Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Hydrology 302 (2005) 307–330

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows

V.V. Srinivas[a,*], K. Srinivasan[b,1]

[a]Department of Civil Engineering, Indian Institute of Science, Bangalore 560 012, India
[b]Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600 036, India

Received 10 December 2003; revised 9 July 2004; accepted 30 July 2004

Abstract

The Hybrid approach introduced by the authors for at-site modeling of annual and periodic streamflows in earlier works is extended to simulate multi-site multi-season streamflows. It bears significance in integrated river basin planning studies. This hybrid model involves: (i) partial pre-whitening of standardized multi-season streamflows at each site using a parsimonious linear periodic model; (ii) contemporaneous resampling of the resulting residuals with an appropriate block size, using moving block bootstrap (non-parametric, NP) technique; and (iii) post-blackening the bootstrapped innovation series at each site, by adding the corresponding parametric model component for the site, to obtain generated streamflows at each of the sites. It gains significantly by effectively utilizing the merits of both parametric and NP models. It is able to reproduce various statistics, including the dependence relationships at both spatial and temporal levels without using any normalizing transformations and/or adjustment procedures. The potential of the hybrid model in reproducing a wide variety of statistics including the run characteristics, is demonstrated through an application for multi-site streamflow generation in the Upper Cauvery river basin, Southern India.

Keywords: Streamflow simulation; Non-parametric; Bootstrap; Hybrid model; Drought analysis

1. Introduction

Integrated river basin planning, design and operation/management studies involve stochastic simulation of contemporaneous/concurrent streamflow sequences at several sites in a basin. These simulations augment the description provided by the observed streamflow records at various sites, which, in most cases are limited in size. These synthetic sequences serve as useful input in the design of reservoirs, the evaluation of alternate operation policies for a system of reservoirs within a river basin considering multiple purposes/objectives, the assessment of risk and reliability of water resources system operation and the analysis of critical droughts

* Corresponding author. Tel.: +91-80-2293-2641; fax: +91-80-2360-0404.

E-mail addresses: vvs@civil.iisc.ernet.in (V.V. Srinivas), vvsrinivas@yahoo.com (V.V. Srinivas), ksrini@iitm.ac.in (K. Srinivasan).

[1] Tel.: +91-44-2257-8300; fax: +91-44-2257-0509.

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2004.07.011

at a regional scale, to mention a few. An ideal model employed for stochastic simulation of multi-site multi-seasonal streamflows should ensure the preservation of summary statistics, dependence structure (linear and non-linear), marginal distributions at various temporal levels at each site, in addition to cross-correlations among sites and summability properties within and across sites.

Ever since Fiering (1964) proposed a two-station multi-variate flow model, a number of multi-variate time series models have been developed in hydrology (e.g. Salas and Pegram, 1977; Bartolini et al., 1988; Haltiner and Salas, 1988; Rasmussen et al., 1996). Readers are referred to Camacho et al. (1983), Salas et al. (1985), and Salas (1993) for review of the development of multi-variate/multi-site time series models in hydrology. Multi-variate vector periodic AR/ARMA models necessitate the estimation of a large number of parameters jointly so as to account for the periodic space-time dependence structure. Moreover, parameter estimates may be unstable and may lead to poor reproduction of some of the important statistics. Furthermore, the goodness-of-fit tests for the full multi-variate models are quite complex. This motivated the development of contemporaneous AR/ARMA models, wherein the diagonalization of the parameter matrices allows 'model decoupling', thereby avoiding the joint estimation of model parameters and enabling univariate models to be fitted at each site (Salas and Pegram, 1977). On the other hand, the constraints imposed by the complex structure of the individual site models may result in infeasible cross-correlations among innovations. Rasmussen et al. (1996) give a lucid presentation on the problems associated with the different parameter estimation methods used in modeling multi-variate data in hydrology.

The need to preserve the statistical properties at different time and space scales directed the development of disaggregation models (Valencia and Schaake, 1973; Mejia and Rousselle, 1976). A major drawback of these disaggregation models is that they require the estimation of excessive number of parameters to achieve this. With a view to reduce the number of parameters of multi-site, multi-season disaggregation models and make them computationally more amenable, staged disaggregation models and condensed disaggregation models were developed. While staged disaggregation models (Lane, 1979, 1982; Salas et al., 1980; Stedinger and Vogel, 1984; Grygier and Stedinger, 1988, 1990; Santos and Salas, 1992) reduce the number of parameters by disaggregating aggregate annual flows at one or more sites to periodic flows at those and other sites in two or more steps (stages or cascades), the condensed disaggregation models (Lane, 1979, 1982; Stedinger and Pei, 1982; Pereira et al., 1984; Oliveira et al., 1988; Stedinger et al., 1985; Grygier and Stedinger, 1988) reduce the number of parameters by explicitly modeling only a selected set of the desired relationships among the seasonal flows (lower-level variables).

The LAST computer package (Lane, 1979; Lane and Frevert, 1990) incorporates staging and the use of condensed models. However, it has two major drawbacks, as pointed out by Grygier and Stedinger (1988), one concerning the preservation of the high-lag serial correlations and the other regarding the summability of flows. The condensed disaggregation model of Stedinger et al. (1985) explicitly reproduces the correlations between monthly flows and annual flows, the lower lag serial correlations within the year, and the cross-correlations of concurrent flows at different sites, but not the higher lag serial correlations even within the same water year. The SPIGOT stochastic streamflow package (Grygier and Stedinger, 1990) uses univariate and multi-variate generalizations of the temporal disaggregation model developed in Stedinger et al. (1985). Furthermore, empirical adjustment procedures suggested by Grygier and Stedinger (1988), have been incorporated into SPIGOT (Grygier and Stedinger, 1990) in order to restore summability of the disaggregate flows to the aggregate flows, in the event of normalizing transformations being applied to flows.

The model structure of the aforementioned linear parametric models cannot reproduce non-linearities inherent in the observed data. Moreover, the parametric models use normalizing transformations, that range from the simple power transformations to computationally involved Log Wilson–Hilferty transformation, the ill-effects of which are well known. In addition, these models are not appropriate for simulating long-term persistence exhibited by the observed streamflows. Models such as fractional Gaussian noise, fast fractional Gaussian noise and broken line are suitable for modeling long-term persistence of hydrological processes (related to the observed tendency of the annual average streamflow to stay above or below the mean value continuously for long periods). However, these models have a number of problems as pointed out by Koutsoyiannis (2000).

In the last decade, further advancements have been made in the parametric modeling front by Koutsoyiannis (1992, 1999, 2000) and Koutsoyiannis and Manetas (1996). Koutsoyiannis (1992) developed a parsimonious non-linear multi-variate dynamic disaggregation model (DDM) that follows a stepwise approach for simulation of hydrologic series. This involved two parts: (i) a linear step-by-step moments determination; and (ii) an independent non-linear partitioning. This model was shown to treat the skewness of the lower-level variables explicitly, without loss of additive property. Koutsoyiannis and Manetas (1996) proposed another simpler multi-variate disaggregation method, that retained the parsimony in model parameters for lower-level variables as in DDM (Koutsoyiannis, 1992), and implemented accurate adjusting procedures to allocate the error in the additive property, followed by repetitive sampling to improve the approximations of the statistics that are not explicitly preserved by the adjustment procedures.

More recently, a generalized mathematical framework for stochastic simulation and forecasting problems in hydrology has been proposed by Koutsoyiannis (2000). A generalized autocovariance function is introduced and is implemented in a generalized moving average generating scheme that yields a new time-symmetric (backward–forward) representation. A notable highlight of this model framework is that unlike in the traditional stochastic models, the number of model parameters, the type of generation scheme and the type of autocovariance function can be decided separately by the modeler. This framework is shown to be appropriate for stochastic processes with either short-term or long-term memory. Koutsoyiannis (2001) proposed a methodology for coupling stochastic models of hydrologic processes applying to different time scales. It is noted that DDM and the further developments (Koutsoyiannis, 1992, 2000, 2001; Koutsoyiannis and Manetas, 1996) perform reasonably well at the verification stage.

In the last decade, non-parametric (NP) models have been proposed in hydrologic literature (Lall, 1995) with a view to model the non-linearities inherent in the data. The NP models resample observed streamflow values either conditionally or randomly with replacement. The resampled values are concatenated to form synthetic replicates. These models are data-driven and hence can efficiently model the skewness and other distributional characteristics of the historical flows. Unlike traditional parametric models, the NP models do not make assumptions regarding the dependence structure and the form of the probability density function of the observed hydrologic data. Examples of models in this category include moving block bootstrap (MBB, Vogel and Shallcross, 1996), *k*-nearest neighbor bootstrap (*k*-NN, Lall and Sharma, 1996; Rajagopalan and Lall, 1999; Kumar et al., 2000), kernel based methods (Sharma et al., 1997; Tarboton et al., 1998). However, raw NP models such as MBB and *k*-NN can neither fill in the gaps between the data points in the observed record nor extrapolate beyond the observed data. In other words, they tend to parse the data (no smoothing effect is seen) and this defeats the purpose of streamflow simulation. On the other hand, kernel based methods could generate large negative values in the tails of the distribution. Furthermore, these methods demand enormous computational effort for the estimation of bandwidth in higher dimensions (Prairie, 2002).

