Journal of Hydrology 542 (2016) 506–531

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Simulation-optimization framework for multi-site multi-season hybrid stochastic streamflow modeling

Roshan Srivastav [a,b], K. Srinivasan [b,*], K.P. Sudheer [b]

[a] School of Civil and Chemical Engineering, VIT University, Vellore, Tamil Nadu 632014, India
[b] EWRE Division, Dept. of Civil Engineering, IIT Madras, Chennai, Tamil Nadu 600036, India

ARTICLE INFO

Article history:
Received 23 January 2016
Received in revised form 12 August 2016
Accepted 8 September 2016
Available online 15 September 2016
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Stochastic streamflow models
Simulation-optimization
NSGA-II
Evolutionary algorithms
Hybrid matched block bootstrap

ABSTRACT

A simulation-optimization (S-O) framework is developed for the hybrid stochastic modeling of multi-site multi-season streamflows. The multi-objective optimization model formulated is the driver and the multi-site, multi-season hybrid matched block bootstrap model (MHMABB) is the simulation engine within this framework. The multi-site multi-season simulation model is the extension of the existing single-site multi-season simulation model. A robust and efficient evolutionary search based technique, namely, non-dominated sorting based genetic algorithm (NSGA - II) is employed as the solution technique for the multi-objective optimization within the S-O framework. The objective functions employed are related to the preservation of the multi-site critical deficit run sum and the constraints introduced are concerned with the hybrid model parameter space, and the preservation of certain statistics (such as inter-annual dependence and/or skewness of aggregated annual flows). The efficacy of the proposed S-O framework is brought out through a case example from the Colorado River basin. The proposed multi-site multi-season model AMHMABB (whose parameters are obtained from the proposed S-O framework) preserves the temporal as well as the spatial statistics of the historical flows. Also, the other multi-site deficit run characteristics namely, the number of runs, the maximum run length, the mean run sum and the mean run length are well preserved by the AMHMABB model. Overall, the proposed AMHMABB model is able to show better streamflow modeling performance when compared with the simulation based SMHMABB model, plausibly due to the significant role played by: (i) the objective functions related to the preservation of multi-site critical deficit run sum; (ii) the huge hybrid model parameter space available for the evolutionary search and (iii) the constraint on the preservation of the inter-annual dependence. Split-sample validation results indicate that the AMHMABB model is able to predict the characteristics of the multi-site multi-season streamflows under uncertain future. Also, the AMHMABB model is found to perform better than the linear multi-site disaggregation model (MDM) in preserving the statistical as well as the multi-site critical deficit run characteristics of the observed flows. However, a major drawback of the hybrid models persists in case of the AMHMABB model as well, of not being able to synthetically generate enough number of flows beyond the observed extreme flows, and not being able to generate values that are quite different from the observed flows.

1. Introduction

Starting with Fiering (1964) and Matalas (1967), there have been a number of attempts in hydrology to model multi-site/multi-variate streamflows. These belong to one of the two basic types, (i) parametric and (ii) non-parametric models. A detailed review of the parametric type of multivariate/multi-site time series models used in hydrology is presented by Salas et al. (1980), Salas (1993) and McLeod and Hipel (1978), while the various types of non-parametric models in use are reviewed by Lall (1995), Lall and Sharma (1996), Srinivas and Srinivasan (2005) and Salas and Lee (2010).

The parametric type of models may be classified as: (i) Periodic vector AR/ARMA models; (ii) contemporaneous AR/ARMA models; (iii) Disaggregation models. The PAR/PARMA models need to estimate a large number of parameters jointly, to account for the periodic space-time dependence, especially at shorter time scale, with the available historical samples of limited record length. Moreover, the parameter estimates may be unstable and may lead to poor reproduction of some of the important statistics. This motivated the development of a simplified set of models known as contempo-

* Corresponding author.
E-mail address: ksrini@iitm.ac.in (K. Srinivasan).

http://dx.doi.org/10.1016/j.jhydrol.2016.09.025
0022-1694/

raneous AR/ARMA models (CAR/CARMA), wherein the preservation of dependence structure of concurrent streamflows at the various stations was effected through model decoupling (Stedinger et al., 1985; Salas et al., 1985). However, the complex structure of some of the individual site models could impede the exact preservation of the spatial cross-correlations of flows (Rasmussen et al., 1996).

The need to preserve the statistical properties at more than one level necessitated the development of disaggregation models in the hydrologic literature (Harms and Campbell, 1967; Valencia and Schaake, 1973; Mejia and Rousselle, 1976). Preservation of a wide range of statistical relationships between both multiple time scales and space scales needs accurate estimation of a large number of parameters in case of disaggregation models which may not be feasible with the limited hydrologic data available. Hence, staged disaggregation models (Lane, 1982; Stedinger and Vogel, 1984; Grygier and Stedinger, 1988; Santos and Salas, 1992) and condensed disaggregation models (Lane, 1982; Stedinger and Pei, 1982; Pereira et al., 1984; Oliveira et al., 1988; Stedinger et al., 1985; Grygier and Stedinger, 1988) were developed with a view to reduce the number of parameters to make them computationally more amenable. Moreover, empirical adjustment procedures were suggested by Grygier and Stedinger (1988) to restore the summability of the disaggregated flows to the aggregate flows, especially when normalizing transformations were applied to flows. The traditional linear parametric models of streamflows of the AR/ARMA (Box-Jenkins) type including the multi-site parametric disaggregation models can provide only a linear control system representation of watershed processes, while the various physical components of streamflow such as snowmelt runoff, soil water retention as well as soil drainage are dynamic, non-linear processes. Also, non-stationarity trends owing to the underlying dynamics of the physical processes may not be captured effectively.

