Refinements to the Srikanthan-McMahon Stochastic Model for Daily Rainfall

T.G. Chapman

School of Civil and Environmental Engineering,
The University of New South Wales, Sydney 2052, Australia (tomc@civeng.unsw.edu.au)

Abstract: The Srikanthan-McMahon stochastic model for daily rainfall, which uses monthly sets of parameters, has been shown to provide satisfactory fit to both rainfall occurrences and rainfall amounts, and also to rainfall amounts classified by the adjoining number of wet days. However it underestimates the standard deviation of both monthly and annual rainfall totals, and Boughton [1999] has introduced a correction factor which compensates for this shortcoming in regard to the annual totals. It will be shown here that this does little to improve the standard deviation of the simulated monthly totals, and that the model cannot reproduce observed sequences of wet→wet and dry→dry years, and performs poorly in relation to the predicted number of 'wet' years (those with above average annual rainfall). Appropriate adjustments are developed in this paper, using data from 11 rainfall stations covering a wide spread of the Australian climatic environment.

Keywords: Daily rainfall; Stochastic modelling; Variance; Persistence; Linear regression

1. INTRODUCTION

The main objective in developing a stochastic model of daily rainfall is to provide long input sequences to rainfall-runoff models. Provided these sequences retain the statistical characteristics of the historical record, they can result in a range of scenarios which may differ markedly from the details of that record, and so enable more confident design of hydraulic structures ranging in scale from roof-tank water supplies to agricultural, flood control and water supply dams.

Most stochastic models of daily rainfall can be divided into two parts: a model for rainfall occurrence, which provides a sequence of dry and wet days, and a model of rainfall amounts, which simulates the amount of rainfall occurring on each wet day. A comprehensive review of approaches to modelling daily rainfall is given in Woolhiser [1992].

Srikanthan and McMahon [1985], in a study covering the main climatic regions of Australia, followed Allan and Haan [1975] by extending the Markov chain structure commonly used for rainfall occurrence into a multi-state model or transition probability matrix, in which the daily rainfalls are grouped into up to seven classes of given magnitude ranges (see Table 1), and the probabilities are calculated for transition from each class to any other. The lowest class gives the occurrences of dry days, the top class is modelled by a skewed normal distribution (requiring estimation of 3 parameters), and intermediate classes are modelled by a linear distribution. Separate parameter values are calculated for each month, so that when all 7 classes are used, the total number of parameters is 12 x (7 x 6 + 3) = 540. However, when there are insufficient data in Class 7 for a particular month, the number of classes may be reduced, the top class having the lower limit shown in Table 1, but no upper limit.

Table 1. Daily rainfall class limits in the Srikanthan-McMahon (SM) model.

Class	Lower limit (mm)	Upper limit (mm)
1	0	0.0
2	0.1	0.9
3	1.0	2.9
4	3.0	6.9
5	7.0	14.9
6	15.0	30.9
7	31.0	-

The parameter set for each month is calculated from the historical record by counting the number of occurrences of transitions between classes, and dividing by the total number of transitions to obtain the transition probabilities. The rainfall amounts in the top class are then adjusted for skew, and fitted to a normal distribution.

A similar approach has more recently been used by Gregory et al. [1993], in a study of area-average daily rainfalls for 3-month seasonal periods in Britain. They noted that the transition probability matrix was better able to simulate the standard deviation of longer-period totals than models which separate rainfall occurrence from rainfall amount.

Chapman [1998], in a study covering 66 rainfall stations in Australia, South Africa, Pacific islands and North America, found that there are significant differences between the distributions of daily rainfall amounts, when they are classified according to the number of adjoining wet days (0,1 or 2). Thus Class 0 comprises solitary wet days, Class 1 comprises rainfalls on days at the beginning or end of a wet spell (of at least 2 days' duration), and Class 2 comprises rainfalls in the interior of a wet spell (which is therefore of at least 3 days' duration). For the Australian stations, the average ratio of Class 0 rainfall to mean daily rainfall was 0.40, while the ratio of Class 2 rainfall to mean daily rainfall was 2.33. Failure to model such differences would clearly have a major impact on the output of a rainfall-runoff model.

