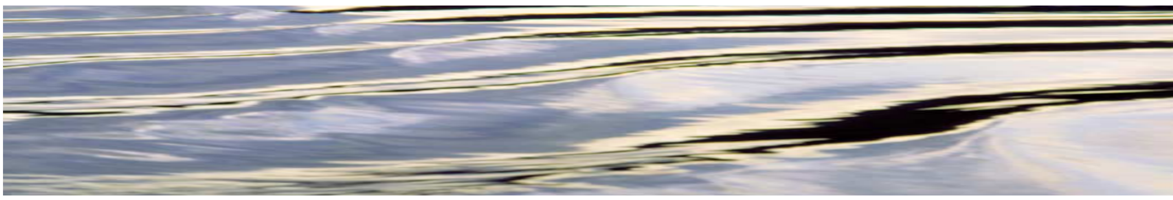
COOPERATIVE RESEARCH CENTRE FOR **CATCHMENT HYDROLOGY**

STOCHASTIC GENERATION OF CLIMATE DATA: A REVIEW

TECHNICAL REPORT
Report 00/16
December 2000

Ratnasingham Srikanthan / Tom McMahon



COOPERATIVE RESEARCH CENTRE FOR
CATCHMENT HYDROLOGY

Stochastic Generation of Climate Data: A Review

Ratnasingham Srikanthan and Tom McMahon

Cooperative Research Centre for Catchment Hydrology

December, 2000

Preface

This is the first of a series of reports expected from CRC for Catchment Hydrology Program 5, Climate Variability. This program aims to reduce management uncertainty by developing space-time models for Australia, improving the representativeness of surface hydrology in numerical weather prediction models, developing methods to forecast rainfall and streamflow several hours to several months ahead, and establishing a robust set of stochastic models for the generation of climate and streamflow.

This report deals with the latter aspect, and is part of CRCCH Project 5.2 (National Data Bank of Stochastic Climate and Streamflow Models). The recommendations outlined in the report will form the basis of the research program to be conducted in this project.

Russell Mein
Director

Acknowledgements

We wish to thank the Project 5.2 review panel (George Kuzcera, Russell Mein, Rory Nathan and Geoff Pegram) for their contribution to this review.

Summary

The purpose of the report is to review the state of research and practice in the stochastic generation of annual, monthly and daily climate data. This review forms part of the Cooperative Research Centre for Catchment Hydrology Project 5.2: National data bank of stochastic climate and streamflow models.

The generation of rainfall and other climate data needs a range of models depending on the time and spatial scales involved. There are three broad types of rainfall models, namely, empirical statistical models, models of dynamic meteorology and intermediate stochastic models. The models included in the review are empirical statistical models only.

Surprisingly, very little work has been done since 1985 in stochastic generation of annual and monthly rainfall data. Because the generation of annual and monthly rainfall and streamflow are similar, streamflow generation models have been included in this part of the review. Most of the models used in the past do not take into account the year to year variations in the model parameters. They were assumed to be constant from year to year 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 and this needs to be considered in the model structure. Recently, Thyer and Kuczera, University of Newcastle developed a hidden state Markov model to account for the long term persistence in annual rainfall. The review looked at the traditional time series models first and then the more complex models, which take account of long term persistence in the data.

Monthly rainfall data have been successfully generated by using the method of fragments. The main criticism of this approach is the repetition of the same yearly pattern when there is only a limited number of years of historical data. This deficiency was overcome by using synthetic fragments but this brings in an additional problem of generating the right number of months of zero rainfall. The disaggregation schemes are an effective way to obtain monthly data but the main problem is the large number of parameters to estimate when dealing with a large number of sites. Several simplifications are proposed to overcome this problem. Here again, one needs to take into account year to year variation in model parameters and this has not been done before.

Models for generating daily rainfall are well developed and a great deal of progress has been made recently in developing techniques for parameter estimation. 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. 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 in other parts of the world 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 constrained a two part daily model within a monthly model and it appears to perform well. Also, Boughton has adjusted the generated daily rainfalls by a trial and error procedure to match the variance of the observed annual rainfall. As an alternative, conditioning model parameters on monthly amounts or perturbing the model parameters with the SOI result in better agreement between the variance of the simulated and observed annual rainfall and these approaches should be investigated further.

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. As climate data are less variable than rainfall, but are correlated among themselves and with rainfall, mutisite models have been used successfully to generate annual data. The monthly climate data can be obtained by disaggregating the generated annual data. On a daily time step at a site, climate data has been generated by 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 successfully modelled by truncated power normal distribution then the model can be easily extended to generate daily climate data at several sites simultaneously.

Concerns over climate change caused by increasing concentration of CO_2 and other trace gases in the atmosphere has increased in recent years. A major effect of climate change may be alterations in regional hydrologic cycles and changes in regional water availability. The main source of climate change projections is the 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 providing regional scale information that are necessary for hydrological studies. As a result, the 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.

Most of the early work on the impacts of climate change used historical data adjusted for the climate change. In recent studies, the stochastic daily weather generation models are adapted for generation of synthetic daily time series consistent with assumed future climates. The assumed climates were specified by the monthly means and variances of rainfall and temperature. The greatest uncertainty in modelling climate data under climate change conditions is the uncertainty in the future climate predictions. The GCMs at present are able to provide either scenarios or projections of the future climate. If the future climate conditions are known with sufficient accuracy, the stochastic climate models available at present can be adapted to generate climate for the new conditions.

Since the rainfall and climate data are much less variable and less correlated than streamflow, the existing models can be used to generate these at annual and monthly level for single and multisites. As these models do not take into account of long term persistence, hidden state Markov and other models need to be investigated. Regarding daily rainfall, the transition probability matrix method performs well, but is not suitable for regionalisation and with limited length of data. Wang and Nathan approach appears to be promising. The review makes recommendations on models to be adopted and models that should be further tested. The recommended models can be used to generate climate data under climate change conditions by adjusting the parameters appropriately.

Notations/Abbreviations

A	Coefficient matrix
B	Coefficient matrix
B	Beta model generator
C	Coefficient matrix
C_{ik}	Amplitude of the k^{th} harmonic
d	Threshold for defining a wet or dry day
e	Error term
F	Adjustment factor
f(u)	Probability density function of the random variable u
F(u)	Probability distribution function of the random variable u
G	Systemic multiplier to include spatial variation in rainfall
$G_i(t)$	Value of the i^{th} parameter on day t
G_{i0}	Mean value of the i^{th} parameter
GCM	General circulation model
$g_{ij}(t)$	Logit transform of probability pij(t) for day t
HSM	Hidden state Markov
I_i	Mean intensity of rainfall for wet days for weather type i
L_o	Outer length scale
M	Observed mean annual rainfall
$\mathbf{M_0}$	Lag zero cross correlation matrix
$\mathbf{M_1}$	Lag zero cross correlation matrix
MAR	Multivariate autoeregressive
MCi	Markov chain of order i
m	Number of sites
m_i	Maximum number of harmonics
NHMM	Nonhomogeneous hidden state Markov model
n_{ij}	Number of transitions from state i to state j
P	Transition probability matrix used in HSM model
P(W|D)	Probability of transition from state dry to wet
P(W|W)	Probability of transition from state dry to wet
P(W)	Probability of wet day
P_{ij}	Probability of transition from state i to state j (elements of **P**)
p	number of climatic variables
$p_{ij}(t)$	Probability of transition from state i to state j on day t
R	Rainfall amount for a wet day
R_o	Average rainfall intensity at the outer length scale
r	Lag one autocorrelation coefficient
S_t	State of a year t
s_x	Standard deviation of annual rainfall
s_y	Standard deviation of monthly rainfall for the last month of a year
$\mathbf{S_{XY}}$	Matrix of the cross product of X and Y
T_i	Generated annual rainfall for year i
U	A uniform random variable between 0 and 1
$\mathbf{u_t}$	Vector of uniformly distributed random variates (0,1)
$\mathbf{v_t}$	Vector of uniformly distributed random variates (0,1)
W	cascade generators
$\mathbf{X_t}$	A vector of standardised annual rainfall or climate data
X_t	Annual rainfall or climate data in year t having zero mean and unit variance
x_t	Annual rainfall for year t
X(t)	State of day t (wet or dry)
Y	Monthly rainfall or climate data for the present year
Y	Log normal variable
y_{k-1}	Disaggregated monthly rainfall for the last month of the year k-1
$y_{i,\tau}$	Monthly rainfall for year i and month τ
Z	Monthly rainfall or climate data for the previous year
$\alpha(t)$	Shape parameter of gamma distribution for day t
α	Mixing fraction for the mixed exponential distribution
α_i	A coefficient
$\beta(t)$	Scale parameter of gamma distribution for day t
β	Parameter of the exponential distribution
β_i	A coefficient
δ	Parameter of the exponential distribution
ε_t	Random number with zero mean and unit variance
$\boldsymbol{\varepsilon_t}$	Matrix of random numbers with zero mean and unit variance
Ω	Variance - covariance matrix
Φ	Normal probability distribution function
ϕ_{ik}	Phase angle of the k^{th} harmonic for the i^{th} parameter
ϕ	Normal cumulative distribution function
γ	Skewness of e
γ_ε	Skewness of annual data
η	Normally distributed random number with zero mean and unit variance
κ	Shape parameter of the gamma distribution
μ_{dry}	Mean of dry year rainfalls
μ_{wet}	Mean of wet year rainfalls
π_1	Probability of being wet in year 1
$\theta^{(i)}_{ks}$	Coefficients in the rainfall regression equation
σ_{dry}	Standard deviation of dry year rainfalls
σ_{wet}	Standard deviation of wet year rainfalls
τ	Season or month
$\omega(k,l)$	Correlation between the Gaussian random variates at sites k and l, which forces the occurrence process
$\xi(k,l)$	Correlation between the binary occurrence variates at sites k and l
ψ	Matrix of random numbers
$\zeta(k,l)$	Correlation between the rainfall amounts at sites k and l
$\Gamma(\,)$	Gamma function

1 Introduction

1.1 Purpose of the report

The purpose of this report is to review the present state of research and practice in the stochastic generation of climate data. This review forms part of the Cooperative Research Centre for Catchment Hydrology Project 5.2: National data bank of stochastic climate and streamflow models. The stochastic generation of streamflow data is not reviewed in this report and will be carried out later.

