Journal of Hydrology 511 (2014) 904–919

Contents lists available at ScienceDirect

Journal of Hydrology

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

The examination of reproducibility in hydro-ecological characteristics by daily synthetic flow models

Gene Jiing-Yun You *, Bing-Hao Thum, Fu-Hua Lin

Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan

ARTICLE INFO

Article history:
Received 22 October 2013
Received in revised form 6 February 2014
Accepted 12 February 2014
Available online 24 February 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Peter F. Rasmussen, Associate Editor

Keywords:
Synthetic flow model
Indicators of Hydrologic Alteration (IHA)
Shot noise
Markov-based
Flow duration curve
k-Nearest neighbor (k-NN)

SUMMARY

Research into synthetic streamflow generation previously focused primarily on engineering purposes, while disregarding hydro-ecological implications with regard to daily flow. This study investigated the applicability of daily synthetic flow models in the representation of hydro-ecological characteristics. This study applied flow duration curves and three performance indices as well as a method of statistical testing to identify whether Indicators of Hydrologic Alteration (IHA) derived from historical and synthetic flow data have the same distribution. This enabled us to evaluate the capacity of the synthetic flow models in capturing other important hydrological characteristics. This study examined various representative methods, most of which proved effective in simulating the magnitude of monthly flow and high flow within short durations. However, low flow conditions remain problematic, particularly over short durations. Among the various models, annual flow simulation using annual–daily disaggregation demonstrated good reproducibility, but its applicability with regard to within year fluctuations may be limited by the relatively low degree of freedom. Direct simulation models (shot noise and Markov-based) proved suitable in dealing with engineering problems based on monthly averages and annual daily maximum flow. Linear regression models, PARMA, and nonparametric models, modified *k*-NN, were shown to be well-suited to the representation of monthly stream flow. In conjunction with nonparametric bootstrap disaggregation, these methods proved effective in representing daily flow patterns. Coupling these methods with the shot noise disaggregation model was the least recommended due to bias resulting from numerical errors associated with repeated convolutions.

1. Introduction

Reliable hydrologic records are fundamental to the planning and utilization of water resources; however, records of sufficient length are seldom available. As a result, synthetic streamflow models are commonly used to generate representative data. Retaining the statistical characteristics of historical flow time series make it possible for synthetic hydrologic sequences to reproduce hydrologic features and generate variability sufficient to meet the needs of water resource engineers. Conventional synthetic hydrologic sequences are generated using linear and parametric models, such as autoregressive moving average models (ARMA) and lag-1 periodic autoregression models (PAR) (Salas, 1985). These stochastic hydrological models provide hydrological sequence simulations over relatively long timescales, such as monthly, seasonally or annually; however, they tend to be somewhat limited in the simulation of daily discharge.

Beard (1967) and Green (1973) were the first to attempt daily synthetic flow simulation. Beard used a second-order Markov model and Log-Pearson Type III distribution to generate daily hydrological sequences; however, these efforts proved insufficient in the reproduction of storm hydrographs and baseflow recession. Green developed a method based on linear interpolation using five-day averages for the synthesis of daily flow; however, this approach failed to provide sufficient accuracy for the simulation of high-flow conditions. Both of these approaches were only able to reproduce a limited number of hydrological characteristics of discharge (Treiber and Plate, 1977). Weiss (1977) took into account the characteristics of rainfall runoff and adopted a filtered Poisson method to develop a shot noise (SN) model for synthetic flow simulation. That study provided a comprehensive discussion of the fundamental concepts underlying the model as well as the corresponding processes for the calibration of various parameters. A disaggregation model based on the same conceptual approach has also been proposed. Treiber and Plate (1977) furthered the shot noise model by utilizing pulse processes and system transform functions to simulate runoff associated with daily synthetic

* Corresponding author. Tel./fax: +886 2 33664238.
E-mail address: genejyu@ntu.edu.tw (G.J.-Y. You).

http://dx.doi.org/10.1016/j.jhydrol.2014.02.047
0022-1694/© 2014 Published by Elsevier B.V.

streamflow. They employed a Markov chain and autoregressive processes to produce correlated pulses. Variations on this approach using various techniques for the estimation of pulse and transform function parameters were also developed by Hino and Hasebe (1981, 1984), Battaglia (1986), and Wang and Vandewiele (1994).

Schneider and Schultz (1982) and Kron et al. (1990) modified Treiber and Plate's (1977) approach and applied it to various regions. Subsequent researchers have focused on overcoming the limitations of the Treiber and Plate model and shot noise models to extend their applicability (Claps and Murrone, 1994; Claps et al., 1993; Murrone et al., 1997). More recently, Claps et al. (2005) discovered that using discharge increment pulses (DIP) to derive corresponding response functions can lead to the overestimation of flow events, with small pulses generating a considerably greater number of events than actually occur. Thus, the filtered peaks over threshold (FPOT) approach was proposed to filter pulses, such that only significant peak values are retained. A case study conducted on seven catchment areas in northern Italy demonstrated the superior simulation results of the FPOT method.

Another daily flow model utilizes Markov models to determine the probability of excess rainfall (wet/dry), prior to estimating the magnitude of runoff in accordance with hydrologic and statistical characteristics (Xu et al., 1997, 2002, 2003). Aksoy and Bayazit (2000a, 2000b) applied second-state (2000a) and third-state (2000b) Markov models to conduct similar hydrologic synthetic flow simulations. Claps et al. (2005) compared these Markov-based methods with a shot noise model according to their complexity and the number of associated parameters, concluding that shot noise models benefit from model parsimony when taking both the quality and efficiency of the simulations into account. However, all of these models are based primarily on parametric methods, which are affected by assumptions related to probability distribution, thereby leading to inconsistencies between simulation and historical data.

