# Stochastic generation of daily, monthly and annual precipitation


By

Said Obakrim

Supervisor: Prof. Abdelhak Zoglat

Master of statistics and econometrics

Departement of mathematics, Faculty of sciences Rabat


May 17, 2019

# Contents

# Introduction

Meteorological data are commonly needed as inputs to hydrological and agricultural modeling systems. To evaluate these models, it is rarely sufficient to use only the observed weather data. Not only because observed ground-based climate data are inadequate in terms of their length and completeness, but the use of observed sequences gives a solution based on only one realization of the weather process, hence it provides us with only one scenario.
To address this issue, it is beneficial to generate synthetic sequences of climate data based on the stochastic structure of the weather process. These statistical models are known as "weather generators" and the most effort in their creation has been dedicated to rainfall processes because precipitation is the most crucial meteorological variable for many applications.

Precipitation is a mixed variable that contains both discrete and continuous sequences and is very often exactly zero. Thus, there is a strong discontinuity in the probability distribution of rainfall data between the zero and the nonzero observations. Therefore, most models contain separate treatments of the precipitation occurrence and amounts processes. The precipitation occurrence process has two weather states, wet or dry and the precipitation amounts process simulate the nonzero observations.

# Part I

# Parametric precipitation generation

# 1

# Stochastic generation of daily precipitation

## 1.1 Single site stochastic generation of daily precipitation

Denote $X_t(k)$ the binary event of wet and dry day at location $k$ on day $t$ such that

$$X_t(k) = \begin{cases} 0 & \text{if day t is dry at location } k \\ 1 & \text{if day t is wet at location } k \end{cases} \tag{1.1}$$

Here day t is classified as wet day if the precipitation amount is above some fixed threshold. Denote $r_t(k)$ the nonzero precipitation amounts at location $k$ on day $t$.
Thus, the time series of precipitation amounts at location $k$ is then

$$Y_t(k) = X_t(k) r_t(k) \tag{1.2}$$

### 1.1.1 Single site precipitation occurrence process

Assuming that the probability of rainfall on any day depends only on whether the previous day was wet or dry, the precipitation occurrence is modeled using the first-order two-stage Markov process. The Markov process is defined by the transition probabilities $p_{01}(k)$ and $p_{11}(k)$ such that

$$p_{01}(k) = P(X_t(k) = 0 | X_{t-1}(k) = 1) \text{ and } p_{11}(k) = P(X_t(k) = 1 | X_{t-1}(k) = 1),$$

which will be estimated from the observed binary sequences of the precipitation occurrence process using maximum likelihood estimation (see Appendix A).
The other conditional probabilities are derived from these two transition probabilities by

$$p_{00}(k) = 1 - p_{01}(k) \text{ and } p_{10}(k) = 1 - p_{11}(k).$$

To simulate the $X_t(k)$ series we first generate a uniform $[0, 1]$ random number $u_t(k)$ for each day $t$ and location $k$ and the state of each day is determined using the transition probabilities $p_{01}(k)$ and $p_{11}(k)$. If the previous day $(t-1)$ was dry, then day $t$ is simulated to be wet if $u_t(k) \leq p_{01}(k)$, and otherwise it is also dry. If the previous day was wet, then the current day is simulated to be wet if $u_t(k) \leq p_{11}(k)$, and is dry otherwise.
To summarize, let $p_c(k)$ be the critical probability such that

$$p_c(k) = \begin{cases} p_{01}(k) & \textbf{if } X_{t-1} = 0 \\ p_{11}(k) & \textbf{if } X_{t-1} = 1 \end{cases} , \tag{1.3}$$

the next value of the $X_t(k)$ is determined as

$$X_t(k) = \begin{cases} 1 & \text{if } u_t(k) \leq p_c(k) \\ 0 & \text{otherwise} \end{cases} . \tag{1.4}$$

### 1.1.2 Single site precipitation amounts process

The precipitation amounts process simulates the nonzero sequences which are typically strongly skewed to the right, with many small values and few but important large precipitation amounts (Figure 1.1) that are most important to hydrological and agricultural models.