An alternate approach that is useful in accounting for non-linearity in historical streamflows is model based resampling. This approach demands the same amount of effort involved in parametric modeling framework for pre-whitening streamflows, since the residuals extracted from the fitted parametric model are resampled considering them as independent and identically distributed. Pereira et al. (1984) and Oliveira et al. (1988) randomly resampled residuals for a multi-site disaggregation model of lag-one annual flows. Tasker and Dunne (1997) suggest blocked bootstrapped multi-site PARMA model, wherein the residuals extracted from periodic autoregressive moving average model with log-transformation (PARMA(1,1)-LT) fitted to observed monthly streamflows at each site, are resampled using contemporaneous 12-month (or 1 year) non-overlapping blocks, with a view to model the complex cross-correlation structure of the innovations. The synthesized streamflows have been used in a position analysis model for a water-supply storage and delivery system in central New Jersey, USA. This method requires the appropriate stochastic model (normalizing transformation and model order) to be identified to fit observed flows at the various sites. In a typical case, normalizing transformation and model order required may differ from site-to-site and from month-to-month. Identification of the most appropriate model (including transformation) to fit historical monthly flows at all sites may not be an easy task for a modeler, particularly often handicapped with the limited length of observed records. If the exact normalizing transformation for each period at each site is applied, the issue of variable degree of non-linearity arises during the inverse transformation process, and this in turn, may distort the contemporaneous cross-correlation structure present in the observed flows between the various sites.

Though a variety of methods have been proposed in hydrologic literature for multi-site simulation of streamflows, none of the methods seem to have gained universal acceptability among practising engineers for various water resources applications. This may either be due to lack of confidence in the existing models, or the inability to adopt models proposed in literature because of their complexity or both. Consequently, the practising hydrologists have resorted to simple techniques that may not model the data adequately. A candid example is the index sequential method (ISM), a simple resampling strategy adopted by the United States Bureau of Reclamation (USBR) to generate synthetic hydrology for the Colorado River Simulation System (Kendall and Dracup, 1991; Ouarda et al., 1997), despite its limitations. The Colorado River basin policymakers have been using the ISM for more than a decade to evaluate the effects of various policies (both water quantity and water quality related) that may be suggested for implementation (Prairie, 2002). Thus, there is a pressing need for identification of a multi-site simulation model that is efficient and at the same time computationally simple to be readily adopted by practising hydrologists in river basin simulation and multi-reservoir operation studies.

Herein, with a view to develop a computationally simple and efficient model for simulation of multi-site streamflows, we have extended the post-blackening approach (Davison and Hinkley, 1997) introduced by Srinivas and Srinivasan (2000, 2001a,b) for single-site modeling of streamflows. This hybrid approach is simple in the sense that it does not have to deal with the uncertainties associated with identification of normalizing transformation, model selection and parameter estimation tasks of the conventional parametric approach. The hybridization helps in acquiring complementary strengths from both the parametric and the NP parts, enabling the reproduction of skewness and other marginal distributional features and the preservation of the serial correlation structure and the month-to-annual correlations. Successful validation of single-site hybrid model has been presented in Srinivas and Srinivasan (2000, 2001a,b) at annual and periodic time scales, which include preservation of long term drought and storage characteristics. Moreover, Izzeldin and Murphy (2000) have reported that the hybrid moving block bootstrap is able to replicate the long-term time dependencies in the observed time series. They have suggested the use of this method for obtaining finite sample critical values of modified rescaled range, which is used to detect long memory in financial, economic and hydrologic time series. In the context of multi-site multi-season simulation of streamflows, the contemporaneous resampling of overlapping yearly blocks of residuals is expected to reproduce the cross-correlation structure present in the historical streamflows, without trading off on the preservation of a wide variety of statistics of interest at each site. Thus, the resulting multi-site hybrid model is believed to result in reasonable level of prediction of multi-site critical run/drought characteristics.

The paper is organized into four sections. Algorithms of the proposed Hybrid multi-site model (HMM) and Tasker and Dunne (1997)'s multi-site model (TDMM) are presented in Section 2. Following this, HMM is used to simulate multi-site periodic streamflows at selected gauging stations in the upper Cauvery river basin, India, and a comparison of the performance of the same with TDMM in stimulating the stream flows is presented in terms of reproduction of a wide variety of statistical attributes (verification) and prediction of multi-site drought characteristics (validation). Section 4 is devoted to summary and conclusions.

2. Algorithms

This section presents the algorithm of the *H*ybrid *M*oving block bootstrap *M*ulti-site model (HMM) introduced in this study and the *M*ulti-site model of Tasker and Dunne (1997) abbreviated as TDMM. In the presentation of the algorithms, bold upper case letters will represent vectors and lower case letters will represent the elements of the vectors.

2.1. Hybrid moving block bootstrap multi-site model (HMM)

Let the observed (historical) streamflows be represented by the vector $\mathbf{Q}^k_{v,\tau}$, where the superscipt k denotes the site index ($k=1,\ldots,n_s$), v is the index for year ($v=1,\ldots,N$) and τ denotes the index for season (period) within the year ($\tau=1,\ldots,\omega$), n_s refers to the number of sites, N represents the number of years of historical record and ω denotes the number of periods within the year. Note that, for $n_s=1$, this algorithm reduces to single-site hybrid periodic streamflow model proposed by Srinivas and Srinivasan (2001a). The modeling steps involved are as follows:

(i) Standardize the elements of the vector $\mathbf{Q}^k_{v,\tau}$ using Eq. (1)

$$y^k_{v,\tau} = \frac{q^k_{v,\tau} - \bar{q}^k_\tau}{s^k_\tau} \tag{1}$$

where $\bar{q}^k_\tau$ and s^k_τ are, respectively, the mean and the standard deviation of the observed streamflows in the period τ at the kth site. The observed streamflows are not subjected to normalizing transformation. This is done with a view to retain the distributional features and non-linearity of the observed flows as such in the residuals, allowing the MBB to capture these features. Thus, the advantage of the NP model is utilized effectively in the hybrid multi-site model.

(ii) Pre-whiten the standardized historical streamflows, $\mathbf{Y}^k_{v,\tau}$, partially, using a simple periodic autoregressive model of order one (PAR(1)), and extract the residuals $\Xi_k(\Xi_k\{\varepsilon^k_{v,\tau},\ v=1,\ldots,N;\ \tau=1,\ldots,\omega\})$, at each site k.

$$\varepsilon^k_{v,\tau} = y^k_{v,\tau} - \phi^k_{1,\tau} y^k_{v,\tau-1} \tag{2}$$

In Eq. (2), $\phi^k_{1,\tau}$ is the periodic autoregressive parameter of order one for period τ, at the kth site. For parameter estimation, the method of moments (Salas et al., 1980) has been used. Depending on the complexity of the dependence structure inherent in the historical record of the seasonal flows, any other parsimonious linear parametric model can also be tried for the partial pre-whitening suggested. Partial pre-whitening of standardized streamflows is suggested with a view to utilize the potential of the NP model, MBB, in capturing the weak linear dependence structure and the non-linear dependence structure present in the residuals.

(iii) Obtain the simulated innovations, Ξ^*_k, at each site k, by randomly resampling the contemporaneous blocks of residuals (Fig. 1) obtained from partial pre-whitening stage, B^k_i, using the MBB (Künsch, 1989) method. In our earlier works (Srinivas and Srinivasan, 2000, 2001a,b), it is demonstrated that resampling blocks of residuals extracted from the fitted stochastic model at a site enables the hybrid model to simulate temporal dependence structure at the site. Resampling contemporaneous blocks of residuals is useful in capturing the site-to-site cross-correlations (dependence across sites).

For the purpose of resampling, we suggest using overlapping blocks of size l, taken as an integral multiple of the number of periods (ω) within the year. The size of the block to be chosen, primarily depends on the structure existing in the residuals. On the other hand, longer blocks used for resampling residuals, may result in loss of variability in the statistics of interest and may reduce smoothing and extrapolation.

Let B^k_i denote the ith among the N_b over-lapping blocks formed from the residuals extracted for the site k, (Ξ_k). Further, let l_b represent (l/ω).

$$B^k_i = (\varepsilon^k_{i,1},\ldots,\varepsilon^k_{i,\omega},\ldots\varepsilon^k_{i+l_b-1,1},\ldots\varepsilon^k_{i+l_b-1,\omega})$$
$$1 \le i \le N_b; \tag{3}$$
$$N_b = N - l_b + 1$$

Each of the overlapping blocks starts with the first period in a hydrological water year. This is done with a view to capture the within-year dependence structure of streamflows at each site. For example, the block sizes of residuals in monthly streamflow

