

---

# Stochastic generation of annual, monthly and daily climate data: A review

R. Srikanthan<sup>1</sup> and T.A. McMahon<sup>2</sup>

<sup>1</sup>Hydrology Unit, Bureau of Meteorology, Melbourne, Australia / University of Melbourne, Parkville, Australia

<sup>2</sup>Cooperative Research Centre for Catchment Hydrology, Clayton, Australia

Email for corresponding author: R.Srikanthan@BOM.GOV.AU

---

## Abstract

The generation of rainfall and other climate data needs a range of models depending on the time and spatial scales involved. Most of the models used previously do not take into account year to year variations in the model parameters. Long periods of wet and dry years were observed in the past but were not taken into account. Recently, Thyer and Kuczera (1999) developed a hidden state Markov model to account for the wet and dry spells explicitly in annual rainfall. This review looks firstly at traditional time series models and then at the more complex models which take account of the pseudo-cycles in the data. Monthly rainfall data have been generated successfully by using the method of fragments. The main criticism of this approach is the repetitions of the same yearly pattern when only a limited number of years of historical data are available. This deficiency has been overcome by using synthetic fragments but this brings an additional problem of generating the right number of months with zero rainfall. Disaggregation schemes are effective in obtaining monthly data but the main problem is the large number of parameters to be estimated when dealing with many sites. Several simplifications have been proposed to overcome this problem. Models for generating daily rainfall are well developed. The transition probability matrix method preserves most of the characteristics of daily, monthly and annual characteristics and is shown to be the best performing model. The two-part model has been shown by many researchers to perform well across a range of climates at the daily level but has not been tested adequately at monthly or annual levels. A shortcoming of the existing models is the consistent underestimation of the variances of the simulated monthly and annual totals. As an alternative, conditioning model parameters on monthly amounts or perturbing the model parameters with the Southern Oscillation Index (SOI) result in better agreement between the variance of the simulated and observed annual rainfall and these approaches should be investigated further. As climate data are less variable than rainfall, but are correlated among themselves and with rainfall, multisite-type models have been used successfully to generate annual data. The monthly climate data can be obtained by disaggregating these annual data. On a daily time step at a site, climate data have been generated using a multisite type model conditional on the state of the present and previous days. The generation of daily climate data at a number of sites remains a challenging problem. If daily rainfall can be modelled successfully by a censored power of normal distribution then the model can be extended easily to generate daily climate data at several sites simultaneously. Most of the early work on the impacts of climate change used historical data adjusted for the climate change. In recent studies, stochastic daily weather generation models are used to compute climate data by adjusting the parameters appropriately for the future climates assumed.

## Introduction

This paper reviews the state of research and practice in the stochastic generation of annual, monthly and daily climate data. One of the major gaps in the design and operation of hydrological systems is the quantification of uncertainty as a result of climatic variability. This applies whether the systems are complex water resources systems or simple planning models of catchment behaviour. For very simple systems, analytical techniques of estimating uncertainty may suffice but, for the majority of systems, one has to resort to system simulation using stochastically generated data. In

addition to quantifying uncertainty, stochastically generated data have applications such as the design and operation of water resources systems, design of urban drainage systems and land management changes.

Generated annual data have little direct application but are used indirectly in disaggregation schemes to obtain monthly data. Estimation of water demand and the simulation of water supply systems generally need monthly data. For rainfall-runoff and crop growth models, daily data are required. This review covers the generation of rainfall and climate data such as temperature, solar radiation and

evaporation. Rainfall and climate data are measured at point locations but are assumed to represent the surrounding areas. In small to medium size catchments, areally averaged data can take spatial variations into account. In large catchments, it is necessary to model the spatial variation explicitly especially in rainfall.

Though numerous stochastic models are available in the literature, few have been tested adequately with regard to characteristics at different time scales or at a number of locations with different climates. For instance, a proper daily model should preserve monthly and annual characteristics in addition to preserving the daily characteristics.

In the past, data generation models assumed that there was no variation in the model parameters between years and only the seasonal or monthly variations within a year were taken into account. However, there is a growing awareness of long term persistence in the climatic data in the form of wet and dry years or ENSO cycles, so the parameters of the models should be varied in some way to model the wet and dry spells. There has been very little research on this aspect.

The generation of rainfall and other climate data needs a range of models depending on the time and spatial scales involved. Cox and Isham (1994) presented three broad types of rainfall models, namely, empirical statistical models, models of dynamic meteorology and intermediate stochastic models, a classification based on the amount of physical realism incorporated into the model structure. In empirical statistical models, empirical stochastic models are fitted to the data available. The models for the generation of annual, monthly and daily rainfall and climate data are of this type. In the models of dynamic meteorology, large systems of simultaneous nonlinear partial differential equations representing, fairly realistically, the physical processes involved, are solved numerically. These are generally used for weather forecasting rather than for data generation. In intermediate stochastic models, few parameters are used to represent the rainfall process, the parameters being intended to relate to underlying physical phenomena such as rain cells, rain bands and cell clusters. These types of models are used for the analysis of data collected at short time intervals such as hourly. The models reviewed in this paper are of the first type only, namely, empirical statistical models.

Because of the different types of models used for different time scales, the review is carried out for different time scales separately. The generation of rainfall data is reviewed first and that of climate data follows. In each section, the models for the generation of data at a single site are reviewed first, followed by models for a number of sites over an area.

## Annual and monthly rainfall data

Surprisingly, very little work has been done since 1985 on stochastic generation of annual and monthly rainfall data; the year to year variations in the model parameters were ignored and only the within-year seasonal variations in parameters were taken into account. Long periods of wet and dry years were observed in the past (Warner, 1987; Srikanthan and Stewart, 1992) and this phenomenon needs to be considered in the model structure. Recently, Thyer and Kuczera (1999, 2000) developed a hidden state Markov model to account for the long term persistence in annual rainfall.

### ANNUAL RAINFALL DATA AT A SITE

Thomas and Fiering (1962) were the first to propose a first order Markov model with Wilson-Hilferty transformation to generate streamflow data. Srikanthan and McMahon (1985) recommended a first order Markov model incorporating the Wilson-Hilferty transformation to generate annual rainfall data. This model degenerates into a white noise model when the coefficient of skewness and the lag one auto-correlation coefficient are close to zero.

Thyer and Kuczera (1999) developed a Hidden State Markov (HSM) model with Bayesian inference to generate annual rainfall data for Sydney, Australia. The model assumes that the climate is composed of two states, either a dry state (low rainfall year) or a wet state (high rainfall year). Each state has separate Normal annual rainfall distributions. The transition from one state to the other is governed by the transition probabilities. If the transition probabilities to other states are sufficiently low then the climate may persist in one state for a number of years. This provides an explicit mechanism for the HSM model to simulate the influence of a quasi-periodic phenomenon such as El Nino.

In the HSM model, the simulation of annual rainfall is a two step process. In the first step, the state at year  $t$  is simulated by a Markov process,

$$S_t | S_{t-1} \sim \text{Markov}(\mathbf{P}, p_t) \quad (1)$$

where  $\mathbf{P}$  is the transition probability matrix whose elements  $p_{ij}$  are defined by,

$$p_{ij} = \text{Pr}(S_t = i | S_{t-1} = j) \quad i, j = \text{wet or dry} \quad (2)$$

and  $p_t$  is the probability distribution vector of the wet and dry states at year  $t$ .

Once the state for the year  $t$  is known, the annual rainfall is simulated using,

$$x_t = \begin{cases} N(\mu_{wet}, \sigma_{wet}^2) & \text{if } S_t = \text{wet} \\ N(\mu_{dry}, \sigma_{dry}^2) & \text{if } S_t = \text{dry} \end{cases} \quad (3)$$

where  $N(\mu, \sigma^2)$  denotes a normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

They compared the results from the HSM model with those from an AR(1) model and found that the dry spell persistence identified by the HSM model produced higher and more realistic drought risks. The Sydney rainfall data supported strongly the assumption of a two state climate model with the average residence time similar to the quasi-periodicity of the ENSO phenomenon.

#### ANNUAL RAINFALL DATA AT A NUMBER OF SITES

Annual rainfall at a number of sites can be generated by using a multi-site model (Matalas, 1967; Young and Pisano, 1968).

$$\mathbf{X}_t = \mathbf{A} \mathbf{X}_{t-1} + \mathbf{B} \mathbf{e}_t \quad (4)$$

where  $\mathbf{X}_t = (n \times n)$  matrix of standardised annual rainfall data for year  $t$ ,

$\mathbf{e}_t = (n \times 1)$  vector of independent random deviates with zero mean and unit variance, and

$\mathbf{A}$ ,  $\mathbf{B} = (n \times n)$  matrices of constant coefficients to preserve the auto-correlations and the cross-correlations.

If the annual rainfall data are skewed, they can be normalised using a three-parameter Log-normal transformation. The parameters in the log domain are obtained by using the Matalas moment transformation equations (Matalas, 1967). Pegram and James (1972) and Hipel (1985) suggest that the lag one cross-correlations can be ignored. This simplifies the correlation matrices and the solutions to  $\mathbf{A}$  and  $\mathbf{B}$ .

#### MONTHLY RAINFALL DATA AT A SITE

Thompson (1984) derived a model for monthly rainfall based on a Poisson process of rainfall occurrence with an independent exponentially distributed rainfall amount. The model has two parameters and can preserve the mean and variance of monthly rainfall explicitly. The model ignores the skewness and the serial correlation between months. The model was used not for data generation but to test the homogeneity of rainfall records at a station without the need for data from neighbouring stations.

Srikanthan and McMahon (1985) recommended the method of fragments for the generation of monthly rainfalls. The observed monthly rainfall data are standardised year

by year so that the sum of the monthly rainfalls in any year equals unity. This results in  $n$  sets of fragments of monthly rainfalls from a record of  $n$  years. The generated annual rainfalls are disaggregated by selecting a set of fragments at random and multiplying the annual rainfall generated by each of the 12 fragments to give 12 generated monthly rainfalls. A major limitation of this procedure is that the monthly correlation between the first month of a year and the last month of the previous year will not be preserved. However, this problem can be avoided by choosing the hydrological year starting with a month with the minimum serial correlation.

