

Multisite rainfall and evaporation data generation for the Macquarie Valley

Michael Leonard, Seth Westra and Bree Bennett



Disclaimer

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.

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.

The Intelligent Water Decisions research group maintains no responsibility for the misrepresentation of results due to incorrect usage of the information contained within this report.

This report should remain together and be read as a whole.

This report has been prepared solely for the benefit of the client NSW Department of Climate Change, Energy the Environment and Water.

Department Reference Number: PUB24/326

No liability is accepted by the Intelligent Water Decisions research group with respect to the use of this report by third parties without prior written approval.

Contact

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

Acknowledgements

We acknowledge Forugh Dorani and Dushmanta Dutta from the New South Wales Department of Planning and Environment, as well as Owen Droop from ODHydrology for the provision of data and constructive conversations on the relevant modelling requirements. A/Prof. Mark Thyer provided considerable guidance and invaluable discussions on the climatic modelling elements of this project.

Contents

Executive summary	8
1 Introduction.....	11
1.1 Project scope	11
1.2 Background	12
2 Methodology.....	14
2.1 Observation data.....	14
2.1.1 Rainfall data.....	14
2.1.2 Evaporation data.....	15
2.1.3 Instrumental records of low-frequency climatic variability	18
2.1.4 Paleo records of low-frequency climatic variability.....	18
2.1.5 Analysis of IPO-partitioned rainfall data.....	19
2.1.6 Analysis of IPO-partitioned evaporation data	21
2.2 Model specification – base model.....	23
2.2.1 At-site rainfall model.....	24
2.2.2 At-site evaporation model	26
2.2.3 Multisite rainfall model.....	30
2.2.4 Multisite evaporation model	31
2.2.5 Joint simulation of rainfall and evaporation.....	33
2.3 Extended model specification – climatic variability and future climate.....	35
2.3.1 IPO dwell-time distribution from instrumental and paleoclimatic record.....	35
2.4 Model calibration.....	36
2.4.1 Step 1 Rainfall: single-site rainfall distribution	36
2.4.2 Step 2 Rainfall: single-site temporal correlation.....	36
2.4.3 Step 3 Rainfall: spatial correlation	36
2.4.4 Step 4 Evaporation: single-site evaporation distribution.....	37
2.4.5 Step 5 Evaporation: single-site temporal correlation	37
2.4.6 Step 6 Evaporation: spatial correlation.....	37
2.4.7 Step 7 Rainfall–evaporation: spatial correlation	37
2.4.8 Step 8 Space–time correlation.....	37
2.4.9 Step 9 IPO calibration.....	37
2.5 Model evaluation	38
2.5.1 Evaluation of distribution quantiles.....	38
2.5.2 Evaluation of distribution of monthly totals.....	40
2.5.3 Pooling performance over multiple sites.....	40
3.1 Model evaluation summary	42
3.2 Detailed evaluation of distributions.....	44
3.2.1 Multiyear annual rainfall totals (1, 2, 5 and 10 years).....	44

3.2.2	Mean and standard deviation of monthly rainfall totals.....	46
3.2.3	Annual wet day proportion.....	47
3.2.4	Monthly wet day proportion (mean and standard deviation).....	50
3.2.5	Annual rainfall maximums (1, 2 and 3 day).....	51
3.2.6	Multiyear annual evaporation totals (1, 2, 5 and 10 years).....	52
3.2.7	Mean and standard deviation of monthly evaporation totals.....	56
4	Discussion and recommendations.....	58
4.1	Model summary.....	58
4.2	Summary of model evaluation.....	58
4.3	Recommendations.....	60
5	References.....	61
	Appendix A – Finalised list of sites	62
	Appendix B – Traffic light comparison for Model A (base), Model B (instrumental IPO) and Model C (paleoclimatic IPO).....	70

List of figures

Figure 1	Locations of rainfall stations in Macquarie Valley.....	14
Figure 2	Distribution of average annual rainfall in Macquarie catchment (left) and monthly totals (right).....	15
Figure 3	Locations of evaporation stations in Macquarie Valley.....	16
Figure 4	(left) Morton Wet evaporation (middle) IQQM evaporation and (right) FAO56 evaporation.....	17
Figure 5	Daily IQQM station evaporation characteristics.....	18
Figure 6	Time series of Interdecadal Pacific Oscillation (red) showing positive and negative states (black).....	18
Figure 7	Comparison of the instrumental IPO time series with the combined paleo IPO time series, 1550 to 2000 CE.....	19
Figure 8	Distribution of dwell times for phases of the IPO; solid lines represent the estimated distribution from the instrumental IPO record, dashed lines represent the estimated distribution from paleoclimatic IPO reconstructions.....	19
Figure 9	Mean annual rainfall at sites in Macquarie catchment reported by IPO phase.....	20
Figure 10	Yearly time series of annual total rainfall.....	21
Figure 11	Mean annual evaporation at sites in Macquarie catchment reported by evaporation type and IPO phase.....	22
Figure 12	Yearly time series of annual total evaporation (Morton Wet sites only).....	22
Figure 13	Average annual rainfall vs average annual evaporation for evaporation sites, stratified by IPO phase.....	23
Figure 14	Schematic of latent variable concept.....	24
Figure 15	Distribution of lag 1 daily autocorrelation parameter across rainfall sites for each month...25	25
Figure 16	Distribution of rainfall; observed (red symbol) versus 100 simulated replicates (grey lines) monthly means (left) and monthly standard deviations (right).....	25
Figure 17	Distribution of daily evaporation for a representative site for each month.....	26
Figure 18	Schematic of method used to reproduce skewness in distribution of daily evaporation.....	26
Figure 19	Daily evaporation for a representative site (top) fitted mean trend, (bottom) residuals after removing the mean.....	27
Figure 20	Fitted trends to mean-corrected residuals (top) positive and (bottom) negative.....	28
Figure 21	Standardised residuals of evaporation.....	29
Figure 22	Autocorrelation of standardised residuals for a representative site, horizontal axis represents daily lags.....	29
Figure 23	Example simulation from single-site model of evaporation.....	30
Figure 24	Distribution of rainfall. Observed (red symbol) versus 100 simulated replicates (grey lines) monthly means (left) and monthly standard deviations (right).....	30
Figure 25	Rainfall sample correlation values with distance (km) for a representative site and month...31	31

Figure 26 Rain–rain correlation matrix for 100 sites.....	31
Figure 27 Structure of evaporation correlations (left) shows sample correlation with distance for a representative site for different evaporation types	32
Figure 28 Evaporation–evaporation correlation matrix for 45 sites	32
Figure 29 Rainfall–evaporation sample correlation values with distance for a representative site and month, observations (black) and fitted correlation function (red).....	33
Figure 30 Nested block correlation matrix of a joint simulation of 100 rain sites and 45 evaporation sites.....	34
Figure 31 Nested block correlation matrix joint simulation of 145 sites and 2 timesteps; t and $t + 1$	35
Figure 32 Example simulation of 1,000 years of IPO states from the alternating renewal model with parameters from the IPO instrumental record.....	36
Figure 33 Flow chart of performance classification for annual distributions into 3 categories (Good, Fair, Poor) using the criteria specified, according to 2 tests.....	39
Figure 34 Flow chart of performance classification of simulation into 3 categories (Good, Fair, Poor) using the criteria specified, according to 2 tests	41
Figure 35 Distribution of annual rainfall totals for 1-, 2-, 5- and 10-year totals for a representative location.....	45
Figure 36 Distribution of (left) means of monthly totals; (right) standard deviations of monthly totals	46
Figure 37 Paired comparison of full distribution of monthly totals for one representative site, R1 50018, and one replicate.....	46
Figure 38 Distribution of the proportion of wet days within a given year for 4 representative sites with different performance outcomes (Good, Fair and Poor).....	49
Figure 39 Distribution of means of monthly proportions of wet days (left) and standard deviations of monthly proportions of wet days (right) for 3 different representative sites with different performance classifications (Good, Poor, Fair)	51
Figure 40 Frequency analysis of (left) annual maximum daily rainfall (middle) annual maximum 2-day rainfall (right) annual maximum 3-day rainfall for 4 representative sites R25–R28 for different performance outcomes.....	53
Figure 41 Distribution of annual evaporation totals for 1-, 2-, 5- and 10-year totals for a representative location with Fair/ Good performance; simulation shows 90% confidence interval	54
Figure 42 Distribution of annual rainfall totals for 1-, 2-, 5- and 10-year totals for a representative location with Poor performance; simulation shows 90% confidence interval.....	55
Figure 43 Distribution of monthly evaporation totals; (left) means of monthly totals and (right) standard deviations of monthly totals	57

List of tables

Table 1	Rainfall evaluation summary of performance, Paleoclimatic IPO model variant; 117 sites, 129 years length, 77 replicates.....	9
Table 2	Evaporation evaluation summary of performance, Paleoclimatic IPO model variant; 79 sites, 129 years length, 77 replicates.....	9
Table 3	Cases of simulated data outputs; Model C is intended as the primary output for evaluation of historical climate conditions.....	12
Table 4	Aggregate performance categorisation criteria.....	40
Table 5	Rainfall evaluation summary of performance, Model C, paleoclimatic IPO model variant, 117 rainfall sites 129 years length, 77 replicates.....	43
Table 6	Evaporation evaluation summary of performance, Model C, paleoclimatic IPO model variant, 79 evaporation sites 129 years length, 77 replicates.....	44
Table 7	Observed versus simulation comparison for 4 representative sites.....	47
Table A.1	List of 100 rainfall sites used in the study.....	63
Table A.2	List of 45 evaporation sites used in the study (31 unique locations because E32–E45 are co-located).....	67
Table B.1	Rainfall evaluation summary of performance for 3 different model variants.....	70
Table B.2	Evaporation evaluation summary of performance for 3 different model variants.....	72

Executive summary

The production of long stochastic time series of climate variables such as rainfall and evapotranspiration is often used to supplement the historical climate record when conducting drought risk assessments. While historical data provides one realised set of climatic conditions, stochastic models enable the generation of extended synthetic climatic conditions which are just as plausible as those occurring in the past.

To this end, 10,000 years of jointly simulated stochastic data have been generated for 100 rainfall sites and 45 evaporation sites in the Macquarie River catchment, with 7 different model variants developed to simulate the stochastic sequences, each with different assumptions regarding the role of natural climate variability and anthropogenic climate change. The 'historical' climate runs comprise Model A, which is the base model without climate partitioning; Model B, which accounts for shifts based on the instrumental record of the Interdecadal Pacific Oscillation (IPO); and Model C, which includes paleoclimate information to improve the estimation of the dwell time in each IPO phase. Although all 3 model variants reflect the historical climate, Model C is identified as the best representation of historical observations.

The performance of Model C is evaluated in detail relative to historical rainfall and evaporation over the period from 1890 to 2018. A high-level summary is provided in Table 1 for rainfall sites and Table 2 for evaporation sites. The labels 'Overall Good', 'Overall Fair' and 'Overall Poor' are detailed in Section 2.5 and arise from the consistent application of a defined set of tests to all sites and to a set of 20 relevant variables. Of the statistics presented here, 12 are identified as Overall Good, 7 are identified as Overall Fair and 1 is identified as Overall Poor. An intermediate level summary can be found in the Section 3.1, and plots for each site can be found in the annex documentation corresponding to each statistic.

Table 1 Rainfall evaluation summary of performance, Paleoclimatic IPO model variant; 117 sites, 129 years length, 77 replicates

Statistic	Overall evaluation of model performance
Distribution of annual total rainfall	Good
Distribution of 2-year rainfall totals	Good
Distribution of 5-year rainfall totals	Good
Distribution of 10-year rainfall totals	Good
Mean of monthly rainfall totals	Good
Standard deviation monthly rain totals	Fair
Distribution of annual proportion of wet days	Poor
Mean of monthly proportion of wet days	Good
Standard deviation monthly proportion of wet days	Fair
Annual 1-day rainfall maximum distribution	Good
Annual 2-day rainfall maximum distribution	Fair
Annual 3-day rainfall maximum distribution	Fair

Traffic light criterion specified by systematic evaluation method (Section 2.5).

Table 2 Evaporation evaluation summary of performance, Paleoclimatic IPO model variant; 79 sites, 129 years length, 77 replicates

Statistic	Overall evaluation of model performance
Distribution of annual total evaporation	Good
Distribution of 2-year evaporation totals	Good
Distribution of 5-year evaporation totals	Good
Distribution of 10-year evaporation totals	Good
Mean of monthly evaporation totals	Good
Standard deviation monthly evaporation totals	Good

Traffic light criterion specified by systematic evaluation method (Section 2.5).

The outcomes of the stochastic modelling, and anticipated implications for water security assessments, can be summarised as follows:

- **Multiyear annual totals rainfall/evaporation** – the model reproduces these statistics well, which is critical for application to drought assessment.
- **Monthly totals rainfall/evaporation** – the means of the monthly totals are considered Good, and the standard deviations are considered Fair. Discrepancies in simulated data are discussed in the report and show for a representative rainfall site that the Fair classification of the standard deviation of monthly totals is due to occasional simulated wet months that inflate the standard deviation. For evaporation, some months may have a standard deviation which is 2 mm lower

than observed. For a reported example at one site, the observations range from 85 mm to 115 mm (90% confidence interval) while the simulations range from 88 mm to 112 mm.

- **Proportion of wet days** – this statistic is classified as Overall Poor in the annual distribution, but Overall Good for mean of monthly proportions and Overall Fair for the standard deviations of monthly proportions. Interpreting this performance, the proportion of wet days is unbiased, but has insufficient variability at the annual scale. For the driest year on record from a representative site, this translates to an extra 10 wet days in the median simulation compared to the observation. However, the process of rainfall amounts compensates for this lack of variability (by simulating less rainfall per wet day in these situations) so that rainfall totals are unbiased and there is good reproduction of the variability of annual totals.
- **Annual maximums** – the annual maximums are Overall Good for 1-day maximums and Overall Fair for 2-day or 3-day maximums. Where there is Fair performance, it is shown that the observed values are not far outside the border of the 90% confidence interval of simulated extremes. The performance of these statistics is mostly relevant to flood studies, for catchments that have a response time in the order of several days.

Compared to single-site rainfall models, the class of models that are able preserve multisite statistics from daily to interannual scales across both rainfall and evaporation variables is not large. Based on the performance summary, the model outlined in this report was able to account for key attributes of the multisite rainfall and evaporation observations for the historical record.

The generated daily time series of 10,000 years length are recommended for use in hydrological modelling studies of the Macquarie River catchment. Model C is recommended for use as the best representation of the historical climate, because it includes information on the IPO from paleoclimate records.

1 Introduction

Water in the Macquarie Valley serves many purposes, including irrigation, watering stock, domestic use, town supply, ecological and recreational demands. The valley supports a diversity of agricultural activity, including a wide variety of crops: cotton, cereals, wine grapes, oil seed, legumes and tree crops such as olives, nuts and cherries. However, flows in the Macquarie River are highly variable, and show evidence of prolonged periods of below-average flow. For example, during the Federation drought, storages fell to 4% of full supply volume, causing suspension of the water sharing plan and significant water restrictions in townships such as Dubbo.

The New South Wales Department of Planning and Environment has developed a risk-based method to assess impacts of water scarcity on the regional economy and ecosystem assets of the Macquarie Valley. The method is based on hydrological and economic modelling, in which the Integrated Quantity and Quality Model (IQQM) water resources model will be used to generate daily water balance simulations over a long climatic period.

To facilitate daily water balance modelling, stochastic inputs are required to account for variability and potential projected future changes to rainfall and evaporation at many sites. To simulate key hydrological features of interest, the stochastic inputs must reproduce variability and gradients in the rainfall and evaporation across all sites within the region, and on multiple timescales – daily, monthly, annual and multiannual.

1.1 Project scope

The project scope is to deliver stochastically generated rainfall and evaporation at multiple sites within the Macquarie Valley and surrounding region. The list of sites is provided in Appendix A and comprises 117 rainfall sites and 79 evaporation sites. There are 196 time series to be generated, but only 117 unique locations because the 79 evaporation sites are also rainfall sites. A 129-year common period (1890–2018) is used to calibrate and evaluate the rainfall/evaporation model.

A single replicate of length 10,000 years is required for several climatic cases (Table 2). In each case the climate model uses long-term stationary assumptions.

- Model A is calibrated against daily and seasonal variability from the observed record but does not account for interannual sources of variability that are available from climate indices.
- Model B uses separate parameters calibrated against the positive and negative phases of the Interdecadal Pacific Oscillation (IPO) to account for interannual variability, using IPO data from the instrumental record (that is, from 1890–2018).
- Model C uses paleoclimate information about the IPO to inform the distribution of possible dwell times in each phase of the IPO. Model C uses the most complete available information on long-term climate variability by blending both historical and paleoclimate information, so is intended as the primary characterisation of historical variability, whereas Models A and B are mostly intended for comparisons.

The generated time series for Models A–C are intended to characterise statistical features of the historical rainfall, including interannual and multidecadal variability, with a view to replicating the climatic drivers of drought. The output is generated as a single time series (for example, 10,000 years), which provides maximum flexibility – it can be used as a single extended time series or broken

up into shorter replicates to suit modelling requirements (for example, 100 × 100-year replicates instead of a single 10,000-year replicate). To statistically evaluate the performance of the model, the 10,000-year simulations are partitioned into 77 replicates of 129 years each to match the length of the historical record; this allows direct comparison of statistical quantities. The evaluated statistics include the proportion of wet days, means and standard deviations of monthly and annual totals of rainfall, variability in multiannual totals (2-, 5- and 10-year totals) and an analysis of 1-day, 2-day and 3-day annual maximums.

Table 3 Cases of simulated data outputs; Model C is intended as the primary output for evaluation of historical climate conditions

Model	Case description
A	Base case: no climate partitioning
B	Instrumental IPO: conditioned on the Interdecadal Pacific Oscillation, instrumental record only
C	Paleo IPO: conditioned on the IPO using paleo proxies to inform climate states

1.2 Background

The Macquarie River in central western New South Wales forms part of the Murray–Darling basin. The catchment headwaters are in the Great Dividing Range east of Bathurst with catchment peaks approaching 1,400 mAHD and an annual average rainfall greater than 1,200 mm. The river flows northwest for 960 km towards the Barwon River near Brewarrina, with an elevation of about 100 mAHD. Rainfall over most of the lower catchment averages 300–500 mm/year, with summer months the wettest. Annual average evaporation ranges from 1,100 mm to 1,800 mm across the catchment.

The Macquarie River has highly variable flows; for example, at Dubbo the annual flows have historically ranged from 24.1 ML to 10,113 GL (NSW Department of Primary Industries, 2016), with annual average flow 1,175 GL. The low flows typically occur during extended (that is, multiyear) drought periods (for example, 1935–42, 2001–2009). Because of decreasing channel capacity, irrigation and the presence of numerous effluent channels, annual average flow reduces downstream of Dubbo, and is less than 330 GL above the Macquarie marshes and less than 150 GL below the marshes (NSW Department of Primary Industries, 2016). The catchment is hydrologically complex, with numerous anabranches in the lower reaches and surface water/groundwater interactions in the upper Macquarie (NSW Department of Industry, 2018).

There are 2 significant dams in the headwaters, Windamere Dam (368 GL, on the Cudgegong River, built in 1984) and Burrendong Dam (1,678 GL, on the Macquarie River, built in 1967). The dams provide water for irrigators, stock, domestic use and town supply as well as flood storage capacity. The catchment population is less than 200,000 people, concentrated in the town centres of Bathurst, Dubbo, Mudgee, Orange and Wellington. The towns have higher priority access to water than do irrigators. Over 80% of the land is used for agriculture – mainly sheep and cattle grazing as well as a range of broadacre and orchard/higher value horticulture (for example, cotton, cereals, wine grapes, oil seed and legumes, and tree crops such as olives, nuts and cherries). The river also supports an array of recreational activities including fishing, water sports, bushwalking and camping.

The quantity and quality of water are equally significant for the biodiversity of the Macquarie River ecosystem. The Macquarie marshes are listed under the Ramsar convention and are one of the

largest inland semi-permanent wetlands in the Murray-Darling Basin. The marshes include a diverse range of habitat and vegetation types, hosting numerous endangered and threatened species of native birds and fish. There are numerous additional ecosystem assets including lakes, billabongs, wetlands and flood-dependent forests.

Given the high degree of variability in the climatic system there are significant challenges in managing the provision of water for irrigation and ecosystem functions. Future climate change poses additional risks to the region, and though uncertain, it is expected that surface water availability is more likely to decrease than increase across the basin (van Dijk et al., 2016). In particular, there are significant consequences if future droughts occur with similar or greater magnitude than historical droughts. If the water security reaches critically low levels there is the potential for high water restrictions, loss of ecosystem assets, biodiversity stress, no water allocated for agricultural activity, reduced mining activity and significant indirect impacts on local townships and the regional economy.

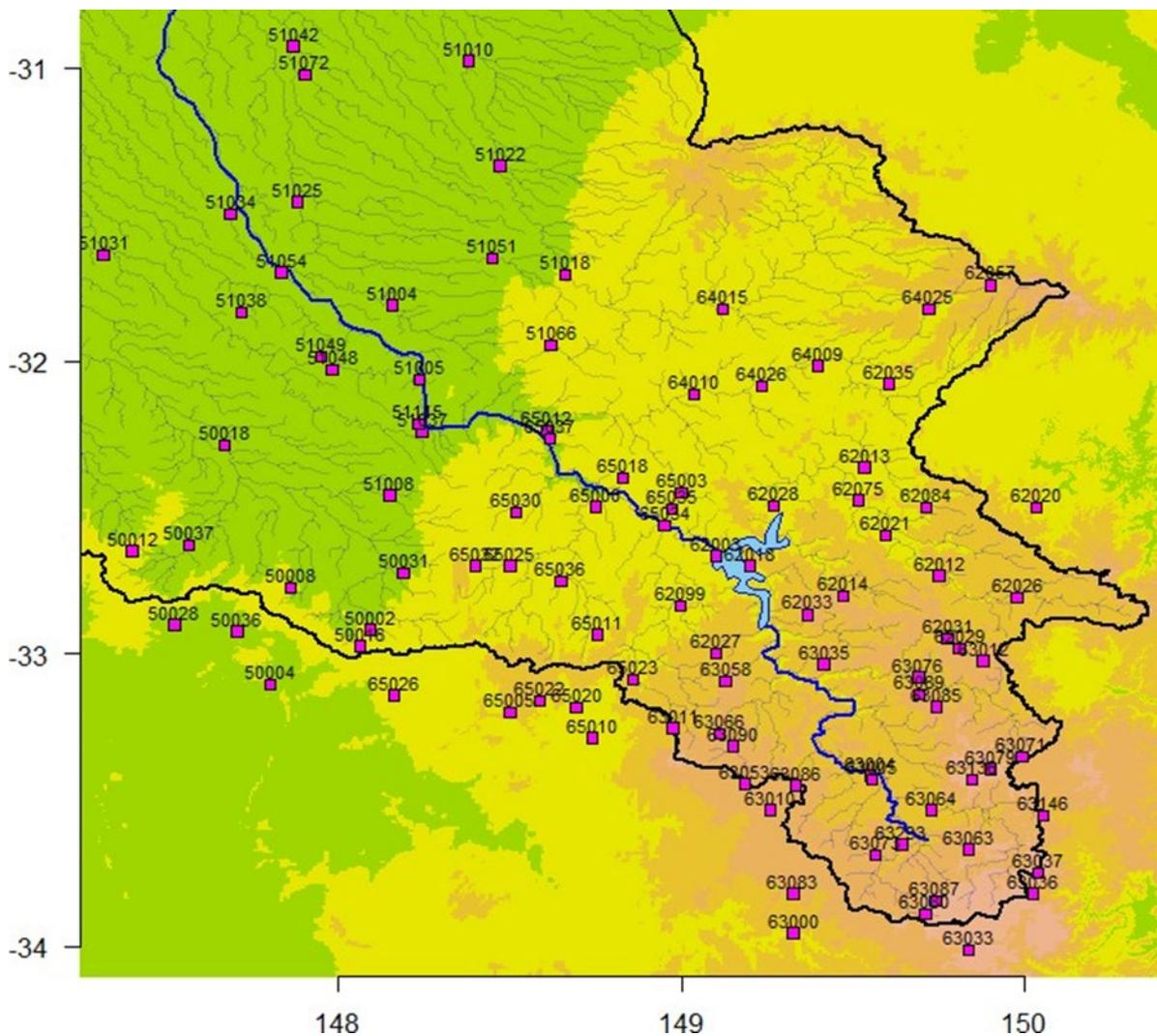
An IQQM water balance model was developed to inform the Water Sharing Plan for the Water Sharing Plan for the Macquarie and Cudgegong, and for diversion compliance purposes. The Macquarie IQQM model takes into account contemporary infrastructure, water access entitlements and water sharing rules. The Macquarie IQQM model will underpin risk-based assessment, with critical outputs including estimates of water availability and allocation for various scenarios of economic and environmental impact. This project exists to provide stochastic data of climatic forcings as a necessary input for water balance modelling.

