# Stochastic Generation of Spatially Consistent Dail Rainfall

**Srikanthan, R.**[1] **and G.G.S. Pegram** [2]

[1] eWater CRC, Hydrology Unit, Bureau of Meteorology, Melbourne, AUSTRALIA
[2] Civil Engineering Programme, University of KwaZulu-Natal, Durban, SOUTH AFRICA
Email: r.srikanthan@bom.gov.au, pegram@ukzn.ac.za

***Keywords:*** *Stochastic generation, spatial daily rainfall, nested multi-site model*

## EXTENDED ABSTRACT

As the historical record only provides a single realisation of the underlying climate variability, stochastically generated data are used to assess the impact of climate variability on water resources and agricultural systems. The generation of climate data at a single site is a well researched area in the hydrological and climatological literature. The assessment of hydrological and land management changes over larger catchments or regions however requires that the spatial dependence between the climate data generated at multiple sites to be preserved. This is particularly important to the simulation of rainfall, which displays the largest variability in time and space. Wilks (1998) proposed a multi-site daily rainfall model using a number of single site two-part models driven by a cross correlated set of random numbers. Even though the model preserved the statistical characteristics at the daily level, it failed to preserve them at the monthly and annual time scales. In order to improve these statistics at higher time scales, the multi-site daily model was nested in multi-site monthly and annual models. The nested model was evaluated using daily rainfall data from two regions.

The first region is the Woady Yaloak Catchment located in southwest Victoria, Australia and has three rainfall stations. The area of the catchment is 1157 km$^2$. Eighty three years of rainfall data were used covering the period 1919 to 2001. The second region is around Sydney which extends from Newcastle in the north to Canberra in the south. Thirty rainfall stations are in the region with 43 years of rainfall data covering the period 1960 to 2002. One hundred replicates each of length equal to the historical record length were generated for the two regions. A number of statistics at daily, monthly and annual time scales were calculated from each of the replicates and averaged for comparison with the corresponding historical values. The results showed that the nested multi-site model preserved the statistics well including the spatial correlations as shown in Figure 1 for the Sydney region. This shows that the nested multi-site model is effective in preserving the spatial correlations in all three time scales.