Porter and Pink (1991) reported that the use of the method of fragments resulted in the conspicuous repetition of monthly patterns when generating data much longer than the historical data. They proposed to obtain the monthly fragments from a generated monthly flow sequence. It appears that the monthly values were generated independently at each site. Each generated annual rainfall was disaggregated using the monthly fragments from the generated monthly rainfall for which the generated annual rainfall is closer to the annual value obtained from the generated monthly rainfalls. This overcomes the problem of repetition but does not preserve the monthly correlation between the first month of a year and the last month of the previous year. Maheepala and Perera (1996) proposed a modification to the Porter and Pink (1991) model which allows the preservation of monthly correlation across consecutive years. They have compared this modified procedure with the above two methods of fragments, which use historical and synthetic fragments using streamflow data from five rivers in Victoria. The results showed that the modified model preserved the monthly correlations across consecutive years. The method used for the generation of synthetic monthly flows is not clear from Maheepala and Perera (1996).

Since rainfall data are less variable and have smaller skewness than streamflow data, the extended disaggregation scheme proposed by Mejia and Rousselle (1976) can be used to disaggregate the generated annual rainfall into monthly rainfall. Lane (1979) developed an approach which essentially sets to zero several parameters of the model which are not important. The model considers one month at a time. The seasonal values are then adjusted to match the annual values. Stedinger *et al.* (1985) developed a condensed version of the disaggregation model in which the monthly flow sequences are described by a set of coupled regression equations. The performance of these models to sites with a considerable number of zero monthly rainfalls is not known.

Recently Tarboton *et al.* (1998) proposed a disaggregation procedure based on nonparametric density (NPD) estimation

which models adaptively complex relationships between aggregate and disaggregate flows. The primary shortcoming of the NPD procedure is that it is data and computationally intensive. For less than 30 years of data, the NPD procedure is not expected to disaggregate monthly data better than parametric models.

#### MONTHLY RAINFALL DATA AT A NUMBER OF SITES

Monthly rainfall data at a number of sites can be generated by disaggregating the generated annual rainfall using the method of fragments (Srikanthan McMahon, 1984), the method of synthetic fragments (Porter and Pink, 1991) or the modified method of synthetic fragments (Maheepala and Perera, 1996).

The extended model developed by Mejia and Rousselle (1976) can be used to disaggregate the generated annual rainfalls to monthly rainfalls. Here again, the condensed form of the model developed by Lane (1979) can be used at the expense of not preserving some of the cross correlations. One of the main drawbacks with the disaggregation approach is the large number of parameters to be estimated from the historical data. The number of parameters in the model for the generation of monthly data at  $N$  sites is  $156N^2$  for the basic (Valencia and Schaake, 1973),  $168N^2$  for the extended (Mejia and Rousselle, 1976) and  $36N^2$  for the condensed (Lane, 1979) schemes. Salas *et al.* (1980) give a parsimony guide for disaggregation modelling.

### Daily rainfall data

Long sequences of daily rainfall are required increasingly, not only for hydrological purposes but also to provide inputs for models of crop growth, landfills, tailing dams, land disposal of liquid waste and other environmentally-sensitive projects. Rainfall is generally measured at the daily time scale and this forms the basis for monthly and annual rainfall series. Because daily data form this basic data set, modelling of the daily rainfall process has attracted a lot of interest in the past.

#### DAILY RAINFALL DATA AT A SITE

Daily rainfall data generation models can be classified broadly into four groups, namely, two-part models, transition probability matrix models, resampling models and time series models of the ARMA type.

#### TWO-PART MODELS

Most stochastic models of daily rainfall consist of two parts:

a model for the occurrence of dry and wet days and a model for the generation of rainfall amount on wet days. The seasonal variation in rainfall is an important factor and several approaches have dealt with seasonality, assuming that parameters vary either as a step function for each month/season or as a periodic function (such as Fourier series) to provide intra-annual variation of parameters.

#### RAINFALL OCCURRENCE MODELS

Models of rainfall occurrence are of two main types, those based on Markov chains and those based on alternating renewal processes.

##### *Markov chains*

Markov chains specify the state of each day as 'wet' or 'dry' and develop a relation between the state of the current day and the states of the preceding days. The order of the Markov chain is the number of preceding days taken into account. Most Markov chain models referred in the literature are first order (lag one) (Gabriel and Newmann, 1962; Caskey, 1963; Weiss, 1964; Hopkins and Robillard, 1964; Feyerherm and Bark, 1965, 1967; Lowry and Guthrie, 1968; Selvalingam and Miura, 1978; Stern, 1980a,b; Garbutt *et al.*, 1981; Richardson, 1981; Stern and Coe, 1984). Models of second or higher orders have been studied by Chin (1977), Coe and Stern (1982), Gates and Tong (1976), Eidsvik (1980), Pegram (1980) and Singh *et al.* (1981). The results varied with the climate characteristics of the rainfall stations investigated, with the statistical tests used and with the length of record.

The Akaike information criterion (AIC), introduced by Akaike (1974), was widely used to determine the order of the Markov chains. Katz (1981) derived the asymptotic distribution of the (AIC) estimator but found that the estimator is inconsistent. The Bayesian information criterion (BIC) proposed by Schwarz (1978) was shown to be consistent and asymptotically optimal. However, Hurvich and Tsai (1989) provided a correction for AIC for model selection in small samples and the corrected AIC does not over fit the models as the AIC tends to do.

Jimoh and Webster (1996) determined the optimum order of a Markov chain model for daily rainfall occurrences at five locations in Nigeria using AIC and BIC. The AIC consistently gave a higher order for the Markov chain than the BIC. The optimum order was also investigated by the generation of synthetic sequences of wet and dry days using zero-, first- and second-order Markov chains. They found that the first-order model was superior to the zero-order model in representing the frequency distribution of wet and

dry spells and there was no discernible difference between performances of the first- and second-order models. It was concluded that caution is needed with the use of AIC and BIC for determining the optimum order of the Markov model and the use of frequency duration curves can provide a robust alternative method of model identification.

Jimoh and Webster (1999) investigated the intra-annual variation of the Markov chain parameters for seven sites in Nigeria. They found that there was a systematic variation in  $P_{01}$  (probability of a wet day following a dry day) as one moves northwards and a limited regional variation in  $P_{11}$ .

A general conclusion is that a first-order model is adequate for many locations but a second- or higher order model may be required at other locations or during some times of the year.

#### *Alternating renewal process*

In the alternating renewal process, the daily rainfall data is considered as a sequence of alternating wet and dry spells of varying length. The wet and dry spells are assumed to be independent and the distributions may be different for wet and dry spells. Distributions investigated include the logarithmic series (Williams, 1947), a modified logarithmic series (Green, 1964), truncated negative binomial distribution (Buishand, 1977), and the truncated geometric distribution (Roldan and Woolhiser, 1982). Roldan and Woolhiser (1982) compared the alternating renewal process with truncated geometric distribution of wet sequences and truncated negative binomial distribution of dry sequences with a first-order Markov chain. For five US stations with 20-25 years of record lengths, the first-order Markov chain was superior to the alternating renewal process according to the Akaike information criterion (Akaike, 1974). The parameters of the distributions were assumed to be either constant within seasons or to vary according to Fourier series. One of the disadvantages of the alternating renewal process is that the seasonality is difficult to handle. The starting day of the sequence is usually used to determine the season to which the sequence belongs.

Small and Morgan (1986) derived a relationship between a continuous wet-dry renewal model with Gamma distributed dry intervals and a Markov chain model for daily rainfall occurrence. The Markov process model was shown to provide a good representation in certain parts of the United States while in other areas, where the Markov model is inappropriate due to event clustering or other phenomena, the Gamma model provides an improved characterisation of the relationship between continuous and discrete rainfall occurrence.

Foufoula-Georgiou and Lettenmaier (1987) developed a

Markov renewal model for rainfall occurrences in which the time between rainfall occurrences were sampled from two different geometric distributions. The transition from one distribution to the other was governed by a Markov chain. Smith (1987) introduced a family of models termed Markov-Bernoulli processes that might be used for rainfall occurrences. The process consists of a sequence of Bernoulli trials with randomised success probabilities described by a first-order, two-state Markov chain. At one extreme the model is a Bernoulli process, at the other a Markov chain.

A binary discrete autoregressive moving average (DARMA) process was first used by Buishand (1977). He found that an alternating renewal process was superior to the DARMA model for the data from The Netherlands but the DARMA model looked more promising in tropical and monsoonal areas. Chang *et al.* (1984) and Delleur *et al.* (1989) used four seasons for two stations in Indiana (USA) and found that either the first-order autoregressive or the second-order moving average model was appropriate for different seasons. Buishand (1977) pointed out that the properties of the rainfall in New Delhi cannot be preserved by a model with constant parameters — stochastic parameters are required. This observation may be generally valid for regions with monsoonal climates.

Chapman (1994) compared five models, namely, Markov chains of orders 1, 2 and 3 (MC1, MC2 and MC3), truncated negative Binomial distribution (TNBD) and the truncated Geometric distribution (TGD) with separate parameter values for each month using data from 17 Australian rainfall stations. Three of the above models (MC1, TNBD and TGD) were also compared with parameters varying smoothly throughout the year according to a Fourier series having 0, 1 and 2 harmonics. The Fourier series representation with one harmonic for parameter variation throughout the year using the MC1 or TGD model was successful for five stations with high rainfall in southern Australia. The monthly MC2 model or monthly TNBD model fitted the remaining stations best. Different record lengths appear to affect the selection of the best model, particularly when wet and dry spells are considered separately. For combined results, different models were selected for the 20-year and 50-year records in four out of ten cases, for the 20- and 100-year records in two out of five cases, and for the 50- and 100-year records in one case out of five. He concluded that the prospects for regionalisation of parameters are poor unless there is a good sample of long records. In a later study, Chapman (1997) compared the above distributions and Markov chain for the rainfall stations from 22 islands in the Western Pacific. He concluded that a first-order Markov chain or truncated geometric distribution with a Fourier series representation for parameter variation over months

was successful for stations with a latitude greater than  $14^\circ$ . For the stations close to the Equator, the seasonal regularity was less important and the models with individual monthly values or constant value throughout the year performed well.