A one-day workshop of stakeholders was held at the Bureau of Meteorology, Melbourne on 19 March 2000 to come up with a list of climate variables and the time intervals. The participants agreed on the following list of climate variables to be included in the project:

- **Single site (point)**
  1. daily rainfall (monthly and annual)
  2. daily mean temperature
  3. daily global radiation
  4. daily wind run
  5. daily vapour pressure deficit
  6. short duration rainfall
- **Multisite (catchment)**
  1. monthly rainfall
  2. monthly number of raindays
  3. monthly potential evapotranspiration

The objectives of this review are:

- to survey the literature and identify the latest developments in stochastic data generation of the above listed climate variables
- to recommend models which perform adequately and identify areas for further model development

1.2 Background

One of the major gaps identified by industry and researchers is the need to quantify the uncertainty in hydrologic systems as a result of climatic variability. This need 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 the systems one has to resort to system simulation using stochastically generated data. In addition to quantifying the uncertainty, stochastically generated data has many applications such as the design and operation of water resources systems, design of urban drainage systems and land management changes.

What is stochastic data? Stochastic data are random numbers that are modified so that they have the same characteristics (in terms of mean, variance, etc and auto-correlation structure) as the data set on which they are based. For example, in 1954 Frank Barnes of the then Melbourne Metropolitan Board of Works generated 1000 years of stochastic (or synthetic) annual streamflows for the Upper Yarra Dam investigation (Barnes, 1954). He did this by assuming the annual flows were independent and normally distributed and using a table of random numbers was able to generate the long time series. This was the first occasion in Australia in which stochastic data were used in hydrologic investigation.

Even though there are a number of stochastic models available in the literature, most of them have not been adequately tested 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 the monthly and annual characteristics in addition to preserving the daily characteristics.

In the past, all the data generation models generally assumed no variation in the parameters between years. Only the seasonal or monthly variations within a year have been taken into account and the same set of parameter values were used for all the years. There is a growing awareness of long term persistence in the climatic data in the form of wet and dry years or ENSO cycles and to take this information into account, the parameters of the models should be varied in some way to model the long term persistence. There is very little research work done on this and this project plans to cover this aspect as well. Another important aspect which did not receive much attention in the past is the quantification of the uncertainty in model parameters. Recent developments in the Bayesian analysis such as Markov Chain Monte Carlo method are being used with the stochastic data generation models (Thyer and Kuczera, 1999; Wang 2000) and this project will attempt to quantify the uncertainty in the developed model parameters.

The overall goal of the project is to identify and develop a robust set of stochastic models for the generation of climate and streamflow data anywhere in Australia at different time scales and to provide parameter values with known levels of uncertainty for the developed models. This report covers only the generation of rainfall and other climate data up to daily time interval. The generation of short duration rainfall data is not included in this review but will be dealt with separately. Since there is a growing concern on the impacts of climate change on water resources, the generation of climate data under changed climate conditions are briefly reviewed.

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. The idea behind this classification is the amount of physical realism incorporated into the model structure. In the empirical statistical models, empirical stochastic models are fitted to the available data. 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 and not for data generation. In intermediate stochastic models, a modest number of 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 interval such as hourly. The models reviewed in this report are of the first type only.

1.3 Layout of the report

Because of the different type of models used for different time scales, the review is carried out for different time scales separately. Chapters 2 and 3 cover the generation of rainfall data while chapter 4 covers the generation of climate data. In each chapter the models for the generation of data at a single site are reviewed first followed by models for a number of sites or over an area. Chapter 2 deals with the generation of annual and monthly rainfall data at single and multi-sites. The models for the generation daily rainfall are reviewed in chapter 3. The generation of annual, monthly and daily climate data is reviewed in chapter 4. Chapter 5 briefly reviews the generation of climate data under climate change conditions. The conclusions from the review along with the recommendations are given in chapter 6.

2 Annual and Monthly Rainfall Data

Surprisingly, very little work has been done since 1985 in stochastic generation of annual and monthly rainfall data. However, the generation of annual and monthly streamflows is relevant for modelling large catchments and is therefore included in the review. Most of the models used in the past do not take into account the year to year variations in the model parameters. They were assumed to be constant from year to year and only the within-year seasonal variations in parameters were taken into account. Long period of wet and dry years were observed in the past (Warner, 1987; Srikanthan and Stewart, 1992) and this need to be considered in the model structure. Recently, Thyer and Kuczera (1999, 2000) developed a hidden state Markov model to account the long term persistence in annual rainfall. The following review looks at the traditional time series models first and then the more complex models, which take into account of long term persistence in the data.

2.1 Annual rainfall data at a site

Srikanthan and McMahon (1985) recommended a first order Markov model incorporating the Wilson-Hilferty transformation to generate annual rainfall data.

$$X_t = r\,X_{t-1} + (1-r^2)^{1/2}\varepsilon_t \tag{2.1}$$

$$\varepsilon_t = \frac{2}{\gamma_\varepsilon}\left\{\left(1+\frac{\gamma_\varepsilon\eta_t}{6}-\frac{\gamma_\varepsilon^2}{36}\right)^3-1\right\} \tag{2.2}$$

where X_t = standardised rainfall in year t having zero mean and unit variance

r = lag one autocorrelation coefficient

η_t = normally distributed random number with zero mean and unit variance

ε_t = random number with zero mean, unit variance and coefficient of skewness γ_ε which is related to the skewness of the annual data, γ, through

$$\gamma_\varepsilon = \frac{(1-r^3)}{(1-r^2)^{3/2}}\gamma \tag{2.3}$$

This model degenerates into a white noise model when the coefficient of skewness and the lag one autocorrelation coefficient are close to zero.

The annual rainfall amount is then obtained from

$$x_t = \bar{x} + sX_t \tag{2.4}$$

where $\bar{x}$ and s are respectively the mean and standard deviation of the annual rainfall x_t.

Thyer and Kuczera (1999) developed a hidden state Markov (HSM) model with Bayesian inference technique to generate annual rainfall data for Sydney. 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 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 quasi-periodic phenomenon such as El Nino.

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:

$$S_t|S_{t-1} \sim \text{Markov}\ (P,\pi_1) \tag{2.5}$$

where **P** is the transition probability matrix whose elements p_{ij} are defined by:

$$p_{ij} = \Pr(S_t = i|S_{t-1} = j) \quad i, j = \text{wet or dry} \tag{2.6}$$

and π_1 is the probability distribution of the wet and dry states at year 1.

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

$$x_t = \begin{cases} N(\mu_{wet}, \sigma^2_{wet}) & \text{if } S_t = \text{wet} \\ N(\mu_{dry}, \sigma^2_{dry}) & \text{if } S_t = \text{dry} \end{cases} \qquad (2.7)$$

where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^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 drought risks. The Sydney rainfall data strongly supported the assumption of a two state climate model with the average residence time similar to the quasi-periodicity of the ENSO phenomenon.

2.2 Annual rainfall data at a number of sites

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

$$\mathbf{X}_t = \mathbf{A}\,\mathbf{X}_{t-1} + \mathbf{B}\,\varepsilon_t \qquad (2.8)$$

where $\mathbf{X}_t$ = (m x 1) vector of standardised annual rainfall at m sites

ε_t = (m x 1) vector of random deviates with zero mean and unit variance

A, B = (m x m) matrices of constant coefficients to preserve the cross correlations.

The matrices **A** and **B** are obtained from:

$$\mathbf{A} = \mathbf{M}_1\,\mathbf{M}_0^{-1} \qquad (2.9)$$

$$\mathbf{B}\mathbf{B}^T = \mathbf{M}_0 - \mathbf{M}_1\,\mathbf{M}_0^{-1}\,\mathbf{M}_1^T \qquad (2.10)$$

where $\mathbf{M}_0$ and $\mathbf{M}_1$ are respectively the lag zero and lag one cross correlation matrices.

If the annual rainfall data is skewed, then they can be normalised using a 3 parameter log normal transformation. The parameters in the log domain are obtained by using the Matalas moment transformation equations (Matalas, 1967).

If the annual rainfall data are not serially correlated, then **A** = 0 and the model becomes

$$\mathbf{X}_t = \mathbf{B}\,\varepsilon_t \qquad (2.11)$$

Pegram and James (1972) and Hipel (1985) claim that the lag one cross correlations are not that important and could be ignored. This simplifies the correlation matrices and the solutions to **A** and **B**. Thyer and Kuczera (1999) reported that work was in progress in extending the HSM model for multi-site and we are not aware of any outcome.

2.3 Monthly rainfall data at a site

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 generated annual rainfall 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.

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. The modified model is described in the following steps.

Step 1 Generate a monthly rainfall series using a suitable monthly data generation model such as the Thomas-Fiering model.

Step 2 Generate an annual rainfall series using a suitable annual data generation model.