Chapman [1998] also showed that the Srikanthan-McMahon (SM) model successfully models the rainfall in each class, as shown for a typical station in Figure 1. This feature cannot be reproduced in models which simulate rainfall occurrences and rainfall amounts separately, unless three different distributions are used for the rainfall amounts.

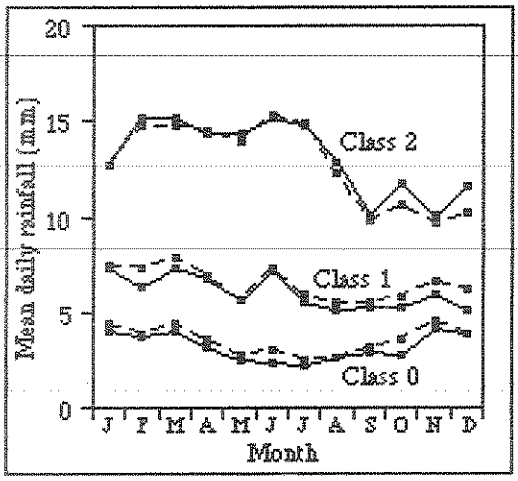

Figure 1. Observed (full line) and simulated (dashed line) mean daily rainfalls for Sydney, classified according to number of adjoining wet days (0, 1 or 2). Simulation is by SM model.

However, the SM model generally underestimates the standard deviation (SD) of monthly and annual rainfalls, which would result in underestimation of extreme events. This is a feature common to most stochastic models of daily rainfall, and is termed 'overdispersion' [Katz and Parlange, 1998; Katz and Zheng, 1999; Wilks, 1999]. Boughton [1999] proposed a correction for the SD of annual rainfall in the SM model, in which the simulated daily rainfall in each year is multiplied by the following ratio:

$$\text{ratio} = \{M + F(T_i - M)\}/T_i \qquad (1)$$

where T_i is the annual total for year i, M is the observed mean annual rainfall, and F is an adjustment factor, which can be shown to be the ratio of the observed to the simulated SD, SD_O/SD_S. This correction forces the simulated annual SD to be exactly equal to the observed value, which is an artificial construct, as the SD of the historical sample is unlikely to be exactly the same as that of the population.

This paper addresses the problem of underestimation by the SM model of the SD of monthly and annual rainfall, and also some shortcomings of the model in simulating wet→wet and dry→dry annual sequences, where 'wet' and 'dry' are defined as being above and below average rainfall respectively.

2. DATA

The data used in this study are from the files prepared by Srikanthan and McMahon [1985] for their model development and demonstration. Stations with at least 40 years of data have been selected from the full set. Locations of the stations are shown in Figure 2, and details of the records are given in Table 2. Only complete calendar years of record have been analysed, and the record has been taken as continuous across missing years. All rainfall amounts are to the nearest 0.1 mm.

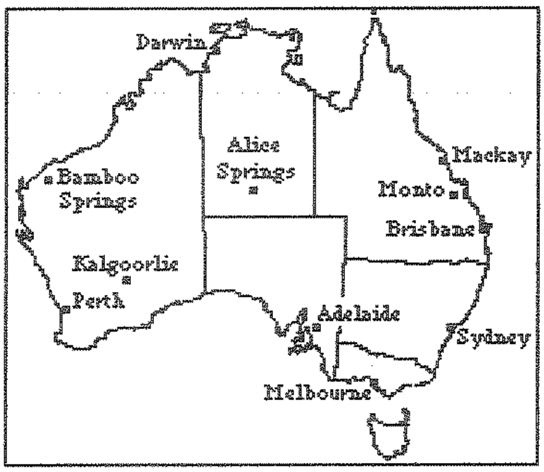

Figure 2. Location of rainfall stations.

Table 2. Details of rainfall records.

