# Reliability of Daily and Annual Stochastic Rainfall Data Generated from Different Data Lengths and Data Characteristics

[1]Chiew, F.H.S., [2]R. Srikanthan, [2]A.J. Frost and [1]E.G.I. Payne

[1]Department of Civil and Environmental Engineering, The University of Melbourne
[2]Australian Bureau of Meteorology
**Email:** f.chiew@civenv.unimelb.edu.au

***Keywords:*** *stochastic data, stochastic model, annual rainfall, daily rainfall, climate variability, Australia*

**EXTENDED ABSTRACT**

This paper assesses the performance of the single-site stochastic daily rainfall model, TPMb, using data from 101 locations across Australia, as a function of the historical rainfall characteristics, geographical locations and the length of historical data used to calibrate the model.

TPMb (Transition Probability Matrix with Boughton's correction) is one of the more robust and commonly used stochastic daily rainfall models in Australia. TPMb is one of the stochastic models in SCL (Stochastic Climate Library, www.toolkit.net.au), a software product in the Catchment Modelling Toolkit designed to facilitate the generation of stochastic climate data. The results here therefore also provide a perspective of the model performance that can be expected for different locations and historical data characteristics.

Stochastic rainfall data provide alternative realisations that are equally likely to have occurred, and are often used as inputs into hydrological models to quantify uncertainty in environmental systems associated with climatic variability, allowing informed risk-based design and system operations decisions to be made.

The TPMb model is used to generate 1000 replicates of 100-year daily rainfall time series for 101 locations across Australia, and the model performance is assessed by comparing key statistics in the stochastic replicates with those of the historical data. Two annual rainfall characteristics/statistics (mean annual rainfall and 5-year low rainfall total) and four daily rainfall characteristics (mean wet day rainfall, mean wet spell length, mean maximum 3-day rainfall and mean dry spell length) that are not used in calibrating the model are compared.

The results indicate that TPMb can generally reproduce the historical rainfall data characteristics satisfactorily. The average of the statistic in the 1000 stochastic replicates is generally within 10% of the statistic of the historical data, and the 2.5$^{th}$ percentile and 97.5$^{th}$ percentile of the statistic in the stochastic replicates are almost always lower and higher respectively than the statistic in the historical data.

Some of the main observations are: TPMb slightly overestimates the mean annual rainfall and mean wet day rainfall (but only when the coefficient of variation of annual rainfall is smaller than 0.4); the higher overestimations are in the southern parts of Australia; TPMb simulation of mean maximum 3-day rainfall is poorer for smaller mean maximum 3-day rainfall (<20 mm); TPMb simulation of mean dry spell length is poorer for higher mean dry spell length (>20 days); apart from these, there is no obvious relationship between the model performance and the historical data characteristics or geographical locations; and long historical records are required to derive reliable stochastic rainfall data, particularly the statistics that reflect longer-term variability.

## 1. INTRODUCTION

Rainfall is a key input into hydrological models that estimate runoff, sediment, pollutant loads and other hydrological fluxes and state variables. Rainfall is highly variable over various time scales – daily, seasonal, inter-annual, inter-decadal and longer time scales. The use of a historical rainfall time series as input into hydrological models provides results that are based on only one realisation of the past climate. Stochastic rainfall data provide alternative realisations that are equally likely to have occurred, and are often used as inputs into hydrological models to quantify uncertainty in environmental systems associated with climatic variability, allowing informed risk-based design and system operations decisions to be made (McMahon et al., 1996).

Stochastic rainfall data are generated from the characteristics of the historical data. In short, stochastic rainfall data are random numbers that are modified so that they have the same characteristics (in terms of mean, variance, skew, long-term persistency, etc…) as the historical data from which they are based. Each stochastic replicate (sequence) is different and has different characteristics compared to the historical data, but the average of each characteristic from all the stochastic replicates should be the same as the historical data.