Step 3 Disaggregate the generated annual rainfall in Step 2 using the monthly fragments from Step 1. The appropriate monthly fragments for a given year, k, is selected by considering the closeness of the generated annual rainfall data and the monthly rainfall for the last month of the previous year of the already disaggregated data and the generated monthly data. This is achieved by selecting the monthly fragments of a year, i, in the generated monthly series that produces a minimum value for ($\alpha_i + \beta_i$) where α_i and β_i are defined below:

$$\alpha_i = \left(\frac{x_k - x_i}{s_x} \right)^2 \tag{2.12}$$

$$\beta_i = \left(\frac{y_{k-1} - y_{i-1}}{s_y} \right)^2 \tag{2.13}$$

where x_k = generated annual rainfall for year k
s_x = standard deviation of the annual rainfall
y_{k-1} = disaggregated monthly rainfall for the last month for year k-1
s_y = standard deviation of the monthly rainfall for the last month of the year

They have compared this modified procedure with the above two methods of fragments, which use historical and synthetic fragments using streamflow data from 5 rivers in Victoria. The results showed that the modified model was able to preserve 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 is less variable and has smaller skewness than streamflow data, the extended disaggregation scheme proposed by Mejia and Rousselle (1976) can be used to disaggregate the generated annual rainfall (X) into monthly rainfall (Y).

$$\mathbf{Y} = \mathbf{A}\mathrm{X} + \mathbf{B}\varepsilon + \mathbf{CZ} \tag{2.14}$$

where
Z = a column matrix containing as many monthly values from the previous year as are desired
A, B, C = coefficient matrices

Lane (1979) has developed an approach which essentially sets to zero several parameters of the model which are not important. The model considers one month (τ) at a time and the model equation is written as:

$$y_{t,\tau} = a_\tau x_t + b_\tau \varepsilon + c_\tau y_{t,\tau-1} \tag{2.15}$$

The seasonal values are then adjusted to match the annual values.

2.4 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 et al 1985), the method of synthetic fragments (Porter and Pink, 1991) or the modified method of synthetic fragments (Maheepala and Perera 1996). In the last method, α_i and β_i are defined as:

$$\alpha_i = \sum_{j=1}^{N} \left(\frac{x_k^j - x_i^j}{s_x^j} \right)^2 \tag{2.16}$$

$$\beta_i = \sum_{j=1}^{N} \left(\frac{y_{k-1}^j - y_{i-1}^j}{s_y^j} \right)^2 \tag{2.17}$$

where j refers to the site and N is the number of sites.

The extended model developed by Mejia and Rousselle (1976) can be used to disaggregate the generated annual flows to monthly flows.

$$\mathbf{Y} = \mathbf{AX} + \mathbf{B}\varepsilon + \mathbf{CZ} \tag{2.18}$$

where

Z = a column matrix containing as many monthly values from the previous year as are desired

A, B, C = coefficient matrices

The coefficient matrices are estimated from:

$$\mathbf{A} = (\mathbf{S}_{yx} - \mathbf{S}_{yz}\mathbf{S}_{zz}^{-1}\mathbf{S}_{zx})(\mathbf{S}_{xz}\mathbf{S}_{zz}^{-1}\mathbf{S}_{zx})^{-1} \tag{2.19}$$

$$\mathbf{B} = (\mathbf{S}_{yz} - \mathbf{AS}_{xz})\mathbf{S}_{zz}^{-1} \tag{2.20}$$

$$\mathbf{BB}^T = \mathbf{S}_{yy} - \mathbf{AS}_{xx}\mathbf{S}_{zz}^{-1}\mathbf{A}^T - \mathbf{AS}_{xz}\mathbf{C}^T - \mathbf{CS}_{zx}\mathbf{A}^T - \mathbf{CS}zz\mathbf{C}^T \tag{2.21}$$

or equivalently,

$$\mathbf{BB}^T = \mathbf{S}_{yy}\mathbf{AS}_{xy} - \mathbf{CS}_{zy} \tag{2.22}$$

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. The model equation for month τ is written as:

$$\mathbf{Y}_{t,\tau} = \mathbf{A}_\tau X_t + \mathbf{B}_\tau\varepsilon + \mathbf{C}_\tau\mathbf{Y}_{t,\tau\text{-}1} \tag{2.23}$$

The coefficient matrices are estimated from

$$\mathbf{A}_\tau = [\mathbf{S}_{yx}(\tau,\tau) - \mathbf{S}_{yy}(\tau,\tau\text{-}1)\mathbf{S}_{yy}^{-1}(\tau-1,\tau-1)\mathbf{S}_{yx}(\tau-1,\tau)]$$
$$[\mathbf{S}_{xx}(\tau,\tau) - \mathbf{S}_{xy}(\tau,\tau-1)\mathbf{S}_{yy}^{-1}(\tau-1,\tau-1)\mathbf{S}_{yx}(\tau-1,\tau)]^{-1} \tag{2.24}$$

$$Ct = [\mathbf{S}_{yy}(\tau,\tau-1) - \mathbf{A}_\tau\mathbf{S}_{xy}(\tau,\tau-1)]\mathbf{S}_{yy}^{-1}(\tau-1,\tau-1) \tag{2.25}$$

$$\mathbf{B}_\tau\mathbf{B}_\tau^T = \mathbf{S}_{yy}(\tau,\tau) - \mathbf{A}_\tau\mathbf{S}_{xy}(\tau,\tau) - \mathbf{C}_\tau\mathbf{S}_{yy}(\tau-1,\tau) \tag{2.26}$$

One of the main drawbacks with the disaggregation approach is the large number of parameters that need 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), $168\ N^2$ for the extended (Mejia and Rousselle, 1976) and $36N^2$ for the condensed (Lane 1979) schemes. Salas et al (1980) gives a parsimony guide for disaggregation modelling.

Table 1. Parsimony guide for disaggregation modelling (Salas et al, 1980)

Ratio (R) of observations to parameters	Comment
R < 1	Impossible
1 ≤ R < 3	Foolish
3 ≤ R < 5	Poor
5 ≤ R < 10	Fair
10 ≤ R < 20	Good
20 ≤ R	Very good

2.5 Summary

If we ignore the year to year variations or long term persistence, 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 multi sites. The estimation of parameters under a Bayesian frame work needs to be investigated to quantify the parameter uncertainties. In the case of monthly rainfall, the procedures of Porter and Pink (1991) and Maheepala and Perera (1996) appear to be 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 many, the condensed version of Lane (1979) should be investigated. For the months with a high coefficient of variation (> 1), appropriate transformation will be used to eliminate the generation of negative values. If we are not using the method of fragments, 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).

3 Daily Rainfall Data

Long sequences of daily rainfall are increasingly required, 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 daily time scale and this forms the basis for monthly and annual rainfall. Because of this basic nature, modelling of daily rainfall process has attracted a lot interest in the past. The generation of daily rainfall at a site is reviewed first, followed by a review of the generation of daily rainfall at a number of sites.

3.1 Daily rainfall data at a site

The daily rainfall data generation models can be broadly classified into four groups, namely, two-part models, transition probability matrix models, resampling models and time series models.

3.1.1 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 been used to deal with seasonality: assume that parameters vary as a step function for each month, season or as a periodic function such as Fourier series to provide daily 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 (Gabriel and Newman, 1962; Caskey, 1963; Weiss, 1964; Hopkins and Robillard, 1964; Feyerherm and Bark, 1965, 1967; Lowry and Guthrie, 1968; Selvalingam and Miura, 1978; Stern, 1980; 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) 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. Katz (1981) derived the asymptotic distribution of the Akaike's information criterion (AIC) estimator and showed that the estimator is inconsistent. The Bayesian information criterion (BIC) proposed by Schwarz (1978) was shown to be consistent and asymptotically optimal.

Jimoh and Webster (1996) determined the optimum order of a Markov chain model for daily rainfall occurrences at 5 locations in Nigeria using AIC and BIC. The AIC consistently gave 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 differences 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 7 sites in Nigeria. The found that there was a systematic variation in P_{01} (probability of a wet day following a dry day) as one moves northwards and limited regional variation in P_{11}.

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

Alternating renewal process

The alternating renewal process consists of alternating wet and dry spells which are assumed independent. The distributions may be different between 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 either assumed to be 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 is shown to provide a good representation in certain parts of the United States while in other areas, where the Markov model in 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 (1978), and later by Chang et al (1984) and Delleur et al (1989). Buishand found that an alternating renewal process was superior to the DARMA model for the data from 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 and found that the first order autoregressive or the second order moving average model were appropriate for different seasons. Buishand (1977) pointed out an important factor that 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 through out 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 MC1 or TGD model was successful for higher rainfall stations (Perth, Adelaide, Lerderderg, Melbourne and Cowra) in southern Australia. The monthly MC2 model was the best fit for Alice Springs PO, Darwin, Broome, Onslow and Bamboo Springs (with 20 year record). For the remaining stations (Alice Springs Airport, Brisbane, Kalgoorlie, Mackay and Monto), the best fit was obtained with the monthly TNBD model. Different record lengths appear to have an effect on the selection of the best model, particularly when wet and dry spells are considered separately. For combined results, different models would be selected for the 20 year and 50 year records in 4 out of 10 cases, for the 20 and 100 year records in 2 out of 5 cases, and for the 50 and 100 year records in 1 case out of 5. He concludes that the prospects for regionalisation of parameters are poor unless there is a good sample of long records.

Rainfall amount models

Models used for rainfall amounts include the two parameter gamma distribution (Jones et al 1972; Goodspeed and Pierrehumbert, 1973; 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 (1978) 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.

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 contiguous United States. It was 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.

Chapman (1994) 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 is generated month by month using the normal two part model using 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 and found that 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 the skewness of the annual maximum daily rainfall.

3.1.2 Transition probability matrix models

Allan and Haan (1975) used a multi-state (7 x 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 number of 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 the months. Selvalingam and Miura (1978) modified the above procedure by having twelve parameters for the exponential distribution, but they obtained the parameter empirically by trial and error until the model adequately reproduced the daily maximum monthly rainfalls. Srikanthan and McMahon (1985) used a linear distribution for the intermediate classes and Box-Cox transformation for the last state. The number of states in each month was varied to obtain adequate number of data items in the last state.

Chapman (1994) compared the Srikanthan and McMahon model with the best selected rainfall occurrence and amount models and found that the latter performed better than the former in 5 out of 15 twenty year records, for 2 of the 10 fifty 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 TPM model 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.

$$\text{Ratio}_i = \{M + (T_i - M)F\}/T_i \qquad (3.1)$$

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

In this model, the log-Boughton distribution was used in place of the Box-Cox trabsformation for the largest state in the TPM. The parameter of the log-Boughton distribution was estimated by a trial and error procedure.