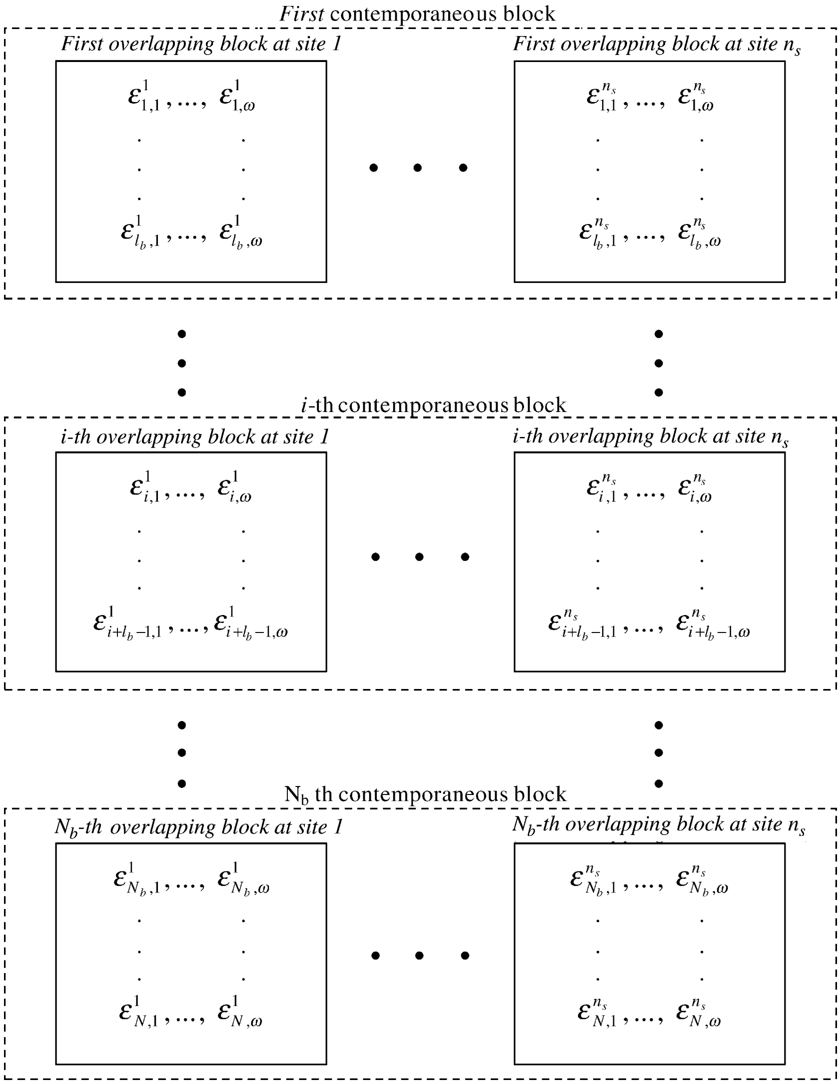

Fig. 1. Contemporaneous blocks of residuals (shown in dotted boxes) that are randomly resampled with replacement to generate innovation series. $l_b = (l/\omega)$ and $N_b = N - l_b + 1$.

modeling context would be 12, 24, 36, and so on (abbreviated as $l = \omega$, $l = 2\omega$, and $l = 3\omega$, and so on). Note that when the block length l is n years long, the overlap is $(n-1)$ years, so that when it is 1 year long there is no overlap. Blocks are drawn at random with replacement from the N_b contemporaneous blocks formed from the residuals. The overlapping blocks corresponding to a site k in the resampled contemporaneous blocks are concatenated to form innovation series at site $k(\Xi^*_k,\ k = 1,\ldots n_s)$. Note that each contemporaneous block consists of one overlapping block from each one of the n_s sites (Fig. 1). Let $\varepsilon^{k^*}_{v,\tau}$ denote an element corresponding to year v and period τ in the bootstrapped innovation series at site k.

(iv) The bootstrapped innovation series Ξ^*_k at each site is then 'post-blackened' to obtain the sequence $\mathbf{Z}^k_{v,\tau}$ (Eq. (4)).

$$z^k_{v,\tau} = \phi^k_{1,\tau} z^k_{v,\tau-1} + \varepsilon^{k^*}_{v,\tau} \tag{4}$$

Herein, to start the generation process, $z^k_{1,0}$ is assumed to be equal to zero for all k. The 'burn-in' or 'warm-up' period is chosen to be large enough to remove any initial bias. The elements of $\mathbf{Z}^k_{v,\tau}$ are then inverse standardized (using Eq. (5)) to obtain the synthetic streamflow replicate $\mathbf{X}^k_{v,\tau}$, at site k.

$$x^k_{v,\tau} = (z^k_{v,\tau} \cdot s^k_\tau) + \bar{q}^k_\tau \tag{5}$$

2.2. Multi-site model of Tasker and Dunne (TDMM)

The modeling steps involved are as follows:

(i) Normalize the observed streamflows $\mathbf{Q}^k_{v,\tau}$ using appropriate normalizing transformation function $\hat{g}_\tau(\cdot)$.

Let $\xi^k_{v,\tau} = \hat{g}_\tau(q^k_{v,\tau})$ for $k=1,\ldots,n_s$; $v=1,\ldots,N$; $\tau=1,\ldots,\omega$.

(ii) Standardize the normalized streamflows to obtain $\mathbf{Y}^k_{v,\tau}$.

$$y^k_{v,\tau} = \frac{\xi^k_{v,\tau} - \bar{\xi}^k_\tau}{\zeta^k_\tau} \tag{6}$$

where $\bar{\xi}^k_\tau$ and ζ^k_τ are, respectively, the sample mean and the sample standard deviation of the transformed streamflows ($\xi^k_{v,\tau}, v=1,\ldots,N$) for period τ at the kth site.

(iii) Fit an appropriate linear periodic stochastic model to $\mathbf{Y}^k_{v,\tau}$, separately to each site k. Parameter estimation is done using method of moments (Salas et al., 1982). Method of maximum likelihood (Vecchia, 1985a,b) can also be used in place. Suppose the model fitted to the standardized transformed data is PARMA(1,1), then:

$$y^k_{v,\tau} = \phi^k_{1,\tau} y^k_{v,\tau-1} - \theta^k_{1,\tau}\varepsilon^k_{v,\tau-1} + \varepsilon^k_{v,\tau} \tag{7}$$

In Eq. (7), $\phi^k_{1,1},\ldots,\phi^k_{1,\omega}$ are the periodic autoregressive parameters and $\theta^k_{1,1},\ldots,\theta^k_{1,\omega}$ are the periodic moving average parameters for the month τ at the kth site.

(iv) Extract the residuals $\Xi_k(\Xi_k = \{\varepsilon^k_{v,\tau},\ v=1\ldots,N,\ \tau=1\ldots,\omega\})$, at each site k, using Eq. (8):

$$\varepsilon^k_{v,\tau} = y^k_{v,\tau} - \phi^k_{1,\tau} y^k_{v,\tau-1} + \theta^k_{1,\tau}\varepsilon^k_{v,\tau-1} \tag{8}$$

Herein, the starting value of $\varepsilon^k_{v,\tau-1}$ is assumed to be equal to zero for all k. These residuals are supposedly nearly independent (since it is assumed that the appropriate stochastic model is fitted in step (iii)), but they may not be identically distributed in time.

(v) Contemporaneous 12-month, non-overlapping blocks of residuals are sampled with replacement to generate innovations for each site. For each site, the residuals for the 12 months of the same randomly selected water year are chosen. According to Tasker and Dunne (1997), this is done with a view to preserve the lag-zero cross-covariance in the generated multi-site streamflow traces. The preservation of the same may not be achieved, if the block size used for resampling the contemporaneous residuals is less than 12 months. An added advantage of this contemporaneous resampling of residuals is that the within-year serial correlation structure at each site may also be reproduced since 12-month block size is used at each site.

(vi) Generate the synthetic replicates $\mathbf{Z}^k_{v,\tau}$ (in the transformed domain) each of size equal to that of the observed record.

$$z^k_{v,\tau} = \phi^k_{1,\tau} z^k_{v,\tau-1} - \theta^k_{1,\tau}\varepsilon^{k^*}_{v,\tau-1} + \varepsilon^{k^*}_{v,\tau} \tag{9}$$

Herein, to start the generation process, $z^k_{1,0}$ is assumed to be equal to zero for all k. The burn-in or warm-up period is chosen to be large enough to remove any initial bias. The elements of $\mathbf{Z}^k_{v,\tau}$ are then inverse standardized, followed by inverse transformation (using Eq. (10)) to obtain the synthetic streamflow replicate $\mathbf{X}^k_{v,\tau}$, at site k

$$x^k_{v,\tau} = \hat{g}^{-1}_\tau[(z^k_{v,\tau} \cdot \zeta^k_\tau) + \bar{\xi}^k_\tau] \tag{10}$$

where $\hat{g}^{-1}_\tau(\cdot)$ denotes the inverse of the normalizing transformation function $\hat{g}_\tau(\cdot)$.

It is to be noted that TDMM requires the appropriate stochastic model (normalizing transformation and model order) to be identified to fit observed flows at the various sites. In a typical case, normalizing transformation and model order required may differ from site-to-site and from season-to-season. Fitting the most appropriate model (including transformation) for all months at all sites, may not be an easy task for a modeller, particularly with the limited length of observed records. If the exact normalizing transformation for each period at each site is applied, the issue of variable degree of non-linearity arises during the inverse transformation process, and this in turn, may distort the contemporaneous cross-correlation structure present in the observed flows between the various sites. If transformations with lower bound

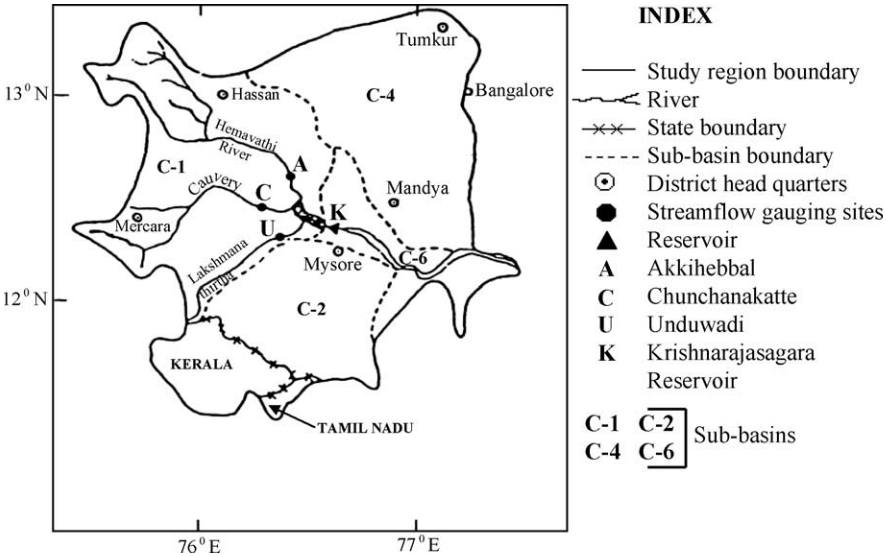

Fig. 2. Location of Gauging stations in the study region of the Cauvery river basin, India.

are to be applied to normalize the flows, then, this may introduce additional bias even in the summary statistics of the simulated flows. Also, this model may not be effective in reproducing the statistical attributes at aggregated annual level.

