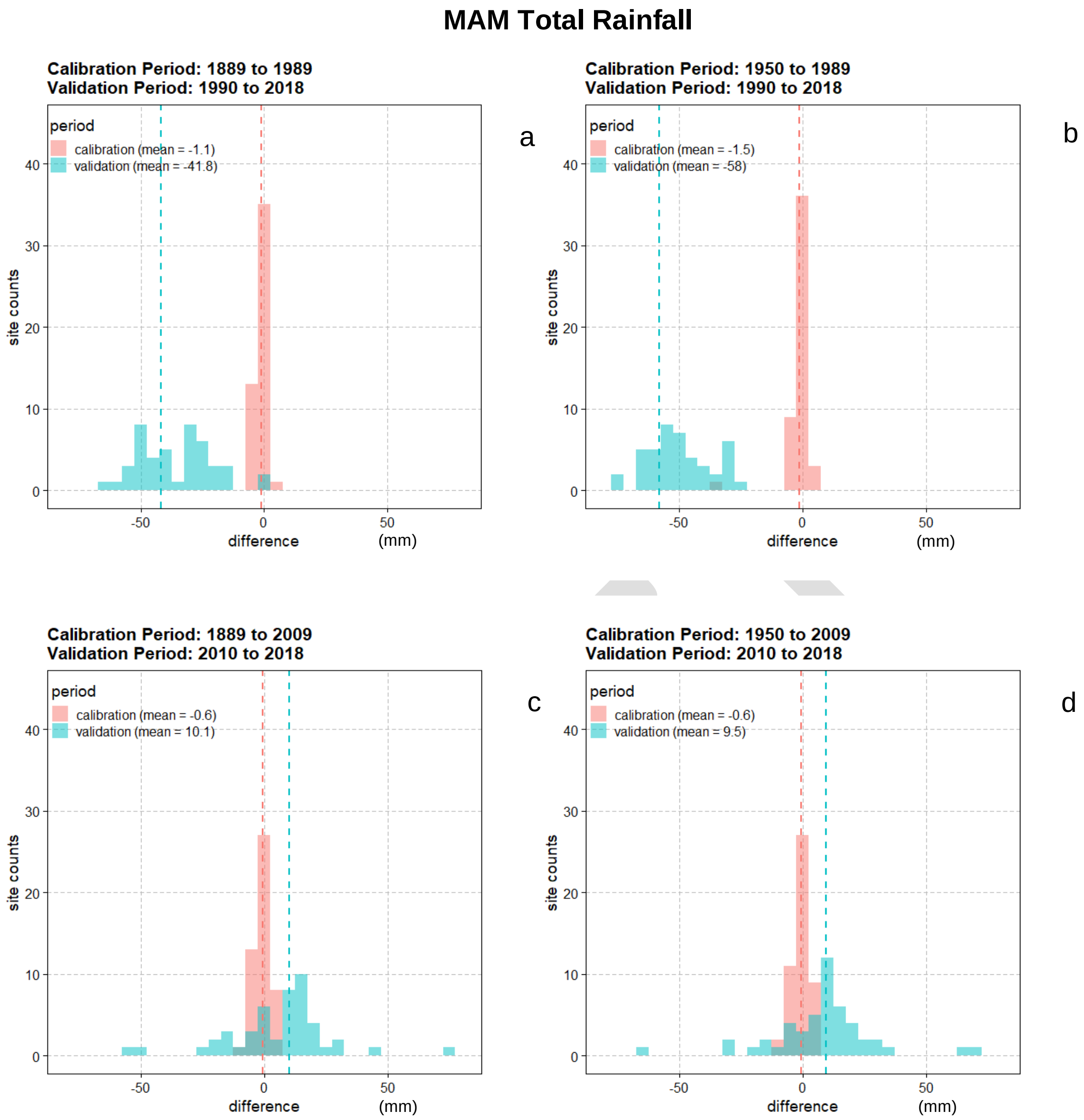

Figure 12 Histograms of the mean differences in total rainfall in MAM during the calibration and validation time periods from stochastic simulations (in mm) at 49 pilot sites (a) 1990-reference split sample test using a long calibration period, (b) 1990-split sample test using a short calibration period, (c) drought-reference split sample test using a long calibration period, and (d) drought-reference split sample test using a short calibration period. The dashed vertical lines mark the means of the respective histograms.

The results of the split sample tests in the MAM seasonal rainfall totals are consistent with the significant decreasing trends detected from the analysis of trends at the pilot sites. The mean total rainfall during MAM, JJA and SON (Appendix A) are lower during the recent validation periods compared to the earlier calibration periods used for the simulations consistent with the negative trends detected in these seasons.

While the inclusion of the drought in the calibration period reduces the biases in MAM total rainfall in the simulations, there are increases in biases in some other statistics. The histograms of the total rainfall during DJF is shown in Figure 11. The simulations underestimate the rainfall during DJF in the 1990-reference experiments (-32mm, -16%). After the inclusion of data up to the year 2009 in the calibration, the biases increase (-72 mm, -29%). This change is in general consistent with the long term positive trend in total DJF rainfall. But the large positive difference during the recent validation period (+69 to 71 mm) could be due to statistically insignificant changes in the most recent records.

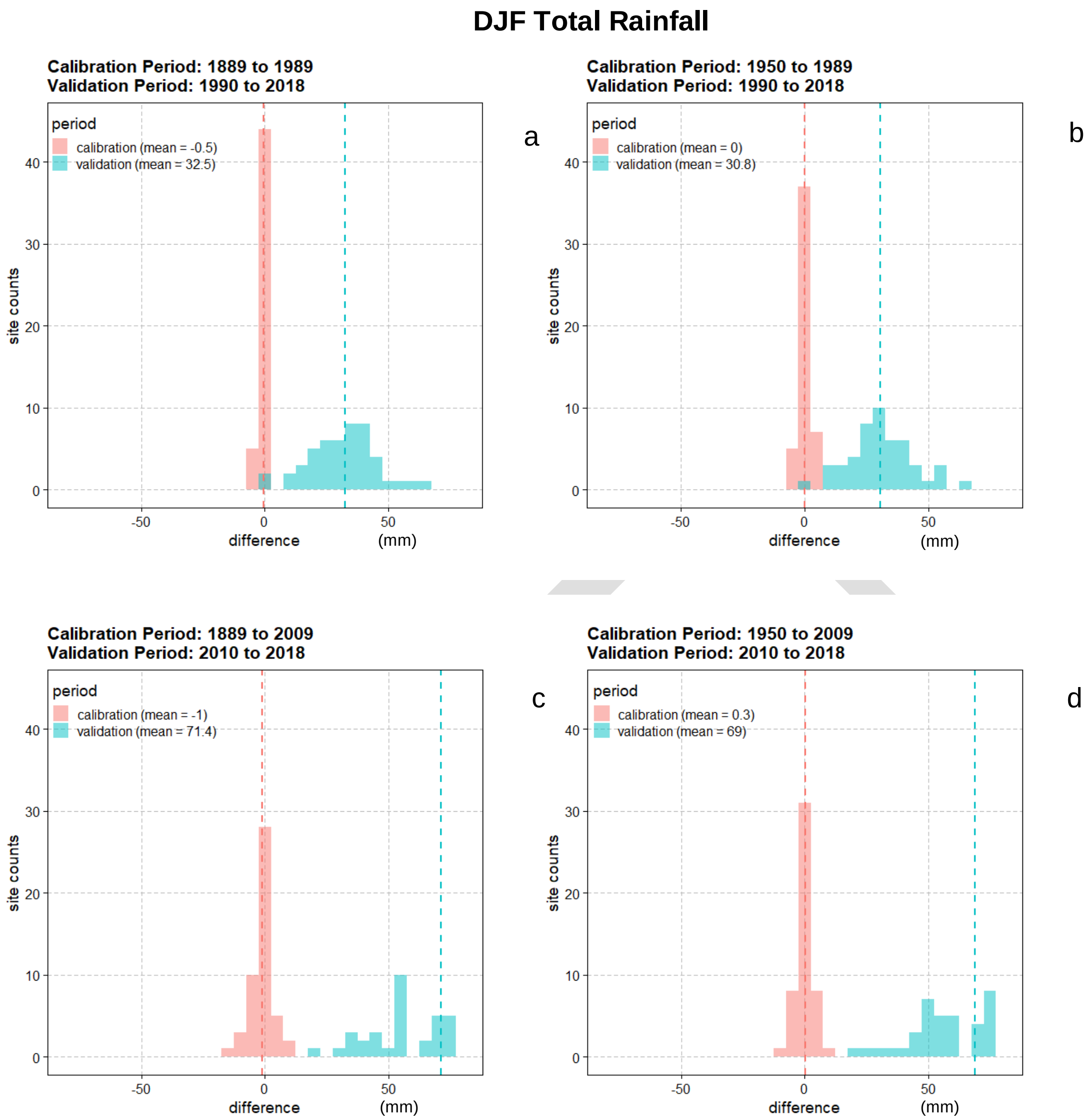

Figure 13 Similar to Figure 10, but for total rainfall during DJF (in mm).

Figure 12 shows the histograms of the differences in the number of wet days during MAM during the calibration and validation time periods from the stochastically generated data. When compared to the data during the calibration time period, the stochastic simulations underestimate the number of wet days in all four experiments. The number of wet days is not a statistic that is used in the calibration of the model, and the simulated time series show a slight bias in this statistic with respect to the calibration period data (1 to 1.4 days per season).

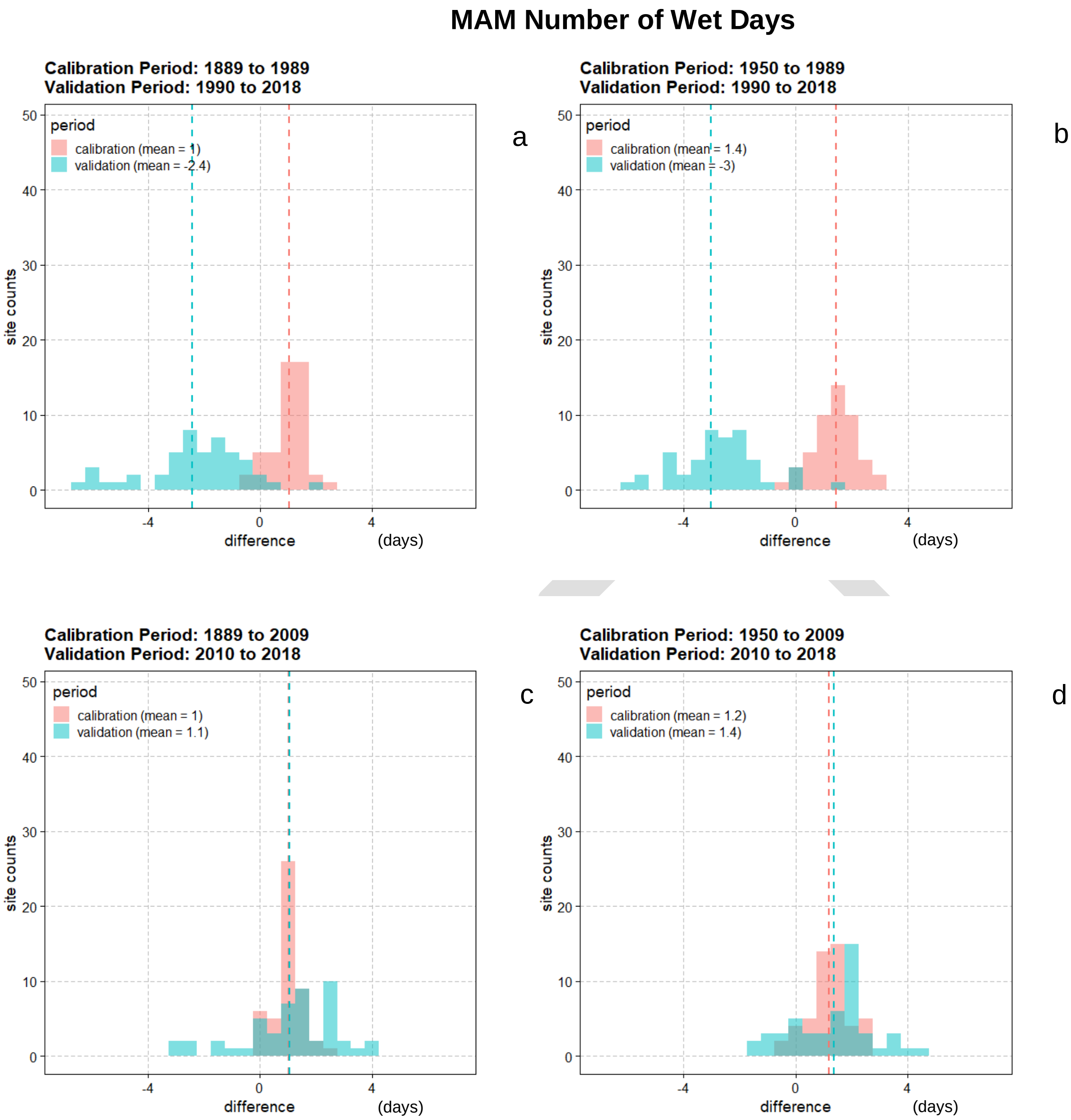

Figure 14 Similar to Figure 10, but shows the histograms of the mean differences in the number of wet days in MAM during the calibration and validation time periods from stochastic simulations (in days) at 49 pilot sites.

In the 1990 reference experiments, the mean number of wet days during the validation period are lower than that during the calibration period, consistent with the negative trends detected in this attribute (Table 8). The mean number of wet days during the validation period is lower than that in the simulations by 2.4 to 3 days per season. When the millennium drought is included in the calibration period, the number of wet days during the calibration and validation time periods are very similar. Hence in the drought reference split sample tests, the simulations underestimate the number of wet days with respect to both the calibration and validation time period statistics.