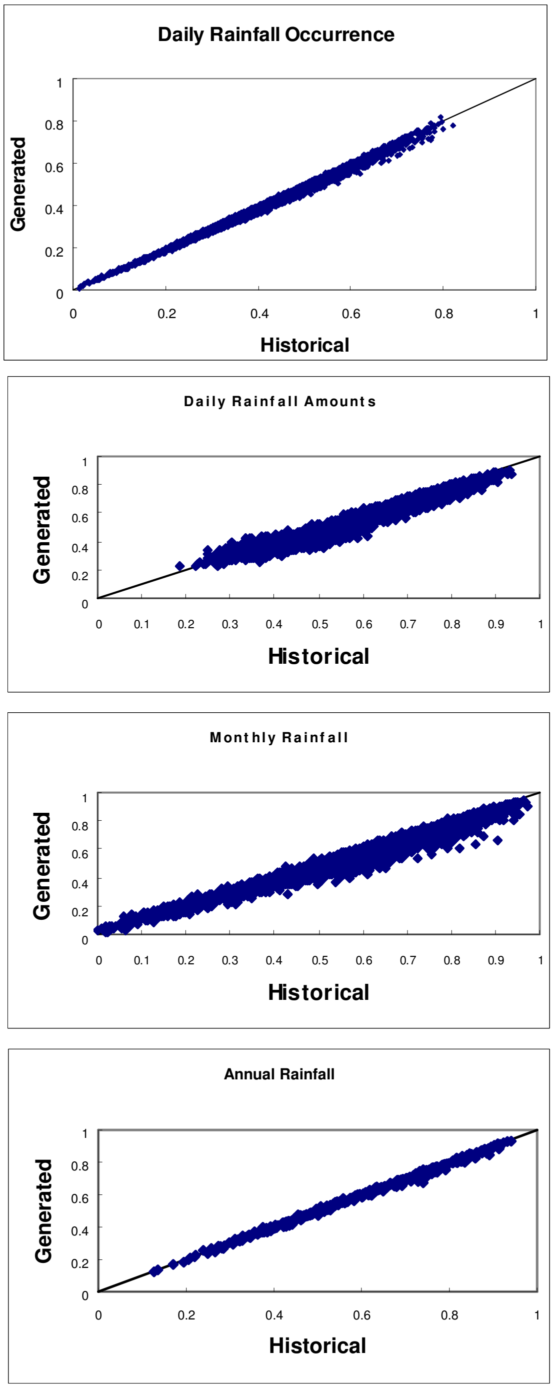


**Figure 1**. Comparison of modelled and observed spatial correlations - Sydney region.

## 1. INTRODUCTION

Climate data, particularly rainfall data, are a major input to water resources and agricultural modelling systems. As the historical record only provides a single realisation of the underlying climate variability, stochastically generated data are used to assess the impact of climate variability on water resources and agricultural systems. The generation of climate data at a single site is a well researched area in the hydrological and climatological literature (Srikanthan and McMahon 2001) and a two-part model has been widely used to generate daily rainfall data. The assessment of hydrological and land management changes over larger catchments or regions however requires that the spatial dependence between the climate data generated at multiple sites to be preserved. This is particularly important to the simulation of rainfall, which displays the largest variability in time and space. There are a number of approaches (conditional models, extension of Markov chain models, random cascade models and nonparametric models) proposed recently to generate rainfall data at multiple sites.

Conditional models generate the occurrence and the amount of rainfall using surface and upper air data (Zucchini and Guttorp, 1991; Bardossy and Plate, 1991, 1992; Wilson and Lettenmaier, 1993; Hughes et al., 1999; Charles et al. 1999). Wilks (1998) extended the familiar two part model, consisting of a two–state, first-order Markov chain for rainfall occurrences and a mixed exponential distribution for rainfall amounts, to generate rainfall simultaneously at multiple locations by driving a collection of individual models with serially independent but spatially correlated random numbers. He applied the model to 25 sites in the New York area. Jothityangkoon et al. (2000) constructed a space-time model to generate synthetic fields of space-time daily rainfall. The model has two components: a temporal model based on a first-order, four-state Markov chain which generates a daily time series of the regionally averaged rainfall and a spatial model based on a nonhomogeneous random cascade process which disaggregates the regionally averaged rainfall to produce spatial patterns of daily rainfall. The cascade used to disaggregate the rainfall spatially is a product of stochastic and deterministic factors; the latter enables the model to capture systematic spatial gradients exhibited by measured data. Buishand and Brandsma (2001) used nearest neighbour resampling for multi-site generation of daily precipitation and temperature at 25 stations in the German part of the Rhine basin. Mehrotra and Sharma (2005) applied the k-nearest neighbour technique to simulate rainfall conditional upon atmospheric variables simultaneously at 30 stations around Sydney.

Conditional models are both data and computationally intensive. All models reviewed (Srikanthan and McMahon 2001) were only applied in one area and were not tested adequately. The random cascade models also require a large amount of data to characterise the spatial dependence at different levels in the cascade as it generates rainfall data over a grid. The nonparametric model is being developed at the University of New South Wales by Mehrotra and Sharma (2005). The extended two part model of Wilks (1998) which is an extension of the Markov chain model appears to be a relatively simple model and at the same time, it has the potential to perform well. A comparison with two other approaches (hidden state Markov model and the k-nearest neighbour model) to model rainfall occurrence has shown that this approach performed the best (Mehrotra et al. 2005). Hence this method was chosen for further development.