Koutsoyiannis (1999) developed a parsimonious nonlinear multi-variate dynamic disaggregation model (DDM) that followed a two-step approach for simulation of hydrologic time series. Following this, a generalized mathematical framework for stochastic simulation and forecasting problems in hydrology was proposed by Koutsoyiannis (2000) for modeling stochastic processes with short- or long-term memory structure, in which a generalized autocovariance function was implemented within a generalized moving average generating scheme. Although the DDM and the further developments (Koutsoyiannis, 2000, 2001) were reported to reproduce long-term dependence, and were validated for practical water resources use, the computational complexity involved was high. Langousis (2006) proposed an approach that directly deals with the hydrologic data at the seasonal time scale, but still preserves both the seasonal and the annual statistics and the over-year scaling behavior without restoring to disaggregation techniques. However, it is reported to be quite complex owing to several steps of nonlinear multi-variate optimization (Langousis, 2006). Recently, Efstratiadis et al. (2014) have presented a multi-variate parametric stochastic modeling framework that preserves the important statistical characteristics of the data at multiple sites and at daily, monthly and annual time scales which also involves a number of computational complexities concerning multi-variate parameter estimation and the adjustments and refinements required to reduce the biases in the simulations. The limitations of the parametric stochastic disaggregation models concerning preservation of complex spatial and temporal dependence structure and reproduction of the non-standard marginal distributions have been brought out by Sharma and ÓNeill (2002). Recently, copula-based multisite stochastic simulation models have been proposed by Chen et al. (2015). The spatial and temporal dependencies were modeled by combining bivariate copulas and conditional probability distributions. The main advantages of this method are (i) the parameters of the model can be easily estimated and (ii) the computational time is less.

On the other hand, non-parametric models can provide more accurate representation of the non-linear dynamics of the physical watershed processes by way of effectively modeling the complex dependence structure present in the streamflow data. Also, they can successfully mimic the bi-modality present in the marginal distributions in certain months that may be caused due to different runoff generating mechanisms. A unique feature of the non-parametric technique is to reproduce the empirical structure of multi-variate datasets without recourse to assumptions about data or model structure. Moreover, the complexities associated with parameter estimation are not experienced. Silverman (1986) discusses a wide range of non-parametric methods, while Lall(1995) provides a review of the non-parametric techniques applied to a variety of water and environmental applications. Non-parametric methods have been applied to a wide variety of hydro-climate modeling problems that include stochastic daily weather generation (Rajagopalan and Lall, 1999; Yates et al., 2003), streamflow simulation (Lall and Sharma, 1996; Sharma et al., 1997; Prairie et al., 2006), streamflow forecasting (Grantz et al., 2006; Singhrattna et al., 2005), and flood frequency estimation (Moon and Lall, 1994). Some of the non-parametric techniques that are often used in hydrology are: moving block bootstrap (MBB) (Vogel and Shallcross, 1996);k-nearest neighbor (k-NN) bootstrap (Lall and Sharma, 1996) and its variations and improvements (Prairie et al., 2007; Lee et al., 2010; Salas and Lee, 2010); kernel based methods (Sharma et al., 1997; Tarboton et al., 1998); and matched block bootstrap (MABB) (Srinivas and Srinivasan, 2005). Salas and Lee (2010) have presented a review of the non-parametric models used in streamflow modeling, clearly bringing out the limitations of each model.

Prairie et al. (2006) proposed a modified k-NN approach that enables the simulation of values not seen in the historical record, which has recently been improved by Li and Singh (2014) through the implementation of a multi-model simulation scheme. Also, Salas and Lee (2010) have employed the k-nearest neighbor resampling algorithm with gamma kernel perturbation to generate the seasonal data by conditioning the annual data. Although these models perform well in simulating multi-season streamflows, they are applicable only to modeling single site data.

Prairie et al. (2007) presented a parsimonious non-parametric disaggregation model for space-time simulation of streamflows at river basin level, extending the single site temporal disaggregation scheme of Tarboton et al. (1998) by replacing the tedious kernel based methods with the k-NN approach. Although this method captures the distributional characteristics and the spatial dependencies well, a number of limitations have been pointed out by Lee et al. (2010) concerning underestimation of critical drought characteristics and repetitious nature of the data patterns being generated. Lee et al. (2010) have proposed a spatio-temporal disaggregation model that generates the higher level variable (e.g., annual flow data) based on any parametric or non-parametric model, then generates the lower level sequence (e.g., seasonal flow data) by applying k-nearest neighbor resampling in such a way that their sum is close to the higher level generated flow data. Moreover, genetic algorithm based mixing is implemented to achieve variety in the generated data. This multi-site multi-season non-parametric disaggregation model is reported to yield better simulations than that of Prairie et al. (2007). More recently, based on the maximum entropy bootstrap (MEB) modeling approach proposed by Vinod (2006) for economic time series. Srivastav and Simonovic (2014) have developed a computationally less demanding and simple procedure to model multi-site, multi-season stream flows. The orthogonal transformation is used with MEB to capture the spatial dependence present in the multi-site collinear data.