The histograms of the differences in the number of wet days during the other seasons (JJA, SON and DJF) are available in Appendix A. Similar to the results during MAM, the mean number of wet days from the stochastic simulations are biased lower compared to the calibration period (by 0.8 to 2.7 days per season) for all seasons and both time periods of calibration. The differences of the mean statistics during the validation period vary. The mean number of wet days in JJA and SON are lower during the validation time periods compared to the calibration time periods, consistent with the trends detected in the pilot sites. In contrast, the number of wet days in DJF during the validation period are higher especially during the 2010 to 2018 validation period (+4.1 to +4.3 days per season). While the long term trend in the number of wet days in DJF are positive at the pilot sites, there are no significant short term trends during this season. So this positive signal in the number of wet days in DJF during the most recent validation period could potentially be due to statistically insignificant trends in the recent record.

The histograms of annual extreme rainfall intensity are available in Appendix A. The annual mean extreme rainfall intensity during the calibration period is close to that in the simulated data. During the validation period, the rainfall intensity is higher compared to the simulations, consistent with the positive trends at the pilot sites.

To summarize,

- **The millennium drought is a 'high-leverage' event:** the statistics of the simulated rainfall can vary significantly depending on whether the drought is included in the calibration period or validation period.
- **The inclusion of the millennium drought improves some statistics during the validation period, but leads to a deterioration in others:** The simulations show biases with the respect to the validation period data, and the sign and magnitude of biases vary with season and time period. In the 1990-reference split sample test, both the validation period mean MAM rainfall and the MAM number of wet days are biased higher in the simulations. The statistics match better in the drought-reference tests once the millennium drought is included in the calibration period. However, other statistics like the total rainfall during DJF and annual extreme rainfall intensity show larger biases in the drought reference experiments.

5.4.2 Evapotranspiration

The mean annual total evapotranspiration from the stochastic simulations is compared with the mean annual evapotranspiration during the calibration and validation time periods in Figure 13. The simulations match the annual evapotranspiration values during the calibration time period in all the simulations. In the 1990 reference split sample tests, the mean evapotranspiration during the validation period are higher than the simulated totals (and calibration period) by 18 to 20 mm. When the model is calibrated using data up to year 2009, the annual totals from the simulations are closer to the annual totals during the validation period (mean difference = 8 to 11 mm). The split sample tests using shorter and longer calibration periods do not show any major influences on the differences between the simulations and the validation period.

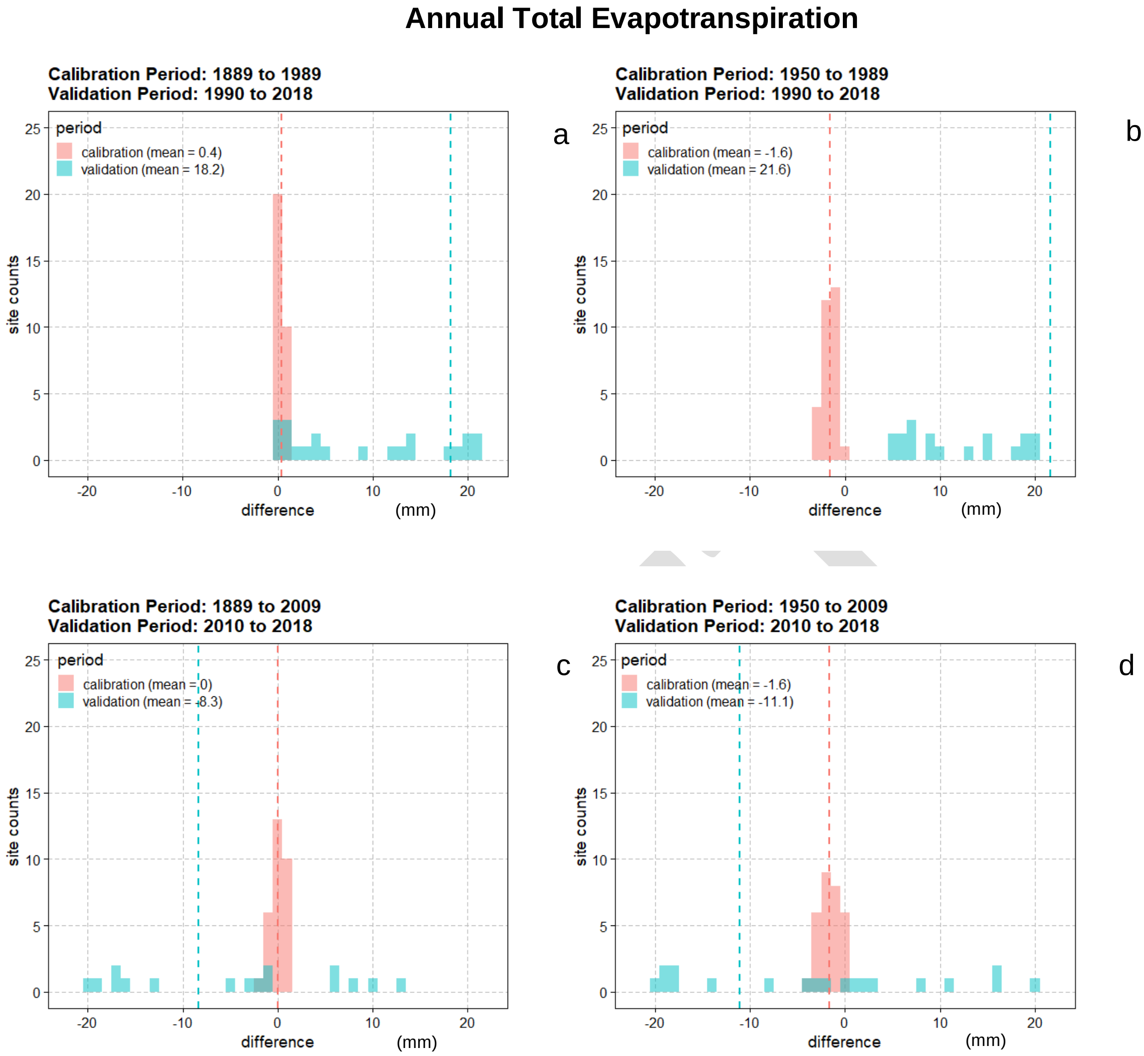

Figure 15 Histograms of the mean differences in total annual evapotranspiration during the calibration and validation time periods from stochastic simulations (in mm) at 30 pilot sites (a) 1990-reference split sample test using a long calibration period, (b) 1990-split sample test using a short calibration period, (c) drought-reference split sample test using a long calibration period, and (d) drought-reference split sample test using a short calibration period. The dashed vertical lines mark the means of the respective histograms.

6 Summary of key findings from assessment of non-stationarity

Based on an analysis of multiple lines of evidence, comprising a review of available literature and the assessment of pilot sites in Ovens, Upper Murray and Snowy catchments using both Mann Kendall and split sample tests, there is evidence of non-stationarity in the record for rainfall and temperature. The non-stationarity in evapotranspiration is not as well established, likely in part due to the different processes driving the different 'types' of evapotranspiration (e.g. pan versus Morton's). Specific findings varied depending on the specific dataset used (e.g. gridded versus point; whether the data has been homogenized or not; the spatial domain of the data; the temporal period of analysis) but broad conclusions are:

- **Temperature records are non-stationary:** Literature documents temperature increases in this region, especially post 1960 (Jones, 2012; CSIRO and BOM, 2015). The 'Climate Change in Australia' initiative reports an increase of 0.8°C in the Murray basin cluster (which contains the pilot catchments) over the 1910-2013 time period assuming a linear trend, with higher trends for temperature minima than for maxima; this is broadly consistent with a mean temperature increase over Australia by just over 1°C during the slightly longer period from 1910 to 2018. Climate projections indicate that increases of 0.6 to 1.3 degrees are expected in the Murray Basin cluster in the near-term (2020 to 2039) with respect to a baseline of 1986-2005. The homogenized temperature sites near the pilot region analysed here also show statistically significant increases. The increases amount to 0.8-1.5°C in maximum temperatures and 1.9-2.6°C in minimum temperatures during the period 1913-2018.

- **Cool-season and annual rainfall totals are non-stationary:** Literature documents declines in cool season (April to October) rainfall by 10-20% in southeast Australia since the mid-1990s, predominantly in autumn and early winter. The trend assessment using data from the pilot sites are in agreement with literature and shows short-term decreases in autumn, winter, and spring rainfall totals. The short term (1950 to 2018) trends are strongest in autumn – the median decline in autumn rainfall at the pilot sites amount to 5.5%/decade for the period 1950 to 2018; the median declines in annual total rainfall at the pilot sites amount to 2.1%/decade for the period 1950 to 2018.
- **There are trends in multiple attributes of rainfall:** Literature reports decreasing trends in the number of wet days during the cool season; the signal also exists in the data from the pilot sites. In addition, the pilot sites show a short-term decline in spring (SON) rainfall. This result is not consistent with other large scale studies in this region, but is consistent with regional studies in nearby catchments. There is an increasing trend in annual extreme rainfall intensity at the pilot sites. There is less consensus in literature on extreme rainfall intensity; however, the trend in the pilot sites is consistent with reported trend in a nearby catchment.
- **Non-stationarity in the evapotranspiration is not as well established:** Literature documents negative trends in pan evaporation over the period 1975-2002, whereas studies using more recent data report insignificant/increasing trends. There are long term increases in annual Morton wet evapotranspiration at the pilot sites; the short-term trends are not statistically significant. The pilot study results are not directly comparable with available literature based on pan evaporation data, and so some uncertainty remains regarding the non-stationarity of evapotranspiration data.
- **Multiple climate drivers influence the rainfall in southeast Australia:** Literature documents that the declining cool season rainfall is associated with an expansion of the tropics, increasing intensity of the sub-tropical ridge over the continent and positive trends in the Southern Annular Mode (SAM). Literature indicates that these changes in large scale patterns during the cool season are at least partly attributable to climate change (CSIRO, 2012; Hope et al., 2017). Other climate drivers, notably El Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD), influence the inter-annual variability regional rainfall, primarily affecting rainfall in winter and during the warm season (CSIRO and BOM, 2015; Hope et al., 2017).
- **The period used for calibration influences the statistics of the simulated stochastic data:** The split sample tests show that the millennium drought is a ‘high-leverage’ event, in the sense that statistics of the simulated rainfall can vary significantly depending on whether the drought is included in the calibration period or validation period. When the drought is included in the calibration period, the match between the simulations and observed data post the drought is dependent upon the statistic and season under consideration. The inclusion of the drought in the calibration period brings the autumn rainfall and number of wet days in the simulated data close to recent (2010-2018) observations, but results in a larger deviations in simulated summer rainfall, number of wet days in summer, and extreme rainfall intensity. For example, when the drought is excluded from calibration, the simulations show mean biases in the MAM rainfall during the validation period of +42 mm (+24%). When the drought is included in calibration the bias reduces; the simulations show biases in total MAM rainfall during the validation period of -10 mm (-4%). But in the same tests, the bias in DJF rainfall increases from a mean of -32 mm (-16%) to -71 mm (-29%) with the inclusion of the drought in the calibration period.

Although the pilot analysis does not seek to attribute trends to a specific driver (whether it be natural climate variability, anthropogenic climate change or other potential drivers of change), several studies have attributed the temperature shifts/trends and cool season rainfall trends to anthropogenic influences (CSIRO, 2012; Jones, 2012; CSIRO and BOM, 2015, Hope et al., 2017). The SEACI synthesis report states that "*the decline in rainfall across south-eastern Australia was at least partly attributable to climate change*" (CSIRO, 2012). The 'Climate Change in Australia' report notes that the literature on rainfall attribution in southern Australia is based on "*implied attribution*" using inferred causality from the attribution of large-scale drivers. The report states that more formal attribution to definitive causes are yet to be established due to a range of uncertainties, but at the same time noting that the "*drying across southern Australia cannot be explained by natural variability alone*" (CSIRO and BOM, 2015).

Future regional projections indicate a warmer and drier climate with respect to a baseline of 1986-2005. In the Ovens Murray catchment, VCP19 projections (Clarke et al. 2019b) report annual rainfall decreases during the 2020 to 2039 period. The projected median change is -6% (range of -12% to -4%) under the medium emissions scenario, and -11% (range of -18% to -3%) under high emissions scenario with further declines anticipated in subsequent decades. Climate projections for Murray Basin cluster indicate increases in potential evapotranspiration in all seasons; VCP19 projects 8 to 10% increases in pan evaporation by 2030s (2020-2039). The projected changes for the Murray Basin cluster is based on Morton's wet potential evapotranspiration, whereas the pan evaporation projections from VCP19 is based on pan evaporation modelled directly by the climate model.

