Methodology report for multisite rainfall and evapotranspiration data generation of the Northern Basins

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 New South Wales Department of Climate Change, Energy the Environment and Water.

Department Reference Number: PUB24/327

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 Jon Midwinter, Forugh Dorani, Rachel Blakers and Dushmanta Dutta from the New South Wales Department of Industry.

Contents

List of figures

List of tables

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 designed to be just as plausible as those occurring in the past.

This document outlines the method for generating time series of rainfall and evapotranspiration at many sites. The methods used build on the methodology developed in previous work. The model must:

- simulate 318 rainfall and potential evapotranspiration time series for a period of 10,000 years at sites from multiple catchments including the Border rivers, Gwydir, Namoi and Western regions
- use climate states from a prior 10,000-year simulation, developed during work for the Macquarie region; the simulated time series assumes a stationary climate state but allows for oscillations according to the Interdecadal Pacific Oscillation
- maintain correlations with 196 rainfall and potential evapotranspiration time series previously simulated for the Macquarie region at interannual, annual and monthly timescales; all simulated data will share 'joint' characteristics despite having been simulated in separate stages
- maintain correlations between 5 different variant formulations of evapotranspiration (IQQM PET, SILO Morton Wet, SILO FAO56, SILO Pan, SILO Morton Potential), 2 variants of reconstructed rainfall (SILO Rain, IQQM Rain) and a temperature time series
- maintain correlations between different variables that are located at the same coordinates
- preserve key statistical properties across all timescales, including interannual and multidecadal rainfall variability and the assessment of extreme droughts.

To meet these criteria, a hierarchical approach was used to accommodate different models at the interannual, annual, monthly and daily timescales. Spatially consistent data are important for drought assessment. The various catchments are part of the same basin and must therefore be treated consistently for potential use in risk assessments at scales larger than a single catchment. Achieving consistency through separate simulations, such that latter simulations are conditioned on earlier ones, adds considerable complexity when compared to methods that simulate all of the time series 'jointly' – that is, in a single simulation. Given the complexity, a post-processing step was added to correct for any artefacts in the simulated time series and to ensure simulated means are identical to the historical record (to avoid biases in flow simulations).

The following document reviews the available data and outlines the model developed to generate the stochastic time series. Separate annex documents and discussion are provided for each catchment to evaluate the generated data. In general, the model is able to reproduce multiyear annual totals of rainfall and potential evapotranspiration, which are relevant for drought assessment. The model preserves seasonality, with a post-processing step used to ensure the mean of the distribution of monthly totals matches the monthly means from the observed record. As noted in previous work, the model matches the median of the number of wet days but underestimates the variability in the number of wet days. The maximums from 1-, 2- and 3-day totals show sound reproduction compared to observations at the majority of sites.

1 Introduction

The New South Wales Government has committed to delivering 12 regional water strategies for the state. A key input to the strategy is the generation of stochastic paleoclimatic data to better reflect climate variability and to help the department understand the impacts of this variability on water resource management.

1.1 Project scope

The project requires the generation of time series of rainfall, evapotranspiration and (in one case) temperature data at multiple sites. This is done by building on the methodology developed in previous work (Leonard and Westra 2019; Leonard et al. 2019). The method has been extended to allow for conditioning rainfall amounts from new locations to be consistent (that is, spatially and temporally correlated) with any existing simulated data. The requirements for the Northern Basin are outlined below, and specify the variable types, catchments and key relationships of interest. The data must:

- simulate 318 rainfall and potential evapotranspiration time series for a period of 10,000 years at sites from multiple catchments including the Border rivers, Gwydir and Namoi catchment regions
- use climate states from a prior 10,000-year simulation, developed during work for the Macquarie region (Leonard et al. 2018); the simulated time series assumes a stationary climate state allowing for oscillations according to the Interdecadal Pacific Oscillation
- maintain correlations with 196 rainfall and potential evapotranspiration time series previously simulated for the Macquarie region (Leonard et al. 2018); at interannual, annual and monthly timescales; all simulated data will share 'joint' characteristics despite having been simulated in separate stages
- maintain correlations between 5 different variant formulations of evapotranspiration (IQQM PET, SILO Morton Wet, SILO FAO56, SILO Pan, SILO Morton Potential), 2 variants of reconstructed rainfall (SILO Rain, IQQM Rain) and a temperature time series
- maintain correlations between different variables that are located at the same coordinates
- preserve key statistical properties across all timescales, including interannual and multidecadal rainfall variability and the assessment of extreme droughts. Evaluation should include mean and variance of monthly totals, frequency analysis of daily to multiday extremes and analysis of the number and distribution of rain days.

The model calibrates separate parameters to positive and negative phases of the Interdecadal Pacific Oscillation (IPO) to account for interannual variability (Leonard et al. 2018). The rainfall and evapotranspiration data are partitioned based on the IPO instrumental record from 1890 to 2018. Paleoclimate information is used to inform the distribution of dwell times in each phase of the IPO.

Model developments have been implemented to ensure that spatial relationships are preserved with respect to existing simulated data from other catchments, especially at longer timescales, so that multi-catchment analyses of risk are feasible. A hierarchical structure is used so that annual totals are simulated according to the phase of the IPO, then monthly and daily values are downscaled based on those annual totals.

The outputs are generated as individual time series (of 10,000 years). To provide maximum flexibility, they can be used as a single block or broken 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. In either case, the simulated data represent long-term stationary assumptions.

This document summarises the methodology used to generate and evaluate the stochastic time series of rainfall and evapotranspiration. Separate but related annex reports are provided to summarise data and performance in each basin (for example, Gwydir, Namoi).

2 Data and methodology

2.1 Observation data

Figure 1 shows the locations of rainfall, evapotranspiration and temperature data across the region.

