Stochastic Generation of Daily Rainfall Data

Srikanthan, R.
CRC for Catchment Hydrology, Bureau of Meteorology, Melbourne, Australia,

E-Mail: r.srikanthan@bom.gov.au

Keywords: *Stochastic generation; daily rainfall; nested model*

EXTENDED ABSTRACT

As the historical record provides a single realisation of the underlying climate, stochastically generated data are used to assess the impact of climate variability on water resources and agricultural systems. A widely used approach in other parts of the world to modelling daily rainfall has been a two part model in which the first part describes the rainfall occurrence (dry-wet) process and the second part describes the distribution of rainfall amounts on wet days. Even though the model preserves the daily rainfall characteristics, the monthly and annual characteristics are not preserved. Recently, a daily monthly mixed algorithm (Wang and Nathan, 2002) was proposed to preserve the monthly rainfall characteristics explicitly. However, the model fails to preserve the annual rainfall characteristics. By nesting the two-part daily model in monthly and annual models, the characteristics of rainfall at daily, monthly and annual levels can be preserved (Srikanthan 2004).

The transition probability matrix (TPM) method with Boughton's adjustment has been shown to preserve most of the statical characteristics of the rainfall data and is widely used in Australia. However, it has been found that this approach slightly overestimates the mean annual rainfall and fails to preserve the monthly serial correlation. To overcome these deficiencies, the TPM model is nested in monthly and annual models so that the characteristics of rainfall at daily, monthly and annual levels are preserved (Srikanthan 2005). As an alternative to the nested TPM model, a simple adjustment is proposed to the TPM model to preserve both the annual mean and standard deviation.

The main objective of the paper is to compare the performance of the above three models using a wide variety of daily rainfall data. In earlier studies (Srikanthan 2004, 2005), comparisons were made using only Australian rainfall data. In this paper, data from North America and South Africa were used in addition to Australian data.

A number of statistics at the daily, monthly and annual time scales were used to assess the performance of the models. The results showed that all the models preserved the daily statistics well while the nested models also preserved the annual (Figure 1) and monthly correlations and the skewness of annual rainfall data.

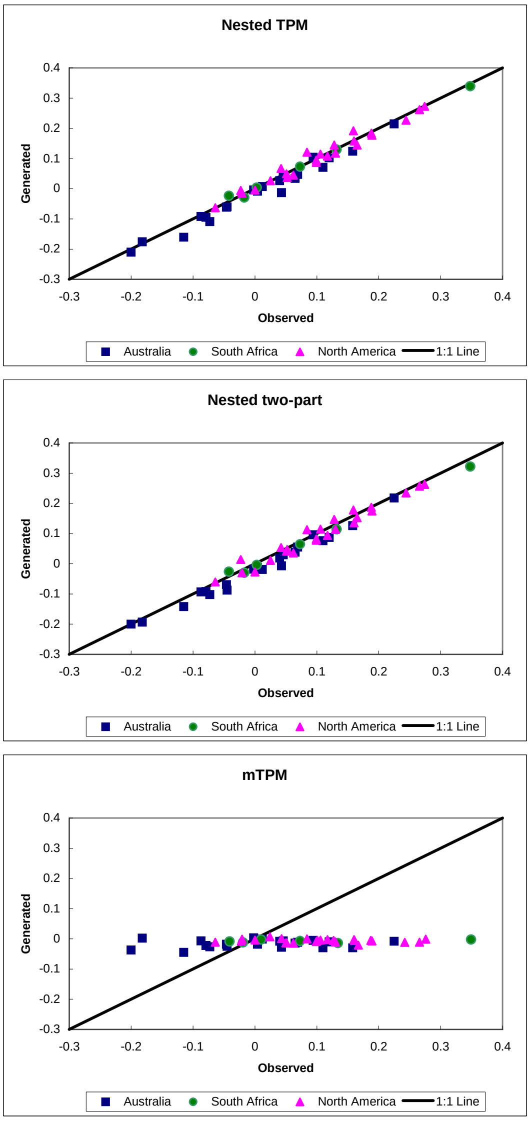

Figure 1. Comparison of annual lag one autocorrelation coefficients.