7 Options for stochastic time series generation in the Southern Basin

Due to reported trends in the historical record for the Southern Basin, which is at least partly attributable to anthropogenic influences, there is a need to consider the implications of non-stationarity on the methodology utilized for the stochastic risk assessment. The working definition of 'non-stationarity' used here is that the parameters of the stochastic model (and thus the statistics associated with the stochastically generated data) vary with time. This time-variation means that stochastic results need to be reported relative to a time period, with the following terminology used in this report:

- **Historical record**. This is used synonymously with the instrumental record (although noting that paleo-records of the IPO are used to inform the stochastic sequences in previous studies). For the purposes of this report, the 'historical record' is usually interpreted to mean the period from 1890-2018 as this is the period of in-filled data available from the SILO database.
- **Baseline period**. The term 'baseline' is usually used in context with climate projections, which are usually reported relative to a climatological baseline. The specific baseline period will depend on the climate models used and various other factors, with the World Meteorological Organization suggesting defining climate baselines of at least 30 years in duration (WMO, 1989). Further studies document that a baseline of 40 years is required for hydrological investigations in southeast Australia to capture the significant year-to-year variability associated with key processes (Potter et al., 2016).
- **Current climate**. This is interpreted here to mean the risk at the current time (e.g. the year 2020, which is the date this report is written) of for a window centered on the current time.
- **Future climate**. This can refer to any future period for which projections are available, and is usually reported as a 'static' estimate using some window, typically of length 20-30 years. Consistent with WMO applications, hydrological applications usually benefit from longer windows to increase the signal-to-noise ratio; however noting that there may be significant changes over this time period (i.e. the projected data within a 30 year window may not be stationary). It is noted that near-term future climate projections have started encompassing the current climate; for example VCP19 uses a future window of 2020-2039 and thus encompasses the current year.

To help illustrate these terms, the temperature in Victoria is reported to have increased 0.5° over the historical record from 1910 to 1995, and the current (2018) temperature is about 0.6° warmer than the 1986 to 2005 baseline (Clarke et al. 2019a). Future temperature projections suggest increases of between 0.5° and 1.3° for the period 2020 to 2039, relative to this same baseline.

The stochastic method that has been used for the northern basins is based on the historical record and simulated to be conditional to the Interdecadal Pacific Oscillations, with the implicit assumption that the climate is stationary over this period. These simulations are intended to capture variability in the baseline historical record, which was composed of (SILO infilled) climatic measurements from the period 1890 to 2018 as well as IPO distribution dwell times informed by paleoclimatic records. DPIE developed additional methodology to amalgamate the variability from the stochastic replicates with climatic signals informed by RCMs. This method applies NARCliM based scaling factors to the stochastic model outputs.

There are a number of trade-offs associated with modifying this methodology to incorporate climate non-stationarity, balancing the complexity of the climate signal, the strength of various non-stationarity elements as compared to the underlying variability, and the variety of methods that may be available to explicitly accommodate known trends. Here we detail a range of options available to account for the presence of non-stationarity in stochastics time series generation. The options vary in complexity.

1. Use the entire 'historical' record to calibrate the stochastic model.

This is the simplest option, forming a naïve default. The advantage of this option is that it makes full use of the available historical data (1889 to 2018), which improves the statistical precision of parameters used in the stochastic model. The disadvantage is that where there are significant long-term shifts the historical record (as opposed to 'natural' variability including interdecadal variations associated with the IPO), the generated sequences may not be representative of 'current' climate. Moreover, it is assessments of future risk based on climate model results are invariably derived with respect to a historical baseline, and the extent to which the full (1889 to 2018) historical record is reflective of any given climatological baseline used for climate projections is unclear and is likely to depend on the specific climate variable, as well as the statistics for those variables (given the different findings for including/excluding the millennium drought depending on whether one was looking at MAM seasonal total, MAM number of wet days, JJA seasonal total or extreme rainfall intensity).

2. Use a climatological baseline to calibrate the stochastic model.

This option involves using a shorter climatological baseline for calibration of the model, which can then be scaled using climate model outputs to derive both 'current' and 'future' estimates of variability. The baseline could be chosen to be consistent with the NARCliM scaling factors that are used for future projections by DPIE. The benefit of this approach is that it maintains basic consistency with NARCliM by having an identical baseline. The NARCliM 1.0 baseline is 1990 to 2009, and the NARCliM 1.5 baseline is 1950 to 2005. The method is thus suitable for NARCliM scaling based current and future climate applications.

The tradeoff is that the initial years of data (say, 1890 to 1950/1990) are no longer utilised and there is a corresponding loss of precision in the parameters used as the basis of the stochastic model. The significant interannual and interdecadal variability in particular means that depending on whether one includes (say) the Millenium drought as part of the baseline will make a significant difference on the simulated results.

It is noted that this baseline will neither reflect 'current' or 'future' climate, and thus climate model outputs will be required to adjust the stochastic sequences to reflect these time periods. This requires decisions on the climatological baselines to be used for analysis, and if the NARCliM outputs are to be used, then this will depend on the specific version of NARCliM. In particular, the NARCliM 1.0 baseline of 20 years and, for the reasons described above, is too short to capture the 'historical' climate. To address, as already discussed the World Meteorological Organization recommends a minimum baseline length of 30 years (WMO, 1989). Guidelines developed under the Victorian Climate Change Initiative (VicCI) recommends a baseline of at least 40 years, given the high inter-annual variability in rainfall and runoff in this region

(Potter et al., 2016). Another disadvantage of the method is that the use of simple scaling to represent future climate would not reflect changes in statistics like the number of wet days or extremes.

Another challenge associated with this approach is that scaling of the 'baseline' climate using climate models may be present significant challenges. In particular, multiplying rainfall by seasonal 'change factors' will mean that the number of wet days will stay constant and the extremes will change in the same manner as seasonal rainfall totals, which is inconsistent with the findings of the non-stationarity analysis presented above. It may be that these other statistics of 'second order' importance compared to seasonal and annual totals depending on the hydrological application (e.g. long-term catchment yield assessments), although this would need to be tested. Alternatively, more complex scaling options such as quantile-based approach could be used to jointly change the seasonal totals, wet days and extremes in a manner that reflects the climate projections.

There are several further variants that could be considered for option (2) that involves using a 'hybrid-baseline', but they present additional challenges that are likely to over-complicate the methodology. We detail these variants as option 3 below.

3. Use a hybrid-baseline of the "historical" climate to calibrate the stochastic model.

One option is to utilize a hybrid-baseline period. For example, some statistics may have negligible or no shift between a baseline and a longer historical period. Therefore, additional precision would be achieved by calibrating selected statistics to a longer record, while reserving only some key statistics (such as the mean value) to the baseline period. The challenge with this approach is that the term "baseline" is no longer informative, since it is not clear what period is being represented in the calibration.

Another option is to scale the historical record up to the values observed in the baseline period. This is one of the approaches to address non-stationarity recommended by the expert review panel. The benefit of this approach is that it provides a mechanism for utilizing the entire record. A challenge with this approach is that it requires the scaling of the prior historical record to be representative. This procedure is unlikely to be straight forward, because:

Multiple statistics are changing in different ways, so that scaling by seasonal or annual totals may not lead to appropriate adjustments of the other statistics. In particular, the pilot study identified negative trends in cool season totals, negative trends in number of wet days and positive trends in extreme intensity, which would require care in how the historical observations are mapped.

Historical non-stationarity is likely to encompass a combination of 'natural' climate variability and anthropogenic climate change, so one would need to assess the magnitude of historical change *specifically associated with anthropogenic climate change* in order to scale the historical values appropriately. If this is not done, then aspects of the signal associated with natural variability may also inadvertently be removed (at least in part), which is unlikely to be desirable.

4. Use the 'inverse' method to generate a 'current' or 'future' climate.

This approach could be used to target statistic values representative of a particular period, such as the 'current' or 'future' climate. Unlike traditional stochastic generation methods in which the stochastic generator parameters are estimated to achieve the best performance over a calibration dataset, the 'inverse' method defines a set of 'target' statistics (e.g. total annual rainfall, seasonal rainfall, number of wet days, intermittency, extremes, etc) and then calibrates the stochastic generator to those statistics. The advantage of this method is that it is able to target the most recent values corresponding to current climate, whereas other methods (as in option 2) would use a baseline of 'historical' climate spanning the selected years, and which may not reflect 'current' climate.

The key challenge of this method is defining the target statistics reflecting current or future climate. This is not a trivial exercise, and may require a combination of examining historical climate trends together with climate modelling evidence. Moreover, this approach has not yet been implemented in large-scale real-world analyses, and thus would require a period of method testing prior to widespread implementation.

5. Use weather typing to capture specific mechanisms of generation of precipitation.

This method involves stratifying the historical record by synoptic types (such as ECLs) to provide a scaling method that can explicitly account for mechanisms with strong expected change. The methodology would involve additional complication introduced by the multiple weather types. There is also a challenge that it may 'double count' the effects of scaling when deployed alongside methods that are needed to scale other components. For example, if ECLs were introduced as a scaling category with their own relationships to future climate, other scaling relationships (e.g. shift in seasonality, shift in extremes) would need to be developed for the specific case of non-ECL weather events. This method is likely to be highly prospective given the potential of 'double counting' the effects of synoptic meteorology if they are also explicitly included in climate model projections.

The trade-offs involved in the five options detailed above are summarized in Table 12 below. *It is noted that the discussion of method advantages and disadvantages are predicated on the assumption that the historical record is non-stationarity and that at least part of that non-stationarity is attributable to anthropogenic climate change.*

Table 15 Summary of options for stochastic time series generation in the Southern basin, with advantages/disadvantages described in the context of a non-stationarity climate signal

Option	Advantages	Disadvantages	Application
1. Entire record	Long record, historical enables increased precision of parameter estimation	Neither reflective of current climate nor a climatological baseline	Reflection of long term historical climate
2. Historical baseline	Matches NARCliM baseline and thus can be used as the basis for 'current' and 'future' climate assessments	Need to choose which NARCliM, noting that NARCliM 1.0 is too short relative to WMO & VicCI recommendations Not reflective of current climate and thus requires some level of subsequent processing Simple seasonal scaling would not reflect changes in wet days and other statistics, although there may be alternative (e.g. quantile-based) methods that could address this issue	NARCliM-derived 'current' and 'future' climate applications
3. Hybrid historical baseline	Uses full historical record Possible extra precision in some attributes compared to baseline only	Advantage of extra data before baseline is unclear There are multiple complicating factors associated with implementation of this method, including complex variations with key attributes, and the need to separate the natural and anthropogenic components of any	Adjustments would enable stochastic data to reflect 'current' climate

		historical trends prior to adjustments	
4. Inverse method	Can be implemented to match current and future climate Allows changes in all key attributes	Difficulties Method development and testing required	Could be designed to reflect 'current' and 'future' climate
5. Weather typing	Allows for specific mechanisms (ECLs)	Additional complication introduced by multiple types Scaling is complicated by types, and there exists a significant possibility of 'double counting'	Specific allowance for future changes to weather types rather than simple scaling

Another point raised by the independent review panel is the role of multiple climate drivers in the southern region. Evidence in literature indicates a strong trend in the SAM during the cool season, while the warm season is influenced by ENSO and IOD. Note that previous stochastic generation methodology based on climatic relationship to the IPO relied on partitioning the climate under the assumption of stationarity in the IPO. This is conceptually different from assuming a trend signal in the relevant climate drivers, since it does not require projection of the trend associated with the climate driver. Also, in locations where the partition between IPO positive and negative shows negligible difference, the partition collapses back to the underlying marginal model (i.e. the model does not become 'worse' by inclusion of the IPO in areas where the IPO is not important).

Although there is the potential to develop a non-stationary stochastic model that directly conditions on trend properties (such as SAM), this is inadvisable. Apart for the many technical methodological challenges of developing the non-stationary model and conditioning it on an RCM, any application of RCM-derived scaling factors of meteorological variables (such as temperature, rainfall and evapotranspiration) would then need to exclude the role of that driver as part of the analysis to avoid double counting this effect (given that the intent of the stochastic conditioning is to have the influence of the relevant driver already included in the analysis). Therefore, this is not presented as a viable option.

7.1 Recommendation for stochastic generation

SECTION TO BE COMPLETED

8 References

Alexander, L. V., & Arblaster, J. M. (2009). Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. In *International Journal of Climatology* (Vol. 29, pp. 417–435). John Wiley & Sons, Ltd. https://doi.org/10.1002/joc.1730

Alexander, L. V., & Arblaster, J. M. (2017). Historical and projected trends in temperature and precipitation extremes in Australia in observations and CMIP5. *Weather and Climate Extremes*, *15*, 34–56. https://doi.org/10.1016/j.wace.2017.02.001

Alexander, L. V., Hope, P., Collins, D., Trewin, B., Lynch, A., & Nicholls, N. (2007). Trends in Australia's climate means and extremes: A global context. *Australian Meteorological Magazine*, *56*(1), 1–18.

Allen, K. J., Anchukaitis, K. J., Grose, M. G., Lee, G., Cook, E. R., Risbey, J. S., et al. (2019). Tree-ring reconstructions of cool season temperature for far southeastern Australia, 1731–2007. *Climate Dynamics*, *53*(1–2), 569–583. https://doi.org/10.1007/s00382-018-04602-2

Ashcroft, L., & Karoly, D. J. (2012). Temperature variations in southeastern Australia, 1860-2011. *Australian Meteorological and Oceanographic Journal*, *62*, 227–245.

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. https://doi.org/10.1016/j.jhydrol.2016.12.043

BOM, & CSIRO. (2019). *State of the Climate 2018*.

Clarke, J., Grose, M., Thatcher, M., Hernaman, V., Heady, C., Round, V., et al. (2019a). *Victorian Climate Projections 2019 Technical Report*. Melbourne, Australia.

Clarke, J., Grose, M., Thatcher, M., Round, V., & Heady, C. (2019b). *Ovens Murray Climate Projections 2019*. Melbourne, Australia.

CSIRO. (2012). *Climate and water availability in south-eastern Australia A synthesis of findings from Phase 2 of the South Eastern Australian Climate Initiative (SEACI)* (Vol. 41). Australia. Retrieved from http://www.seaci.org

CSIRO, & BOM. (2015). *Climate Change in Australia Information for Australia's Natural Resource Management Regions: Technical Report*.