The multisite two-part model of Wilks (1998) was nested in single site monthly and annual models and its performance was evaluated in an earlier study (Srikanthan 2005, 2006). The nested model preserved all the at-site statistics and the cross correlations at the daily level but under-estimated the cross correlations between the sites at monthly and annual levels (Srikanthan 2005, 2006). The reason for this is that the nesting was carried out individually at each site. The model is herein further enhanced by nesting the daily generated amounts in a cascade of multi-site monthly and annual models. The performance of the enhanced model was evaluated using data from two regions with 3 and 30 rainfall stations.

## 2. NESTED MULTISITE DAILY RAINFALL MODEL

The nested multi-site daily rainfall model consists of three parts, namely, occurrence, amounts and nesting. These three parts are briefly described below. For a more detailed description and derivation, readers are referred to Srikanthan (2005).

### 2.1. Multi-site rainfall occurrence model

A first-order two-state Markov chain is used to determine the occurrence of rainfall at each site. For each site k, the Markov chain has the two transition probabilities: $p^k_{W|D}$ and $p^k_{W|W}$, respectively the conditional probabilities of a wet day given that the previous day was dry or wet. The individual models are driven by serially independent but cross correlated random numbers to preserve the spatial correlation in the rainfall occurrence process.

Given a network of N locations, there are N(N - 1)/2 pair-wise correlations that should be maintained in the generated rainfall occurrences. This is achieved by using correlated uniform random numbers ($u_t$) in simulating the occurrence process. The cross-correlated uniform variates $u_t(k)$ can be derived from standard Gaussian variates $w_t(k)$ through the quantile transformation

$$u_t(k) = \Phi[w_t(k)] \quad (1)$$

where $\Phi[.]$ indicates the standard normal cumulative distribution function. Let the correlation between the Gaussian variates, $w_t$, for the station pair $k$ and $l$ be

$$\omega(k,l) = Corr[w_t(k), w_t(l)] \quad (2)$$

Together with the transition probabilities for stations $k$ and $l$, a particular $\omega(k,l)$ will yield a corresponding Bernoulli correlation between the synthetic binary series $(Y_t)$ for the two sites.

$$\xi(k,l) = Corr[Y_t(k), Y_t(l)] \quad (3)$$

Let $\xi^o(k,l)$ denote the observed value of $\xi(k,l)$, which will have been estimated from the observed binary series $Y_t^o(k)$ and $Y_t^o(l)$ at stations $k$ and $l$. Hence the problem reduces to finding the $N(N-1)/2$ correlations of $\omega(k,l)$ which together with the corresponding pairs of transition probabilities reproduces $\xi^o(k,l) = \xi(k,l)$ for each pair of stations. Direct computation of $\omega(k,l)$ from $\xi^o(k,l)$ is not possible. In practice, one can invert the relationship between $\omega(k,l)$ and $\xi(k,l)$ using a nonlinear root finding algorithm or obtain $\omega(k,l)$ by simulation as suggested by Wilks (1998). In the earlier study, the correlation between the corresponding normal variates is obtained by an iterative method using simulation and the method of bisection (Srikanthan, 2005). In this paper, an efficient root finding algorithm (Srikanthan and Pegram, 2006) is used to determine the correlation between the normal variates.

Realisations of the vector $w_t$ may be generated from the multivariate normal distribution having mean vector 0 and variance-covariance matrix $\Omega$, whose elements are the correlations $\omega(k,l)$.

The multivariate normal variates are generated from

$$w_t = B\varepsilon_t \quad (4)$$

where $B$ is a coefficient matrix and $\varepsilon_t$ independent normal vector.

The coefficient matrix is obtained from

$$BB^T = \Omega \quad (5)$$

The elements of $B$ can be obtained by Cholesky's decomposition for a small number of rainfall stations (up to 5). For a larger number of rainfall stations, the Cholesky's decomposition frequently fails as the matrix $\Omega$ tends to become non-positive definite for sequences of different lengths or due to infilling. In such cases the elements of $B$ can be obtained by singular value decomposition, a method that is robust even if the matrix $\Omega$ is ill-conditioned. The seasonality in daily rainfall occurrence is taken into account by considering each month separately.

