Journal of Hydrology 538 (2016) 22–32

Contents lists available at ScienceDirect

Journal of Hydrology

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

Hydrologic regionalization using wavelet-based multiscale entropy method

A. Agarwal [a,b], R. Maheswaran [a], V. Sehgal [a,c], R. Khosa [a], B. Sivakumar [d,e,*], C. Bernhofer [b]

[a] Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110 016, India
[b] IHM, Technical University Dresden, 01737 Dresden, Germany
[c] Department of Biological Systems Engineering, Virginia Polytechnic and State University, VA 24061, USA
[d] School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia
[e] Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 19 January 2016
Received in revised form 12 March 2016
Accepted 14 March 2016
Available online 6 April 2016
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Hydrologic regionalization
Ungaged catchments
Wavelet transform
k-means clustering
Multiscale entropy

SUMMARY

Catchment regionalization is an important step in estimating hydrologic parameters of ungaged basins. This paper proposes a multiscale entropy method using wavelet transform and k-means based hybrid approach for clustering of hydrologic catchments. Multi-resolution wavelet transform of a time series reveals structure, which is often obscured in streamflow records, by permitting gross and fine features of a signal to be separated. Wavelet-based Multiscale Entropy (WME) is a measure of randomness of the given time series at different timescales. In this study, streamflow records observed during 1951–2002 at 530 selected catchments throughout the United States are used to test the proposed regionalization framework. Further, based on the pattern of entropy across multiple scales, each cluster is given an entropy signature that provides an approximation of the entropy pattern of the streamflow data in each cluster. The tests for homogeneity reveals that the proposed approach works very well in regionalization.

1. Introduction

Estimates of streamflow are a prerequisite for solving a number of engineering and environmental problems. These include design or dimensioning a water control structure, economic evaluation of flood protection projects, land use planning and management, water quality control, and stream habitat assessment, among others. When the availability of streamflow records is limited at the site of interest, it is a common practice to apply regionalization techniques to derive the streamflow quantile estimates at the sites where records are limited or in ungaged catchments (Kokkonen et al., 2003). Regionalization can be defined as the transfer of information from one catchment to another (Blöschl and Sivapalan, 1995). This information may comprise characteristics describing hydrologic data or models. To have greater confidence in extrapolating hydrologic behavior from catchments with flow records to an ungaged catchment, all these catchments should form a relatively homogeneous group (Pilgrim et al., 1988; Nathan and McMahon, 1990; Post and Jakeman, 1999). The homogeneity is not only in terms of geographic contiguity but also in terms of hydrologic similarity.

Some of the common approaches for regionalization in hydrology include: the method of residuals (MOR) (Choquette, 1988), the region of influence (ROI) approach (Zrinji and Burn, 1994, 1996), principal component analysis (PCA) (Singh et al., 1996), and cluster analysis and its extensions (Rao and Srinivas, 2006a,b; Isik and Singh, 2008; Srinivas et al., 2008; Satyanarayana and Srinivas, 2011); see also Razavi and Coulibaly (2013) for a review of regionalization methods for streamflow prediction in ungaged basins, and Sivakumar et al. (2015) for a comprehensive account of catchment classification more broadly. Nathan and McMahon (1990) used a combination of multiple regression, cluster analysis, principal component analysis, and graphical representation of eighteen physical catchment variables for predicting the low-flow characteristics of 184 catchments in south-eastern Australia. Notwithstanding their ability to provide reasonable outcomes, these approaches have an important disadvantage in that they mainly rely on the pre-conceived notion of the factors that are thought to influence the behavior of the streamflow from a catchment and that these factors are measurable (Zoppou et al., 2002). In reality, however, the streamflow is a resultant of integrated effects of

* Corresponding author at: School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia. Tel.: +61 2 93855072; fax: +61 2 93856139.
E-mail address: s.bellie@unsw.edu.au (B. Sivakumar).

http://dx.doi.org/10.1016/j.jhydrol.2016.03.023
0022-1694/

many factors, such as topography, lithology, climate, and many others (Yadav et al., 2007), which are occurring at a whole range of time (and space) scales. Therefore, it would be more appropriate to analyze the streamflow across different scales and group the catchments using their signatures.

In recent years, wavelet analysis has become a common tool for analyzing the highly irregular, complex, and intermittent time series often encountered in geophysics (Torrence and Compo, 1998; Smith et al., 1998; Labat et al., 2000; Labat, 2005, 2008; Özger et al., 2010, 2011; Niu, 2013; Niu and Sivakumar, 2013; Chen et al., 2014; Niu et al., 2014; Sehgal et al., 2014a,b; Shoaib et al., 2014, 2015). Several studies have also combined the wavelet transform with other methods to improve our ability to capture the features of geophysical signals (e.g. Adamowski and Sun, 2010; Nourani et al., 2009, 2012). In wavelet transform, by decomposing a time series into time–frequency space, it is possible to determine both the dominant modes of variability and how those modes vary in time. Hence, wavelet transform proves to be a useful tool for analyzing localized variations of power within a time series. Many studies have demonstrated the utility of wavelet analysis in regionalization. For example, Saco and Kumar (2000) used wavelets with rotated principal component analysis of the wavelet spectra, to cluster streamflow stations in the United States. A similar approach using k-means clustering was adopted by Zoppou et al. (2002) to regionalize 286 catchments throughout Australia. They used the wavelet power spectra as the characterizing variable for the cluster analysis and the linear Pearson's correlation coefficient for measuring the degree of similarity between the clusters. The results revealed the capability of wavelets in quantifying the temporal variability of streamflow and, thereby, aiding in regionalization of different catchments.