DELWP. (2016). *Guidelines for assessing the impact of climate change on water supplies in Victoria*.

Do, H. X., Westra, S., & Leonard, M. (2017). A global-scale investigation of trends in annual maximum streamflow. *Journal of Hydrology*, *552*, 28–43. https://doi.org/10.1016/j.jhydrol.2017.06.015

Dowdy, A. J., Mills, G. A., Timbal, B., Griffiths, M., & Wang, Y. (2013). *Understanding rainfall projections in relation to extratropical cyclones in eastern Australia*. *Australian Meteorological and Oceanographic Journal* (Vol. 63).

Dowdy, A. J., Pepler, A., Di Luca, A., Cavicchia, L., Mills, G., Evans, J. P., et al. (2019). Review of Australian east coast low pressure systems and associated extremes. *Climate Dynamics*, *53*(7–8), 4887–4910. https://doi.org/10.1007/s00382-019-04836-8

Fiddes, S. L., Pezza, A. B., & Barras, V. (2015). Synoptic climatology of extreme precipitation in alpine Australia. *International Journal of Climatology*, *35*(2), 172–188. https://doi.org/10.1002/joc.3970

Gallant, A. J. E., Hennessy, K. J., & Risbey, J. (2007). Trends in rainfall indices for six Australian regions: 1910-2005. *Australian Meteorological Magazine*, 223–239. Retrieved from https://www.researchgate.net/publication/242601092

Hewaarachchi, A. P., Li, Y., Lund, R., & Rennie, J. (2017). Homogenization of daily temperature data. *Journal of Climate*, *30*(3), 985–999. https://doi.org/10.1175/JCLI-D-16-0139.1

Hope, P. ., Timbal, B. ., Hendon, H. ., Ekström, M. ., & Potter, N. (2017). *A synthesis of findings from the Victorian Climate Initiative (VicCI)*. Retrieved from http://www.bom.gov.au/research/projects/vicci/

Jones, R. N. (2012). Detecting and attributing nonlinear anthropogenic regional warming in southeastern Australia. *Journal of Geophysical Research: Atmospheres*, *117*(D4), n/a-n/a. https://doi.org/10.1029/2011JD016328

Karoly, D. J., & Braganza, K. (2005). A new approach to detection of anthropogenic temperature changes in the Australian region. *Meteorology and Atmospheric Physics*, *89*(1–4), 57–67. https://doi.org/10.1007/s00703-005-0121-3

Lavender, S. L., & Abbs, D. J. (2013). Trends in Australian rainfall: Contribution of tropical cyclones and closed lows. *Climate Dynamics*, *40*(1–2), 317–326. https://doi.org/10.1007/s00382-012-1566-y

Leonard, M., & Westra, S. (2020). *Methodology Report for Multisite Rainfall and Evaporation Data Generation of the Northern Basin*.

Di Luca, A., Evans, J. P., Pepler, A., Alexander, L., & Argüeso, D. (2015). Resolution sensitivity of cyclone climatology over Eastern Australia using six reanalysis products. *Journal of Climate*, *28*(24), 9530–9549. https://doi.org/10.1175/JCLI-D-14-00645.1

Di Luca, A., Evans, J. P., Pepler, A., Alexander, L., & Argüeso, D. (2016). Evaluating the representation of Australian East Coast Lows in a regional climate model ensemble. *Journal of Southern Hemisphere Earth Systems Science*, *66*, 108–124. https://doi.org/10.22499/3.6602.003

Marohasy, J. J., & Abbot, J. W. (2016). Southeast Australian Maximum Temperature Trends, 1887–2013. In *Evidence-Based Climate Science* (pp. 83–99). Elsevier. https://doi.org/10.1016/B978-0-12-804588-6.00005-7

Nguyen, H., Hendon, H. H., Lim, E. P., Boschat, G., Maloney, E., & Timbal, B. (2018). Variability of the extent of the Hadley circulation in the southern hemisphere: a regional perspective. *Climate Dynamics*, *50*(1–2), 129–142. https://doi.org/10.1007/s00382-017-3592-2

Nicholls, N. (2010). Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958-2007. *Climate Dynamics*, *34*(6), 835–845. https://doi.org/10.1007/s00382-009-0527-6

Pepler, A. S., Di Luca, A., Ji, F., Alexander, L. V., Evans, J. P., & Sherwood, S. C. (2015). Impact of identification method on the inferred characteristics and variability of Australian east coast lows. *Monthly Weather Review*, *143*(3), 864–877. https://doi.org/10.1175/MWR-D-14-00188.1

Pepler, A. S., Di Luca, A., Ji, F., Alexander, L. V., Evans, J. P., & Sherwood, S. C. (2016). Projected changes in east Australian midlatitude cyclones during the 21st century. *Geophysical Research Letters*, *43*(1), 334–340. https://doi.org/10.1002/2015GL067267

Potter, N., Chiew, F., Zheng, F., Ekström, M., & Zhang, L. (2016). *Hydroclimate projections for Victoria at 2040 and 2065*. Retrieved from https://publications.csiro.au/rpr/pub?pid=csiro:EP161427

Risbey, J. S., McIntosh, P. C., & Pook, M. J. (2013). Synoptic components of rainfall variability and trends in southeast Australia. *International Journal of Climatology*, *33*(11), 2459–2472. https://doi.org/10.1002/joc.3597

Roderick, M. L., & Farquhar, G. D. (2004). Changes in Australian pan evaporation from 1970 to 2002. *International Journal of Climatology*, *24*(9), 1077–1090. https://doi.org/10.1002/joc.1061

Squintu, A. A., van der Schrier, G., Brugnara, Y., & Klein Tank, A. (2019). Homogenization of daily temperature series in the European Climate Assessment & Dataset. *International Journal of Climatology*, *39*(3), 1243–1261. https://doi.org/10.1002/joc.5874

Stephens, C. M., McVicar, T. R., Johnson, F. M., & Marshall, L. A. (2018). Revisiting Pan Evaporation Trends in Australia a Decade on. *Geophysical Research Letters*, *45*(20), 11,164-11,172. https://doi.org/10.1029/2018GL079332

Taschetto, A. S., & England, M. H. (2009). An analysis of late twentieth century trends in Australian rainfall. *International Journal of Climatology*, *29*(6), 791–807. https://doi.org/10.1002/joc.1736

Theobald, A., McGowan, H., & Speirs, J. (2016). Trends in synoptic circulation and precipitation in the Snowy Mountains region, Australia, in the period 1958–2012. *Atmospheric Research*, *169*, 434–448. https://doi.org/10.1016/j.atmosres.2015.05.007

Timbal, B., & Hendon, H. (2011). The role of tropical modes of variability in recent rainfall deficits across the Murray-Darling Basin. *Water Resources Research*, *47*(12). https://doi.org/10.1029/2010WR009834

Timbal, B., Abbs, D., Jonas, B., Chiew, F., Church, J., Ekström, M., et al. (2015). *Murray Basin Cluster Report, Climate Change in Australia Projections for Australia's Natural Resource Management Regions: Cluster Reports, eds. Ekström, M. et al.* Australia.

Trewin, B. (2013). A daily homogenized temperature data set for Australia. *International Journal of Climatology*, *33*(6), 1510–1529. https://doi.org/10.1002/joc.3530

Trewin, B., Braganza, K., Fawcett, R., Grainger, S., Jovanovic, B., Jones, D., et al. (2020). An updated long-term homogenized daily temperature data set for Australia. *Geoscience Data Journal*, gdj3.95. https://doi.org/10.1002/gdj3.95

Ukkola, A. M., Roderick, M. L., Barker, A., & Pitman, A. J. (2019). Environmental Research Letters Exploring the stationarity of Australian temperature, precipitation and pan evaporation records over the last century Recent citations Exploring the stationarity of Australian temperature, precipitation and pan evaporation r. *Environ. Res. Lett*, *14*, 124035. https://doi.org/10.1088/1748-9326/ab545c

Vincent, L. A., Hartwell, M. M., & Wang, X. L. (2020). A Third Generation of Homogenized Temperature for Trend Analysis and Monitoring Changes in Canada's Climate. *Atmosphere-Ocean*, 1–19. https://doi.org/10.1080/07055900.2020.1765728

van Wijngaarden, W. A., & Mouraviev, A. (2016). Seasonal and Annual Trends in Australian Minimum/Maximum Daily Temperatures. *The Open Atmospheric Science Journal*, *10*(1), 39–55. https://doi.org/10.2174/1874282301610010039

WMO. (1989). *Calculation of monthly and annual 30-year standard normals* (WMO/TD). Geneva: World Meteorological Organization.

9 Appendix A: Supplementary Material

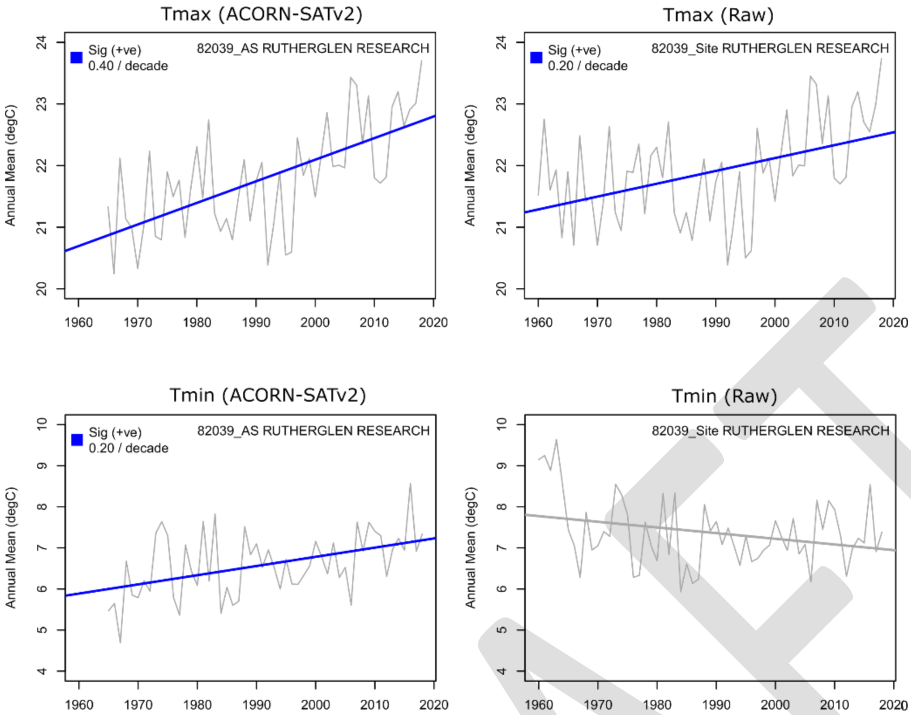

Figure A1 Annual mean short-term (post 1960) Tmax and Tmin at site 82039 (Rutherglen Research) from raw station data and homogenized ACORN SATv2 data with linear trend lines. Blue trend lines indicate the presence of significant trends (at 5% level) using Mann-Kendall trend test. The data from 1960 to 1964 is missing in the ACORN-SATv2 data.

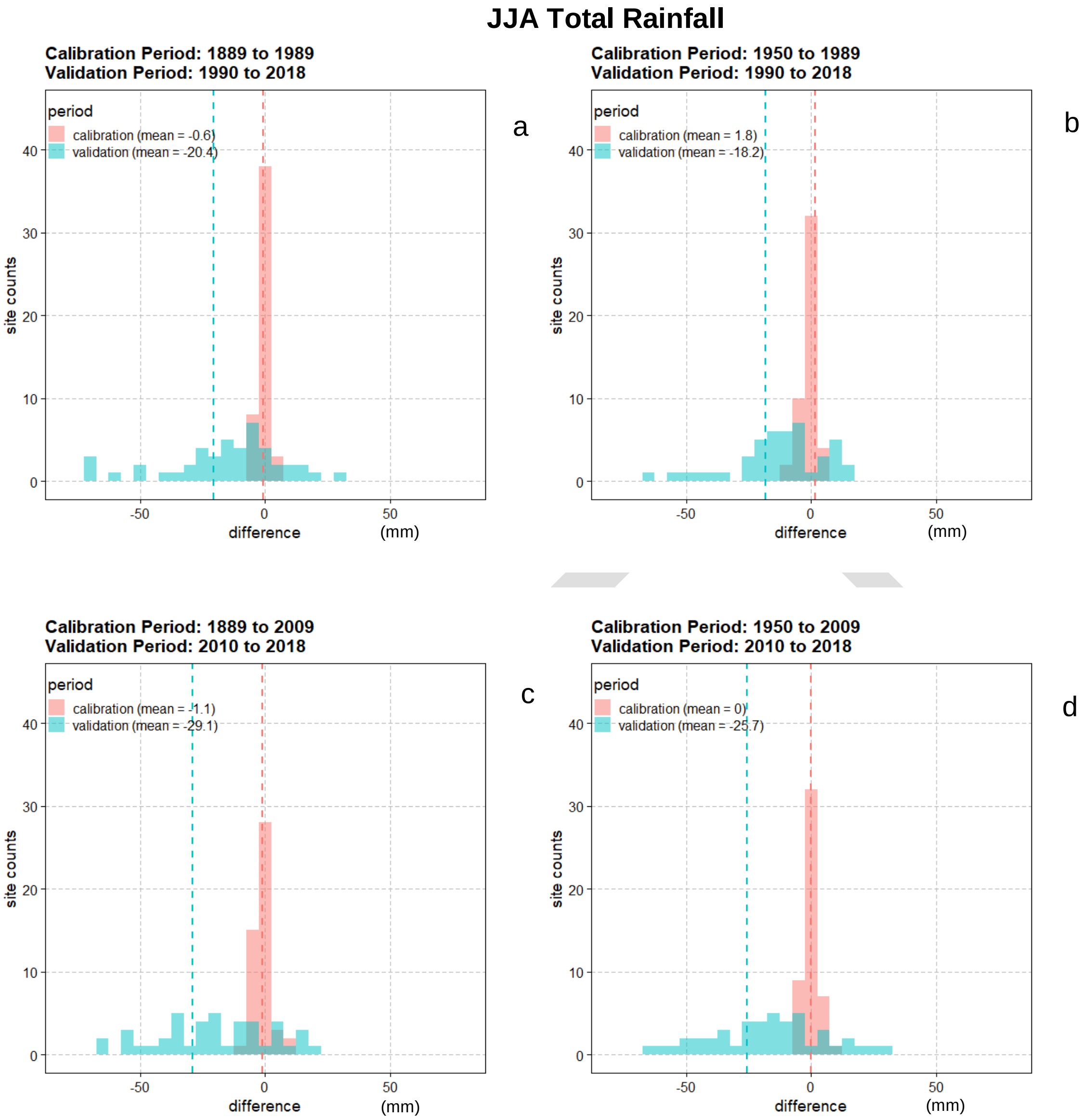

Figure A2. Histograms of the mean differences in total rainfall in JJA during the calibration and validation time periods from stochastic simulations (in mm) at 49 pilot sites (a) 1990-reference split sample test using a long calibration period, (b) 1990-split sample test using a short calibration period, (c) drought-reference split sample test using a long calibration period, and (d) drought-reference split sample test using a short calibration period. The dashed vertical lines mark the means of the respective histograms.