The performance of a stochastic model is often assessed by comparing the average (and distribution) of each characteristic from all the stochastic replicates with that of the historical data. Different characteristics are important for different applications, for example, extreme rainfall for floods, and dry spell and long-term persistency for droughts. Most stochastic models can reproduce the historical characteristics that are used directly by the model to generate the stochastic data, but are less capable in reproducing characteristics that are not used to calibrate the model. Models also tend to perform better at the time step on which they are based, for example, a daily stochastic rainfall model tends to simulate the daily characteristics better than the annual characteristics. The use of long historical record to calibrate a stochastic model also generally leads to more reliable stochastic rainfall data.

This paper assesses the performance of a single-site stochastic daily rainfall model using data from 101 locations across Australia, in terms of the ability of the model to reproduce the daily and annual historical rainfall characteristics. The paper investigates the model performance as a function of the historical rainfall characteristics, geographical locations and the length of historical data used to calibrate the model. The results from the paper also provide a perspective of the model performance that can be expected for different locations and historical data characteristics.

## 2. SINGLE-SITE STOCHASTIC DAILY RAINFALL MODEL – TPMb

The single-site stochastic daily rainfall model, TPMb (Transition Probability Matrix with Boughton’s correction), is used in this study. TPMb is one of the more robust and commonly used single-site stochastic daily rainfall models, particularly in Australia. TPMb is one of the stochastic models in SCL (Stochastic Climate Library, www.toolkit.net.au/scl), which is a software product in the Cooperative Research Centre for Catchment Hydrology’s Catchment Modelling Toolkit, designed to facilitate the generation of stochastic climate data. A brief description of TPMb is given here, whilst more detailed descriptions of the model can be found in Siriwardena et al. (2002).

In TPMb, daily rainfall can occur in one of up to seven states (state 1 is dry, state 2 is rainfall from 0.1-0.9 mm, state 3 is 0.9-2.9 mm, state 4 is 2.9-6.9 mm, state 5 is 6.9-14.9 mm, state 6 is 14.9-30.9 mm and state 7 is >30.9 mm). The shifted Gamma distribution is used to model rainfall amounts in the unbounded last state, while a linear distribution is used for states 2 to 6. The parameters in the model, which are estimated from the historical data, are therefore the transition probabilities of being in a particular state given the state on the previous day, and the two parameters of the Gamma distribution for the last state. The seasonality in occurrence and magnitude of daily rainfall are taken into account by considering each month separately.

To ensure that TPMb reproduces the inter-annual variability in the historical data, the generated daily rainfall is scaled by an adjustment factor, $[\{G + (Ti - G) F\} / Ti]$, where G is the generated mean annual rainfall, Ti is the stochastically generated annual rainfall for year i in the replicate, and F is the ratio of the standard deviation of annual historical rainfall and average of the standard deviations of the stochastically generated replicates.

## 3. DATA

100 years of daily rainfall time series data (1902-2001) from 101 locations across Australia are used for this study (see map in Figure 1).

The 101 sites are chosen to provide a good coverage across Australia. Ninety-nine of the 101 sites are in the list of high quality Australian Bureau of Meteorology rainfall sites identified by Lavery et al. (1992), with the two extra sites added to aid spatial coverage in the Northern Territory.

Sixty-one of the 101 sites have complete records, while the other 40 sites have less than eight years affected by missing data. To obtain a complete record for the 40 sites, the SILO patch point data sets (www.nrm.qld.gov.au/silo/ppd/ and Jeffrey et al., 2001) are used. Given the small amount of missing data in the historical records, the SILO data are considered to accurately represent the climatic variability at the sites.

## 4. MODELLING EXPERIMENTS AND RESULTS

TPMb is used to generate 1000 replicates of 100 years of daily rainfall time series for the 101 locations. The plots in Figures 1, 2 and 3 compare some of the key statistics in the stochastically generated replicates with those of the historical rainfall data.