Station name	First year	Last year	Years of record
Adelaide	1853	1978	125
Alice Springs PO	1887	1965	73
Bamboo Springs	1917	1976	55
Brisbane	1869	1981	112
Darwin PO	1872	1939	66
Kalgoorlie	1940	1981	41
Mackay PO	1889	1949	61
Melbourne	1856	1980	125
Monto	1931	1980	50
Perth	1880	1980	101
Sydney	1859	1981	123

All analyses were performed with the full period of record available, and all simulations were for a period of 10000 years. For each station, the number of classes (see Table 1) in each month was adjusted so that there were at least 10 samples in the top class, in order to provide for reasonable fitting of the skewed normal distribution. This was accomplished by maximum likelihood fitting of data normalised by the Box-Cox transformation

$$y = [(x+c)^{\lambda} - 1] / \lambda, \qquad \lambda \neq 0$$

$$y = \log(x+c), \qquad \lambda = 0 \qquad (2)$$

where the constant c has been taken as 0.1.

3. QUANTIFICATION OF PROBLEMS

Figure 3 shows the relation in the SM model between observed and simulated annual SD for all stations. It will be seen that, with one exception (Darwin), the simulated SD is underestimated, with the error increasing with higher SD. The effect of the Boughton correction is to force all the points on to the 1:1 line.

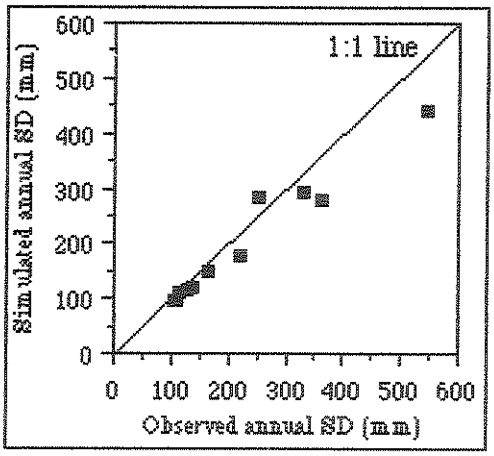

Figure 3. Relation between observed and simulated annual SD for all stations.

The magnitude of the underestimation is given by the regression (excluding Darwin):

$$SD_s = 0.78\ SD_o + 15.3 \quad (R^2 = 0.99) \qquad (3)$$

The monthly SD's are also underestimated by the original model, as shown in Figure 4. The regression for all stations and months is

$$SD_s = 0.81\ SD_0 + 6.5 \quad (R^2 = 0.97) \qquad (4)$$

When the Boughton correction was applied to the annual SD's, the regression coefficient for the

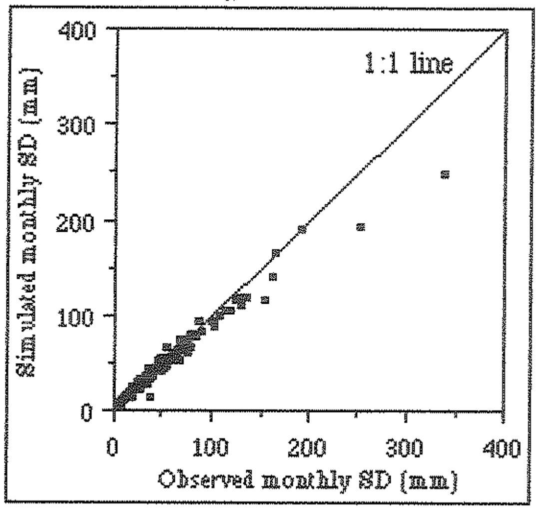

Figure 4. Relation between observed and simulated monthly SD for all stations.

monthly SD's increased from 0.81 to 0.85, leaving most of the bias uncorrected.

The model as developed by Srikanthan and McMahon is unable to simulate observed sequences of 'wet' and 'dry' months or years. This is illustrated in Figures 5 and 6.

A feature that was not anticipated is the relative constancy of the simulated percentage of wet years, as illustrated in Figure 7.