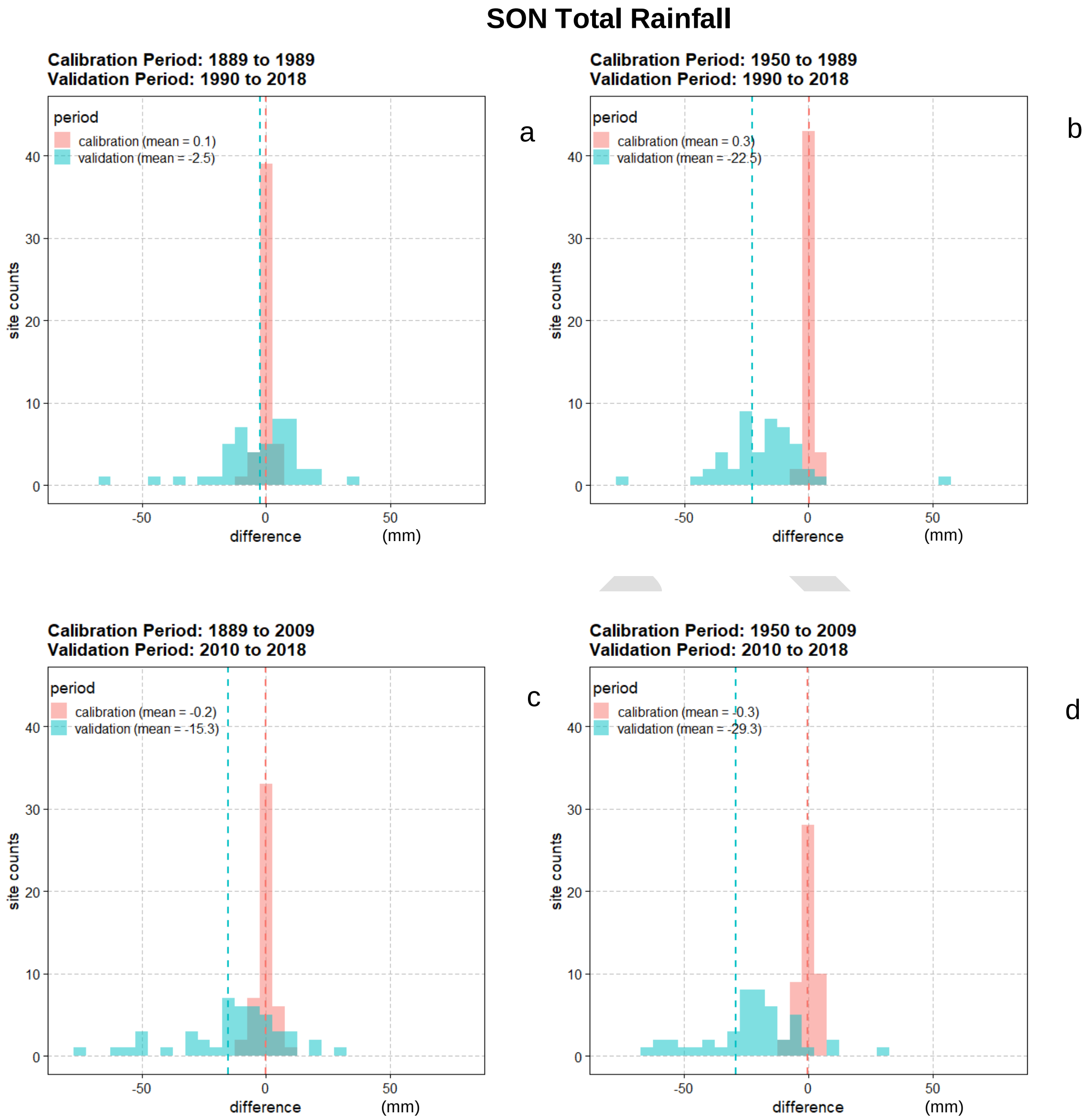

Figure A3. Similar to Figure A2, but for total rainfall during SON (in mm).

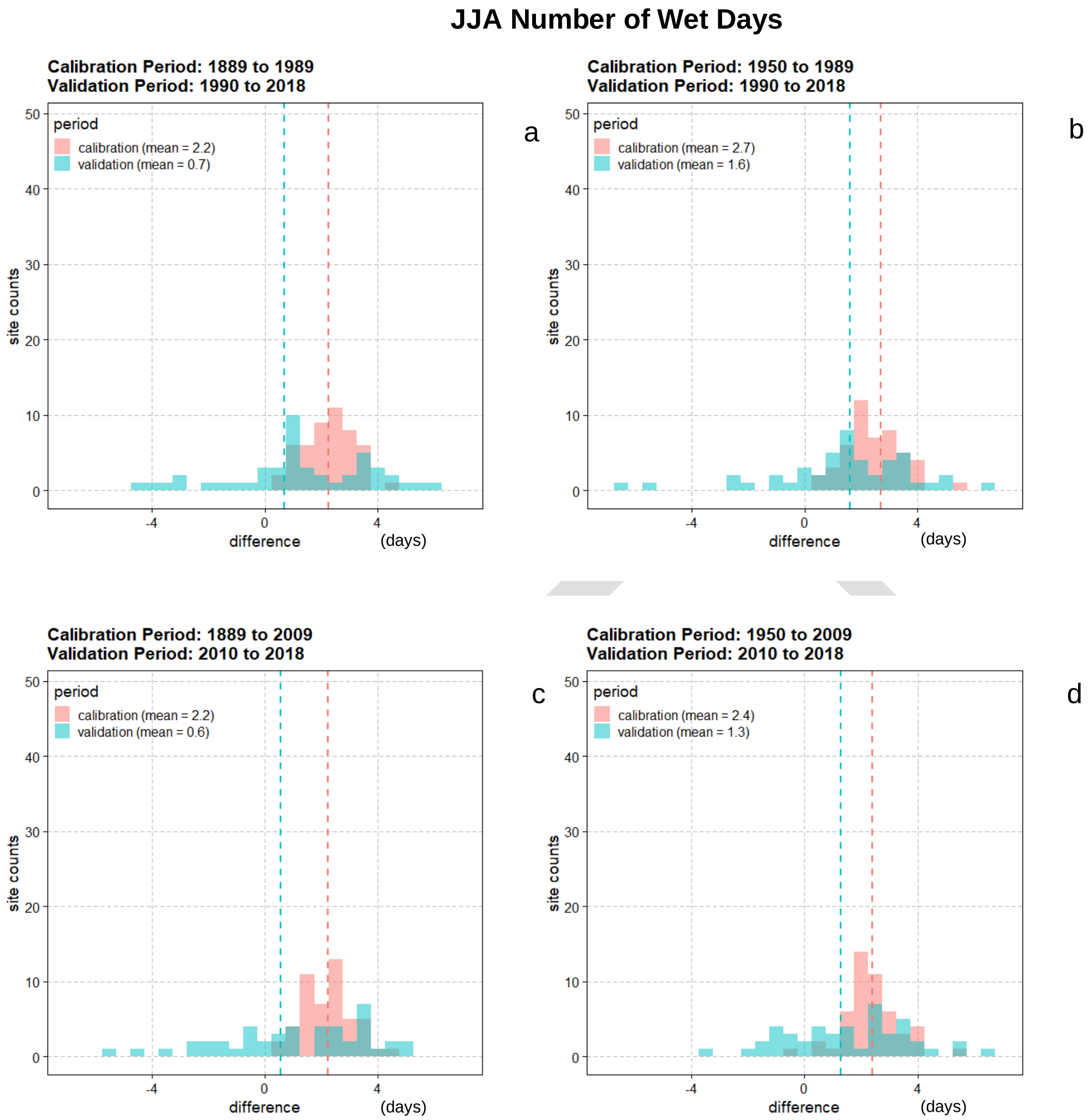

Figure A5. Histograms of the mean differences in number of wet days in JJA during the calibration and validation time periods from stochastic simulations (in days) at 49 pilot sites (a) 1990-reference split sample test using a long calibration period, (b) 1990-split sample test using a short calibration period, (c) drought-reference split sample test using a long calibration period, and (d) drought-reference split sample test using a short calibration period. The dashed vertical lines mark the means of the respective histograms.

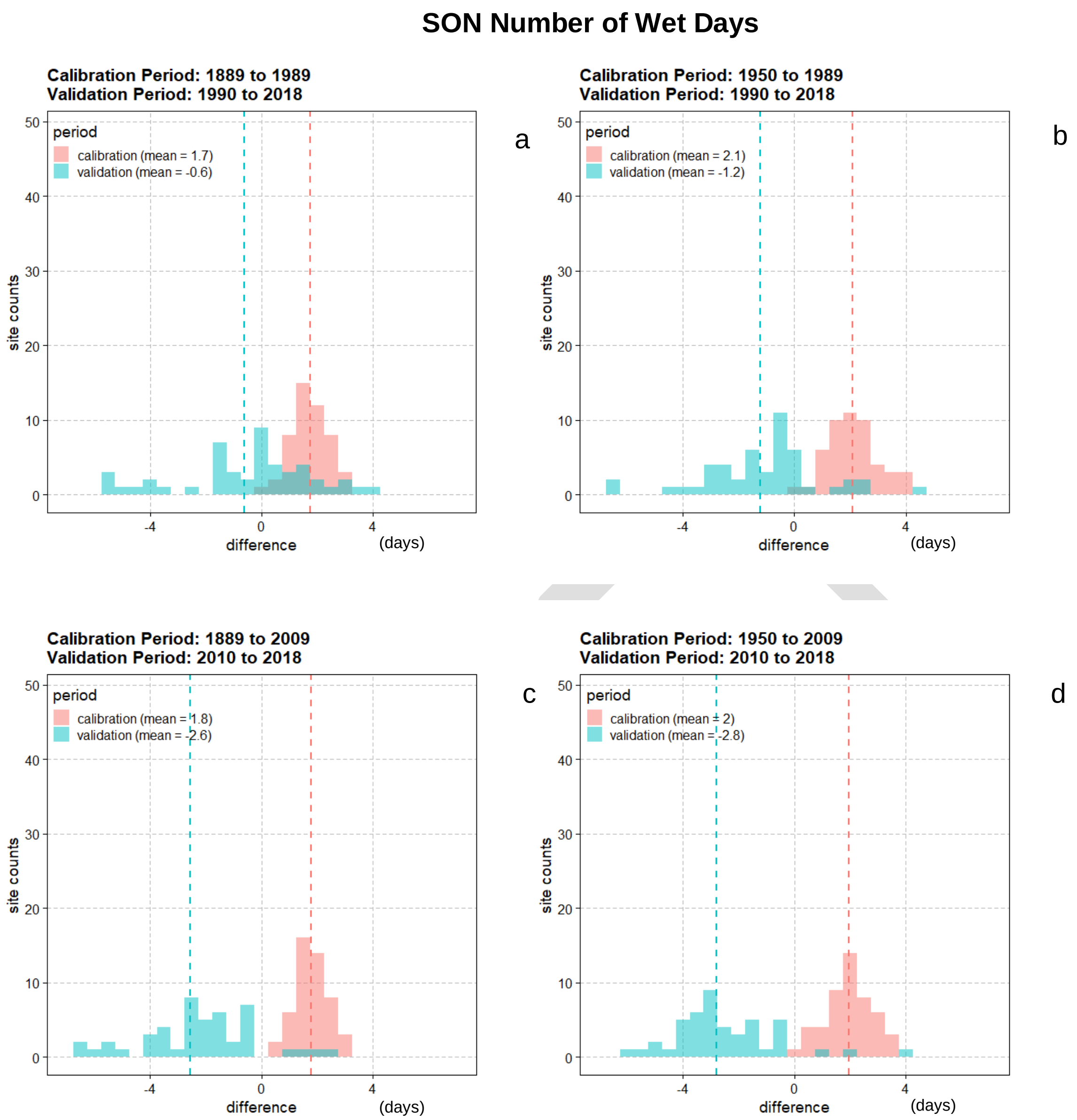

Figure A6. Similar to Figure A5, but for number of wet days during SON (in days).

