Journal of Hydrology (2007) **345**, 102–120

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

Preserving low-frequency variability in generated daily rainfall sequences

R. Mehrotra *, Ashish Sharma

The University of New South Wales, Sydney, Australia

Received 25 April 2007; received in revised form 17 July 2007; accepted 1 August 2007

KEYWORDS
Markov chain;
Rainfall;
Stochastic weather generator;
Aggregated time scale rainfall characteristics;
Low-frequency rainfall characteristics;
Spatial correlation structure;
Kernel density estimation

Summary A stochastic modeling framework for multisite generation of daily rainfall is developed with an aim of representing both short and higher time scale dependence in the generated rainfall sequences. The framework simulates rainfall at individual locations using separate models for rainfall occurrences and rainfall amounts on the simulated wet days. The spatial correlations in the generated occurrences and amounts are induced using spatially correlated yet serially independent random numbers. The rainfall occurrence model is based on a modification of the transition probabilities of the traditional Markov model through an analytically derived factor that represents the influence of rainfall aggregated over long time periods in an attempt to incorporate low-frequency variability in simulations. The rainfall amounts on the wet days are generated using a nonparametric conditional simulation approach. The utility of the proposed method is illustrated by applying the model on a network of 30 raingauge stations around Sydney, Australia, and comparing a range of statistics describing daily and higher time scale distribution and dependence attributes. The analyses of the results show that the method adequately captures daily as well as aggregated higher time scale rainfall characteristics at individual locations including the spatial distribution of rainfall over the region.

Introduction

Stochastic modeling of daily time series of hydrologic variables has been an active topic of research. Stochastic models for variables such as daily rainfall and temperature, sometimes also known as 'weather generators' are commonly used to generate synthetic sequences that are statistically consistent with the historical record. Such generators fall into two categories. The first category consists of models that treat rainfall as composed of two random processes – rainfall occurrence (denoting whether a day is wet or dry) and rainfall amount on a wet day. Rainfall occurrence is modeled as a Markov process, the state (wet or dry) of the current and a few prior days deciding the state of the day that follows (Katz, 1977; Todorovic and Woolhiser,

* Corresponding author. Tel.: +61 293855140; fax: +61 293856139.
E-mail address: raj@civeng.unsw.edu.au (R. Mehrotra).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2007.08.003

1975; Katz and Parlange, 1993; Stern and Coe, 1984), while amounts on wet days are modeled through appropriately specified probability distributions. The second category consists of approaches that represent the length of the current dry or wet spell as a random variable and model it as an alternating renewal process (Buishand, 1978; Foufoula-Georgiou and Lettenmaier, 1987; Woolhiser and Roldan, 1982). While the above two alternatives have traditionally been modeled using parametric probability distributions to characterize the conditional or transition probabilities needed, there has been a surge in nonparametric equivalents that use conditional resampling and simulation (Lall et al., 1996; Rajagopalan et al., 1996; Brandsma and Buishand, 1998; Rajagopalan and Lall, 1999; Sharma and Lall, 1999; Buishand and Brandsma, 2001; Sharma and O'Neill, 2002; Harrold et al., 2003a; Harrold et al., 2003b) as the basis for generating rainfall occurrences and wet day precipitation amounts.

The assumption of Markovian dependence is the basis for all modeling alternatives mentioned above. It has been observed that assuming a low-order Markov dependence, in general, undersimulates long dry spells (runs of consecutive dry days) (Buishand, 1978; Guttorp, 1995; Racsko et al., 1991; Semenov and Porter, 1995; Wilks, 1999) and the variance at aggregated monthly, seasonal and annual time scales, an effect termed as ''overdispersion'' (Katz and Parlange, 1998). These limitations have implications for applications that are sensitive to proper representation of low-frequency persistence, or, the representation of sustained droughts and periods of above average rainfall (or above average wet days) in the generated record.