4. MODIFICATION OF MODEL

It was recognised that adjustment of SD's, whether applied on a monthly or annual basis, would have no effect on the persistence characteristics illustrated in Figures 5 and 6. However, the effect on SD's of adjusting the persistence characteristics was unknown. An attempt to improve the persistence characteristics was therefore given first priority.

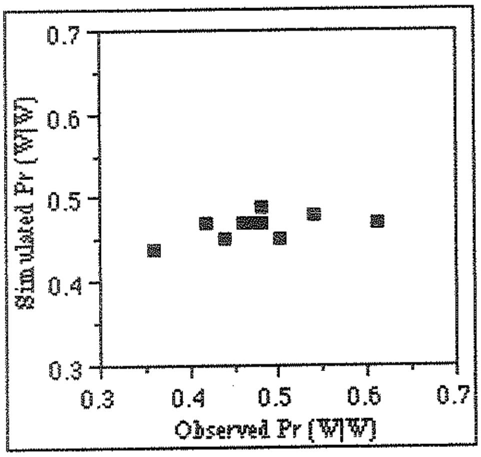

Figure 5. Relation between observed and simulated probability of a wet year following a wet year.

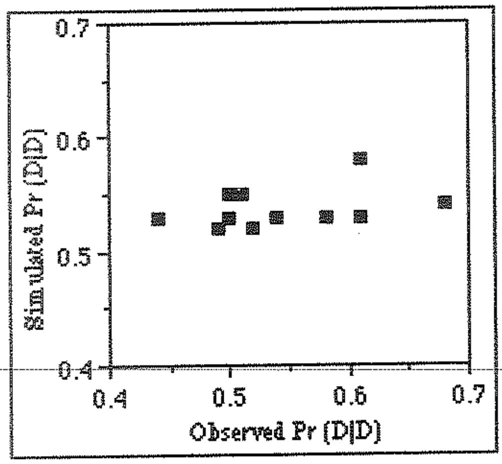

Figure 6. Relation between observed and simulated probability of a dry year following a dry year.

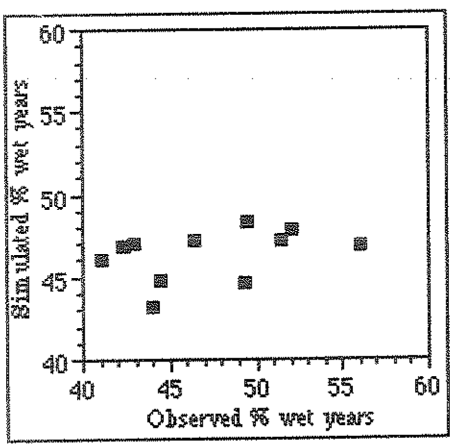

Figure 7. Relation between observed and simulated percentage of wet years.

An obvious approach to this problem is to obtain two sets of parameters for the model, one for wet years and one for dry. In simulation, the decision on which set of parameters to use for a particular year is then based on the observed probabilities Pr (W|W) and Pr (D|D). The results of this procedure (Figures 8-10) are acceptable in terms of the persistence characteristics, but there remains a significant underestimate of the simulated percentage of wet years. This has not been further addressed at this time.

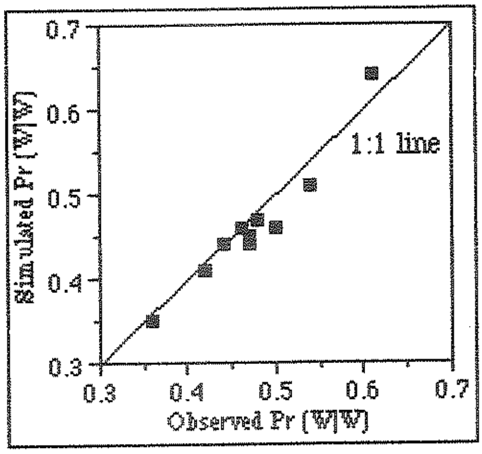

Figure 8. Relation between observed and simulated probability of a wet year following a wet year, after model modification.