In contrast, non-parametric methods do not require assumptions related to the distribution of simulated flow, making them more flexible in their applicability. A number of studies have investigated non-parametric flow synthesis (Yakowitz, 1973, 1979, 1985, 1993; Schuster and Yakowitz, 1979; Karlsson and Yakowitz, 1987a,b; Smith, 1991; Smith et al., 1992). The bootstrap approach is a widely accepted non-parametric technique. Lall and Sharma (1996) developed a monthly flow generation model based on a *k*-nearest neighbor (*k*-NN) bootstrap method using historical data. In their model, the *k*-NN method takes the events of the previous period to obtain the k number of historical data nearest the simulation values. By assigning a weighted factor with respect to distance, the model resamples from the data set of the k number to formulate events in the current period. Lall and Sharma (1996) demonstrated that the *k*-NN approach outperformed other parametric methods in fitting the trends and peak values of monthly streamflow. Variations of the *k*-NN approach have also been applied to the simulation of univariate and multivariate weather data, such as rainfall, radiation, wind speed, dew point, humidity, and daily maximum and minimum temperatures (Rajagopalan and Lall, 1999; Yates et al., 2003; Harrold et al., 2003; Prairie et al., 2007).

The bootstrap approach cannot be used to simulate observed values that have not previously occurred, which severely limits its capacity to reproduce variability in data. Prairie et al. (2005, 2006) improved on the original method with the development of the modified *k*-NN approach, which uses residuals as sampling standards to enhance the variability of simulated data. The applicability of the modified *k*-NN method has been verified in a number of subsequent studies (Grantz et al., 2005; Singhrattna et al., 2005).

Another approach to daily streamflow simulation involves simulating total flow over longer time steps, such as monthly or annually (ex. ARMA, *k*-NN, and modified *k*-NN), and then applying a disaggregation model to divide the total flow over the period into days. Valencia and Schakke (1973) began work into disaggregation models, followed by numerous other researchers (Mejia and Rousselle, 1976; Tao and Delleur, 1976; Stedinger and Vogel, 1984; Maheepala and Perera, 1996; Koutsoyiannis and Manetas, 1996). Most disaggregation methods have been based on parametric statistical theory and various non-parametric methods, such as the kernel-based approach, have proven inefficient in high-dimensional problems resulting from calculation complexity (Lall et al., 1996; Sharma et al., 1997; Tarboton et al., 1998). To overcome the limitations of the kernel-based approach, Prairie et al. (2007) developed the *k*-NN based disaggregation approach, which replaced the kernel density function with the *k*-NN bootstrap technique. This modification simplified computation and avoided the prerequisite of Gaussian assumptions and boundary issues. However, the application of this approach to daily flow disaggregation remains problematic. Kumar et al. (2000) adopted a *k*-NN bootstrap technique in conjunction with an optimization scheme for the spatial and temporal disaggregation of monthly streamflow to daily flow. This optimization scheme significantly increases the number of parameters required. In the estimation of discharge at five stations, over 1500 parameters were required for calibration. This demonstrates the fact that complexity associated with daily flow characteristics is exceedingly difficult to describe using parametric methods, suggesting that non-parametric methods might be more suitable. Using the San Juan River as an example, Nowak et al. (2010) demonstrated a *k*-NN-based disaggregation model from annual to daily streamflow. This model resampled historical daily flow proportion vectors to maintain the continuity of daily flow behavior in the results. Unfortunately, parameters relevant to synthetic flow data, such as the mean, skewness, maximum, and minimum, did not differ significantly from previous trends. As a result, flow simulation may suffer from a lack of variability.

Computer software for the analysis and simulation of stochastic hydrology was first introduced in the 1970s, in conjunction with theoretical developments in this field. The US Bureau of Reclamation (USBR) completed Lane's applied stochastic techniques (LAST) in the late 1970s. LAST was used to generate flow conditions for multiple stations within a given watershed for the purpose of engineering (Lane, 1979; Lane and Frevert, 1988). LAST has also been widely adopted by other governmental agencies and academic/research institutions. Another software package, SPIGOT (Grygier and Stedinger, 1990), made a number of improvements to the algorithms employed by LAST, and applied univariate/multivariate statistical generation methods with disaggregation models for the simulation of streamflow. Stochastic analysis modeling and simulation (SAMS) is a more comprehensive model recently developed by Colorado State University (Salas et al., 2006; Sveinsson et al., 2007). The model provides comprehensive functionality for the analysis of hydrologic time series using single or multiple stations. This program is applicable to monthly, seasonal, and annual flow simulation; however, it still lacks complete functionality with regard to daily flow analysis.

Until recently, synthetic streamflow generation has focused primarily on engineering purposes, largely disregarding the ecological implications. In the last few decades however, ecological issues have been receiving increased attention. The integrity of ecosystems is closely bound to hydrologic flow conditions. Effective ecological management requires that the existing regime be characterized using ecologically relevant hydrological parameters. The Indicators of Hydrologic Alteration (IHA) is a hydrological analysis tool commonly used for ecological evaluation (Richter et al., 1996, 1997). By processing daily hydrologic records, the IHA characterizes within-year variations in streamflow on the basis of a series