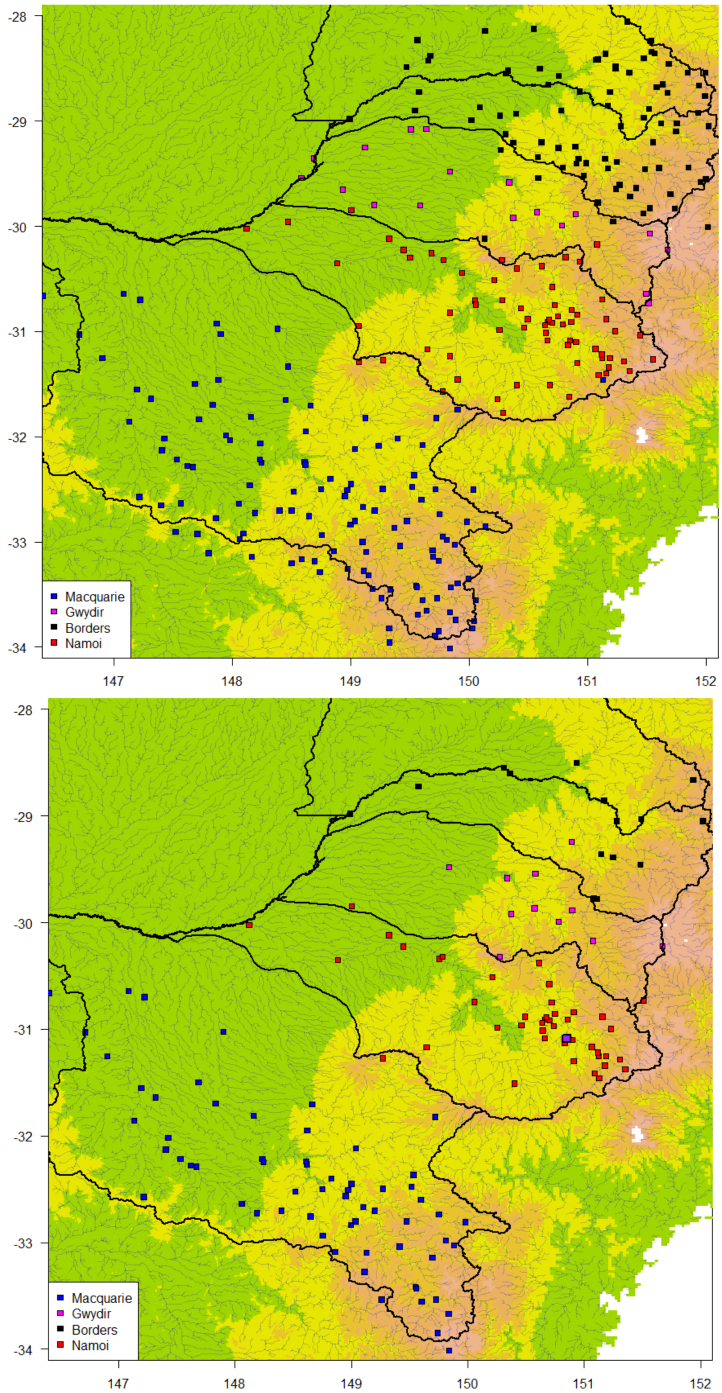


Figure 1 Location of rainfall sites (top panel) and evapotranspiration sites (bottom panel); the single temperature site (Tamworth airport) is shown in the bottom panel as a purple symbol in the Namoi region

Table 1 shows the distribution and types of historical data. Note that the historical data come from sources that have reconstructed or interpolated time series with varying degrees of modelling and infilling. Data labelled IQQM have been generated by the department for modelling as part of earlier IQQM water balance assessments, whereas data labelled SILO have been externally developed (www.longpaddock.qld.gov.au). In both instances the provided data are considered representative and are used as-is with only preliminary checks for consistency.

Table 1a shows the data split according to existing and new time series. The existing series are from the Macquarie region (Leonard et al. 2018) which had 117 rainfall time series and 79 evapotranspiration time series. Due to the method of conditioning developed in this report, these existing 196 time series do not need to be re-simulated, but they are important for subsequent simulations to ensure spatial and temporal correlations are preserved. The simulation for the Macquarie region involved the generation of a 10,000-year synthetic series of positive and negative states of the IPO. The same time series of simulated IPO states is required for simulation over all additional regions. The additional catchments represent 318 new time series, comprising 185 rainfall time series, 132 evapotranspiration time series and 1 temperature time series.

Table 1a Historical data used in the study

Type	Rain	Evapotranspiration	Temperature	Total
Total existing	117	79	0	196
Total new	185	132	1	318

Table 2b List of historical data used in the study

Region	SILO Rain	IQQM Rain	SILO Mwet	SILO FAO56	SILO Pan	SILO Mpot	IQQM PET	SILO Temp	Total
Macquarie	117	0	64	7	0	0	8	0	196
Gwydir	24	0	10	13	0	0	13	0	60
Namoi	63	5	36	13	7	0	20	1	145
Border	93	0	13	5	0	2	0	0	113

Table 1b provides an additional level of detail according to the specific catchments and the different types of variables. There are 2 types of rainfall product, although there are only minor discernible differences in their properties, based on the attributes of the time series. The IQQM Rain sites represent only those sites located at unique coordinates – differing from the SILO Rain sites. A preliminary analysis was conducted showing that for sites located at the same coordinates but having different type (SILO or IQQM), the values were indistinguishable in terms of summary statistics. Figure 2 shows a 1:1 relationship for most values, with only a small number of departures (noting that there are over 47,000 data points from a 129-year record). For models requiring IQQM Rain, wherever there is a co-located SILO Rain site, the latter can be used without loss of efficacy. Some rainfall sites had identical ID tags but differing coordinates and differing values (for example, 52020 and 52020SID SILO Rain). These represent 2 different interpolations in which the flag SID is used to denote the time series used for 'Source' modelling. Both time series are generated in the simulated data.

Table 1 shows that the Macquarie region had 3 types of evapotranspiration data (SILO Morton Wet, SILO reference crop FAO56 and 'IQQM' derived). The new basins require simulation of 2 additional evapotranspiration variables (SILO Pan, SILO Morton potential) and a single temperature time series.

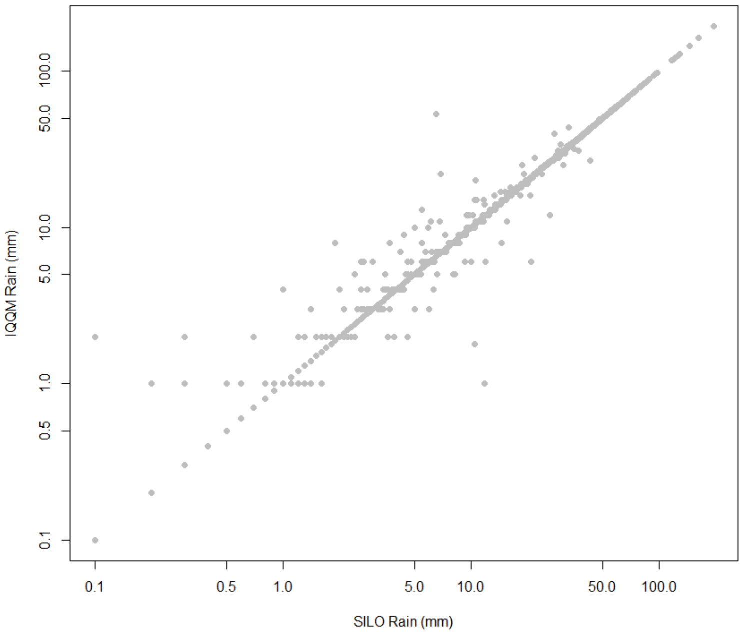
The SILO time series all follow 'smooth' seasonal trends, and the quality of the IQQM time series was discussed with reference to the Macquarie region (Leonard et al. 2018).