Figures 1 and 2 show comparisons of the mean annual rainfall and 5-year low rainfall total statistics respectively. The y-axis gives the percentage difference between the statistic in the stochastically generated replicates and that of the historical data. The solid squares show results for the average of the statistic from the 1000 replicates for the 101 locations, and the dotted lines show results for the 2.5$^{th}$ and 97.5$^{th}$ percentiles of the statistic from the 1000 replicates. Figures 1a and 2a show the results plotted against mean annual rainfall, and Figures 1b and 2b show the results plotted against the coefficient of variation (Cv, standard deviation of annual rainfall divided by mean annual rainfall). Figures 1c and 2c show results for the model calibrated against only the last 25 years of data (1977-2001) – except for Figures 1c and 2c, results throughout the paper are presented for the model calibrated against all 100 years of data (1902-2001). The maps in Figures 1d and 2d summarise the difference between the average of the statistic from the 1000 replicates and that of the historical data, for the 101 locations across Australia.

The plots in Figure 3 compare four daily rainfall statistics from the stochastically generated replicates with those of the historical data, for January and for July – mean wet day rainfall, mean wet spell length, mean maximum 3-day rainfall and mean dry spell length. The results are presented as in Figures 1 and 2, with each plot plotted against the statistic being compared.

## 5. DISCUSSION OF RESULTS

### 5.1 Annual rainfall characteristics

TPMb is a stochastic daily rainfall model, and except for the standard deviation, the annual characteristics of the historical data are not used to calibrate the model. As described at the end of Section 2, the stochastically generated daily rainfall data are adjusted to reproduce the historical standard deviation of annual rainfall, and therefore a comparison of the standard deviations of the stochastic replicates and historical data is not presented here.

The results indicate that TPMb slightly overestimates the mean annual rainfall (Figure 1a). This overestimation is less than 5% in all the 101 locations modelled here, and less than 3% in two thirds of the locations, with all the higher overestimations occurring in the southern parts of Australia (Figure 1d). In about 5% of the locations, the 2.5$^{th}$ percentile mean annual rainfall from the 1000 replicates is greater than the historical mean annual rainfall. There is no relationship between the model performance and the mean annual rainfall (Figure 1a), but the percentage difference between the average of the mean annual rainfalls from the stochastic replicates and the historical mean annual rainfall is smaller for higher Cvs, particularly Cvs greater than 0.4 (Figure 1b). There is also a larger spread in the mean annual rainfall in the 1000 stochastic replicates for locations with higher Cvs.

TPMb generally reproduces the 5-year low rainfall total, with the average of the 5-year low rainfall totals from the 1000 replicates being within 10% of the 5-year low rainfall total of the historical data in more than 80% of the locations and within 5% in more than half of the locations (Figure 2a). In about 5% of the locations, the 2.5$^{th}$ percentile 5-year low rainfall total from the 1000 replicates is greater than the 5-year low rainfall total of the historical data. There is no obvious relationship between the model performance and mean annual rainfall (Figure 2a), Cv of annual rainfall (Figure 2b) and geographical location (Figure 2d). The 5-year low rainfall total reflects persistent low rainfall conditions over several years, and is an important characteristic in water resources systems and drought studies.

