Journal of Hydrology 521 (2015) 1–17

Contents lists available at ScienceDirect

Journal of Hydrology

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

Configurational entropy theory for streamflow forecasting

Huijuan Cui [a,*], Vijay P. Singh [b,c]

[a] Water Management and Hydrological Science, Texas A&M University, College Station, TX 77843-2117, USA
[b] Department of Biological & Agricultural Engineering, Texas A&M University, College Station, TX 77843-2117, USA
[c] Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-2117, USA

ARTICLE INFO

Article history:
Received 27 August 2014
Received in revised form 18 November 2014
Accepted 22 November 2014
Available online 29 November 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Configurational entropy
Entropy spectral analysis
Spectral density
Burg entropy
Streamflow forecasting

SUMMARY

This study develops configurational entropy theory (CET) for monthly streamflow forecasting. The theory is comprised of three main parts: (1) determination of spectral density (2) determination of parameters by cepstrum analysis, and (3) extension of autocorrelation function. Comparison with the Burg entropy theory (BET) shows that CET yields higher resolution spectral density with more accurate location of spectral peaks. Cepstrum analysis yields more accurate parameters than the Levinson algorithm in the autoregressive (AR) method and the Levinson–Burg algorithm in BET. CET is tested using monthly streamflow data from 19 river basins covering a broad range of physiographic characteristics. Testing shows that CET captures streamflow seasonality and satisfactorily forecasts both high and low flows. High flows are satisfactorily forecasted with the coefficient of determination (r^2) higher than 0.92 for one year ahead of time, with r^2 higher than 0.85 for two years ahead of time, and up to 60 months ahead with r^2 higher than 0.80. However, low flows are forecasted with r^2 higher than 0.50 for one year ahead time. When relative drainage area is considered for analyzing streamflow characteristics and spectral patterns, it is found that upstream streamflow is forecasted more accurately (r^2 = 0.84) than downstream streamflow (r^2 = 0.75). Residuals of forecasted values relative to observed values are found to follow normal distribution.

1. Introduction

Streamflow forecasting is needed for river training and management, river restoration, reservoir operation, power generation, irrigation, and navigation. Time series analysis is often used for forecasting monthly streamflow (Hipel and McLeod, 1994; Noakes et al., 1985). Monthly streamflow time series are stochastic but exhibit seasonal and periodic patterns. Revealing the correct seasonality and correlation structure are two main aspects of forecasting streamflow. Spectral analysis is applied to characterize patterns of streamflow variation (Labat et al., 2005; Smith et al., 1998), identify the periodicity of streamflow (Cengiz, 2011; Hameed, 1984; Sang et al., 2009, 2012), analyze streamflow discontinuity (Adamowski and Prokoph, 2014), separate base flow (Eckhardt, 2005; Spongberg, 2000), and examine the climatic influence on streamflow variability (Andreo et al., 2006; Kuhnel et al., 1990; Prokoph et al., 2012). Thus, spectral analysis permits to extract significant information for understanding the streamflow process and prediction thereof (Fleming et al., 2002; Ghil et al., 2002; Labat, 2005; Marques et al., 2006; Molenat et al., 1999). For forecasting streamflow, spectral analysis has, however, not yet been widely applied.

Burg (1975) defined entropy in the frequency domain and developed what is now called Burg entropy theory (BET). He applied the theory to develop "maximum entropy spectral analysis (MESA)" for time series forecasting. MESA is used to extend autocorrelation in a manner that maximizes the entropy of the underlying process. For a stationary random process BET computes spectral power from autocorrelation of given lags, without assuming autocorrelation of unknown lags as zero (Edward and Fitelson, 1973). It has been widely applied to spectral analysis of geomagnetic, climate indices, surface air temperature, tide levels, precipitation and runoff series (Currie, 1973; Dalezios and Tyraskis, 1989; Ghil et al., 2002; Hasanean, 2001; Padmanabhan and Rao, 1988; Pardo-Iguzquiza and Rodriquez-Tovar, 2006; Sang et al., 2009, 2012; Tosic and Unkasevic, 2005; Wang et al., 2004). BET has also been employed for long-term streamflow forecasting and real-time flood forecasting (Krstanovic and Singh, 1991a,b, 1993a,b) and has been shown to have an advantage in long-term streamflow forecasting over traditional stochastic methods, but has not been found to be superior for short-term forecasting.