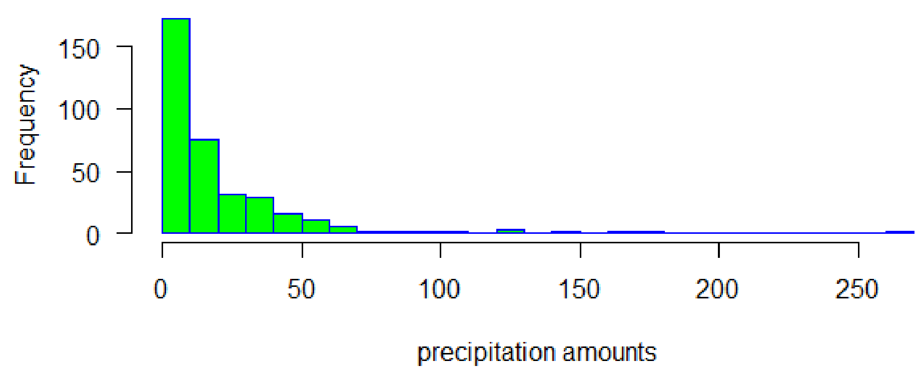


Figure 1.1: Histogram of daily precipitation amounts

Under the assumption that precipitation amounts on wet days are independent and follow the same distribution, a number of models have been proposed for the distribution of daily precipitation amounts given the occurrence of a wet day. The most common distributions are the exponential, gamma, and mixed exponential distribution.

Figure 1.2 illustrates the difference between the exponential and the gamma distributions fitted to nonzero rainfall amounts at a particular location. In this case, the gamma distribution is more consistent compared to the exponential distribution (AIC for the gamma distribution is 2830.327 and for the exponential distribution 2841.764).

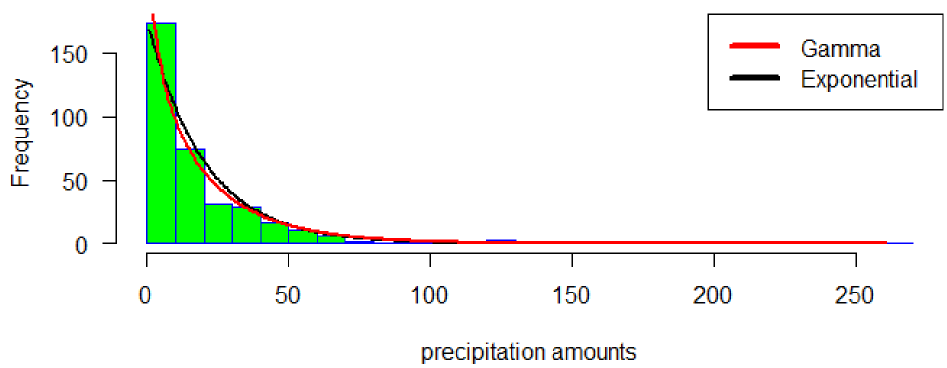


Figure 1.2: Comparison of the histogram for precipitation amounts with the corresponding fitted exponential and gamma probability density functions

Having decided which distribution fits well the nonzero precipitation amounts, simulations are accomplished by generating uniform $[0, 1]$ random variates and using the inversion theorem.
Figure 1.3 illustrates the single site two-part model consisting of the precipitation occurrences and amounts process.

## 1.2 Multisite stochastic generation of daily precipitation

Wilks (1998) extended the single site stochastic model described in the previous section to a multisite model. The basic idea in his model is to simultaneously simulate precipitation occurrences using a uniform $[0, 1]$ vector $u_t$ and the precipitation amounts using a uniform $[0, 1]$ vector $v_t$ in the way in which the elements of $u_t(k)$ and $v_t(k)$, respectively, are correlated so that

$$Corr[u_t(k), u_t(l)] \neq 0 \text{ and } Corr[v_t(k), v_t(l)] \neq 0$$

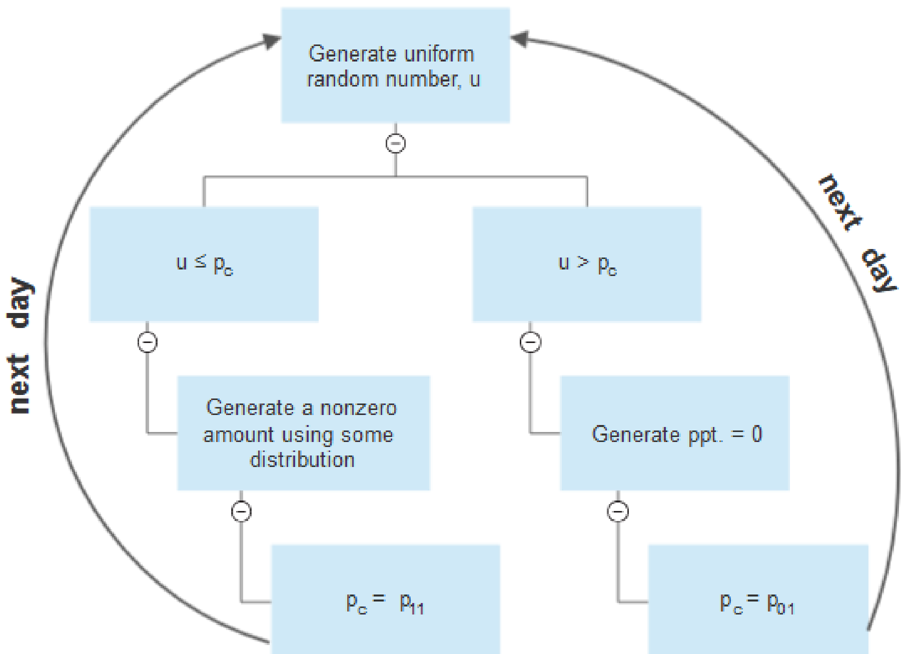


Figure 1.3: Flowcharts for single site dialy precipitation generation using two-part model