For all of the time series, there were no 'internal' missing values, and the data cover the period 1 January 1890 to 31 December 2018. Some sites were missing 6 months of data at the start or end of the record (for example, 52020SID SILO Rain), which was infilled by duplicating the identical period from a neighbouring year. This represents a negligible discrepancy and enables an identical record length to be used across all sites.

The temperature time series can be accommodated into the same methodology as the evapotranspiration data because it has a similar seasonal structure, which can be fitted using sinusoids (Figure 3).

As with the Macquarie region analysis, the Hadley SST version of the IPO was used to derive positive and negative phases. 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).



The IQQM and SILO rainfall are highly similar, and any differences are assumed to be from differing interpolation algorithms. There are numerous discrepancies from the 1:1 relationship, but they are negligible within a 129-year record (and indistinguishable in terms of summary statistics).

Figure 2 Comparison of rainfall located at station ID 54003, with identical coordinates, but different types

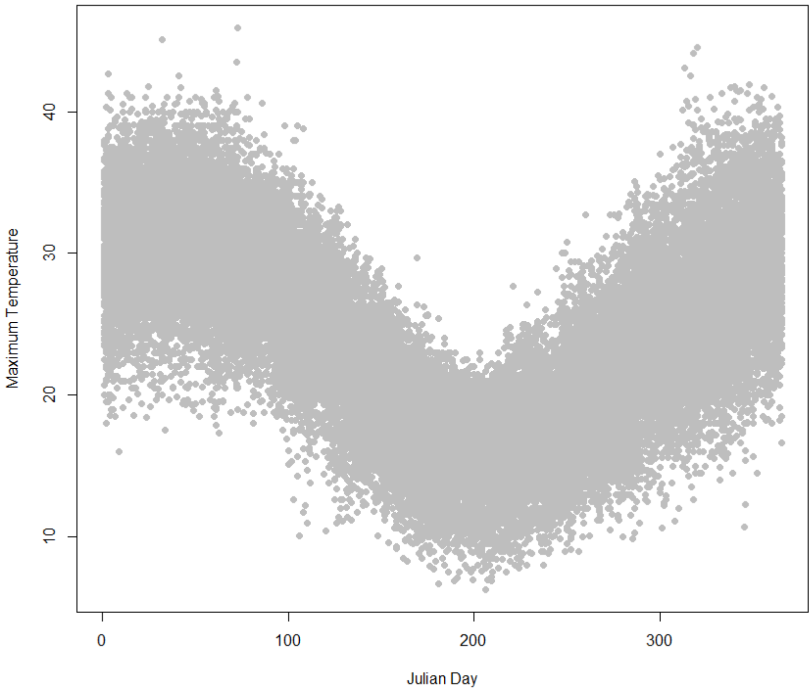


Figure 3 Time series of maximum temperature at Tamworth Airport

2.2 Model specification – base model

2.3 Overview of the existing multisite stochastic generator

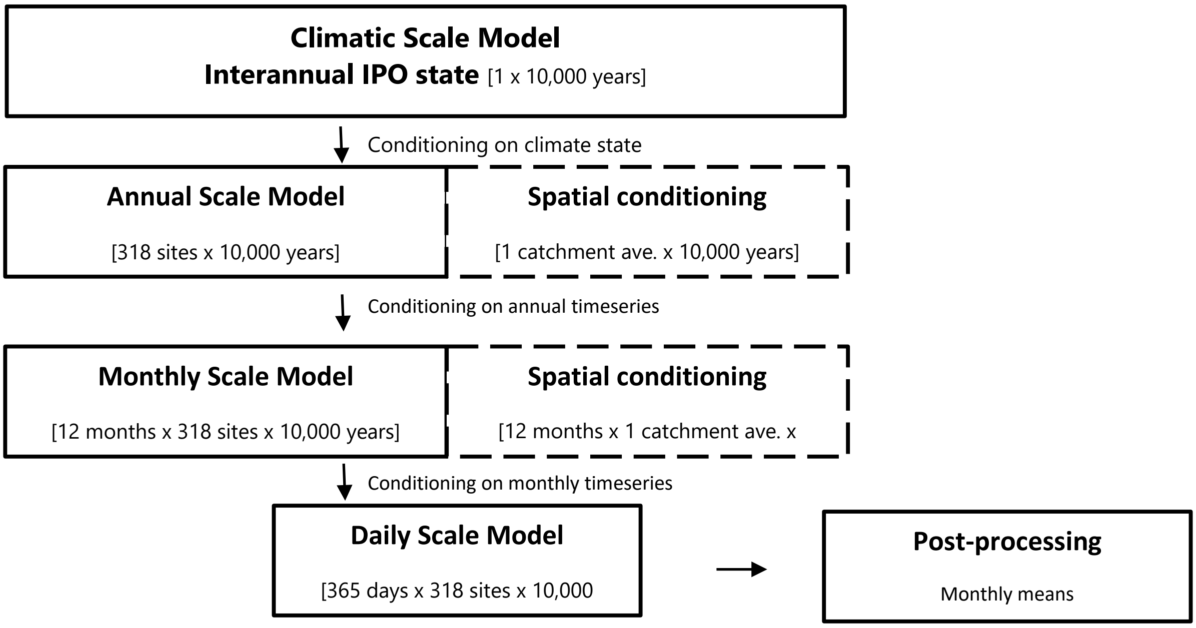
The method implemented for the Macquarie Regional Water Strategy generated 10,000-year time series for 117 rainfall sites and 79 evapotranspiration sites. The model was built as a highly dimensional Gaussian probability distribution, with transformations to ensure key features of rainfall and evapotranspiration were preserved (Leonard and Westra 2019; Leonard et al. 2019). A benefit of this approach is that it is a natural method for implementing correlations in space and time and can handle a large number of sites. These benefits are exploited in this project to ensure the data can be simulated jointly for all locations of interest.

Figure 4 shows the structure of the modelling methodology used in this project. The following subsections provide details of each level. The existing simulation for the Macquarie region included a stochastic simulation of the IPO for 10,000 years, which allows the partitioning of above- and below-average wet and dry periods. The same stochastic series of IPO states must therefore be used for subsequent simulations.

The annual scale model is calibrated with different parameters for each state of the IPO. At the annual scale, the simulated annual totals are spatially correlated to the catchment average annual totals of the Macquarie region. The conditional structure of the Gaussian distribution is used for this purpose. This ensures the annual totals across the basin are synchronised. The method is generic so that additional new regions can be conditioned on existing data from multiple catchments.

The simulation uses a 'downscaling' approach to generate monthly values that directly add up to the simulated annual total, but which are also correlated with monthly catchment average values from the prior simulated catchment. The downscaling method relies on non-dimensionalising the monthly values as a multiplier (proportion) of the annual total.

