Environmental Modelling & Software 59 (2014) 59–75

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

An analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrapping

Roshan K. Srivastav*, Slobodan P. Simonovic

Department of Civil and Environmental Engineering, The University of Western Ontario, London, Ontario, Canada

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form
8 May 2014
Accepted 9 May 2014
Available online

Keywords:
Empirical cumulative distribution
Orthogonal transformation
Entropy
Bootstrap
Multi-site

ABSTRACT

Stochastic time series models are very useful in many environmental domains. In this paper, an analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrap stochastic model (M3EB) is developed that can implicitly preserve both the spatial and temporal dependence structure, in addition to the other statistical characteristics present in the historical time series. The proposed model is computationally less demanding and simple in terms of modeling complexity. The maximum entropy bootstrap (MEB) generates random samples from the empirical cumulative distribution function (ECDF) and rearranges the random series based on the rank ordering of the historical time series. The modeling structure of MEB implicitly satisfies the ergodic theorem (preservation of summary statistics) and guarantees the reproduction of the time dependent structure of an underlying process. The orthogonal transformation is used with M3EB to capture the spatial dependence present in the multi-site collinear data. The performance of M3EB is verified by comparing the statistical characteristics between the observed and synthetically generated streamflows. Three case studies from Colorado River Basin, USA; Red River Basin, USA and Canada; and Cauvery River Basin, India; are used to demonstrate the advantages of M3EB. The statistical measures adopted for evaluation of M3EB performance include monthly statistics (mean, standard deviation and skewness), temporal and spatial correlation, smoothing (flows other than present in historical data) and extrapolation (flows outside the range of historical data). The M3EB model shows (i) a high level of accuracy in preserving the statistics; and (ii) a high computational efficiency. Since M3EB can be used for multiple variable problems, the model can be easily extended to other environmental or hydroclimatic time series data.

Software availability section

The source code for M3EB model developed in this paper is available. The implementation of the program is discussed later in the paper.:
Name of the software: M3EB
Developer: Facility for Intelligent Decision Support (FIDS), The University of Western Ontario, Canada
Contact: Roshan K. Srivastav; Slobodan P. Simonovic, Department of Civil and Environmental Engineering, The University of Western Ontario, Canada. roshan1979@gmail.com; simonovic@uwo.ca
Year first available: 2013
Availability: Upon request for research purpose
Implementation: R, MATLAB

1. Introduction

Uncertainty in environmental or hydrologic systems results from the variability associated with the natural processes, not consistent and sufficiently accurate measurements, and deficiency in our knowledge of physical processes and their interactions. The uncertainty present in the time series is captured by synthetically generating the likely patterns (replicates) that mimic the underlying process of the historical time series. These patterns are required for modeling in many environmental domains including air pollution, management of storage reservoirs, wind load studies, water drainage infrastructure design etc (Keyser et al., 2010; Srivastav et al., 2011; Ailliot and Monbet, 2012; Rodriguez et al., 2013). In this paper we consider generation of multi-site multi-variate stochastic streamflows which are useful in water resources infrastructure planning, design and operations.

The stochastic models used in the simulation of multi-site multi-season streamflows can be classified as: (i) parametric models which have linear dependence structure (Box and Jenkins,

* Corresponding author.
E-mail addresses: roshan.srivastav@uwo.ca, roshan1979@gmail.com (R.K. Srivastav).

http://dx.doi.org/10.1016/j.envsoft.2014.05.005
1364-8152/

1976; Salas and Boes, 1980; Bras and Rodriguez-Iturbe, 1985; Bender and Simonovic, 1994); (ii) parametric disaggregation models which attempt to model the dependence structure at the various spatial and temporal scales (Valencia and Schaake, 1973; Mejia and Rousselle, 1976; Grygier and Stedinger, 1988; Lane and Frevert, 1990); (iii) semi-parametric models (Srinivas and Srinivasan, 2001a,b, 2006; Lee et al., 2010; Salas and Lee, 2010; Srivastav et al., 2011); (iv) non-parametric models (Lall and Sharma, 1996; Sharma et al., 1997; Srinivas and Srinivasan, 2005; Westra et al., 2007; Keylock, 2012; Lee, 2012) and (v) non-parametric disaggregation models (Tarboton et al., 1998; Prairie et al., 2007).

Although many developments have taken place in the field of stochastic simulation of multi-site multi-season streamflow time series over the last two decades, the acceptance of these models among the researchers and practicing engineers is found to be limited. The potential reasons may be: (i) modeling complexity (steps involved); (ii) mathematical complexity; (iii) model selection (order of model); (iv) parameter estimation; and (v) computational burden (CPU time). In addition, accurate preservation of statistical characteristics and spatial and temporal dependence structure in case of multi-site multi-season streamflows has been posing a persistent challenge to the modelers. In this study, the authors extend the multi-site multi-variate MEB based weather generator introduced by Srivastav and Simonovic (2014) and present an analytical procedure for its implementation. The main characteristics of the multi-site multi-season maximum entropy bootstrap (M3EB) based stochastic streamflow model are (i) simple and robust model structure; (ii) reliable performance (ability to preserves complex underlying processes); and (iii) computationally efficient algorithm. Further, the proposed method guarantees limited extrapolation beyond extremes. The seasonality present in the historical data is implicitly handled by the proposed model without seasonal detrending.