2 Methodology

2.1 Observation data

2.1.1 Rainfall data

Figure 1 shows the locations of rainfall data, which were provided at 100 locations (see Appendix A for complete list). The data were originally sourced from the SILO patch point database, so there were no missing values. All sites had considerable length of data prior to infilling methods from SILO. The start dates are either 1 January 1889 or 1 January 1890 and the end date is 11 September 2018. All available data from 1890 onwards was used for calibration. For evaluation, the last portion of 2018 was padded with data from 2017 to give an indicative total for 2018 rather than to discard 9 months of data within the evaluation. This means all evaluation replicates 129 years long.



See Appendix A for site details.

Figure 1 Locations of rainfall stations in Macquarie Valley

Analysis of the observed data shows that the highest rainfalls are concentrated in the mountainous south-east corner of the catchment (Figure 2, left). There is a significant gradient over the catchment with a majority of the lower catchment receiving less than 500 mm rainfall on average. The seasonal distribution (Figure 2, right) shows that there is considerable variability across the sites. The highest rainfalls are in summer, autumn rainfall is less than summer and winter is more variable than other seasons.

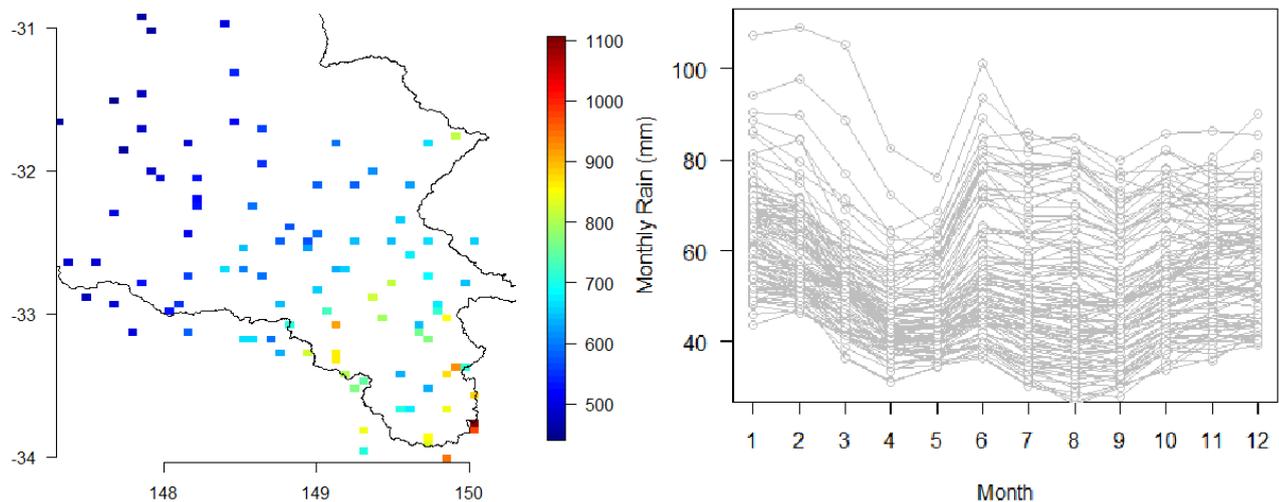


Figure 2 Distribution of average annual rainfall in Macquarie catchment (left) and monthly totals (right)

2.1.2 Evaporation data

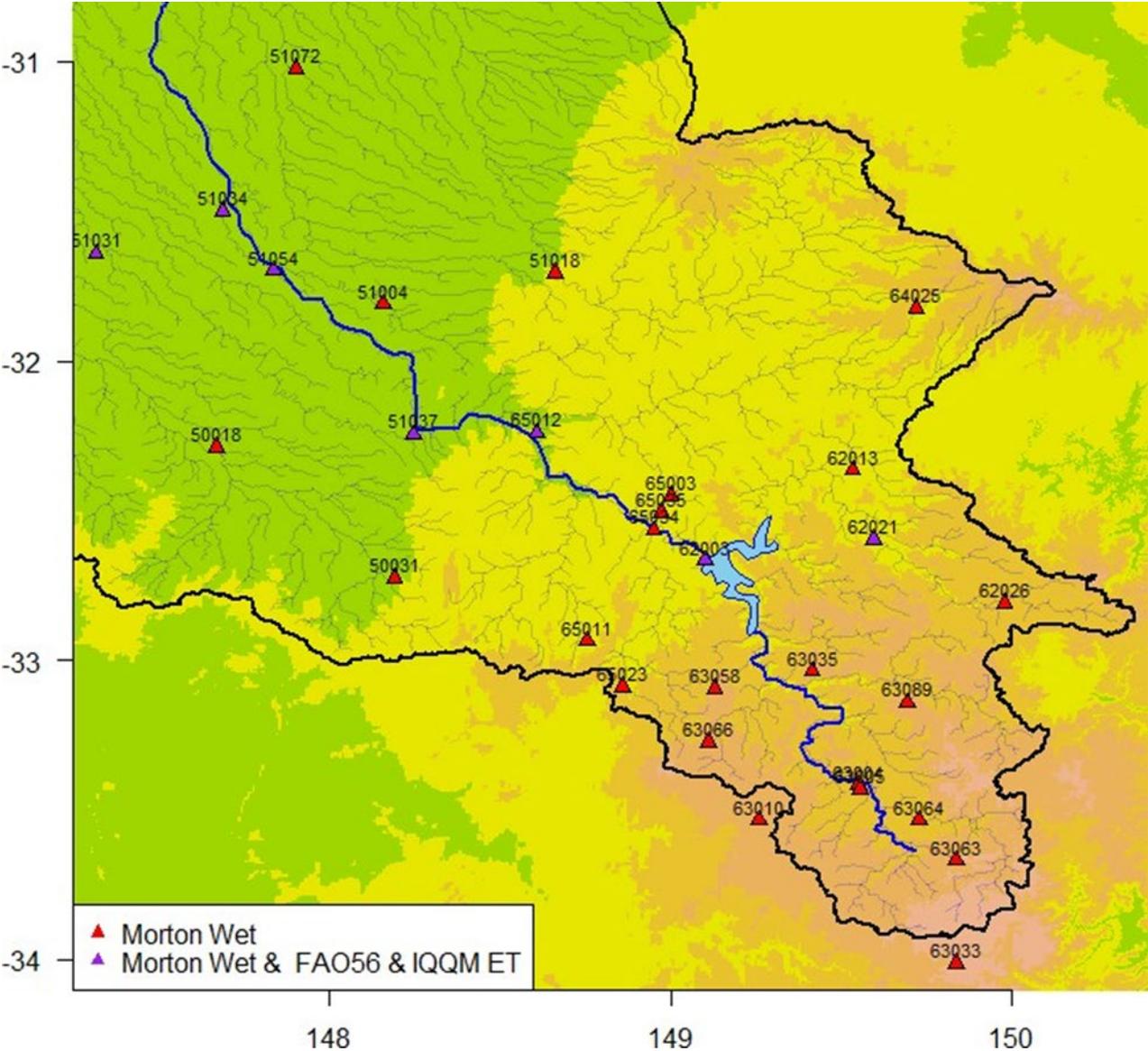
Data were provided for 31 Morton Wet sites, 7 IQQM evaporation sites and 7 FAO56 reference crop sites (the term evaporation is used throughout to mean evapotranspiration). There are 45 separate evaporation time series, but only 31 unique locations, because the 7 IQQM and 7 FAO56 time series were co-located at Morton Wet locations. The sites are listed in Appendix A and shown in Figure 3. The Morton Wet and FAO56 reference crop data were obtained from the SILO database, so no values were missing and the data cover the period 1 January 1889 to 11 September 2018. The IQQM estimates span the period 1 January 1890 to 28 August 2017. All available data were used in the model calibration.

The evaporation ranges from 1,100 mm in the mountainous south-east corner (Figure 4, top row) to 1,800 mm inland. There is a strong seasonal cycle, with highest evaporation in summer and lowest evaporation in winter (middle and bottom rows at monthly and daily timescales respectively). The 3 columns in Figure 4 show differences between the Morton Wet (left) IQQM (middle) and FAO56 (right) evaporation variants. The Morton Wet and FAO56 variants have similar features, with a gradient from east to west and similar magnitudes. The IQQM variant is noticeably different from the other versions, in that:

- the magnitude is significantly higher and shows a different spatial pattern (middle column, top row)
- the magnitude and the variability is higher (middle column, middle row)
- the daily distribution is 'blocky' at the monthly scale with stippling showing persistence of the same value for many days within the month (middle column, bottom row).

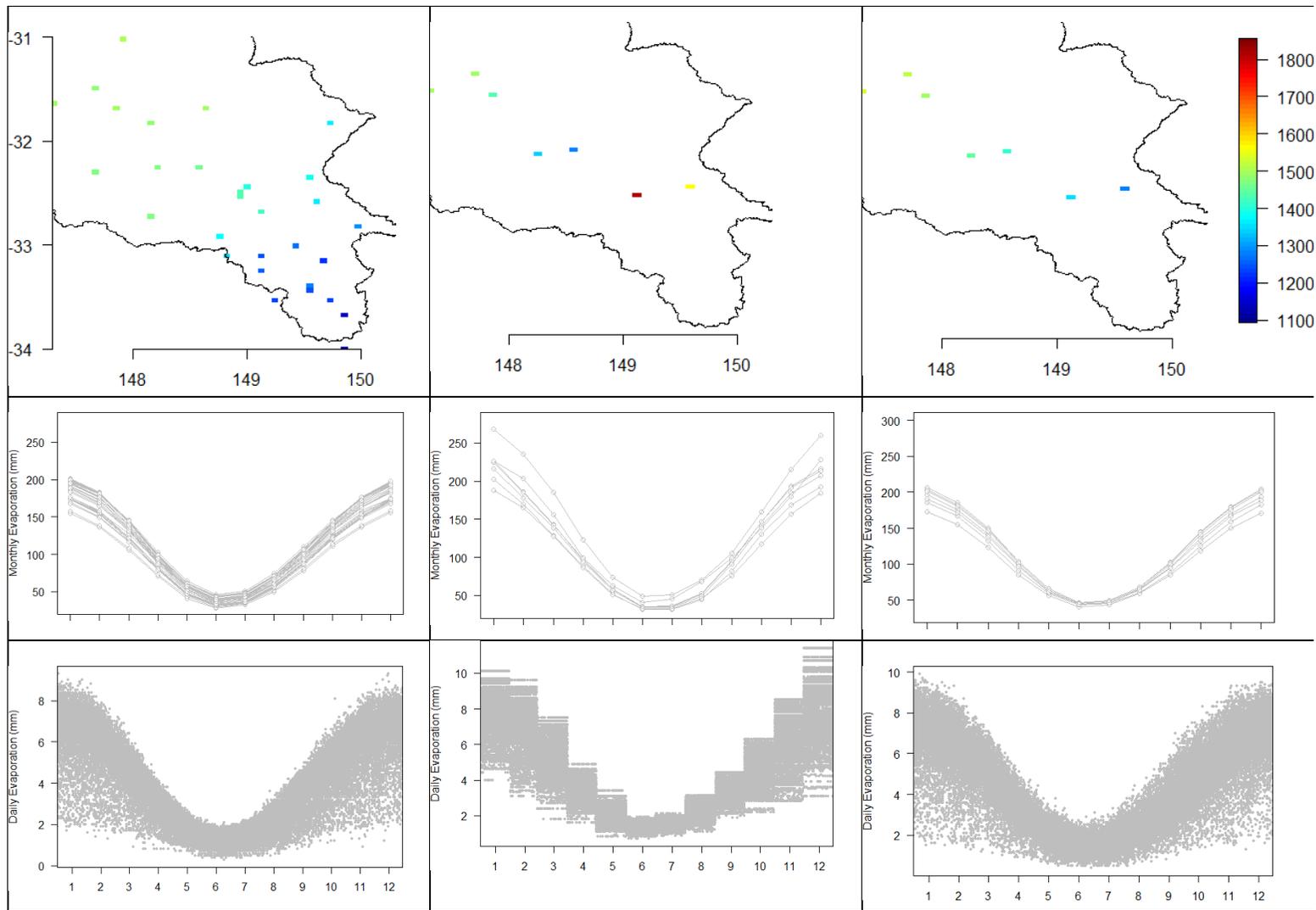
The provenance of the IQQM data is pan evaporation data extended using longer co-located rainfall records. The statistical model used correlates total monthly evaporation with monthly rain days for each month, and extends using the longer rainfall data set, introducing a random component to preserve monthly standard deviation. This monthly total is then disaggregated to give a value for dry days and a value for wet days (equal to 0.7 times the dry day value). Distinct features of the modelled data are evident in the daily time series (Figure 5).

The IQQM evaporation data are reasonable in that they are assumed to preserve the monthly distribution of evaporation, the seasonal cycle and known correlation with rainfall. The daily persistence of evaporation values has the implication that subsequent models based on these 'data' will have an anomalously high correlation at the daily scale. Also, although the mean of monthly evaporation totals is preserved, the variability of monthly totals is likely to be spuriously high. The interpretation for hydrological modelling is that, although the distribution of evaporation is preserved across multiple years, within any given year it could be biased due to persistence in daily evaporation (especially for blocks of dry days). Although this effect is noticeable in the evaporation data, it is not clear whether this is of practical significance to subsequent modelled flows.



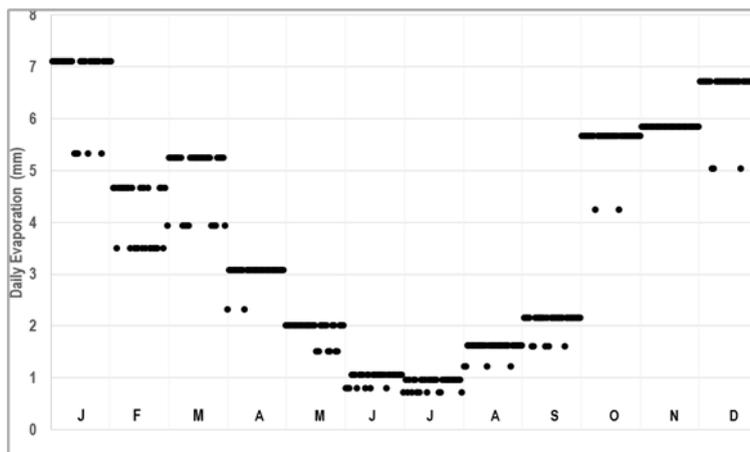
See Appendix A for site details.

Figure 3 Locations of evaporation stations in Macquarie Valley



Top row shows the spatial distribution, middle row shows monthly averages for each year on record, bottom row shows distribution of daily evaporation for a representative site. IQQM data have a different spatial distribution, higher seasonal variation and 'blocky' daily distribution.

Figure 4 (left) Morton Wet evaporation (middle) IQQM evaporation and (right) FAO56 evaporation



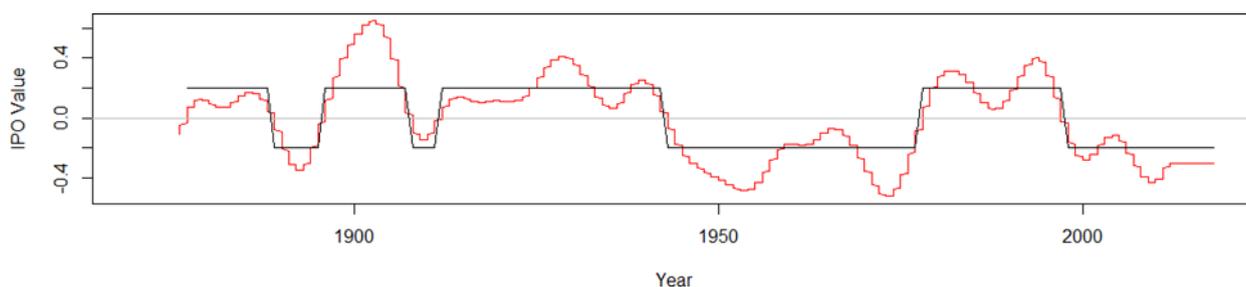
The IQQM data have distinctive characteristics based on how they were generated. The total for an individual month is determined from an empirical relationship (for that month) of total monthly evaporation and number of rain days. The total for that month is then distributed across the month as one value for days without rain and another for days with rain (set to 70% of the dry-day value).

Figure 5 Daily IQQM station evaporation characteristics

2.1.3 Instrumental records of low-frequency climatic variability

Data for the IPO were obtained from Henley et al. (2015) for the period 1854–2018. The Hadley SST version of the IPO was used, and is plotted in Figure 6, showing periods when the IPO is in a positive or negative state. Because the IPO is constructed as a low-pass filter, estimates are not available in the first 5 or last 5 years of the record. The IPO estimates end in 2012, so the state from 2012 to 2018 was assumed rather than excluding this period from the calibration. The IPO was used to partition data and calibrate the model separately to each partition. The partition years were:

- positive phase: 1877–1888, 1896–1907, 1912–1942, 1978–1997
- negative phase: 1889–1895, 1908–1911, 1943–1977, 1998–2012 (+ 2013–2018, assumed).



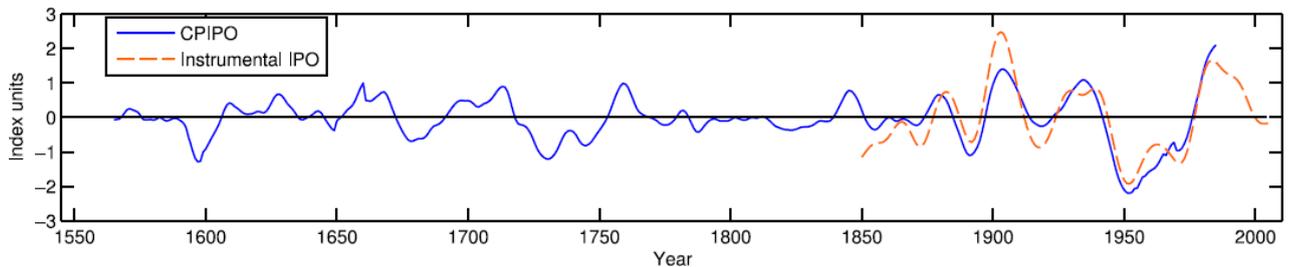
Source: Henley et al. (2015)

Figure 6 Time series of Interdecadal Pacific Oscillation (red) showing positive and negative states (black)

2.1.4 Paleo records of low-frequency climatic variability

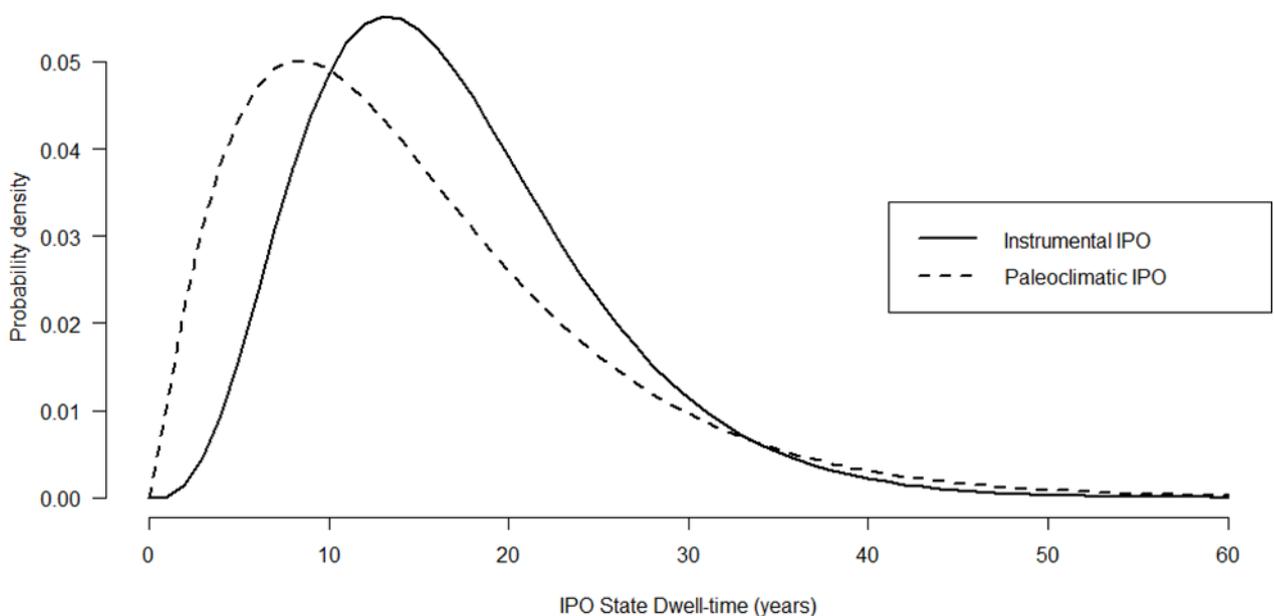
Henley et al. (2011) developed a weighted average of 7 paleoclimatic time series, including those from tree rings and coral from the Pacific Ocean, to produce a combined paleo IPO signal (referred to as the CPIPO index). Figure 7 compares the instrumental IPO time series with the CPIPO time series; the comparison is favourable (Nash–Sutcliffe efficiency 0.75, and comparable distributions of

run lengths). The distribution of run lengths was analysed and found that the gamma distribution was the most appropriate model to represent the dwell time in each IPO phase (Henley et al. 2011). Figure 8 compares the distribution of dwell times from the instrumental and the paleo records. The paleoclimatic IPO distribution has a lower mean, but higher variance. Although the longest dwell time in the instrumental record was 35 years (Figure 7, 1943–1977, IPO negative), both distributions show that it is possible to achieve dwell times of much greater duration (upper tail of Figure 8).



Source: Figure 1 from Henley et al. (2011)

Figure 7 Comparison of the instrumental IPO time series with the combined paleo IPO time series, 1550 to 2000 CE



Source: from Henley et al. (2011)

Figure 8 Distribution of dwell times for phases of the IPO; solid lines represent the estimated distribution from the instrumental IPO record, dashed lines represent the estimated distribution from paleoclimatic IPO reconstructions

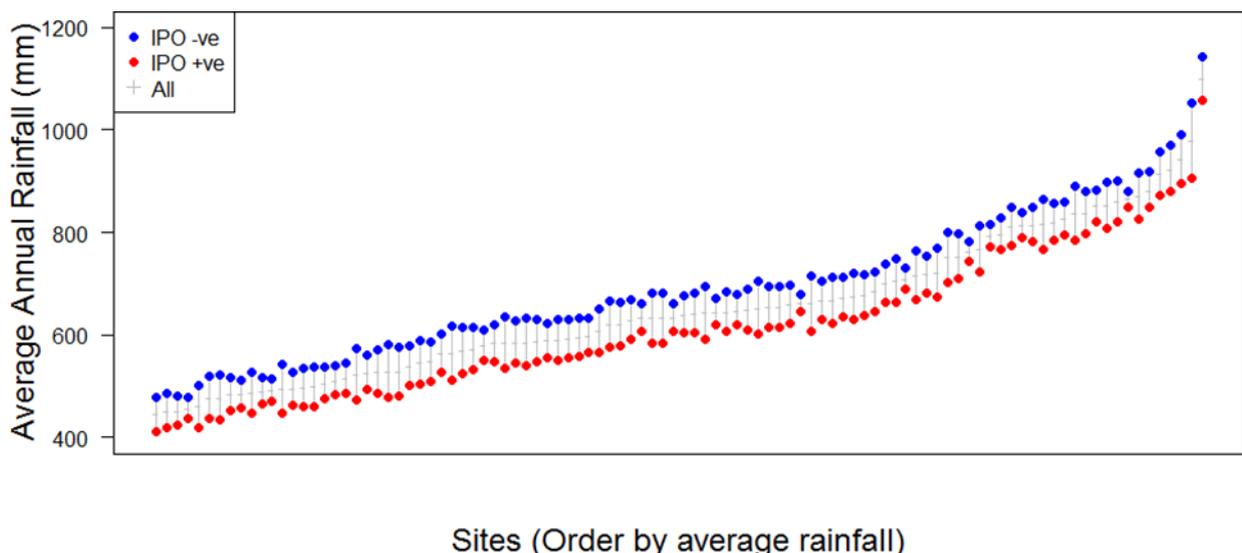
2.1.5 Analysis of IPO-partitioned rainfall data

Partitioning the rainfall time series by the IPO (time periods listed in Section 2.2.3) demonstrates a systematic shift in amounts between the 2 climate states. Figure 9 shows that on average the difference in annual rainfall between the IPO positive and negative states is 90 mm and that the phenomenon is consistent across all sites.