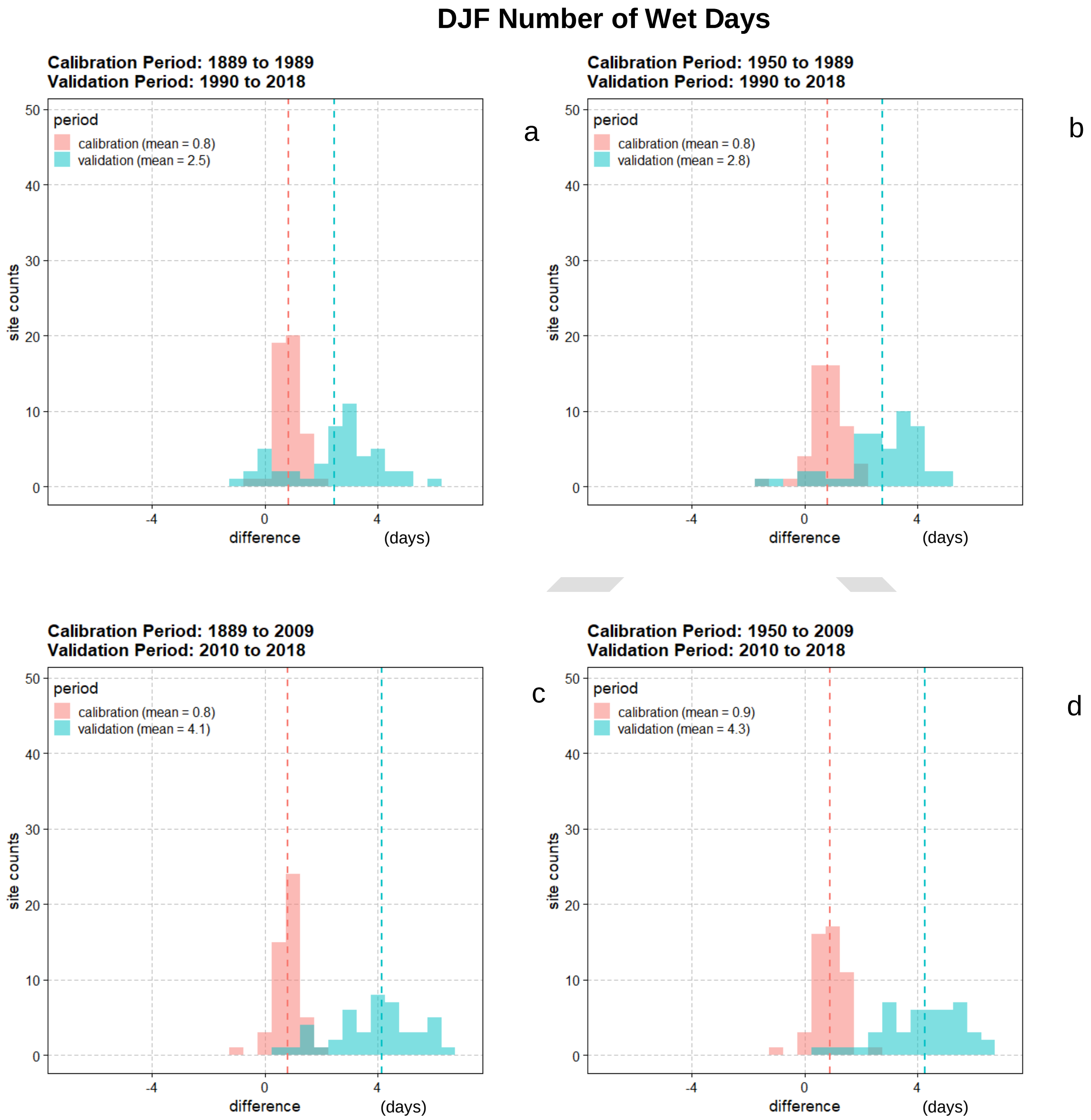

Figure A7. Similar to Figure A5, but for number of wet days during DJF (in days).

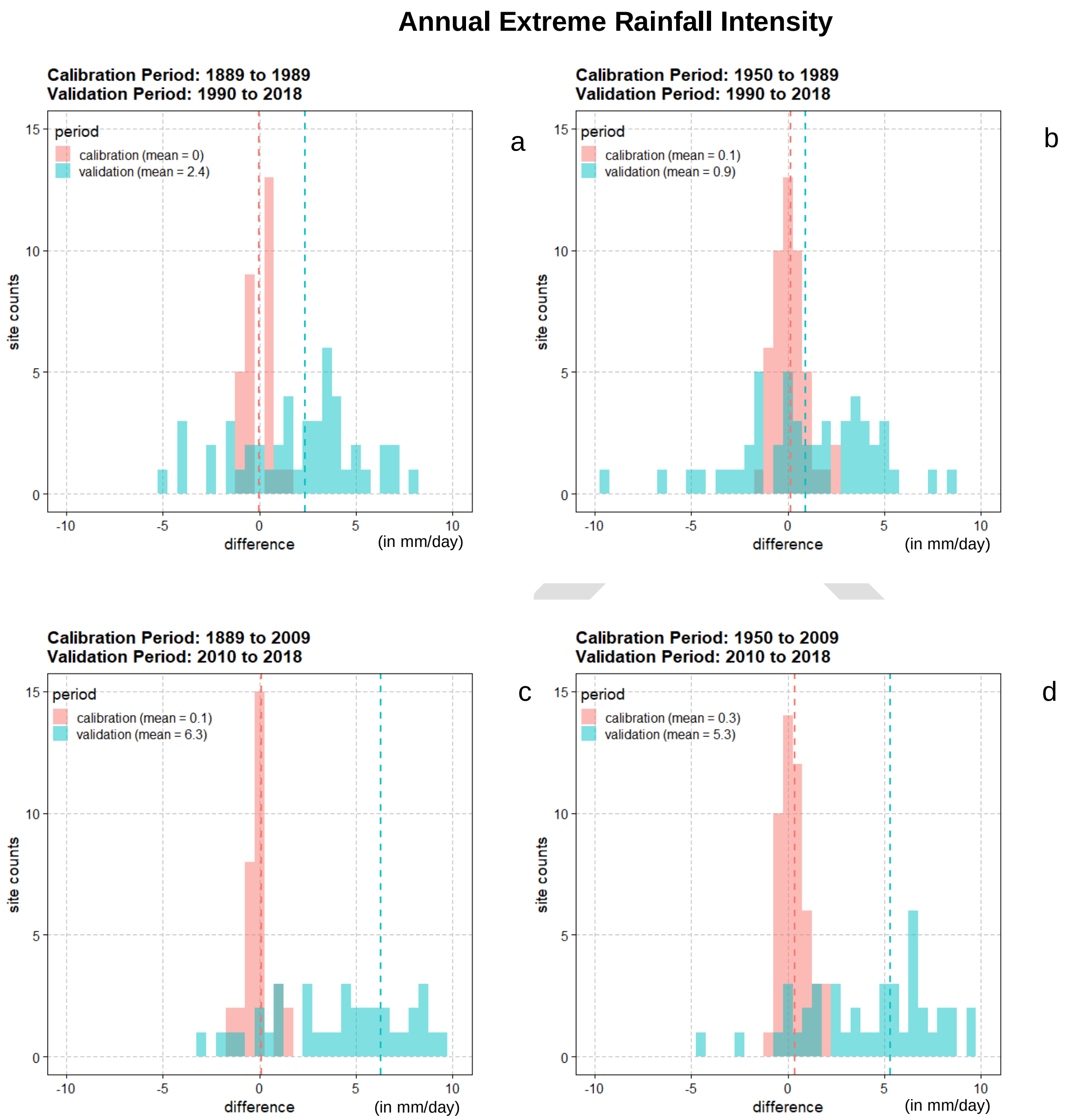

Figure A8. Histograms of the mean differences in annual extreme rainfall intensity during the calibration and validation time periods from stochastic simulations (in mm/day) at 49 pilot sites (a) 1990-reference split sample test using a long calibration period, (b) 1990-split sample test using a short calibration period, (c) drought-reference split sample test using a long calibration period, and (d) drought-reference split sample test using a short calibration period. The dashed vertical lines mark the means of the respective histograms.

APPENDIX 2: PROPOSED UOA METHODOLOGY

Water Research Centre

University of Adelaide, SA, 5005

Mr Jon Midwinter

NSW Department of Planning, Industry & Environment

28 August 2020

Dear Jon Midwinter,

Proposal – Provision of stochastic climatic data for the Southern Basin representing historical, current and future climate conditions

Thank you for the opportunity to submit a quote to provide stochastic data for the Southern basin, whichwe understand includes the New South Wales/Victorian Murray, and Murrumbidgee basins.

This document describes the proposed method for generation of the stochastic data considering the full historical record as well as several alternatives to account for climate change, reflecting Options 1-3 as described in the *"Implications of Non-Stationarity for Stochastic Time Series Generation in the Southern Basins"* report. The proposal has been broken into a set of distinct stages to provide flexibility and enablecost-effective testing of alternative approaches as part of overall program design.

The team that has been assembled to complete this work will comprise Dr Michael Leonard, Dr Anjana Devanand and myself, who each have extensive experience in stochastic hydrology, climate data analyses,and hydrological modelling. We believe that the proposed methodology represents the latest research in stochastic generation for hydrological modelling, and meets the requirements for simulating rainfall, evapotranspiration and temperature in the Southern basin catchments.

Yours sincerely,

Prof Seth Westra

Phone: (08) 8313 1538

Mobile: 0414 997 406

1 Background and Project Appreciation

The Department of Planning, Infrastructure, Environment (DPIE) has undertaken a risk-based methodology to account for climate variability and change in developing the Regional Water Strategies (RWS). A core component of the methodology involves the use of stochastically generated long term sequences of climate data to characterize the historical climate, and application of scaling factors to the stochastic data to generate future climate projections. The stochastic modelling is based on historical (observed/reconstructed) recordsof daily rainfall, evapotranspiration and temperature to generate synthetic data for 10,000 years that reflect variability over the instrumental record.

This document outlines a proposal for generation of 10,000 year time series of rainfall, evaporation and temperature data at multiple sites in the Southern basin utilising the methodology developed in previous work on the northern basins; namely Macquarie, Border rivers, Gwydir, Namoi, Far North Coast, North Coast, and the Western catchments [1,2,3]. An expert panel independent review of the DPIE climate risk methodology recommended careful consideration of the implications of climatic non-stationarity on stochastic time series generation in the Southern basin. Subsequently, a pilot assessment was undertaken to analyse data from pilot sites in the Southern basin, and collate available information from a review of literature and future projections for the region [4]. Multiple lines of evidence indicate that the historical rainfall record in the Southern basin is non-stationary, and the observed changes are at least partly attributable to climate warming [5,6]. The non-stationarity in the historical record has potential implications on stochastic time series generation in the region, as outlined in the pilot study report [4]. The methodology for stochastic time series generation proposed in this document accounts for climate change in the historical record.

The project is structured in three stages:

Stage 1 – Southern basin (~450 time series) – stochastic simulations based on the full record (1890 to 2018) under the stationarity assumptions is to be completed as a first priority given DPIE's timeline and operational needs. This adopts consistent methodology with earlier stochastic generation work and will preserve correlations with previously generated stochastic sequences. An optionaltask of modifying the tail distribution to incorporate potential 'outlier' events such as occurred in 2019 for the Border region is also included as part of this Stage.

Stage 2 – Assessment of future changes from NARCliM 1.5 data including provision of confidence intervalsfor changes to a range of key climate attributes, and qualitative comparison with projected changes from other sources of information on climate change projections as articulated in [4]. This will include the development of an agreed set of scaling factors based on best-available understanding of likely change in the Southern Basin.

Stage 3 – Accounting for climate change in the stochastic sequences - stochastic simulations based on historical baseline consistent with either NARCliM 1.5 or adjusted scaling factors from Stage 2 will be developed using two alternate approaches (options #2 and #3 in pilot report). The implications of alternative approaches on the statistics of the stochastic sequences for rainfall,evapotranspiration and temperature will be assessed, and an optional task of also assessing implications directly on runoff is included as part of this Stage.

The generated timeseries will characterise key statistical properties of rainfall, evaporation and temperature as necessary for hydrological response and water planning. The assessment will include inter-annual and multi-decadal rainfall variability including the assessment of extreme droughts.

The following section reviews the existing model based on the northern basin project, and outlines the methodology for stochastic generation to account for climate change in the Southern Basin. It is expected that the simulated historical data will share 'joint' characteristics despite being simulated in separate stages. The joint nature of the simulation means that similar climatic conditions co-occur across all catchments (e.g. drought conditions). The joint simulation can be achieved via an approach of conditioning across all regions and timescales to ensure that subsequent stages relate to data simulated in earlier stages. The methodology to account for climate change includes an assessment of the NARCliM 1.5 data, stochastic simulations corresponding to the NARCliM 1.5 historical baseline and application of seasonal or quantile based scaling factors to generate current and future stochastic simulations.

2 Proposed Method and Outputs

2.1 Stage 1: Stochastic Simulations using the full historical record

2.1.1. Stochastic model development and data generation for the Southern basin

The generation of stochastic data for the southern basin will use the multi-site generator employed in the northern basin catchments. This method has already been implemented for the other regions in New South Wales [1,2] to generate 10,000 year timeseries for a combined total of 754 rainfall and 569 evapotranspiration sites in Macquarie, Border Rivers, Namoi, Gwydir, North Coast, Far North Coast, and Western Regions of New South Wales.