Daily values are generated based on monthly totals via a downscaling approach. The downscaling approach uses a brute-force simulation to match daily values to nearest monthly totals while preserving spatial correlations. Due to artefacts from the brute-force matching, a post-processing step is added to ensure key statistical properties are preserved.



The climatic scale model time series is fixed from previous work (Leonard and Westra 2019; Leonard et al. 2019). The annual scale model simulates spatially correlated annual totals for 10,000 years with separate parameters for each IPO state. The monthly scale model simulates spatially correlated monthly totals that match the simulated annual total. The daily scale model simulates spatially correlated daily values conditioned on the monthly scale, yielding a complete set of time series. A post-processing step is added to ensure monthly means are matched. The annual and monthly scales are also spatially conditioned to ensure that they are correlated to the catchment average rainfall for the region so that wet years and months are synchronised).

Figure 4 Structure of modelling methodology showing a hierarchical model from climate scale to daily scale

2.3.1 Climatic scale model

Figure 5 shows a sample of the first 1,000 years of the IPO climate state simulation from the Macquarie region, which is used across all new regions. The histogram shows the simulated dwell time spent in the positive and negative states of the IPO. The sequence of states defines the parameters that are used at all subsequent scales, where the parameters embody the magnitude of the effect (difference in rainfall and evapotranspiration). It can be seen that while many IPO periods are shorter than a decade, based on this simulation, it is possible for IPO states to persist for several decades.

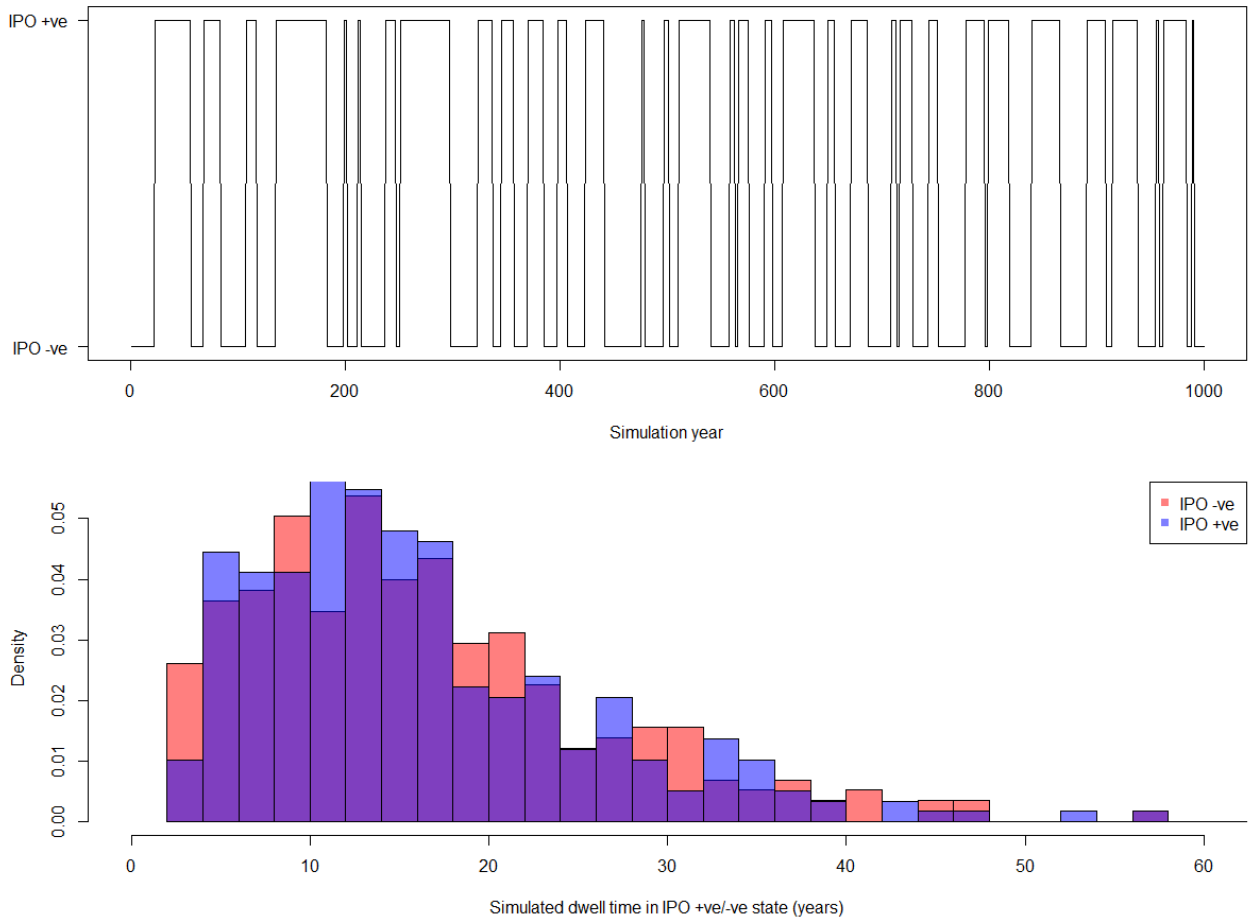


Figure 5 Sample of climate state simulation shown for first 1,000 years (top panel) and distribution of periods spent in each IPO state (bottom panel)

2.3.2 Annual scale model

When there are many sites over a large area and potential for significant variation in site characteristics across the region, a key challenge is that all sites must be simulated jointly to capture spatial dependence between them. Additional challenges are to ensure simulations of new regions maintain consistency with the simulated time series from the existing Macquarie sites. Similarly, if using a staged approach to the simulation (for example, Gwydir conditioned on Macquarie, then other regions must be conditioned on Gwydir and Macquarie), each newly added simulation must maintain consistency with all prior simulations. Gaussian-based models are well positioned to achieve joint simulation under these constraints because they can factorise a joint distribution into a set of conditional distributions.

The method outlined in the Macquarie Valley report (Leonard et al. 2018) demonstrates that both rainfall and evapotranspiration can be represented via a multivariate Gaussian distribution, in which each site is a separate dimension of the distribution. At the daily scale, the distribution is highly non-Gaussian and a power transformation is used to match the observations. At the annual scale, the distribution is less skewed but nonetheless demonstrates some departure from Gaussian behaviour. Figure 6 shows an example for a selected site where there is positive skewness with proportionally more values sampled in the upper tail than in a Gaussian distribution. The method used in this project to account for the skewness is to apply a quantile mapping based on a kernel density estimate (KDE) of the observations (using automatic bandwidth selection). The KDE can represent the shape of the marginal distribution at each site, while the quantile mapping ensures that the correlation structure and efficient simulation of the Gaussian distribution can be exploited.