#### RAINFALL AMOUNT MODELS

Models used for daily rainfall amounts include the two-parameter Gamma distribution (Jones *et al.*, 1972; Goodspeed and Pierrehumbert, 1975; Coe and Stern, 1982; Richardson, 1981; Woolhiser and Roldan, 1982), mixed Exponential distribution (Woolhiser and Pegram, 1979; Woolhiser and Roldan, 1982, 1986), a skewed Normal distribution (Nicks and Lane, 1989) and a truncated power of Normal distribution (Bardossy and Plate, 1992; Hutchinson *et al.*, 1993). Cole and Sherriff (1972) applied separate models to rainfalls for a solitary wet day, the first day of a wet spell and the other days of a wet spell, while Buishand (1977) related the mean rainfall amount on a wet day to its position in a wet spell such as a solitary wet day, wet day bounded on one side by a wet day and wet day bounded on each side by a wet day. Recently, Chapman (1998) investigated the impact of adjoining wet days on the distribution of rainfall amounts and found that the models which take this into account resulted generally in a better fit than the models which lump the data together.

Chin and Miller (1980) examined the possible conditional dependence of the distribution of daily rainfall amounts on the occurrence of rainfall on the preceding day using 25 years of daily rainfall data at 30 stations in the continental United States. They concluded that, except for the winter season in the Pacific northwest, the distribution of daily rainfall did not depend on whether the preceding day was wet or dry.

Recently, Menabde and Sivapalan (2000) used Levy-stable distributions to fit the storm duration and rainfall totals. They showed that this distribution having a fat tail fits the storm duration and amounts better than the Exponential or Gamma in the tail. The traditional goodness-of-fit tests may not be able to show this as the small number of large values are outweighed by the large number of smaller values.

Chapman (1994, 1998) compared the following five models for rainfall amounts, the Exponential (one parameter), the mixed Exponential (three parameters), the Gamma (two parameters), a skewed Normal (three parameters) and the Kappa distribution (two parameters). Based on the AIC, the ranking of the models was consistent, the best being the skewed Normal distribution, followed by the mixed Exponential, the Kappa, the Gamma, and last the Exponential. There was also consistency in the model

selected for different groups of data (solitary wet days, first day of a wet spell etc.). He observed little variation in the coefficient of variation between different groups and relatively little between months. Yevjevich and Dyer (1983) suggested that the latter feature may be a general characteristic of daily rainfall series and this could lead to a significant parsimony in the number of parameters to model seasonal variations.

Wang and Nathan (2000) developed a daily and monthly mixed (DMM) algorithm for the generation of daily rainfall. Daily rainfall data are generated month by month using the usual two-part model with two sets of parameters for the Gamma distribution, one estimated from the daily rainfall data and the other from monthly rainfall data. The monthly total is obtained by summing the daily values generated from the monthly Gamma parameters and adjusted for serial correlation. The generated daily rainfalls from the daily Gamma parameters are linearly scaled to match the serially correlated monthly rainfalls. Results for the Lake Eppalock catchment rainfall and for six other sites around Australia showed that the DMM algorithm reproduced the mean, coefficient of variation and skewness of daily, monthly and annual rainfall. The results were examined in detail for the Lake Eppalock catchment (in southern Australia); the algorithm worked well in reproducing the mean, coefficient of variation and skewness of monthly maximum daily rainfall, but not as well for the annual maximum rainfall. For the other six sites, the algorithm worked well in reproducing the mean and coefficient of variation but not as well in reproducing the skewness of the annual maximum daily rainfall.

#### TRANSITION PROBABILITY MATRIX MODELS

Allen and Haan (1975) used a multi-state ( $7 \times 7$ ) Markov chain model and employed a Uniform distribution for each of the wet states except for the last, for which an Exponential distribution was used. Due to the lack of sufficient data items in the last class for each month, the values in this class were lumped together and only one value of the Exponential parameter was estimated to generate the rainfall depth in the last class for all twelve months. Selvalingam and Miura (1978) modified the above procedure by having twelve parameters for the Exponential distribution, but they obtained the parameters empirically until the model reproduced the daily maximum monthly rainfalls adequately. Srikanthan and McMahon (1983a, 1985) used a linear distribution for the intermediate classes and a Box-Cox transformation for the last state. The number of states in each month was varied to obtain an adequate number of items in the last state.

Chapman (1994, 1997, 1998) compared the Srikanthan and McMahon model with the best selected rainfall occurrence and amount models; the latter performed better than the former in five out of 15 twenty year records, for two of the ten 50-year records and for none of the 100-year records. He also found that the Srikanthan and McMahon model, which does not use any wet day classification, was successful in reproducing the mean, standard deviation, skew and the number of wet days in each class.

Boughton (1999) observed that the Transition Probability Matrix (TPM) model (Srikanthan and McMahon, 1985) underestimates the standard deviation of annual rainfall and proposed an empirical adjustment to match the observed standard deviation. The adjustment factor ( $F$ ) is obtained by trial and error until the standard deviation of the generated and observed annual rainfall matches. The generated daily rainfall in each year is multiplied by the following ratio:

$$\text{Ratio}_i = \{M + (T_i - M)F\}/T_i \quad (5)$$

where  $M$  = the observed mean annual rainfall, and  
 $T_i$  = the generated annual rainfall for year  $i$ .

This adjustment simply matches the observed standard deviation by decreasing the rainfall in below-mean years and increasing the rainfall in above-mean years. It does not improve the persistence in the annual rainfalls as in Thyer and Kuczera (1999).

#### RESAMPLING MODELS

Lall *et al.* (1996) developed a non-parametric, wet-dry spell model for re-sampling daily rainfall at a site. All marginal, joint and conditional probability densities of interest (dry spell length, wet spell length, precipitation amount and wet spell length given prior to dry spell length) are estimated non-parametrically using at-site data and kernel probability density estimators. The model was applied to daily rainfall data from Silver Lake station in Utah (USA) and the performance of the model was evaluated using a number of performance measures. The model reproduced satisfactorily the wet day precipitation, wet spell length and dry spell length.

Rajagopalan *et al.* (1996) presented a non-homogeneous Markov model for generating daily rainfall at a site. The first-order transition probability matrix was assumed to vary smoothly day by day over the year. A kernel estimator was used to estimate the transition probabilities through a weighted average of transition counts over a symmetric time interval centred at the day of interest. The rainfall amounts on each wet day were simulated from the kernel probability

density estimated from all wet days that fall within a time interval centred on the calendar day of interest over all the years of available data. Application of the model to daily rainfall data from Salt Lake City, Utah, showed that the wet- and dry-spell attributes and the rainfall statistics were reproduced well at the seasonal and annual time scales.

Sharma and Lall (1997, 1999) used a nearest-neighbour conditional bootstrap for re-sampling daily rainfall for Sydney. The dry spell lengths were conditioned on the number of days in the previous wet spell and the wet spell lengths were conditioned on the number of days in the previous dry spell. The rainfall amounts were conditioned on two variables, the rainfall amount on the previous day and the number of days from the start of the current spell. Results from the model showed its ability to simulate sequences that are representative of the historical record.

A limitation of the non-parametric density estimation approach is the rather limited extrapolation of daily rainfall values beyond the largest value recorded. The simulations from the k-nearest-neighbour method do not produce values that have not been observed in the historical data and this is a major limitation if extreme values outside the available record are of interest (Rajagopalan and Lall, 1999). Sample sizes needed for estimating the probability density function of interest are likely to be larger than for parametric estimation. The non-parametric methods have been tested on a limited range of sites and testing over a greater range of climates is needed for broader applicability. It may be less amenable to direct regionalisation which is done in terms of the parameters of a parametric model.

#### TIME SERIES MODELS OF THE ARMA TYPE

In this approach, time series models similar to stream flow data generation are used to generate daily rainfall data. Adamowski and Smith (1972) used a first-order Markov model to generate standardised daily rainfall data. The major problem with this procedure is the cyclical standardisation which occurs if there are numerous zero daily values. A truncated power of Normal distribution has been suggested to model daily rainfall (Hutchinson *et al.*, 1993; Hutchinson, 1995). The underlying Normal distribution can be put into a simple first-order autoregressive scheme to account for the day-to-day persistence of wet and dry days. The lag-one autocorrelation can be specified by matching the conditional probability  $P(D|D)$ . The correlations in the amounts of rainfall on successive wet days from this model were found to be much larger than the observed correlations in the rainfall and, to a first approximation, could be ignored (Hutchinson, 1995). Such systematic differences between correlations based on occurrence and intensity have not been

recognised in the applications of such models as described in Bardossy and Plate (1992).

#### CONDITIONAL DAILY RAINFALL MODELS

Stochastic models of daily rainfall with annually varying parameters usually do not preserve the variance of monthly and annual precipitation (Buishand, 1977; Zucchini and Adamson, 1984; Woolhiser *et al.*, 1993; Boughton, 1999). This underestimation may be due to real long-term trends in rainfall, changes in the data collection techniques or in rain gauge exposure, model inadequacies, and/or the existence of large-scale atmospheric circulation patterns that do not exhibit annual periodicities (Woolhiser, 1992). One such phenomenon that has attracted recent scientific interest is the Southern Oscillation Index (SOI). Woolhiser (1992) proposed a technique to identify the effects of ENSO on rainfall. The rainfall occurrence was described by a first-order Markov chain and the mixed Exponential distribution was used for the rainfall amount on wet days. The parameters of the Markov chain and the mixed Exponential distribution are perturbed by a lagged linear function of SOI.

$$G'_i(t) = G_i(t) + b_i S(t-t_i) \quad (6)$$

where  $b_i$  and  $t_i$  are parameters to be estimated from the data and  $S(t)$  is the SOI on day  $t$ . Both the parameters of the Markov chain and the mean,  $m(t)$ , of the mixed Exponential distribution were affected by the SOI. A monthly SOI series was used, so that  $S(t)$  is represented as a step function.

Data from 11 stations in Arizona, Idaho and Oregon (USA) were analysed by Woolhiser (1992). Perturbing the periodic logits of the transition probabilities resulted in a minimum AIC for six stations, with the Arizona stations being affected most strongly. The signs of the coefficients are fairly consistent with previous studies, with a negative SOI leading to more rainfall in the southwest and the opposite effect in the Pacific northwest. The perturbed mean precipitation resulted in the minimum AIC for all stations and the sign of the coefficient was consistent with expectations except for the station Bose, Indiana. The most common lag was about 90 days.