3.1.3 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 and the performance of the model was evaluated using a number of performance measures. The model reproduced the wet day preciptiation, wet spell length and dry spell length well.

Rajagopalan et al (1966) 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 well reproduced at the seasonal and annual time scales.

Sharma and Lall (1997) used a nearest neighbour conditional bootstrap for resampling daily rainfall for Sydney. The dry spells were conditioned on the number of days in the previous wet spell and the wet spells 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 numbr of days from the start of the current spell. Results from the model showed the ability of the model to simulate sequences that are representative of the historical record.

A limitation of the non-parametric density estimation approach used here 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 pdf of interest are likely to be larger than for parametric estimation. On the question of regionalisation and portability of the method, the non-parametric approach clearly enjoys the broader applicability than its parametric competitors. On the other hand, it may be less amenable to direct regionalisation as is done in terms of the parameters of a parametric model.

3.1.4 Time series models

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 is a large number of 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 could to a first approximation be ignored (Hutchinson, 1995). Such systematic differences between correlations based on occurrence and intensity have not been recognised in the existing applications such as Bardossy and Plate (1992).

3.1.5 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, 1988; 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 (SO). 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.

Let

$$X(t) = \begin{cases} 0 & \text{if day t has rain} < d,\ t = t_1, t_2, \ldots, t_T \\ 1 & \text{if day t has rain over a threshold, d} \end{cases} \tag{3.2}$$

$$p_{ij}(t) = P[X(t)=j|X(t-1)=i] \qquad i, j = 0, 1 \tag{3.3}$$

Let Y(t) be the rainfall on a wet day t and the random variable U(t) = Y(t) – d be distributed as a mixed exponential distribution.

$$f_t(u) = \alpha(t)/\beta(t)\exp\{-u/\beta(t)\} + \{1 - \alpha(t)\}/\delta(t)\exp\{-u/\delta(t)\} \tag{3.4}$$

where d is a threshold, $0 < \alpha(t) < 1$, $0 < \beta(t) < \delta(t)$

The mean $\mu(t)$ is given by

$$\mu(t) = \alpha(t)\,\beta(t) + \{1 - \alpha(t)\}\,\delta(t) \tag{3.5}$$

To account for the seasonal variability of the parameters of the model, the parameters $p_{00}(t)$, $p_{10}(t)$, $\alpha(t)$, $\beta(t)$ and $\mu(t)$ are written in the polar form of a finite Fourier series:

$$G_i(t) = G_{i0} + \sum_{k=1}^{m_i} [C_{ik} \sin(2\pi tk/365 + \phi_{ik}] \tag{3.6}$$

where i = 1, 2, .., 5
$G_i(t)$ = the value of the i^{th} parameter on day t
m_i = the maximum number of harmonics
G_{i0} = the mean value of the i^{th} parameter
C_{ik} = the amplitude of the k^{th} harmonic
φ_{ik} = the phase angle of the k^{th} harmonic for the i^{th} parameter

To avoid imposing constraints, the logit transform of the transition probabilities are fitted with Fourier series (Stern and Coe, 1984):

$$g_{ij}(t) = \log\{p_{ij}(t)/[1-p_{ij}(t)]\} \quad -\infty < g_{ij}(t) < +\infty \tag{3.7}$$

The periodic parameters are perturbed by a lagged linear function of SOI.

$$G'_i(t) = G_i(t) + b_i S(t-\tau_i) \tag{3.8}$$

where b_i and τ_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, $\mu(t)$, 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 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 most strongly affected. 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, ID. 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 overrunning and warm frontal overrunning. Transitions from one weather type to another weather type was 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 as an exponential distribution alone underestimated the variance of the daily rainfall. The rainfall amounts were modelled as:

$$R = [I_i(-\log(U)](1 + e) \qquad (3.9)$$

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

e = the error term, a uniform random variable between –1 and 1

When there were less than 10 days of recorded rainfall for a given weather type and month, the rainfall amounts were modelled using:

$$R = I_i(1 + e) \qquad (3.10)$$

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

Wilks (1989) has 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 using data from 10 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.

3.1.6 Uncertainty in model parameters

Chaouche and Parent (1999) used a Bayesian framework to estimate the uncertainties of the parameters of a daily rainfall model utilising a Markov chain and gamma distribution. They used 69 years of data from Dedougou, a station in the Sudan-Sahel zone of Burkina-Faso, but modelled only the rainy season (day 100 to day 300).

A Fourier series with 2 harmonics was used to model the logits of the Markov chain probabilities and this resulted in 10 parameters. The log likelihood of the transition probabilities is given by:

$$L\{p_{01}, p_{11} \mid X(t)\} =$$

$$\sum_{t=t_1+1}^{t_2} n_{01}(t)\log p_{01}(t) + \sum_{t=t_1}^{t_2-1} n_{00}(t)\log[1-p_{01}(t)]$$

$$+ \sum_{t=t_1+1}^{t_2} n_{11}(t)\log p_{11}(t) + \sum_{t=t_1}^{t_2-1} n_{10}(t)\log[1-p_{11}(t)] \qquad (3.11)$$

where $n_{ij}(t)$ is the number of transitions between days t-1 and t from the state i to the state j.

The shape parameter (κ) of the gamma distribution was assumed constant throughout the season and only the scale parameter (β_t) was allowed to vary. A Fourier series with 2 harmonics was fitted to the scale parameter resulting in five parameters for generating rainfall amounts.

$$\log(\beta_t) = \sum_{j=0}^{2} \lambda_j \cos\left(\frac{2\pi}{365} jt\right) + \eta_j \sin\left(\frac{2\pi}{365} jt\right) \qquad (3.12)$$

The log likelihood for the mean depth of rainfall is given by:

$$L(\lambda, \eta, \kappa \mid y_t) = \sum_{t=t_1}^{t_2} \sum_{y_{st}>0} \log f(y_{st} \mid \beta, \kappa)$$

$$= \sum_{t=t_1}^{t_2} \sum_{y_{st}>0} \left[-\log\Gamma(\kappa) + \kappa\log\kappa - \kappa\log\mu_t \right. \qquad (3.13)$$

$$\left. + (\kappa-1)\log y_{st} - \frac{\kappa y_{st}}{\mu_t} \right]$$

The 16 (10+1+5) model parameters were estimated within a Bayesian framework using the Markov chain Monte Carlo simulation with the Metropolis-Hastings algorithm (Gelman et al 1995).

3.1.7 Regionalisation of daily rainfall model parameters

Richardson and Wright (1984) have computed the monthly transitional probabilities for a two state Markov chain based on 20 years of data for each of the 31 locations in the United States. Geng et al (1986) developed empirical equations for the parameters of a two-part model using the data from the Netherlands (Wangeningen), the Phillipines (Los Banos) and the United States (Colombia, Boise, Miami, Phoenix and Boston).

$$P(W|D) = 0.75\ P(W) \qquad (3.14)$$

$$P(W|W) = 0.25 + 0.75\ P(W) \qquad (3.15)$$

$$\beta = -2.16 + 1.83\ \mu_w \qquad (3.16)$$

$$\alpha = \mu_w / \beta \qquad (3.17)$$

They proposed that these empirical equations allow rainfall simulation models to be used for crop growth studies in many areas where too little 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 carefully screened to ensure consistency of data. The model parameters for four test stations were more closely estimated 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 interpolated parameters for the four test sites were significantly different from parameters estimated from rainfall records.

3.1.8 Summary

Models for generating daily rainfall are well developed and a great deal of progress has been made recently in developing techniques for parameter estimation. 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). 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 in other parts of the world 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 by a trial and error procedure to match the variance of the observed annual rainfall. As an alternative, conditioning model parameters on monthly amounts or perturbing the model parameters with the SOI result in better agreement between the variance of the simulated and observed annual rainfall and these approaches should be investigated.

3.2 Daily rainfall data at a number of sites

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

3.2.1 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, US.

Bardossy and Plate (1991) developed a semi-Markov chain model for atmospheric circulation patterns and linked to the occurrence of rainfall using transition probabilities. Using the model, several series of circulation patterns and corresponding rainfall occurrences were simulated. Statistics of the simulated and the observed data were found to be 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 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/nonoccurrence 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. It was observed that the model was able reproduce the probability distribution of daily rainfall amounts reasonably well, but with some downward bias.

Charles et al (1999) extended the nonhomogeneous 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, …, N). The conditional distribution consisted of regressions of inverse normal transformed amounts at a give site on rainfall occurrence at neighbouring sites within a given radius (δ km). An automatic variable selection procedure was used to identify the key neighbouring sites. The precipitation model can be expressed as

$$z_s^{(i)} = \theta_{0s}^{(i)} + \sum_{k \in n_i(\delta)} \theta_{ks}^{(i)} r^{(k)} + \varepsilon_s^{(i)} \qquad i = 1, ..., n \tag{3.18}$$

where the $\theta_{ks}^{(i)}$ are regression parameters, $n_i(\delta)$ 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

$$z_s^{(i)} = \Phi^{-1}\{F(y_s^{(-1)})\} \tag{3.19}$$

in which Φ 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 satisfactorily reproduced the dry and wet spells. Only the Spearman rank inter-site correlations were compared for the rainfall amounts. The results for the summer season was not presented.

Pegram and Seed (1998) developed a space-time model for the generation of daily rainfall in Bethlehem, South Africa. The model has two components:

- a climate generator in the form of a 3-state Markov chain with periodically varying parameters
- bins of rainfall data (a collection 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.

Table 3.1 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

Starting from a known current state, the models first determines 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 from the collection of scattered rain days in the appropriate month an historical date at random. Look up the set of data for that date which in the current model is the mean value of rainfall and record it. If the type is general note how many general rain days are in the current 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 1 to 5 days and record that sequence as it occurred historically. The model will thus produce a sequence of daily averages of rainfall based on historical record. The actual rainfalls that fell at all active rain gauges on rain days are obtained from the historical data.