Figure 10 shows the time series of annual total rainfall for each year on record. The top panel shows the annual rainfall for each site (wettest on average shown at the top to driest at the bottom) along

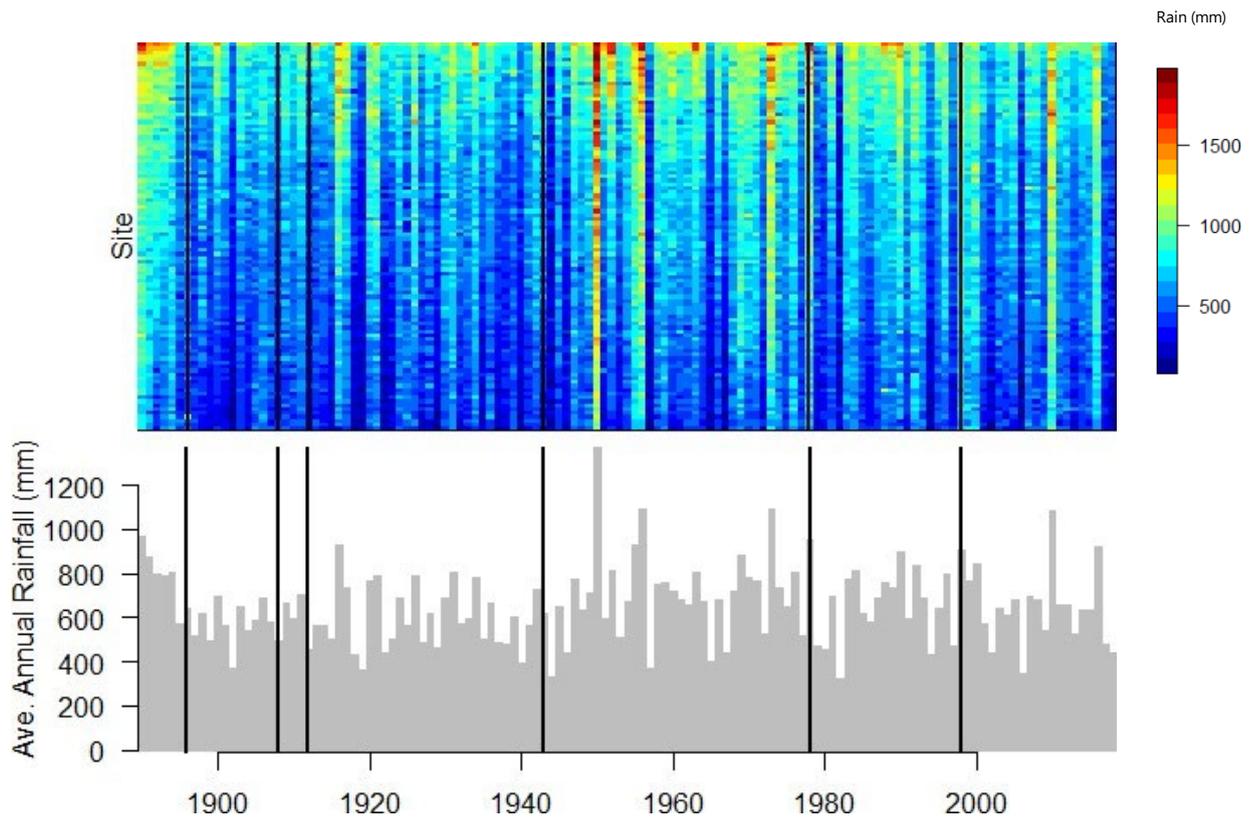
with black lines that mark transitions in the IPO state. The bottom panel shows the average across all sites. A number of features of the rainfall are evident:

- There is a strong rainfall gradient. Within a single year, annual totals can range from less than 300 mm to more than 1,800 mm.
- There is significant variation from year to year. The annual total can shift substantially and does not necessarily persist with similar values for multiple years.
- In addition to the overall variation in annual totals, there is the possibility for significantly above- or below-average periods to persist for multiple years (for example, 1890–1895 above average, 1879–1982 below average).
- The strength of spatial correlation is varied, but overall, very strong. The wettest years tend to have all sites with significantly increased rainfall (for example, 2010), whereas other 'wet' years have only a small number of very wet sites (for example, 1890s). To see this, consider the spread of yellow and red pixels, which represent the wettest sites: some years are very striped, showing that all sites have high rainfall, whereas other years show only a cluster of high-rainfall sites at the top of the top panel plot.
- Although the IPO partitions yield a 90 mm average difference for each site (Figure 10), the distribution of annual totals shows that there is significant variability within the IPO state. For example, the period 1997–2018 is an IPO negative phase (nominally the 'wet' phase) but has a similar average to the preceding IPO positive period 1978–1997.



Each vertical bar represents 1 site. The IPO negative state (blue symbols) show about ~90 mm more rainfall per year than the IPO positive state (red symbols).

Figure 9 Mean annual rainfall at sites in Macquarie catchment reported by IPO phase



Annual evaporation for each site (top panel) with sites sorted by their long-term average, with site with highest average at the top and lowest average at the bottom. Arithmetic average across all sites (bottom panel). Black lines show years that IPO state changed.

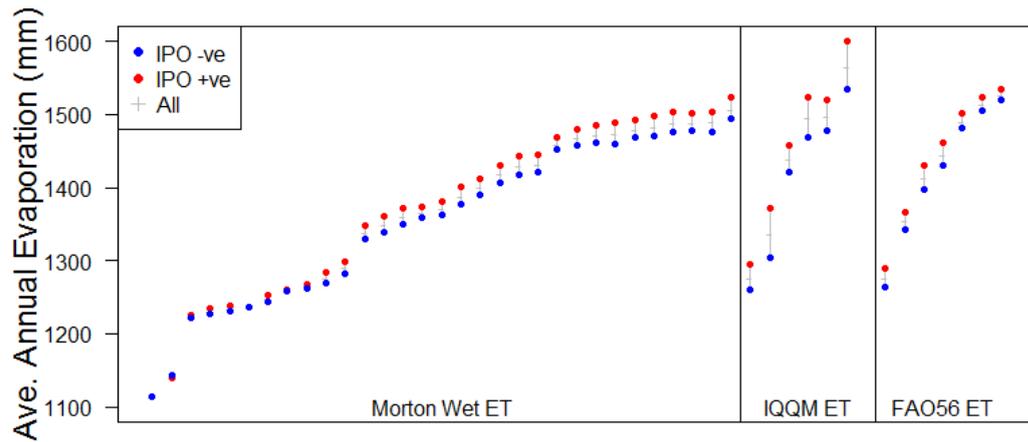
Figure 10 Yearly time series of annual total rainfall

2.1.6 Analysis of IPO-partitioned evaporation data

Figure 11 shows that for evaporation the effect of partitioning by the IPO is less pronounced than for rainfall. There is a 17 mm difference in evaporation between the positive and negative states.

Figure 12 shows the time series of annual total evaporation for each year on record. The top panel shows the annual evaporation for each site (highest average evaporation shown at the top) along with black lines that mark transitions in the IPO state. As with the rainfall, there is significant variation from year to year, a pronounced gradient and spatial correlation leading to years when the majority of sites have above or below average evaporation.

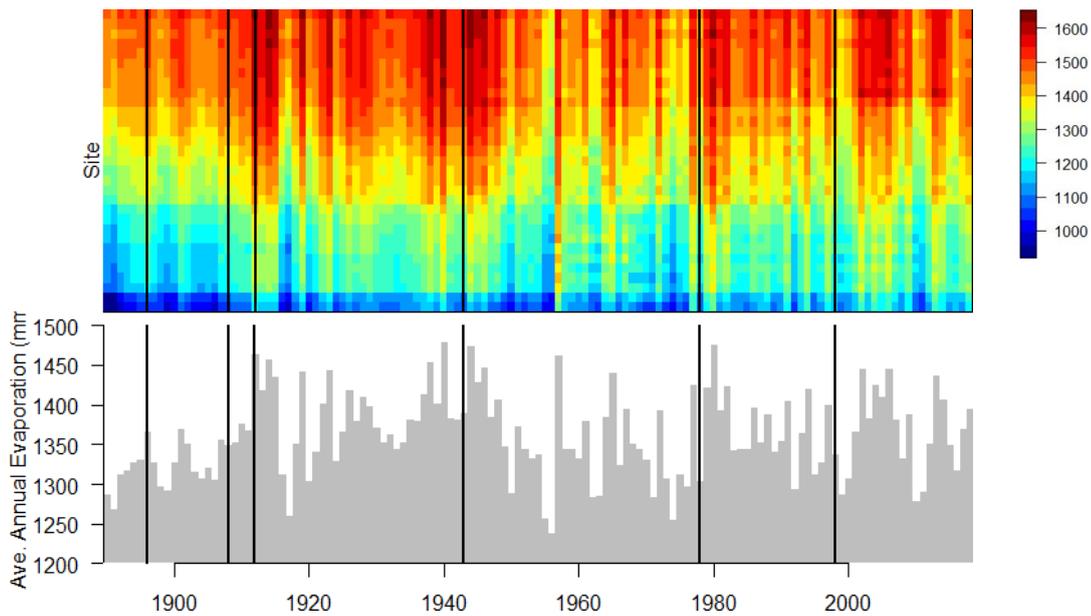
Figure 13 shows a scatterplot of annual rainfall with annual evaporation. There is a clear negative relationship between them. The plot shows years stratified by IPO phases; the 2 series overlap significantly, except for years with very high rainfall.



Sites (Order by average evaporation)

Each vertical bar represents 1 site. The IPO negative state (blue symbols) show, approximately 17 mm less average annual evaporation than the IPO positive state.

Figure 11 Mean annual evaporation at sites in Macquarie catchment reported by evaporation type and IPO phase



Annual evaporation for each site (top panel) with sites sorted by their long-term average, with site with highest average at the top and lowest average at the bottom. Arithmetic average across all sites (bottom panel). Black lines show years that IPO state changed.

Figure 12 Yearly time series of annual total evaporation (Morton Wet sites only)

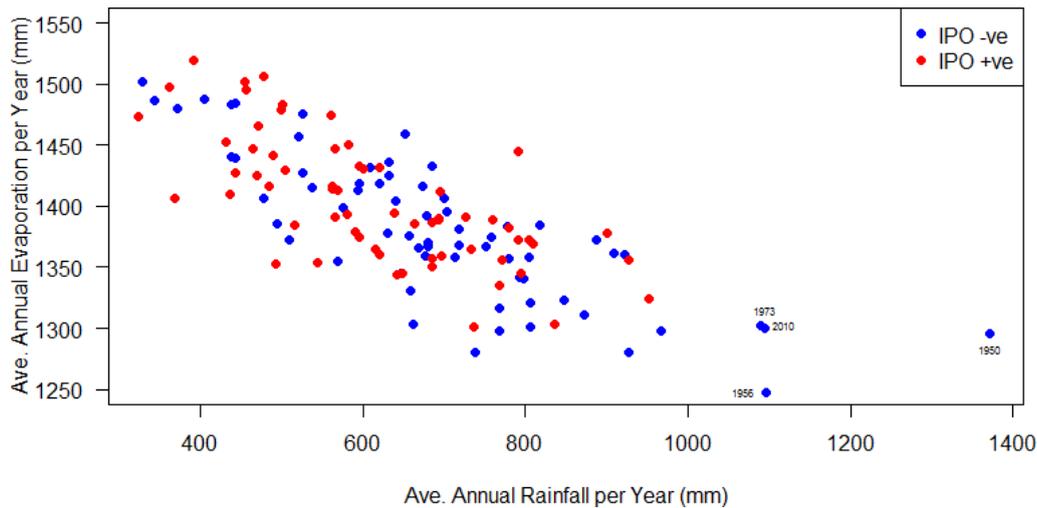


Figure 13 Average annual rainfall vs average annual evaporation for evaporation sites, stratified by IPO phase

2.2 Model specification – base model

To simulate daily rainfall and evaporation at multiple sites, it is important to have a model that can reproduce the marginal distribution at each site as well as the correlations between all sites. The model requirements are different for rainfall and evaporation:

- **Rainfall** – A large fraction of days are dry, and on rainy days the amount of rain follows a skewed distribution. There is temporal correlation in both the pattern of wet and dry values as well as the magnitudes. It is important to preserve this correlation because hydrological response can depend on the successive wetting and drying of a catchment. The correlation of wet values is also important for rainfall extremes, because many larger rural catchments have a flood response in the order of 1 or more days. Replication of both month-to-month and year-to-year variation in rainfall is important, so that the distribution of annual totals has appropriate variability. Lastly, it is possible for rainfall anomalies (above or below average periods) to persist for multiple years (or even decades), which is an essential consideration for drought studies. Spatially, it is important for a model to be flexible enough to permit differences at each site (for example, to represent trends across the catchment) as well as account for the correlation between sites (to reproduce catchment totals).
- **Evaporation** – Unlike rainfall, evaporation is always positive, which makes fitting a distribution to the data easier. Nonetheless, at the daily scale evaporation can follow a nonsymmetric distribution and have a complicated correlation structure: there is a known negative relationship with rainfall, but also considerable persistence over many days. The seasonal cycle of evaporation is well defined and shows smooth variation over time. The mean is large compared with the variance. As with rainfall, annual and multiannual totals are important for modelling drought. Spatially, evaporation varies smoothly (compared with rainfall, which can be patchy).

One popular approach for generating daily rainfall is a 2-step method that first simulates the wet–dry occurrences and then the conditional rainfall amounts (Kleiber et al. 2012; Wilks 1998). A challenge with this approach is to parsimoniously condition the amounts (whether rainfall or evaporation) on the wet–dry pattern, which can be challenging at multiple sites given the many wet/dry combinations. An alternative approach is to use a transformed latent (that is, hidden) variable that maps the wet and dry occurrences to a single distribution: dry values stem from the lower truncated

portion and the amounts stem from the upper portion (Baxevani and Lennartsson, 2015). A multisite version of this model was introduced by Rasmussen (2013), which has a convenient calibration structure to simplify the identification of parameters (the at-site means and standard deviations can be fitted separately from the autocorrelations and spatial correlations).

The following sections provide a description of the model used in this study. A full technical explanation of the rainfall model can be found in Bennett et al. (2018), so the following sections provide a brief conceptual explanation of the rainfall component. A more detailed explanation is provided here for the evaporation and evaporation–rainfall relationship because this is not covered in Bennett et al. (2018). Given the complexities due to evaporation data having 3 different types (Morton Wet, IQQM, FAO56), details are provided on how the correlation between the types was structured.

2.2.1 At-site rainfall model

The rainfall model uses a latent variable concept, which proceeds by sampling from a normally distributed ‘hidden’ variable. This concept is shown in Figure 14. A latent variable can be transformed to a rainfall value by truncating values below zero and by rescaling values above zero to match the distribution of rainfall observations. The transformation of latent variable value to rainfall uses a power transformation.

Figure 14 Schematic of latent variable concept

Let r_t^i be the rainfall at site i , where $i = 1, \dots, N$ and $t = 1, \dots, T$ is the time (days). For example, a 100-year simulation would have $T = 365 \times 100$ days (ignoring leap years). The rainfall amount can be related to a normally distributed latent variable, l_t^i , via truncation and power transformation,

$$r_t^i = \begin{cases} (l_t^i)^{\beta_t^i} & l_t^i > 0 \\ 0 & \text{otherwise} \end{cases}, \text{ where } l_t \sim N(\mu_{Rt}^i, \sigma_{Rt}^i) \quad (1)$$

where r_t^i is rainfall and β_t^i is a power transformation parameter. Note that the distribution is specified by 2 parameters, μ_{Rt}^i and σ_{Rt}^i , the mean and standard deviation for each site and timestep. Here, the parameters are varied monthly, so all timesteps within the same month have identical parameters. Simulating from this distribution reproduces the daily distribution of rainfall for a given time period (for example, month), including the proportion of zeros. An advantage of the model is that it has parameters to match the mean and variability of daily rainfall. However, the transformation is not always perfect because it needs to match the moments of the rainfall distribution as well as the proportion of zero values.

To model sequences of rainfall values, autocorrelation of the latent variable is considered. Because the variable is Gaussian, it is possible to use a single autocorrelation parameter for a given site to reproduce sequences of wet and dry values as well as correlation in wet day amounts. The temporal structure of the latent variable at a site is modelled via an autoregressive AR(1) process. That is, the value at time t depends on that at time $t - 1$ (sometimes referred to as ‘lag 1 autocorrelation’):

$$l_t = \mu_t + \varphi_{Rt}(l_{t-1} - \mu_{Rt-1}) + \epsilon_t \quad (2)$$

where φ_{Rt} is the autoregressive parameter and the autoregressive error distribution ϵ_t is normally distributed, which is to say

$$\epsilon_t \sim N\left(0, \sqrt{1 - \varphi_{Rt}^2} \sigma_{Rt}\right). \quad (3)$$

Figure 15 shows the autoregressive parameter for 100 sites for each month. There is a significant variation in the parameter values within a given month. It is possible to accommodate these differences when single sites are generated independently, but it is difficult to preserve this feature in a multisite setting (Rasmussen 2013). For this reason, only the average value is used for each month. The result of this assumption is a loss of variability at the daily timescale, which affects wet-dry patterns and wet and dry spell durations.

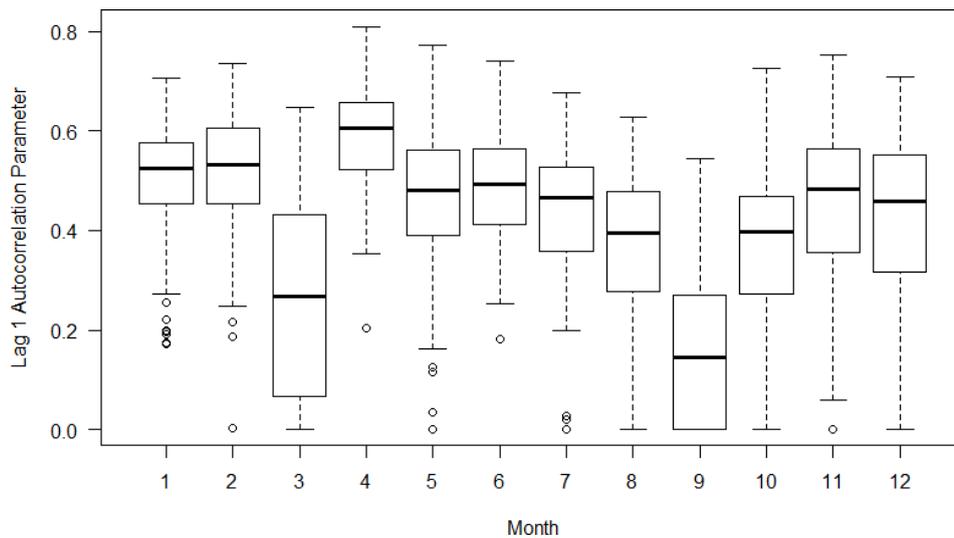


Figure 15 Distribution of lag 1 daily autocorrelation parameter across rainfall sites for each month

A simulation of the single-site daily rainfall model is illustrated in Figure 16 for 100 replicates aggregated to monthly totals and compared to the corresponding observations (red symbol).

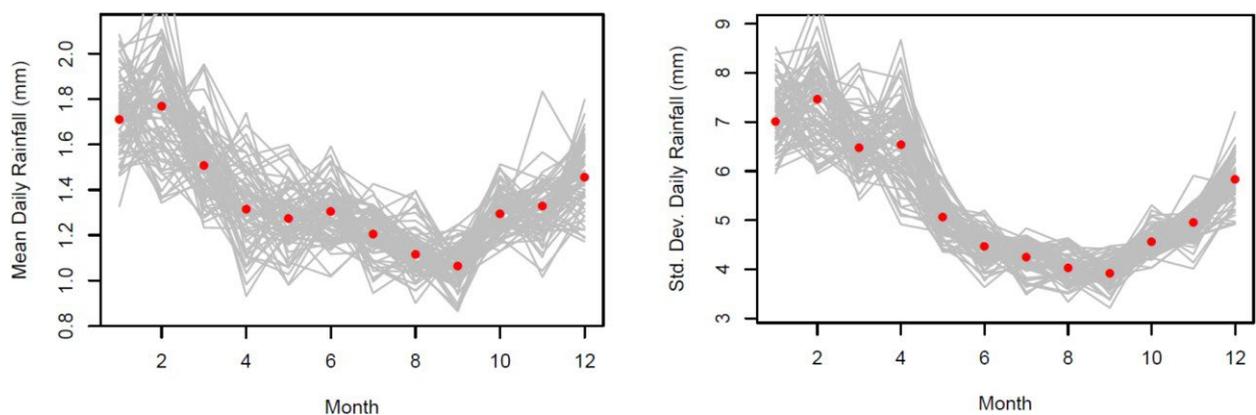


Figure 16 Distribution of rainfall; observed (red symbol) versus 100 simulated replicates (grey lines) monthly means (left) and monthly standard deviations (right)

2.2.2 At-site evaporation model

Whereas rainfall is highly variable, has a skewed distribution and complex wet-dry pattern, evaporation is a continuous variable (no zeros) and is significantly less skewed. Figure 17 shows a box plot summary of daily evaporation for each month at a representative site. While the seasonal variation is obvious, a subtle feature of these data is that summer evaporation is skewed, as suggested by the outliers. This is explored in more detail below.

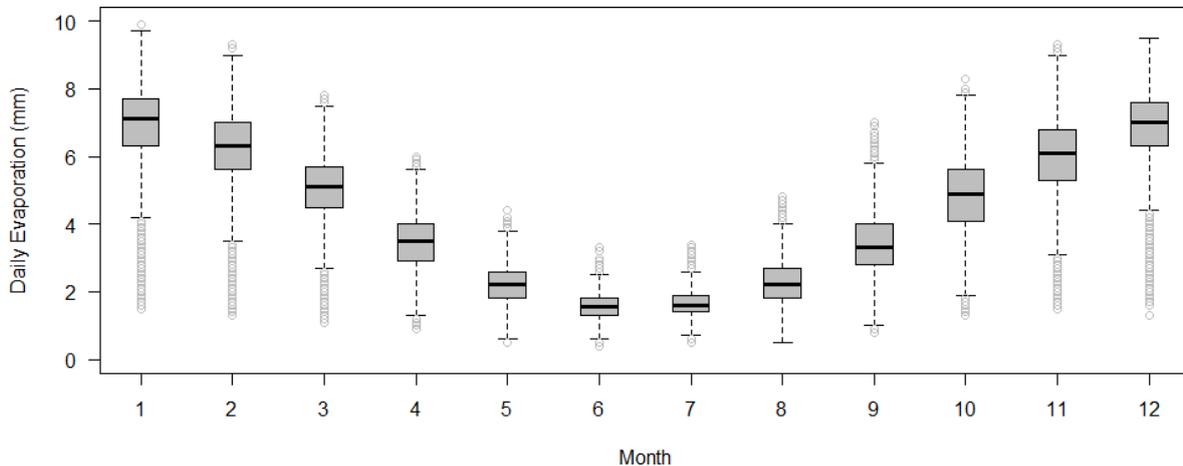


Figure 17 Distribution of daily evaporation for a representative site for each month

Figure 18 depicts the method used to generate the skewed distribution of daily evaporation, especially for summer months. A split normal distribution is used in which the half above the mode has a different standard deviation from the half below the mode.