### 2.2. Rainfall amounts model

The rainfall amounts on wet days are generated by using a Gamma distribution, which has been found to fit better than the routinely used distributions, exponential and Weibull. As was detailed above for the occurrences model, the spatial correlation in the daily rainfall amounts is preserved by using a vector of suitably cross-correlated uniform variates $v_t$ obtained from a corresponding realisation of correlated standard normal variates $z_t(k)$: $v_t(k) = \Phi[z_t(k)]$. This vector $z_t$ is drawn from a multivariate normal distribution with mean 0 and variance-covariance matrix $Z$, whose elements are

$$\zeta(k,l) = Corr[z_t(k), z_t(l)] \quad (6)$$

As was the case in finding the binary $\Omega$, direct computation of $Z$ is not feasible since the $z_t$ are not observed. The correlations in Eq (6) can be estimated by an iterative procedure using simulation and the method of bisection. The correlated multivariate normal variates are obtained from independent normal variates through a similar transformation to that using equations (4) and (5).

The generated daily rainfall amounts when aggregated into monthly and annual totals will not in general preserve the monthly and annual characteristics. Hence, the daily amount model is nested in a single site monthly and annual model Srikanthan 2005, 2006). This procedure will only improve the monthly and daily at-site characteristics of the generated rainfall and will have no effect on the spatial correlation for the monthly and annual rainfall. An outline of the new spatial adjustment procedure follows.

Once the daily rainfalls at all sites are generated for a given month, the monthly rainfall totals, $\tilde{x}_i^k$, at each site are obtained by summing the daily rainfall values. Their cross-correlations are calculated and the monthly totals modified by using a multi-site monthly model (Srikanthan and Pegram 2007) to preserve the monthly spatial and serial correlations.

$$X_i = A_i X_{i-1} + B_i a_i \quad (7)$$

where $A_i$ and $B_i$ are coefficient matrices and $X_i$ is the adjusted standardised adjusted monthly rainfall (zero mean and unit variance) vector for month $i$. The matrices $A_i$ and $B_i$ can be calculated from the lag zero ($M_0$) and lag one ($M_1$) spatial correlation of the observed monthly rainfall and the lag zero ($C_0$) cross correlation of the standardised, aggregated, already generated monthly rainfall ($a_i$).

$$A = M_1 M_0^{-1} \quad (8)$$

$$FF^T = M_0 - M_1 M_0^{-1} M_1^T \quad (9)$$

$$DD^T = C_0 \quad (10)$$

$$B = FD^{-1} \quad (11)$$

where $D$ and $F$ are intermediate matrices used in the computation The details of the above estimation procedure appears in Srikanthan and Pegram (2007).

After the adjustment, the monthly rainfall, $x_i^k$, at each site is obtained by putting back the mean and standard deviation. Once the values for the twelve months of a year ($j$) have been adjusted, the generated monthly values are aggregated to obtain the annual values ( $\tilde{z}_j^k$ ). The aggregated annual values are standardised to have zero mean and unit variance and then modified as above, by using a multi-site model to preserve the annual characteristics.

$$Z_j = P Z_{j-1} + Q b_j \quad (12)$$

where $P$ and $Q$ are coefficient matrices to preserve the lag zero and lag one cross correlations, $b_j$ is the already generated standardised annual value before adjustment and $Z_j$ is the adjusted standardised annual rainfall (zero mean and unit variance) vector. After the adjustment, the annual rainfall at each site is again obtained by appropriate scaling and shifting.

Each generated monthly rainfall value is multiplied by the ratio $z_j^k / \tilde{z}_j^k$. 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 $x_i^k z_j^k / \tilde{x}_i^k \tilde{z}_j^k$.