The model uses a hierarchical structure to ensure consistency across the region at monthly to inter-annual timescales. The Macquarie region was simulated at the daily timescale with the only hierarchy being the climate state model for inter-annual variability. The hierarchical structure in the model developed for subsequent regions have separate models for the interannual (IPO climate model as per [1]), annual (autoregressive lag-1) and monthly (autoregressive lag-1) scales [3]. The model employs a scaling of the neighbourhood size using a Gaussian Markov Random Field while retaining the correlation structure of the entire region. The model has been developed to maintain consistency with the simulated timeseries from existing sites in the previously simulated regions.

A multi-site generator based on the model applied for the northern basins will be developed for the Southern basin under stationary assumptions using the full historical record from 1890 to 2018. The simulated data will be evaluated in detail relative to historical rainfall and evaporation over the period from 1890 to 2018, using the established evaluation methodology followed for the northern basins. The multi-site and multi- variable simulation requires extensive evaluation because it is important to reproduce variability across a wide range of scales (daily, monthly, seasonal, annual, decadal), and a wide range of statistics (amounts, dry- occurrences, extremes, dry and wet spells, excursions above/below a threshold, correlations, etc). It is important to carefully evaluate the model tradeoffs across the multiple sites, variables, statistics and scales.Detailed reporting and systematic evaluation of sites has been demonstrated based on simulated data for the northern region [1] and will follow a similar style for this project.

The method uses a systematic evaluation to classify each statistic as 'Good', 'Fair' or 'Poor' according to a defined set of tests at all sites including:

- distribution of annual rainfall/evaporation totals (1, 2, 5, 10 years)
- mean and standard deviation of monthly rainfall/evaporation totals
- proportion wet days, and
- extreme rainfalls.

Simulations were developed in the northern region for historical climate scenarios for partitions of the IPO estimated from the instrumental record and paleoclimate proxies. In the southern region, the differences of partitioning the observed record based on IPO are anticipated to be low or non-existent given the lower influence of this mode of variability in this region, but will be tested directly on the data and the partitioning will be implemented if required. If indicated by the analysis, the developed model may exclude a partitioning based on the IPO, and this is not anticipated to affect the daily correlations with other regions.

2.1.2. Option 1: Model development to better represent the lower tail of the annual total rainfall

The recent (2017 **–)** drought in the Border rivers region has resulted in record low annual total rainfall at time scales of 18 months to 3 years across the region [7]. A preliminary investigation showed that inclusion of theyear 2019 annual rainfall in the calibration period for that region influenced the stochastic simulations of minimum rainfall. Moreover, the stochastic analysis without inclusion of 2019 was not able to simulate annual totals as low as 2019.

To ensure that future stochastic sequences yield credible low annual total years, the lower tail of the annual rainfall probability distribution used for stochastic simulation may be modified to achieve higher probabilityof low rainfall years to align with recent observations. This model development would examine the use of alternate probability distributions with heavier tails to represent the annual totals, and implement an appropriate distribution to represent the low rainfall years more accurately in the existing hierarchical stochastic generator.

The developed model will be applied for simulation in the southern basin. As part of this development, it is proposed that the existing stochastic data generated for the northern regions be post-processed such that the annual totals from the simulations exhibit higher probability in the lower tails, consistent with the updated probability distribution.

2.1.3. Outputs

- The 10,000 year timeseries for all requested sites in suitable electronic format based on the full historical climate
- Comprehensive evaluation of at-site statistics for each site, ranging from daily through to inter-annual timescale, including a wide range of statistics and multiple variables
- Summary assessment of key statistics and regional statistics across all sites
- Option 1 also includes 10,000 year timeseries for sites in the northern basin generated throughpost-processing of the existing data to match the implemented alternative tail probability distribution

2.2 Stage 2: Assessment of current and future climate conditions using NARCliM1.5

An assessment of NARCliM 1.5 data will be undertaken to derive scaling factors for future projections. Bias correction of the NARCliM data would not be necessary for this assessment since the focus is on obtaining ‘change factors’ representing future changes relative to the NARCliM historical simulations, which are then applied to the stochastic sequences. However, significant biases if identified will be described as part of the reporting, as large discrepancies might indicate broader issues with the projections. This analysis would involve the assessment of future changes projected for the Southern basin for a range of hydrologically relevant attributes of rainfall, evaporation, and temperature including:

- Annual and seasonal rainfall totals

- Annual and seasonal evaporation totals
- Annual and seasonal mean temperatures
- Standard deviation of annual and seasonal rainfall totals
- Number of wet days and heavy rainfall days
- Extreme rainfall intensity and frequency

As part of this analysis, future changes for multiple future time slices using the NARCliM 1.5 data relative tothe historical climate baseline in all the attributes for the southern basin will be documented, as well as the uncertainty ranges of the projections. The analysis may include 2 or 3 time windows (minimum 30 years long) potentially including a ‘current’ climate window (i.e. a window centred on 2020) and one or two future windows, with the specific windows to be decided based on discussion with DPIE. The spatial homogeneity of the projected changes over the Southern basin will be examined to assess whether spatially uniform scaling factors may be used to generate stochastic future projections.

The projected changes from NARCliM 1.5 will be compared with future projections from other published lines of evidence [eg: 8], as has been documented in the pilot study report [4]. Depending on the consistency between the NARCliM 1.5 findings and the other lines of evidence, adjustments in the scaling factors may be indicated to reflect best available understanding of likely climatic changes. Overall, the workundertaken in Stage 2 would identify the scaling factors and their uncertainties for future changes in rainfallattributes in the southern basin.

2.1. Outputs

Report detailing (1) the fidelity of NARCliM results in simulating historical climate, (2) future scaling factorsand (3) uncertainty from the NARCliM 1.5 for hydrologically relevant climate attributes, as well as comparison to other sources of projections based on the lines of evidence reviewed in the pilot study report [4].

2.3 Stage 3: Generation of stochastic future projections

Stage 3 involves two key steps: (1) generation of stochastic timeseries corresponding to the NARCliM 1.5 baseline using the model developed for the Southern region in Stage 1, and (2) application of current andfuture scaling factors based on the analyses in Stage 2 to the generated data. An optional step of testing the implications of scaling factors on runoff series in selected pilot sites is also included in the analysis.

2.3.1. Stochastic simulations corresponding to the NARCliM 1.5 baseline

This step involves the generation of stochastic data consistent with the NARCliM 1.5 baseline (years 1950 to 2005). The generation of stochastic data may follow two alternate approaches as outlined in the pilot study report [4] using:

a) observed data corresponding to the NARCliM 1.5 baseline to calibrate the stochastic model (Option 2 presented in the pilot study report [4])
b) observed hybrid baseline which involves de-trending the historical data prior to the NARCliM 1.5 baseline, and use of the extended observed record to calibrate the stochastic model (Option 3 presented in the pilot study report [4])

Both of these approaches will be employed to obtain two alternate sets of ‘baseline’ stochastic simulations corresponding to the NARCliM 1.5 baseline. The use of observed data corresponding to the NARCliM baseline for calibration of the stochastic model is straightforward. De-trending the observed data to generate the extended observed data corresponding to the NARCliM 1.5 baseline is more intricate. It is proposed that the seasonal and annual totals of rainfall will be de-trended and adjusted to reflect the NARCliM baseline to create the extended record, applying a bootstrap resampling method to ensure variance of the data is preserved. The method will not involve de-trending other statistics like the numberof wet days. The stochastic model developed for the Southern basin in Stage 1 would be calibrated to shorter/de-trended observed data to generate two alternate 'baseline' stochastic simulations corresponding to the NARCliM 1.5 baseline.

2.3.2. Scale the 'baseline' stochastic sequences to generate future projections

This step involves scaling the 'baseline' stochastic simulations using the scaling factors obtained from the analysis of NARCliM data and examination of other lines of evidence in Stage 2. The scaling approach would employ two alternative methods: seasonal scaling factors or quantile scaling. It is proposed that multiple approaches will be considered to scale the different attributes of rainfall and generate stochastic current and future projections that reflect the expected changes in attributes, noting that quantile scaling will enable adjustment of the full daily distribution including extremes as well as number of wet days. The scaling method would incorporate future changes in annual and seasonal totals and could include changes in wet days. However, the method would not account for future changes to wet/dry spell lengths other than what would occur through adjustment of wet day frequency.

The uncertainty estimates from NARCliM 1.5 (or adjusted NARCliM 1.5) in Stage 2 will be used in the scaling approaches, to generate current and future stochastic simulations that incorporate a range of conditions.

The scaling method would be applied to both alternatives of the 'baseline' stochastic data generated usingthe approaches outlines in section 2.3.1.

2.3.3. Option 2: Evaluation of alternative options on runoff data

The implications of the alternate approaches of current and future stochastic simulations on the simulated runoff will be assessed through hydrological modelling of pilot catchments in the southern basin. The catchments will be selected based on discussion with DPIE, and it is assumed that model and parameters will be made available by DPIE. Hydrological simulations will be performed using alternate stochastic futureprojections generated as outlined in section 2.2.3. The results of this assessment will help understand the implications of alternative approaches directly on the runoff data.

2.3.4. Outputs

- Comprehensive evaluation of at-site statistics for each site, ranging from daily through to inter-annual timescale, including a wide range of statistics and multiple variables
- Summary assessment of key statistics and regional statistics across all sites
- Alternate sets of 10,000 year timeseries for all requested sites in suitable electronic format generated by application of both scaling methods on both 'baseline' stochastic time series. Time series also will be provided for up to three time horizons (e.g. a current and two future horizons),so the total output consist of up to 12 alternate sets of 10,000 year timeseries.
- Script to implement scaling method in a scripting language suitable for DPIE (e.g. R or Python)
- Option 2: Report on the implications of alternate approaches on runoff
- Recommendation on the most suitable baseline period and scaling method to use for implementation purposes based on the outcomes of testing (ideally including outcomes of Option2)

2.4 Summary of Tasks

The main tasks involved in the methodology are summarised below. The work will be implemented by DrAnjana Devanand with significant input and discussion from Dr Michael Leonard and Prof. Seth Westra.

Task	Description
1.1	Stochastic model development and data generation for the southern basin (Stage 1)
1.2	Model development to better represent the lower tail of the annual total rainfall (Stage 1,Option 1)
2.1	Assessment of current and future climate conditions using NARCliM 1.5 (Stage 2)
3.1	Stochastic simulations corresponding to the NARCliM 1.5 baseline (Stage 3)
3.2	Scale the ‘baseline’ stochastic sequences to generate future projections (Stage 3)
3.3	Evaluation of alternative options on runoff data (Stage 3, Option 2)

2.5 Project Reporting

Four project reports will be developed summarising details of the methodology, assumptions, results, discussion of the outputs and any observed limitations.

Stage 1	Report on ‘historical’ stochastic time series generation corresponding to the full historicalrecord in the southern basin
Stage 2	Report on NARCliM 1.5 projected future changes in the southern basin and comparison with other lines of evidence
Stage 3	Report on ‘baseline’ stochastic time series generation corresponding to the NARCliM 1.5 baseline based on two approaches (post-1950 observed record, de-trended observed record)
Stage 3, Option 2	Report on the implications of alternate approached of baseline data generation on runoff in the southern basin

The reports on the stochastic time series generation will include appendices containing detailed visual summaries across each location will be provided to accompany the main report. The results section will include description of the following statistics:

- Number and distribution of rain days
- Rainfall depth
- Intensity, frequency and duration of extreme events
- Severity and duration of below average rainfall
- Variability over multiple timescales from inter-annual to multi-decadal
- Multi-site rainfall dependency
- Multi-variate (rainfall verse evaporation) relationship
- Evaporation totals
- Mean temperatures
- The report on the NARCliM 1.5 projected changes will include documentation of the changes in all attributes listed in section 2.2.

The reports will be written by Dr Anjana Devanand and internally reviewed by all members of the project team.

2.6 Project Management

Note: Section 2.6 has been redacted.

3 Project Timeline and Price Estimate

Note: Section 3 has been redacted

4 References

Leonard M., Westra S, Thyer M, Bennett B (2019) Multisite Rainfall and Evaporation Data Generation for the Macquarie Water Infrastructure Project, Report 2019/01

Leonard M., Westra (2019) Multisite Rainfall and Evaporation Data Generation for the Macquarie Water Infrastructure Project,

Addendum – Updated Simulation with Additional Sites, Report 2019/02

Leonard, M., Westra, S., Bennett, B (2020). Methodology Report for Multisite Rainfall and Evaporation Data Generation of theNorthern Basin, Report 2020/01

Devanand, A, Leonard, M, Westra, S (2020). Implications of Non-Stationarity for Stochastic Time Series Generation in the Southern Basins, Report 2020/07

CSIRO. (2012). Climate and water availability in south-eastern Australia A synthesis of findings from Phase 2 of the South Eastern Australian Climate Initiative (SEACI) (Vol. 41). Australia. Retrieved from http://www.seaci.org

Hope, P, Timbal, B, Hendon, H, Ekström, M, Potter, N. (2017). A synthesis of findings from the Victorian Climate Initiative (VicCI).

Retrieved from http://www.bom.gov.au/research/projects/vicci/

BoM. (2019). Special Climate Statement 70 update—drought conditions in Australia and impact on water resources in the Murray–Darling Basin.

Clarke, J., Grose, M., Thatcher, M., Hernaman, V., Heady, C., Round, V., et al. (2019). Victorian Climate Projections 2019 Technical Report. Melbourne, Australia.

APPENDIX 3: PROPOSAL UON METHODOLOGY

Proposal prepared by University of Newcastle 27th November 2020 to *Provide stochastic climatic data for the Greater Sydney Region representing historical, current and future climate conditions* for DPIE Water.

Task 1: Inception meeting.