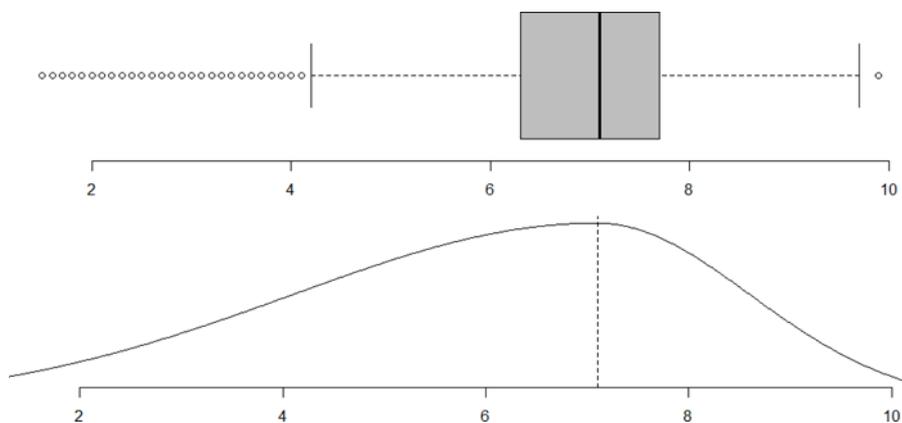


Figure 18 Schematic of method used to reproduce skewness in distribution of daily evaporation

Given the strong seasonal signal and the smooth transition in evaporation between seasons, a sinusoidal method was used to model daily evaporation (whereas the rainfall model has sets of 12 parameters to represent each month). Figure 19 (top) shows the distribution of daily evaporation for all years from a representative site along with a mean sinusoid trend fitted to the data. Subtracting this trend from the data yields residuals with zero mean, but with evidence of seasonality in the variation.

The sinusoidal equations for the mean evaporation, $\mu_{E,t}$ on a given day, t , is specified by the linear regression:

$$\mu_{E,t} = \theta_1 + \theta_2 \cos(2\pi t/365.25) + \theta_3 \sin(2\pi t/365.25) \quad (4)$$

where the θ_i are the regression coefficients and t is the day. After the mean model is fitted, the residuals of the evaporation are calculated as

$$E'_t = E_t - \mu_{E,t} \quad (5)$$

where the fitted mean is subtracted away from the observed evaporation, E_t , to give the residuals, E' .

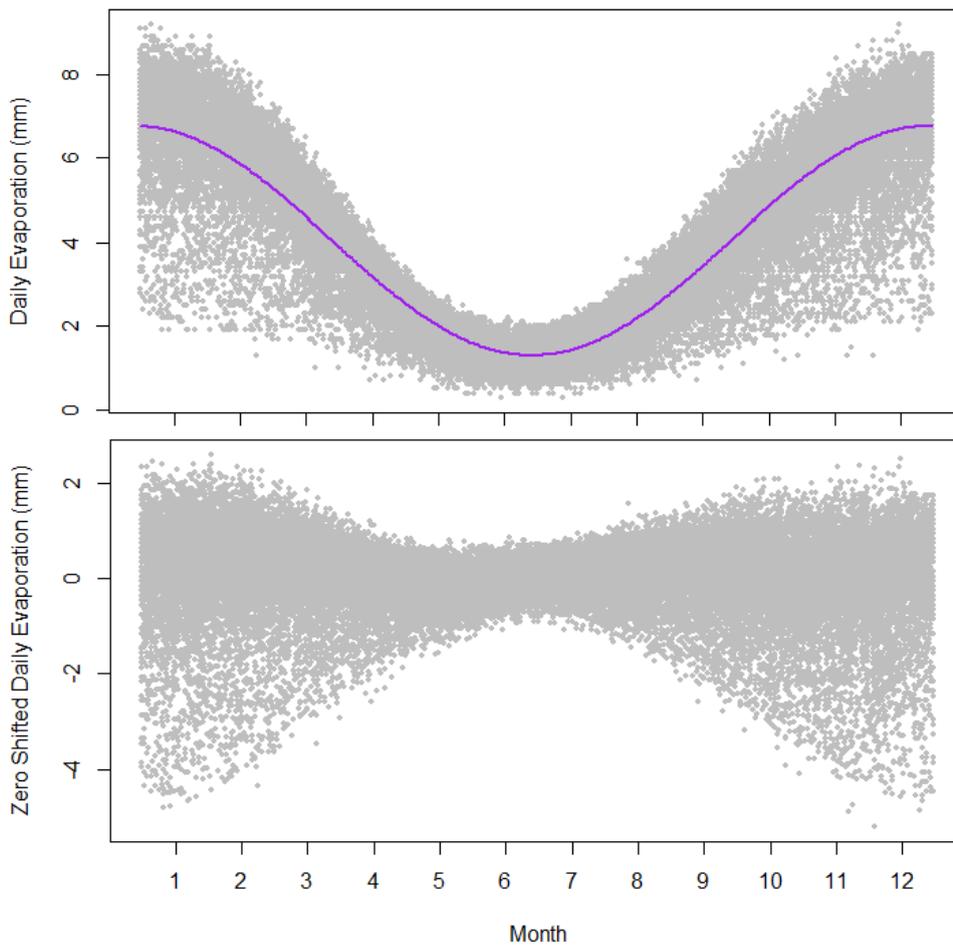


Figure 19 Daily evaporation for a representative site (top) fitted mean trend, (bottom) residuals after removing the mean

Sinusoidal regression equations can be fitted to the positive and negative aspects of the residuals (Figure 20):

$$\sigma_{E,t}^+ = \theta_4 + \theta_5 \cos(2\pi t/365.25) + \theta_6 \sin(2\pi t/365.25) \quad (6)$$

$$\sigma_{E,t}^- = \theta_7 + \theta_8 \cos(2\pi t/365.25) + \theta_9 \sin(2\pi t/365.25) \quad (7)$$

where $\sigma_{E,t}^+$ is the standard deviation parameter of the positive residuals for a given day and $\sigma_{E,t}^-$ is the counterpart for the negative residuals.

The standardised residuals, E'' , are obtained by dividing the evaporation residuals, E' , by the relevant standard deviation parameter for a given day:

$$E''_t = \begin{cases} E'_t / \sigma_{E,t}^+ & E'_t > 0 \\ E'_t / \sigma_{E,t}^- & E'_t \leq 0 \end{cases} \quad (8)$$

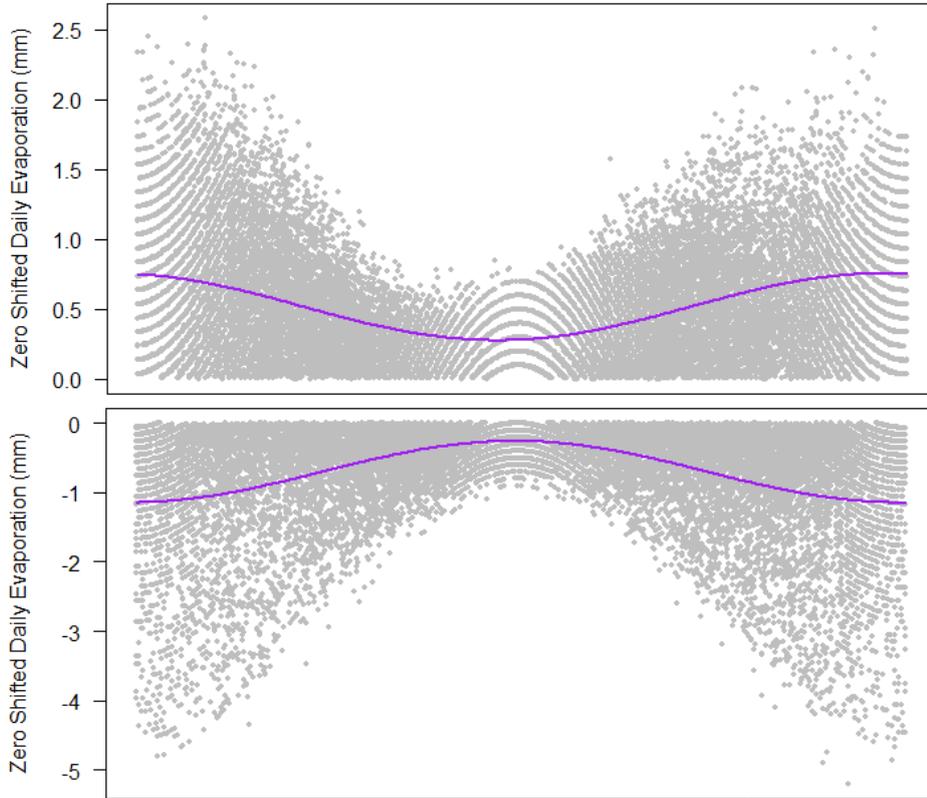


Figure 20 Fitted trends to mean-corrected residuals (top) positive and (bottom) negative

An example of the standardised residuals is shown in Figure 21. The overall spread of the distribution is approximately normal, but there are some noticeable artefacts. The distribution is not identically distributed throughout the year, because the sinusoid is not a perfect fit for the seasonality. There is a 'fingerprint' stippling effect in the middle of the year, due to (i) the sinusoidal transformation (causing the curve) and (ii) the digitised input (values are rounded to only 1 decimal place), along with the small range of values in winter (Figure 19) causing the transformed values to appear quantised. None of these artefacts is significant because the normal distribution is a good approximation to the standardised residuals.

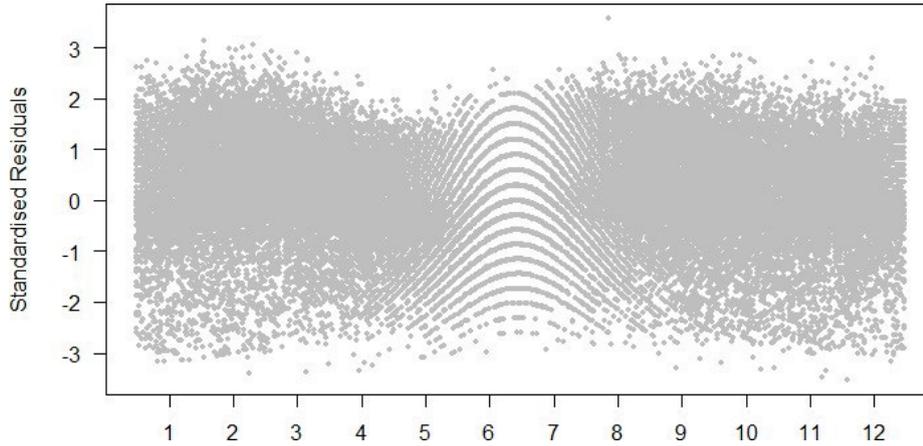


Figure 21 Standardised residuals of evaporation

After the standardised residuals are obtained, it is necessary to determine the autocorrelation structure of the residuals. Figure 22 shows the autocorrelation for a representative site. An AR(1) autoregressive model is used to represent the temporal correlation structure of evaporation:

$$E_t'' = 0 + \varphi_{Et}(E_{t-1}'' - 0) + \epsilon_t \quad (9)$$

where the mean of the residuals is 0, φ_{Et} is the autoregressive parameter and the autoregressive error distribution ϵ_t is normally distributed – that is

$$\epsilon_t \sim N\left(0, \sqrt{1 - \varphi_{Et}^2}\right). \quad (10)$$

The autoregressive model can reproduce the main correlations in the first few lags, but it does not reproduce the low levels of correlation present up to lag 40.

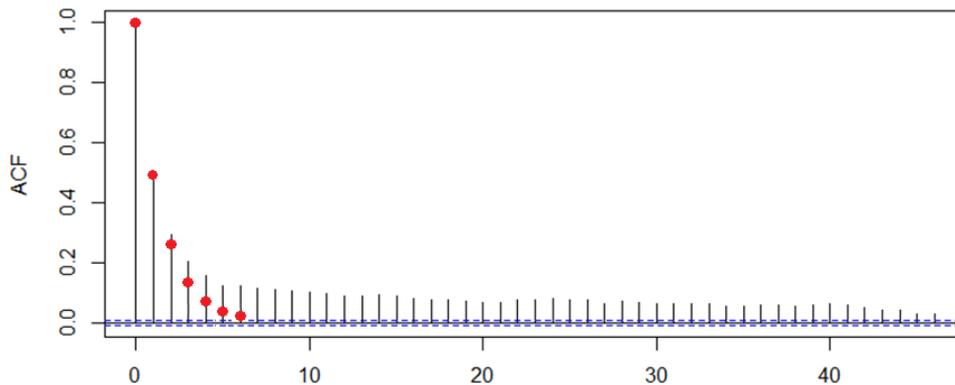
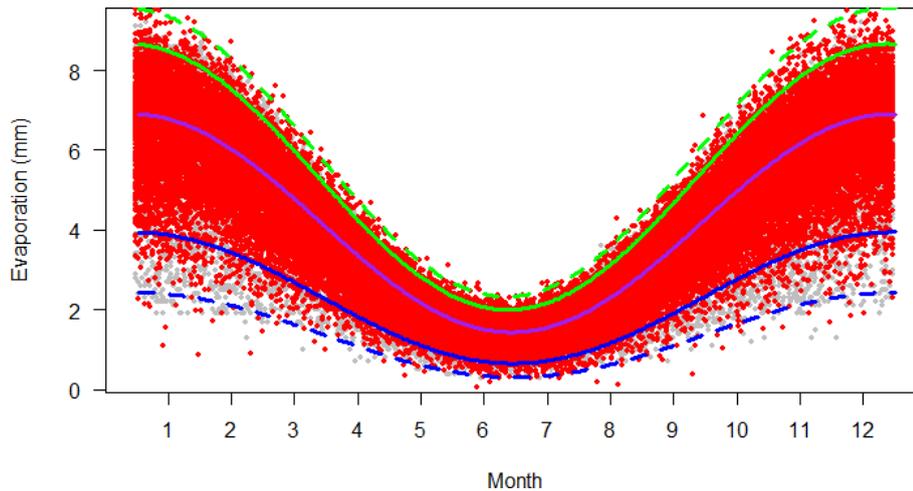


Figure 22 Autocorrelation function (ACF) of standardised residuals for a representative site, horizontal axis represents daily lags

A simulation of 100 replicates of the single-site evaporation model is shown at the daily scale in Figure 23 and at the monthly scale in Figure 24, which gives monthly mean and monthly standard deviation. The dashed lines show the 99.7% confidence intervals of the simulations, which agree with the underlying distribution of observations (grey symbols). Even with the split normal distribution, there is a noticeable discrepancy in the lower tail of the summer months (observed grey values can be seen to be below the red simulated values). This is caused by a limitation of the sinusoid linear regression: an explanatory variable with a sharper transition than a sinusoid would be required to allow for higher variability in the summer months.



Grey symbols are observations, red symbols are simulation. Solid purple shows the simulated mean, solid blue/green show the lower/upper 95% interval and dashed blue/green show the lower/upper 99.7% interval

Figure 23 Example simulation from single-site model of evaporation

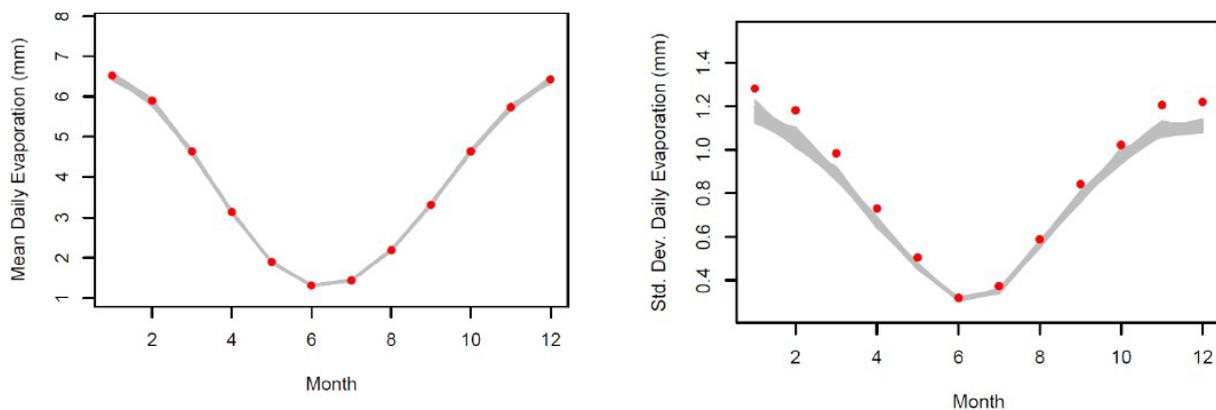


Figure 24 Distribution of rainfall. Observed (red symbol) versus 100 simulated replicates (grey lines) monthly means (left) and monthly standard deviations (right)

2.2.3 Multisite rainfall model

Developing the multisite model from the single-site model requires the spatial cross-correlation between sites. An example of the correlation with distance is shown for 100 pairs of rainfall sites for a selected month (all the months look similar). There is a noticeable scatter in the data due to variation within the region, but it is nonetheless very high (for example, even at 250 km separation, the correlation is 0.6) (Figure 25). Because the model is a multisite model, it is possible to fit the sample correlations exactly. In other words, because it is not required to infill or interpolate between the gauges, there is no need to fit a smooth correlation function to the data (which would reduce the modelled variability). Estimating the pairwise correlation for all sites yields the correlation matrix shown in Figure 26. The main point of interest is the blue stripes that persist in a couple of the rows and columns, indicating these sites are weakly correlated to the rest of the region (for example, sites R10, R12, R13 and R22). Given there are 100 sites, the lower correlation at these sites is not likely to have a big impact on the overall catchment rainfall.

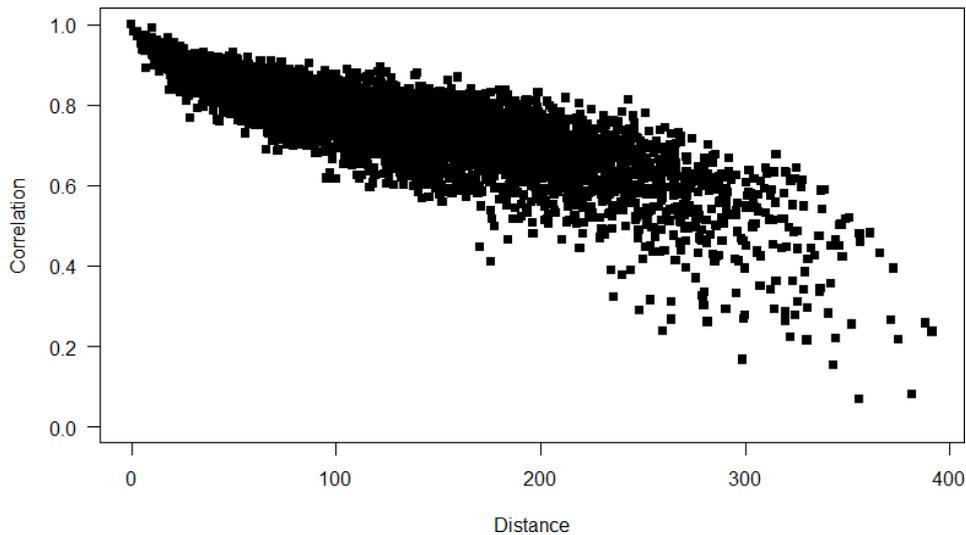


Figure 25 Rainfall sample correlation values with distance (km) for a representative site and month

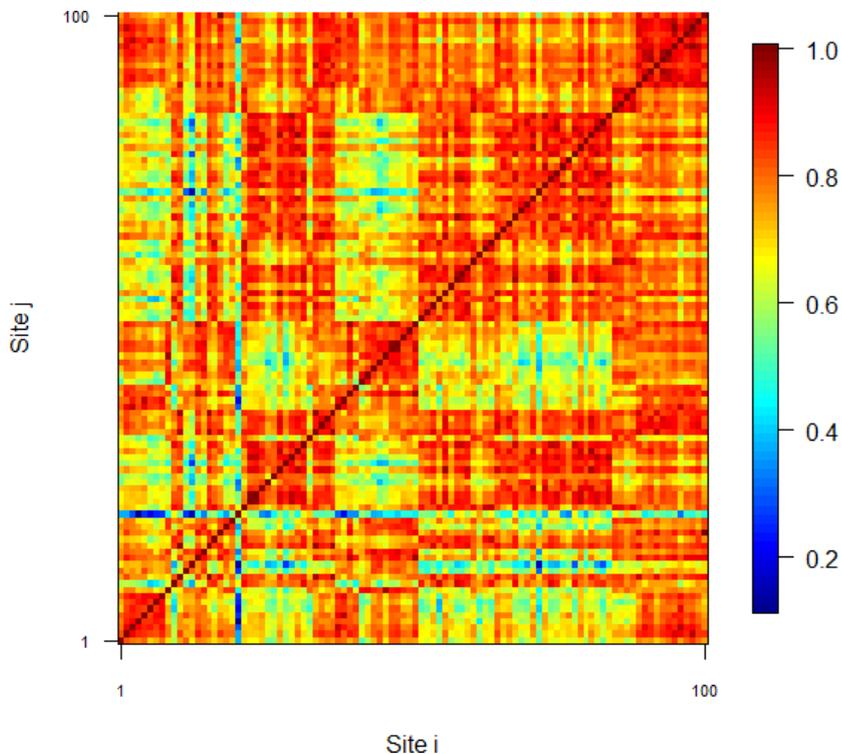
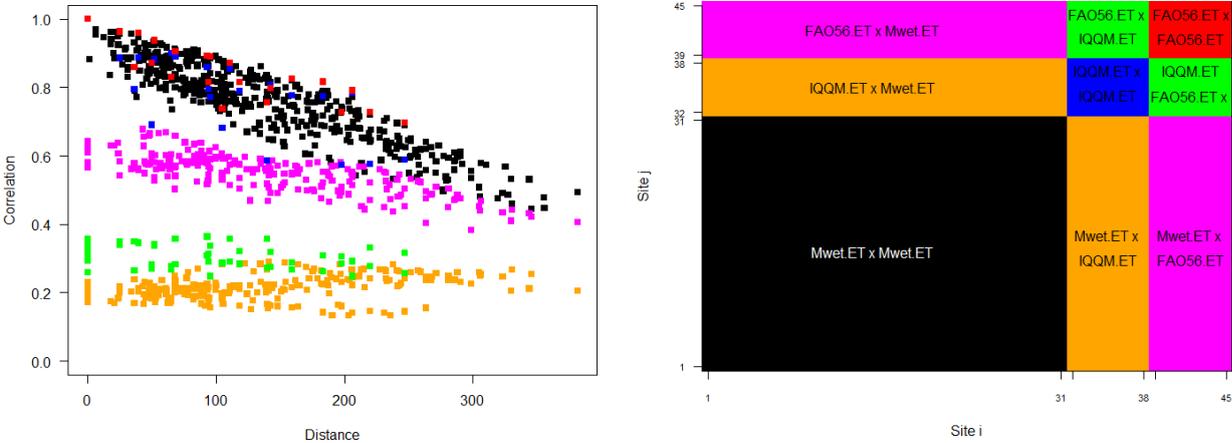


Figure 26 Rain-rain correlation matrix for 100 sites

2.2.4 Multisite evaporation model

The correlation structure of the evaporation data is more complicated because of the 3 different types. As with the rainfall, it is not necessary to fit a smoothed correlation function to the data because a multisite model is being used (rather than a model continuous in space). An example of the evaporation correlations is shown in Figure 27. A colour-coded matrix is provided in the right panel of Figure 27 to show the pairings of the various types. As with rainfall, there are significant correlations with distance, and the correlation at 250 km separation is approximately 0.6 when each type (Mwet – black, IQQM – blue, FAO56 – red) is compared to itself. When there is a cross-correlation between 2 different types of evaporation, the correlation is lower (for example, FAO56

with Mwet – magenta). The IQQM evaporations have the lowest correlation to other data (Mwet – orange, FAO56 – green), suggesting these data are not strongly related to the other evaporation data types for the region. This can also be seen in the correlation matrix (Figure 28).