Woolhiser *et al.* (1993) applied the above procedure to 27 stations in California, Nevada, Arizona and New Mexico. Perturbations of the logits of the dry-dry transition probabilities resulted in statistically significant improvements in the log likelihood functions for 23 stations and perturbations of the mean daily rainfall resulted in significant increases for 18 stations. The most common lag identified was 90 days, suggesting the possibility of conditional simulations of daily precipitation. Even though

the simulated rainfall sequences with model parameters perturbed by the SOI exhibited greater monthly and annual variances than those simulated with purely periodic parameters, these variances were still underestimated.

Hay *et al.* (1991) presented a method of modelling rainfall as a function of weather type. A Markov-based model was used to generate temporal sequences of six daily weather types, high pressure, coastal return, maritime tropical return, frontal maritime tropical return, cold frontal over-running and warm frontal over-running. Transitions from one weather type to another weather type were modelled using a Markov chain. The length of time, in days, a given weather type persisted was modelled by a Geometric distribution. Observed monthly probabilities of rainfall for each weather type were used to classify a day as wet or dry. The rainfall amounts were modelled using the product of an Exponential random variable and a Uniform random variable because an Exponential distribution alone underestimated the variance of the daily rainfall. The rainfall amounts were modelled as:

$$R = [I_i(-\log(U))](1 + e) \quad (7)$$

where  $I_i$  = the mean intensity of rainfall for wet days for the given weather type  $i$ ,

$U$  = a uniform random variable between 0 and 1, and

$e$  = the error term, a Uniform random variable between -1 and 1.

When there were less than ten days of recorded rainfall for a given weather type and month, the rainfall amounts were modelled using,

$$R = I_i(1 + e) \quad (8)$$

A Monte Carlo simulation consisting of 50 replicates of 30-year sequences reproduced daily weather type and precipitation sequences similar to those of the observed record.

Wilks (1989) developed a daily rainfall model in which the parameters of the Markov chain and the Gamma distribution were estimated separately for months in the lower 30% (dry), middle 40% (near normal) and the upper 30% (wet) of the distribution of monthly rainfall. The transitions among dry, near normal and wet months were modelled by a three-state, first-order Markov chain. This conditional model was compared to the usual unconditional model derived from the entire data set using data from ten North American stations. It was found that the unconditional models produced too few dry and wet months compared to the observations, while the conditional model reproduced



the climatological distribution of the monthly rainfall. Wilks used generalised likelihood ratio tests to show that the increase from four to ten parameters per month was justified by the data.

#### REGIONALISATION OF DAILY RAINFALL MODEL PARAMETERS

The main purpose of regionalisation of model parameters is to obtain the parameters for locations with no observation data as input to rainfall simulation models. Geng *et al.* (1986) developed empirical equations for the parameters of a two-part model using the data from The Netherlands (Wageningen), the Phillipines (Los Banos) and the USA (Colombia, Boise, Miami, Phoenix and Boston).

$$P(W|D) = 0.75 P(W) \quad (9)$$

$$P(W|W) = 0.25 + 0.75 P(W) \quad (10)$$

$$b = -2.16 + 1.83 m_w \quad (11)$$

$$a = m_w / b \quad (12)$$

where  $a$  and  $b$  are the shape and scale parameters of a Gamma distribution and  $m_w$  is the average amount of rainfall per wet day. They proposed that these empirical equations allow rainfall simulation models to be used for crop growth studies in many areas where too few weather data were available.

Woolhiser and Roldan (1986) investigated the seasonal and regional variability of parameters of stochastic daily precipitation models for South Dakota, USA. Fourier series were used to describe the seasonal variation of the five parameters of the Markov chain mixed Exponential model fitted to 16 rainfall stations. A concise description of seasonal variations of parameters was obtained by using from 15 to 27 coefficients. Semivariograms calculated for the mean Markov chain parameters showed a nugget effect. The large nugget variance was attributed to real differences in precipitation regime and to inconsistencies in the records due to methodological differences affecting small precipitation amounts. Time of observation appeared to be an important factor. They suggested that rainfall records for use in regional parameter mapping must be screened carefully to ensure consistency of data. The model parameters for four test stations were estimated more closely by arithmetic averages of six nearby stations than by three other interpolation techniques, including nearest-neighbour, spline fitting and linear interpolation. They also found that the parameters interpolated for the four test sites were

significantly different from parameters estimated from rainfall records.

#### DAILY RAINFALL DATA AT A NUMBER OF SITES

If hydrological and land management changes are required simultaneously across regions, then the spatial dependence between the weather inputs at different sites has to be accommodated. This is particularly important to the simulation of rainfall, which displays the largest variability among meteorological variables in time and space. The model used to generate daily rainfall at a number of sites can be grouped broadly into three categories: conditional models, extension of Markov-chain models and random cascade models.

#### CONDITIONAL MODELS

Zucchini and Guttorp (1991) constructed a hidden Markov model for the occurrence/nonoccurrence of rainfall at  $N$  sites by assuming a different probability of events at the sites for each of a number of unobservable climate states. The climate process is assumed to follow a Markov chain. The method was illustrated by applying it to data for one site in Washington and to data from five sites in the Great Plains, USA.

Bardossy and Plate (1991) developed a semi-Markov chain model for atmospheric circulation patterns and linked it to the occurrence of rainfall using transition probabilities. Several series of circulation patterns and corresponding rainfall occurrences were simulated and the statistics of the simulated and the observed data were similar.

Wilson *et al.* (1992) developed a stochastic model of weather states and daily rainfall at multiple rainfall sites. Four classification techniques were investigated to obtain a single index of the regional weather state for each day of the study period. Once the weather classification scheme was selected, the daily occurrence process of the weather states was modelled by a semi-Markov model with either geometrically or mixed geometrically distributed lengths of stay in each weather state. A hierarchical modified Polya Urn model was developed to model the rainfall occurrence at multiple stations. The hierarchical structure comes about by conditioning the first station on the day's weather class, the second station on the weather class and the occurrence/non-occurrence of rain at the first station and so on. The rainfall amounts were modelled using a mixed Exponential distribution for each station within each season within each weather class. The rainfall amounts for each station were simulated simultaneously, based on the correlation structure between the station amounts. The model was able to

reproduce the probability distribution of daily rainfall amounts reasonably well but with some downward bias.

Charles *et al.* (1999) extended the non-homogeneous hidden-state Markov model (NHMM) of Hughes *et al.* (1999) by incorporating rainfall amounts. The joint distribution of daily rainfall at  $n$  sites was evaluated through the specification of  $n$  conditional distributions for each weather state ( $s = 1, \dots, N$ ). The conditional distribution consisted of regressions of inverse Normal transformed amounts at a given site on rainfall occurrence at neighbouring sites within a given radius ( $d$  km). An automatic variable selection procedure was used to identify the key neighbouring sites. The precipitation model can be expressed as

$$z_s^{(i)} = \theta_{0s}^{(i)} + \sum_{k \in n_i(d)} \theta_{ks}^{(i)} r^{(k)} + \varepsilon_s^{(i)} \quad i = 1, \dots, n \quad (13)$$

where the  $\theta_{ks}^{(i)}$  are regression parameters,  $n_i(d)$  denotes the set of indices of the key neighbouring sites for site  $i$ ,  $\varepsilon_s^{(i)}$  is an error term modelled stochastically by assuming  $\varepsilon_s^{(i)} \sim N(0, \sigma_s^2(i))$ , and

$$z_s^{(i)} = \Phi^{-1} \{ F(y_s^{(i)}) \} \quad (14)$$

in which  $F$  denotes the Normal cumulative distribution function and  $F(y_s^{(i)})$  is the empirical distribution function of  $y_s^{(i)}$ , the rainfall amount on days with  $r^{(i)} = 1$ .

The above method was applied to a network of 30 daily rainfall stations and historical atmospheric circulation data in southwestern Australia. A year was divided into winter (May – October) and summer (November – April) seasons. A six-state NHMM was found adequate and reproduced the dry and wet spells satisfactorily. Only the Spearman rank inter-site correlations were compared for the rainfall amounts. The results for the summer season were not presented.

Pegram and Seed (1998) developed a space–time model for the generation of daily rainfall over a 256 km<sup>2</sup> region near Bethlehem, South Africa. The model has two components,

- a climate generator in the form of a three-state Markov chain with periodically varying parameters
- bins of rainfall data (a collection of historical rainfall dates on which the various types of rain occurred).

The daily weather was classified into three types based on the number of rain gauges reporting rainfall (see Table 1).

Starting from a known current state, the model determines

Table 1. Daily weather classification (Pegram and Seed, 1998)

Dry	< 3 % gauges report rain
Scattered	> 3 % gauges report rain, but < 50 % report > 5 mm rain
General	> 50 % gauges report > 5 mm rain

the state of the following day using the transition probabilities. If the state is dry, assign zero rainfall to all gauges. If the state is scattered, select an historical date at random from the collection of scattered rain days in the appropriate month. Look up the set of data for that date which in the current model is the mean areal value of rainfall and record it. If the type is general, note how many general rain days are in the current generated sequence of general rain days and resample the state for the day after next. If the current state is general and the next one is other than general, select a rain-day sequence from the sets of runs of one to five days and record that sequence as it occurred historically. The model will thus produce a sequence of daily averages of rainfall based on the historical record. The actual rainfalls that fell at all active rain gauges on rain days are obtained from the historical data. The model is a form of extended Bootstrap.

Bardossy and Plate (1992) developed a multi-dimensional stochastic model for the space-time distribution of daily rainfall linked to atmospheric circulation patterns using conditional distributions and conditional spatial covariance functions. The model is a transformed multivariate first-order autoregressive model with parameters depending on the atmospheric circulation patterns. The negative values are declared as dry days.

The model was applied to the rainfall data recorded at 44 stations in the Ruhr River catchment (5000 km<sup>2</sup>) using the classification scheme of the German weather service. The parameters were estimated from the moments of the observations. The model reproduced the point and spatial rainfall statistics, including rainfall covering only part of the total area under study.

