A Multisite Stochastic Downscaling Model of Daily Rainfall Occurrences with Long Term Persistence

Mehrotra, R. ¹, and A. Sharma ¹

¹ School of Civil and Environmental Engineering, The University of New South Wales, Sydney
Email: raj@civeng.unsw.edu.au; a.sharma@unsw.edu.au

Keywords: Statistical downscaling, rainfall, spatial and temporal rainfall structure, low frequency variability

EXTENDED ABSTRACT

Direct use of outputs from the General Circulation Models (GCMs) for climate change impact assessment is often limited by their incapability at representing local features and dynamics at spatial scales finer than the in-built GCM grid scale. This has led to the development of downscaling techniques for transfer of coarse GCM simulated weather output to finer spatial resolutions. However, the downscaling models are often suspected for their validity in the future climate conditions and their inability to simulate the hydrologic extremes. This paper presents a stochastic downscaling model for simulation of multi-site daily rainfall occurrences with the aim of proper simulation of rainfall extremes. At-site rainfall occurrences are modelled using a Modified Markov model (MMM) as described in Mehrotra and Sharma (2007) that defines the temporal persistence in the rainfall occurrence by updating at each time step the Markovian transition probabilities on the basis of recent past rainfall behaviour and the selected atmospheric variables. The spatial dependence across the rainfall occurrence field is maintained through spatially correlated random numbers and atmospheric and other variables defining the history of rainfall in the recent past, common across the stations. The proposed model is applied for downscaling of rainfall occurrences at a network of 45 raingauge stations around Sydney in Australia and its performance evaluated. The analyses of the results show that the scheme of updating the transition probabilities of the at-site rainfall occurrence model and the logic of providing spatial treatment separately, imparts considerable accuracy and flexibility in the representation of characteristics of interest in hydrologic studies. These characteristics include representation of spell patterns, spatial distribution, and low and higher time scale persistence characteristics and as generic indicators of water balance and variability that are of importance in a catchment scale water balance simulation.

Figure 2: Scatter plots of observed and model simulated log-odds ratios of rainfall occurrences for each station pair.

Figure 3: Scatter plots of observed and model simulated means and standard deviations of rainfall occurrences on monthly, seasonal and annuals basis at each station. SD stands for standard deviation.
1. INTRODUCTION

General Circulation Models (GCMs) are commonly used to obtain detailed climate information needed for the assessment of the various consequences of future climate changes on ecosystems and societies, both in space and time (e.g. Bergström et al. 2001, Varis et al. 2004). Inspite of the fact that the spatial resolution of the GCMs is steadily improving, most GCMs operate at large spatial scales (>10^4 km²) and provide a reasonable representation of global and continental scale processes. However, they are incapable of representing local sub-grid-scale features and dynamics (IPCC, 2001; Charles et al., 2004; Vicuna et al., 2007). As a consequence techniques have been developed to post process the GCM-output by means of downscaling in order to meet the need for detailed information at local and regional levels for use in modelling at the catchment scale. Statistical downscaling is a way to infer local information from coarse scale information by applying empirical statistical links between large scale fields and local conditions (e.g. von Storch et al. 2000, Yarnal et al. 2001). Such statistical links may be used both to validate global and regional climate models, and to develop detailed local climate scenarios based upon the output from such climate models. Statistical downscaling has been especially recommended in areas with complex topography (Kattenberg et al. 1996). A diverse range of statistical downscaling techniques have been developed over the past few years, with regression based and weather state based methods are quite popular (Wilby and Wigley, 1997; Hughes et al., 1999; Charles et al., 2004; Bartholy et al., 1995; Stehlík and Bárdossy, 2002; Mehrotra and Sharma, 2005). Yarnal et al. (2001), Charles et al. (2004), IPCC (2001) and Prudhomme et al. (2002) provide an excellent review and discussion of various downscaling techniques.

It has been noted that the weather patterns producing rainfall over a region introduce spatial and temporal dependence structure into the rainfall record. Also, persistent longer-term climate variation like the El Nino Southern Oscillation (ENSO) or similar climatic anomalies influence the low-frequency persistence structure of rainfall including the representation of sustained droughts and periods of above average rainfall (or above average wet days) in the rainfall record (Madden et al., 1999; Shea and Madden, 1990; Singh and Kripalani, 1986; Harrold et al., 2003). However, commonly available statistical downscaling approaches are limited in their representation of the spatio-temporal structure of the downscaled rainfall field. The common logic used for downsampling of the rainfall occurrence field, while adequate at representing spatial attributes, is often found wanting at representation of temporal dependences, specifically of low frequency rainfall variability, of great importance in the simulation of hydrologic extremes (floods, droughts) of interest (Bartholy et al., 1995; Stehlík and Bárdossy, 2002; Charles et al., 1999). It has been shown that the regression methods and some weather-typing approaches under-predict climate variability to varying degrees, since only part of the regional and local climate variability is related to large scale climate variations. For example, Conway et al. (1996) compared two downscaling approaches and found that mean daily precipitation probabilities, wet-day amounts and persistence were well represented, but variations in the interannual rainfall totals were not modelled to the same standard at either of the two reference sites.