Banding of the correlation is evident due pairing of the different types of evaporation: Mwet–Mwet (black); Mwet–IQQM (orange); Mwet–FAO56 (magenta); IQQM–IQQM (blue); IQQM–FAO56 (green) and FAO56–FAO56 (red). The right image gives a key for the different pairings of evaporation type in the structure of the correlation matrix.

Figure 27 Structure of evaporation correlations (left) shows sample correlation with distance for a representative site for different evaporation types

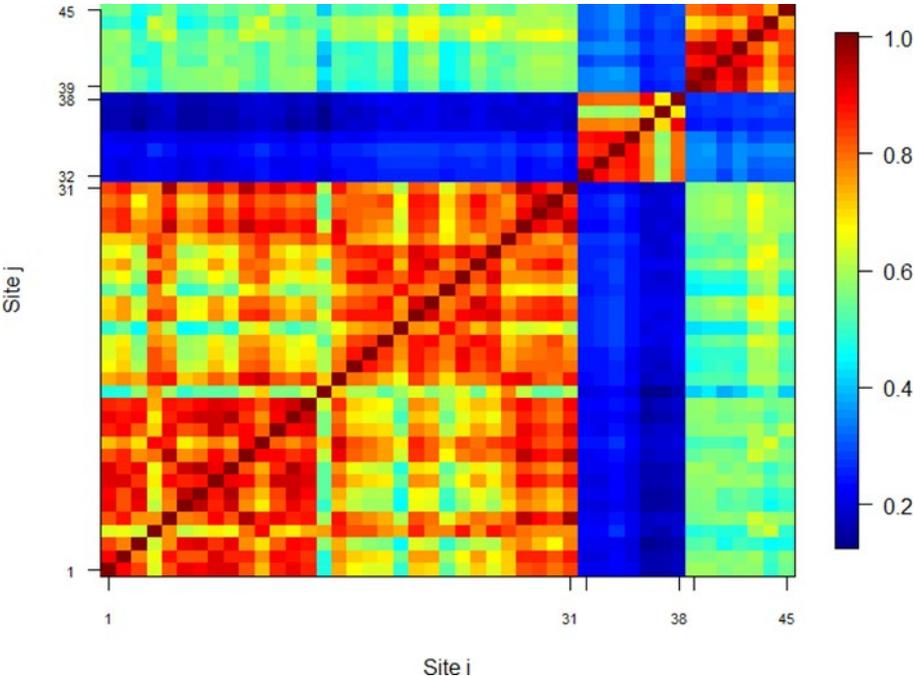


Figure 28 Evaporation–evaporation correlation matrix for 45 sites

2.2.5 Joint simulation of rainfall and evaporation

Using the multisite specifications in the prior sections, it is possible to simulate multisite rainfall at 100 sites, or multisite evaporation at 45 sites, but not (yet) possible to jointly simulate rainfall and evaporation at 145 sites. To achieve this, the correlation between rainfall and evaporation must be accounted for. Figure 29 shows a representative analysis of rain–evaporation correlation with distance. The correlation is negative and typically in the range -0.2 to -0.3 , even for distances 250 km apart. As with Figure 27, there is a small degree of striping, which is due to the different evaporation types.

An exponential correlation function is fitted (red line) to the rainfall–evaporation cross-correlations

$$\text{COR}(d) = C \exp(-d/\alpha) \quad (11)$$

where $\text{COR}(d)$ is the correlation, d is the distance between a pair of locations, α is the range parameter and C is the scale parameter. A model is fitted to the cross-correlations to ensure the robustness of the overall correlation model (when using sample correlations there can be issues with the positive definite structure of the correlation matrix). Using a fitted model introduces some smoothness into this element of the correlations, but it should not be significant because the correlations are relatively low and because they are relatively consistent (from -0.2 to -0.3).

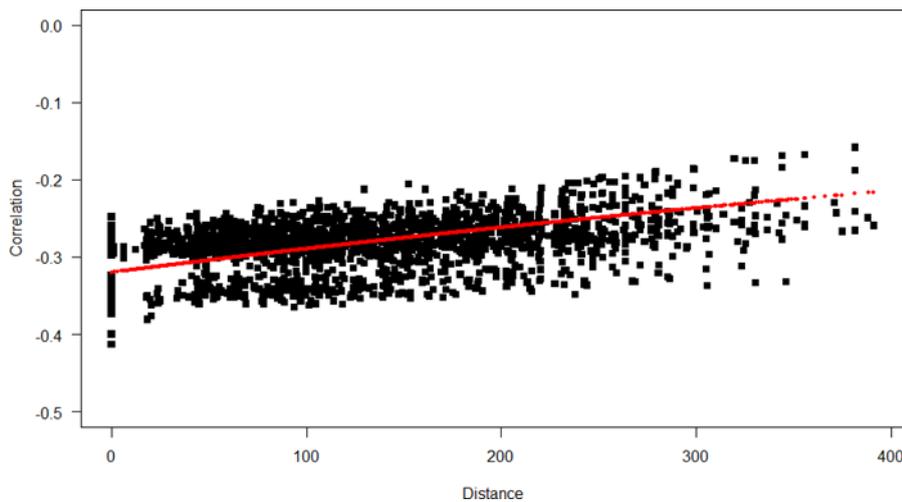
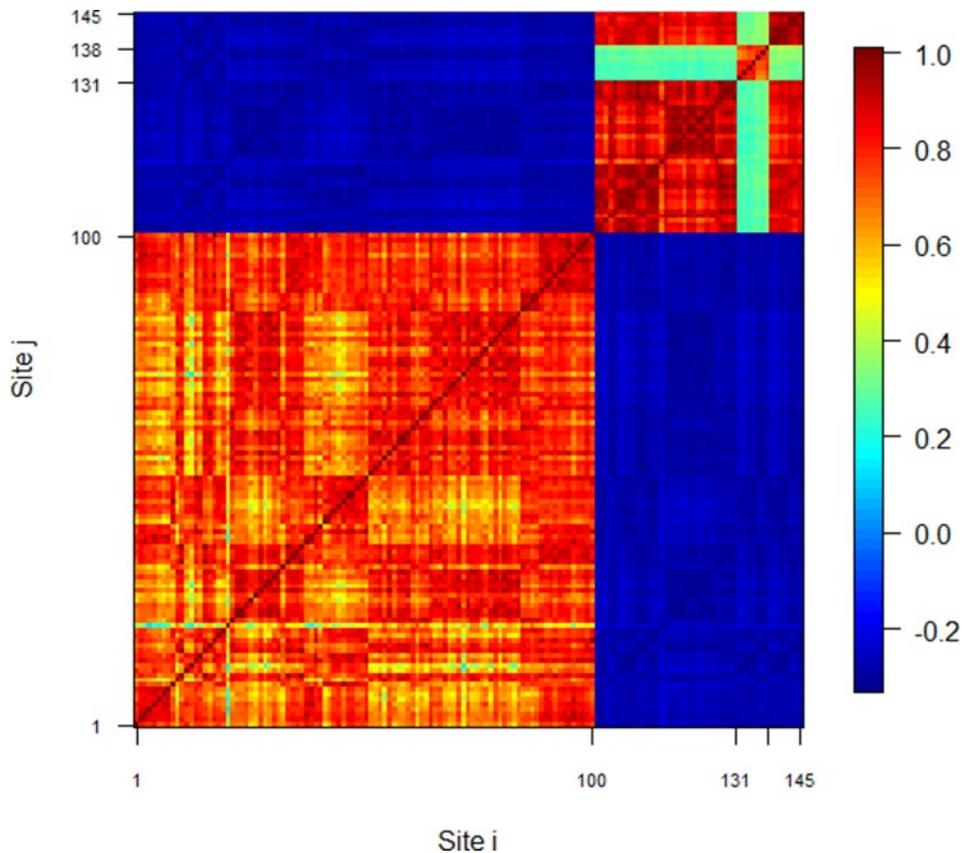


Figure 29 Rainfall–evaporation sample correlation values with distance (km) for a representative site and month, observations (black) and fitted correlation function (red)

Figure 30 shows a jointly populated rainfall–evaporation correlation matrix in which the diagonal terms reproduce the existing rain–rain (Figure 27) and evaporation–evaporation (Figure 28) correlation matrices. The off-diagonal elements are based on correlations from Figure 29.

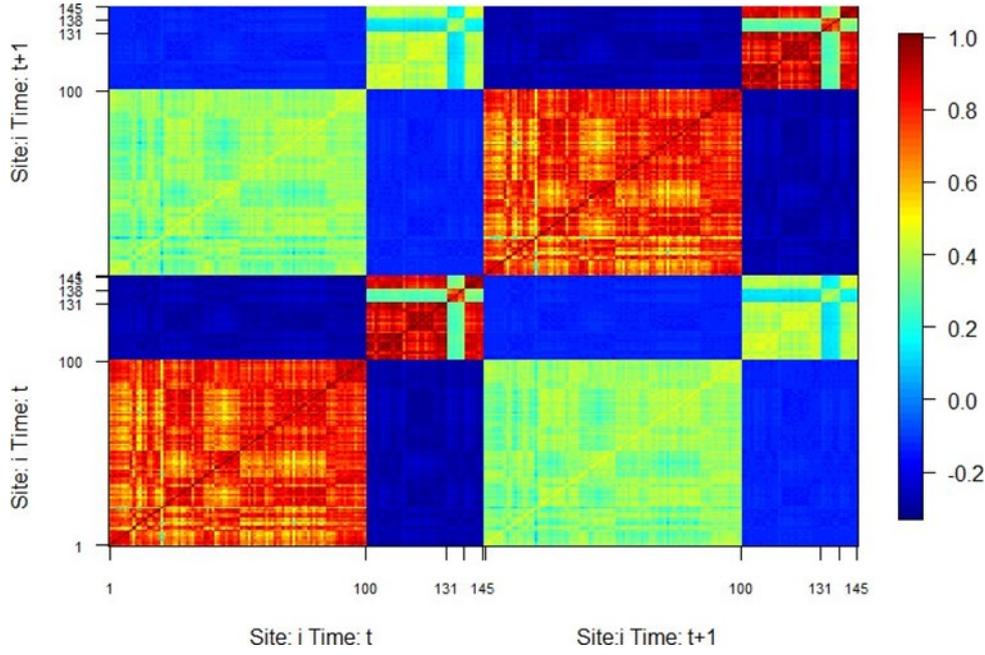
Figure 31 shows the extension of the spatial rainfall–evaporation correlation model (Figure 29) to include the lag-1 correlation parameter. The diagonal terms give the lag-0 spatial correlations for timesteps t and $t + 1$, respectively. The off-diagonal terms give the lag-1 space–time correlations (for example, site i at time t with respect to site j at time $t + 1$). The space–time correlations are obtained by assuming a contemporaneous model structure (Rasmussen, 2013). This means that the space–time correlation can be decomposed into a spatial correlation multiplied by a single temporal correlation (for that month), which is the same at all sites. The benefit of this assumption is model stability. The limitation of this assumption is that all sites must use the same autocorrelation parameter (as noted in Figure 15, it is variable). An implication of this assumption is a reduction in

the ability to accurately represent different sites (they will all have the same autocorrelation for the latent variable). This means that statistics strongly associated with autocorrelation at the daily scale may lack variability (for example, length of daily wet and dry spells). The correlation matrix in Figure 31 represents a full specification of the correlation structure needed to simulate the model across all sites and multiple timesteps for a given calibration period.



Rain–rain correlation is the bottom-left block (Figure 26); evaporation–evaporation is the top-right block (Figure 28) and the off-diagonal blocks showing negative correlations are obtained from the rain–evaporation function (Figure 29)

Figure 30 Nested block correlation matrix of a joint simulation of 100 rain sites and 45 evaporation sites



Lower-left block – time (t, t) – and upper-right block – $(t + 1, t + 1)$ – are nested matrices from Figure 30 of static spatial correlations; the off-diagonal blocks represent the lag-1 in time $(t, t + 1)$ cross-correlations between pairs of sites. The off-diagonals are rescaled from the spatial correlations by a constant autocorrelation parameter across all sites (space–time separable covariance structure).

Figure 31 Nested block correlation matrix joint simulation of 145 sites and 2 timesteps; t and $t + 1$

2.3 Extended model specification – climatic variability and future climate

2.3.1 IPO dwell-time distribution from instrumental and paleoclimatic record

The IPO is modelled as being in 1 of 2 states – negative and positive. Let $\mathbf{D}^- = (D_1^-, D_2^-, \dots)$ denote the dwelling time in the negative IPO state and $\mathbf{D}^+ = (D_1^+, D_2^+, \dots)$ denote the dwelling time in the positive IPO state, in years. The IPO is modelled as an alternating renewal process in which the system starts in an arbitrary state (for example, the positive state) and persists in that state for a duration D_1^+ before transitioning to the negative state for a period D_1^- , and then back to positive for the duration D_2^+ , and so on. The oscillation continues for a number of cycles, nc , until the total duration equals the length of the simulation; that is, $\sum_{j=1}^{nc} (D_j^+ + D_j^-) = T/365$, where T is total simulation length in days. All dwelling times are considered to be independent and identically distributed, and to come from a gamma distribution, $f(x)$ (Figure 4 in Henley et al. 2011), defined as

$$f(x) = \frac{1}{s^a \Gamma(a)} x^{a-1} \exp\left(-\frac{x}{s}\right) \quad (12)$$

where a is the shape parameter, s is the scale parameter and $\Gamma(a)$ is the gamma function. For the j th year, the dwelling time, whether in the positive or negative state, is independently sampled from the gamma distribution

$$D_j^\mp \sim \text{Gamma}(a, s) \quad (13)$$

where the parameters are related to the distribution mean, m , and standard deviation, s , by $a = (m/s)^2$ and $s = (m/s)^2/m$.

For the instrumental IPO record, the dwelling time distribution for IPO phases has the properties, $m = 17$ and $s = 8$, and for the paleoclimatic IPO record, $m = 15$ and $s = 10$ (Henley et al. 2011). An example simulation is shown for 1,000 years, based on parameters from the IPO instrumental record. A state persisting for 58 years can be seen between years 88 and 144 (Figure 32).

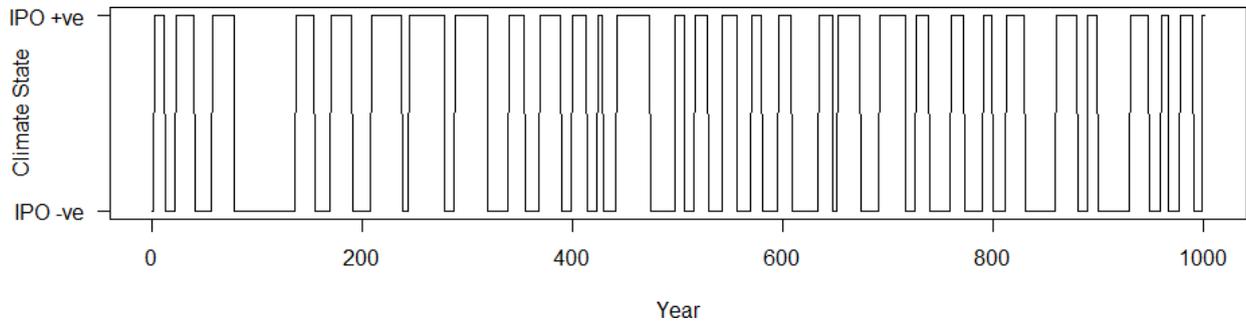


Figure 32 Example simulation of 1,000 years of IPO states from the alternating renewal model with parameters from the IPO instrumental record

2.4 Model calibration

Details of the calibration procedure are provided in Bennett et al. (2018); therefore, only an overview is provided here. At all stages, the method of moments and least squares are used to determine parameters.

2.4.1 Step 1 Rainfall: single-site rainfall distribution

The data for each site are partitioned monthly, where separate parameters are found for each month on record. The mean and standard deviation of rainfall amounts, as well as the proportion of dry days, are calculated. These statistics are matched to the corresponding properties of the truncated, power-transformed normal distribution from Eq. (1), having parameters μ_{RT}^i , σ_{RT}^i , and β . Thus, there are 3×12 parameters per site.

2.4.2 Step 2 Rainfall: single-site temporal correlation

The autocorrelation is calculated for each site based on the rain-day periods for a given month. To determine the parameter φ_{Rb} this statistic is transformed to have correlation equivalent to the underlying latent variable by accounting for the effects of truncation (Bennett et al. 2018). There is 1 parameter per month per site. However, the contemporaneous multisite autoregressive model requires a common autocorrelation parameter for all sites. The parameter value is averaged across all sites to yield 1 parameter per month.

2.4.3 Step 3 Rainfall: spatial correlation

The procedure for estimating spatial correlation is identical to single-site autocorrelation, except that it is calculated for the lag-0 in time cross-correlation for a pair of sites. Because the model is not continuous in space, it is not necessary to fit a correlation function. Instead, the sample correlations

for each pair of sites can be used to parameterise the correlation matrix. For $n = 100$ sites there are $n(n - 1)/2 = 4,950$ unique pairs of sites (that is, 4,950 parameters per month).

2.4.4 Step 4 Evaporation: single-site evaporation distribution

The regression function for the mean daily evaporation is fitted first to the daily observations using Eq. (4) and then subtracted from the observations to produce a residual time series, Eq. (5). The residuals are split according to positive and negative values, with a separate regression function fitted to each partition, Eq. (6) and Eq. (7). On dividing through by the fitted model, standardised residuals are obtained using Eq. (8). The remaining time series is assumed to be normally distributed. There are 3 sets of regression equations, each having 3 parameters, giving 9 parameters per site.

2.4.5 Step 5 Evaporation: single-site temporal correlation

The lag-1 autocorrelation is calculated from the standardised residual observed time series (obtained from Step 4). This parameter, φ_{Et} is set equal to the lag-1 sample autocorrelation. As with the rainfall autocorrelation, this value is averaged across all sites, yielding a contemporaneous version of the evaporation model. Thus, there is only 1 parameter per month.

2.4.6 Step 6 Evaporation: spatial correlation

The lag-0 cross-correlation for all pairs of evaporation sites is calculated. Because the model is not continuous in space, it is not necessary to fit a correlation function. Instead, the sample correlations can be used for each pair of sites to parameterise the correlation matrix. For $n = 45$ sites there are $n(n - 1)/2 = 990$ unique pairs of sites (that is, 990 parameters per month). As discussed with respect to Figure 27, the different types of evaporation have different correlation properties, but these are directly represented in the correlation matrix because each pair of sites has its own sample correlation.

2.4.7 Step 7 Rainfall–evaporation: spatial correlation

The lag-0 in time cross-correlation is calculated between each of the 100 rainfall sites and the 45 evaporation sites. The parameters of the correlation function (range α and scale C in Eq. (11)) are fitted to the cross-correlations using least squares. There are 2 parameters fitted per month. The correlation function is used instead of sample correlations for improved model stability.

2.4.8 Step 8 Space–time correlation

There are no extra steps required to fit the space–time correlations; that is, the lag-1 in time cross-correlation between sites. These correlations are assumed to follow a space–time separable structure. In other words the lag-1 cross-correlation is obtained from the lag-1 autocorrelation (Steps 2 and 5) multiplied by the lag-0 cross-correlation (Steps 3, 6 and 7).

2.4.9 Step 9 IPO calibration

Steps 1 to 8 specify the model calibration requirements for Model A – the base model. To calibrate the model for Model B, the time series are partitioned into the IPO positive and negative states

indicated in Section 2.1.3. A separate set of parameters is fitted to each partition. The parameters for the dwelling time in each IPO phase are obtained from Henley et al. (2011).

2.5 Model evaluation

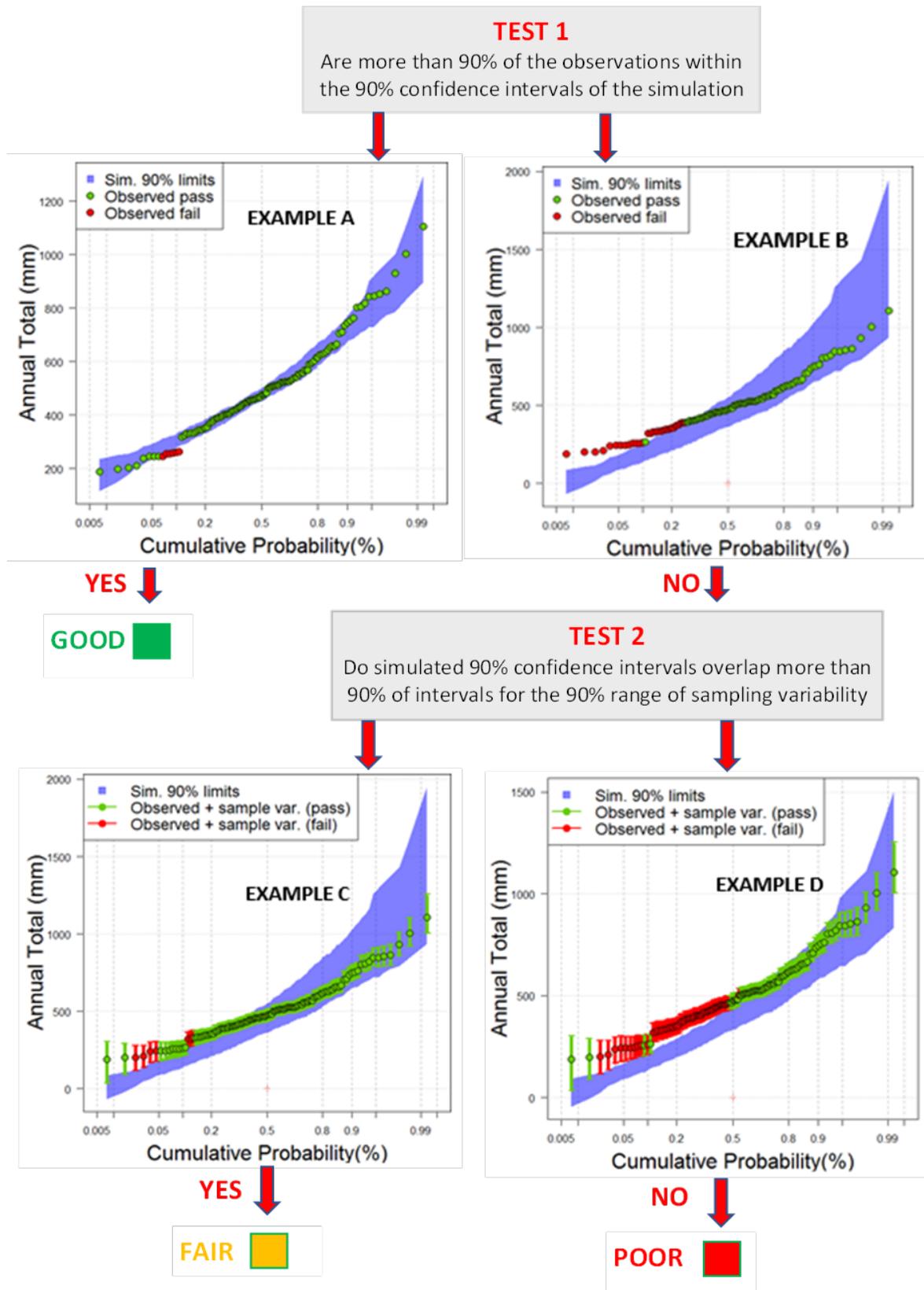
Evaluation of stochastically generated weather can involve many different aspects including a range of timescales, locations and statistics of relevance. Many evaluation methods rely on visual inspection and qualitative assessment. Bennett et al. (2018) demonstrated the benefit of using a systematic quantitative assessment, which provides consistent evaluation and the ability to pool evaluations over a larger grouping.