The basic differences between TDMM and HMM are: (i) HMM does not use any normalizing transformation to the original data, while TDMM uses normalizing transformations; (ii) HMM does not attempt to fit the best parametric model to the data, while TDMM does; and (iii) HMM may choose to use an overlapping block of size ω or higher for resampling the residuals depending on the structure present in the residuals, while TDMM uses only a non-overlapping block size of ω.

3. Performance investigation

In this section, the multi-site hybrid model (HMM) introduced, is used for the spatial simulation of the monthly streamflows at four gauging stations located in the upper reaches of the Cauvery river basin, southern India (Fig. 2). Table 1 provides the details of the gauging stations on the upper Cauvery river and its tributaries, namely the Hemavathi

Table 1
Multi-site periodic modelling of streamflows—details of gauging stations

Name of river basin	Location of basin	River	Gauging station	Location of gauging station	Catchment area (km^2)	Period of record	Mean annual discharge (m^3/s)
Cauvery	Karnataka State, India	Cauvery	Chunchanakatte	About 48 km upstream of KRS	2966	1916–1974	91.58
		Hemavathi	Akkihebbal	About 40 km upstream of KRS, clear of reservoir spread	5198	1916–1974	80.78
		Lakshma-nathirtha	Unduwadi	About 40 km upstream of KRS, clear of reservoir spread	1502	1916–1974	13.47

KRS, Krishnarajasagara dam (downstream key station).

and the Lakshmanathirtha rivers that have been considered for modeling. In the discussion to follow, the gauging station at Krishnarajasagara (KRS) dam will be referred to as the *Key station* and the upstream gauging stations, namely Chunchanakatte, Akkihebbal and Unduwadi, will be referred to as *substations*. The historical unregulated monthly streamflows at these gauging stations are available for the period 1916–1917 to 1973–1974 from the Water Resources Development Organization of the Karnataka state (WRDO, 1976). In this part of the basin, more than 90% of the annual flow occurs during the southwest monsoon period (June to October).

The river Cauvery rises on Brahmagiri Hills of the Western Ghats in Coorg district of southwestern Karnataka State, India at an elevation of about 1275 m (above MSL). It flows in a southeasterly direction across the plateau of Mysore through the states of Karnataka and Tamil Nadu and confluences with the Bay of Bengal in the state of Tamil Nadu (Fig. 3). Just upstream of the KRS dam, the Cauvery is joined by two tributaries, the Hemavathi and the Lakshmanathirtha. This dam has been under operation since 1931 with a live storage capacity of about 1258.6 Million m^3. One more dam has been constructed on the tributary Hemavathi and it is in operation since the early 1980s with a live storage capacity is about 962 Million m^3 (Srinivasan et al., 1992). These two dams serve the irrigation demands of the region for two seasons in a water year (June to May). The streamflows in the upper basin sites display strong cross-correlation structures. Moreover, the marginal distribution of streamflows in a few months exhibit bi-modality. The dry and the wet periods within the water year are well defined owing to the monsoon climate.

In this study, the hybrid multi-site model proposed (HMM) is fitted to the streamflow data of the upper cauvery basin in southern India. The parametric constituent of this model is PAR(1) with no transformation and the parameter estimation is done using the method of moments (Salas et al., 1980). The NP constituent is the MBB with $L=2\omega$ ($=24$ months). The results from this model are compared with those from the recently proposed multi-site stochastic model of Tasker and Dunne (TDMM, 1997). For TDMM, the best low-order linear model,

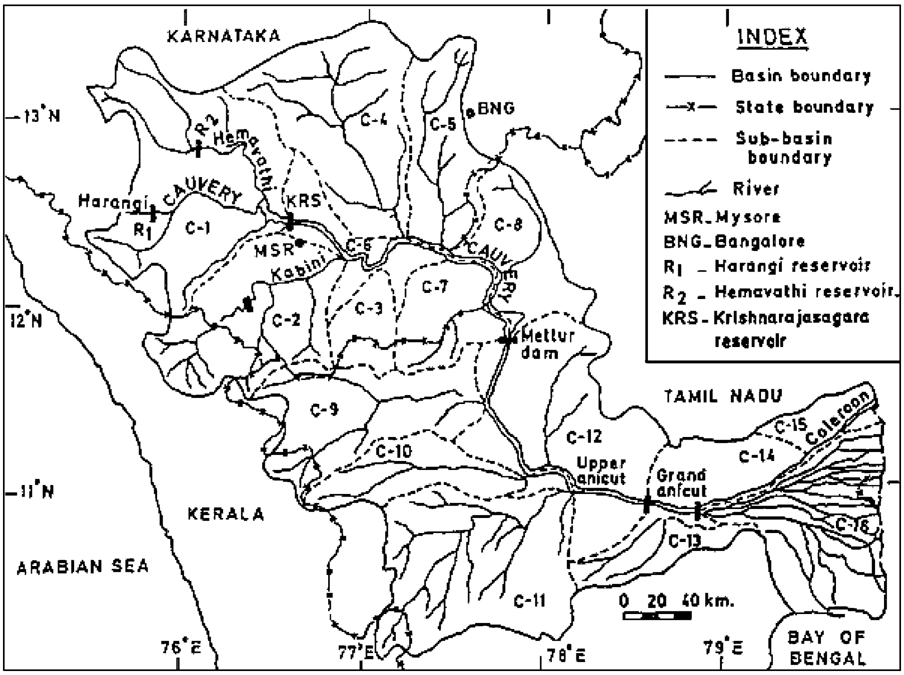

Fig. 3. Index map of the Cauvery river basin, India.

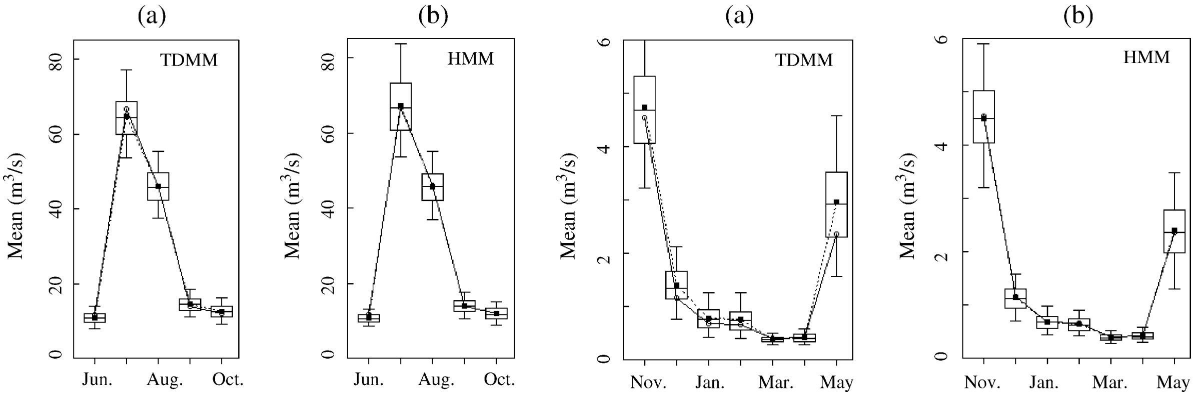

Fig. 4. Preservation of the mean monthly streamflows for the river Lakshmanathirtha at station Unduwadi by (a) Tasker and Dunne's multi-site model (TDMM) and (b) hybrid multi-site model (HMM) graphed using box plots: a line in the middle of the box represents median. The historical statistic is represented by a circle, and the mean of the generated statistic over 500 replicates is represented by a solid square. The solid line that joins the circles indicates the historical trend, while the dotted line connecting the solid squares depicts the mean synthetic trend.

PAR(2), (from among PAR(1), PAR(2) and PARMA(1,1)), was fitted to the standardized monthly streamflows. The Wilson–Hilferty transformation was adopted for normalizing the streamflows (out of the standard transformations considered), based on the diagnostic checking of the residuals.

3.1. Preservation of summary statistics

The mean monthly flows (MMF) at each station are well reproduced by both TDMM and HMM, though TDMM shows some inflation in the preservation of the mean flows for December and May months in case of the river Lakshmanathirtha (Fig. 4). The standard deviations of observed flows are not reproduced by TDMM in quite a few months, while HMM shows a consistent performance. For brevity, preservation of the statistic at Lakshmanathirtha and the aggregate (key) site KRS are shown (Figs. 5 and 6).

At the aggregated annual level, the mean flow is well reproduced by both TDMM and HMM (not shown for brevity). However, it is evident from Fig. 7 that TDMM underpredicts the standard deviation of observed aggregated annual flows at the gauging stations Akkihebbal, Chunchanakatte and the key site KRS, while HMM is seen to reproduce the statistic at all stations.