1. INTRODUCTION

Daily rainfall is a major input to the design of water resources and agricultural systems. As historical data provides only one realisation of the underlying climate, stochastically generated data is used to assess the impact of climate variability on water resources and agricultural systems. Rainfall data generation is a well researched area in the hydrological and climatological literature (Buishand 1978; Chapman, 1997; Sharma and Lall 1999; Srikanthan and McMahon 1985; Srikanthan and McMahon 2001; Woolhiser 1992).

A common approach to modelling daily rainfall has been a two part model in which the first part describes the rainfall occurrence (dry-wet) process and the second part describes the distribution of rainfall amounts on wet days (Woolhiser, 1992). Rainfall occurrence is represented in two ways: either as a Markov process, the assumption being that the rainfall state on the next day is related to the state of rainfall on a finite number of previous days; or as an alternating renewal process for dry and wet sequences, the approach being to stochastically generate the dry and wet spell lengths. Once a day has been specified as wet, rainfall amount is then generated using a Gamma or mixed Exponential distribution. Even though the model preserves the daily rainfall characteristics, the monthly and annual characteristics are not preserved. Wang and Nathan (2002) proposed a daily monthly mixed algorithm to preserve the monthly rainfall characteristics explicitly. In this model, two daily rainfall sequences are generated using daily and monthly parameters and the daily rainfall sequences generated from the daily parameters are adjusted using the other sequence generated from the monthly parameters. This adjustment ensures that the monthly characteristics are preserved in the generated daily rainfall sequences. However, the model fails to preserve the annual rainfall characteristics. A nested two-part daily rainfall model was developed to preserve the daily, monthly and annual characteristics (Srikanthan 2004).

The transition probability matrix (TPM) model (Srikanthan and McMahon, 1985) is widely used in Australia for stochastic generation of daily rainfall, and it appears to preserve most of the characteristics of daily, monthly and annual rainfall. While it performs better than many alternative models, it consistently under represents the variances of the observed monthly and annual rainfall. Boughton (1999) proposed an empirical adjustment to match the observed annual standard deviation (TPMb). This adjustment improves the variability in the annual rainfall by scaling the rainfall amounts on wet days. However, not all the monthly and annual characteristics are preserved by this model (Srikanthan et al 2003). In order to preserve the monthly and annual characteristics, the TPM model is nested in a monthly annual model. The generated daily rainfall data are used to drive the monthly model and the resulting monthly rainfalls are used to drive an annual model (Srikanthan 2005).

The TPMb model consistently overestimates the mean rainfall (Srikanthan et al. 2003). A simple adjustment is proposed to correct for the overestimation of mean rainfall and evaluated in this paper. The modified model is referred to as the mTPM.

The main objective of the paper is to compare the performance of the above three models using a wide variety of daily rainfall data. In earlier studies (Srikanthan 2004, 2005), comparisons were made using only Australian rainfall data. In this paper, data from North America and South America were used in addition to Australian data to assess the performance of nested two-part, nested TPM and mTPM models.

2. DAILY RAINFALL DATA

Daily rainfall data from 21 Australian, 24 North American and 6 South African sites were used. The stations are uniformly distributed in each country and represents a wide range of climates ranging from dry climate with annual number of wet days as low as 30 days to wet climates with climate with annual number of wet days as high as 160 days. A brief summary of the daily rainfall data is given in Table 1.

Table 1. Summary of daily rainfall data.