Spectral analysis, based on minimum relative entropy (MRE), also called minimum cross-entropy (MCE), was developed by

* Corresponding author. Tel.: +1 979 4223359.
E-mail address: hj.cui@tamu.edu (H. Cui).

http://dx.doi.org/10.1016/j.jhydrol.2014.11.065
0022-1694/

Shore (1979, 1981). The MRE-based spectra are reported to have higher resolution and are more accurate in detecting peak location than other methods for spectral computation (Papademetriou, 1998). However, this method has only been applied for forward modeling and for solving inverse problems in groundwater (Woodbury and Ulrych, 1993, 1996, 1998), but has not been applied to surface hydrology yet.

Frieden (1972) was the first to use configurational entropy in image reconstruction and Gull and Daniell (1978) applied it to radio astronomy. Configurational entropy was later applied for spectral analysis and shown to have better resolution than BET for autoregressive moving average (ARMA) and moving average (MA) processes, and comparable to BET for the autoregressive (AR) process (Nadeu et al., 1981). On the contrary, Burg entropy appears to be better for white noise, as suggested by experiments on speech synthesis (Johnson and Shore, 1983; Katsakos-Mavromichalis et al., 1985). However, neither an explicit solution nor an equivalent extrapolating model had been developed until Wu (1983) used cepstrum analysis to derive an explicit solution for the extended Burg entropy or configurational entropy. This study draws from the study by Wu (1983) for developing the configurational entropy theory (CET) for streamflow forecasting with parameter estimation by cepstrum analysis.

2. Background

There exists a multitude of methods for streamflow forecasting, including deterministic as well as stochastic. Deterministic methods include rainfall–runoff models (Singh, 1988; Wang et al., 2011); watershed models (Singh, 1995; Singh and Frevert, 2002a,b, 2006), such as the simulation hydrology model (SIMHYD) (Chiew et al., 2002), the Systeme Hydrologique Europeen (SHE) (Abbott et al., 1986), TOPMODEL (Beven and Freer, 2001; Beven et al., 1984) and hydrologic model based ensemble streamflow prediction framework (Cloke and Pappenberger, 2009; Wood et al., 2005). For most of the deterministic methods, streamflow is forecasted by simulating soil moisture and groundwater storage with future rainfall, and the accuracy is highly dependent on the determination of initial catchment state (Wood and Lettenmaier, 2008) and forecasting of rainfall is highly uncertain. Streamflow entails a high degree of stochasticity which makes it difficult to forecast streamflow entirely deterministically. In hydrology, stochastic methods, primarily based on time series analysis, are usually employed for forecasting future events and determining the distribution of errors in forecasts.