Bardossy and Plate (1992) developed a multidimensional stochastic model for the space-time distribution of daily rainfall using atmospheric circulation patterns.

Let A_t be the random variable describing the atmospheric circulation pattern, taking its values from the set of possible patterns $\{\alpha_1, \ldots, \alpha_n\}$. Let the daily rainfall at location u and time t be the random function Z(t,u). Z(t,u) is related to a normally distributed random function, W(t,u), through the following transformation:

$$Z(t,u) = \begin{cases} 0 & W(t,u) \le 0 \\ W^{\beta}(t,u) & W(t,u) > 0 \end{cases} \qquad (3.20)$$

where β is a positive exponent needed as the distribution of rainfall amounts is generally much more skewed than the truncated normal distribution. The probability of rainfall at time t and point u depends on A_t:

$$P[W(t,u) > 0|A_t = \alpha_i] = P[Z(t,u) > 0|\ A_t = \alpha_I] = p_i(u) \qquad (3.21)$$

The distribution of daily rainfall amount at point u, $F_i(z|u)$ is:

$$P[Z(t,u) < z|\ A_t = \alpha_I,\ Z(t,u) > 0] = F_i(z|u) \qquad (3.22)$$

The expected daily rainfall at point u for circulation pattern α_I is

$$m_i(u) = E[Z(t,u) \mid A_t = \alpha_i] = \int_0^{\infty} \tau f_i(\tau \mid u) d\tau \qquad (3.23)$$

where $f_i(\tau|u)$ is the density function corresponding to $F_i(\tau|u)$. The expression for the corresponding W(t,u) is

$$w_i(u) = E[W(t,u)|\ A_t = \alpha_I]$$

To simplify the subsequent development, the following notation is used.

$$\mathbf{W}(t) = \{W(t,u_1), \ldots, W(t,u_n)\}$$

$$\mathbf{Z}(t) = \{Z(t,u_1), \ldots, Z(t,u_n)\}$$

$$\mathbf{w}_i = \{ w_i(u_1), \ldots , w_i(u_n)\}$$

The random process describing $\mathbf{W}(t)$ is assumed to be a multivariate autoregressive process in the case of a persisting atmospheric circulation pattern α_i:

$$\mathbf{W}(t) = \mathbf{B}_i[\mathbf{W}(t\text{-}1)\text{-}\mathbf{w}_i)] + \mathbf{C}_i\Psi(t) + \mathbf{w}_i \qquad (3.24)$$

where $\mathbf{B}_i$ and $\mathbf{C}_i$ are (n x n) matrices and $\boldsymbol{\Psi}(t) = \{\Psi(t,u_1), \ldots , \{\Psi(t,u_n)\}$.

The matrices $\mathbf{B}_i$ and $\mathbf{C}_i$ are related to $\mathbf{W}(t)$ through

$$\mathbf{M}_{0i} = E[\{\mathbf{W}(t) - \mathbf{w}_i\}\{\mathbf{W}^T(t) - \mathbf{w}_i\}| A_t = \alpha_i] \qquad (3.25)$$

$$\mathbf{M}_{li}^T = E[\{\mathbf{W}(t-1) - \mathbf{w}_i\}\{\mathbf{W}^T(t) - \mathbf{w}_i\}| A_t = A_{t-1} = \alpha_i] \qquad (3.26)$$

$$\mathbf{B}_i = \mathbf{M}_{li}\mathbf{M}_{0i}^{-1} \qquad (3.27)$$

$$\mathbf{C}_i\mathbf{C}_i^T = \mathbf{M}_{0i} - \mathbf{M}_{li}\mathbf{M}_{0i}^{-1}\mathbf{M}_{li}^T \qquad (3.28)$$

There is no time continuity if atmospheric circulation pattern changes at time t.

$$\mathbf{W}(t) = \mathbf{D}_i\Psi(t) + \mathbf{w}_i \qquad (3.29)$$

$$\mathbf{D}_i\mathbf{D}_i^T = \mathbf{M}_{0i} \qquad (3.30)$$

The model was applied to the rainfall data recorded at 44 stations in the river Ruhr catchment (5000 km^2) using the classification scheme of the German weather service. The parameters were estimated from the moments of the observed data. It was reported that the model reproduced the point and spatial rainfall statistics including rainfall covering only part of the total area under study.

3.2.2 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 K sites first. The collection of individual site models are driven with vectors of uniform [0,1] variates $\mathbf{u_t}$ and $\mathbf{v_t}$ whose elements, $u_t(k)$ and $v_t(k)$ respectively, are correlated so that corr$[u_t(k), u_t(l)] \neq 0$ and corr$[v_t(k), v_t(l)] \neq 0$, and are serially and mutually independent corr$[u_t(k), v_t(l)]$ = corr$[u_t(k), u_{t+1}(l)]$ = corr$[v_t(k),v_{t+1}(l)] = 0$. Non-zero correlations among the elements of $\mathbf{u_t}$ and $\mathbf{v_t}$ result in intersite correlations between the generated rainfall sequences.

Multisite occurrence process

Given a network of K locations, there are K(K - 1)/2 pairwise correlations that should be maintained in the uniform random numbers **($\mathbf{u_t}$)** forcing the occurrence process. The uniform variates $u_t(k)$ can be derived from standard Gaussian variates $w_t(k)$ through the transformation

$$u_t(k) = \Phi[w_t(k)] \qquad (3.31)$$

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

$$\omega(k,l) = \mathrm{Corr}[w_t(k), w_t(l)] \qquad (3.32)$$

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

$$\xi(k,l) = \mathrm{Corr}[X_t(k), X_t(l)] \qquad (3.33)$$

Let $\xi^0(k,l)$ denote the observed value of $\xi(k,l)$, which has been estimated from the observed binary series $X_t^0(k)$ and $X_t^0(l)$ at stations k and l. Hence the problem reduces to finding the K(K –1) correlations of $\omega(k,l)$

which together with the corresponding pairs of transition probabilities reproduces $\xi^0(k,l) = \xi(k,l)$ for each pair of stations. Direct computation of $\omega(k,l)$ from $\xi^0(k,l)$ is not possible. In practice, one can invert the relationship between $\omega(k,l)$ and $\xi(k,l)$ using a nonlinear rootfinding algorithm. Realisations of the vector $\mathbf{w}_t$ may be generated from the multivariate normal distribution having mean vector $\mathbf{0}$ and variance-covariance matrix Ω, whose elements are the correlations $\omega(k,l)$.

Multisite rainfall amounts process

The probability distributions of rainfall amounts conditional on nearby stations being dry have smaller means than the corresponding distributions conditional on near neighbours being wet. This problem of smooth transition between wet and dry areas was referred to as spatial intermittence by Bardossy and Plate (1992), who point out that failure to address it leads to unrealistically sharp transition between wet and dry areas.

The precipitation mean $\beta_1(k)$ or $\beta_2(k)$ are chosen at each location according to the relationship between the uniform variate for precipitation occurrence $u_t(k)$ and the mixing parameter $\alpha(k)$.

$$\beta_t(k) = \begin{cases} \beta_1(k) & u_t(k)/p_c(k) \le \alpha(k) \\ \beta_2(k) & u_t(k)/p_c(k) > \alpha(k) \end{cases} \quad (3.34)$$

where $u_t(k) = \Phi[w_t(k)]$. Because of the strong spatial correlation in the $w_t(k)$, wet locations in proximity to areas of rainfall for day t will generally result from $u_t(k)$ smaller but near $p_c(k)$, so that Eq (3.31) will choose the smaller of the two mixed exponential scale parameter $\beta_2(k)$ for that day. The locations in the heart of a wet area on day t are more likely to have been forced by a relatively small $u_t(k)$, in which case Eq (3.31) will choose the larger scale parameter $\beta_1(k)$. The suppression of large rainfall amounts at the fringe of wet regions is enhanced by continuously varying the larger of the two scale parameters according to

$$\beta_t(k) = \begin{cases} \beta_2(k) + 2[\beta_1(k) - \beta_2(k)]\left[1 - \dfrac{u_t(k)}{\alpha(k)p_c(k)}\right] & u_t(k)/p_c(k) \le \alpha(k) \\ \beta_2(k) & u_t(k)/p_c(k) > \alpha(k) \end{cases} \quad (3.35)$$

After choosing the scale parameter $\beta_1(k)$, the rainfall amounts are generated from a vector of correlated uniform variates $\mathbf{v}_t$. As in the rainfall occurrence model, it is convenient to obtain the elements of this vector from a corresponding realisation of correlated standard normal variates $z_t(k)$ as $v_t(k) = \Phi[z_t(k)]$. This vector $\mathbf{z}_t$ is drawn from a multivariate normal distribution with mean 0 and variance-covariance matrix $\mathbf{Z}$, whose elements are

$$\zeta(k,l) = \mathrm{Corr}[z_t(k), z_t(l)] \quad (3.36)$$

As was the case of $\mathbf{\Omega}$, direct computation of $\mathbf{Z}$ is not feasible since the $\mathbf{z}_t$ are not observed. The correlations in Eq (3.33) can be estimated using a similar procedure to the one used in the rainfall occurrence model.

The model was applied to a network of 25 rainfall stations in New York state with interstation separations ranging from 10 to 500 km. The model reasonably reproduced various aspects of the joint distribution of daily rainfall at the modelled stations. The mixed exponential distributions provided substantially better fit than the more conventional gamma distribution and found to be 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 well reproduced. In addition, the use of mixed exponential rather than gamma distribution resulted in closer interannual variability to the observed.