for every locations $k$ and $l$. However, those vectors must be mutually and serially uncorrelated so that

$$Corr[u_t(k), v_t(l)] = Corr[u_t(k), u_{t+1}(l)] = Corr[v_t(k), v_{t+1}(l)] = 0.$$

### 1.2.1 Multisite precipitation occurrence process

To simulate precipitation occurrences in a network of $K$ locations, there are $K(K-1)/2$ pairwise correlations that should be maintained. To achieve the desired correlations, we generate correlated uniform $[0,1]$ random variates $u_t$ whose elements, $u_t(k)$, can be derived from standard Gaussian variates $w_t(k)$ through the quantile transformation

$$u_t(k) = \Phi[w_t(k)],$$

where $\Phi$ is the standard normal cumulative distribution function.
Let $w(k,l)$ be the correlation between the station pair $k$ and $l$ so that

$$w(k,l) = Corr[w_t(k), w_t(l)]. \tag{1.5}$$

A particular $w(k,l)$ and transition probabilities $p_{01}(k)$ and $p_{11}(k)$ will yield a particular correlation between the synthetic binary series for the two sites,

$$\xi(k,l) = Corr[X_t(k), X_t(l)]. \tag{1.6}$$

Denote the observed sample of $\xi(k,l)$ as $\xi^o(k,l)$ which will have been estimated from the observed binary sequences $X_t^o(k)$ and $X_t^o(l)$ at locations $k$ and $l$. Therefore, the desired correlations are achieved by individually finding the $K(K-1)/2$ pairwise correlations $w(k,l)$ which, together with the corresponding pairs of transition probabilities, reproduces $\xi^o(k,l) = \xi(k,l)$ for each pair of stations.

Direct computation of $w(k,l)$ from $\xi^o(k,l)$ is impossible, because $w(k,l)$ cannot be observed. However, there is a polynomial relationship between $w(k,l)$ and the resulting $\xi(k,l)$ for a given pairs of transition probabilities. Figure 1.4 shows an example of such a relationship, computed using stochastic simulation.

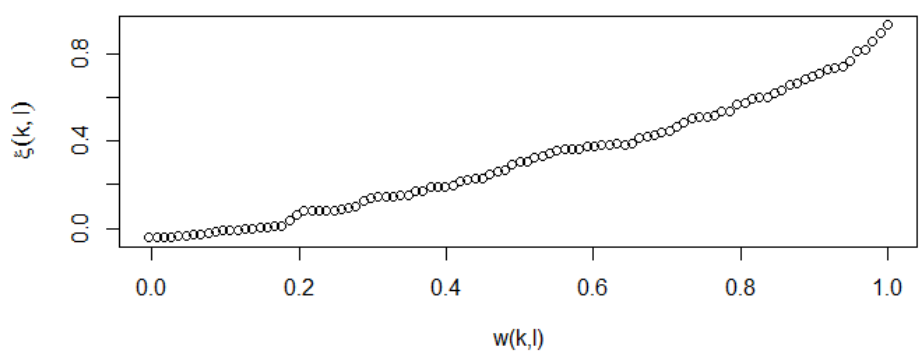


Figure 1.4: An example of the relationship between $w(k,l)$ and $\xi(k,l)$

This relationship can be estimated using polynomial regression. Figure 1.5 illustrates the estimated polynomial function.

Let $f$ be the estimated function of the relationship between $w(k,l)$ and $\xi(k,l)$ so that

$$f[w(k,l)] = \xi(k,l).$$

To solve the first aspect of the precipitation generation problem, we use a nonlinear root finding algorithm by defining the function

$$Q[w(k,l)] = f[w(k,l)] - \xi^o(k,l)$$

and finding the $w(k,l)$ for which $Q[w(k,l)] = 0$.
The $w_t(k)$ are generated using a multivariate normal distribution vector $w_t$ having mean vector 0 and variance-covariance matrix $[\Sigma]$ the elements of which are the correlations $w(k,l)$(see Appendix B).