The general approach of Bennett et al. (2018) is used here. It uses 3 categories, Good, Fair and Poor. A variation has been made to the performance criterion for the Fair category, which had limited utility in Bennett et al. (2018) as a 'borderline' category. The following sections outline the rules of the specific tests. It is not important for the tests to represent a statistical hypothesis test, only that they can reliably differentiate between classes of performance and provide a relative measure. Performance plots of all the relevant statistics are also provided in annexes to this report for each location, which enables visual inspection of the results in conjunction with the formal model evaluation approach. Regardless of the method, interpreting the evaluation requires consideration of the relevance of the statistic to the application of interest.

2.5.1 Evaluation of distribution quantiles

A common case to evaluate is how well the quantiles of a distribution are matched between observations and simulations. Figure 33 provides a schematic illustration of 2 tests used to evaluate 'goodness of fit' of distribution quantiles and classify the fit of the entire distribution into a relevant category. The 2 tests are:

- **Test 1:** Are more than 90% of the observations within the 90% confidence intervals of the simulation?
 - If the first test is passed, a classification of Good is applied. This is shown in example A (Figure 33), in which only a few quantiles are outside the interval.
 - If test 1 is not passed, test 2 is applied. This is shown in example B, in which many quantiles are outside the interval.
- **Test 2:** When comparing the simulated 90% confidence intervals to the 90% range of sampling variability for each statistic, are more than 90% of the intervals overlapping?
 - This test is more lenient than test 1. A bootstrap method can be used to calculate the sample variability of the observed statistic. If more than 90% of quantiles overlap, a classification of Fair is applied (see example C).
 - If both tests are failed, a classification of Poor is applied (see example D).



In example A, >90% of observations are inside the simulated 90% confidence interval (Good). In example B, <90% of observations are inside the simulated confidence interval, so test 2 is used. In example C, >90% of sample statistic intervals overlap the simulated 90% confidence interval (Fair). Example D fails both tests (Poor). Examples used 129 data points.

Figure 33 Flow chart of performance classification for annual distributions into 3 categories (Good, Fair, Poor) using the criteria specified, according to 2 tests

2.5.2 Evaluation of distribution of monthly totals

Another common case is to evaluate the distribution of monthly totals, because the model uses monthly parameters for some aspects and because of the significance of the seasonal cycle. The evaluation in this report considers the mean and standard deviation of monthly totals from 129-year record lengths. Having multiple simulated replicates (for example, 77 replicates) produces a distribution of means and a distribution of standard deviations of monthly totals.

The same tests used on the quantiles (Section 2.5.1) are used on the distribution of means and distribution of standard deviation for the 12 months. The same concept of requiring a 90% match is applied, but in the context of monthly distributions there are only 12 data points, so the criterion is rounded so that 11 of the 12 months must be within the confidence interval.

Figure 34 provides a flow chart for this type of test, which is very similar to the test illustrated in Figure 33. Note that Good and Fair labels can be achieved only if no more than 1 month's observed statistic is outside the simulated limits. As in Figure 33, test 2 is more lenient because it allows for sample variability in the observations. Based on sampling properties of the normal distribution, the sample variability for the mean monthly total can be calculated analytically for the 90% limits as $X_{\mu}^{90\%} = \hat{\mu} \pm 1.64\hat{\sigma}/\sqrt{n}$ and for the 90% limits of the standard deviation of monthly totals as $X_{\sigma}^{90\%} = \hat{\sigma} \pm 1.64\hat{\sigma}/\sqrt{2n}$, where $\hat{\mu}$ is the estimated mean of observed monthly totals, $\hat{\sigma}$ is the estimated standard deviation of the observed monthly totals and n is the number of observations (here $n = 129$, because each month is observed once a year for 129 years).

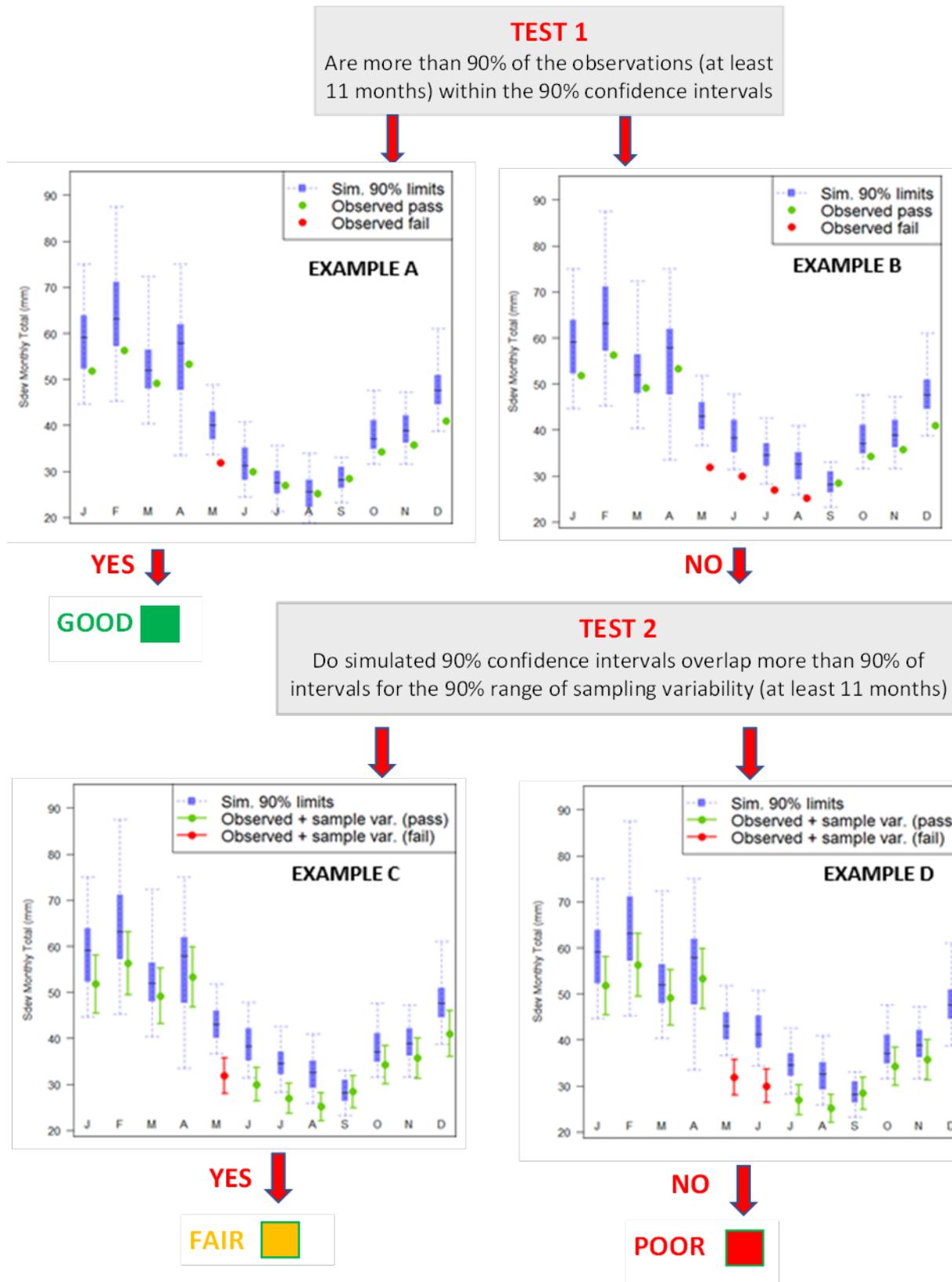
2.5.3 Pooling performance over multiple sites

Table 4 shows the rules used to pool multiple sites and determine an overall summary classification for that statistic. When more than 50% of the individual sites are labelled Good, a classification of Overall Good is applied to that statistic. A similar rule is applied to determine the classifications Overall Fair and Overall Poor. If no single category captures more than 50% of sites, the labels Overall Fair – Good, Overall Fair – Poor and Overall Variable are used, according to the rules outlined in Table 4.

Table 4 Aggregate performance categorisation criteria

Overall performance category	Definition: Sites in performance category (%)	Example: Good model performance (%)	Example: Fair model performance (%)	Example: Poor model performance (%)
Overall Good	Good is > 50%	85	10	5
Overall Fair	Fair is > 50%	5	85	10
Overall Poor	Poor is > 50%	10	5	85
Overall Fair – Good	Fair & Good is > Poor	35	55	10
Overall Fair – Poor	Fair & Poor is > Good	10	35	55
Overall Variable	Good & Poor is > Fair	35	20	45

Source: Bennett et al. (2018)



In example A, >90% of observations fall inside the simulated 90% confidence interval (Good). In example B, <90% of observations fall inside the simulated 90% confidence interval, so test 2 is invoked. In example C, >90% of sample statistic intervals overlap with the simulated 90% confidence intervals (Fair). Example D fails both tests (Poor).

Figure 33 Flow chart of performance classification of simulation into 3 categories (Good, Fair, Poor) using the criteria specified, according to 2 tests

3.1 Model evaluation summary

Performance evaluation in this section focuses on Model C, which is the most informative simulation from the historical record because it includes information based on the paleoclimate IPO. Because Model C is intended as the primary output for representing the historical record, it is evaluated in detail in this section, with discussion of performance at representative sites and reference to annex documentation for assessment of all sites.

The performance summary of Model A (base model) and Model B (instrumental IPO) is provided in Appendix B, but annex documentation for site-by-site analyses of these model variants has been omitted (it can be provided on request). The comparison shows that the overall performance across the model variants is similar.

Table 5 summarises the performance of the rainfall across the 100 sites for Model C. Specifically:

- The strength of the model can be seen in its ability to reproduce totals at monthly, annual, interannual and decadal scales. The 1-, 2-, 5- and 10-year totals are Overall Good, with only 8 sites out of 100 receiving an evaluation of Fair at one or more of these scales.
- The mean of monthly totals is matched perfectly at all sites, but the standard deviation of monthly totals is Overall Fair, and the simulations are typically more variable at the monthly scale.
- The distribution of the proportion of wet days at the annual scale is shown to be Overall Poor. This is a known limitation of the model (Bennett et al. 2018), and is due to simplifying assumptions in the temporal correlation structure. The characteristics of the performance are discussed in detail in Section 3.2.3. The succinct explanation is that the model outputs are less variable than the observations. This can be seen in Table 5 – although the mean of the monthly proportion of wet days is Overall Good (that is, the process is unbiased at most sites), the standard deviation in the monthly proportion of wet days is only Overall Fair.
- The annual maximums are Overall Good for 1-day maximums and Overall Fair for multiday accumulations. Section 3.2.5 will demonstrate that despite the Fair performance at the majority of sites, the simulations do not deviate far from the observations. The Overall Fair performance is a consequence of simplifications in the temporal correlation structure of the model (Section 2.2.1).

Table 5 Rainfall evaluation summary of performance, Model C, paleoclimatic IPO model variant, 117 rainfall sites 129 years length, 77 replicates

Statistic	Overall performance category*	Good model performance (%)	Fair model performance (%)	Poor model performance (%)
Distribution of annual total rainfall	Good	99	1	0
Distribution of 2-year rainfall totals	Good	99	1	0
Distribution of 5-year rainfall totals	Good	98	1	1
Distribution of 10-year rainfall totals	Good	97	3	0
Mean of monthly rainfall totals	Good	100	0	0
Standard deviation of monthly rain totals	Fair	21	79	0
Distribution of annual proportion of wet days	Poor	4	21	75
Mean of monthly proportion of wet days	Good	97	3	0
Standard deviation of monthly proportion of wet days	Fair	26	70	4
Annual 1-day rainfall maximum distribution	Good	51	49	0
Annual 2-day rainfall maximum distribution	Fair	4	90	6
Annual 3-day rainfall maximum distribution	Fair	5	92	3

*See Table 4 for categorisation criteria.

Table 6 provides the performance summary for the evaporation simulations, which is based on 45 sites. Similar to the rainfall evaluation, the performance is Overall Good for the totals at the annual and interannual scales. At the monthly scale, the characteristics of the model are similar to the rainfall, in that the modelling of the mean of monthly evaporation totals is Overall Good but the standard deviation of monthly totals is Overall Fair. The sites with Poor performance in the standard deviation of monthly totals are the IQQM evaporation sites, where the mismatch in performance occurs due to an artefact of the observed data (discussed in Section 3.2.7).

Table 6 Evaporation evaluation summary of performance, Model C, paleoclimatic IPO model variant, 79 evaporation sites 129 years length, 77 replicates

Statistic	Overall performance category*	Good model performance (%)	Fair model performance (%)	Poor model performance (%)
Distribution of annual total evaporation	Good	90	9	1
Distribution of 2-year evaporation totals	Good	91	9	0
Distribution of 5-year evaporation totals	Good	90	10	0
Distribution of 10-year evaporation totals	Good	86	11	3
Mean of monthly evaporation totals	Good	100	0	0
Standard deviation monthly evaporation totals	Good	90	0	10

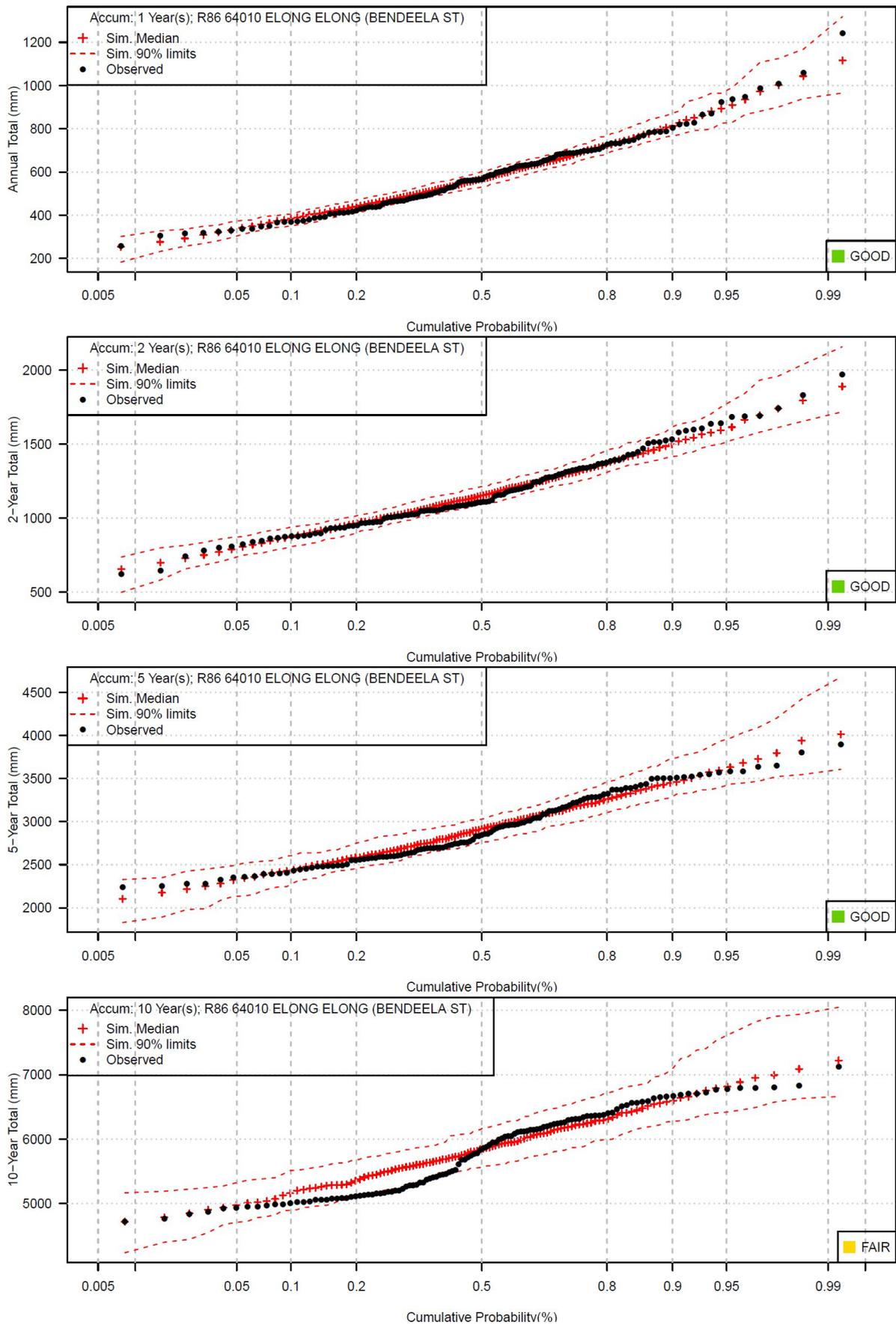
*See Table 4 for categorisation criteria.

3.2 Detailed evaluation of distributions

3.2.1 Multiyear annual rainfall totals (1, 2, 5 and 10 years)

From Table 5, the quality of the modelling of annual rainfall totals at 1 or more years is Overall Good, with almost all sites showing Good performance. Statistics at the annual, multiyear and decadal scales must be well-modelled for adequate drought risk assessment. The Overall Good performance of these statistics indicates that the seasonal component and climatic components of the model are functioning with appropriate levels of variability.

Figure 35 shows an example using site R86, selected because it has an evaluation of both Good (at the 1-, 2- and 10-year scales) and Fair (at the 5-year scale). The Good evaluation is typical of the performance of this statistic and shows that the mean and variability of the simulations match the observations well. Assessing the Fair performance, it is clear that the mean and variability of the simulations is acceptable, with the simulations not deviating far from the observations. From Figure 35, by comparing the lowest total in each panel to the lower 90% limit, the simulations are able to generate rainfall totals below the lowest values in the observation record.



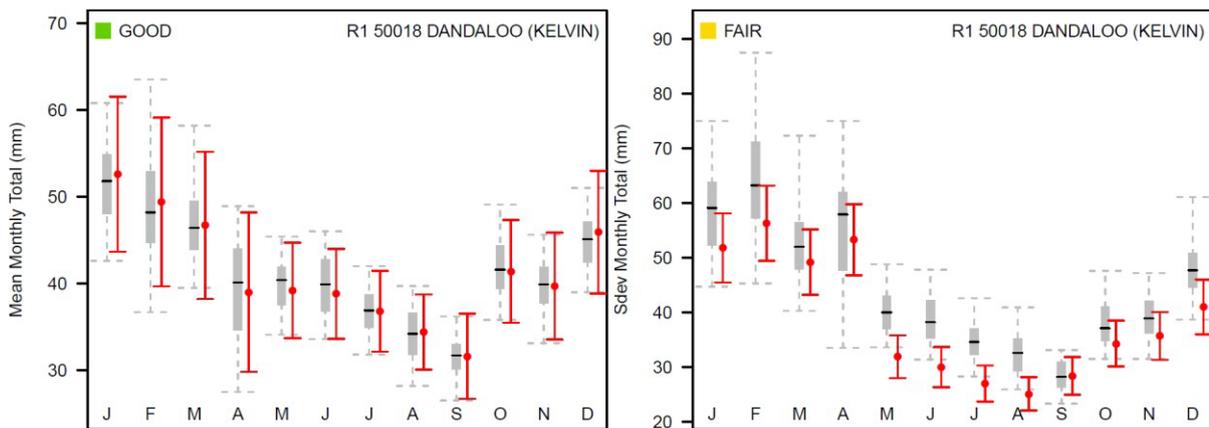
Simulated rainfall and its 90% confidence interval. Performance classifications shown.

Figure 35 Distribution of annual rainfall totals for 1-, 2-, 5- and 10-year totals for a representative location

3.2.2 Mean and standard deviation of monthly rainfall totals

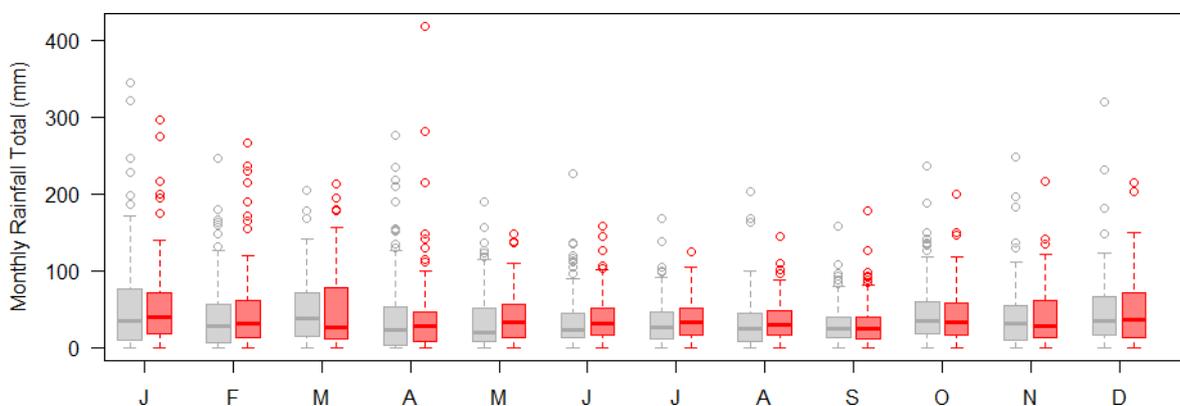
From Table 5, the quality of the modelling of the means of the monthly totals is Overall Good, and that of the standard deviations of the monthly totals is Overall Fair. Figure 36 provides an example for site R1, which is typical for the majority of sites. No site exhibits Poor performance. The model reproduces mean values well, which is expected from this model because each site has its own mean parameter. Although each site has a standard deviation parameter, the standard deviation of monthly totals is also affected by the autocorrelation process at daily and monthly scales.

Figure 36 (left panel) shows that the means of the monthly totals are Good. The right panel shows that the 90% intervals for the standard deviations from the simulations are higher than the observations for the months May to August (leading to the classification of Fair). To help with interpreting this difference, Figure 37 provides the full distribution of monthly totals for one replicate at this site. The interquartile range of the boxes is very similar for all months, but the upper tail of the distribution (the whisker and outliers) is typically heavier in the simulation (grey boxes). A practical interpretation is that the simulation has occasional very wet months that are wetter than the observations suggest are likely. Because the distribution of monthly totals is skewed, standard deviations in Figure 36 are sensitive to the behaviour of the tail.



Observed in red, simulated in grey. Whiskers of grey box plot extend to the 90% interval. Performance classifications shown.

Figure 36 Distribution of (left) means of monthly totals; (right) standard deviations of monthly totals



Grey monthly boxplot has 129 years of simulated data from one replicate and red monthly boxplot has 129 years of observations.

Figure 37 Paired comparison of full distribution of monthly totals for one representative site, R1 50018, and one replicate

3.2.3 Annual wet day proportion

From Table 5, the distribution of the proportion of wet days at the annual scale is Overall Poor. Figure 38 provides example for 4 representative sites: R1, R2, R3 and R4. In all examples, the distributions are unbiased, but the variability of the simulations is typically less than that of the observations.

To illustrate this performance, Table 7 compares the observed number of wet days for the driest year on record to the equivalent value from simulated replicates. This comparison corresponds to the lowest observation in each plot from Figure 38 relative to the simulated values at that quantile (where the number of wet days is obtained from the proportion by multiplication by 365.25). Interpreting Table 7, for the driest year, the observed numbers of wet days at the Poor sites R3 and R4 were respectively 11 and 10 fewer wet days than the median. Taking the 90% limits of the simulations as representative of the spread of the model, the observed statistic is not inside this range at the Poor sites.

Table 7 Observed versus simulation comparison for 4 representative sites

Site	Classification	Number of wet days (observed)	Number of wet days (simulated) Lower limit	Number of wet days (simulated) Median	Number of wet days (simulated) Upper limit
R1	Fair	28	26	32	38
R2	Good	38	28	37	43
R3	Poor	38	49	59	67
R4	Poor	28	29	38	45

The number of wet days in the driest year (per 129 record length). The simulation is summarised by the median and the 5% and 95% limits.

Although the overall simulations are unbiased for the median number of wet days per year (Figure 38), for the driest year on record the simulations are slightly wetter at most sites – Table 7 shows approximately 10 more wet days at Poor sites). A similar analysis of the upper tail would show that, for the wettest year on record, the simulations are slightly drier at most sites. Despite the discrepancy in this statistic, it is not necessarily a significant practical concern. The reason is that because of the threshold, a day is classified as ‘wet’ for even 0.01 mm rain. From comparisons of the annual totals (Figure 35 and Annex A) the simulations show Good reproduction of annual totals at nearly all sites and for all portions of the distribution (the driest years, the median year and the wettest years). In other words, although there is a bias in the number of wet days in the driest year, there is not a bias in the rainfall amount (because the process of rainfall amounts compensates). This could be loosely interpreted as showing that the simulations have slightly (for example, 10 days) more light-rain ‘drizzle’ in the driest year on record. For the wettest year on record there are fewer wet days in the simulation.