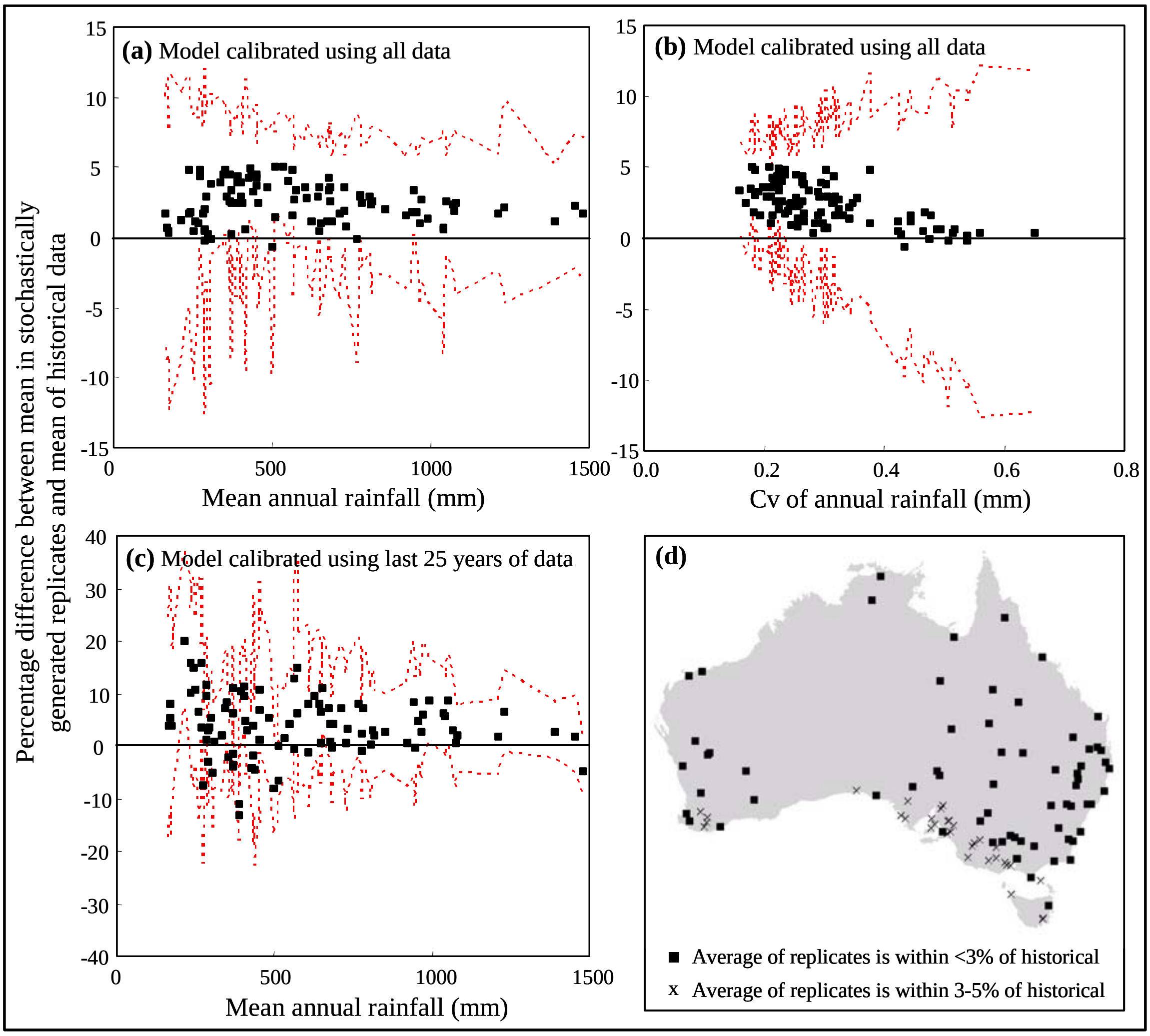


**Figure 1.** Comparison of mean annual rainfall in the stochastic replicates and mean annual rainfall of the historical data (see Section 4 for description of plots)

The plots in Figures 1c and 2c illustrate limitations in the use of short historical records to calibrate a stochastic model. They highlight that the characteristics in a short historical record can be very different from those derived from a longer historical record. In particular, it is difficult to obtain a meaningful 5-year low rainfall total statistic from a 25-year historical rainfall series (Figure 2c). As such, stochastic data derived from long historical records will be more reliable than those derived from short records, particularly for statistics that reflect longer-term variability (assuming stationarity in the climate).

### 5.2 Daily rainfall characteristics

The plots in Figure 3 compare four daily rainfall statistics from the stochastically generated replicates with those of the historical data, for January and for July. The TPMb model simulates daily rainfall using the transition probability concept (see Section 2), and does not use these four daily rainfall characteristics directly in calibrating the model. The mean wet day rainfall, mean wet spell length and mean maximum 3-day rainfall are important in the estimation of runoff volume, with the mean maximum 3-day rainfall also important in estimating extreme runoff amounts. The mean dry spell length affects daily low flow characteristics that may impact on various ecosystems.