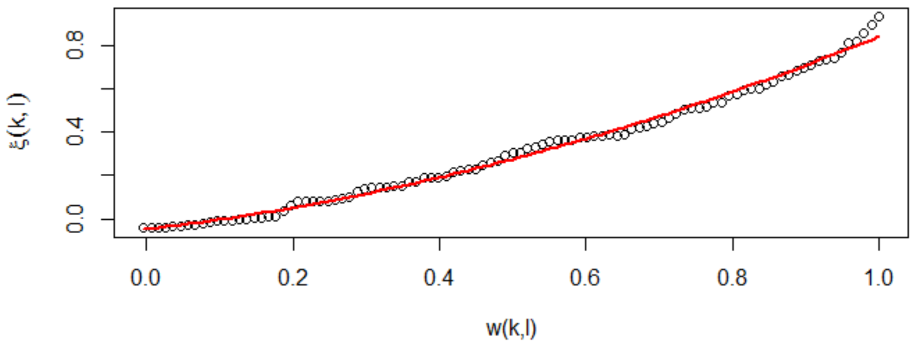


Figure 1.5: Estimation of the relationship between $w(k, l)$ and $\xi(k, l)$ using polynomial regression

The resulting individual $w_t(k)$ will be distributed as univariate standard normal distribution and may be used to generate precipitation occurrence time series $X_t(k)$ using

$$X_t(k) = \begin{cases} 1 & \text{if } w_t(k) \leq \Phi^{-1}[p_c(k)] \\ 0 & \text{otherwise} \end{cases} . \tag{1.7}$$

This process of simultaneously simulate occurrences using multivariate normal distribution defined by the $K(K-1)/2$ pairwise correlations, was referred to as the hidden covariance model (Srikanthan and Pegram, 2006).

### 1.2.2 Multisite precipitation amounts process

As was the case with the single site precipitation amounts model, the rainfall amounts on wet days at multiple locations are modeled by using distributions which fits well the observed sequences.

The probability distributions of rainfall amounts conditional on nearby stations being dry have smaller means than the corresponding distributions conditional on nearby locations being wet. Consequently, simulation of rainfall amounts $r_t$ independently of the occurrence process $X_t$ is inappropriate in the multisite model.

To address this problem, Wilks (1998) proposed to model nonzero amounts in multiple locations with the mixed exponential distribution, which has the probability density function

$$f[r(k)] = \frac{\alpha(k)}{\beta_1(k)} \exp[\frac{-r(k)}{\beta_1(k)}] + \frac{1-\alpha(k)}{\beta_2(k)} \exp[\frac{-r(k)}{\beta_2(k)}] \tag{1.8}$$

$$0 < \alpha(k) \leq 1, \;\; \beta_1(k) \geq \beta_2(k) > 0,$$

because it allows to model the nonzero amounts with smaller means, $\beta_2(k)$, in locations close to dry areas and greater means, $\beta_1(k)$, in locations close to wet areas.

The precipitation means $\beta_1(k)$ and $\beta_2(k)$ are chosen at each location $k$ according to the relationship between the uniform random variate, $u_t(k)$, for the precipitation occurrences and the parameter $\alpha(k)$ such that

$$\beta_t(k) = \begin{cases} \beta_1(k) & \textbf{if } u_t(k)/p_c(k) \leq \alpha(k) \\ \beta_2(k) & \textbf{otherwise} \end{cases}, \tag{1.9}$$

On day $t$, Equation (1.9) will choose the smaller of the two mixed exponential scale parameters, $\beta_2(k)$, in wet locations close to dry areas. Because the simulation of rainfall occurrence $X_t(k)$ in this day will generally result from $u_t(k)$ smaller than but near to $p_c(k)$.
whereas, locations close to wet areas on day $t$ are more likely to have been forced by $u_t(k)$ smaller than but far from $p_c(k)$. Consequently, Equation (1.9) will choose the larger scale parameter $\beta_1(k)$.

After choosing the scale parameters for each location on each day, we generate a vector of random uniform variates $v_t$ whose elements $v_t(k)$ are cross correlated but serially independent and can be derived from standard Gaussian variates $z_t(k)$ through the quantile transformation $v_t(k) = \Phi[z_t(k)]$. Then the precipitation amounts are generated by inversion using

$$r_t(k) = r_{min} - \beta \log[v_t(k)], \tag{1.10}$$

where $\beta$ is either $\beta_1(k)$ or $\beta_2(k)$ and $r_{min}$ is the threshold above which a day is recorded as wet day.
The vector $z_t$ is drawn from a multivariate normal distribution with mean vector 0 and variance-covariance matrix $[\Omega]$, whose elements are:

$$\zeta(k,l) = Corr[z_t(k), z_t(l)], \tag{1.11}$$

Direct computation of $[\Omega]$ from the observed precipitation data is not possible. Let $\eta(k,l)$ be the cross correlations between daily precipitation at locations $k$ and $l$ such that

$$\eta(k,l) = Corr[Y_t(k), Y_t(l)] \tag{1.12}$$

Denote the observed sample of $\eta(k,l)$ as $\eta^o(k,l)$ which will have been estimated from the observed sequences $Y_t^o(k)$ and $Y_t^o(l)$ at locations $k$ and $l$. As in the case of simulating precipitation occurrences, in a network of $K$ locations, there are $K(K-1)/2$ pairwise correlations that should be preserved in the simulation process.