If the lag one cross correlations (monthly or annual) are all small, a contemporaneous multi-site model can be used for nesting. In this case, the matrix ***A or P*** becomes a diagonal matrix with diagonal elements being the lag one autocorrelations. In this case, one only needs to estimate the other matrix ***B or Q***.

## 3. MODEL EVALUATION

The model was evaluated using a number of statistics and cross correlations at the daily, monthly and annual levels. The daily, monthly and annual statistics used are listed in the following sections. One hundred replicates each of length equal to the historical data were generated and statistics were estimated from each of the replicates and averaged for comparison Due to lack of space, only a few results are presented here for the Sydney region.

### 3.1. Daily statistics

The cross correlations of daily rainfall occurrences and amounts between the sites were preserved for both the regions. The cross correlations for the Sydney region are presented in Figure 1.

In addition to the cross correlations, 17 other daily at site statistics were used to evaluate the model. The daily statistics include:

- Mean, standard deviation and coefficient of skewness of daily rainfall
- mean daily rainfall for different types of wet days; solitary wet day (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 per month
- maximum daily rainfall in each month
- mean, standard deviation and coefficient of skewness of dry spell length
- mean, standard deviation and coefficient of skewness of wet spell length

The mean number of wet days per month, maximum daily rainfall in each month, mean and standard deviation of daily rainfall per month are presented in Figure 2 for the Sydney region. There is a slight under-estimation of the maximum daily rainfall when the historical values are greater than 100 mm. The other statistics were satisfactorily preserved except the coefficient of skewness of wet and dry spells (Figure 2). The results for the other region are similar. The problem with the preservation of skewness will be addressed in further research.

### 3.2. Monthly statistics

The cross correlations of monthly rainfall amounts between the sites were well preserved for the Sydney region (Figure 1) and the Woady Yaloak