In this paper, we present a new approach that aims to incorporate low-frequency variability in the generated sequences. The approach models rainfall as consisting of a two-stage process (occurrence and amount), where the occurrence is assumed to follow a Markov process, structured so as to represent low-frequency variability through a ''wetness indicator'' that is based on aggregated rainfall over past time periods to inform on the overall ''wetness'' the sequence has exhibited prior to the current time step. Alternate model structures, formed using combinations of wetness indicators representing the wetness over varying periods of aggregation and the order of Markov process, are investigated with the aim of improving rainfall occurrence attributes deemed important in stochastic generation. Rainfall amounts are generated using a nonparametric stochastic model. Seasonal variations in occurrence and amount generation processes are accommodated using a moving window approach. The model structure found most suitable is then used to generate rainfall amounts at 30 rainfall locations in a region near Sydney, Australia and results presented.

The paper is organised as follows. In the next section, a brief discussion on the representation of longer time scale variability in rainfall is presented. Section ''Methodolgy'' covers details of the methodology we propose while Section ''Application of occurence and amounts models'' presents the details on selection of wetness indicators, application of models and the study area considered. This is followed by presentation and comparison of results. Conclusions from the study are drawn in the last section.

Representation of low-frequency variability in rainfall

As noted in the introduction, the commonly used low-order Markov models for daily rainfall undersimulate very long dry spells and variances of aggregated monthly, seasonal and annual wet days and rainfall amounts. Extended versions of the binary Markov chains using a higher order of dependence may improve performance (Pegram, 1980; Stern and Coe, 1984; Katz and Parlange, 1998).

It may be noted that the number of parameters required to characterize two-state Markov chains increases exponentially with the order of the processes and in general is 2^k for kth order dependence. To reduce the number of parameters of the higher order Markov models of rainfall, Pegram (1980) devised a Markov chain modeling structure which allows the specification of a lag-k dependence structure with $k+1$ parameters rather than 2^k. Similarly, Stern and Coe (1984) have suggested Markov chains of hybrid order requiring only $k+1$ rather than 2^k parameters. Other alternative replaces the sequence of individual states of these time lags by conditional probabilities based on aggregation of states of previous time steps (Sharma and O'Neill, 2002); a set of wetness indices based on previous days wetness state (Harrold et al., 2003a; Mehrotra and Sharma, 2007); discrete states of other variables governing rainfall such as indices of atmospheric circulation (Katz and Parlange, 1993) and; as a function of the covariates (Hughes and Guttorp, 1994; Hughes et al., 1999; Mehrotra et al., 2004).

It has been suggested that day to day and lower-frequency variations in the rainfall dominate the observed interannual variability and use of more complex occurrence and amount models is of limited help in improving the unexplained higher time scale variance (Katz and Parlange, 1998; Wilks, 1999). Other researchers have viewed this unexplained variance as the portion of the interannual variance associated with climatic nonstationarity, or longer-term variations (Madden et al., 1999; Shea and Madden, 1990; Singh and Kripalani, 1986; Harrold et al., 2003a). Another possible explanation for the underestimation of variance is that most of the available models are not sufficiently nonlinear. The variability of a nonstationary climate can not be effectively reproduced by the simple stationary models of daily rainfall under the assumption of a linear relationship, and incorporation of some mechanism into rainfall generation process explaining the underlying climatic variations and nonlinearity is necessary to improve the underestimation of the variance of aggregated series and extended wet and dry spells (Katz and Parlange, 1993, 1998). Following this, the approaches allowing the variations in the stochastic model parameters by conditioning on exogenous variable(s) (variables that cannot be formed from the generated rainfall values alone such as large-scale atmospheric variables) (Hughes and Guttorp, 1994; Hughes et al., 1999; Mehrotra et al., 2004; Katz and Parlange, 1993; Katz and Zheng, 1999; Wallis and Griffiths, 1997; Wilks, 1989) are expected to explain better the observed interannual variability in the generated series. It may, however, be noted that as these approaches consider