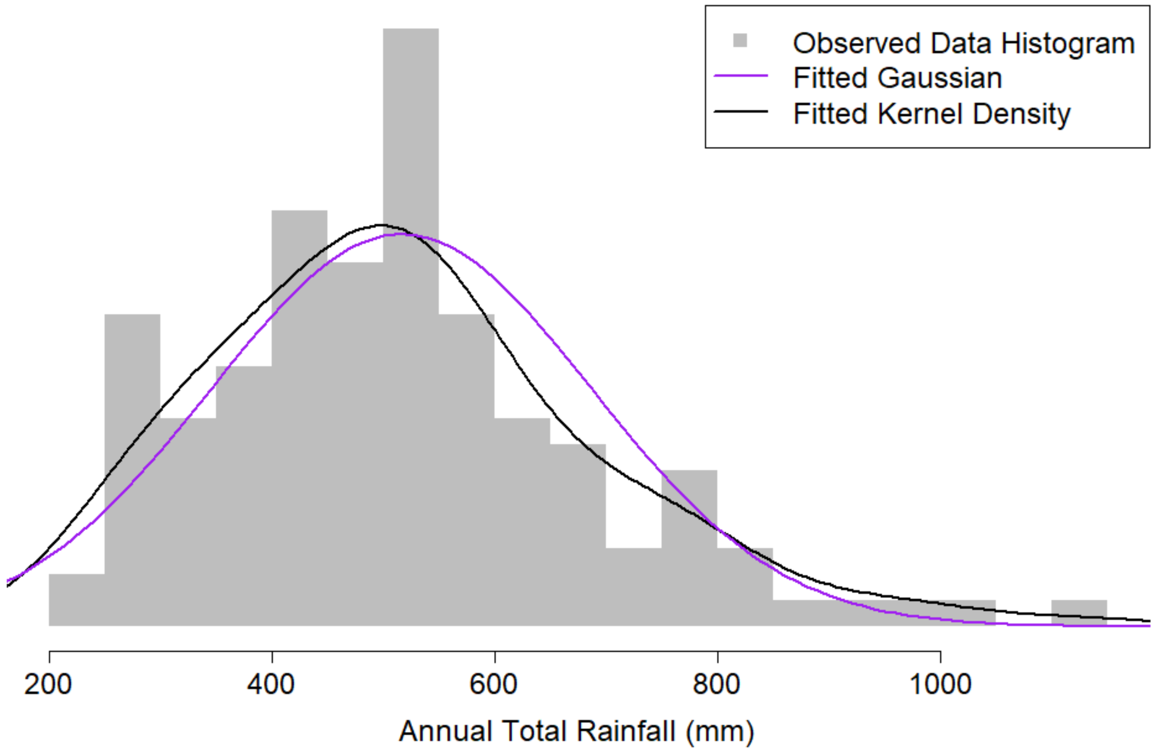


Figure 6 Annual scale distribution for representative site showing comparison of Gaussian fitted density and a kernel density estimate to observed histogram

In the following equations, the superscript prime is used to indicate a variable at the annual scale.

Let $g'(\mathbf{Z}')$ represent a transformation of the true rainfall distribution into Gaussian space at the annual scale, where in this instance a KDE transformation is applied to map quantiles of the Gaussian distribution.

$$\mathbf{X}' = g'(\mathbf{Z}') \tag{1}$$

The KDE-transformed annual rainfall totals, $\mathbf{X}'$, are modelled according to a multivariate Gaussian distribution that is defined by parameters for the mean, $\boldsymbol{\mu}'$, and covariance, $\boldsymbol{\Sigma}'$

$$\mathbf{X}' \sim N(\boldsymbol{\mu}', \boldsymbol{\Sigma}') \tag{2}$$

where the parameters for new (N) and existing (E) regions can be partitioned into 2 groups

$$\boldsymbol{\mu}' = \begin{bmatrix} \boldsymbol{\mu}'_N \\ \boldsymbol{\mu}'_E \end{bmatrix} \boldsymbol{\Sigma}' = \begin{bmatrix} \boldsymbol{\Sigma}'_{NN} & \boldsymbol{\Sigma}'_{NE} \\ \boldsymbol{\Sigma}'_{EN} & \boldsymbol{\Sigma}'_{EE} \end{bmatrix}. \tag{3}$$

This partition enables the use of a standard multivariate regression of the Gaussian distribution to define the conditional relationship (N|E) between the new and existing regions, with parameters

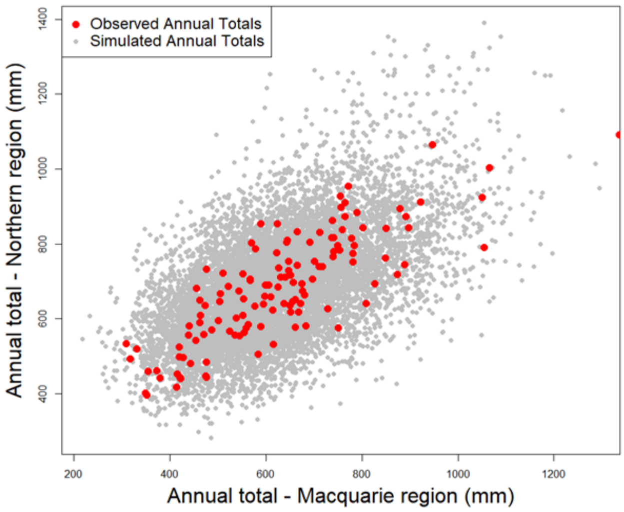
$$\boldsymbol{\mu}'_{t,N|E} = \boldsymbol{\mu}'_N + \boldsymbol{\Sigma}'_{NE}\boldsymbol{\Sigma}'^{-1}_{EE}\left(\mathbf{X}'_{t,E} - \boldsymbol{\mu}'_E\right) \text{ and } \boldsymbol{\Sigma}'_{N|E} = \boldsymbol{\Sigma}'_{NN} - \boldsymbol{\Sigma}'_{NE}\Sigma'^{-1}_{EE}\boldsymbol{\Sigma}'_{EN} \tag{4}$$

that depend on the simulated values in the existing region for any given timestep t, $\mathbf{X}'_{t,E}$.

The regression relationship preserves the spatial correlation across regions. The temporal correlation is modelled using a separable covariance matrix, such that the temporal correlation from one year to the next is the same at all sites (here, separable means that the space–time covariance matrix is the product of the spatial covariance with the scalar temporal correlation parameter). The temporal correlations are simulated according to an autoregressive process AR(1), with a parameter φ' for the lag-1 correlation in time, averaged over all sites in the new region. The simulated values at timestep $t+1$ at all sites in the new region ($\mathbf{X}'_{t+1,N}$) are obtained via autoregression on values at time t with a spatially correlated innovation of the conditional Gaussian distribution;

$$\mathbf{X}'_{t+1,N} = \varphi'\left(\mathbf{X}'_{t,N} - \boldsymbol{\mu}'_{t,N|E}\right) + \boldsymbol{\epsilon}'_{t,N|E} \quad \text{where } \boldsymbol{\epsilon}'_{N|E} \sim N\left(0, (1-\varphi'^2)\boldsymbol{\Sigma}'_{N|E}\right). \tag{5}$$

Figure 7 shows a comparison of simulated and observed annual totals from this model. The red values are the 129 pairs of observed annual totals averaged over the Macquarie catchment (horizontal axis) and northern catchments (vertical). The grey values are the corresponding 10,000 values from the simulated model. The comparison shows that the simulated and observed data have similar correlations in annual totals at the catchment scale.