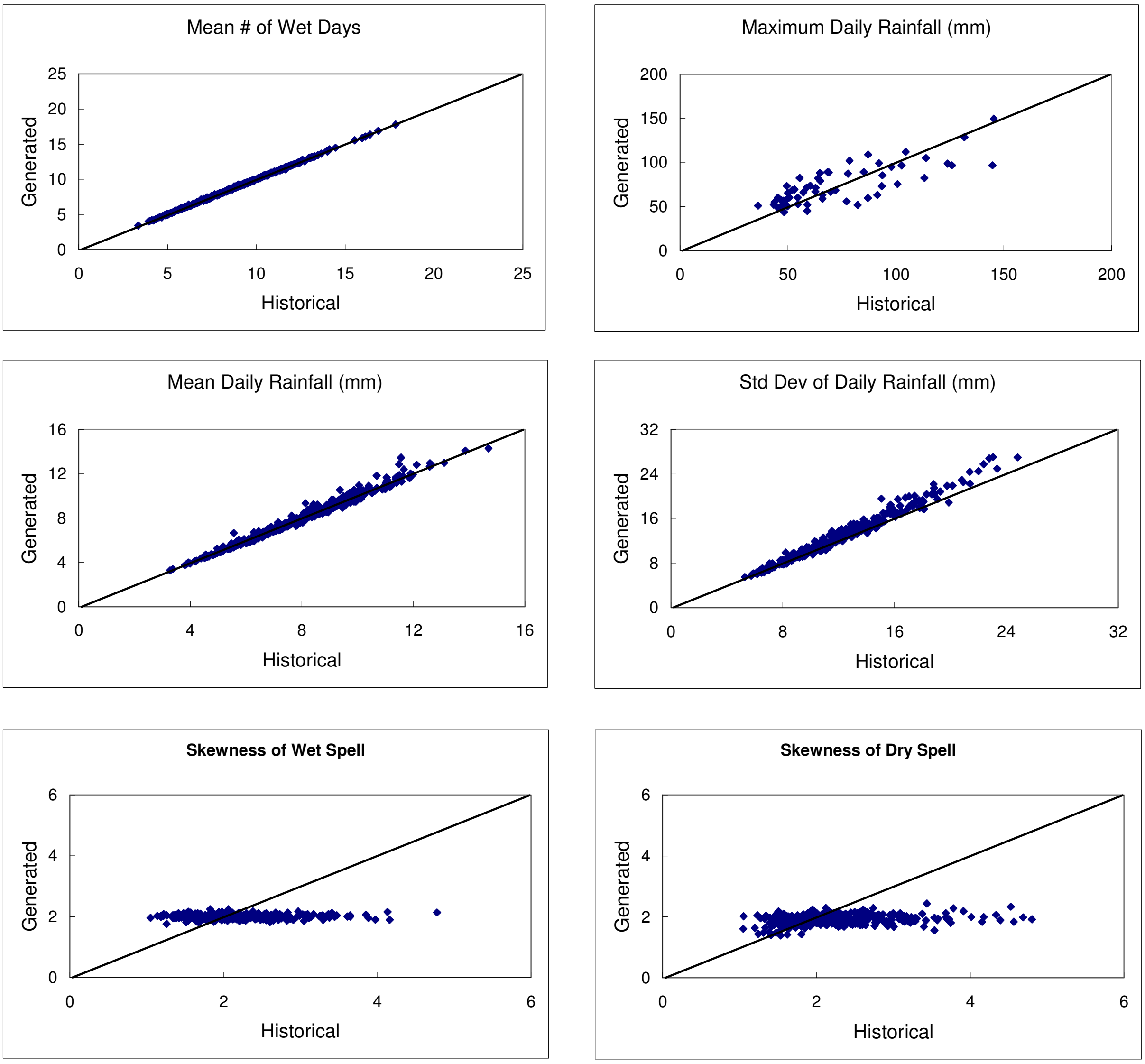


**Figure 2.** Selected daily statistics for the Sydney region.

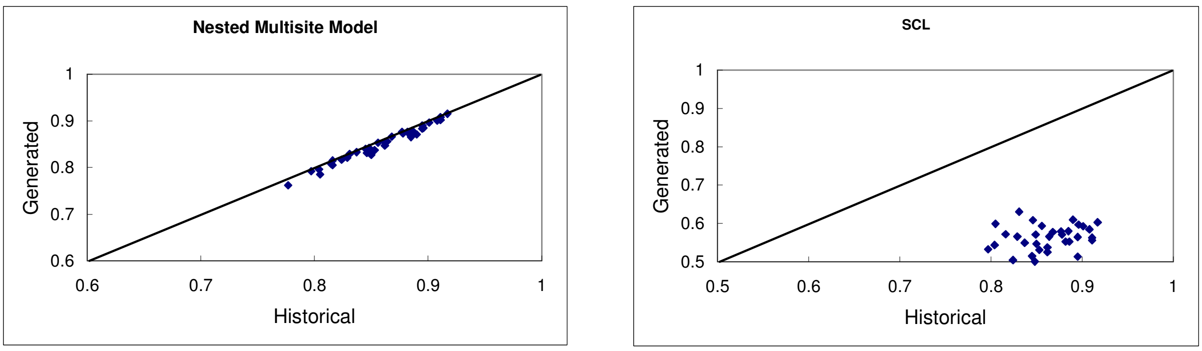


**Figure 3.** Comparison of monthly cross-correlation for the Woady Yaloak Catchment - the current model on the left and the SCL model on the right. Note the improvement gained by 'nesting' in a multi-site model.