Country	Mean record length (years)	Mean annual Rainfall (mm)
Australia	42 - 125	180 - 1490
North America	74 - 122	215 - 1279
South Africa	85 - 115	214 - 1020

3. NESTED TWO-PART MODEL

In the two-part model, the occurrence of rainfall is determined by using a first order Markov chain using the two transition probabilities: $p_{W|D}$, the conditional probability of a wet day given that the previous day was dry; $p_{W|W}$, the conditional probability of a wet day given that the previous day was wet. The unconditional probability of a wet day can be derived as

$$\pi = \frac{p_{W|D}}{1 + p_{W|D} - p_{W|W}} \tag{1}$$

For wet days, the rainfall depth is obtained from a Gamma distribution whose probability density function is given by

$$f(x) = \frac{(x/\beta)^{\alpha-1}\exp(-x/\beta)}{\beta\Gamma(\alpha)} \tag{2}$$

where α is the shape parameter and β the scale parameter. The mean and variance of the Gamma distribution are given by

$$\mu(x) = \alpha\beta \tag{3}$$

$$\sigma^2(x) = \alpha\beta^2 \tag{4}$$

The seasonality in daily rainfall is taken into account by considering each month separately. Once the daily rainfall is generated for a month, the monthly rainfall is obtained by summing the daily rainfall values. The generated monthly rainfall value, $\widetilde{X}_i$, is modified by using the Thomas-Fiering monthly model to preserve the monthly characteristics

$$\frac{X_i - \mu(X_i)}{\sigma(X_i)} = \rho_{i,i-1}\frac{X_{i-1} - \mu(X_{i-1})}{\sigma(X_{i-1})} + (1 - \rho_{i,i-1}{}^2)^{1/2}\frac{\widetilde{X}_i - \mu'(X_i)}{\sigma'(X_i)} \tag{5}$$

where $\rho_{i,i-1}$ is the correlation coefficient between months i and $i-1$. The theoretical mean and variance of the rainfall total, X, over a month of N days is given by

$$\mu(X) = N\pi\alpha\beta \tag{6}$$

$$\sigma^2(X) = N\pi\alpha\beta^2\left[1 + \alpha(1-\pi)\frac{1 + p_{W|W} - p_{W|D}}{1 - p_{W|W} + p_{W|D}}\right] \tag{7}$$

The subscript i for all the variables in Eq. (6) and (7) is omitted for clarity. The generated daily rainfall data is multiplied by the ratio $X_i / \widetilde{X}_i$.

Once the values for the twelve months of a year (k) have been generated, the generated monthly values can be aggregated to obtain the annual value. The aggregated annual value, $\widetilde{Z}_k$, is modified by using a lag one autoregressive model to preserve the annual characteristics.

$$\frac{Z_k - \mu(Z\ )}{\sigma(Z\ )} = \rho(Z\ )\frac{Z_{k-1} - \mu(Z\ )}{\sigma(Z\ )} + [1 - \rho^2(Z\ )]^{1/2}\frac{\widetilde{Z}_k - \mu'(Z\ )}{\sigma'(Z\ )} \tag{8}$$

where ρ is the lag one autocorrelation coefficient. If the annual rainfall data exhibits significant skewness, then the noise term in Eq. (8) is modified by using the Wilson-Hilferty transformation (1931). The theoretical values of the mean and variance of the aggregated annual rainfall are given by

$$\mu(Z\ ) = \sum_{j=1}^{12}\mu(X_j) \tag{9}$$

$$\begin{aligned}\sigma^2(Z) \approx &\sum_{j=1}^{12}\sigma^2(X_j) + 2\sum_{j=2}^{12}\sigma(X_j)\sigma(X_{j-1})\rho_{j,j-1} \\ &+ 2\sum_{j=3}^{12}\sigma(X_j)\sigma(X_{j-2})\rho_{j,j-1}\rho_{j-1,j-2} \\ &+ 2\sum\sigma(X_j)\sigma(X_{j-3})\rho_{j,j-1}\rho_{j-1,j-2}\rho_{j-2,j-3}\end{aligned} \tag{10}$$

The generated monthly rainfall value is multiplied by the ratio $Z_k / \widetilde{Z}_k$. This will preserve the annual characteristics. The modified monthly rainfall values are used to adjust the daily rainfall values. Rather than adjusting the daily rainfall values twice, the adjustment to the daily rainfall values can be carried out in one step by multiplying the generated rainfall values for each month (i) by the ratio $X_i Z_k / \widetilde{X}_i \widetilde{Z}_k$.