Even though the wavelet power spectrum has successfully been used for capturing the streamflow behavior, it becomes difficult to use the wavelet spectrum in case of limited data or incomplete data (Yiou et al., 2000). For instance, while the global power spectrum (see below for details) provides the variability of power across different scales, it does not provide any information about the characteristics of the features at a given scale. Therefore, use of the global power spectrum alone does not allow one to infer information about some important features of the signals, such as intermittency, and time variability. However, entropy provides information about the uncertainty at a given scale, which can be corroborated to the level of variation present at that scale. Further, entropy enables determination of least-biased probability distribution with limited signal knowledge and data. Entropy theory can serve as a better approach to study hydrologic and meteorologic processes (Singh, 1997). Numerous studies have used the entropy concept to study a wide variety of problems in hydrology and water resources. Singh and Rajagopal (1987) presented new perspectives for potential applications of entropy theory in water resources. A historical perspective on entropy applications in water resources was presented by Singh and Fiorentino (1992). Harmancioglu and Alpaslan (1992) discussed the use of entropy in water resources, especially for the design and evaluation of water quality monitoring network design. Comprehensive reviews of the applications of entropy theory in hydrology and water resources are available in Singh (1997, 2011), among others.

The concept of entropy, when applied in conjunction with wavelet analysis, can be used to determine the randomness (i.e. level of uncertainty) of a time series at different timescales. At a given scale, maximum entropy is possible when the information is evenly spread across time, and minimum entropy occurs when all the information is contained in a single location. The Wavelet-based Multiscale Entropy (WME), which is a measure of the degree of order/disorder of the signal and carries information associated with multi-frequency signal, can provide useful information about the underlying dynamic processes associated with the signal and can help in regionalization studies (Cazelles et al., 2008). This provides the motivation for the present study to develop a robust regionalization tool based on WME. In this study, the WME method is applied to monthly streamflow data observed at 530 stations in the contiguous United States. Continuous Wavelet Transform (CWT) is applied to each of the observed streamflow time series using the Morlet wavelet to capture the temporal multiscale variability of the streamflow in the form of wavelet coefficients. These wavelet coefficients for each scale are utilized to obtain the entropy for the respective scales. The spectral organization of this multi-spectral variability in terms of WME is identified using k-means clustering.

The rest of the paper is organized as follows. Section 2 describes the proposed methodology with description of the wavelet transform, WME, and k-means clustering technique. Details of the study area and dataset are presented in Section 3. Section 4 presents the application of the proposed methodology to streamflow data, followed by a discussion of the results. Finally, Section 5 presents some of the important conclusions and scope for further research.

2. Methodology

2.1. Wavelet transform

The Continuous Wavelet Transform (CWT) W_n of a discrete sequence of observations x_n is defined as the convolution of x_n with a scaled and translated wavelet $\Psi(n)$ that depends on a non-dimensional time parameter η with zero mean and localized in both frequency and time (Farge, 1992; Torrence and Compo, 1998), and is written as:

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[\frac{(n-n')\delta t}{s} \right] \tag{1}$$

where n is the localized time index, n' is the time variable, s is the wavelet scale, δt is the sampling period, N is the number of points in the time series, and the asterisk (*) indicates the complex conjugate. By varying the wavelet scale s and translating along the localized time index n, one can construct a picture showing both, i.e. amplitude of any features versus the scale and how this amplitude varies with time. The choice of the wavelet function $\Psi(n)$ is neither unique nor arbitrary. The mother wavelet function may be chosen from one of several functions having certain admissibility requirements. Maheswaran and Khosa (2012) provided detailed guidelines for selecting the mother wavelet function. Since the definition of multi-scale entropies is based on the distribution of the activity in the time–frequency domain, a high degree of time–frequency localization allows an accurate measure of entropy. In this study, the Morlet wavelet function is used because of its better time–frequency localization when compared to the other commonly used wavelets, such as the Mexican Hat and the Daubechies wavelets; see Addison (2002) and Maheswaran and Khosa (2012) for some details.

Generally, the result of the wavelet transform is displayed by plotting the amplitude of the wavelet coefficients ($|W_n(s)|$) obtained (Farge, 1992; Meyers et al., 1993). However, the shortcoming of this method is that it is not directly comparable with Fourier spectrum. To resolve this limitation and for direct comparison of spectra, the amplitude squared spectrum, $|W_n(s)|^2$, i.e. wavelet power spectrum, can be used, as is done in the present study.