The discussion in this section is a good example of the strengths and weaknesses of quantitative performance metrics:

- The strengths are that the evaluation is comprehensively applied to all sites, it is consistently applied, and therefore able to highlight differences in performance (for example, R1: Fair; R2: Good; R3: Poor; R4: Poor).
- The weakness is that the practical significance of an evaluation is not in itself clear. The annual proportion of wet days receives an Overall Poor rating, but it is important to consider the implications of this performance classification for practical applications of the time series. Put another way, which statistics are most important to reproduce successfully depends on the application.

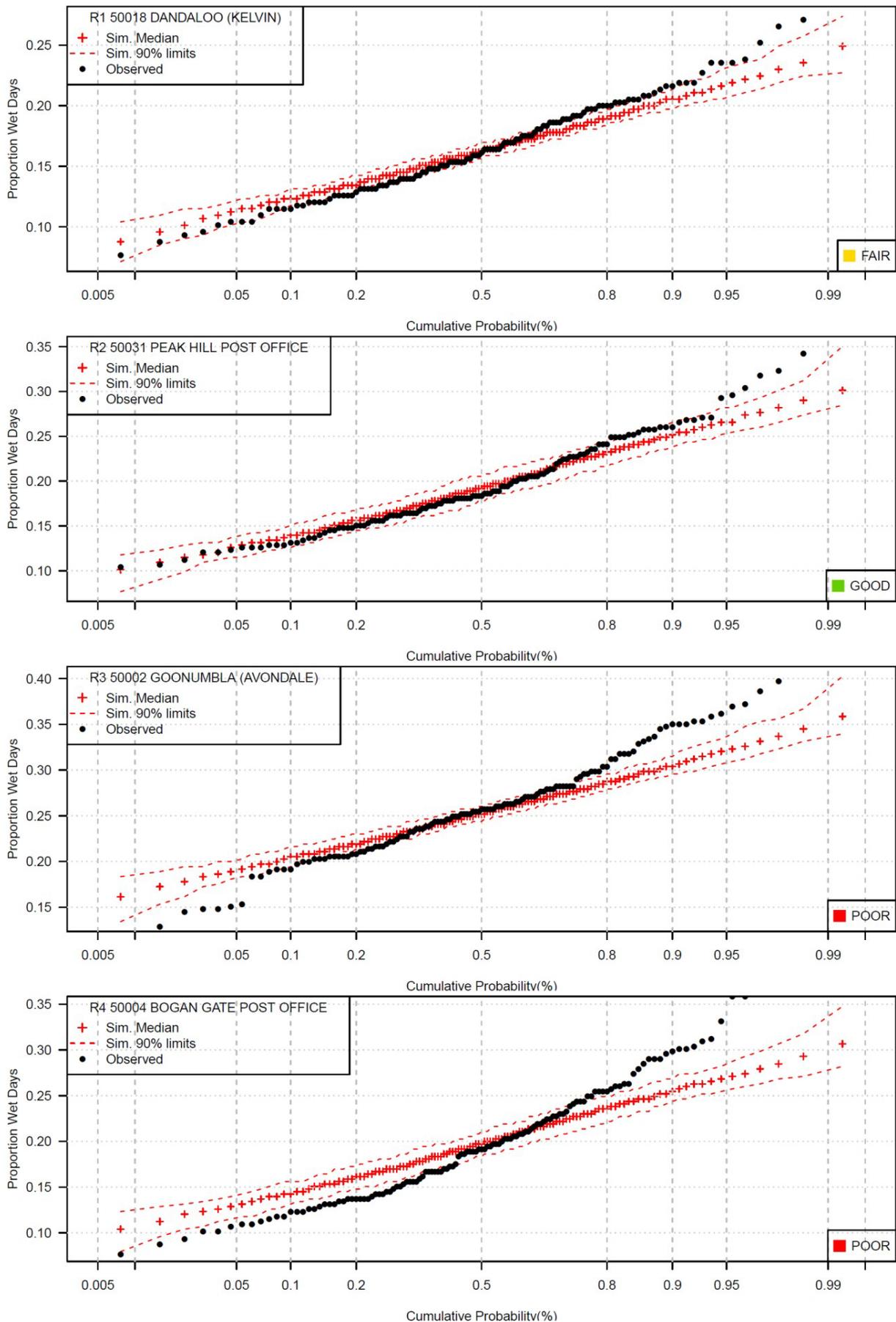
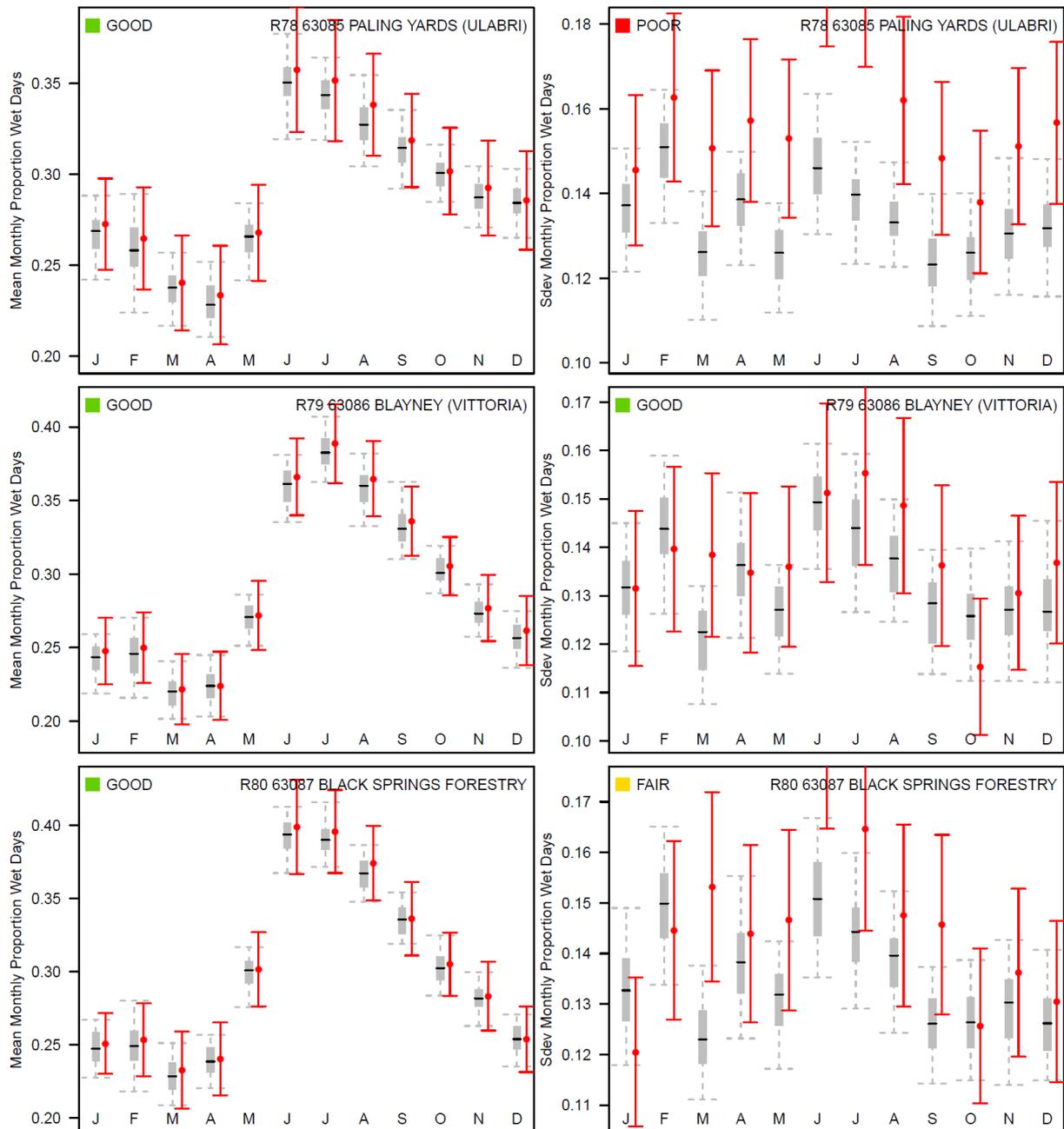


Figure 38 Distribution of the proportion of wet days within a given year for 4 representative sites with different performance outcomes (Good, Fair and Poor)

3.2.4 Monthly wet day proportion (mean and standard deviation)

The monthly summary of the proportion of wet days from Table 7 shows that the mean of this statistic is Overall Good and that the standard deviation is Overall Fair. Examples are shown in Figure 39 for 3 sites R78, R79 and R80 and it is clear that the means are consistent in the simulations for each month. The performance of the standard deviation of monthly proportions is Fair across most sites, but some sites can have multiple months that do not match well (for example, June, July at site 78 in Figure 39). Following the discussion in Section 3.2.3 the lack of variability in the proportion of wet days, while noticeable, is not necessarily a significant practical concern when compared to other statistics such as the mean and variability of monthly totals. Even though the proportion of wet days is less variable in the simulations (Figure 39), comparing back to Figure 36, it can be seen that the monthly total rainfalls are typically more variable in the simulations than in the observations. This shows that the distribution of rainfall amounts compensates for the lack of variability in number of wet days.



Whiskers of grey box plots extend to the 90% interval. Red shows the observed mean/standard deviation along with error bars showing the standard error of each observed statistic.

Figure 39 Distribution of means of monthly proportions of wet days (left) and standard deviations of monthly proportions of wet days (right) for 3 different representative sites with different performance classifications (Good, Poor, Fair)

3.2.5 Annual rainfall maximums (1, 2 and 3 day)

From Table 5, the performance of 1-day annual maximums is Overall Good and the performance of 2-day and 3-day maximums is Overall Fair.

Figure 40 provides an example for 4 representative sites R25, R26, R27, R28, where each row is a different site and the 3 columns are the 3 different accumulated 1-, 2- and 3-day maximums. Inspecting this figure shows that for the Good and Fair cases the simulated distribution of annual maximums is similar to the observations. Despite the large number of Fair sites, it is worth noting

that it is very difficult for a model to match the extremes because the calibration and parameters are related to moments of the distribution rather than the extremes. Furthermore, multiday totals are an emergent feature of temporal correlation structure in addition to the daily marginal distribution.

For 2-day maximums there are 11 Poor sites and for 3-day maximums there are 13 Poor sites. For the example of site R27 in Figure 40, the site is labelled Poor because of a discrepancy in the middle of the distribution but the upper tail has a reasonable match. For site 28, the upper tail does not match well. The limitation at these sites is mostly relevant to flood studies. This may be of concern for studies that are centred on sub-catchments near the Poor performing gauges, especially if those catchments have a response time in the order of several days. If this were the case, it would be possible to post-process the simulated data with quantile mapping to the observed extremes. This would improve the extremes at these sites without significantly affecting other features of the data (for example, there are only approximately 10% – 15% Poor sites, the annual maximum(s) affects only a small number of data points per year, and only a fraction of the years would require shifting).

3.2.6 Multiyear annual evaporation totals (1, 2, 5 and 10 years)

Table 6 summarises the performance of the modelled evaporation data. As with the rainfall totals, the 1-, 2-, 5- and 10-year totals have Overall Good performance. At the annual scale there are 15 sites with Fair performance and at the 5- and 10-year scale there is one site with Poor performance.

Figure 41 shows a representative site, R31, which has Fair performance at the annual scale and Good performance at the multiyear scale. As with the rainfall totals, where a year is labelled Fair the distribution of the simulations remains reasonably close to the observations (that is, there are no large departures of the distribution). Figure 42 shows performance at R32, which is the sole site having poor reproduction of the annual evaporation totals. Notably, this site is one of the IQQM evaporation sites, and is a site where artefacts in the record have been commented on at the daily and monthly timescale (Section 2.1.2). Figure 42 shows that the distribution is not biased at all timescales, that it has Good performance at the annual scale, but that it is too variable at the 5- and 10-year timescales. This result could be addressed with post-processing.

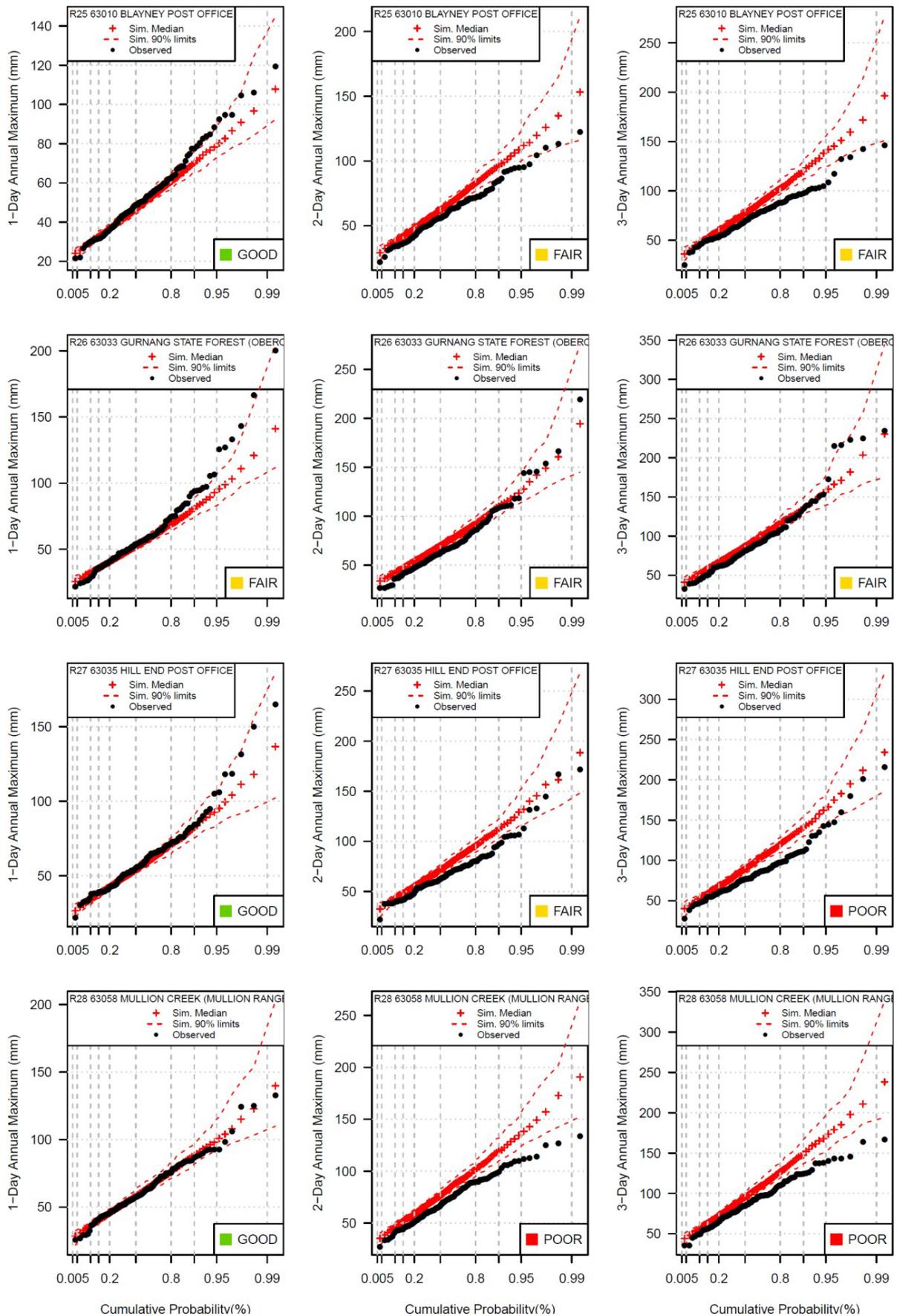


Figure 40 Frequency analysis of (left) annual maximum daily rainfall (middle) annual maximum 2-day rainfall (right) annual maximum 3-day rainfall for 4 representative sites R25–R28 for different performance outcomes

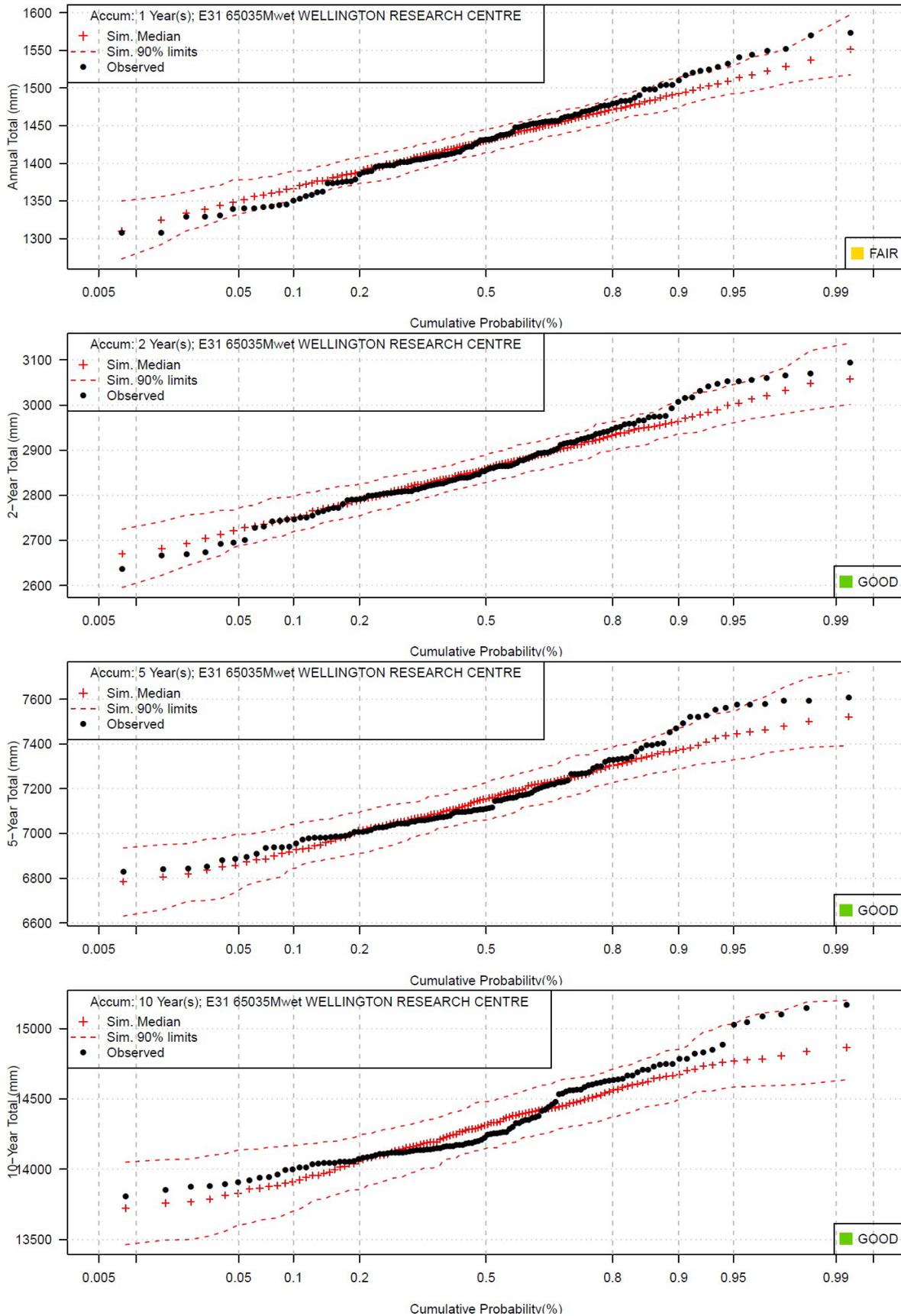


Figure 41 Distribution of annual evaporation totals for 1-, 2-, 5- and 10-year totals for a representative location with Fair/ Good performance; simulation shows 90% confidence interval

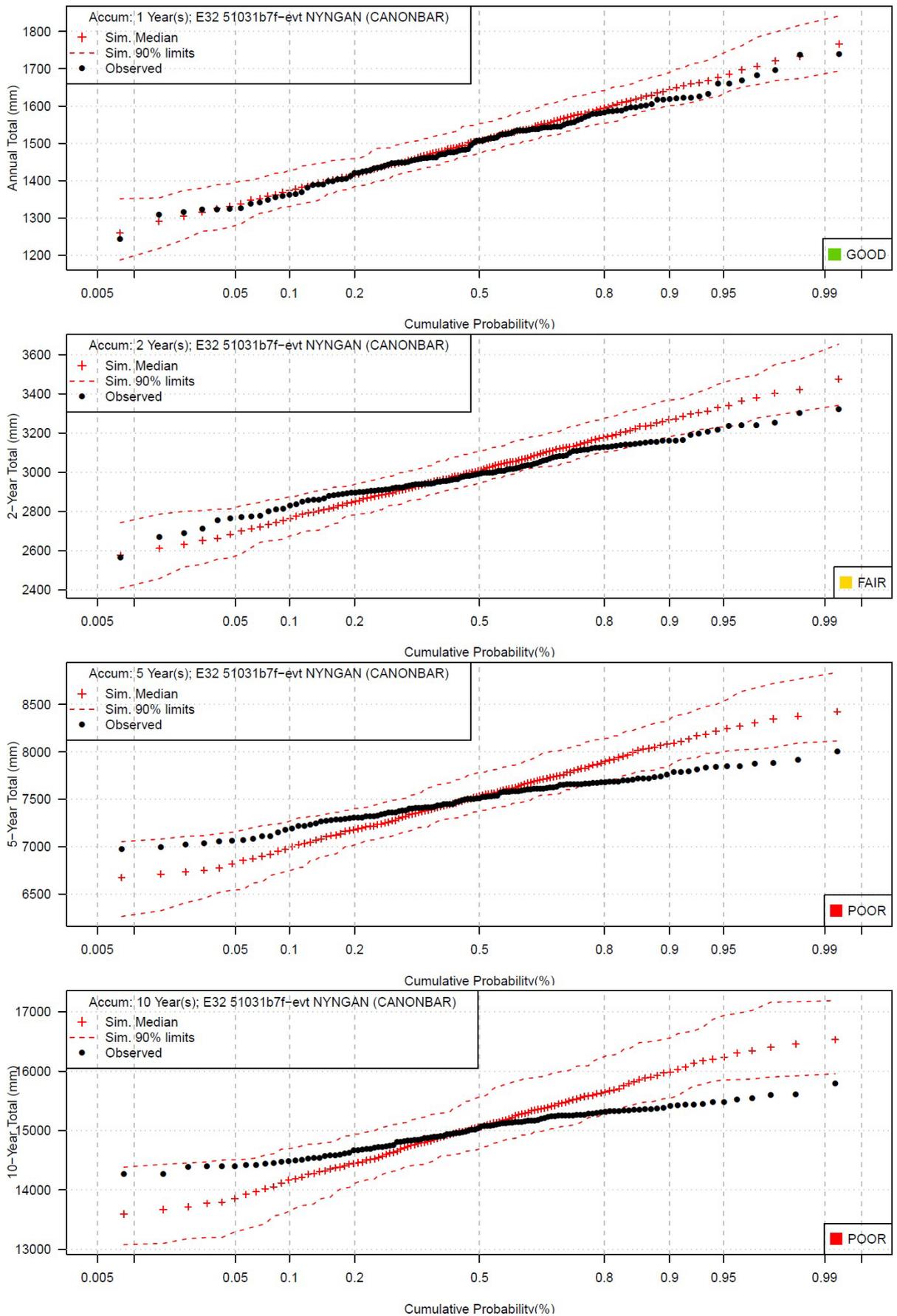


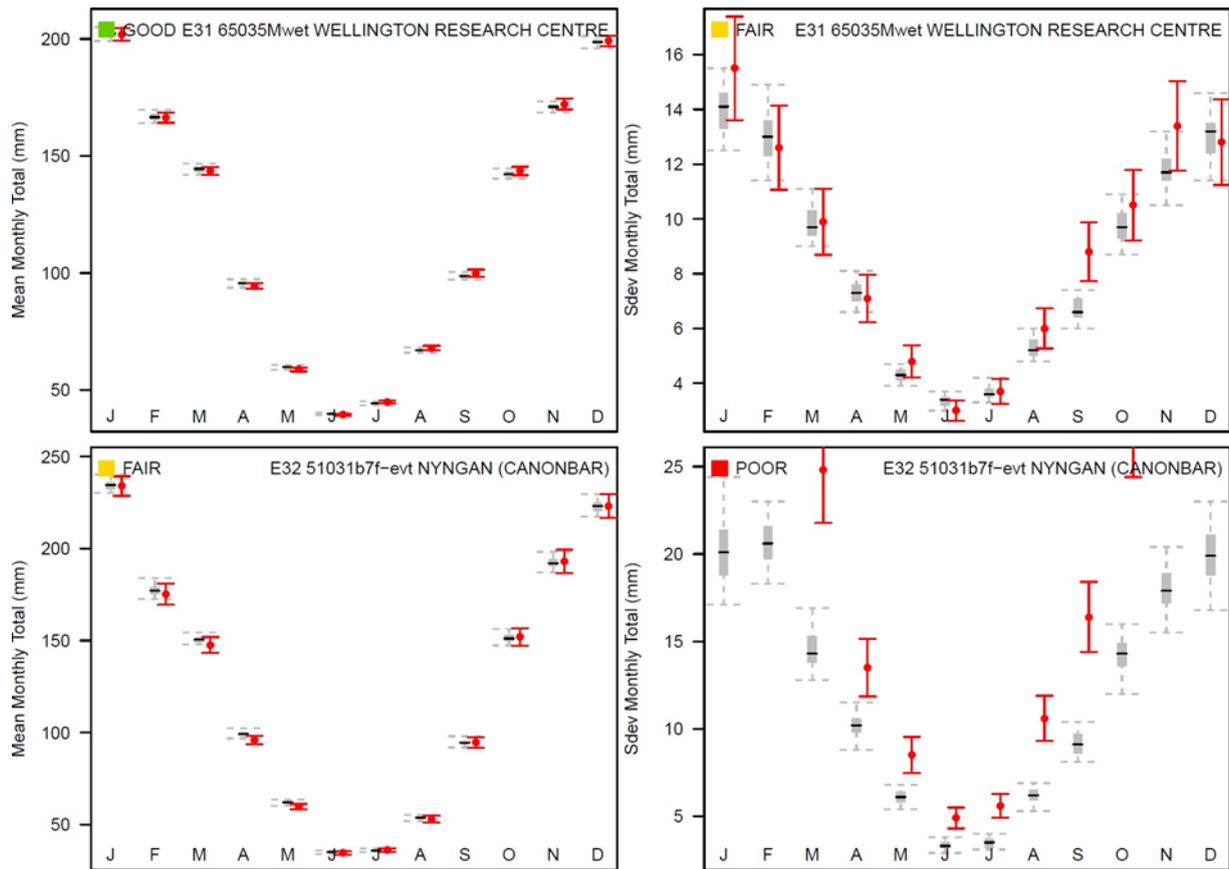
Figure 42 Distribution of annual rainfall totals for 1-, 2-, 5- and 10-year totals for a representative location with Poor performance; simulation shows 90% confidence interval