The results indicate that TPMb slightly overestimates the mean wet day rainfall (also reflected in the mean annual rainfall, as discussed in Section 5.1). This overestimation is less than 10% in all the 101 locations, and less than 5% in more than 80% of the locations. The percentage difference between the average mean wet day rainfalls in the stochastic replicates and the mean wet day rainfall of the historical data is smaller at locations with higher mean wet day rainfall.

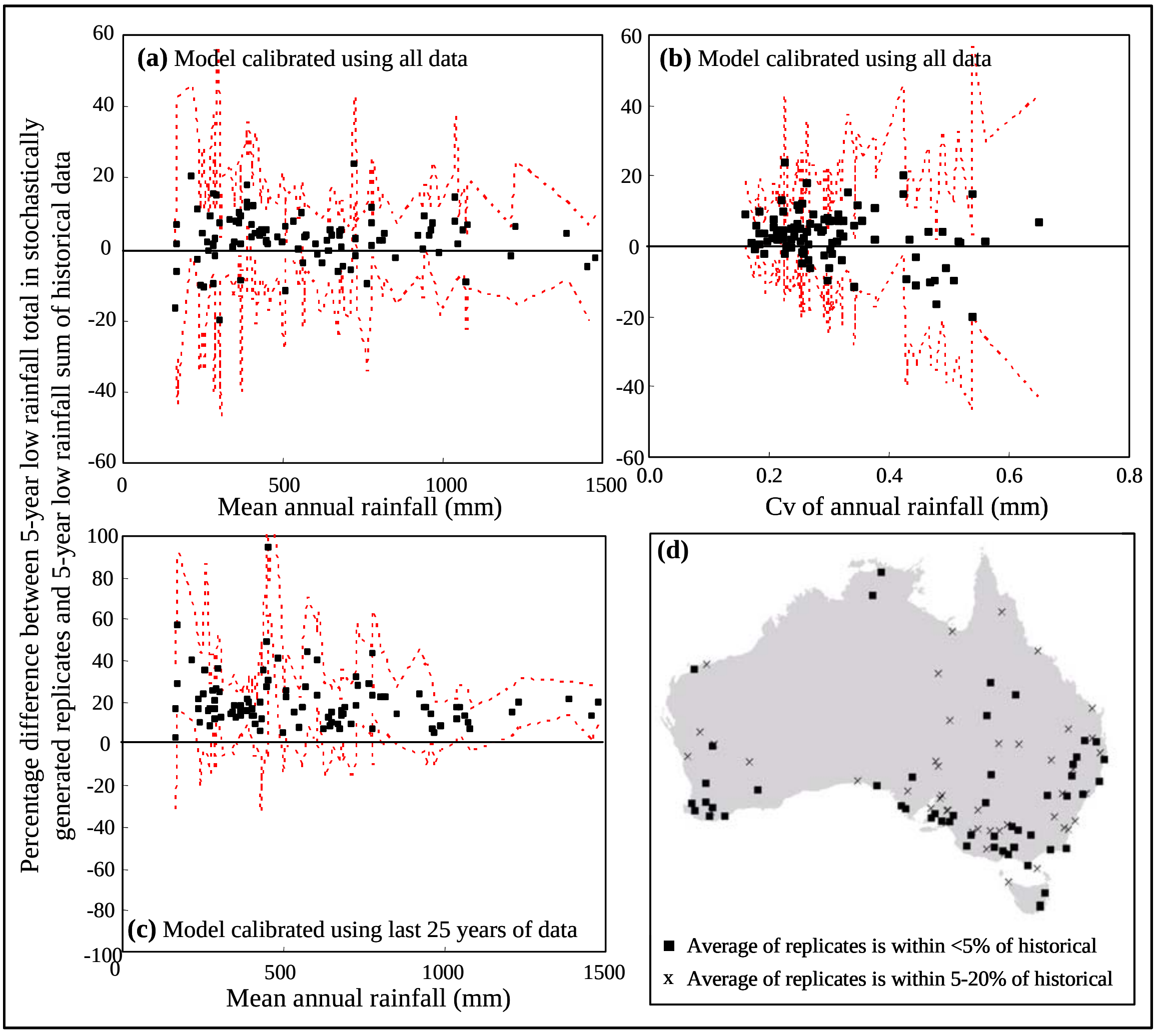


**Figure 2.** Comparison of 5-year low rainfall total in the stochastic replicates and 5-year low rainfall total of the historical data (see Section 4 for description of plots)

The TPMb model simulates the mean wet spell length very well with the average of the mean wet spell lengths in the stochastic replicates generally within 2% of the mean wet spell length of the historical data.