Wilks (1998) has found that determining the correlations $\zeta(k,l)$, in Equation (1.11), pairwise in the manner described in for $w(k,l)$, yields matrices $[\Omega]$ that are not positive definite.

To address this issue, Wilks proposed to obtain the correlations $\zeta(k,l)$ using a smooth function of the distances between the station pairs such that:

$$\zeta(k,l) = \exp[-d_1 Dis(k,l)^{d_2}], \tag{1.13}$$

where $Dis(k,l)$ is the horizontal distance between the station pair $k$ and $l$, and parameters $d_1$ and $d_2$ are chosen for each month which minimize

$$\sum_{k,l} [\eta(k,l) - \eta^o(k,l)]^2. \tag{1.14}$$

The function of the form of Equation (1.13) can also be fitted to the correlations $w(k,l)$ for the precipitation occurrence process.

2

# Stochastic generation of monthly and annual precipitation

Generation of monthly and annual rainfall data is important for water resources systems and the estimation of water yield from large catchments. One can suggest to aggregate the generated daily rainfall sequences, detailed in the previous chapter, to monthly and annual rainfall values. However, in doing so the model does not take into consideration the month to month and year to year variations.

To address this problem, Srikanthan (2005, 2006, 2009) proposed a nested model, an extension of the two-part model for monthly and annual generation for both single and multiple sites, that ensures the monthly and annual covariances are preserved.

## 2.1 Single site monthly and annual models

### 2.1.1 Single site monthly model

As previously mentioned, the generated rainfall amounts when aggregated into monthly totals will not, in general, preserve the monthly characteristics. Thus, the daily amounts model was nested and modified by Srikanthan (2005, 2006) in a single site monthly model to improve the monthly at-site characteristics of the generated rainfall data.

Once the daily rainfall amounts at station $k$ are generated for a given month, the monthly rainfall totals, $a_i(k)$, are obtained by summing the daily rainfall values. Then the monthly rainfall totals are modified by a lag-one autoregressive monthly model.

Let $x_i(k)$ be the standardized rainfall at station $k$ on month $i$ having zero mean and unit variance. Therefore, using a lag-one autoregressive process, we have

$$x_i(k) = c_i(k)x_{i-1}(k) + d_i(k)\epsilon_i \;\; i = 2, 3, ..., n \tag{2.1}$$

where $c_i(k)$ and $d_i(k)$ are coefficient numbers defining the temporal correlations between monthly total values, $\epsilon_i$ is a random component that has zero mean and unit variance and is independent of $x_{i-1}(k)$ and $n$ is the number of years.

The coefficients $c_i(k)$ and $d_i(k)$ are defined for each month because each month is considered separately in the model in order to preserve the seasonal characteristics.

If both sides of Equation (2.1) are multiplied by $x_{i-1}(k)$ such us

$$x_i(k)x_{i-1}(k) = c_i(k)x_{i-1}^2(k) + d_i(k)\epsilon_i x_{i-1}(k) \tag{2.2}$$

Applying the expectation leads to

$$\begin{aligned} E[x_i(k)x_{i-1}(k)] &= c_i(k)E[x_{i-1}^2(k)] + d_i(k)E[\epsilon_i x_{i-1}(k)] \\ &\Longleftrightarrow \ c_i(k) = r_i(k) \end{aligned} \tag{2.3}$$

where $r_i(k)$ is the lag-one autocorrelation coefficient for station $k$ on month $i$. Hence the Equation (2.1) becomes

$$x_i(k) = r_i(k)x_{i-1}(k) + d_i(k)\epsilon_i. \tag{2.4}$$

If both sides of Equation (2.4) are multiplied by $x_i(k)$ such us

$$x_i(k)^2 = r_i(k)x_{i-1}(k)x_i(k) + d_i(k)\epsilon_i x_i(k) \tag{2.5}$$

Again, applying the expectation leads to

$$\begin{aligned} &E[x_i(k)^2] = r_i(k)E[x_{i-1}(k)x_i(k)] + d_i(k)E[\epsilon_i x_i(k)] \\ &\quad\Longleftrightarrow\ 1 = r_i(k)E[x_{i-1}(k)x_i(k)] + d_i(k)E[\epsilon_i(r_i(k)x_{i-1}(k) + d_i(k)\epsilon_i)] \\ &\quad\Longleftrightarrow\ 1 = r_i(k)E[x_{i-1}(k)x_i(k)] + d_i(k)r_i(k)E[\epsilon_i x_{i-1}(k)] + d_i^2(k)E[\epsilon_i^2] \\ &\quad\Longleftrightarrow\ 1 = r_i(k)E[x_{i-1}(k)x_i(k)] + d_i^2(k) \\ &\quad\Longleftrightarrow\ 1 = r_i^2(k) + d_i^2(k) \end{aligned}$$