#### EXTENSION OF SINGLE SITE MARKOV CHAIN MODELS

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. Individual models are fitted to each of the sites ( $k$ ) first. The collection of individual site models is driven by vectors of Uniform  $[0, 1]$  variates  $\mathbf{u}_i$  and  $\mathbf{v}_i$  whose elements,  $u_i(k)$  and  $v_i(k)$  respectively, are cross-correlated so that  $corr[u_i(k), u_i(l)] \neq 0$  and  $corr[v_i(k), v_i(l)] \neq 0$ , and are serially and mutually independent  $corr[u_i(k), v_i(l)] = corr[u_i(k), u_{i+1}(l)] = corr[v_i(k), v_{i+1}(l)] = 0$ . Non-zero correlations among the elements of  $\mathbf{u}_i$  and  $\mathbf{v}_i$  result in inter-site correlations between the generated rainfall sequences.

For a network of  $N$  stations, there are  $N(N-1)$  pairs of correlations for each of the rainfall occurrence and rainfall amount processes. The correlated uniform variates required,  $u_i(k)$ , are obtained from the Gaussian variates,  $w_i(k)$ , through the transformation  $u_i(k) = F[w_i(k)]$  where  $F$  is the normal cumulative function. The estimation of the correlations between the Gaussian variates is carried out by simulation.

The model was applied to a network of 25 rainfall stations in New York state (USA) with inter-station separations ranging from 10 to 500 km. The model reproduced reasonably well the various aspects of the joint distribution of daily rainfall at the modelled stations. The mixed Exponential distributions provided a substantially better fit than the more conventional Gamma distribution and was convenient for representing the tendency for smaller amounts at locations near the edge of the wet areas. Means, variances and interstation correlations of monthly rainfalls were also reproduced well. In addition, the use of mixed Exponential rather than Gamma distribution resulted in the interannual variability being closer to that observed.

RANDOM CASCADE MODELS

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 non-homogeneous random cascade process which disaggregates the above 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 enable the model to capture systematic spatial gradients exhibited by measured data. If the initial area (at level 0) is assigned an average intensity  $R_o$  (in mm/day, as simulated by the temporal model), this gives an initial volume  $R_o L_o^2$ , where  $L_o$  is the outer length scale. At the first level, the initial area is subdivided into four sub-areas denoted by  $\Delta_1^i$ ,  $i = 1, \dots, 4$ . At the second level, each of the above sub-areas is further subdivided into four further subareas denoted by

$\Delta_2^i$ ,  $i = 1, \dots, 16$ . When the process of subdivision is continued, the volume  $\mu_n(\Delta_n^i)$  in the sub-areas at the  $n^{\text{th}}$  level of subdivision ( $\Delta_n^i, i = 1, \dots, 2^n$ ) are given by

$$\mu_n(\Delta_n^i) = R_o L_o^2 2^{-n} \prod_{j=1}^n W_j^i \tag{15}$$

where for each  $j$ ,  $i$  represents the sub-areas along the path to the  $n^{\text{th}}$  level sub-areas and the multipliers  $W$  are non-negative random cascade generators with  $E[W] = 1$ . The so-called beta-lognormal model was used for the generation of the cascade generators  $W$  (Over and Gupta, 1994, 1996).

$$W = BY \tag{16}$$

where  $B$  is a generator from Beta model and  $Y$  is drawn from a Lognormal distribution (Gupta and Waymire, 1993). To include the systematic spatial variation in the rainfall,  $W$  in (22) is modified to include a systematic multiplier  $G$

$$W = BYG \tag{17}$$

with the condition that the average value of  $G$  over the respective sub-areas is equal to 1 at every discretisation step.

The model was applied to a 400×400 km region encompassing the Swan-Avon River Basin in the southwest of Western Australia. The model parameters were estimated from 11 years of daily rainfall data observed at 490 rain gauges located in the region. The generated regionally-averaged rainfall was disaggregated progressively down to the scale of 12.5 km. The model was able to reproduce (1) the spatial patterns of long-term mean daily, monthly and annual rainfall; (2) the spatial patchiness characteristics of daily rainfall, estimated in terms of a wet fraction; (3) statistical characteristics relating to storm arrival and inter-arrival times at a selected number of stations; and (4) probability distributions and exceedence probabilities of rainfall at selected stations for selected months. The model under-predicted the mean number of wet days and the mean wet spell lengths, especially during the winter months. A possible reason given for this is the exclusion of space-time correlations in the model.

Climate data

One major use of climate data in conjunction with rainfall data is in computer simulation of hydrological and agricultural systems. Rainfall-runoff models and crop growth models require, in addition to rainfall, net radiation or evaporation as a measure of energy input. In irrigation simulation studies, both rainfall and evaporation are also required.

A special characteristic that must be preserved in stochastic modelling of climate data is the cross-correlation between variables. The models for generating climate data at annual, monthly and daily time intervals are reviewed in this section.

#### ANNUAL AND MONTHLY CLIMATE DATA

Annual climate data for a single site and at a number of sites can be generated by using a multi-site type model (Matalas, 1967; Young and Pisano, 1968). Monthly climate data can be obtained by disaggregating the generated annual data as mentioned under monthly rainfall data, but there does not appear to be much use for such sequences.

#### DAILY CLIMATE DATA AT A SINGLE SITE

Jones *et al.* (1972) hypothesised that daily temperature and evaporation could be obtained from the time of the year and the occurrence of rainfall on both the day in question and the preceding day. Daily temperature and evaporation were simulated by Monte Carlo type sampling from a Normal distribution, with parameters chosen according to the time of the year and to the state of the present and preceding days. The main drawback with this procedure is that the skewness, cross-correlations and auto-correlations of daily temperature and evaporation values are ignored.

Edelsten (1976) proposed a similar model with additional day states, which depended on temperature as well as rainfall, and fitted a second order Markov model. He also incorporated significant cross-correlations and auto-correlations for minimum and maximum temperature. The model simulated adequately most of the cross-correlations and lag one auto-correlations. The drawback with this model is the large number of parameters needed for a second order Markov model (Hutchinson, 1987).

Nicks and Harp (1980) generated daily temperature and solar radiation data using a first order Markov model dependent on the state of the present and preceding days. A Normal distribution was assumed for the temperature and solar radiation whose means and standard deviations were conditioned on the type of day and month of year. This model can be modified to account for the skewness (Srikanthan and McMahon, 1983b), but the modified model will not preserve the cross-correlations.

Bruhn *et al.* (1980) presented a relatively parsimonious model which modelled, on a monthly basis, only those cross-correlations and lag one auto-correlations which were found to be significant. The maximum and minimum temperatures were conditioned on the wet/dry status of the preceding day only, while the solar radiation was conditioned on the wet/

dry status of the present day only. The Normal distribution was used for all variables since only solar radiation on dry days appeared to deviate significantly from normality.

To reduce the number of parameters, Larsen and Pense (1981) fitted three parameter Sine curves to the mean daily maximum and minimum temperatures conditioned on the wet/dry state of the present day. The residuals of these two variables from their mean values for the two types of days were modelled by two bi-variate Normal distributions. A Gamma distribution for dry days and a Beta distribution for wet days were used to model differences of solar radiation from the theoretical clear day values, which depend on the latitude and the time of year. The model did not take into account the cross-correlations.

Richardson (1981) adopted a weakly stationary multivariate model to generate the residual series of maximum and minimum temperatures and solar radiation. The residuals were assumed normally distributed and conditioned only on the state of the present day. This model preserves the cross-correlations and auto-correlations as follows:

$$X_t = A X_{t-1} + B e_t \quad (18)$$

where  $X_t = (3 \times 3)$  matrix of standardised daily climate data for day  $t$

$e_t = (3 \times 1)$  vector of independent random deviates with zero mean and unit variance

$A, B = (3 \times 3)$  matrices of constant coefficients to preserve the auto-correlations and the cross-correlations.

The matrices  $A$  and  $B$  are determined using the matrices of lag 0 and lag 1 correlations among the three elements of  $X_t$ , using a scheme similar to that of Matalas (1967), and are assumed to be equal for wet and dry days. A common implementation of this algorithm treats  $A$  and  $B$  as being constant in time and equal for all locations in the conterminous United States (Richardson and Wright, 1984). Wilks (1992) claims that this assumption is dubious, particularly for those elements depending strongly on the cross-correlations between solar radiation and temperature variables.

Using Australian data, Guenni *et al.* (1990) also found no general support for the constancy of correlation between air temperature and solar radiation. The correlations are dependent on season and location. They also observed weak dependence of temperature on wet or dry status of the day and stressed the importance of cloudiness on temperature. An alternative, more realistic model, was recommended in which seasonal fractional cloudiness is generated first and then temperature and solar radiation are generated conditional on the amount of cloud cover for a particular day.

Srikanthan (1985) modified the model given by Eqn. (18) so that it is conditional on the present and preceding days and also takes into account the skewness through the Wilson-Hilferty transformation. The model was applied to daily evaporation, maximum and minimum temperatures and solar radiation data for Melbourne. The simulated climate data were found to have similar statistical characteristics to those of the historical data.

Racsco *et al.* (1991) first reduced the four-dimensional weather process (average daily temperature, solar hours, rainfall and relative humidity) to a three-dimensional sub-process because of the very high negative correlation between the relative humidity and the temperature. Rainfall was considered as the independent variable and modelled first. The wet and dry spells were computed on a daily basis but using a characteristic interval  $[d-14, d+14]$ . The seasonal variation was handled through a Fourier series representation. The rainfall amounts were divided into three groups (0.1–0.3, 0.3–20, > 20 mm) and modelled with a Uniform distribution for the first group and an Exponential distribution for the second group. An average value was used for the third group. The average temperature and solar hours were generated separately for wet and dry days using a first order autoregressive model with a Normal distribution. The cross-correlations between the average temperature and solar hours were not modelled. The model was applied to two sites in Hungary.

Young (1994) described a multivariate chain model for simultaneously simulating daily maximum and minimum temperatures and rainfall. The multivariate chain model sets up a discriminant space defined by the daily maximum and minimum temperatures and rainfall. Each day of the historical data set is represented by a point in discriminant space, located by its temperatures and rainfall amount. For a given current day, the following day is selected randomly from a set of nearest-neighbours. The model was tested on daily data for Tucson and Safford, Arizona (USA), for the period 1948–1988. A slight tendency to underestimate the variance of monthly average temperatures was noted. The distribution of monthly temperature extremes was reproduced well, with the exception of a tendency to underestimate the warmest minimum temperatures and the coolest maximum temperatures. There was very little difference between the simulated and observed distributions of the diurnal range. The median and 90th percentile of monthly rainfall were well reproduced. A tendency to underestimate the frequency of dry months was observed. The frequency of runs of dry and wet days of different lengths of simulated data was found not to differ significantly from that of the observed data.