Another important and perhaps the more serious limitation of the application of the downscaling approaches to future climate scenario generation, is the fact that the relationships between the rainfall and associated meteorological properties over a region are seldom constant in time. Wilby et al. (1995) partly attributed this variability to the subtle changes in the dominant precipitation mechanism (stratiform or convective with reference to UK precipitation), whereas Sweeney and O’Hare (1992) have attributed this to the role of changes in the intensity of circulation development, and/or shifts in depression trajectories.

This paper presents a rainfall occurrence downscaling model that tries to address these limitations to a greater extent. The model is termed as Modified Markov Model (MMM) (Mehrotra and Sharma, 2007). It conditionally simulates the rainfall occurrence field based on exogenous atmospheric forcings and aggregated longer time scale variables that represent the low frequency variability of rainfall within the commonly used low order Markov dependence structure. The MMM is nonstationary or dynamic in nature in a sense that at-site Markovian transition probabilities of the model are modified at each time step to accommodate the variations in the circulation variables over the region including the trend of the recent downscaled rainfall series. Spatial correlations of downscaled rainfall occurrence field are maintained by making use of random innovations that are spatially correlated yet serially independent in nature (Wilks, 1998).

2. METHODOLOGY

In the discussions that follow, all multivariable vectors or matrices are expressed as bold and
single variables or parameters using non-bold characters or symbols. We denote rainfall occurrence at a location \( k \) and time \( t \) as \( R_t(k) \) and at the \( p \)th time step before the current as \( R_{t-p}(k) \).

Also, a \( n_s \)-site rainfall vector at time \( t \) is denoted as \( \mathbf{R}_t \), a vector of atmospheric predictor variables at time \( t \) as \( \mathbf{A}_t \), and a vector of predictor variables (consisting of atmospheric and/or other relevant indicators) as \( \mathbf{Z}_t \). Also, the areal averaged rainfall at time \( t \) is expressed as \( \mathbf{R}_t = \frac{1}{n_s} \sum_{i=1}^{n_s} R_{i,t} \).

\[
P(R_t = 1, R_{t-1} = i) = \frac{P(R_t = 1, R_{t-1} = i)}{P(R_{t-1} = i)} f(\mathbf{X}_t | R_t = 1, R_{t-1} = i) = \frac{f(\mathbf{X}_t | R_t = 1, R_{t-1} = i) + f(\mathbf{X}_t | R_t = 0, R_{t-1} = i)P(R_t = 0 | R_{t-1} = i)}{P(R_{t-1} = i)}
\]

In general, the rainfall downscaling problem could be expressed as the conditional simulation of \( \mathbf{R}_t(k) \) given \( \mathbf{Z}_t(k) \) where \( \mathbf{Z}_t(k) \) represents a vector of variables at a location \( k \) and at time \( t \) that in addition to previous time steps values of rainfall imparting daily or short term persistence, also includes atmospheric variables and/or other continuous variables (\( \mathbf{X}_t(k) \)) explaining the higher time scale persistence. In the simplest case of a first order Markov model for rainfall generation, \( \mathbf{Z}_t(k) \) contains \( R_{t-1}(k) \) only.

### 2.1. Modelling at-site temporal persistence

In the following discussions, the details of a method that aims to formulate a generic representation of the conditional simulation of \( \mathbf{R}_t(k) \) given \( \mathbf{Z}_t(k) \) within the framework of Markov process by considering order-one short-term dependence are presented. It may however be noted that the model structure presented can easily be extended to include higher order Markovian dependence. For brevity, site notations are dropped in the subsequent discussions. The parameters (or the transition probabilities) of a model expressing the order one Markovian dependence (first order Markov model) are defined by \( P(R_t | R_{t-1}) \) with \( \mathbf{Z}_t \) consisting of \( R_{t-1} \) only. Inclusion of additional predictors \( \mathbf{X}_t \) in the conditioning vector \( \mathbf{Z}_t \) would modify these transition probabilities as \( P(R_t | R_{t-1}, \mathbf{X}_t) \). The following parameterization is adopted to estimate \( P(R_t | R_{t-1}, \mathbf{X}_t) \):