The proposed method will be presented at a meeting of the Independent Expert Panel chaired by the Deputy Chief Science and Engineer on 4th December 2020. This meeting will also review similar work being undertaken by University of Adelaide on addressing non-stationarity in the southern Murray Darling Basin.

Comments and feedback from this meeting of the Independent Expert Panel will be implemented, and the proposed method revised if required, to ensure consistency and identify any potential synergies with the University of Adelaide approach and analysis.

Task 2: Identify weather systems.

- Key weather systems responsible for rainfall characteristics that are important for determining storage inflows over a range of time scales (e.g. short-term episodic dam filling events, multi-decadal dry periods etc.) will be identified.
- This analysis will (i) provide context for identification of suitable palaeoclimate data sets (used in Task 3) and (ii) help understand the possible nature of future climate change (used in Task 6).
- Key weather systems will be identified for the Greater Sydney region using Self-Organising Map (SOM) Analysis. SOMs have been successfully used to analyse synoptic weather across multiple regions in Australia and internationally. The SOM algorithm is preferred over traditional discrete clustering methods (e.g. k-means) as it accounts for continuity and nonlinearity, providing a more realistic representation of the continuous movement between weather patterns rather than assigning to discrete clusters that may misrepresent extremes. This makes the SOM algorithm preferable when consideration of both extremes and average climatology is important.
- Rainfall distributions associated with each of the key weather systems will be determined as well as the frequency of occurrence for each key weather system.
- Relationships between the identified key weather systems and large-scale ocean-atmospheric processes (e.g. El Niño/Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), Interdecadal Pacific Oscillation (IPO)) will also be investigated.

Task 3: Collate data sets.

- DPIE will identify and supply the historical rainfall/evaporation data sets that will be used in this project. It is envisaged that this will include up to 200 paired data sets of historical rainfall and potential evapotranspiration at a daily time step for 130 years of record.
- Appropriate palaeoclimate data sets will also be obtained to provide insights into pre-instrumental hydroclimatic variability. This will include the ~450 year IPO reconstruction used elsewhere in NSW Regional Water Strategies and the ~2,700 year data set of dry/non-dry periods embodied in the Wombeyan Caves stalagmite geochemistry.
- Check all daily rainfall/evaporation data supplied by DPIE.
- Use the rainfall/evaporation data supplied by DPIE to calculate the annual and daily statistics that need to be realistically simulated by the stochastic modelling. This includes:

- Annual statistics:
    - Mean, maximum, minimum (rainfall and evaporation);
    - Standard deviation (rainfall and evaporation);
    - Lag-one auto-correlation (rainfall);
    - 2-year low rainfall sums;
    - 5-year low rainfall sums;
    - 10-year low rainfall sums;
    - Spatial cross-correlation between rainfall and evaporation sites;
    - Spatial cross-correlation between rainfall and evaporation sites.
- Daily statistics:
    - Number and distribution of rain days;
    - Rainfall depth;
    - Intensity, frequency and duration of extreme events;
    - Severity and duration of below average rainfall;
    - Spatial cross-correlation between Tuross rainfall and evaporation sites;
    - Spatial cross-correlation between Tuross and Bega rainfall and evaporation sites.
    - Prepare inputs for the non-stationarity testing (Task 4) and the stochastic modelling for current (Task 5) and future (Task 6) conditions.

Task 4: Test for non-stationarity.

- The unprecedented low inflows over the three years preceding the February 2020 dam filling event suggest the regional climate in the Greater Sydney region may have already changed. Tests will be conducted on the data provided by DPIE (in Task 3) to determine if there is evidence for non-stationarity in the 130 year instrumental rainfall/evaporation records.
- The approach employed for testing for non-stationarity will be as per what is recommended by the Independent Expert Panel to ensure consistency with the approach previously employed by University of Adelaide when addressing non-stationarity in the southern Murray Darling Basin.
- The exact approach that will be used to test for non-stationarity will be confirmed at the Inception meeting on 4th December 2020 (Task 1).

Task 5: Develop and implement stochastic models – stochastically generate daily rainfall and evaporation data for Sydney catchments and Greater Sydney region.

- The daily hydroclimatic data will be generated for the required sites using a multi-site stochastic generation based on a lag one autoregressive (AR1) Matalas (1967) model. The multi-site Matalas model was originally developed for application at the annual time scale but this has been adapted in previous work (e.g. McMahon et al., 2008; Mortazavi-Naeini et al., 2015) for application at the daily time scale. The modified multi-site Matalas model uses an AR1 model that preserves the lag-zero and lag-one cross correlations between seasons and locations for all variables (i.e. rainfall and evaporation).
- The modified multi-site Matalas model will be calibrated (or trained) using the DPIE provided ~130 years of historical (i.e. instrumental) daily rainfall and evaporation.
- The stochastically generated outputs will then be compared with the historical data to determine if the key statistics observed in the instrumental records (as calculated in Task 3) are accurately reproduced in the stochastically modelled outputs and to gain some insights into the impacts of climate variability in the Greater Sydney region.
- The outputs will be formatted according to DPIE requirements for inputs into the hydrological modelling that DPIE needs to conducted as part of the Greater Sydney Water Strategy.

Task 6: Perturb data sets for climate change.

The following approaches will be used to perturb data sets to account for the projected impacts of climate change:

- As per the approach used for the NSW South Coast (Bega and Tuross) which involved altering the seasonal frequency of East Coast Lows (based on plausible projections from NARCliM1.0 and information obtained in recent literature) which in turn changes the distribution of daily rainfall/evaporation. These ‘climate change impacted’ rainfall/evaporation sequences are then used as input to the stochastic models developed in Task 5 to produce plausible scenarios for how rainfall/evaporation could change in the future which are then used to stress-test options for ensuring/improving water security in the Greater Sydney region. For this Greater Sydney region study, the NSW South Coast (Bega and Tuross) method will be updated by using insights from Task 2 and the newly (or soon to be) available NARCliM1.5 data sets to get improved information about how East Coast Lows (and other key weather systems) are projected to change.
- Follow the approach proposed by University of Adelaide for the southern Murray Darling Basin. This approach is designed to make use of the newly (or soon to be) available NARCliM1.5 data sets, on the premise that it is the best available climate model information for NSW. The exact details of this approach are to be confirmed at the Inception meeting on 4th December 2020 (Task 1).

Task 7: Implement quality assurance on outcomes.

- Ensure the stochastically generated data for historical conditions (Task 5) satisfactorily reproduces the key statistics observed in the instrumental records (as calculated in Task 3).
- Work with DPIE modelling staff to ensure that the stochastically generated data produces expected hydrological outcomes.
- Provide all of data from Task 5 and Task 6 to DPIE for hydrological and water resources modelling and work with DPIE to ensure satisfactory reproduction of key statistics (e.g. number of rain days, rain event frequency, means and standard deviations for multiyear periods).

Task 8: Reporting.

- Attend virtual meetings as required.
- Prepare and provide a draft report which details the methodology, assumptions, results and a discussion on the limitations of the approaches taken to generate the data. Information in the results section will include plots, tables and discussion that:
  (a) Demonstrates that the performance of the stochastic model is satisfactory (i.e. the key statistics observed in the instrumental records (as calculated in Task 3) are accurately reproduced in the stochastically modelled outputs); and
  (b) Illustrates the range of rainfall and evaporation conditions that are possible under the plausible climate change scenarios investigated.
- Obtain feedback on draft report from DPIE.
- Finalise report submit it to DPIE.

APPENDIX 4: ROLLING ORIGIN APPROACH TO MODEL EVALUATION

Rolling Origin Approach to Model Evaluation – Testing the Temporal Stability of Model Parameter Estimates and Simulations

Assessments of model performance are typically based on out-of-sample measures. For example, the available sample can be split into two parts. The model is then fitted to the training set and its skill evaluated using one or more error measures on independent test data. The overall aim is to minimize the combined bias and variance by sacrificing some bias in exchange for reducing sampling variance. If model performance is noticeably better during training, overfitting might be an issue (Shumeli 2010).

Evaluating model performance based on a single split may provide a biased estimate of model performance due to the presence of outliers and changes in the mean level of the response variable (e.g. rainfall or streamflow). Application of more sophisticated conventional techniques (such as k-fold cross-validation) to time series can be problematic if there is an implicit temporal dependence in such data (which might include seasonality, low frequency variability and a chronic trend), and the series is nonstationary (due, for example, to temporal variations in the mean and variance).

Rolling origin evaluation is a resampling method in which the origin of the test set is updated successively, and model simulations are produced from each origin (Tashman 2000; Petropoulos et al. 2017). There are different options of how this can be done. One approach is illustrated in the schematic diagram below. For convenience assume that the length of the series of interest $y_t, t = 1, 2, \ldots, N$, is exactly divisible by a user-specified integer h. Let $M = N/2$ denote the number of observations in the training set for each iteration. During the first iteration, the model is calibrated to the first M observations. It is then used to produce an ensemble of simulations for which the mean is calculated for $t = M+1, M+2, \ldots, N$. The means are stored for later retrieval. During the second iteration the first h observations are discarded, and the training set is moved forward to maintain the training set size (M). The model parameter estimates are updated using this training set and a new ensemble of simulations produced. Means of the ensemble are obtained for $t = M+h+1, M+H+2, \ldots, N$ and stored. This process is repeated until the last observation is used for training. Each test set can be segmented into subseries of length h. Model performance can be evaluated by applying similarity/distance measures to the stored means and corresponding observations for each subseries.

For a given iteration, the marked changes in model performance across the subdivided test series is an indicator of temporal instability. Across iterations and for a given subperiod of the test set, performance measure values can be compared to assess the sensitivity of model simulations to variations in the temporal origin of the training data.

Finally, if the number of parameters is small, XY plots of the parameter estimates against iteration number may be informative. Otherwise and when appropriate, the spatial-temporal stability of model parameter estimates can be gauged by comparing parameter estimate maps for each iteration.

References

Petropoulos F, Goodwin P, Fildes R (2017) Using a rolling training approach to improve judgmental extrapolations elicited from forecasters with technical knowledge. Int Forecast 33:314-324. https://doi.org/10.1016/j.ijforecast.2015.12.006

Shmueli G (2010) To explain or to predict? Stat Sci 25:289-310. https://doi.org/10.1214/10-STS330
Tashman LJ (2000) Out-of-sample tests of forecasting accuracy: an analysis and review. Int J Forecast 16:437-450. https://doi.org/10.1016/S0169-2070(00)00065-0

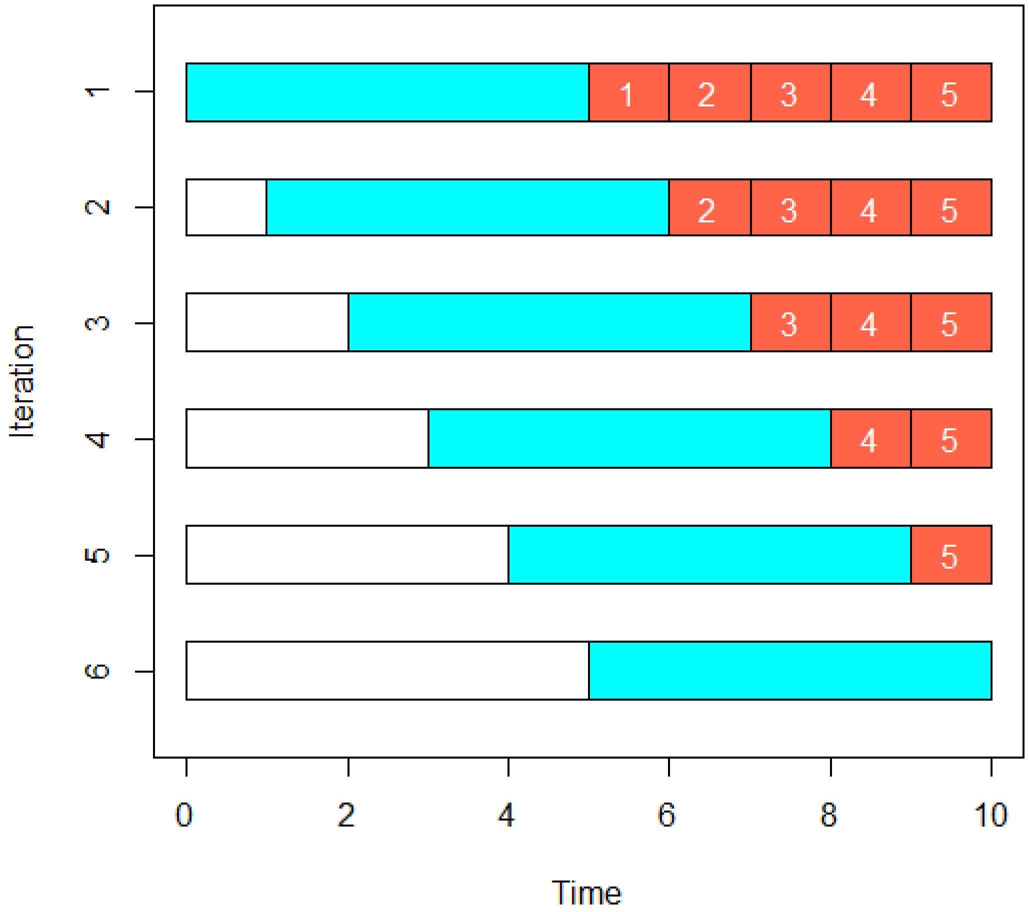

Figure 16:Schematic of rolling origin procedure for model evaluation.

For a given iteration: the filled cyan box indicates observations included in the training set; the filled red box indicates the test set, with each numbered interior box containing approximately h observations; and the white box indicates observations that are temporarily excluded from the training or test set.