4. NESTED TPM MODEL

The daily rainfall data are first generated by the TPM model. The seasonality in occurrence and magnitude of daily rainfall is taken into account by considering each month separately. The daily rainfall is divided into a number of states, up to a maximum of seven states. State 1 is dry (no rainfall) and the other states are wet. The state boundaries for rainfall amounts are given in Table 2. If the number of states for a month is less than seven, then the upper limit of the last state is assumed to be infinite.

The shifted Gamma distribution is used to model the rainfall amounts in the highest state, while a linear distribution is used for the intermediate states. The latter is chosen because daily rainfall usually exhibits a reverse J shape distribution. The parameters of the Gamma distribution are estimated by using Eq. (3) and (4).

Table 2. State boundaries used for the TPM model.

State number	Upper state boundary limit (mm)
1	0.0
2	0.9
3	2.9
4	6.9
5	14.9
6	30.9
7	∞

The transition probabilities are estimated from

$$p_{ij}(k) = \frac{f_{ij}(k)}{\sum_{j=1}^{C} f_{ij}(k)} \quad i, j = 1,\ldots, C;\ k = 1,\ldots, 12 \quad (11)$$

where $f_{ij}(k)$ is the historical frequency of transition from state i to j within month k and C the number states.

Once the daily rainfall is generated for a month, the monthly rainfall is obtained by summing the daily rainfall values. As above, the generated monthly rainfall is modified by using Eq (5). Expressions for the mean and standard deviation of monthly rainfall obtained from the TPM model are not available. Hence, these are estimated from a number of the generated monthly totals and averaged. Adjusted monthly values are then summed to obtain the annual value and adjusted using Eq (8). Finally, the generated daily rainfall was adjusted with respect to the adjusted monthly and annual rainfall values as before.

5. MODIFIED TPM MODEL

Boughton (1999) applied an adjustment to match the standard deviation of the observed annual rainfall. However, it was noted that the resulting sequences over estimated the mean. In the modified TPM model, the generated daily rainfall values are adjusted with respect to both the mean and standard deviation of annual rainfall. An adjustment factor (F) is first obtained from

$$F = \frac{stdev_o}{stdev_g} \quad (12)$$

The standard deviation of the generated annual rainfall is estimated from a number of replicates and averaged. The generated daily rainfall in each year is multiplied by the following ratio:

$$Ratio_i = \frac{\{H + (T_i - G)F\}}{T_i} \quad (13)$$

where G is the generated mean annual rainfall, H the historical mean annual rainfall and T_i the generated annual rainfall for year i.

6. DISCUSSION OF RESULTS

One hundred replicates, each of length equal to the historic record were generated using the above four models for all the 50 stations. The number of states for the North American and South African stations was first decided based on the mean monthly rainfall. It was then adjusted if there were not enough items (> 20) in the largest state. The number of states finally selected for the North American and South African stations is not presented due to lack of space and is available from the author. The number of states for the Australian stations was selected using the guidance given in Srikanthan and McMahon (1985) and is available in Srikanthan (2005).

The performance of the models is evaluated using a number of statistics at the daily, monthly and annual levels. The daily, monthly and annual statistics used are listed in the following sections. Due to lack of space, only a few results are presented here. An overall assessment of the results is presented in Table 4.

6.1. Daily statistics

The daily statistics include:

- Mean, standard deviation and coefficient of skewness daily rainfall
- mean daily rainfall for different types of wet days; solitary (class 1) , bounded only on one side by a wet day (class 2), bounded on both sides by wet days class 3)
- correlation between rainfall depth and duration of wet spells
- mean number of wet days
- maximum daily rainfall
- mean, standard deviation and coefficient of skewness of dry spell length
- mean, standard deviation and coefficient of skewness of wet spell length

All the models preserved the mean and standard deviation of daily rainfall. None of the models preserved the skewness when it was larger than about 6. The mean daily rainfall for different types of wet days was preserved reasonably well except for class 3 when it was larger than 20 mm. All the models preserved the correlation between the rainfall depth and duration. The mean and standard deviation of the dry and wet spells were preserved by all the models. However, the skewness was not preserved by any of the models for wet spell while for dry spell none of the models could preserve the large skewness values (> 6). Shorter maximum dry (< 100) and wet (< 15) spell lengths were preserved but the longer ones were not.