Thus

$$d_i(k) = \sqrt{1 - r_i^2(k)} \tag{2.6}$$

Therefore, the equation (2.1) becomes

$$x_i(k) = r_i(k)x_{i-1}(k) + \sqrt{1 - r_i^2(k)}\,\epsilon_i \quad i = 2, 3, ..., 12n \tag{2.7}$$

Srikanthan (2004) proposed to replace the white noise in Equation (2.7) by $\hat{a}_i(k)$, the standardized generated monthly totals having zero mean and unit variance. Thus the adjusted value,$x_i(k)$ is obtained from

$$x_i(k) = r_i(k)x_{i-1}(k) + \sqrt{1 - r_i^2(k)}\,\hat{a}_i(k) \tag{2.8}$$

$$i = 2, 3, ..., 12n, \ \text{ with } \ x_1(k) = \hat{a}_1(k).$$

### 2.1.2 Single site annual model

As was the case for monthly rainfall, a similar method is used to nest the generated monthly rainfall into a single site annual model.
Having generating the rainfall amounts at station $k$ for a given year, using the nested model, the annual rainfall totals, $b_j(k)$, are obtained by aggregating the monthly precipitation values and the annual rainfall totals are also modified using a lag-one autoregressive annual model.

Denote $y_j(k)$ the standardized precipitation amounts at station $k$ on year $j$ having zero mean and unit variance. Using a lag-one autoregressive process, we have

$$y_j(k) = \rho(k) y_{j-1}(k) + \sqrt{1-\rho(k)^2}\,\hat{b}_j(k) \tag{2.9}$$
$$j = 2, 3, ..., n, \;\; \text{with} \;\; y_1(k) = \hat{b}_1(k)$$

where $\rho(k)$ is the lag-one autocorrelation coefficient for station $k$, $\hat{b}_j(k)$ is the standardized aggregated annual values having zero mean and unit variance and $n$ is the number of years.

## 2.2 Multisite monthly and annual models

### 2.2.1 Multisite monthly model

As was the case with single site monthly precipitation amounts. Having simulating the daily precipitation amounts at all sites, using the multisite two-part model, for a given month, the monthly precipitation totals, $a_i(k)$, at each location are obtained by aggregating the daily rainfall values. Then those monthly totals are standardized by their means and standard deviations and collected in vectors $\hat{a}_i$. Those standardized, aggregated monthly values are modified by using a nested multisite model (Srikanthan and Pegram 2007) to preserve the monthly spatial and serial correlations.

Let $X_i$ be the vector of the standardized adjusted monthly rainfall (zero mean and unit variance) at $K$ station on month $i$, such that

$$X_i = \begin{bmatrix} x_i(1) \\ x_i(2) \\ . \\ . \\ . \\ x_i(K) \end{bmatrix}. \tag{2.10}$$

Using a lag-one autoregressive process, we have

$$X_i = A_i X_{i-1} + B_i \hat{a}_i \quad i = 2, 3, ..., 12n, \;\; \text{with} \;\; X_1 = \hat{a}_1 \tag{2.11}$$

where $A_i$ and $B_i$ are coefficient matrices defining the temporal and spatial correlations between monthly totals for month $i$, which are calculated from the lag-zero, $M_0$, and lag-one, $M_1$, cross-correlation of the observed monthly rainfall and the lag-zero, $C_0$, cross correlation of the standardized and aggregated monthly values $\hat{a}_i$. The coefficient matrices $A_i$ and $B_i$ are derived as follows.

The series $\hat{a}_i$ may itself be spatially and serially correlated and it can be represented by a lag-one autoregressive process such as

$$\hat{a}_i = E_i \hat{a}_{i-1} + F_i \lambda_i \tag{2.12}$$

where $E_i$ and $F_i$ are coefficient matrices which preserve the lag-zero and-lag one cross-correlations in $\hat{a}_i$ and $\lambda_i$ is a vector of mutually independent standardized variates.
Equation (2.11) is equivalent to

$$\hat{a}_{i-1} = B_{i-1}^{-1} X_{i-1} - B_{i-1}^{-1} A_{i-1} X_{i-2}. \tag{2.13}$$

Applying Equations (2.12) and (2.13) successively in Equation (2.11) results in

$$X_i = A_i X_{i-1} + B_i E_i B_{i-1}^{-1} X_{i-1} - B_i E_i B_{i-1}^{-1} A_{i-1} X_{i-2} + B_i F_i \lambda_i \tag{2.14}$$

or

$$X_i = [A_i + B_i E_i B_{i-1}^{-1}] X_{i-1} - B_i E_i B_{i-1}^{-1} A_{i-1} X_{i-2} + B_i F_i \lambda_i \tag{2.15}$$