Rajagopalan *et al.* (1997a) presented a nonparametric

multivariate resampling scheme for generating daily weather variables at a site. The model samples the original data with replacement while smoothing the empirical conditional distribution function. Rainfall is generated from the nonparametric wet/dry spell model (Lall *et al.*, 1996). A vector of solar radiation, maximum temperature, minimum temperature, dew point temperature and wind speed is then simulated by conditioning the vector of these variables on the preceding day and the rainfall amount on the day of interest. The model was applied to 30 years of daily weather data at Salt Lake City, Utah, USA. The results showed that the means and the quantiles are well reproduced. The standard deviation and coefficients of variation and skewness are not well reproduced as the kernel methods inflate the variance by  $(1 + h^2)$  where  $h$  is the band width. This can be corrected by appropriate scaling, but this was not carried out. The correlations from the simulations and historical data seem to be different in a number of cases. Correlations with the rainfall were the most poorly reproduced.

Rajagopalan and Lall (1999) developed a multivariate  $k$ -nearest-neighbour method with lag one dependence for six daily weather variables. This model improves the kernel based approach developed above (Rajagopalan *et al.*, 1997a). A vector of solar radiation, maximum temperature, minimum temperature, dew point temperature and wind speed on a day of interest is resampled from the historic data by conditioning through the vector of the same variables (feature vector) on the preceding day. The re-sampling is done from the  $k$ -nearest-neighbours in state space of the feature vector using a weight function. The model was applied to 30 years of daily weather data at Salt Lake City and the results were compared with those from the application of a multivariate autoregressive (MAR) model similar to that of Richardson (1981). The model reproduced satisfactorily the moments, quantiles, dry and wet spells and the correlations for all four seasons.

#### DAILY CLIMATE DATA AT MULTIPLE SITES

Provided that the problems in accounting for the complicated covariance structure in daily rainfall anomalies based on a truncated power of Normal distribution can be overcome, it is relatively straightforward to incorporate other weather variables using standard multivariate normal models (Hutchinson, 1995). This also depends on the adequacy of the Normal distribution in modelling the remaining variables.

## Uncertainty in model parameters

In most of the above models, the estimated parameter values are assumed to be the “true” values and the sampling errors are ignored. Klemes and Bulu (1979) showed that this traditional approach to generating stochastic data underestimates the range of mean and variance of the likely future sequences. Several researchers have used Bayesian methods to overcome this deficiency.

Vicens *et al.* (1975) incorporated parameter uncertainty into an annual data generation model for a single site using Bayesian techniques. Valdes *et al.* (1977) extended the above model for a multivariate case. McLeod and Hipel (1978) claim that the above methods for the incorporation of parameter uncertainty do not result in generated data that resemble the historical data. McLeod and Hipel (1978) proposed a correct method for incorporating parameter uncertainty based on large sample theory. Thyer (2000) undertook a full Bayesian analysis to derive the posterior distribution of the hidden state Markov model and AR(1) model parameters using Markov chain Monte Carlo (MCMC) method. Recently, a full Bayesian analysis using MCMC simulations was carried out to derive the uncertainties in the parameters of a multisite annual data generation model (Q. J. Wang, pers. comm.). To generate monthly data which reflect the uncertainty in the mean and variance correctly, Stedinger *et al.* (1985) developed a condensed disaggregation model. Chaouche and Parent (1999) used a Bayesian approach to estimate the uncertainty in the parameters of a two-part daily rainfall model utilising a Markov chain and a Gamma distribution.

## Rainfall and climate data under climate change scenario

Concern has increased in recent years over climate change caused by increasing concentration of CO<sub>2</sub> and other trace gases in the atmosphere. A major effect of climate change may be alterations in regional hydrological cycles and changes in regional water availability. The use of modified water balance models offers many advantages in evaluating the regional impacts of global climate change (Gleick, 1989). The main source of climate change projections is general circulation models (GCMs). While current GCMs perform reasonably well in simulating the present climate with respect to annual and seasonal averages over large areas, they are considerably less reliable in specific catchment scale information that are necessary for hydrological studies. As a result, climate change impact studies have had to use a spectrum of climate change scenarios. These are generally constructed using observed records of temperature and

rainfall, adjusted to reflect climate changes obtained from monthly average GCM results.

### ADJUSTMENT OF HISTORICAL DATA

Most of the early work on the impacts of climate change used historical data adjusted for the climate change (Lettenmaier and Gan, 1990; Panagoulia, 1992). Rainfall records were multiplied by the monthly precipitation ratios for the CO<sub>2</sub>-doubling and control runs. The monthly temperature difference between the CO<sub>2</sub>-doubling and control runs was added to the historic temperature data. The potential evapotranspiration (PET) was computed using the Penman equation for two different sets of monthly temperature data for the CO<sub>2</sub>-doubling and control runs, while all other variables (wind speed, humidity, solar radiation, etc.) in the Penman equation remained unchanged. The monthly differences in PET were computed and the resulting differences were then added to the historic PET data (Panagoulia, 1992). Recently, Loaiciga *et al.* (2000) created climate change scenarios as described above to investigate the climate change impacts on a regional karst aquifer in Texas, USA. Mimikou *et al.* (2000) assessed the regional impacts of climate change on water resources by modifying the synthetic series for climate change effects.

### ADJUSTMENT OF MODEL PARAMETERS

Wilks (1992) presented a method to adapt stochastic daily weather generation models for generation of synthetic daily time series consistent with assumed future climates. The climates assumed were specified by the monthly means and variances of rainfall and temperature.

When parameters are changed in a conditional model, certain unanticipated effects can be produced. For instance, modifying the probability of occurrence of daily rainfall not only changes the mean of daily temperature, but also its variance and autocorrelation as well. Katz (1996) derived the theoretical statistical properties of a simplified version of WGEN (Richardson, 1981) and showed how best to adjust the model parameters to obtain the desired climate change.

## Conclusions

Models for the generation of annual, monthly and daily rainfall and climate data were reviewed. If the year-to-year variations or long-term persistence are ignored, models are available to generate annual and monthly rainfall. From past experience, in the case of annual rainfall, a lag-one Markov model is adequate for single site or multiple sites. In the

case of monthly rainfall, the procedures of Porter and Pink (1991) and Maheepala and Perera (1996) are cumbersome and it is not clear how the synthetic monthly data were generated. The modified disaggregation scheme proposed by Mejia and Rousselle (1976) is an elegant option. If the number of parameters in this scheme is too great, the condensed version of Lane (1979) could be used. For the months with a high coefficient of variation ( $> 1$ ), an appropriate transformation should be applied to eliminate the generation of negative values. If the method of fragments is not used, generating the right amount of zero monthly rainfall will be a problem. It is not clear how this aspect was handled by Porter and Pink (1991) and Maheepala and Perera (1996).

There has been a lot of research on the generation of daily rainfall at individual sites. The transition probability method appears to preserve most of the characteristics of daily, monthly and annual characteristics and is shown to be the best performing model (Chapman, 1994, 1997). The main drawback with this method is the large number of parameters, which makes it almost impossible to regionalise the parameters. The two-part model has been shown to perform well by many researchers.

A shortcoming of the existing models is the consistent underestimation of the variances of the simulated monthly and annual totals. Recently, Wang and Nathan (2000) constrained a two-part model within a monthly model and it appears to perform well. Also, Boughton (1999) has adjusted the generated daily rainfalls from the transition probability method by a trial-and-error procedure to match the variance of the observed annual rainfall. As an alternative, conditioning model parameters on monthly amounts (Wilks, 1989) or perturbing the model parameters with the SOI (Woolhiser, 1992) may result in better agreement between the variance of the simulated and observed annual rainfall and these approaches show promise for further investigation.

From the limited amount of work done in generating daily rainfall at a several sites, the approach of Jothityangkoon *et al.* (2000) appears to be promising. The approach of Wilson *et al.* (1992) is hierarchical and becomes difficult to handle for a medium to large number of stations. The method of Bardossy and Plate (1991) uses a censored power Normal distribution and the procedure needs to resolve the problem of correlation based on rainfall occurrences and intensity. The model used for rainfall amounts in Charles *et al.* (1999) is not adequate and is very cumbersome. The extension of single site Markov chain model to multi-sites (Wilks, 1998) appears to be cumbersome in terms of the number of model parameters and in the way the parameters are estimated. The model used by Pegram and Seed (1998) will generate

only the rainfall values, which were present already in the historic record.

As climate data are less variable than rainfall but correlated among themselves and with rainfall, multi-site models have been used successfully to generate annual data. The monthly climate data can be obtained by disaggregating the annual data generated. On a daily time step at a site, climate data have been generated by using a multi-site type model conditional on the state of the rainfall on the present and previous days. The generation of daily climate data at a number of sites remains a challenging problem. If daily rainfall can be modelled successfully by truncated power Normal distribution (Bardossy and Plate, 1992), then the model data can be extended easily to generate daily climate data at several sites simultaneously.

A number of techniques for incorporating parameter uncertainty in annual and monthly models is available. However, estimating the parameter uncertainty in daily models remains a challenging problem.

The greatest uncertainty in modelling climate data under climate change conditions is the uncertainty in future climate predictions. At present, GCMs are able to provide either scenarios or projections of the future climate. If future climate conditions are known with sufficient accuracy, the stochastic climate models available now can be adapted to generate climate for the new conditions.

## References

- Adamowsky, K. and Smith, A. F., 1972. Stochastic generation of rainfall. *J. Hydraul. Eng., ASCE*, **98**, 1935–1945.
- Akaike, H., 1974. A new look at statistical model identification. *IEEE Trans. Auto. Control*, **AC 19**, 716–722.
- Allen, D.M. and Haan, C.T., 1975. *Stochastic simulation of daily rainfall*. Res. Rep. No. 82, Water Resour. Inst., Univ. Kentucky, Lexington, Kentucky, USA.
- Bardossy, A. and Plate, E.J., 1991. Modelling daily rainfall using a semi-Markov representation of circulation pattern occurrence. *J. Hydrol.*, **122**, 33–47.
- Bardossy, A. and Plate, E.J. 1992. Space time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res.*, **28**, 1247–1259
- 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.
- Bruhn, J.A., Fry, W.E. and Fick, G.W., 1980. Simulation of daily weather data using theoretical probability distributions. *J. Appl. Meteorol.*, **19**, 1029–1036.
- Buishand, T.A., 1977. *Stochastic modelling daily rainfall sequences*. Meded. Landbouwhogeschool, Wageningen, **77**, 211pp.
- Caskey, J.E., 1963. A Markov Chain for the probability of precipitation occurrence in intervals of various lengths. *Mon. Weath. Rev.*, **91**, 298–301.
- Chang, T.J., Kavvas, M.L. and Delleur, J.W., 1984. Daily precipitation modelling by discrete autoregressive moving average processes. *Water Resour. Res.*, **20**, 565–580.