The average of the mean maximum 3-day rainfalls in the stochastic replicates is within 10% of the historical value in all the locations with mean maximum 3-day rainfall of more than 20 mm, and within 10% in about 80% of the locations with mean maximum 3-day rainfall of less than 20 mm.

The average of the mean dry spell lengths in the stochastic replicates is also within 10% of the historical value in almost all the 101 locations. The model performance is poorer at locations with longer mean dry spell length, with differences between the average mean dry spell lengths in the stochastic replicates and mean dry spell length of the historical data of more than 20% at three of the five locations with July mean dry spell length of more than 40 days.

## 6. CONCLUSIONS

The paper assesses the performance of the single-site stochastic daily rainfall model, TPMb, using data from 101 locations across Australia. The model performance is assessed by comparing key statistics in the 1000 daily rainfall time series replicates generated by the model with those of the historical data. Two annual rainfall characteristics (mean annual rainfall and 5-year low rainfall total) and four daily rainfall characteristics (mean wet day rainfall, mean wet spell length, mean maximum 3-day rainfall and mean dry spell length) that are not used in calibrating the model are compared.

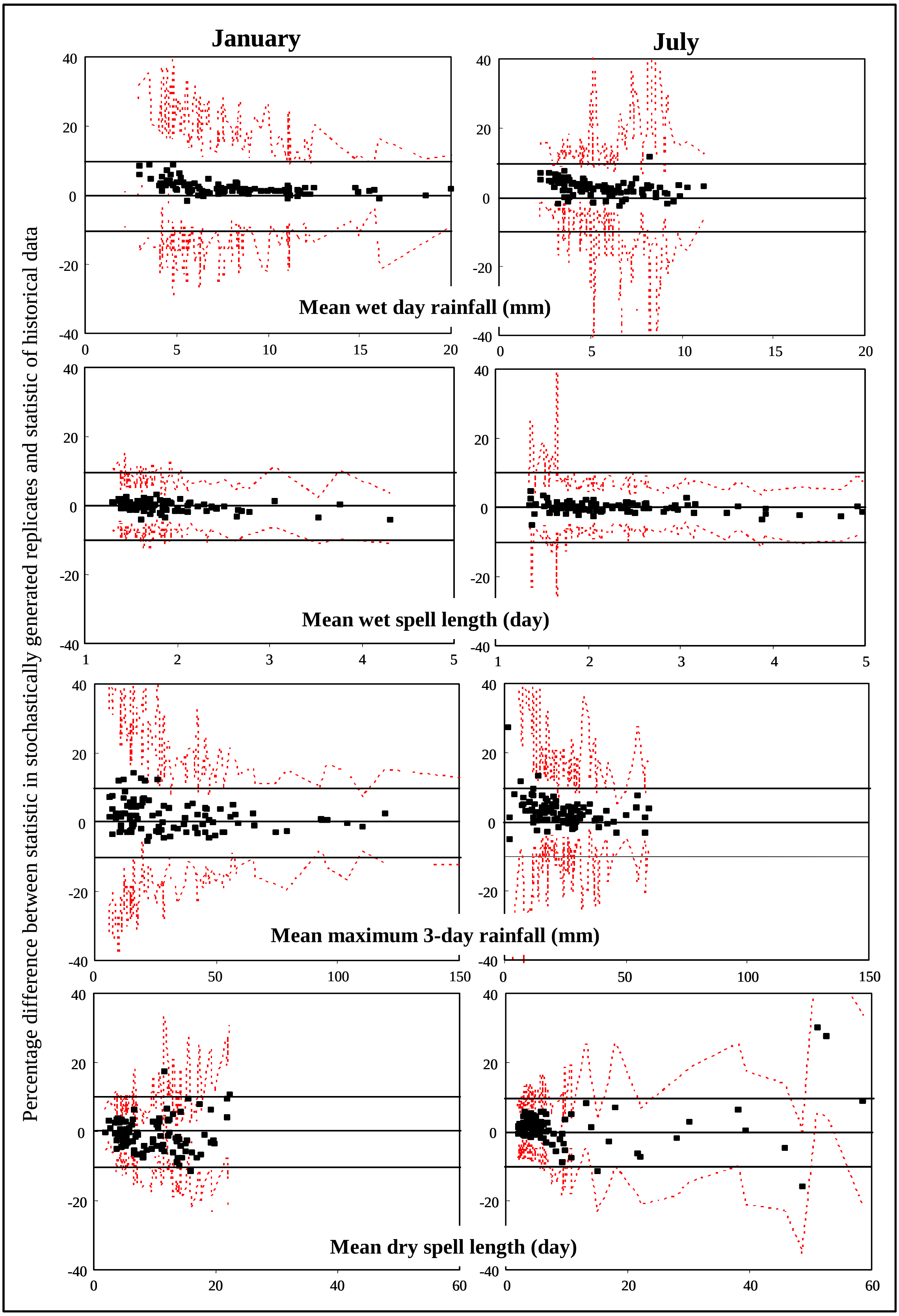


**Figure 3.** Comparison of daily rainfall characteristics in the stochastic replicates and daily rainfall characteristics of the historical data (see Section 4 for description of plots)

The results indicate that TPMb can reproduce the historical rainfall data characteristics satisfactorily. The average of the statistic in the 1000 stochastic replicates is generally within 10% of the statistic of the historical data, and the 2.5$^{th}$ percentile and 97.5$^{th}$ percentile of the statistic in the stochastic replicates are almost always lower and higher respectively than the statistic in the historical rainfall data.