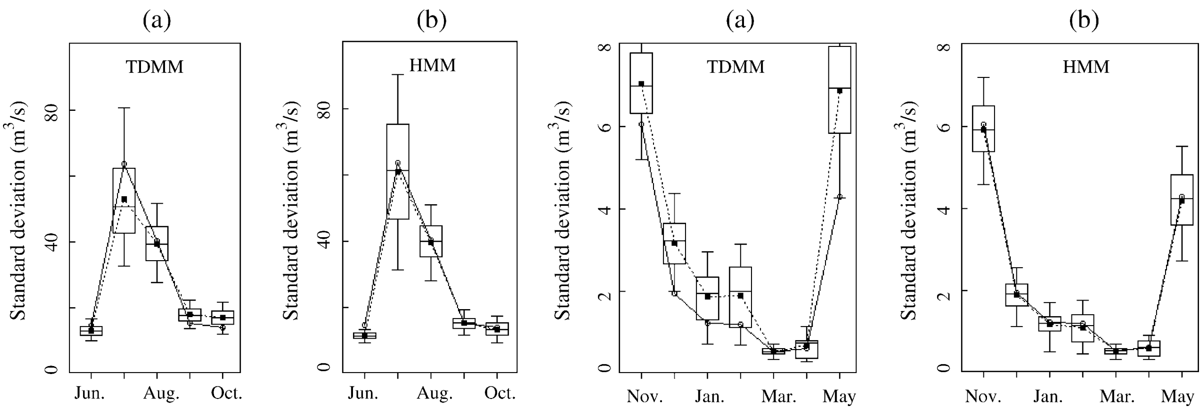

Fig. 5. Preservation of the standard deviation of monthly streamflows for the river Lakshmanathirtha at station Unduwadi: a comparison between (a) Tasker and Dunne's multi-site model (TDMM) and (b) hybrid multi-site model (HMM).

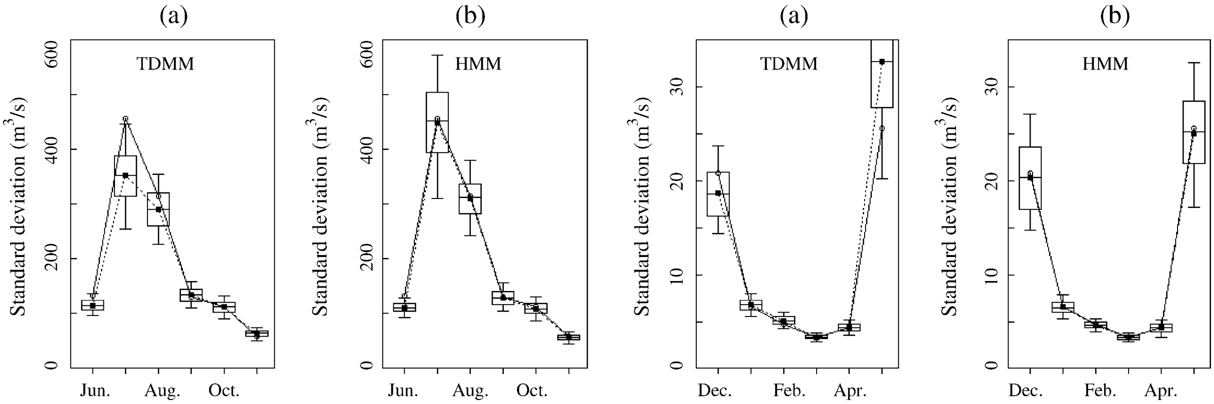

Fig. 6. Preservation of the standard deviation of monthly streamflows for the river Cauvery at station Krishnarajasagara dam: a comparison between (a) Tasker and Dunne's multi-site model (TDMM) and (b) hybrid multi-site model (HMM).

Fig. 8 compares the performance of the models in reproducing skewness of the monthly flows. Both models are seen to reproduce the skewness of monthly streamflows fairly well, except in June. At the aggregated annual level, TDMM underpredicts the skewness considerably at all stations. In contrast, it is interesting to note that HMM is able to reproduce the statistic satisfactorily at all stations, despite no normalizing transformation being applied. It is worth mentioning here that HMM is gaining from its NP constituent (MBB), in reproducing the skewness at both monthly and aggregated annual levels. This is because, at the pre-whitening stage, HMM uses a simple PAR(1) model without any normalizing transformation. Hence, much of the skewness structure of observed flows is retained in the residuals that are subsequently modeled by MBB (see Hybrid Effect in Srinivas and Srinivasan, 2001a). On the other hand, TDMM is relying entirely on its normalizing transformation (Wilson–Hilferty transformation) for modeling skewness at monthly level, like parametric models. Hence, TDMM is able to reproduce skewness at monthly level, but not at the aggregated annual level.

3.2. Preservation of marginal distributions

Fig. 9 shows a typical marginal distribution plot for July month which is characterized by high streamflows in a water year. It may be noted from the figure that HMM preserves the peak density of the marginal distribution of historical flows faithfully, at all stations, while TDMM overpredicts the same. However, TDMM, in general, offers more smoothing in simulations compared to HMM. Moreover, historical streamflows at two of the gauging stations being considered exhibited bi-modality in February, March and April months. Both TDMM and HMM simulated bi-modality observed in historical streamflows fairly well (not shown due to lack of space).

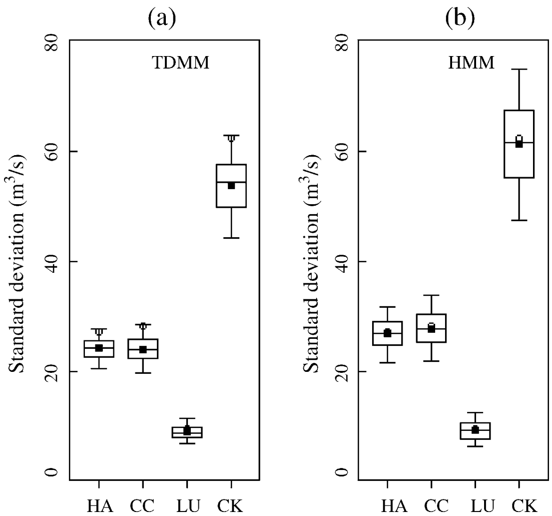

Fig. 7. Preservation of the standard deviation of observed flows at the aggregated annual level for the rivers: a comparison between (a) Tasker's and Dunne's multi-site model (TDMM) and (b) hybrid multi-site model (HMM). HA, Hemavathi at Akkihebbal; CC, Cauvery at Chunchanakatte; LU, Lakshmanathirtha at Unduwadi; CK, Cauvery at KRS.

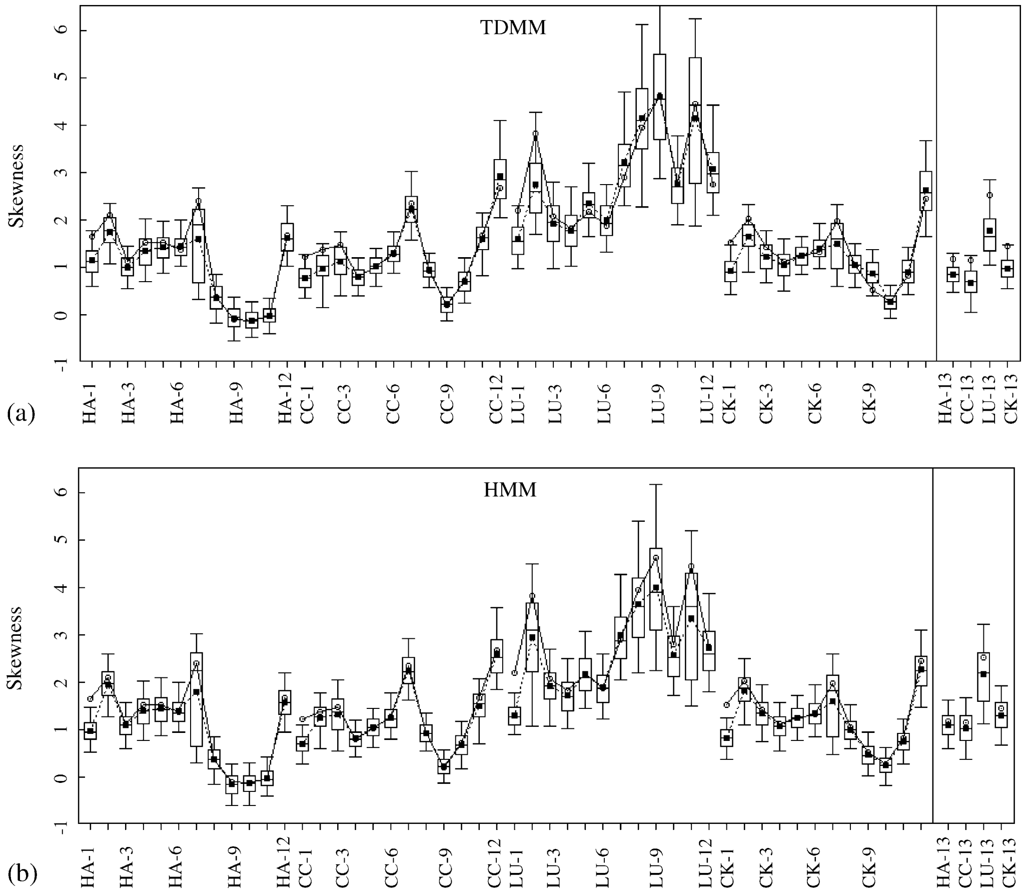

Fig. 8. Preservation of the skewness of observed streamflows at monthly and aggregated annual levels: a comparison between (a) TDMM and (b) HMM. Notation HA-1 on X-axis indicated skew at station HA for the June month of hydrologic water year. Months are numbered according to the water year (1=June, 2=July,…,8=January, 9=February,…,12=May), whereas annual time step is denoted by the number 13. HA, Hemavathi at Akkihebbal; CC, Cauvery at Chunchanakatte; LU, Lakshmanathirtha at Unduwadi; CK, Cauvery at KRS.