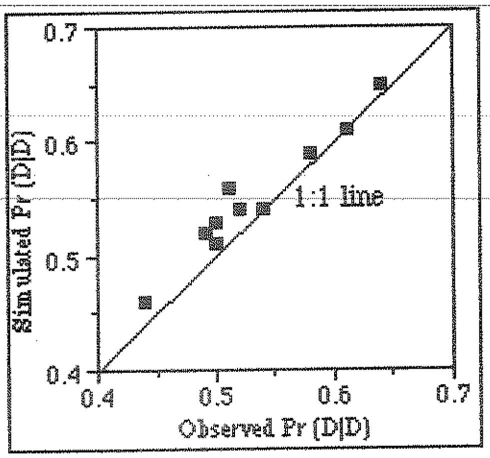

Figure 9. Relation between observed and simulated probability of a dry year following a dry year, after model modification.

The method adopted to adjust the monthly and annual SD's was to use the Boughton algorithm (1) for both statistics, but applied in a way which does not force the SD to be the same as the observed value. Noting the very high correlations (3) and (4) between observed and simulated values, covering all stations and all months, the

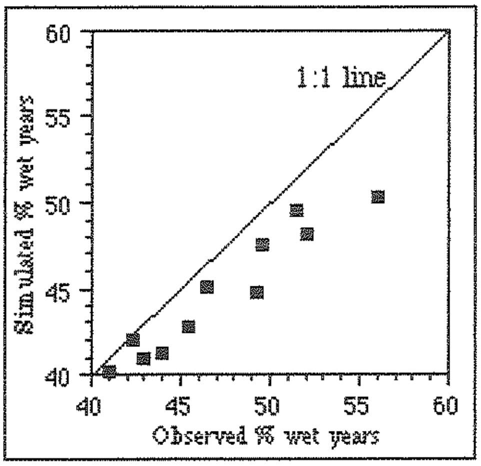

Figure 10. Relation between observed and simulated percentage of wet years, after model modification.

value of F (=SD_O/SD_S) was calculated from the regression equations. However, adjustment of the annual SD's changed the regression relation for the monthly SD's, and vice versa, thus requiring an iterative approach. After 3 cycles of adjustment, the equation for the adjustment factor for annual SD's was

$$F = 0.96\, SD_o^2/(1.27\, SD_o - 3.6)/(1.08\, SD_o - 2.4) \quad (5)$$

while that for the monthly SD's was

$$F = 1.03\, SD_o^2/(0.92\, SD_o + 4.6)/(0.92\, SD_o + 1.6) \quad (6)$$

The results of these adjustments are shown in Figures 11 and 12. Regressions through these data (without the annual SD for Darwin) show that the bias has been completely removed.

5. DISCUSSION

Classifying the parameters of the SM model into separate sets for dry and wet years almost doubles the number of parameters, and may force a reduction in the number of rainfall classes which can be used, as the available data are now split into the two groups. Nevertheless, the improvements in capacity for simulation of the behaviour in successive wet or dry years represent a significant upgrading of the model. This is accompanied by a major improvement in the prediction of the percentage of wet years, a statistic which does not appear to have been previously examined in tests of the SM model.

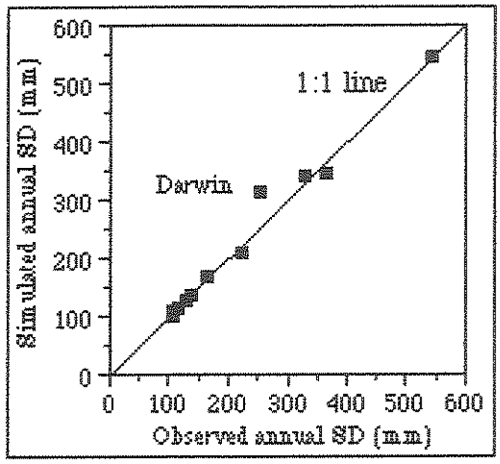

Figure 11. Relation between observed and simulated annual SD, after model modification.

