Available online at www.sciencedirect.com

ScienceDirect

Journal of Hydro-environment Research 9 (2015) 604–613

Journal of Hydro-environment Research

www.elsevier.com/locate/jher

Research paper

Copula-based modeling and stochastic simulation of seasonal intermittent streamflows for arid regions

Changsam Jeong [a], Taesam Lee [b,*]

[a] Dept. of Civil and Environmental Eng., Induk University, Wolgye 2-dong, Nowon-gu, Seoul 139-052, South Korea

[b] Dept. of Civil Engr., ERI, Gyeongsang National University, 501 Jinju-daero, Jinju, Gyeongnam 660-701, South Korea

Received 17 April 2013; revised 28 June 2014; accepted 30 June 2014

Abstract

Streamflow is often intermittent in arid and semi-arid regions. Stochastically simulated data play a key role in managing water resources with intermittent streamflows. The stochastic modeling of intermittent streamflow that incorporates the seasonality of key statistics is a difficult task. In the current study, the product model was tested to simulate the intermittent monthly streamflow by employing the periodic Markov chain (PMC) model for occurrence and the periodic gamma autoregressive (PGAR) and copula models for amount. The copula models were tested in a previous study for the simulation of yearly streamflow, resulting in successful replication of the key and operational statistics of historical data; however, the copula models have never been tested on a monthly time scale. The intermittent models were applied to the Colorado River system in the present study. A few drawbacks of the PGAR model were identified, such as significant underestimation of minimum values on an aggregated yearly time scale and restrictions of the parameter boundaries. Conversely, the copula models do not present such drawbacks but show feasible reproduction of key and operational statistics. We concluded that the copula models combined with the PMC model is a feasible method for the simulation of intermittent monthly streamflow time series.

Keywords: Copula; Drought; Intermittency; Periodic Markov chain; Seasonal streamflow; Stochastic simulation

1. Introduction

Synthetic data obtained from stochastic models play a key role in analyzing extreme events, such as droughts, and in evaluating alternative designs and operating rules of hydraulic structures, especially in arid or semi-arid regions (Lee and Salas, 2006; Salas, 1993; Salas and Abdelmohsen, 1993; Salas et al., 2006; Stedinger et al., 1983; Stedinger and Tasker, 1985). Many alternatives have been developed, which originated from a simple autoregressive model (Koutsoyiannis, 1994; Kwon et al., 2007; Lall, 1995; Lall et al., 1996; Lee and Ouarda, 2010; Lee and Salas, 2011; Lee et al., 2010; Ouarda et al., 1997; Salas and Abdelmohsen, 1993; Salas and Boes, 1980; Salas and Lee, 2010; Srinivas and Srinivasan, 2001, 2006; Stedinger and Vogel, 1984; Sveinsson et al., 2003).

A few characteristics remain difficult to reproduce, such as long-term persistency and intermittency. Long-term persistency (Lee and Ouarda, 2010; Sharma et al., 1998; Young and Holt, 2007) implies the auto-dependency structure of more than a monthly time scale (e.g., yearly). Intermittency (Koutsoyiannis, 2006) is identified by one or more zero values between non-zero values in a time series when values represent events or amounts of streamflows or rainfall. For example, in the Colorado River system, the monthly datasets of some tributary stations display intermittency.

Few effective models have been developed for generating monthly streamflow data with intermittency because of structural difficulties inherent in modeling seasonality and intermittency. Chebaane et al. (1995) applied the product model, which comprises both amount and occurrence. The occurrence process has been described using the periodic discrete autoregressive (PDAR) model; the amount process has been described using the periodic autoregressive moving average (PARMA) model. The occurrence process must be modeled to preserve the statistical behavior of historical records. For example, the occurrence probability of a certain month in historical records should be reproduced by the simulated data.

We hypothesized that the physical process presented by the PARMA model is periodically stationary and that the marginal noise processes are normally distributed. Because monthly streamflow data are not normally distributed, the data needs to be transformed. The process of fitting the model to the transformed data and then back-transforming the data results creates

* Corresponding author. Tel.: +82 55 772 1797; fax: +82 55 772 1799.
E-mail address: tae3lee@gnu.ac.kr (T. Lee).

http://dx.doi.org/10.1016/j.jher.2014.06.001
1570-6443/

bias toward the statistics simulated to describe the original space. Instead of this transformation, a model with a skewed distribution (i.e., gamma) was developed with preserving lag-1 serial correlation on a seasonal time series (Fernandez and Salas, 1986), namely the periodic gamma autoregressive (PGAR) model. However, the PGAR model requires a very complicated procedure to simulate the sequences, and its parameter space is limited. Meanwhile, Lee and Salas (2011) developed the copula-based stochastic model and applied it to a yearly time series. The results showed that the copula-based stochastic simulation model easily captures the key statistics of historical data and does not require any transformation. Furthermore, a number of applications for copula have been popularly reported in literature, including drought analysis (Shiau et al., 2007; Song and Singh, 2010; Wong et al., 2010) and flood frequency (Favre et al., 2004; Kao and Govindaraju, 2008; Zhang and Singh, 2006).

Therefore, we applied the copula-based model to monthly time scale data. In the current study, two models for the amount process (the PGAR and copula models) and one model for the occurrence process (the PDAR model) were tested to simulate intermittent monthly streamflow. The pros and cons of the models were inspected, and the model was applied to the Colorado River system.