In case of the proposed HMM which blends both parametric and NP models, the smoothing and extrapolation in simulations is limited when applied to periodic data. This is the limitation that arises due to shortcomings of MBB (NP constituent of HMM), which requires resampling residuals in large blocks to effectively capture dependence structure in historical flows. As the extrapolation offered by HMM is rather limited, the chance of encountering negative values is very less, except for the months where minimum historical flow value is zero (or near zero). It is suggested that the negative values generated by HMM be treated as zero as done in parametric models. Herein, it is worth mentioning that negative flows are not synthesized by raw NP models such as MBB and *k*-NN, as these models do not offer any smoothing or extrapolation. In contrast, parametric models could generate excessive number of negative flows because of the transformation effects and the tremendous smoothing and extrapolation offered by those models. The problem with parametric models is exacerbated when the historical record comprises of all zero (or near zero) flow values in one or more months of a water year. This situation is quite feasible in ephemeral rivers in arid climate, which are characterized by long dry periods.

For the data set being considered for the study, HMM could generate reasonable extrema below historical minimum value. However, the extrapolation

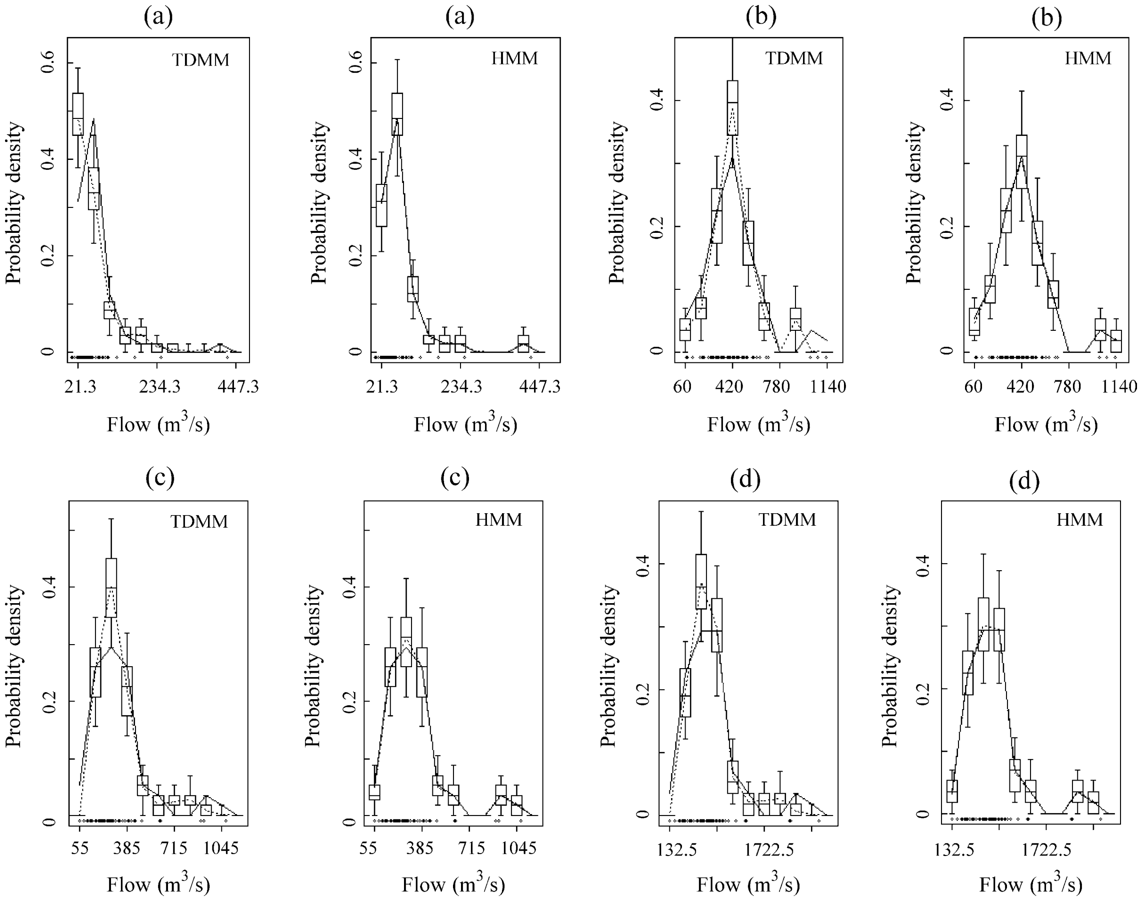

Fig. 9. Preservation of marginal distribution of July flows: a comparison between TDMM and HMM. The dots below the box plots denote the observed flow values. The continuous line connecting the boxes represents the historical trend, while dashed line connecting the boxes depicts the mean synthetic trend of the statistic of concern. (a) Unduwadi; (b) Chunchanakatte; (c) Akkihebbal; and (d) KRS.

beyond historical maxima was found to be very minimal especially for months with high skewness. It is hoped that a better strategy for resampling residuals extracted from pre-whitening stage would help the hybrid model proposed in this paper to overcome the limitation regarding extrapolation and smoothing in simulations.

3.3. Preservation of dependence structure

The monthly lag-zero site-to-site cross-correlations are reproduced by both models (Fig. 10), though HMM performs better. At the aggregated annual level, TDMM shows some underestimation in the preservation of the lag-zero site-to-site cross-correlations, while HMM is seen to be good at reproducing the same (Fig. 10). However, HMM is able to reproduce the lag-1 site-to-site cross-correlations (Fig. 11) (though with some deflation), as a result of using overlapping blocks of size $l=2\omega$ for residual resampling. On the contrary, TDMM is not able to preserve the lag-one cross-correlations, as expected.

The serial correlation structure of periodic streamflows and month-to-annual cross-correlations are well reproduced by both the models, though HMM shows marginally better performance than TDMM (figures are not shown for brevity). The non-linear dependence (expressed through state-dependent correlations) is not found to be significant for the rivers considered for modeling in the upper Cauvery basin system. Hence, reproduction of state-dependent correlations is not presented herein.

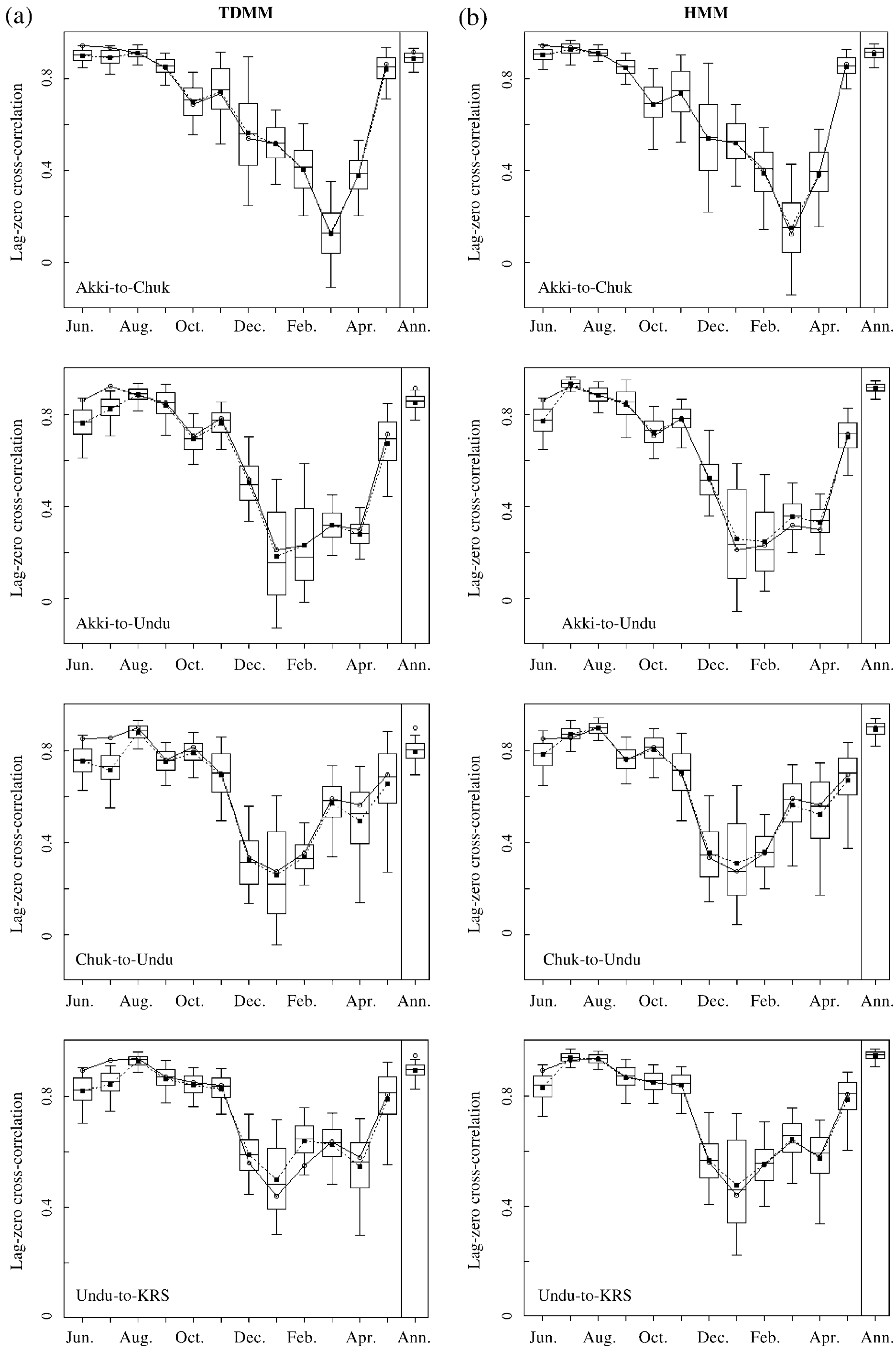

Fig. 10. Preservation of lag-zero site-to-site cross-correlations between flows observed at different stations at both monthly and aggregated annual (Ann.) levels: a comparison between TDMM and HMM. River basin: Cauvery, India. Akki, Akkihebbal; Chuk, Chunchanakatte; Undu, Unduwadi; KRS, aggregate site.