6.2. Monthly statistics

The monthly statistics include:

- mean, standard deviation, coefficient of skewness and serial correlation of monthly rainfall
- maximum and minimum monthly rainfall
- mean number of months of no rainfall

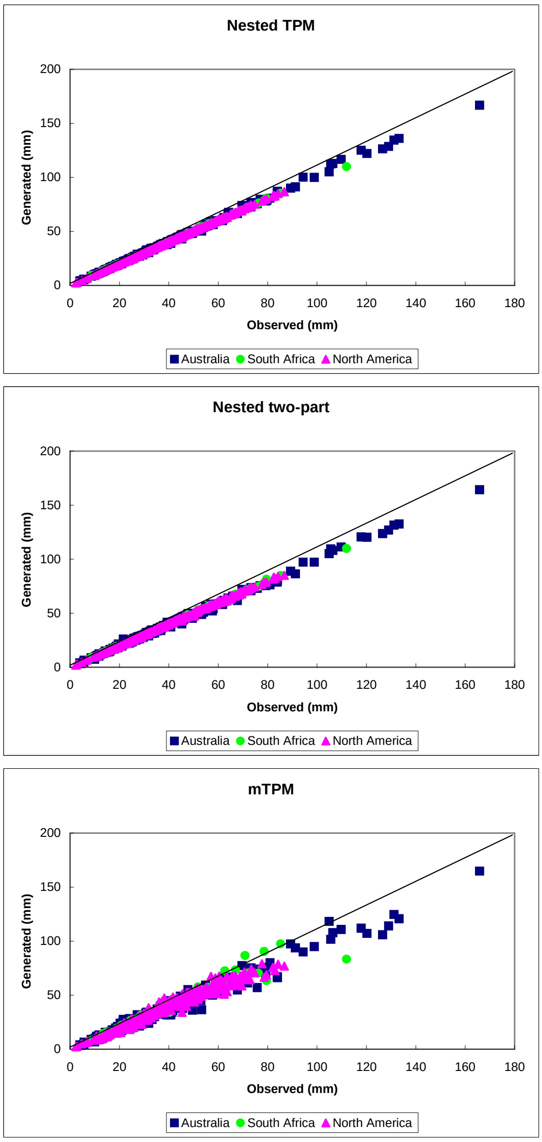

Figure 2. Comparison of standard deviation of monthly rainfall.