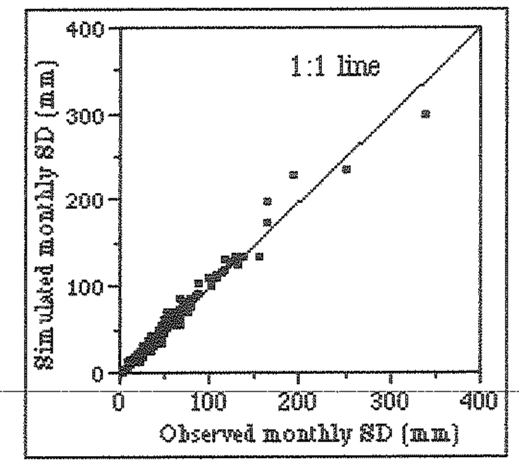

Figure 12. Relation between observed and simulated monthly SD, after model modification.

An unexpected feature of the analysis of monthly and annual SD's was the very strong correlation between the observed and simulated values, covering stations over a wide range of climatic environments and all months of the year. This clearly requires validation from tests with more comprehensive data sets. The nonconforming behaviour of the annual SD for Darwin may be due to the use of the calendar year throughout this study. It would be preferable to use a hydrologic water year for all stations, with the breaks at periods of minimum hydrologic activity.

It should be noted that the use of the regression relations to determine the value of F in the Boughton algorithm obviates the need to run the model first in uncorrected mode. After simulation of daily data for a month, the appropriate monthly correction (which requires the monthly total and the observed mean and SD for that month) is applied. At the end of the year, the corrected monthly totals are added to give the annual total for that year, and the annual correction can then be applied.

This study has concentrated on mean values of the statistics considered, by using a long run of simulated data. Subsequent work should investigate their variability, by using many replicates of runs of the same length as the observed data.

6. CONCLUSIONS

Simple modifications to the SM model enable the bias to be removed from the simulated SD of monthly and annual rainfall. Classifying the parameters of the model according to whether the current year is 'wet' or 'dry' is effective in modelling the persistence of annual rainfalls, and improves the simulation of the percentage of wet years.

Although these results have been obtained by sampling over a wide range of climatic environments, they require validation by use of a more comprehensive data set.

7. REFERENCES

Allan, D.M., and C.T. Haan, Stochastic simulation of daily rainfall, Research Report No. 82, Water Resources Institute, University of Kentucky, 1975.

Boughton, W.C., A daily rainfall generating model for water yield and flood studies, 99/9, CRC for Catchment Hydrology, Melbourne, 1999.

Chapman, T., Stochastic modelling of daily rainfall: the impact of adjoining wet days on the distribution of rainfall amounts, *Environmental Modelling and Software*, 13 , 317-324, 1998.

Gregory, J.M., T.M.L. Wigley, and P.D. Jones, Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals, *Climate Dynamics*, 8 , 299-310, 1993.

Katz, R.W., and M.B. Parlange, Overdispersion phenomena in stochastic modeling of precipitation, *Journal of Climate*, 11 , 591-601, 1998.

Katz, R.W., and X. Zheng, Mixture model for overdispersion of precipitation, *Journal of Climate*, 12 , 2528-2537, 1999.

Srikanthan, R., and T.A. McMahon, Stochastic generation of rainfall and evaporation data, Tech. Paper No. 84, Australian Water Resources Council, Canberra, 1985.

Wilks, D.S., Interannual variability and extreme-value characteristics of several stochastic daily precipitation models, *Agricultural and Forest Meteorology*, 93 , 153-169, 1999.

Woolhiser, D.A., Modeling daily precipitation - progress and problems, In: *Statistics in the Environmental and Earth Sciences*, A.T. Walden, and P. Guttorp (eds), pp. 71-89, Halsted Press, New York, 1992.