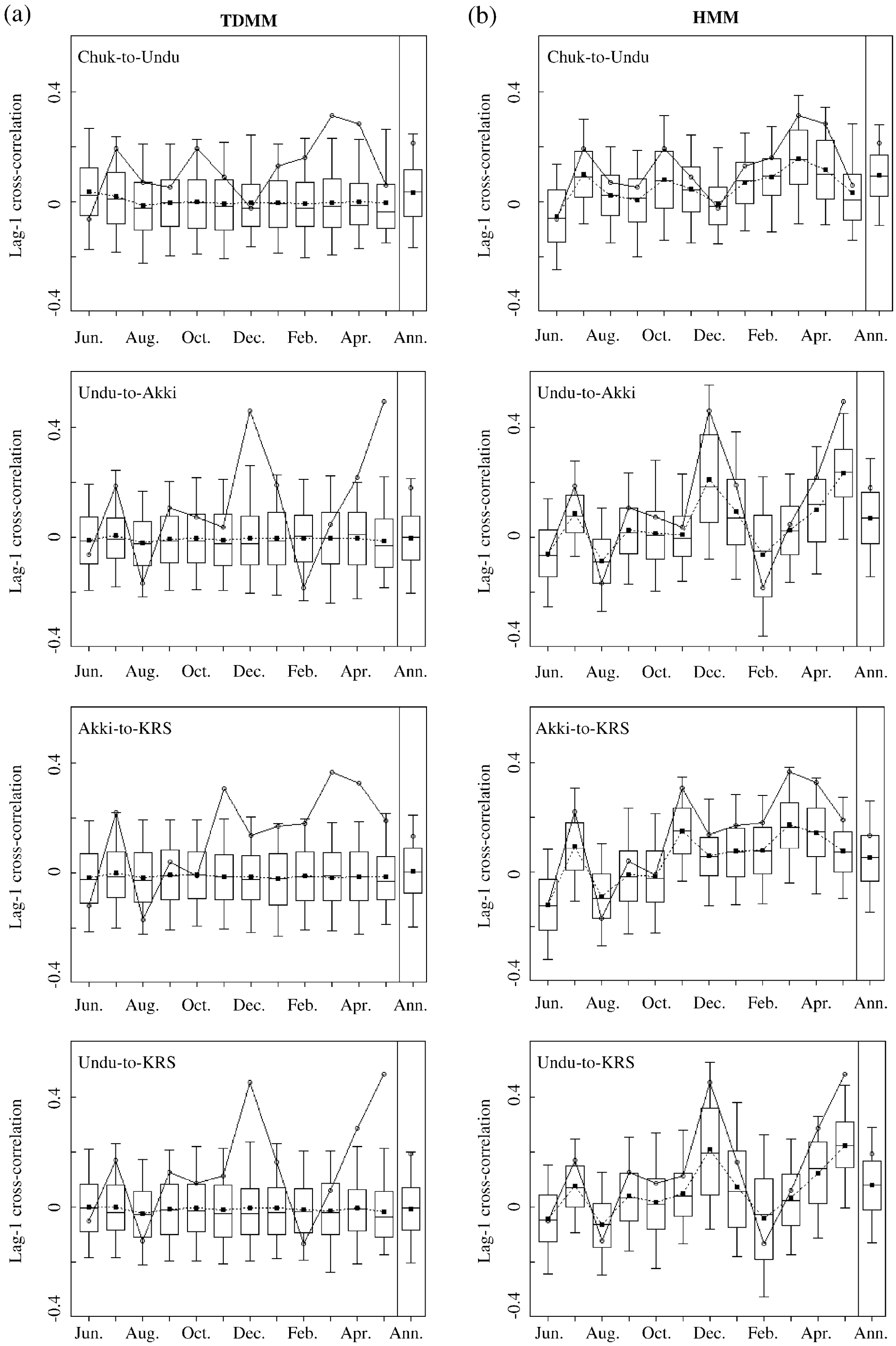

Fig. 11. Preservation of lag-one site-site cross-correlation of observed flows: a comparison between (a) TDMM and (b) HMM. Note that among the two sites mentioned in plots, the second site is being lagged by 1 month. Akki, Akkihebbal; Chuk, Chunchanakatte; Undu, Unduwadi; KRS, aggregate site.

3.4. Preservation of drought characteristics

In the multi-site streamflow modeling context, it is often of interest to examine the ability of a stochastic model to reproduce the critical and the mean run characteristics. Such analysis is believed to provide useful input to regional drought analysis and adequately assess system sensitivity and plan for the operation of reservoirs in a river basin system for the likely future drought scenarios. In the multi-site context, a run is said to begin at a point in time when the flows at all the stations are below their respective truncation levels and is supposed to terminate when the flow at any station becomes greater than the truncation level at that station (Yevjevich, 1972; Haltiner, 1985).

The multi-site run-sum for a particular run is defined as: $s_i = \sum_{k=1}^{n_s} s_i^k$, wherein i denotes the run number, k refers to the station number and n_s represents the total number of stations being modeled. It should be noted that the run-sum for an event i, s_i^k, at a given site k represents only that portion of the total deficit at that site that occurs during the time when all sites considered have flows below their respective truncation levels (Haltiner, 1985). Fig. 12 presents the pictorial description of the same. Herein, the truncation levels have been chosen as percentages of the historical MMF (50–100% MMF at intervals of 5% MMF). The run characteristics considered for the evaluation of the relative performance of the two models are: (i) maximum run length (MARL); (ii) maximum run sum (MARS); (iii) mean run length (MERL); and (iv) mean run sum (MERS).

Let dl_i denote the length and s_i denote the volume of water below a specified monthly truncation level (i.e. deficit volume in Million m^3) for the ith negative run. Then, the aforementioned run characteristics can

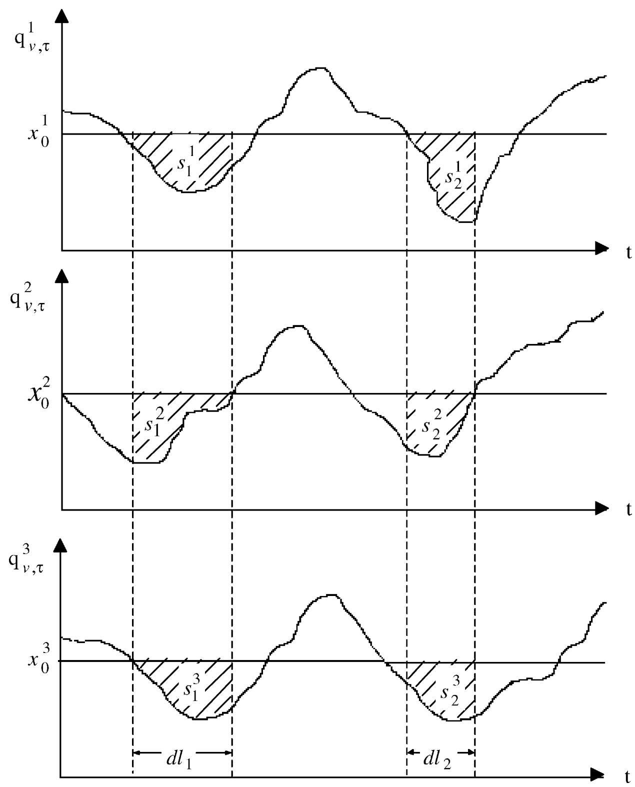

Fig. 12. Graphical representation of multi-site run characteristics: x_0^k is truncation level for site k; dl_i is the length of ith run; s_i^k represents deficit volume at site k for ith run; $q_{v,\tau}^k$ is flow at station k for year v and period τ and t is time axis.

be expresses as

$$\mathrm{MARL} = \max[dl_1, \ldots, dl_{\mathrm{NR}}]; \tag{11}$$

$$\mathrm{MARS} = \max[s_1, \ldots, s_{\mathrm{NR}}]; \tag{12}$$

$$\mathrm{MERL} = \frac{\sum_{i=1}^{\mathrm{NR}} dl_i}{\mathrm{NR}} \tag{13}$$

$$\mathrm{MERS} = \frac{\sum_{i=1}^{\mathrm{NR}} s_i}{\mathrm{NR}} \tag{14}$$

where NR denotes the total number of runs in the flow sequence (historical/synthetic).