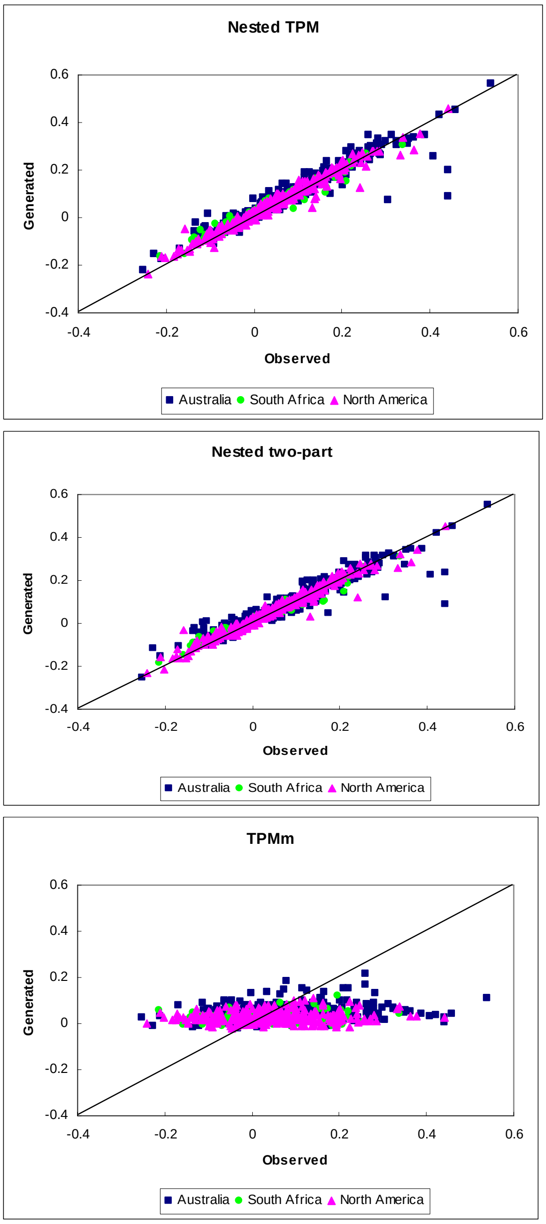

Figure 3. Comparison of monthly correlations.

All the models preserved the monthly means well. The two nested models preserved the standard deviation better than the mTPM due to nesting (Figure 2). Smaller skewness values are preserved but not the larger ones by all the models. The nested models preserved the correlation while the mTPM did not and resulted in almost zero correlation (Figure 3). All the models preserved the number of months of no rainfall.

6.3. Annual statistics

The annual statistics include:

- mean annual rainfall
- standard deviation of annual rainfall
- coefficient of skewness of annual rainfall

- lag one auto correlation
- maximum annual rainfall
- 2-, 5- and 10-year low rainfall sums
- mean annual number of wet days

All the models preserved all the annual statistics except the skewness and lag one autocorrelation. The two nested models preserved the skewness and lag one autocorrelation (Figure 1) while the mTPM model failed to preserve them. However, mTPM model eliminated the slight overestimation of the mean which was a problem with the TPM model with the Boughton's correction (TPMb).

7. CONCLUSIONS

The TPM and two-part models were nested in monthly and annual models to preserve the monthly and annual characteristics. The original TPM model is also modified to match the annual mean and standard deviation. The two nested and the mTPM models were used to generate daily rainfall data for a number of sites in Australia, South Africa and North America. The results showed that all the models preserved the daily statistics except the skewness of the wet spell. The nested models also preserved all the monthly and annual statistics while the mTPM failed to preserve the monthly and annual correlations and annual skewness. However, mTPM is an improvement over TPMb in terms of preserving the annual mean. In general, the nested TPM model performed marginally better than the nested two-part model. However, the nested TPM model needs longer data (generally greater than 30 years) to estimate the transition probabilities properly. For long historical data, the nested TPM and for short historical data, the nested two-part models are recommended for the stochastic generation of daily rainfall data.

8. ACKNOWLEDGMENTS

The South African and North American data were provided by Tom Chapman, University of New South Wales. The South African data were made available by Dr M Deny from the records held at the Computing Centre for Water Research, University of Natal. Most of the North American records were provided by Dr Ben Harding of Hydrosphere Resource Consultants, Boulder, CO. The records for Corvallis and Vancouver, Oregan were given by Jinfan Duan of Oregan State University, from records maintained by the US Forest Service. The data for Montreal were provided by Dr Fred Fabry.

9. REFERENCES

Boughton, W. C. (1999) A daily rainfall generating model for water yield and flood studies. Report 99/9, CRC for Catchment Hydrology, Monash University, Melbourne, 21pp.

Buishand, T.A. (1978), Some remarks on the use of daily rainfall models, *Journal of Hydrology*, 36, pp. 295-308

Chapman, T.G. (1997,) Stochastic models for daily rainfall in the Western Pacific, *Mathematics and Computers in Simulation*, 43, pp. 351-358.

Sharma, A. and U. Lall (1999), A nonparametric approach to daily rainfall simulation, *Mathematics and Computers in Simulation*, 48, pp. 367-371.

Srikanthan, R. (2004), Stochastic generation of daily rainfall data using a nested model. *57th Canadian Water Resources Association Annual Congress*, 16-18 June, 2004, Montreal, Canada.

Srikanthan, R. (2005), Stochastic generation of daily rainfall using a nested transition probability model. *29th Hydrology and Water Resources Symposium*, Engineers Australia, 20 – 23 February, 2005, Canberra

Srikanthan, R. and T.A. McMahon. (1985), *Stochastic generation of rainfall and evaporation data*. AWRC Technical Paper No. 84, 301pp.