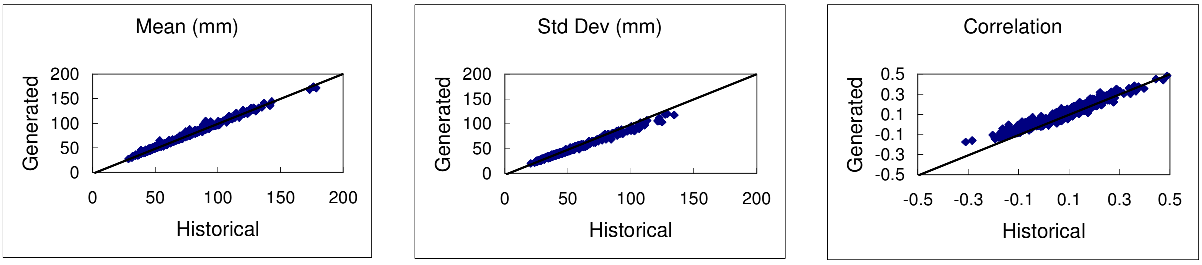


**Figure 4.** Mean, standard deviation and temporal correlation of monthly rainfall for the Sydney region.

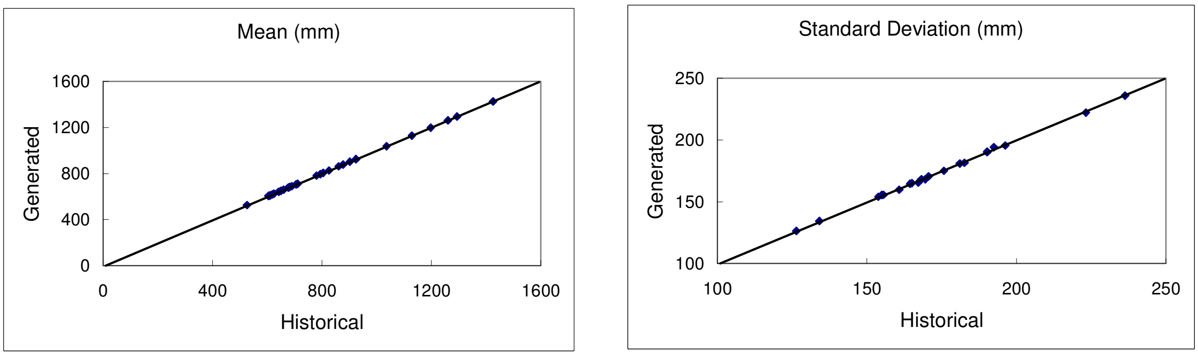


**Figure 5.** Mean and standard deviation of annual rainfall for the Sydney region.

catchment (Figure 3). For comparison, the monthly cross correlations from the spatial daily model in the Stochastic Climate Library (SCL) in the Cooperative Research Centre for Catchment Hydrology toolkit (http://www.toolkit.net.au) and the model described in this paper are presented in Figure 3 for the Woady Yaloak catchment. These figures clearly show a better performance of the model described in this paper compared to the one in SCL.

In addition to the cross correlations, seven other monthly statistics were used. 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

Of the seven statistics used in the monthly comparison, only the monthly mean, standard deviation and correlation are shown in Figure 4 for the Sydney region. The figure shows that these statistics were satisfactorily preserved, as were the rest of the statistics for both the regions.

### 3.3. Annual statistics

The cross correlations of annual rainfall amounts between the sites were well preserved for the Sydney region (Figure 1) and the Woady Yaloak Catchment (Table 1). Table 1 also presents a comparison of the annual cross correlation between the sites for the Woady Yaloak catchment using the nested multi-site model and the one in SCL.

**Table 1**. Comparison of annual cross correlations.for the Woady Yaloak catchment.

| Site pair | Hist | Nested | SCL |
|---|---|---|---|
| 1 - 2 | 0.841 | 0.844 | 0.547 |
| 1 - 3 | 0.794 | 0.798 | 0.548 |
| 2 - 3 | 0.865 | 0.867 | 0.581 |

In addition to the cross correlations, 13 other annual statistics were used to evaluate the model. 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

The results showed that all the statistics were satisfactorily preserved for both regions. Only the mean and standard deviation of the annual rainfall are shown in Figure 5 for the Sydney region. The standard deviation of the annual number of wet days was not compared in this study.