3.2.3 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 nonhomogeneous 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 enables 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 d-1, as simulated by the temporal model), this gives an initial volume $R_oL_o^2$, where L_o is the outer length scale. At the first level, the initial area is subdivided into 4 subareas denoted by Δ^i_1, i = 1, …, 4. At the second level, each of the above subareas is further subdivided into 4 further subareas denoted by Δ^i_2, i = 1, …, 16. When the process of subdivision is continued, the volume $\mu_n(\Delta^i_n)$ in the subareas at the n[th] level of subdivision (Δ^i_n, i = 1,...,2^n) are given by

$$\mu_n(\Delta^i_n) = R_o L_o^2 2^{-n} \prod_{j=1}^{n} W^i_j \qquad (3.37)$$

where for each j, i represents the subareas along the path to the n[th] level subareas and the multipliers W are nonnegative 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).

$$W = BY \qquad (3.38)$$

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 (3.29) is modified to include a systematic multiplier G.

$$W = BYG \qquad (3.39)$$

with the condition that the average value of G over the respective subareas is equal to 1 at every discretisation step.

The model was applied to a 400 x 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) spatial patchiness characteristics of daily rainfall, estimated in terms of a wet fraction; (3) statistical characteristics relating to storm arrival and interarrival times at a selected number of stations; and (4) probability distributions and exceedance probabilities of rainfall at selected stations for selected months. The model underpredicted the mean number of wet days and the mean wet spell lengths, especially during the winter months. A possible reason given to this is the exclusion of space-time correlations in the model.

3.3 Summary

From the limited amount of work done in generating daily rainfall at a several sites, the approach used by Jothityangkoon et al (2000) appears to be promising. The approach used by Wilson et al (1992) is hierarchical and becomes difficult to handle for medium to large number of stations. The method of Bardossy and Plate (1991) uses truncated 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 appears to be very cumbersome. The extension of single site chain model to multisites (Wilks, 1998) appears to be cumbersome in terms of the number of model parameters and the way to estimate the parameters. The model used by Pegram and Seed (1998) will generate only the rainfall values, which were already present in the historical record. At this stage, we propose to try the approach used by Jothityangkoon et al (2000) to Murrumbidgee catchment.

4 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 like those of Boughton (1966), Crawford and Linsley (1966), or the HYDROLOG (Porter and McMahon, 1972) require evaporation data along with rainfall as input. Crop growth models like those of Ritchie (1981) and Saxton and Bluhm (1982) 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 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.

4.1 Annual climate data

Annual climate data for a single site can be generated by using a multi-site type model (Young and Pisano, 1968).

$$\mathbf{X}_t = \mathbf{A}\,\mathbf{X}_{t-1} + \mathbf{B}\,\varepsilon_t \qquad (4.1)$$

where $\mathbf{X}_t$ = (p x 1) vector of standardised annual climate data for year t

ε_t = (p x 1) vector of random deviates with zero mean and unit variance

A, B = (p x p) matrices of constant coefficients to preserve the cross correlations.

p = number of climate variables

The coefficient matrices can be obtained from the correlation matrices as indicated in the Section 2.2.

Annual climate data at multiple sites (m) can also be generated by using Eq (4.1) in which the vector $\mathbf{X}_t$ represents the climate data at the m sites for year t.

4.2 Monthly climate data

Monthly climate data can be obtained by disaggregating the generated annual data as mentioned in Section 2.3.

4.3 Daily climate data

4.3.1 Daily climate data at a single site

Jones et al (1972) hypothesized that daily temperature and evaporation could be obtained from the time of the year and the occurrence of rainfall on the 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 autocorrelations 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 autocorrelations for minimum and maximum temperature. The model adequately simulated most of the cross correlations and lag 1 autocorrelations. 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, 1983), but the modified model will not preserve the cross correlations.

Bruhn et al (1980) presented a parsimonious model which modelled, on a monthly basis, only those cross correlations and lag one autocorrelations 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 cases since only solar radiation on dry days appeared to deviate significantly from normality.

In order 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 autocorrelations.

$$\mathbf{X}_t = \mathbf{A}\,\mathbf{X}_{t-1} + \mathbf{B}\,\varepsilon_t \tag{4.2}$$

where $\mathbf{X}_t$ = (3 x 3) matrix of standardised daily climate data for year t

ε_t = (3 x 1) vector of independent random deviates with zero mean and unit variance

A, B = (3 x 3) matrices of constant coefficients to preserve 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** 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 US (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 generate temperature and solar radiation conditional on the amount of cloud cover for a particular day.

Srikanthan (1985) modified the model given by Eq (4.2) so that it is conditional on the present and preceding days and to take 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.

Racsko 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 negative 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 grouped 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 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-correlation 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 randomly selected from a set of nearest neighbours. The model was tested on daily data for Tucson and Safford, Arizona, 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 quite well reproduced 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 was found to be not significantly different for the observed and simulated data.

Rajagopalan et al (1977) 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, 1995). A vector of solar radiation, maximum temperature, minimum temperature, dew point temperature and wind speed is then simulated by conditioning on 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, 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 with the correlations with the rainfall being 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, 1977). A vector of solar radiation, maximum temperature, minimum temperature, dew point temperature and wind speed on a day of interest is resampled from the historical data by conditioning on the vector of the same variables (feature vector) on the preceding day. The resampling 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, Utah, USA 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 well the moments, quantiles, dry and wet spells and the correlations for all the four seasons. However, only the mean values were reproduced by the MAR model and the variance, skew and quantiles were often biased.

4.3.2 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.

4.4 Summary

As climate data are less variable than rainfall, but correlated among themselves and with rainfall, mutisite models have been used successfully to generate annual data. The monthly climate data can be obtained by disaggregating the generated annual data. On a daily time step at a site, climate data has been generated by 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 successfully modelled by truncated power normal distribution (Bardossy and Plate 1992), then the model data can be easily extended to generate daily climate data at several sites simultaneously.

5 Rainfall and Climate Data Under Climate Change Scenario

Concerns over climate change caused by increasing concentration of CO_2 and other trace gases in the atmosphere has increased in recent years. A major effect of climate change may be alterations in regional hydrologic 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 1986). The main source of climate change projections is the 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 regional scale information that are necessary for hydrological studies. As a result, the climate change impact studies 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.

5.1 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). The rainfall records were multiplied by the monthly precipitation ratios for the CO_2-doubling and control runs. The monthly temperature difference between the CO_2-doubling and control runs was added to the historical temperature data. The potential evapotranspiration (PET) was computed using the Penman equation for two different sets of monthly temperature data for the CO_2-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.

5.2 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 assumed climates were specified by the monthly means and variances of rainfall and temperature.

For a two part model rainfall model with gamma distribution for rainfall amounts, there are four parameters (p_{11}, p_{01}, α, β). The transition probabilities are convenient for Monte-Carlo simulation. However, these are replaced by the unconditional probability of a wet day (π) and a dependence parameter (d).

$$\pi = p_{01}/(1 + p_{01} - p_{11}) \tag{5.1}$$

$$d = p_{11} - p_{01} \tag{5.2}$$

Denoting the parameters for the changed conditions with primes, the ratios of the monthly means and variances result in:

$$\frac{\mu'}{\mu} = \frac{\pi'\alpha'\beta'}{\pi\alpha\beta} \tag{5.3}$$

$$\frac{\sigma'^2}{\sigma^2} = \frac{\pi'\alpha'\beta'^2\left[1+\alpha'(1-\pi')\dfrac{1+d'}{1-d'}\right]}{\pi\alpha\beta^2\left[1+\alpha(1-\pi)\dfrac{1+d}{1-d}\right]} \tag{5.4}$$

The left hand side of the above two equations has the known monthly means and variances under the present and changed conditions. All the variables without prime on the right hand side are known as well. Hence there are four unknowns (π, d, α, β) and two equations. Two additional constraints are required to solve the equations. The nature of these constraints will depend on other available information. One of the simplest form is to assume no change in the precipitation occurrence process so that the only changes are in the gamma distribution parameters. Bates et al (1994) set $\pi' = \pi\pi_2/\pi_1$ and $d' = dd_2/d_1$ where the subscripts 1 and 2 denote the GCM values for a near by GCM grid cell for control and doubled CO_2 runs respectively.

The mean daily temperature, T(t), is generally taken as the average of the maximum and minimum temperature. Hence the mean of the mean daily temperature is given by

$$\mu_T(t) = \{\mu_{max}(t) + \mu_{min}(t)\}/2 \qquad (5.5)$$

A convenient approach is to define separate annual Fourier harmonics for the changes in the maximum and minimum temperatures.

$$\mu'_T(t) - \mu_T(t) = \{\Delta\mu_{max}(t) + \Delta\mu_{min}(t)\}/2$$

$$= \frac{1}{2}\left\{CX_0 + CX_1 \cos\left[\frac{2\pi(t-\phi)}{365}\right] + CN_0 + CN_1 \cos\left[\frac{2\pi(t-\phi)}{365}\right]\right\} \qquad (5.6)$$

where CX_0 and CN_0 are the annual average changes in maximum and minimum temperatures respectively and CX_1 and CN_1 are the corresponding amplitudes of the harmonics. The phase angle φ of the greatest temperature increase is assumed to be the same day for both maximum and minimum temperature components. Since there is a general consensus that warming would be greater in winter than summer (Grotch and MacCracken, 1991; Schlesinger and Mitchell, 1987), φ is assumed to be 21 days for both maximum and minimum temperatures. Appropriate values for φ needs to be chosen for locations in southern hemisphere. The remaining four parameters are estimated from the specification of an average annual temperature increase at a location, the change in temperature range between summer and winter, the change in average diurnal temperature range ($DR = \mu_{max} - \mu_{min}$) in winter and the corresponding change in summer.

$$\Delta\mu = (CX_0 + CN_0)/2 \qquad (5.7)$$

$$\Delta[\mu(JJA) - \mu(DJF)] = -0.895(CX_1 + CN_1) \qquad (5.8)$$

$$\Delta DR(DJF) = CX_0 - CN_0 + 0.895(CX_1 - CN_1) \qquad (5.9)$$

$$\Delta DR(JJA) = CX_0 - CN_0 - 0.895(CX_1 - CN_1) \qquad (5.10)$$

The average of the cosine function in Eq (5.6) results in the constant 0.895 in the above equations.