2. Proposed modeling approach

To model an intermittent streamflow time series, the product of occurrence and amount is denoted as follows:

$$Y_{v,\tau} = X_{v,\tau} Z_{v,\tau} \tag{1}$$

where $v = 1,2,\ldots,N$ and $\tau = 1,2,\ldots,\omega$, representing years and seasons, respectively; N and ω are the numbers of years and seasons, respectively. Note that when the data time scale is monthly, $\omega = 12$. $X_{v,\tau}$ denotes the binary (0 or 1) occurrence process; $Z_{v,\tau}$ denotes the amount process; and $Y_{v,\tau}$ is the product of the two processes. The traditional PDAR(1) model was applied to simulate the occurrence process, $X_{v,\tau}$. The PGAR model was applied to simulate the amount process. The Bivariate Normal Copula with Gamma marginal distribution (BNCG) was also used to simulate the amount process.

2.1. Occurrence process

2.1.1. Order-1 periodic discrete autoregressive (PDAR(1)) model

The PDAR(1) is defined as follows:

$$X_{v,\tau} = V_{v,\tau} X_{v,\tau-1} + (1 - V_{v,\tau}) W_{v,\tau} \tag{2}$$

where $X_{v,\tau}$ is a periodic dependent Bernoulli process; and $W_{v,\tau}$ and $V_{v,\tau}$ are independent Bernoulli processes with probabilities $P[V_{v,\tau} = 1] = \gamma_\tau$ and $P[W_{v,\tau} = 1] = \delta_\tau$, respectively. Chebaane et al. (1995) showed that the PDAR(1) model is equivalent to a periodic Markov chain (PMC) model, in which the elements of the transition probability matrix vary by season as

$$p_\tau(i,j) = P[X_{v,\tau} = j \mid X_{v,\tau-1} = i] \tag{3}$$

where $i, j = 0$ or 1.

The transition probability matrix is expressed as a function of the parameters of PDAR(1) such that

$$\begin{bmatrix} p_\tau(0,0) & p_\tau(0,1) \\ p_\tau(1,0) & p_\tau(1,1) \end{bmatrix} = \begin{bmatrix} \gamma_\tau + (1-\gamma_\tau)(1-\delta_\tau) & (1-\gamma_\tau)\delta_\tau \\ (1-\gamma_\tau)(1-\delta_\tau) & \gamma_\tau + (1-\gamma_\tau)\delta_\tau \end{bmatrix} \tag{4}$$

The limiting distribution $P[X_{v,\tau} = 1] = \mu_\tau$ is given by

$$\mu_\tau = \gamma_\tau \mu_{\tau-1} + (1-\gamma_\tau)\delta_\tau \tag{5}$$

The parameters of PMC are estimated by

$$\hat{p}_\tau(i,j) = \frac{n_\tau(i,j)}{n_\tau(i)} \tag{6}$$

where $n_\tau(i,j)$ is the number of times that the variable $X_{v,\tau}$ in state i at time $\tau - 1$ passes to state j during the period τ; and $n_\tau(i) = n_\tau(i,0) + n_\tau(i,1)$ is the number of times that $X_{v,\tau}$ is in state i at period τ. γ_τ and δ_τ are easily estimated from the relationship outlined in Eqs. (4) and (6).

2.2. Amount process

Two models were applied to simulate the amount process of the product model in the current study: (1) PGAR and (2) bivariate normal copula. These models are not based on Gaussian marginal distributions. Instead, PGAR assumes that the marginal distribution is gamma, while the bivariate normal copula model can have any feasible distribution for a marginal distribution. In practical applications, a gamma marginal distribution is applied for the bivariate normal copula even if there is no prior limitation for selecting marginal distribution because (1) the gamma distribution is one of the most frequently selected distribution types for hydrological frequency analysis; (2) it is comparable to the PGAR model for the marginal distribution; and (3) it is representative of the positively skewed distributions that are typical in historical streamflow data.

For periodic models, Fourier series analysis may be used to model the periodic patterns of certain statistics. Fourier series analysis has been a popular method for use in stochastic models in which the time scale is relatively small, such as weeks and days. Katz and Parlange (1995) used this procedure to capture the diurnal cycle in hourly rainfall data. In the present study, the Fourier series was applied to smooth out periodic patterns, especially the skewness and lag-1 correlation, which have high degrees of uncertainty in parameter estimations.

2.2.1. Periodic gamma autoregressive (PGAR)

The gamma distribution is denoted as

$$f_z = \frac{1}{\alpha\Gamma(\beta)} \left(\frac{z-\lambda}{\alpha}\right)^{\beta-1} \exp\left(-\frac{z-\lambda}{\alpha}\right) \tag{7}$$

where $\alpha > 0$, $\beta > 0$ and $z > \lambda$. If $\lambda = 0$, then Eq. (7) presents a two-parameter gamma distribution represented as $Z \sim \text{gamma}(\alpha, \beta \text{ and } \lambda)$ or $\text{gamma}(\alpha, \beta)$. The relationships between the distribution parameters and the key statistics are

$$\beta = \left(\frac{\mu}{\sigma}\right)^2 \text{ and } \alpha = \frac{\mu}{\beta} \text{ for 2-gamma distribution} \tag{8}$$