Time series analysis uses past observed values to forecast future values by a cause-and-effect approach or self-projecting approach. The cause-and-effect approach generates bi-variate or multivariate analysis by linking the series to be forecasted to one or more other series to which it is related. For instance, streamflow forecasting can be generated with climate indicators in two approaches. Wang et al. (2009) used a dynamic climate model to produce rainfall for forecasting seasonal streamflow by the Bayesian joint probability. Otherwise, statistical relationship between climate variables can be proposed for long term forecasting (Chiew and McMahon, 2002; Sharma, 2000; Sharma et al., 2000; Westra et al., 2008). Involving climate indicators, a longer and more flexible range of forecasting can be made, but the result is sensitive to the predictors so they should be carefully chosen. On the contrary, the self-projecting approach entails univariate analysis and uses only past data to uncover its correlation to forecast future values. It has been widely applied for streamflow forecasting during last decades. The autoregressive (AR) and autoregressive moving average (ARMA) methods are mathematically the simplest for time series forecasting, but their application is limited (Carlson et al., 1970; Haltiner and Salas, 1988; Jones and Brelsfor, 1967; Salas and Obeysekera, 1982). For periodic forecasting, periodic autoregressive moving average (PARMA) or periodic autoregressive (PAR) method (Noakes et al., 1985; Salas and Obeysekera, 1992) is recommended, and seasonal autoregressive and moving average (SARMA) or seasonal autoregressive (SAR) method (Salas et al., 1982) is designed for seasonal forecasting. Its integrated version, referred to as autoregressive integrated moving average (ARIMA) (Frausto-Solis et al., 2008), is used for non-stationary flow; with exogenous input, ARMA can be even extended to forecast streamflow generated by rainfall or snowmelt (Hannan and Kavalieris, 1984). However, the underlying linear assumption of ARMA or AR method is not entirely valid (Elshorbagy et al., 2002). In addition to AR and ARMA methods, Kalman filter was used for both long-term seasonal and short-term forecasting, but all parameter matrices must be known (Jimenez et al., 1989; Kalman, 1960). The nonparametric nearest neighbor method performs better than ARMA in one-step ahead daily discharge forecasting (Galeati, 1990) or is equivalent to ARMA for real-time flood forecasting (Toth et al., 2000). However, the nearest neighbor method is suited for large-sample time series and is limited to predict the values no higher than historic observations (Galeati, 1990; Karlsson and Yakowitz, 1987; Toth et al., 2000). For short-term streamflow forecasting, the artificial neural networks (ANN) or support vector regression (SVR) method has an advantage over the above methods, but neither of them provides an explicit characterization and is unable to quantify physical conditions (Behzad et al., 2009; Frausto-Solis et al., 2008; Wu et al., 2009). The accuracy in forecasting short-term streamflow can be increased by wavelet analysis in conjunction with ANN or SVR, though it is limited to a lead time less than a week (Adamowski, 2008; Kisi, 2009, 2010; Pramanik et al., 2011; Shiri and Kisi, 2010).

The above time series methods are preferable under different conditions but have limited application. The AR method plays a significant role in the time series analysis, as it provides a basis for forecasting. Besides, AR method uses the Durbin–Levinson algorithm or Levinson algorithm, a recursive solution, to determine the coefficients of AR by solving the Yule–Walker equations. Later, Burg improved the recursive method to compute the AR parameters through MESA, which is called Levinson–Burg algorithm. It has an advantage in terms of computational ease, short and smooth spectra with a high degree of resolution, and the robustness and stability of estimates (Burg, 1967, 1975). However, both the Levinson algorithm and the Burg–Levinson algorithm are developed for stationary time series, for monthly streamflow with strong seasonal and periodic characteristics, the algorithm sometimes not work well (Boshnakov and Lambert-Lacroix, 2012). Besides, BET has lower resolution in determining multi-peak spectra, while monthly streamflow hardly possess only one periodicity. Thus, this paper introduces an entropy based efficient method, which is generally applicable to forecast streamflow under different conditions and is capable for characterizing seasonality.

The objective of this study is, therefore, to develop configurational entropy theory (CET) for streamflow forecasting that consists of three main parts: (1) determination of spectral density, (2) determination of parameters using cepstrum analysis, and (3) extension of autocorrelation function. The spectral density is obtained by maximizing the configurational entropy subject to autocorrelations. The maximum entropy-based spectral density contains Lagrange multipliers as parameters that are determined by cepstrum analysis. The autocorrelation function is extended by maximizing entropy for streamflow forecasting. Most forecasting methods emphasize high streamflows while low streamflow has seldom been discussed. In this study, monthly streamflows, both high flows and low flows, are forecasted with configurational entropy theory. The paper is organized as follows. Providing a short