- Chauouche, A. and Parent, E., 1999. Bayesian identification and validation of a daily rainfall model under monsoon conditions. *Hydrol. Sci. J.*, **44**, 199–220 (in French).
- Chapman, T., 1994. *Stochastic models for daily rainfall*. Hydrology and Water Resources Symposium, The Institution of Engineers, Australia, National Conference Publication 94/15, 7–12.
- Chapman, T., 1997. Stochastic modelling of daily rainfall in the Western Pacific. *Math. Comput. Simulat.*, **43**, 351–358.
- Chapman, T., 1998. Stochastic modelling of daily rainfall: the impact of adjoining wet days on the distribution of rainfall amounts. *Environ. Model. Software*, **13**, 317–324.
- Charles, S.P., Bates, B.C. and Hughes, J.P., 1999. A spatiotemporal model for downscaling precipitation occurrence and amounts. *J. Geophys. Res.*, **104**, 31657–31669.
- Chin, E.H., 1977. Modelling daily precipitation occurrence process with Markov chain. *Water Resour. Res.*, **13**, 949–956.
- Chin, E.H. and Miller, J.F., 1980. On the conditional distribution of precipitation amounts. *Mon. Weather Rev.*, **108**, 1462–1464.
- Coe, R. and Stern, R.D., 1982. Fitting models to rainfall data. *J. Appl. Meteorol.*, **21**, 1024–1031.
- Cole, J.A. and Sherriff, J.D.F., 1972. Some single and multisite models of rainfall with discrete time increments. *J. Hydrol.*, **17**, 97–113.
- Cox, D.R. and Isham, V., 1994. Stochastic models of precipitation. In: *Statistics for the environment 2, Water issues*, V. Barnett and K.F. Turkman (Eds.), Wiley, New York, 3–18.
- Delleur, J.W., Chang, T.J. and Kavvas, M.L., 1989. Simulation models of sequences of wet and dry days. *J. Irrig. Drain. Eng.-ASCE*, **115**, 344–357.
- Edelsten, P.R., 1976. A stochastic model of the weather at Hurley in south-east England. *Meteorol. Mag.*, **105**, 206–214.
- Eidsvik, K.J., 1980. Identification of models for some time series of atmospheric origin with Akaike's information criterion. *J. Appl. Meteorol.*, **19**, 357–369.
- Feyerherm, A.M. and Bark, L.D., 1965. Statistical methods for persistent precipitation pattern. *J. Appl. Meteorol.*, **4**, 320–328.
- Feyerherm, A.M. and Bark, L.D., 1967. Goodness of fit of Markov chain model for sequences of wet and dry days. *J. Appl. Meteorol.*, **6**, 770–773.
- Foufoula-Georgiou, E. and Lettenmeier, D., 1987. Markov renewal model for rainfall occurrence. *Water Resour. Res.*, **23**, 875–884.
- Gabriel, K.R. and Neumann, J., 1962. A Markov chain model for daily rainfall occurrences in Tel Aviv, Israel. *J. Roy. Meteorol. Soc.*, **88**, 90–95.
- Garbutt, D.J., Stern, R.D., Dennet, M.D. and Elston, J., 1981. A comparison of rainfall climate of eleven places in West Africa using a two-part model for daily rainfall. *Arch. Met. Geoph. Biokl., Ser. B.*, **29**, 137–155.
- Gates, P. and Tong, H., 1976. On Markov chain modelling to some weather data. *J. Appl. Meteorol.*, **15**, 1145–1151.
- Geng, S., Penning De Vries F.W.T. and Supit, I., 1986. A simple method for generating daily rainfall data. *Agric. Forest Meteorol.*, **36**, 363–376.
- Gleick, P.H., 1989. Climate change, hydrology, and water resources. *Rev. Geophys.*, **27**, 329–344.
- Goodspeed, M. J. and Pierrehumbert, C. L., 1975. Synthetic input data time series for catchment model testing. In: *Prediction in Catchment Hydrology*, T.G.Chapman and F.X.Dunin (Eds.), Aust. Acad. of Sci., Canberra, 359–370.
- Green, J.R., 1964. A model for rainfall occurrence. *J. Roy. Statist. Soc.*, **B26**, 345–353.
- Guenni, L., Rose, C.W., Bradlock, R.D., Hogarth, W. and Charles-Edwards, D., 1990. Seasonal changes in inter-relationships between climatic variables. *Agric. Forest Meteorol.*, **53**, 45–58.
- Gupta, V.K. and Waymire, E.C., 1993. A statistical analysis of mesoscale rainfall as a random cascade. *J. Appl. Meteorol.*, **32**, 251–267.
- Hay, L.E., McCabe Jr., G.J., Wolock, D.M. and Ayers, M.A., 1991. Simulation of precipitation by weather type analysis. *Water Resour. Res.*, **27**, 493–501.
- Hipel, K.W., 1985. *Stochastic research in multivariate analysis*. Fourth International Hydrology Symposium, Fort Collins, Colorado, USA, 2–50.
- Hopkins, J.W. and Robillard, P., 1964. Some statistics of daily rainfall occurrences from the Canadian Prairie province. *J. Appl. Meteorol.*, **3**, 600–602.
- Hughes, J.P., Guttorp, P. and Charles, S.P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. *Appl. Stat.*, **48**, 15–30.
- Hurvich, C.M. and Tsai, C.-L., 1989. Regression and time series model selection in small samples. *Biomertika*, **76**, 297–307.
- Hutchinson, M.F., 1987. Methods of generation of weather sequences. In: *Agricultural Environments, Characterisation, Classification and Mapping*, A.H. Bunting (Ed.), CAB International, Wallingford, UK, 149–157.
- Hutchinson, M.F., 1995. Stochastic space time weather models from ground based data. *Agric. Forest Meteorol.*, **73**, 237–264.
- Hutchinson, M.F., Richardson, C.W. and Dykes, P.T., 1993. *Normalisation of rainfall across different time steps*. Management of Irrigation and Drainage Systems, 21–23 July 1993, Park City, UT. Irrigation and Drainage Division, ASCE, US Department of Agriculture, 432–439.
- Jimoh, O.D. and Webster, P., 1996. Optimum order of Markov chain for daily rainfall in Nigeria. *J. Hydrol.*, **185**, 45–69.
- Jimoh, O.D. and Webster, P., 1999. Stochastic modelling daily rainfall in Nigeria, intra-annual variation of model parameters. *J. Hydrol.*, **222**, 1–17.
- Jones, J.W., Colwick, R.E. and Threadgill, E.D., 1972. A simulated environmental model of temperature, evaporation, rainfall and soil moisture. *Trans. ASAE*, **15**, 366–372.
- 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 Resour. Res.*, **36**, 267–284.
- Katz, R.W., 1981. On some criteria for estimating the order of a Markov chain. *Technometrics*, **23**, 243–249.
- Katz, R., 1996. Use of conditional stochastic models to generate climate change scenarios. *Climatic Change*, **32**, 237–255.
- Klemes, V. and Bulu, A., 1979. Limited confidence in confidence limits derived by operational stochastic hydrologic models. *J. Hydrol.*, **42**, 9–22.
- Lall, U. and Sharma, A., 1996. A nearest neighbour bootstrap for resampling hydrologic time series. *Water Resour. Res.*, **32**, 679–693.
- Lall, U., Rajagopalan, B. and Tarboton, D.G., 1996. A non-parametric wet/dry spell model for resampling daily precipitation. *Water Resour. Res.*, **32**, 2803–2823.
- Lane, W.L., 1979. *Applied stochastic techniques (LAST computer package), user manual*. Division of Planning Technical services, Bureau of Reclamation, Denver, Colorado, USA.
- Larsen, G.A. and Pense, R.B., 1981. *Stochastic simulation daily climate data*. SRS Staff Report AGE810831. Statistical Reporting Service, Washington DC, USDA.
- Lettenmaier, D.P. and Gan, T.Y., 1990. Hydrologic sensitivities of the Sacramento-San Joaquin River basin, California for global warming. *Water Resour. Res.*, **26**, 69–86.
- Loaiciga, H.A., Maidment, D.R. and Valdes, J.B., 2000. Climate-change impacts in a regional karst aquifer, Texas, USA. *J. Hydrol.*, **227**, 173–194.