In calculating the variance terms, the nature of daily temperature autocorrrelations remains the same in the changing climate and this leads to

$$\left[\frac{Var(T')}{Var(T)}\right]^{1/2} = \frac{\sigma'_d}{\sigma_d} \qquad (5.11)$$

The variance of the daily mean temperature is obtained form

$$\sigma_d^2 = \frac{1}{4}(\sigma_{max}^2 + \sigma_{min}^2 + 2\rho_{max-min}\sigma_{max}\sigma_{min}) \qquad (5.12)$$

where $\rho_{max-min}$ is the cross correlation between daily maximum and minimum temperatures.

For the mean daily minimum (μ'_{1j}, j=0,1) and maximum μ'_{2j}, j = 0,1) temperatures under doubled CO_2 conditions, Bates et al (1994) assumed the following:

$$\mu'_{kj}(t) = \mu_{kj}(t) + \mu_{kj}^2(t) - \mu_{kj}^1(t) \qquad (5.13)$$

where the superscripts 1 and 2 refer to the control and doubled CO_2 runs. A Fourier series representation with up to 3 harmonics was used to model the within year variations in 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 its variance and autocorrelation as well. Katz (1996) derived the theoretical statistical properties of a simplified version of Richardson's model and showed how best to adjust the model parameters to obtain the desired climate change.

5.3 Summary

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

6 Conclusions/Recommendations

The models for the generation of annual, monthly and daily rainfall and climate data were reviewed in chapters 2 to 4. Since the rainfall and climate data are much less variable and less correlated than streamflow, the existing models can be used to generate these at annual and monthly level for single and multisites. As these models do not take into account of long term persistence, HSM and other models need to be investigated. Regarding daily rainfall, the transition probability matrix method performs well, but is not suitable for regionalisation and with limited length of data. Wang and Nathan (2000) approach appears to be promising. Based of this review, the following models are recommended for testing and adoption. The recommended models can be used to generate climate data under climate change conditions by adjusting the parameters appropriately.

6.1 Single site models

Annual rainfall model: A Markov model with Wilson-Hilferty transformation. Investigate the possibility of incorporating both parameter uncertainty and year to year variation in parameters. Also, apply the HSM model to generated annual rainfall.

Monthly rainfall model: Modified disaggregation scheme proposed by Mejia and Rousselle (1976) with appropriate transformation. The generation of the appropriate number months of zero rainfall month is a challenge if we do not use the method of fragments. Investigate the possibility of varying the model parameters from year to year.

Daily rainfall model: Compare the approach of Wang and Nathan (2000) with transition probability matrix method. Investigate the year to year variation in model parameters.

Annual climate model: Multisite type model as described above.

Monthly climate model: Disaggregation scheme as described above.

Daily climate model: Multi-site model conditioned on the state of present and previous days. Investigate the means of incorporating correlations between rainfall depths and climate variables into the model.

6.2 Multisite models

Annual rainfall and climate model: Multisite model as discussed above.

Monthly rainfall and climate model: Disaggregation scheme of generated annual values as discussed above.

References

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

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

Bates, B. C., Jakeman, A. J., Charles, S. P., Sumner, N. R. and Fleming, P. M. 1994. Impact of climate change on Australia's surface water resources. In "Greenhouse: Coping with climate change" Bouma, W. J., Pearman, G. I. and Manning, M. R. (Ed.): 248-262.

Bell, T.L. 1987. A space-time stochastic model of rainfall for satellite remote-sensing studies. J. Geophys. Res. 92: 9631-9643.

Bonnel, M. and Sumner, G. 1992. Autumn and winter daily precipitation events in Wales 1982-83 to1986-1987. International Journal of Climatology 12, 77-102.

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. Journal of applied Meteorology, 19: 1029-1036.

Buishand, T. A. 1977. Stochastic modelling daily rainfall sequences. Meded. Landbouwhogesch, Wageningen, 77(3), 211pp.

Chang, T. J., Kavvas, M. L. and Delleur, J. W. 1984. Daily precipitation modelling by discrete autoregressive moving average processes. Water Resources Research, Vol. 20(): 565-580.

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.

Chaouche, A. and Parent, E. 1999. Bayesian identification and validation of a daily rainfall model under monsoon conditions. Hydrol. Sci J, 44(2), 199-220 (in French)

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

Chia, E. H. 1991. Stochastic modelling of the weather – selected aspects. Ph. D. Thesis, Australian National University, Canberra.

Chin, E. H. and Miller, J. F. 1980. On the conditional distribution of precipitation amounts. Monthly Weather Review, 108: 1462-1464.

Cox, D.R. and lsham, V., 1994. Stochastic models of precipitation. In: Barnett, V. Turkman, K.F. (Eds) Statistics for the environment 2: Water issues. Wiley, New York, pp. 3-18

Delleur, J. W., Chang, T. J., and Kavvas, M. L. 1989. Simulation models of sequences of wet and dry days. J. Irrig. And Drainage Engr., 115(3): 344-357.

Eidsvik, K. J. 1980. Identification of models for some time series of atmospheric origin with Akaike's information criterion. Journal of Applied Meteorology, 19(4): 357-369.

Foufoula-Georgiou, E. and Lettenmeier, D. 1986. Continuous-time versus discrete-time point process models for rainfall occurrence series. Water Resources Research, 22 (4):531-542.

Foufoula-Georgiou, E. and Lettenmeier, D. 1987. Markov renewal model for rainfall occurrence. Water Resources Research, Vol. 23(5): 875-884.

Gelman, A., Canlin, J. Stein, H. and Rubin, D. 1995. Bayesian data analysis. Chapman and Hall, London UK.

Geng, S., Penning De Vries F. W. T. and Supit, I., 1986. A simple method for generating daily rainfall data. Agric. For. Meteorol. 36: 363-376.

Gleick, P. H. 1989. Climate change, hydrology, and water resources. Reviews of Geophysics, 27: 329-344.

Guenni, L. Rose, C. W. Bradlock , R. D., Hogarth, W. and Charles-Edwards, D. 1990. Seasonal changes in inter-relationships between climatic variables. Agricultural and Forest Meteorology, 53: 45 –58.

Hay, L. E., McCabe Jr, G. J., Wolock, D. M. and Ayers, M. A. 1991. Simulation of precipitation by weather type analysis. Water Resources Research, 27: 493-501.

Hipel, K. W. 1985. Stochastic research in multivariate analysis. Fourth International Hydrology Symposium, Fort Collins, Colorado, U.S.A: 2-50.

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

Hutchinson, M. F. 1987. Methods of generation of weather sequences. In A H Bunting (Editor) Agricultural Environments: Characterisation, Clasification and Mapping, CAB International, Wallingford, UK: 149-157.

Hutchinson, M. F. 1990b. Robust calibration of seasonally varying stochastic weather models using periodic smoothing splines. Math. Comput. Simul 32:125-130.

Hutchinson, M. F. 1991. Climate analysis in data sparse regions. In R C Muchow and J A Bellamy (Editors) Climatic risk in crop production. CAB International, Wallingford, UK: 55-71.

Hutchinson, M.F. 1995. Stochastic space time weather models from ground based data. Agricultural and Forest Meteorology, 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. Journal of Hydrology 185: 45-69.

Jimoh, O. D. and Webster, P. 1999. Stochastic modelling daily rainfall in Nigeria: intra-annual variation of model parameters. Journal of Hydrology 222:1- 17.

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

Kavvas, L. and Delleur, J. W. 1981. A stochastic cluster model of daily rainfall sequences. Water Resources Research, 17 (4):1151-1160.

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.

Lall, U. and Sharma, A. 1996. A nearest neighbour bootstrap for resampling hydrologic time series. Water Resources Research, 32(3): 679-693.

Lall, U., Rajagopalan, B.and Tarboton, D. G. 1996. A non-parametric wet/dry spell model for resampling daily precipitation. Water Resources Research, Vol. 32(9): 2803-2823.

Lane, W. L. 1979. Applied stochastic techniques (LAST computer package), user manual. Division of Planning Technical services, Bureau of Reclamation, Denver, Colarado.

Larsen, G. A. and Pense, R. B. 1981. Stochastic simulation daily climate data. SRS Staff Report AGE810831. Statistical Reporting Service, Washington DC, USDA.

Leavesley, G. H. 1994. Modelling the effects of climate change on water resources – a review. Climatic Change 28: 159-177.

Lettenmaier, D. P. and Gan, T. Y. 1990. Hydrologic sensitivities of the Sacramento-San Joaquin River basin, Califronia for global warming. Water Resources Research, 26(1): 69-86.

Loaiciga, H. A., Valdes, J. B., Vogel, R., Garvey, J. and Schwarz, H. 1996. Global warming and the hydrologic cycle. Journal of Hydrology, 174: 83-127.

Loaiciga, H. A., Maidment, D. R. and Valdes, J. B. 2000. Climate-change impacts in a regional karst aquifer, Texas, USA. Journal of Hydrology, 227: 173-194.

Maheepala, S. and Perera, C. J. C. 1996. Monthly hydrologic data data generation by disaggregation. Journal of Hydrology, 178: 277-291.

Mejia, J. M. and Rouselle, J. 1976. Disaggregation models in hydrology revisited. Water Resources Research, Vol. 12(2): 185-186.

Mimikou, M. A., Baltas, E., Varanou, E. and Pantaziz, K. 2000. Regional impacts of climate change on water resources quantity and quality indicators. Journal of Hydrology, 234: 95-109.

Over, T. M. and Gupta, V. K. 1996. A space-time theory of meso-scale rainfall using random cascades. Journal of Geophysical Research, 101: 26319-26331.