The first term of (1) defines the transition probabilities \( P(R_t | R_{t-1}) \) of a first order Markov model (representing order one dependence) while the second term signifies the effect of inclusion of predictor set \( \mathbf{X}_t \) in the conditioning vector \( \mathbf{Z}_t \). If \( \mathbf{X}_t \) consists of derived measures (typically linear combinations) of either atmospheric variables or summation of number of wet days in pre-specified aggregation time periods, one could approximate the associated conditional probability \( f(\mathbf{X}_t | R_t = 1, R_{t-1} = i) \) as a multivariate normal distribution. One can consequently derive the conditional probability \( f(\mathbf{X}_t | R_{t-1} = i) \) as a mixture of multivariate normals as specified in equation (1). This leads to the following simplification for \( P(R_t | R_{t-1}, \mathbf{X}_t) \):

\[
\frac{1}{\det(\mathbf{V}_{1,t})^{1/2}} \exp \left\{ \frac{1}{2} (\mathbf{X}_t - \mu_{1,t})^T \mathbf{V}_{1,t}^{-1} (\mathbf{X}_t - \mu_{1,t}) \right\}
\]

Here, the \( \mu_{1,t} \) parameters represent the mean \( E(\mathbf{X}_t | R_t = 1, R_{t-1} = i) \) and \( \mathbf{V}_{1,t} \) is the corresponding variance-covariance matrix. Similarly, \( \mu_{0,t} \) and \( \mathbf{V}_{0,t} \), represent, respectively, the mean vector and the variance-covariance matrix of \( \mathbf{X}_t \) when \( (R_{t-1} = i) \) and \( (R_t = 0) \). The \( p_{ij} \) parameters represent the baseline transition probabilities of the first order Markov model defined by \( P(R_t = 1 | R_{t-1} = i) \) and \( \det() \) represents the determinant operation. Please note that for some applications, the assumption of a multivariate normal may not be sufficient.
In the present application we consider vector $X_t$ as consisting of the aggregated wetness state predictors, $B_t$, over 30 and 365 days and the selected atmospheric variables, $A_t$. Procedure for selection of atmospheric predictor variables is described in the sub-section 3.4.

2.2. Modelling spatial correlations

The rainfall occurrence downscaling model outlined in previous section simulates rainfall at individual stations in isolation, hence resulting in values that are theoretically spatially independent. We induce spatial dependence in the downscaled rainfall occurrences by making use of spatially correlated and serially independent random numbers during generation of rainfall occurrences at individual stations separately. The general logic of estimating the correlation matrices of random numbers for rainfall occurrences is available in Wilks (1998) and Mehrotra et al. (2006).

3. DATASETS AND STUDY AREA

3.1. Study area

The study region is located around Sydney, eastern Australia spanning between 147°E - 153°E longitude and 31°S - 36°S latitude (Figure 1). The physio-geographical conditions in Sydney region cause large climatic gradients even over short distances, e.g. from lowland areas to mountain regions and from the coast to the inland.

3.2. Rainfall

For this study, 43-year continuous record (from 1960 to 2002) of daily rainfall at 45 stations around Sydney, eastern Australia is provided directly by the Sydney Catchment Authority (SCA), Sydney, Australia (Figure 1). The available rainfall amount record is converted into rainfall occurrences (zero or one) by considering a day as wet (one) if observed rainfall on that day is greater than or equal to 0.3 mm, otherwise dry (zero). The dense network of climate stations with sufficient daily records of 43 years or more provides a good base for the development of empirical downscaling models such as MMM.

3.3. Large scale atmospheric variables

The required information about atmospheric variables over 25 grid points covering the study area, is extracted from the National Center for Environmental Prediction (NCEP) reanalysis data provided by the NOAA-CIRES Climate Diagnostics Centre, Boulder, Colorado, USA, from their web site at http://www.cdc.noaa.gov/. These variables are available on 2.5° latitude x 2.5° longitude grids over the study region, on a daily basis for the same period as the rainfall record. As an observed rainfall value represents the total rainfall over a 24-h period ending at 0900 hours (local time, LT) in the morning, the available atmospheric measurements on the preceding day are considered as representative of today’s rainfall.

3.4. Identification of significant predictors

Sea level pressure (SLP) fields, geo-potential heights, air temperatures, humidity, wind speeds or indices derived from these variables (e.g. air flow indices such as zonal and meridional wind and vorticity, vertical and horizontal gradients and thickness of pressure fields) have been demonstrated to account for a large part of the variation and trends in local precipitation (Harpham and Wilby, 2005; Buishand et al., 2004; Charles et al., 1999). Based on the results of these studies, we picked up a large set of atmospheric predictors comprising of circulation and moisture variables at various pressure levels and their horizontal and vertical gradients as the potential predictors (totalling 45 predictors). The predictor identification exercise is carried out at daily time step for each season (MAM, JJA, SON, DJJ). To facilitate the predictor identification exercise, we consider solo predictand as daily area averaged wetness fraction for rainfall occurrence. As some of the predictors might be highly correlated among themselves, initial screening is carried out using linear regression to exclude the highly correlated predictors (having linear correlation of greater than...
0.90). Finally, a nonparametric stepwise correlation analysis based on partial mutual information (Sharma, 2000) is carried out to identify sets of significant atmospheric predictors for each season. Based on this exercise finally, four atmospheric predictors for each season are identified as significant predictors for this region with at least one predictor in each season representing atmospheric moisture.