Assuming that $\hat{a}_i$ is spatially correlated but serially uncorrelated, so that $E_i = 0$ for every month $i$, then

$$\hat{a}_i = F_i \lambda_i. \tag{2.16}$$

And Equation (2.15) reduces to

$$X_i = A_i X_{i-1} + G_i \lambda_i \quad \text{where} \quad G_i = B_i F_i. \tag{2.17}$$

If both sides of Equation (2.17) are multiplied by $X_{i-1}^T$, the transpose of $X_{i-1}$, such that

$$X_i X_{i-1}^T = A_i X_{i-1} X_{i-1}^T + G_i \lambda_i X_{i-1}^T \tag{2.18}$$

applying the expectation leads to

$$E[X_i X_{i-1}^T] = A_i E[X_{i-1} X_{i-1}^T] + G_i E[\lambda_i X_{i-1}^T] \tag{2.19}$$

$$\iff M_1 = A_i M_0 \tag{2.20}$$

Hence

$$A_i = M_1 M_0^{-1} \tag{2.21}$$

where $M_0$ and $M_1$ are, respectively, the lag-zero and lag-one cross-correlation matrices of the observed monthly rainfall.
Thus the Equation (2.17) becomes

$$X_i = M_1 M_0^{-1} X_{i-1} + G_i \lambda_i \tag{2.22}$$

If both sides of Equation (2.22) are multiplied by $X_i^T$, such that

$$X_i X_i^T = M_1 M_0^{-1} X_{i-1} X_i^T + G_i \lambda_i X_i^T \tag{2.23}$$

again, applying the expectation leads to

$$\begin{aligned}
&E[X_i X_i^T] = M_1 M_0^{-1} E[X_{i-1} X_i^T] + G_i E[\lambda_i X_i^T] \\
&\iff M_0 = M_1 M_0^{-1} E[X_{i-1} X_i^T] + G_i E[\lambda_i (M_1 M_0^{-1} X_{i-1} + G_i \lambda_i)^T] \\
&\iff M_0 = M_1 M_0^{-1} E[X_{i-1} X_i^T] + G_i E[\lambda_i (M_1 M_0^{-1} X_{i-1})^T] + G_i E[\lambda_i (G_i \lambda_i)^T] \\
&\iff M_0 = M_1 M_0^{-1} E[X_{i-1} X_i^T] + G_i E[\lambda_i X_{i-1}^T] (M_0^{-1})^T M_1^T + G_i E[\lambda_i \lambda_i^T] G_i^T \\
&\iff M_0 = M_1 M_0^{-1} M_1^T + G_i G_i^T
\end{aligned}$$

Thus

$$G_i G_i^T = M_0 - M_1 M_0^{-1} M_1^T \tag{2.24}$$

Combining Equation (2.16) with (2.17) gives

$$X_i = A_i X_{i-1} + G_i F_i^{-1} \hat{a}_i \tag{2.25}$$

The coefficient matrix $F_i$ is obtained from

$$F_i F_i^T = P_0 \tag{2.26}$$

where $P_0$ is the cross correlation matrix of $\hat{a}_i$. The matrices $F_i$ and $G_i$ can be calculated using singular value decomposition.

### 2.2.2 Multisite annual model

Once the monthly precipitation time series are generated at each location for a given year, using multisite monthly model, a similar method can be used to nest the generated monthly rainfall into a multisite annual model.
At first, the monthly rainfall amounts at each location are aggregated into annual precipitation totals, $b_j(k)$. Then those annual totals are standardized by their means and standard deviations and collected into vectors $\hat{b}_j$, which will be modified using a nested multisite annual model to preserve the annual spatial and serial correlations.
Denote $Y_j$ the vector of the standardized adjusted annual rainfall at $K$ station on year $j$, such that

$$Y_j = \begin{bmatrix} y_j(1) \\ y_j(2) \\ \cdot \\ \cdot \\ \cdot \\ y_j(K) \end{bmatrix}. \tag{2.27}$$

Using a lag-one autoregressive model, $Y_j$ is obtained from

$$Y_j = CY_{j-1} + D\hat{b}_j \tag{2.28}$$

$$j = 2, 3, ..., n \ \text{ with } \ Y_1 = \hat{b}_1$$

where $C$ and $D$ are coefficient matrices which are calculated from the lag-one and lag-zero cross-correlation of the observed annual rainfall and lag-zero cross correlation of the standardized and aggregated, already generated, annual values.