3.2.7 Mean and standard deviation of monthly evaporation totals

From Table 6, the performance of the monthly evaporation totals for the distribution mean is Overall Good and for the standard deviation is Overall Fair. Figure 43 shows the performance for 2 representative sites, E31 and E32. A total of 28 sites have Fair performance and 9 sites have Poor performance for the standard deviation of monthly evaporation totals.

The Fair performance occurs because the variability of the observations about the mean does not follow a sinusoidal function (for example, December observations are lower than November at E31). The sinusoids are fitted to the entire year and therefore smooth out the relationships between months. The implication is that it is possible for 1 or 2 months to show a discrepancy in the standard deviation – for example, the standard deviation of evaporation in September is 2 mm lower than the observations. The observations range from 85 mm to 115 mm (90% interval) whereas the simulations range from 88 mm to 112 mm.

Of the sites with Poor performance, 7 of these use IQQM data. Section 2.1.2 showed that these data were simulated such that the evaporation value was constant within each month within a given year, unless there was a rainy day, in which case a different value was used for all rainy days. This assumption causes an artefact that makes the daily data appear to be more correlated than in reality. The simulation model does not match this temporal correlation structure because the autocorrelation parameter is averaged across all sites (necessary for a contemporaneous model). Without further study of the data from which the IQQM evaporations were derived, it is difficult to conclude how reasonable the simulated data are for this metric at these sites. The other 2 sites with Poor performance are sites E43 and E44. These sites have discrepancies of 1–2 mm lower in the standard deviation for some months. Similarly to the Fair sites, the range of simulated monthly total evaporation at these 2 Poor sites is less than that in the observations by several millimetres.



Whiskers of grey box plot extend to the 90% interval. Red shows the observed mean/standard deviation along with error bars showing the standard error of each observed statistic. Two representative sites are shown (see Section 2.5)

Figure 43 Distribution of monthly evaporation totals; (left) means of monthly totals and (right) standard deviations of monthly totals

4 Discussion and recommendations

4.1 Model summary

A significant amount of data has been stochastically generated for use in hydrological modelling of the Macquarie River catchment. Each model variant has 196 time series, representing rainfall and evaporation data co-located at 117 unique locations: 53 are rainfall only, 56 are rainfall and Morton Wet evaporation and the remaining 7 sites are rainfall with all 3 variants of evaporation (Morton Wet, IQQM and FAO56). Each time series has been output for 10,000 years in a 2-column (date, value) format, including leap years.

Three of the model variants represent historically equivalent simulations. Model C was used for the main output and model evaluation because it incorporates paleoclimatic information of the IPO. The inclusion of IPO states identified that IPO-negative years have 75 mm extra rainfall and 23 mm less evaporation, on average. This information has been built into the simulated data. The longest IPO phase in the instrumental record is 34 years, but simulations of 10,000 years can generate much longer IPO phases (instrumental-calibrated gives 66 years on average for the longest period, paleoclimate-calibrated gives 86 years, based on simulations of the data shown in Figure 9). It should be noted that there is significant variability within each state (see Figure 11 and Figure 13).

Analysis of the observed data identified numerous features of importance to reproduce in the simulated data, including spatial gradients in the amounts of rain and evaporation, seasonal variability, interannual variability and negative correlation between the rainfall and evaporation. Some artefacts were noted in Section 2.2.3 regarding the IQQM evaporation. The data are generated from a model with a relatively coarse resolution and which assumes high persistence in daily evaporation, with the exception of rainfall days.

The model is carefully structured to account for spatial and temporal correlations. The spatial correlations (rain–rain and evaporation–evaporation) can be preserved at each site without any smoothing assumptions (that is, the 100×100 and 45×45 sample spatial correlations are used directly in the model simulations). There is a negative correlation between daily evaporation and rainfall of approximately -0.3 . This correlation is built into the model using a parametric correlation function to ensure stability. The IQQM data have low correlation to other evaporation estimates – even though they were simulated jointly with other evaporation, there is little benefit from so doing. A contemporaneous assumption is employed in the model, which requires the same autocorrelation parameter at all sites. This assumption provides stability to the correlation structure (needed for simulation) but can lead to a simplified temporal structure that can make the model inflexible for some attributes of the daily rainfall intermittency and properties such as multiday extremes.

4.2 Summary of model evaluation

Model evaluation was based on a systematic method of comparison, in which 2 developed tests were specified and applied to all relevant statistics. The tests rely on having replicates of equivalent length to the observations; therefore, the 10,000 years of simulated data were reshaped as 77 replicates, each of length 129 years. Confidence intervals for observed statistics were estimated from standard errors for means and variances and from a bootstrap method for distribution quantiles. All

evaluations require repeated tests across a collection of statistics, such as all months of the year or all quantiles in a distribution. Applying the developed classification rules:

- a label of Good is applied when more than 90% of the observations are within the 90% confidence intervals of the simulation; if this criterion is not met, a second test is applied to determine the classification
- a label of Fair is applied by comparing the simulated 90% confidence interval to the 90% range of observed sampling variability and determining that the intervals overlap in at least 90% of the instances; this is a more lenient test than that for the Good classification
- a label of Poor is established if neither of the above criteria are met.

A traffic light summary was produced by pooling the classifications across the 100 rainfall sites (Table 5) and 45 evaporation sites (Table 6), respectively. An overall classification was applied to a given statistic:

- if more than 50% of the sites have Good performance the statistic is summarised as Overall Good
- if more than 50% of the sites have Fair performance the statistic is summarised as Overall Fair
- if more than 50% of the sites have Poor performance the statistic is summarised as Overall Poor.

Multiyear totals: Given the importance of drought assessment for the Macquarie Valley, the statistics of most interest are the distributions of annual totals across multiple years. Rainfall and evaporation totals were considered for accumulation periods of 1, 2, 5 and 10 years and showed Overall Good performance.

Monthly totals: For the distribution of monthly totals – for both rainfall and evaporation – the means were classified as Overall Good and the standard deviations of the distributions as Overall Fair. The reasons for the Fair classification are different for the rainfall and evaporation. For rainfall, a practical interpretation of the Fair classification is that the simulated data have occasional very wet months that are wetter than the observations suggest is likely. Because the distribution of monthly totals is skewed, the standard deviations are sensitive to the behaviour of the tail. For evaporation, the model uses a sinusoid to match the monthly variability, but the observed data are not perfectly sinusoidal in their variability. An implication of this is that the model cannot match the standard deviation for every month. For example, at one site, for a given month the standard deviation of evaporation was 2 mm lower in the simulation. This is seen in the range – in this example, the observations range from 85 mm to 115 mm (90% interval) the simulations range from 88 mm to 112 mm. This issue arises for only some months due to the smoothing of the sinusoid.

Proportion of wet days: the annual distribution of this statistic was classified as Overall Poor. Inspection shows the distribution is unbiased (so the average number of wet days matches), but the simulations are less variable than the observations. Discussion of representative Poor sites (Table 7) showed that, for the driest year on record, the simulations had approximately 10 more wet days than the observations. An interpretation is that the simulations have slightly more ‘drizzle’ rain in the driest years. This is not necessarily a significant practical issue because the overall number of wet days is unbiased and the annual totals show Good performance (including in the tail regions) due to compensation by the process of rainfall amounts.

Multiday annual maximums: although the daily maximums were classified as Overall Good, the 2-day and 3-day maximums were classified as Overall Fair. Extremes are an emergent feature of the model and are not calibrated, so they can be difficult to match. Where there were discrepancies, the observed maximums were typically not far outside the 90% limits of the simulated maximums. At a

small number of Poor sites, the maximums showed a discrepancy in the upper tail, which could be post-processed using quantile mapping if necessary. The present study is focused on drought conditions rather than floods, so this discrepancy is not of concern here. It might be relevant if the data were used for a flood study in a nearby catchment with a response time in the order of several days.

4.3 Recommendations

A set of output data has been jointly simulated for 100 rainfall sites and 45 evaporation sites, having 10,000 concurrent years at each site. The simulated data are fit for application for drought assessment and show Overall Good reproduction of multiyear rainfall and evaporation totals.

- Of the 3 historical model variants, Model C is recommended as the best model variant for studying the water balance of the Macquarie River catchment. Models A and B are available for comparison if needed. Because all sites were simulated jointly, it is reasonable to use any subset of sites for modelling purposes, providing the years are concurrent (that is, the correlations are preserved). Given the strong gradients in the catchment, care should be taken when selecting sites to ensure they are representative.
- Subsequent hydrological simulations should consider any impact of the IQQM evaporation data. It appears to be qualitatively and quantitatively different from the other types of evaporation data (a different spatial pattern, higher totals, monthly blocks and high daily persistence).

All model variants reproduce a wide range of features including the distributions of amounts from days to multiples of years, key elements of variability such as the seasonal cycle and climatic oscillations, spatial correlations (by using sample estimates), rainfall–evaporation correlations and projected median changes from a selected climate scenario. The generated output time series are fit for the purpose of daily water balance modelling in the Macquarie Valley to facilitate risk assessments related to hydrological functions.

5 References

- Baxevani A and Lennartsson J (2015) 'A spatiotemporal precipitation generator based on a censored latent Gaussian field', *Water Resources Research*, 51(6):4338–4358.
- Bennett B, Thyer M, Leonard M, Lambert M and Bates B (2018) 'A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model', *Journal of Hydrology*, 556:1123–1138.
- Henley BJ, Thyer MA, Kuczera G and Franks SW (2011) 'Climate-informed stochastic hydrological modeling: incorporating decadal-scale variability using paleo data', *Water Resources Research*, 47(11):W11509.
- Henley BJ, Gergis J, Karoly DJ, Power SB, Kennedy J and Folland CK (2015), 'A tripole index for the Interdecadal Pacific Oscillation', *Climate Dynamics*, 45(11–12):3077–3090, doi:10.1007/s00382-015- 2525-1.
- Kleiber W, Katz RW and Rajagopalan B (2012), 'Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes', *Water Resources Research*, 48(1):W01523.
- NSW Department of Industry (2018) *Risk assessment for the Macquarie-Castlereagh water resource plan area (SW11), Part 1, Schedule D*, INT18/162302, New South Wales Government.
- NSW Department of Primary Industries (2016) *Macquarie–Castlereagh water resource plan (surface water), status and issues paper*, New South Wales Government.
- Rasmussen PF (2013), 'Multisite precipitation generation using a latent autoregressive model', *Water Resources Research*, 49(4):1845–1857.
- van Dijk A, Evans R, Hairsine P, Khan S, Nathan R, Paydar A, Winey N and Zhang (2006), *Risks to the shared water resources of the Murray–Darling Basin*, Murray–Darling Basin Commission, Canberra.
- Wilks DS (1998), 'Multisite generalization of a daily stochastic precipitation generation model', *Journal of Hydrology*, 210(1–4):178–191.

Wilks DS (2009), 'A gridded multisite weather generator and synchronization to observed weather data', *Water Resources Research*, 45(10): doi:10.1029/2009WR007902 Appendix A – Finalised list of sites

Table A.1 List of 100 rainfall sites used in the study

Internal ID	Station ID	Station location
R1	50018	Dandaloo (Kelvin)
R2	50031	Peak Hill Post Office
R3	50002	Goonumbla (Avondale)
R4	50004	Bogan Gate Post Office
R5	50012	Burra Burra
R6	50016	Goonumbla (Coradgery)
R7	50028	Trundle (Murrumbogie)
R8	50036	Trundle (Brookview St)
R9	51072	Quambone (Carwell)
R10	62026	Rylstone (Ilford Rd)
R11	65034	Wellington (Agrowplow)
R12	51031	Nyngan (Canonbar)
R13	51034	Warren (Mumblebone)
R14	51037	Narromine (Alagalah St)
R15	51054	Warren (Frawley St)
R16	62003	Mumbil (Burrendong Dam)
R17	62021	Mudgee (George Street)
R18	65012	Dubbo (Darling Street)
R19	51004	Trangie (Old Bundemar)
R20	51018	Gilgandra (Chelmsford Ave)
R21	55041	Nundle Post Office
R22	62013	Gulgong Post Office
R23	63004	Bathurst Gaol
R24	63005	Bathurst Agricultural Station
R25	63010	Blayney Post Office
R26	63033	Gurnang State Forest (Oberon Ya)
R27	63035	Hill End Post Office
R28	63058	Mullion Creek (Mullion Range Frst)
R29	63063	Oberon (Springbank)
R30	63064	O'Connell (Stratford)
R31	63066	Orange (Mclaughlin St)
R32	63089	Wattle Flat General Store
R33	64025	Coolah (Binnia St)
R34	65003	Bodangora Post Office

R35	65011	Cumnock (Willow Park)
R36	65023	Molong (Hill St)
R37	65035	Wellington Research Centre
R38	50008	Peak Hill (Bruie Plains)
R39	50037	Tullamore (Old Post Office)
R40	51005	Narromine (Mumble Peg)
R41	51008	Wyanga (Barcoo)
R42	51010	Coonamble Comparison
R43	51022	Gulargambone (Yalcogrin St)
R44	51025	Warren (Haddon Rig)
R45	51038	Nevertire (Clyde St)
R46	51042	Quambone Station
R47	51048	Trangie Post Office
R48	51049	Trangie Research Station Aws
R49	51051	Gilgandra (Berida)
R50	51066	Eumungerie Post Office
R51	51115	Narromine Airport
R52	62012	Cudgegong (Kiora)
R53	62014	Hargraves (General Store)
R54	62018	Katella
R55	62020	Bylong (Montoro)
R56	62027	Shepherds Creek
R57	62028	Goolma (Brooklyn)
R58	62029	Ilford (Tara)
R59	62031	Ilford (Warrangunyah)
R60	62033	Hargraves (Weeroona)
R61	62035	Leadville (Moreton Bay)
R62	62057	Coolah (Coolah Creek)
R63	62075	Galambine (Gooree Park)
R64	62084	Budgee Budgee (Botobolar Vnyrd)
R65	62099	Stuart Town (Canobla)
R66	63000	Abercrombie (Abercrombie Bridge)
R67	63011	Borenore Store
R68	63012	Running Stream (Brooklyn)
R69	63036	Oberon (Jenolan Caves)

R70	63037	Oberon (Jenolan State Forest)
R71	63053	Millthorpe (Inala)
R72	63071	Portland (Jamieson St)
R73	63073	Rockley Post Office
R74	63076	Sofala Old Post Office
R75	63079	Sunny Corner (Snow Line)
R76	63080	Black Springs (Swatchfield)
R77	63083	Trunkey Creek Black Stump Htl Stn
R78	63085	Paling Yards (Ulabri)
R79	63086	Blayney (Vittoria)
R80	63087	Black Springs Forestry
R81	63090	Wellwood
R82	63136	Yetholme (Kurrawong)
R83	63146	Cheetham Flats (Jundas)
R84	63233	Rockley (Clevelands)
R85	64009	Dunedoo Post Office
R86	64010	Elong Elong (Bendeela St)
R87	64015	Mendooran Post Office
R88	64026	Cobbora (Ellismayne)
R89	65000	Arthurville (Cramond)
R90	65005	Bumberry
R91	65010	Cudal Post Office
R92	65018	Geurie Post Office
R93	65020	Manildra (George St)
R94	65022	Manildra (Hazeldale)
R95	65025	Obley
R96	65026	Parkes (Macarthur Street)
R97	65030	Dubbo (Mentone)
R98	65032	Wandoo Wandong
R99	65036	Yeoval Post Office
R100	65037	Dubbo State Forest

Table A.2 List of 45 evaporation sites used in the study (31 unique locations because E32–E45 are co-located)

Internal ID	Station ID	Station location
E1	50018Mwet	Dandaloo (Kelvin)
E2	50031Mwet	Peak Hill Post Office
E3	51072Mwet	Quambone (Carwell)
E4	62026Mwet	Rylstone (Ilford Rd)
E5	65034Mwet	Wellington (Agrowplow)
E6	51031Mwet	Nyngan (Canonbar)
E7	51034Mwet	Warren (Mumblebone)
E8	51037Mwet	Narromine (Alagalah St)
E9	51054Mwet	Warren (Frawley St)
E10	62003Mwet	Mumbil (Burrendong Dam)
E11	62021Mwet	Mudgee (George Street)
E12	65012Mwet	Dubbo (Darling Street)
E13	51004Mwet	Trangie (Old Bundemar)
E14	51018Mwet	Gilgandra (Chelmsford Ave)
E15	55041Mwet	Nundle Post Office
E16	62013Mwet	Gulgong Post Office
E17	63004Mwet	Bathurst Gaol
E18	63005Mwet	Bathurst Agricultural Station
E19	63010Mwet	Blayney Post Office
E20	63033Mwet	Gurnang State Forest (Oberon Ya)
E21	63035Mwet	Hill End Post Office
E22	63058Mwet	Mullion Creek (Mullion Range Frst)
E23	63063Mwet	Oberon (Springbank)
E24	63064Mwet	O'connell (Stratford)
E25	63066Mwet	Orange (Mclaughlin St)
E26	63089Mwet	Wattle Flat General Store
E27	64025Mwet	Coolah (Binnia St)
E28	65003Mwet	Bodangora Post Office
E29	65011Mwet	Cumnock (Willow Park)
E30	65023Mwet	Molong (Hill St)
E31	65035Mwet	Wellington Research Centre
E32	51031b7f-evt	Nyngan (Canonbar)
E33	51034b7f-evt	Warren (Mumblebone)
E34	51037b7f-evt	Narromine (Alagalah St)

E35	51054b7f-evt	Warren (Frawley St)
E36	62003b17-evg	Mumbil (Burrendong Dam)
E37	62021_17-evg	Mudgee (George Street)
E38	65012b7f-evt	Dubbo (Darling Street)
E39	51031_FAO56	Nyngan (Canonbar)
E40	51034_FAO56	Warren (Mumblebone)
E41	51037_FAO56	Narromine (Alagalah St)
E42	51054_FAO56	Warren (Frawley St)
E43	62003_FAO56	Mumbil (Burrendong Dam)
E44	62021_FAO56	Mudgee (George Street)
E45	65012_FAO56	Dubbo (Darling Street)

The Mwet label indicates data following the Morton Wet equation, -evt/-evg labels indicate the IQQM evaporation data and FAO56 indicates the sites using reference crop data.

Appendix B – Traffic light comparison for Model A (base), Model B (instrumental IPO) and Model C (paleoclimatic IPO)

Table B.1 Rainfall evaluation summary of performance for 3 different model variants

Statistic	Model A: Base Model performance Good (%)	Model A: Base Model performance Fair (%)	Model A: Base Model performance Poor (%)	Model B: Instru mental IPO Model perform ance Good (%)	Model B: Instru mental IPO Model perform ance Fair (%)	Model B: Instru mental IPO Model perform ance Poor (%)	Model C: Paleocli matic IPO Model perform ance Good (%)	Model C: Paleocli matic IPO Model perform ance Fair (%)	Model C: Paleoc limati c IPO Model perform ance Poor (%)
Annual total rainfall nYr=1	93	7	0	99	1	0	96	4	0
Annual total rainfall nYr=2	99	1	0	99	1	0	99	1	0
Annual total rainfall nYr=5	94	5	1	97	2	1	97	3	1
Annual total rainfall nYr=10	89	11	0	96	4	0	95	5	0
Mean of monthly rainfall totals	100	0	0	100	0	0	100	0	0
Sdev of monthly rainfall totals	7	93	0	6	94	0	7	93	0
Distribution prop. wet days	0	15	85	3	24	73	5	20	75
Mean of monthly prop. wet days	98	2	0	99	1	0	98	2	0
Sdev of monthly prop. wet days	15	76	9	19	73	8	22	74	4
Annual 1-day rainfall max.	47	53	0	50	50	0	53	47	0
Annual 2-day rainfall max.	2	89	9	2	90	8	3	86	11
Annual 3-day rainfall max.	4	79	17	4	86	10	4	83	13

Table B.2 Evaporation evaluation summary of performance for 3 different model variants

Statistic	Model A: Base Model performance Good (%)	Model A: Base Model performance Fair (%)	Model A: Base Model performance Poor (%)	Model B: Instrumental IPO Model performance Good (%)	Model B: Instrumental IPO Model performance Fair (%)	Model B: Instrumental IPO Model performance Poor (%)	Model C: Paleoclimatic IPO Model performance Good (%)	Model C: Paleoclimatic IPO Model performance Fair (%)	Model C: Paleoclimatic IPO Model performance Poor (%)
Annual total rainfall nYr=1	80	20	0	87	13	0	84	16	0
Annual total rainfall nYr=2	91	9	0	91	9	0	93	7	0
Annual total rainfall nYr=5	89	9	2	91	4	5	91	7	2
Annual total rainfall nYr=10	98	2	0	93	7	0	98	2	0
Mean of monthly rainfall totals	98	2	0	93	7	0	98	2	0
Sdev of monthly rainfall totals	16	64	20	11	71	18	18	62	20