## 4. CONCLUSION

A nested multi-site two-part model was developed to improve the cross correlation of monthly and annual rainfall aggregated from daily amounts, while preserving the serial correlations at the different time scales. The developed model was evaluated by applying it to 2 catchments/regions with the number of rainfall sites being 3 and 30. A comparison of the historical and generated statistics showed that the model preserves all the important characteristics of rainfall at the daily, monthly and annual time scales. The only exception was the skewness of wet and dry spells.

The nesting of the multi-site daily rainfall in a cascade of multi-site monthly and annual models was effective in preserving the spatial cross correlations at the monthly and annual time scales and it is a major improvement over the model developed earlier (Srikanthan 2005, 2006). Further work is in progress to improve on the skewness of dry and wet spells and to compare the standard deviation of monthly and annual number of wet days.

## 5. REFERENCES


Bardossy, A. and Plate, E.J. (1991) Modelling daily rainfall using a semi-Markov representation of circulation pattern occurrence. Journal of Hydrology, 122: 33-47.

Bardossy, A. and Plate, E.J. (1992) Space time model for daily rainfall using atmospheric circulation patterns. Water Resources Research 28, 1247-1259

Buishand, T. A. and Brandsma, T. (2001) Multisite simulation of daily precipitation and temperature in the Rhine Basin by nearest neighbour resampling, Water Resources Research, 37, 2761 – 2776.

Charles, S. P., Bates, B. C. and Hughes, J. P. (1999) A spatiotemporal model for downscaling precipitation occurrence and amounts. Journal of Geophysical Research, Vol 104(D24): 31657-31669.

Hughes, J. P., Guttorp, P. and Charles, S. P. (1999) A non-homogeneous hidden Markov model for precipitation occurrence. Applied Statistics, 48 Part 1: 15-30.

Jothityangkoon, C., Sivapalan, M. and Viney, N. R. (2000) Tests of a space-time model of daily rainfall in southwestern Australia based on nonhomogeneous random cascades. Water Resources Research, Vol. 36(1): 267-284.

Mehrotra, R. and Sharma, A. (2005) A k-nearest-neighbour approach for downscaling atmospheric circulation indicators to multi-site precipitation occurrence. 29th Hydrology and Water Resources Symposium, Canberra.

Mehrotra, R., Srikanthan, R. and Sharma, A. (2005) Comparison of three approaches for stochastic simulation of multi-site precipitation occurrence. 29th Hydrology and Water Resources Symposium, Canberra.

Srikanthan R. and McMahon T.A. (2001) Stochastic generation of annual, monthly and daily climate data: A review. Hydrology and Earth System Sciences 5(4):653-656.

Srikanthan R. (2005) Stochastic generation of daily rainfall and at a number of sites, Technical Report 05/7, CRC for Catchment Hydrology, Monash University, Victoria.

Srikanthan R. (2006) Stochastic generation of spatially consistent daily rainfall at multiple sites. 30th Hydrology and Water Resources Symposium, Launceston.

Srikanthan, R. and G. G. S. Pegram (2006) Stochastic generation of multi-site rainfall occurrence. Advances in Geosciences, Vol. 6, 1-10.

Srikanthan, R. and G. G. S. Pegram (2007) Nested multisite daily rainfall data generation model. eWater Technical Report (in preparation).

Wilks, D. S. (1998) Multisite generalisation of a daily stochastic precipitation generation model. Journal of Hydrology 210: 178-191.

Wilson, L. L. and Lettenmaier, D. P. (1993) A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. Journal of Geophysical Research 97: 2791-2809.

Zucchini, W. and Guttorp, P. (1991) A hidden Markov model for space-time precipitation. Water Resources Research 27(8): 1917-1923.