- Lowry, W.P. and Guthrie, D., 1968. Markov chains of order greater than one. *Mon. Weath. Rev.*, **96**, 798–801.
- McLeod, A.I. and Hipel, K.W., 1978. Simulation procedure for Box-Jenkins models. *Water Resour. Res.*, **14**, 969–975.
- Maheepala, S. and Perera, C.J.C., 1996. Monthly hydrologic data data generation by disaggregation. *J. Hydrol.*, **178**, 277–291.
- Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. *Water Resour. Res.*, **3**, 937–945.
- Menabde, M. and Sivapalan, M., 2000. Modelling of rainfall time series and extremes using bounded random cascades and Levy-stable distributions. *Water Resour. Res.*, **36**, 3293–3300.
- Mejia, J.M. and Rouselle, J., 1976. Disaggregation models in hydrology revisited. *Water Resour. Res.*, **12**, 185–186.
- Mimikou, M.A., Baltas, E., Varanou, E. and Pantazis, K., 2000. Regional impacts of climate change on water resources quantity and quality indicators. *J. Hydrol.*, **234**, 95–109.
- Nicks, A.D. and Harp, J.F., 1980. Stochastic generation of temperature and solar radiation data. *J. Hydrol.*, **48**, 1–17.
- Nicks, A.D. and Lane, L.J., 1989. Weather generator. In: *USDA – Water Erosion Prediction Project*, L.J. Lane and M.A. Nearing, (Eds.), National Soil Erosion Research Lab., Report No. 2, West Lafayette.
- Over, T.M. and Gupta, V.K., 1994. Statistical analysis of mesoscale rainfall: Dependence of a random cascade generator on large-scale forcing. *J. Appl. Meteorol.*, **101**, 319–326, 331.
- Over, T.M. and Gupta, V.K., 1996. A space-time theory of mesoscale rainfall using random cascades. *J. Geophys. Res.*, **101**, 26319–26331.
- Panagoulia, D., 1992. Impacts of GISS-modelled climate changes on catchment hydrology. *Hydrol. Sci. J.*, **37**, 141–163.
- Pegram, G.G.S., 1980. An auto-regressive model for multi-lag Markov Chains. *J. Appl. Probab.*, **17**, 350–362.
- Pegram, G.G.S. and James, W., 1972. Multivariate autoregressive model for the generation of operational hydrology. *Water Resour. Res.*, **8**, 1074–1076.
- Pegram, G.G.S. and Seed, A.W., 1998. *The feasibility of stochastically modelling the spatial and temporal distribution of rainfields*. WRC Report No 550/1/98, Department of Civil Engineering, University of Natal, RSA. 94 pp.
- Porter, J.W. and Pink, B.J., 1991. *A method of synthetic fragments for disaggregation in stochastic data generation*. Hydrology and Water Resources Symposium, Institution of Engineers, Australia, 187–191.
- Racsko, P., Szeidl, L. and Semenov, M., 1991. A serial approach to local stochastic weather models. *Ecol. Model.*, **57**, 27–41.
- Rajagopalan, B. and Lall, U., 1999. A k-nearest-neighbour simulator for daily precipitation and other weather variables. *Water Resour. Res.*, **35**, 3089–3101.
- Rajagopalan, B., Lall, U. and Tarboton, D.G., 1996. A nonhomogeneous Markov model for daily precipitation simulation. *J. Hydrol. Eng.-ASCE*, **1**, 33–40.
- Rajagopalan, B., Lall, U., Tarboton, D.G. and Bowles, D.S., 1997a. Multivariate non-parametric resampling scheme for generation of daily weather variables. *Stoch. Hydrol. Hydraul.*, **11**, 65–93.
- Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resour. Res.*, **17**, 182–190.
- Richardson, C.W. and Wright, D.A., 1984. *WGEN, A model for generating daily weather variables*. ARS-8, US Department of Agriculture, 83pp.
- Roldan, J. and Woolhiser, D.A., 1982. Stochastic daily precipitation models, 1. A comparison of occurrence processes. *Water Resour. Res.*, **18**, 1451–1459.
- Salas, J.D., Delleur, J.W., Yevjevich, V. and Lane, W.L., 1980. *Applied modelling of hydrological time series*. Water Resources Publications, Littleton, Colorado, USA. 484pp.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.
- Selvalingam, S. and Miura, M., 1978. Stochastic modelling of monthly and daily rainfall sequences. *Water Resour. Bull.*, **14**, 1105–1120.
- Sharma, A. and Lall, U., 1997. *A nearest neighbour conditional bootstrap for resampling daily rainfall*. 24<sup>th</sup> Hydrology and Water Resources Symposium, The Institution of Engineers, Australia, 439–444.
- Sharma, A. and Lall, U., 1999. A nonparametric approach for daily rainfall simulation. *Math. Comput. Simulat.*, **48**, 361–371.
- Singh, S.V., Kripalani, R.H., Saha, P., Ismail, P.M.M. and Dahale, S.D., 1981. Persistence in daily and 5-day summer monsoon rainfall over India. *Arch. Met. Geoph. Biokl., Ser. A*, **30**, 261–277.
- Small, M.J. and Morgan, D.J., 1986. The relationship between a continuous-time renewal model and a discrete Markov chain model of precipitation occurrence. *Water Resour. Res.*, **22**, 1422–1430.
- Smith, J.A., 1987. Statistical model of daily rainfall occurrences. *Water Resour. Res.*, **23**, 885–893.
- Srikanthan, R., 1985. *Stochastic simulation of daily climatic data*. Fourth International Hydrology Symposium, Fort Collins, Colorado, USA, 542–554.
- Srikanthan, R. and McMahon, T.A., 1983a. Stochastic simulation of daily rainfall for Australian stations. *Trans. ASAE*, **26**, 754–759, 766.
- Srikanthan, R. and McMahon, T.A., 1983b. Stochastic simulation of evaporation data in Australia. *Nord. Hydrol.*, **14**, 207–228.
- Srikanthan, R. and McMahon, T.A., 1984. Synthesizing daily rainfall and evaporation data as input to water balance-crop growth models. *J. Aust. Inst. Agr. Sci.*, **50**, 51–54.
- Srikanthan, R. and McMahon, T.A., 1985. *Stochastic generation of rainfall and evaporation data*. AWRC Technical Paper No. 84, 301pp.
- Srikanthan, R. and Stewart, B.J., 1992. *Analysis of Australian rainfall and rainday data with respect to climatic variability and change*. HRS Report No. 2, Hydrology Branch, Bureau of Meteorology, Melbourne, 254pp.
- Stedinger, J.R., Pel, D. and Cohn, T.A., 1985. A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. *Water Resour. Res.*, **21**, 665–675.
- Stern, R.D., 1980a. Analysis of daily rainfall at Samaru, Nigeria, using a simple two-part model. *Arch. Met. Geoph. Biokl., Ser. B*, **28**, 123–135.
- Stern, R.D., 1980b. Computing probability distribution for the start of the rains from a Markov chain model for precipitation. *J. Appl. Meteorol.*, **21**, 420–423.
- Stern, R.D. and Coe, R., 1984. A model fitting analysis of daily rainfall data. *J. Roy. Statist. Soc. A*, **147**, Part 1, 1–34.
- Tarboton, D.G., Sharma, A. and Lall, U., 1998. Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resour. Res.*, **34**, 107–119.
- Thomas, H.A. and Fiering, M.P., 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. Chapter 12 in: *Design of water resources systems* A. Maass, S. Marglin and G. Fair (Eds.), Harvard University Press, Cambridge, Massachusetts, USA.
- Thompson, C.S., 1984. Homogeneity analysis of rainfall series: an application of the use of a realistic rainfall model. *J. Climat.*, **4**, 609–619.
- Thyer, M.A., 2000. *Modelling long-term persistence in hydrological time series*. Ph D Thesis, University of Newcastle, Australia.

- Thyer, M.A. and Kuczera, G., 1999. *Modelling long-term persistence in rainfall time series, Sydney rainfall case study*. Hydrology and Water Resources Symposium, Institution of Engineers, Australia, 550–555.
- Thyer, M.A. and Kuczera, G., 2000. Modelling long-term persistence in hydro-climatic time series using a hidden state Markov model. *Water Resour. Res.*, **36**, 3301–3310.
- Valdes, J.B., Rodriguez-Iturbe, I. and Vicens, G.J., 1977. Bayesian generation of synthetic streamflows 2. The multivariate case. *Water Resour. Res.*, **13**, 291–295.
- Valencia, D. and Schaake Jr, J.C., 1973. Disaggregation processes in stochastic hydrology. *Water Resour. Res.*, **9**, 580–585.
- Vicens, G.J., Rodriguez-Iturbe, I. and Schaake, J.C., 1975. Bayesian generation of synthetic streamflows. *Water Resour. Res.*, **11**, 827–838.
- Wang, Q.J. and Nathan, R.J., 2000. A daily and monthly mixed model algorithm for stochastic generation of rainfall time series. Unpublished manuscript.
- Warner, R.F., 1987. The impacts of alternating flood- and drought-dominated regimes on channel morphology at Penrith, New South Wales, Australia. *IAHS. Publication no. 168*, 327–338.
- Weiss, L.L., 1964. Sequences of wet and dry days described by a Markov chain model. *Mon. Weath. Rev.*, **92**, 169–176.
- Wilks, D.S., 1989. Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resour. Res.*, **23**, 1429–39.
- Wilks, D.S., 1992. Adopting stochastic weather generation algorithms for climate change studies. *Climatic Change*, **22**, 67–84.
- Wilks, D.S., 1998. Multisite generalisation of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**, 178–191.
- Williams, C.B., 1947. The log series and its applications to biological problems. *J. Ecol.*, **34**, 253–272.
- Wilson, L.L., Lettenmaier, D.P. and Skillingstad, E., 1992. A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. *J. Geophys. Res.*, **97**, 2791–2809.
- Woolhiser, D.A., 1992. Modelling daily precipitation – progress and problems. In: *Statistics in the environmental & earth sciences*, A. T. Walden and P. Guttorp, (Eds.), Edward Arnold, UK. 71–89.
- Woolhiser, D.A. and Pegram, G.G.S., 1979. Maximum likelihood estimation of fourier coefficients to describe seasonal variation of parameters in stochastic daily precipitation models. *J. Appl. Meteorol.*, **18**, 34–42.
- Woolhiser, D.A. and Roldan, J., 1982. Stochastic daily precipitation models, 2. A comparison of distribution of amounts. *Water Resour. Res.*, **18**, 1461–1468.
- Woolhiser, D.A. and Roldan, J., 1986. Seasonal and regional variability of parameters for stochastic daily precipitation models. *Water Resour. Res.*, **22**, 965–978.
- Woolhiser, D.A., Keefer, T.O. and Redmond, K.T., 1993. Southern oscillation effects on daily precipitation in the southwestern United States. *Water Resour. Res.*, **29**, 1287–1295.
- Yevjevich, V. and Dyer, T.G.J., 1983. Basic structure of daily precipitation series. *J. Hydrol.*, **64**, 49–67.
- Young, K.C., 1994. A multivariate chain model for simulating climatic parameters from daily data. *J. Appl. Meteorol.*, **33**, 661–671.
- Young, G.K. and Pisano, W.C., 1968. Operational hydrology using residuals. *J. Hydraul. Div.-ASCE*, **94**, 909–923.
- Zucchini, W. and Adamson, P.T., 1984. *The occurrence and severity of drought in South Africa*. Department of Civil Engineering, University of Stellenbosch and Department of Water Affairs, WRC Report No. 91/1/54, 198pp.
- Zucchini, W. and Guttorp, P., 1991. A hidden Markov model for space-time precipitation. *Water Resour. Res.*, **27**, 1917–1923.