The grey values are the simulated totals and correlation, the red values are the observed values.

Figure 7 Correlation at the annual scale between the existing region (Macquarie) and new region (Northern region catchments)

2.3.3 Monthly scale model

To condition monthly totals on annual totals, a hierarchical structure was adopted to ensure consistency across the region at monthly and annual timescales. In contrast, the Macquarie region was simulated at the daily timescale with the only hierarchy being the climate state model for interannual variability. Having additional hierarchies at the monthly and annual timescales provides an opportunity to directly calibrate the monthly and annual scales.

The monthly scale model is obtained by non-dimensionalising the monthly totals with respect to the annual total. In this manner, each year of 12 months is represented as 12 multipliers on the scale (0, 1). Using a multiplicative structure means that the simulated annual total is preserved by ensuring that the sampled multipliers sum to 1. For rainfall, the distribution of multipliers is similar across the months of the year, with considerable overlap in the distributions and with a median value of approximately 0.085 (that is, 1/12). A typical distribution of multipliers for rainfall is shown in Figure 8 for a selected site. It shows the skewed relationship, with small values being by far the most common – some multipliers have zero as a value (that is, dry months).

Figure 9 shows the multipliers for evapotranspiration for selected months – they are strongly seasonal. Although the evapotranspiration multipliers are close to following a Gaussian distribution, KDE is nonetheless used because some months exhibit skewness (for example, January).

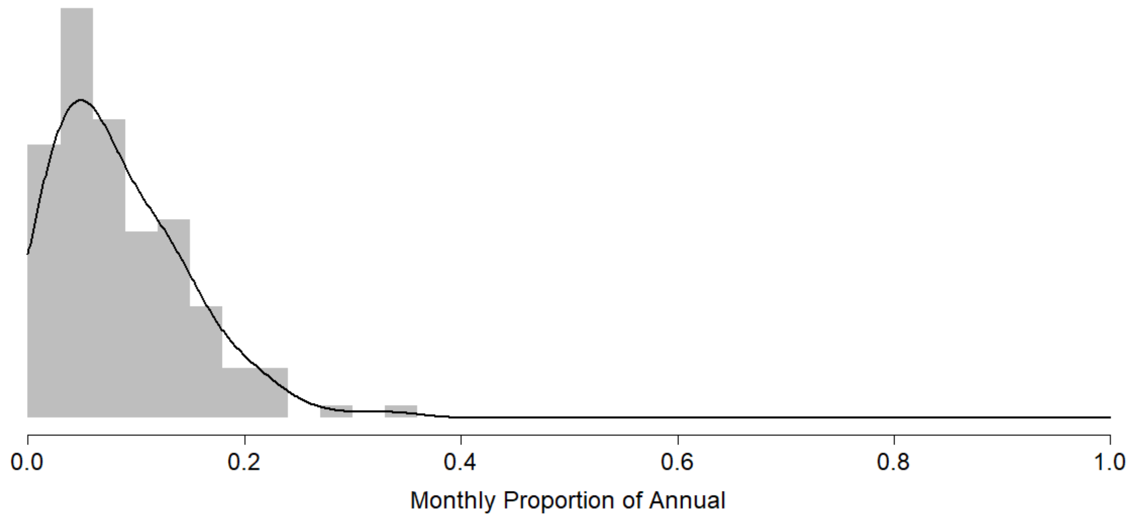


Figure 8 Distribution and kernel density estimate of monthly multipliers for rainfall

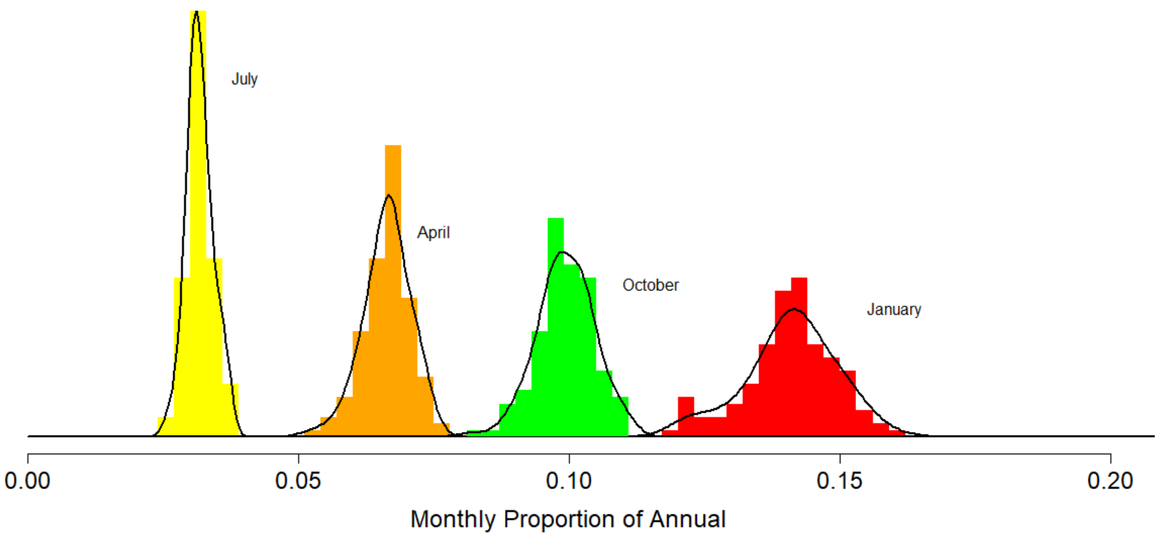


Figure 9 Distribution and kernel density estimate of monthly multipliers for evapotranspiration (selected months shown) to illustrate strong seasonal cycle

In the following, the superscript double-prime is used to indicate a variable at the monthly scale.