Pegram, G. G. S. and James, W. 1972. Multivariate autoregressive model for the generation of operational hydrology. Water Resources Research 8(4): 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, 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.

Rajagopalan, B., Lall, U. and Tarboton, D. G. 1996. A nonhomogeneous Markov model for daily precipitation simulation. ASCE, Journal of Hydrologic Engineering, 1(1): 33-40.

Rajagopalan, B., Lall, U., Tarboton, D. G. and Bowles, D. S. 1997. Multivariate non-parametric resampling scheme for generation of daily weather variables. Stochastic Hydrology and Hydraulics, 11(1): 65-93.

Rajagopalan, B., Lall, U. and Tarboton, D. G. 1997. Evaluation of kernel density estimation methods for daily precipitation resampling. Stochastic Hydrology and Hydraulics, 11(1): 523-547.

Rajagopalan, B. and Lall, U. 1999. A k-nearest-neighbour simulator for daily precipitation and other weather variables. Water Resources Research 35(10): 3089-3101.

Racsko, P., Szeidl, L. and Semenov, M. 1991. A serial approach to local stochastic weather models. Ecol. Modelling 57: 27-41.

Richardson, C. W. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. Water Resources Research, 17: 182-190.

Richardson, C. W. and Wright, D. A. 1984. WGEN: Amodel for generating daily weather variables. ARS-8, US Department of Agriculture, 83pp.

Robock, A., Turco, R. P., Harwell, M. A., Ackerman, T. P., Andressen, R., Chang, H-s. and Sivakumar, M. V. K. 1993. Use of general circulation model output in the creation of climate change scenarios for impact analysis. Climatic Change, 23(4): 293-335.

Roldan, J. and Woolhiser, 1982. Stochastic daily precipitation models: 1. A comparison of occurrence processes. Water Resources Research, 18(5): 1451-1459.

Rosbjerg, R. and Madsen, H., 1992. Prediction in partial duration series with generalised Pareto distributed exceedences. Water Resources Research, Vol. 28 (11): 3001-3010.

Salas, J. D., Delleur, J. W., Yevjevich, V. and Lane, W. L. 1980. Applied modelling of hydrological time series. Water Resources Publications, Littleton, Colorado.

Sanso, B., Guenni, L. 1999. A stochastic model for tropical rainfall at a single location. Journal of Hydrology, 214, 64-73.

Sanso, B., Guenni, L. 1999. Venezuelan rainfall data analysis by using a Bayesian space-time model. Applied Statistics, 48 Part 3: 345-362.

Sanso, B., Guenni, L. 1999. Venezuelian rainfall data analysed using a Bayesian space-time model. Appl. Stat (in press).

Schwarz, G. 1978. Estimating the dimension of a model. Annal of Statistics 6(2): 461-464.

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 Resources research 22(10): 1422-1430.

Smith, J. A. 1987. Statistical model of daily rainfall occurrences. Water Resources Research, Vol. 23(5):885-893.

Smith, J. A. and Karr, A. F. 1983. A point process model of summer season rainfall occurrences. Water Resources Research, 19(1): 95-193.

Smith, J. A. and Karr, A. F. 1985. Parameter estimation for a model of space-time rainfall. Water Resources Research, 21: 1251-1257.

Soltani, A., Latif, N. and Nasiri, M. 2000. Evaluation of WGEN for generating long term weather data for crop simulations. Agricultural and Forest Meteorology 102: 1-12.

Srikanthan, R. 1985. Stochastic simulation of daily climatic data, Fourth International Hydrology Symposium, Fort Collins, Colorado, U.S.A: 542-554.

Srikanthan, R. and McMahon, T.A. (1978a): A review of lag one Markov models for generation of annual flows. Jour. Hydrology, 37: 1-12.

Srikanthan, R. and McMahon, T.A. (1978b): Comparison of fast fractional Gaussian Noise and Broken Line models for generating annual flows. Jour. Hydrology, 38: 81-92.

Srikanthan, R. (1979): Stochastic generation of annual and monthly volumes. PhD Thesis, Monash University.

Srikanthan, R. and McMahon, T.A. (1980a): Stochastic generation of monthly flows for ephemeral streams. Jour. Hydrology, 47: 19-40.

Srikanthan, R. and McMahon, T.A. (1980b): Stochastic time series modelling of arid zone streamflows. Hydrological Sci. Bull., 25 (4): 423-437.

Srikanthan, R. and McMahon, T.A. (1980c): Stochastic generation of annual streamflows. Jour. Hydr. Div. ASCE, 106 (12): 2011-2028.

Srikanthan, R. and McMahon, T.A. (1982a): Stochastic generation of monthly streamflow. Jour. Hydr. Div., ASCE, 108 (3): 419-441.

Srikanthan, R. and McMahon, T.A. (1982b): Stochastic generation of annual and monthly rainfalls - a preliminary study at five Australian stations. Jour. Appl. Meteorology, 21: 1472-1479.

Srikanthan, R. and McMahon, T.A. (1983a): Stochastic simulation of daily rainfall for Australian stations. Trans. ASAE, 26 (3): 754-759, 766.

Srikanthan, R. and McMahon, T.A. (1983b): Stochastic simulation of evaporation data in Australia. Nordic Hydrology, 14 (4): 207-228.

Srikanthan, R. and McMahon, T.A. (1983c): Sequential generation of short time interval rainfall data. Nordic Hydrology, 14 (5): 277-306.

Srikanthan, R. and McMahon, T.A. (1984): Synthesizing daily rainfall and evaporation data as input to water balance-crop growth models. Jour. Aust. Inst. Agr. Sci., 50 (1): 51-54.

Srikanthan, R., McMahon, T.A., Codner G.P. and Mein, R.G. (1984): Theory and application of multi-site streamflow generation models. Civil Engg. Trans., I.E.Aust., CE26 (4):272-279.

Srikanthan, R. and T.A. McMahon. 1983. Stochastic generation of hourly rainfall data. Agricultural Engineering Research Report No. 60/83, Department of Civil Engineering, University of Melbourne, Australia.

Srikanthan, R. and T.A. McMahon. 1983. Stochastic generation of six minute interval rainfall data at Australian locations. Agricultural Engineering Research Report No. 62/83, Department of Civil Engineering, University of Melbourne, Australia.

Srikanthan, R. and T.A. McMahon. 1983. Stochastic generation of annual, monthly and daily evaporation data for Australia. Agricultural Engineering Research Report No. 63/83, Department of Civil Engineering, University of Melbourne, Australia.

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

Srinivasan, K. and Srinivas, V. V. 2000. Discussion on "Application of gamma autoregressive model to analysis of dry periods" by H. K. Cigizoglu and M Bayazit. Journal of Hydrologic Engineering, ASCE, Vol 5(1): 107-111.

Stedinger, J. R., Lettenmaier, D. P. and Vogel, R. M. 1985. Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. Water Resources Research 21(4): 497-509.

Stedinger, J. R., Pel, D. and Cohn, T. A. 1985. A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. Water Resources Research 21(5): 665-675.

Thyer, M. and Kuzcera, G. 1999. Modelling long-term persistence in rainfall time series: Sydney rainfall case study. Hydrology and Water Resources Symposium, Institution of Engineer, Australia: 550-555.

Thyer, M. and Kuzcera, G. 2000. Modelling long-term persistence in hydro-climatic time series using a hidden state Markov model. Submitted to Water Resources Research.

Tomasino, M. and Dalla Vale, F. 2000. Natural climatic changes and solar cycles: an analysis of hydrologic time series. Hydrological Sciences –Journal, 48(3): 477-489.

Tong, H. 1975. Determination of the order of a Markov chain by Akaike'e information criterion. Journal of Applied Probability 12: 486-497.

Wang, Q. J. 2000. Development of a stochastic data generation software package for Melbourne Water. Preliminary Report, Melbourne Enterprises International Ltd.

Wang, Q. J. and Nathan, R. J. 2000. A daily and monthly mixed model algorithm for stochastic generation of rainfall time series. Manuscript.

Wilby, R. L. 1997. Non-stationarity in daily precipitation series : Implications for GCM down-scaling using atmospheric indices. International Journal of Climatology, Vol 17: 439-454.

Wilby, R. L., Hay, L. E. and Leavesley, G. H. 1999. A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colarado. Journal of Hydrology, 225: 67-91.

Wilks, D. S. 1988. Estimating the consequences of CO_2-induced climate change on north American grain agriculture using general circulation model information. Climatic Change 13: 19-42.

Wilks, D. S. 1989. Conditioning stochastic daily precipitation models on total monthly precipitation. Water Resources Research 23(6): 1429-39

Wilks, D. S. 1989. Statistical specification of local surface weather elements from large-scale information. Theor. Appli. Climatol., 40: 119-134.

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. Journal of Hydrology 210: 178-191.

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

Woolhiser, D. A. 1992. Modelling daily precipitation – progress and problems. Ch. 5 in "Statistics in the environmental & earth sciences, A T Walden and P Guttorp (Ed), Edward Arnold. 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. Journal of Applied Meteorology, 18(1): 34-42.

Woolhiser, D. A. and Roldan, J. 1982. Stochastic daily precipitation models: 2. A comparison of distribution of amounts. Water Resources Research, Vol. 18(5): 1461-1468.

Woolhiser, D. A. and Roldan, J. 1986. Seasonal and regional variability of parameters for stochastic daily precipitation models. Water Resources Research, Vol. 22(6): 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 Resources Research, Vol. 29(4): 1287-1295.

Yevjevich, V. and Dyer, T. G. J. 1983. Basic structure of daily precipitation series. Journal of Hydrology, 64: 49-67.

Young, G. K. and Pisano, W. C. 1968. Operational hydrology using residuals. Journal of Hydraulics Division, ASCE, 94(HY4): 909-923.

Young, K. C. 1994. A multivariate chain model for simulating climatic parameters from daily data. Journal of Applied Meteorology, 33: 661-671.

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 Resources Research 27(8): 1917-1923.