# A

# Maximum likelihood estimation of transition probabilities

Consider a Markov chain ($X_1^\infty = X_t,\, t = 0, 1, ...$) with $m$ states, we want to estimate the transition matrix from the observed data $x_1^n \equiv x_1, x_2, ..., x_n$. The elements of the transition matrix $p_{ij}$ are defined as

$$p_{ij} = Pr(X_{t+1} = j | X_t = i).$$

The probability of the realization of $X_1^\infty$ is

$$Pr(X_1^n = x_1^n) = Pr(X_1 = x_1) \prod_{t=2}^{n} Pr(X_t = x_t | X_1^{t-1} = x_1^{t-1}) \tag{A.1}$$

$$= Pr(X_1 = x_1) \prod_{t=2}^{n} Pr(X_t = x_t | X_{t-1} = x_{t-1}) \tag{A.2}$$

$$= Pr(X_1 = x_1) \prod_{t=2}^{n} p_{x_{t-1} x_t} \tag{A.3}$$

Equation (A.1) uses the definition of conditional probabilities and Equation (A.2) uses the Markov propriety.
Denote $p$ the transition probabilities such that

$$p = (p_{ij},\ i = 1, ..., m,\ j = 1, ..., m)$$

Define the transition counts $n_{ij} \equiv$ number of times $i$ is followed by $j$ in $X_1^n$. Thus, the likelihood function of a given transition probabilities is defined as

$$L(p) = Pr(X_1 = x_1) \prod_{i=1}^{m} \prod_{j=1}^{m} p_{ij}^{n_{ij}} \tag{A.4}$$

and the log-likelihood function as

$$\mathcal{L}(p) = \log[L(p)] = log[Pr(X_1 = x_1)] + \sum_{i,j} n_{ij} \log(p_{ij}) \tag{A.5}$$

We want to maximize $\mathcal{L}$ under the $m$ constraint equations

$$\sum_j p_{ij} = 1, \;\; i = 1, ..., m. \tag{A.6}$$

Using $m$ Lagrange multipliers, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_m)$, the new objective function is

$$f(p, \lambda) = \mathcal{L}(p) - \sum_{i=1}^{m} ( \sum_j p_{ij} - 1 ) \tag{A.7}$$

The resulting estimations of transition probabilities, after maximizing the objective function $f$, are

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}} \tag{A.8}$$

# $\mathcal{B}$

## Simulation of multivariate normal distribution

Let $z_1, ..., z_m$ be independent and identically distributed normal variates with mean 0 and variance 1. If for constants $a_{ij}$, $i = 1, ..., n \;\; j = 1, ..., m$

$$
\begin{aligned}
w_1 &= a_{11}z_1 + a_{12}z_2 + ... + a_{1m}z_m \\
w_2 &= a_{21}z_1 + a_{22}z_2 + ... + a_{2m}z_m \\
&. \\
&. \\
&. \\
w_n &= a_{n1}z_1 + a_{n2}z_2 + ... + a_{nm}z_m
\end{aligned}
$$

which is equivalent to $w = Az'$, where $A = (a_{ij})_{\,i=1,...,n\; j=1,...,m}$, then the vector $w = (w_1, w_2, ..., w_n)$ is said to have a multivariate normal distribution such that

$$E(w_i) = 0 \;\text{ and }\; Cov(w_i, w_j) = \sum_{k=1}^{m} a_{ik}a_{jk} \forall \; 1 \leq i, j \leq n \tag{B.1}$$

Equation (B.1) is equivalent to $\Sigma = A^t A$, where $\Sigma$ is the variance-covariance matrix.
Therefore, to generate the multivariate normal distribution, we first find a matrix $A$ such that $\Sigma = A^t A$ using Choleski decomposition or singular value decomposition and then generate independent standard normal variates $z_1, z_2, ..., z_n$ and set $w = Az'$.

# Bibliography


[1] Matalas, N.C. (1967) *Mathematical assessment of synthetic hydrology. Water Resources Research 3 (4), 937–945.*

[2] Martin Haugh (2007) *Generating Random Variables and Stochastic Processes*

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

[4] Srikanthan, R. (2006) *Stochastic generation of spatially consistent daily rainfall at multiple sites.* $30^{th}$ *Hydrology and Water Resources Symposium, Launceston.*

[5] Srikanthan, R. and G. G. S. Pegram (2006) *Stochastic generation of multisite rainfall occurrence. Advances in Geosciences, Vol. 6, 1-10.*

[6] Srikanthan, R. and G. G. S. Pegram *StochasticGeneration of Spatially Consistent Dail Rainfall.*

[7] Srikanthan, R. and G. G. S. Pegram (2009) *A nested multisite daily rainfall stochastic generation model. Journal of Hydrology 371 142–153*

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

[9] Wilks, D. S. and Wilby, R. L. (1999) *The weather generation game: a review of stochastic weather models, Prog. Phys. Geogr., 23, 329–357.*