Srikanthan, R. and T.A. McMahon (2001), Stochastic generation of annual, monthly and daily climate data: A review, *Hydrology and Earth System Sciences*, 5(4), pp. 653-670.

Srikanthan R, T.I. Harrold, A. Sharma and T.A .McMahon (2003), Comparison of two approaches for generation of daily rainfall data, *MODSIM2003*, Townsville, 14 – 17 July 2003, pp. 106 – 111.

Wang, Q.J. and R.J. Nathan (2002) A daily and monthly mixed algorithm for stochastic generation of rainfall time series. *Hydrology and Water Resources Symposium*, Melbourne, 20-23 May 2002.

Wilson, E.B. & M.M. Hilferty (1931), Distribution of Chi-square. *Proc. National Academy of Science*, 17, pp. 684-688.

Woolhiser D.A. (1992), 'Modeling daily precipitation -progress and problems'. In: A.T. Walden and P. Guttorp (Editors), *Statistics in the environmental and earth sciences*. Edward Arnold, London, U.K., 306 pp.

Table 4. Evaluation of the three daily rainfall data generation models.

	Nested TPM			Nested two-part			mTPM		
Daily	Aus	SA	NA	Aus	SA	NA	Aus	SA	NA
Mean number of wet days	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum daily rainfall	✓	✓-	✓	✓	✓-	✓	✓	✓-	✓
Mean daily rainfall	✓	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation	✓	✓	✓	✓	✓	✓	✓	✓	✓
Skewness	✓	✓-	✓	✓-	✓-	✓-	✓	✓-	✓
Mean on class 1 wet day	✓-	✓	✓	✓-	✓	✓	✓-	✓	✓
Mean on class 2 wet day	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean on class 3 wet day	✓-	✓	✓	✓-	✓	✓	✓-	✓	✓
Correl b/w depth & duration	✓-	✓	✓	✓-	✓	✓	✓-	✓	✓
Mean dry spell length	✓	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation	✓	✓	✓	✓	✓	✓	✓-	✓-	✓
Skew ness	✓	✓-	✓	✓	✓-	✓	✓	✓-	✓
Mean wet spell length	✓	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation	✓	✓	✓	✓	✓	✓	✓	✓	✓
Skewness	×	×	×	×	×	×	×	×	×
Max dry spell length	✓	✓-	✓	✓-	✓-	✓	✓-	✓-	✓
Max wet spell length	✓	✓	✓	✓	✓	✓	✓	✓	✓
Monthly									
Mean monthly rainfall	✓	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation	✓	✓	✓	✓	✓	✓	✓	✓	✓
Skewness	✓-	✓	✓	✓-	✓	✓	✓-	✓	✓
Correlation	✓	✓	✓	✓	✓	✓	×	×	×
Maximum	✓	✓	✓	✓	✓	✓	✓	✓	✓
Minimum	✓	✓	✓	✓	✓	✓	✓	✓	✓
No of no rainfall months	✓	✓	✓	✓	✓	✓	✓	✓	✓
Annual									
Mean annual rainfall	✓	✓	✓	✓	✓	✓	✓	✓	✓
Standard deviation	✓	✓	✓	✓	✓	✓	✓	✓	✓
Skewness	✓	✓	✓	✓	✓	✓	×	×	×
Correlation	✓	✓	✓	✓	✓	✓	×	×	×
Maximum	✓	✓	✓	✓	✓	✓	✓	✓	✓
Minimum	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted range	✓	✓	✓	✓	✓	✓	✓	✓	✓
2-year low rainfall sum	✓	✓	✓	✓	✓	✓	✓	✓	✓
5-year low rainfall sum	✓	✓	✓	✓	✓	✓	✓	✓	✓
10-year low rainfall sum	✓	✓	✓	✓	✓	✓	✓	✓	✓
Average annual # of wet days	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓- - A few points may be away from the 45 degree line.