4. RESULTS

In all the results that follow, the statistics reported are ascertained by simulating 100 realisations of the rainfall occurrences from the MMM. The performance of the downscaling model is evaluated on a daily, monthly, seasonal and annual basis for its ability to simulate the observed spatial and temporal characteristics of rainfall including those of importance in water resource management. For a few statistics, the at-site time distribution plots are presented for two representative stations namely, station 19 located inland and representing a drier region, and station 38 located in the coastal area representing a wetter region (Figure 1).

4.1. Spatial correlations

The log-odds ratio, reflecting the spatial correlation between rainfall occurrences at each pair of stations provides a measure of accurate reproduction of the overall wet and dry days between the station pair (Mehrotra and Sharma, 2005). Figure 2 presents this statistic at all stations. The model accurately reproduces the dependence between the stations. Also, the area averaged wetness fraction on daily and higher time scales, has important implications in calculating the flows from a catchment. Table 1 provides the observed and model simulated means and standard deviations details of area averaged wetness fractions at varying time scales. The model accurately simulates these statistics.

4.2. Number of wet days

It is vital that the average number of wet days or wet day probabilities and day to day occurrence of the rainfall at raingauge network be reproduced accurately before using the downscaled rainfall series as an input to any water balance modeling exercise. The first column of Figure 3 presents scatter plots of observed and modelled number of wet days at all stations on monthly, seasonal and annual basis. As can be seen from the graph, the model provides a good fit to the number of wet days at all stations. However, for efficient design and management of water resource projects, not only the number of wet days but their accurate distribution and variations in the downscaled series are also important. Second column of Figure 3 compares the standard deviation of aggregated number of wet days on monthly, seasonal and annual time scales whereas Figure 4 provides the distribution plots of wet days in the observed and downscaled rainfall series for two representative stations. The 5th percentile, median, and 95th percentile values are shown as continuous lines while the historical values are superimposed as circles. As shown in both the plots, the model adequately reproduces the averages, distribution and variation of wet days at varying time scales barring a small underestimation of standard deviation of annual wet days.

Table 1: Mean and standard deviation of area averaged wetness fraction at varying timescales.

<table>
<thead>
<tr>
<th>Time Scale</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Simulated</td>
</tr>
<tr>
<td>Daily</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td>Monthly</td>
<td>9.32</td>
<td>9.67</td>
</tr>
<tr>
<td>Annual</td>
<td>111.86</td>
<td>114.02</td>
</tr>
</tbody>
</table>

4.3. Wet and dry spell characteristics

Sustained periods of wet and dry spells form the basis of reservoir design and operation, and agricultural studies. Therefore, it is vital that wet and dry spells and their distribution be accurately represented by any downscaling approach. For the obvious reason, we choose here to present the results of station 19 (representing a drier region) for dry spells and station 38 (representing wetter
region) for wet spells (Results for other stations are available with authors). The two plots of the top row of Figure 5 compare the maximum wet and dry spells at all stations. The second row compares the distribution of annual maximum dry spells (for station 19) while the bottom row shows the distribution of maximum wet spells for station 38. As can be seen from these graphs, maximum wet and dry spells and their distributions are reproduced well by the model with the exception of slight over estimation of the maximum wet spells at majority of stations.

Figure 5: Maximum wet and dry spells and their distribution at representative stations.

5. CONCLUSION

This paper has demonstrated the applicability of a relatively simple stochastic downscaling framework for multi-site rainfall. The approach down scales rainfall occurrences at all stations using a modified Markov model (MMM) with spatially dependent forcing of uniform random numbers. Downscaling models having the capability to simulate rainfall at a network of stations whilst maintaining the appropriate spatial dependence attributes are best suited for use in catchment management practice, where the nature of spatial variations in rainfall has important influences on streamflow and flooding. Also, important temporal attributes of rainfall like distribution of wet and dry spells, number of wet days at individual stations have a significant impact in crop simulation studies and drought management applications. Such spatio-temporal rainfall attributes assume even more importance when the downscaling procedure is applied for investigating possible changes that might be experienced by hydrological, agricultural and ecological systems in future climates. The results of the MMM downscaling model indicate that the model reproduces fairly well the desired spatio-temporal statistics of the observed rainfall record at all sites and can be used to investigate the possible changes in the rainfall in the warmer climate.

6. REFERENCES