The main objectives of the paper are to: (i) present the M3EB stochastic model for generation of multi-site multi-season streamflows (ii) demonstrate model's ability to reproduce the complex statistical characteristics present in the historical streamflows; (iii) present analytical implementation procedure in R-programming environment; and (iv) illustrate model implementation using three case studies. The M3EB model combines the strengths of maximum entropy bootstrap with orthogonal transformations to capture the underlying dynamics of multi-site multi-season streamflows. The maximum entropy bootstrap (MEB) is intended to capture the temporal dependence and other statistical characteristics of the data. In order to capture the spatial correlation present in the multi-site collinear data we adopt orthogonal transformation.

Maximum Entropy Bootstrap (MEB) is a nonparametric model that can generate replicates from non-stationary time series and does not depend on reshuffling/resampling of data (either random/conditional), which is the most common approach used by traditional bootstrapping methods. Generation of synthetic replicates using MEB model involves two main steps: (i) use of empirical cumulative distribution function (ECDF) for random sampling; and then (ii) rearranging the samples to preserve temporal dependence structure. The modeling structure of MEB allows it to mimic short-term, long-term and/or cyclic temporal dependence structure present in the historical time series. The use of ECDF allows limited extrapolation and smoothing of the generated replicates. Vinod (2006) first introduced MEB modeling approach to generate synthetic replicates using economic time series. Subsequently, due to its inherent advantages over traditional bootstrap techniques, MEB has been applied in various studies (Cook and Buckley, 2009; Vinod and Lopez-de-Lacalle, 2009; Barbosa et al., 2011; Cook et al., 2013; Yalta, 2013). Very few studies have adopted MEB to model hydrological time series and have been restricted to assessment of uncertainty (Cook and Buckley, 2009; Barbosa et al., 2011; Cook et al., 2013). Cook and Buckley (2009) applied MEB to assess the uncertainty in the cumulative probability distributions for single-site precipitation series. Barbosa et al. (2011) using MEB modeled changes in air temperature and assessed the uncertainty in estimation of distribution of slopes from quantile regression. Recently, Cook et al. (2013) adopted MEB to capture the uncertainty in the reconstructed Upper Indus river flow from tree rings data. The strengths of MEB modeling are: (i) ability to reproduce any temporal correlation structure (Vinod, 2006; Vinod and Lopez-de-Lacalle, 2009); (ii) computational efficiency; and (iii) no need for detrending or differencing of the time series. Recently, the authors developed a novel multivariate MEB modeling approach (Srivastav and Simonovic, 2014) to model weather data and demonstrated its superiority to kNN based weather generator in terms of preservation of statistical characteristics of data and computational efficiency. In this study, the multivariate MEB modeling approach is extended to generation of multi-site multi-season streamflow time series. The model combines (i) MEB to capture the temporal and other statistics and (ii) orthogonal transformation to capture the spatial statistics.

Orthogonal transformation converts the multivariate collinear data variables into linearly decorrelated variables (principal components) arranged in terms of explained variance (highest to lowest). The use of orthogonal transformation for generation of synthetic data has very limited application in hydrology. Smith et al. (1996) applied principal component analysis (PCA) to generate monthly sea surface temperature (SST) in frequency domain, later extended by Caron and O'Brien (1998). They showed that the method is able to preserve historical statistics very well. All the principal components were used for generation of synthetic data and were divided into red noise and white noise components. Similarly, Dreveton and Guillou (2004) used PCA for single-site mean temperature series, in which all the principal components are assumed to be random process. The mean temperature series had to be detrended before the application of PCA. The identification of red noise and white noise components leads to subjectivity and adds to computational burden directly proportional to a number of variables. Recently, Westra et al. (2007) used both principal component analysis and independent component analysis to generate multivariate replicates for hydrologic time series. It was found that the above approach would lead to underestimation of temporal correlation (Lee, 2012). To overcome this limitation, Lee (2012) adopted autoregressive (AR(1)) model for generation of synthetic time series. The PCA based methods use all the components for generation of synthetic data and therefore are computationally intensive for larger number of variables – very common in multi-site or multi-variate hydrologic time series (Westra et al., 2007; Lee, 2012). In this paper, we use the orthogonal transformation to capture the functional relationship between the multivariate data (spatial correlation between the streamflow sites) and decorrelate the data in transformed space.

To best of our knowledge, no studies have been reported to have used MEB in modeling of multi-site multi-season streamflow time series. The proposed model is based on a computationally very efficient algorithm. The inherent strengths of the model include ability to reproduce any time dependence structure (short-term and/or long-term) and implement a simple set of modeling steps.

The rest of the paper is organized as follows. Section 2 presents the M3EB algorithm and software details. Section 3 provides the details of three case studies used. The results and discussion of the proposed M3EB streamflow generator is presented in Section 4. In this section the efficacy of the M3EB model is brought out in terms