Some of the main observations are: TPMb slightly overestimates the mean annual rainfall and mean wet day rainfall (but only when the Cv of annual rainfall is smaller than 0.4); the higher overestimations are in the southern parts of Australia; TPMb simulation of mean maximum 3-day rainfall is poorer for smaller mean maximum 3-day rainfall (<20 mm); TPMb simulation of mean dry spell length is poorer for higher mean dry spell length (>20 days); apart from these, there is no obvious relationship between the model performance and the historical data characteristics or geographical locations; and long historical records are required to derive reliable stochastic rainfall data, particularly the statistics that reflect longer-term variability.

TPMb is one of the more robust and commonly used single-site stochastic daily rainfall models in Australia. TPMb is also one of the stochastic models in the software SCL in the Catchment Modelling Toolkit. The results here therefore also provide a perspective of the model performance that can be expected for different locations and historical data characteristics.

## 6. ACKNOWLEDGEMENTS

This study was carried out as part of the Cooperative Research Centre for Catchment Hydrology's Climate Variability Program. We like to thank Alan Beswick and Keith Moodie from the Queensland Department of Natural Resources and Mines for providing the patch point rainfall data.

## 7. REFERENCES

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

Lavery, B., Kariko, A. and Nicholls, N. (1992), A historical rainfall data set for Australia, Australian Meteorological Magazine, 40: 33-39.

McMahon, T.A., Pretto, P.B. and Chiew, F.H.S. (1996), Synthetic hydrology – where's the hydrology? Stochastic Hydraulics '96 (Editors: Tickle et al.), Proceedings of the Seventh IAHR International Symposium, Mackay, Queensland, July 1996, A.A. Balkema, Rotterdam, pp. 3-14.

Siriwardena, L., Srikanthan, R. and McMahon, T.A. (2002), Evaluation of Two Daily Rainfall Data Generation Models, Cooperative Research Centre for Catchment Hydrology, Technical Report 02/14, 75 pp.