Let $\mathbf{M}'' = \mathbf{Z}''/\mathbf{Z}'$ be the monthly multiplier that is the ratio of the monthly total to the annual total. Let $g''(\mathbf{M}'')$ represent the transformation of the monthly multiplier distribution into Gaussian space at the monthly scale (that is, into the variable $\mathbf{X}''$), where in this instance a KDE transformation is applied to map quantiles of the Gaussian distribution. Using the multiplicative structure, the monthly correlations are represented as correlations in the proportional weighting of monthly rainfall in a given year, regardless of whether the year was a high- or low-rainfall year. Separate kernel density estimates are fitted to each month.

$$\mathbf{X}'' = g''(\mathbf{M}'') \tag{6}$$

The multipliers are simulated at the monthly scale according to the same spatial and temporal model at the annual scale, which allows for spatial conditioning of the model as well as autocorrelation in simulated values. An additional set of monthly parameters are used at this scale for marginal properties ($\boldsymbol{\mu}''$), spatial correlation

$$\mathbf{X}'' \sim \mathrm{N}(\boldsymbol{\mu}'', \boldsymbol{\Sigma}'') \tag{7}$$

The parameters at the monthly scale are also partitioned according to new (N) and existing (E) regions

$$\boldsymbol{\mu}'' = \begin{bmatrix} \boldsymbol{\mu}''_N \\ \boldsymbol{\mu}''_E \end{bmatrix} \boldsymbol{\Sigma}'' = \begin{bmatrix} \boldsymbol{\Sigma}''_{NN} & \boldsymbol{\Sigma}''_{NE} \\ \boldsymbol{\Sigma}''_{EN} & \boldsymbol{\Sigma}''_{EE} \end{bmatrix} \tag{8}$$

As at the annual scale monthly conditional relationship between regions can be expressed; they depend on the simulated values in the existing region for any given timestep t, $\mathbf{X}''_{t,E}$.

$$\boldsymbol{\mu}''_{t,N|E} = \boldsymbol{\mu}''_N + \boldsymbol{\Sigma}''_{NE}\boldsymbol{\Sigma}''^{-1}_{EE}\left(\mathbf{X}''_{t,E} - \boldsymbol{\mu}''_E\right) \text{ and } \boldsymbol{\Sigma}''_{N|E} = \boldsymbol{\Sigma}''_{NN} - \boldsymbol{\Sigma}''_{NE}\Sigma''^{-1}_{EE}\boldsymbol{\Sigma}''_{EN} \tag{9}$$

The temporal correlations are simulated according to an autoregressive process AR(1) with a parameter φ'' for the lag-1 correlation in time, averaged over all sites in the new region. The simulated values at timestep $t+1$ at all sites in the new region ($\mathbf{X}'_{t+1,N}$) are obtained via autoregression on values at time t with a spatially correlated innovation of the conditional Gaussian distribution. Based on preliminary testing, the parameter for monthly correlation in time is made common for all pairs of months (for example, January-to-February temporal correlations in multiplicative weights are the same as February-to-March correlations):

$$\mathbf{X}''_{t+1,N} = \varphi''\left(\mathbf{X}''_{t,N} - \boldsymbol{\mu}''_{t,N|E}\right) + \boldsymbol{\epsilon}''_{t,N|E} \text{ where } \boldsymbol{\epsilon}''_{N|E} \sim N\left(0, (1 - \varphi''^2)\boldsymbol{\Sigma}''_{N|E}\right). \tag{10}$$

The simulated multiplier values are summed to one by means of a sampling method, rather than by directly building this constraint into the simulated values for each subsequent month of the year.

2.3.4 Daily scale model

The daily model in this project is identical to the daily model used in the Macquarie region (Leonard et al. 2018) and its details are therefore not reproduced here. Rather than directly conditioning the daily model on the monthly model, for numerical stability the original model was simulated independently. Each month was then sampled to find the pattern of monthly totals in the independent simulation that most closely matched the monthly model from the conditional simulation based on the existing region at the monthly hierarchy (Section 2.3.2). This method is conceptually simple but computationally intensive.

Based on preliminary testing, periods of 2,500 independent months were simulated, and from these 1 month of closely matching totals was obtained. Since the spatial pattern in the independent simulation is not exactly the same as the monthly totals to be matched at all sites, the daily totals were rescaled (multiplicatively) to ensure that the monthly totals match exactly at each site. Post-processing was applied to ensure that the simulated monthly means matched the exact average values from the historical record.

2.4 Model evaluation

Evaluation of the stochastic replicates follows the same methodology as used in the Macquarie region ((Leonard et al. 2018). An explanation is reproduced here for reference.

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 were produced for each location, which enabled visual inspection of the results in conjunction with the formal model evaluation. Regardless of the method, interpreting the evaluation requires consideration of the relevance of the statistic to the application of interest.

2.4.1 Evaluation of distribution quantiles

A common case to evaluate is how well the quantiles of a distribution are matched between observations and simulations. Figure 10 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 11), 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 test than test 1. A bootstrap method can be used to calculate the sample variability of the observed statistic. If most quantiles overlap, then a classification of Fair is applied (see example C).
  - If both tests are failed, a classification of Poor is applied (see example D).

2.4.2 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 2.4.1) 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 11 provides a flow chart for this type of test, which is very similar to the test illustrated in Figure 10. Note that Good and Fair labels can be achieved only if not more than 1 month's observed statistic is outside the simulated limits. As in Figure 10, test 2 is more lenient because it allows for sample variability in the observation. 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).

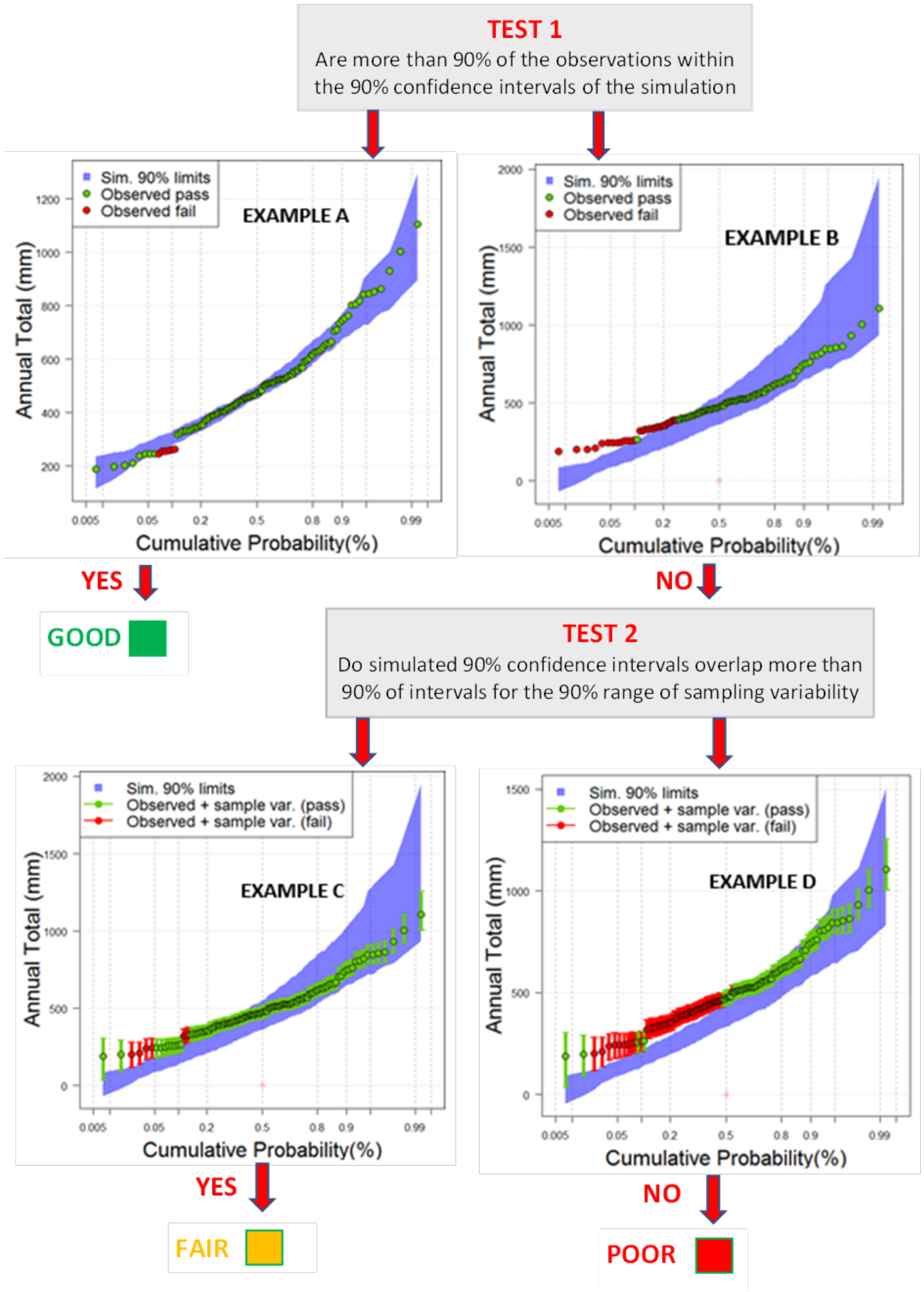
2.4.3 Pooling performance over multiple sites

Table 2 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 of Overall Fair-Good, Overall Fair-Poor and Overall Variable are used, according to the rules outlined in Table 2.

Table 3 Aggregate performance categorisation criteria

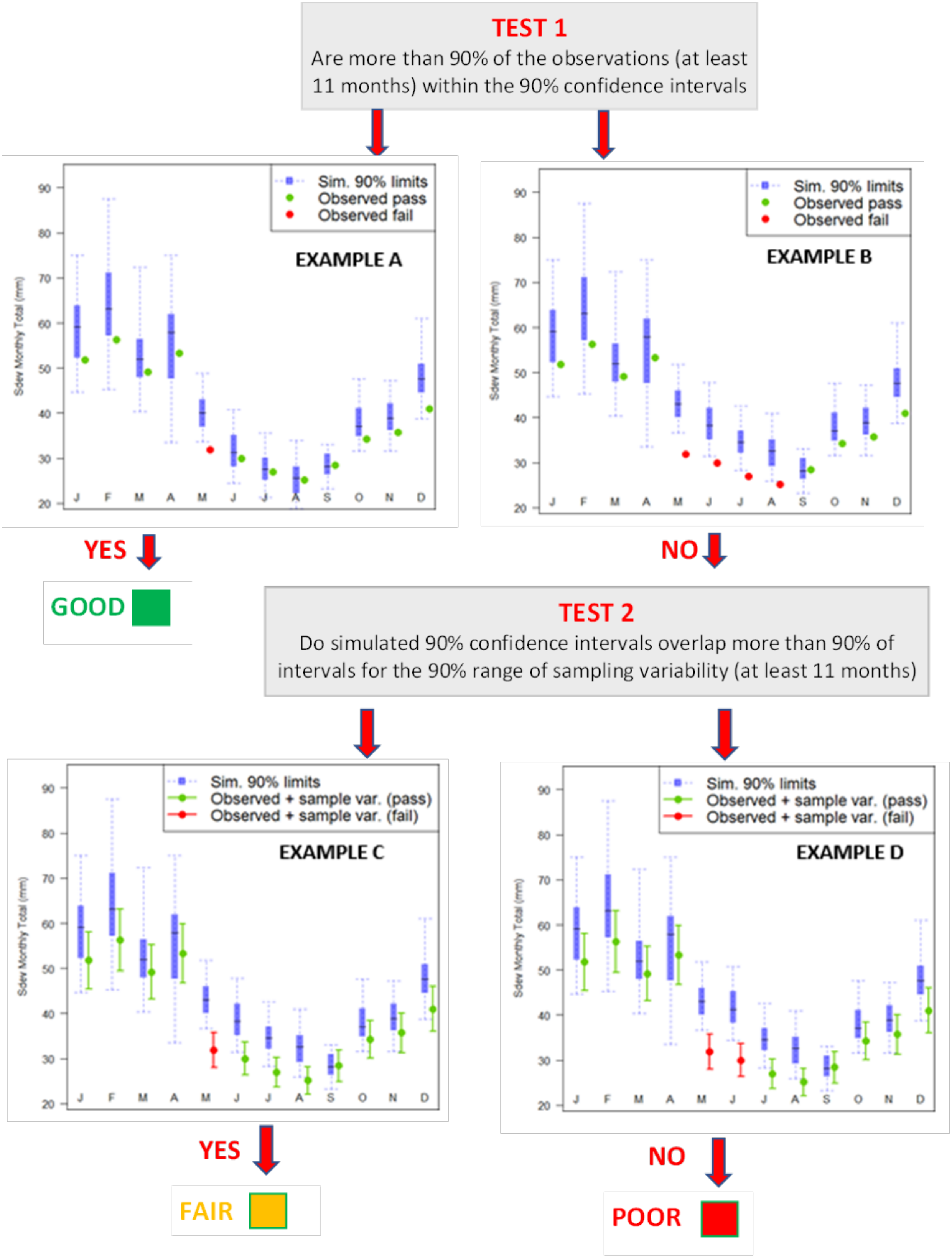
Overall performance category	Definition: Sites in performance category (%)	Example: Good model performance (%)	Example: Fair model performance (%)	Example: Poor model performance (%)
Overall Good	Good is > 50%	85	10	5
Overall Fair	Fair is > 50%	5	85	10
Overall Poor	Poor is > 50%	10	5	85
Overall Fair – 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 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 10 Flow chart of performance classification for annual distributions into 3 categories (Good, Fair, Poor) using the criteria specified, according to 2 tests



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

3 References

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.

Leonard M, Westra S, Thyer M and Bennett B (2019), Multisite rainfall and evaporation data generation for the Macquarie Valley, Report 2019/01, report prepared by the University of Adelaide for the NSW Department of Industry (unpublished).