Fig. 13 compares the performance of TDMM and HMM in reproducing the critical run characteristics. It is evident that neither of the models is consistent in reproducing the critical run length at all truncations, though HMM is seen to be relatively better. Furthermore, TDMM is not able to model the critical run sum, while HMM exhibits a reasonable performance (Fig. 13). From Fig. 14, it may be noted that HMM is more consistent in reproducing MERL, than TDMM. Moreover, TDMM consistently underpredicts the historical MERS, while HMM is seen to be good at reproducing the same. However, both

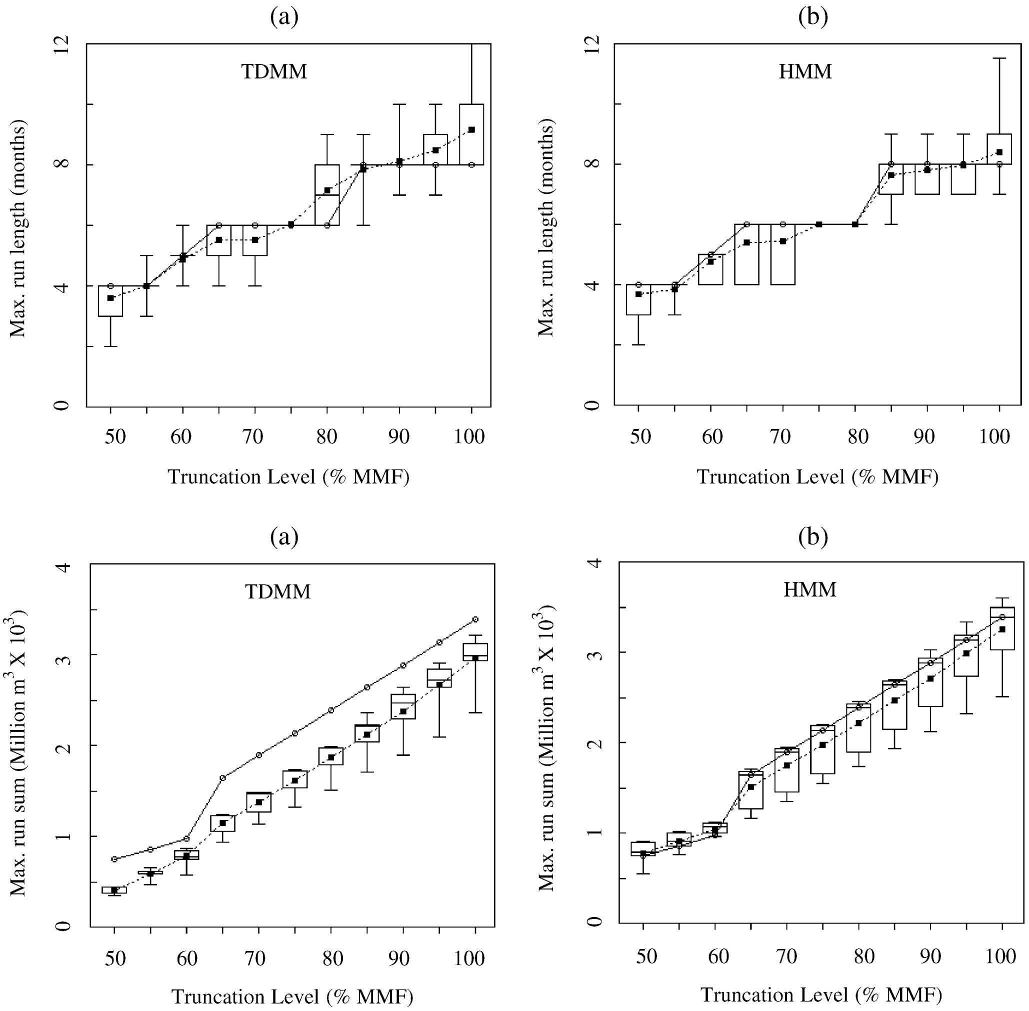

Fig. 13. Preservation of historical critical run characteristics based on monthly truncations: a comparison between (a) TDMM and (b) HMM.

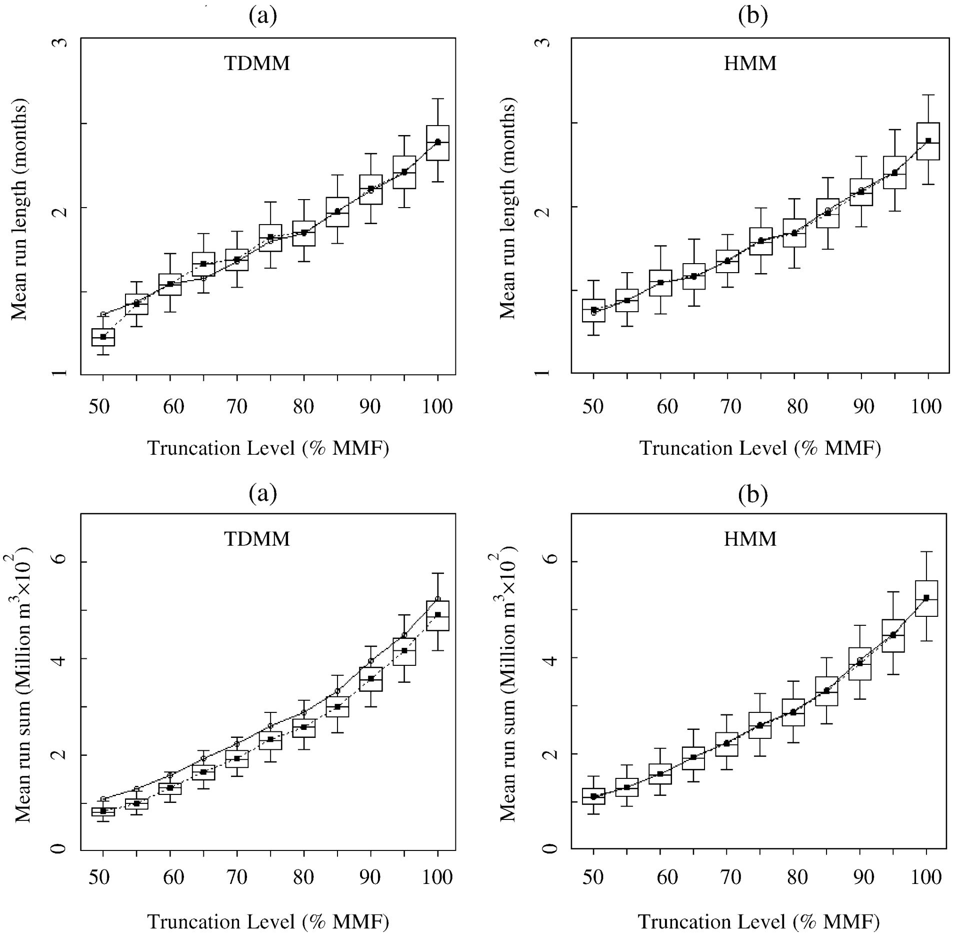

Fig. 14. Preservation of historical mean run characteristics based on monthly truncations: a comparison between (a) TDMM and (b) HMM.

the models could simulate the number of runs in the historical record fairly well, which is not shown here for brevity. The better performance of HMM can be mainly attributed to the better reproduction of cross-correlations (Figs. 10 and 11).

It is felt that the definition of run provided by Haltiner (1985) may not be effective in reflecting the severity of droughts with increase in the number of sites considered for modeling, particularly for rivers that experience low flows over considerable periods of the year (as in the Cauvery basin). This is because, the truncation levels in the low flow months are low and even a small increase in monthly inflows at any one of the stations modeled would imply end of the dry flow run, which may not be realistic at all. The frequency of encountering such a situation is high in the dry flow periods and as a result of this, the criticality of run characteristics appears to be much lower than that in reality. Future research needs to evolve a more appropriate and realistic definition of run that should be able to identify the true beginning and the termination of drought, in the context of (multi-site) periodic streamflows. The works of Herbst et al. (1966) and Tallaksen et al. (1997) may provide some lead in this direction. However, in this study, no such attempt is made to refine the definition of run. In any case, to gain a better idea of the performance of the models in reproducing the critical and the mean run characteristics, two more scenarios are considered, wherein the accumulation of the run characteristics (run length and run sum) are done with a view to minimize the errors of underestimation owing to the identification of unrealistic interruptions of the runs.

The first scenario considers a water year to be consisting of three seasons (i.e. each season is 4 months long) for the sake of drought analysis. Accordingly, truncation levels are defined as percentage of the historical mean seasonal flows (MSF) (50–100% MSF at intervals of 5% MSF). The performance of the two multi-site monthly streamflow models in reproducing the run characteristics at the aggregated seasonal level is compared in Figs. 15 and 16. Herein, it is to be noted that both critical and MERLs are reported in terms of months. This is done with a view to appreciate the extent of differences in the estimation of the run characteristics between month-based and season-based truncations. As expected, the seasonal analysis of runs has resulted in more severe droughts (Figs. 15 and 16) than those that had resulted from the analysis at monthly level (Figs. 13 and 14). Here too, TDMM is not able to model the historical MARS and bias is seen in MERL at a few truncation levels. Furthermore, the historical MERS is underestimated at all the lower and intermediate truncation levels. In contrast, HMM is able to reproduce all the run characteristics fairly well. The HMM is seen to be better than TDMM in reproducing number of runs as well (Fig. 17).

The second scenario considers the water year to be consisting of eight periods, for the purpose of drought analysis. The first seven periods are of 1 month duration each (June to December), while the eighth

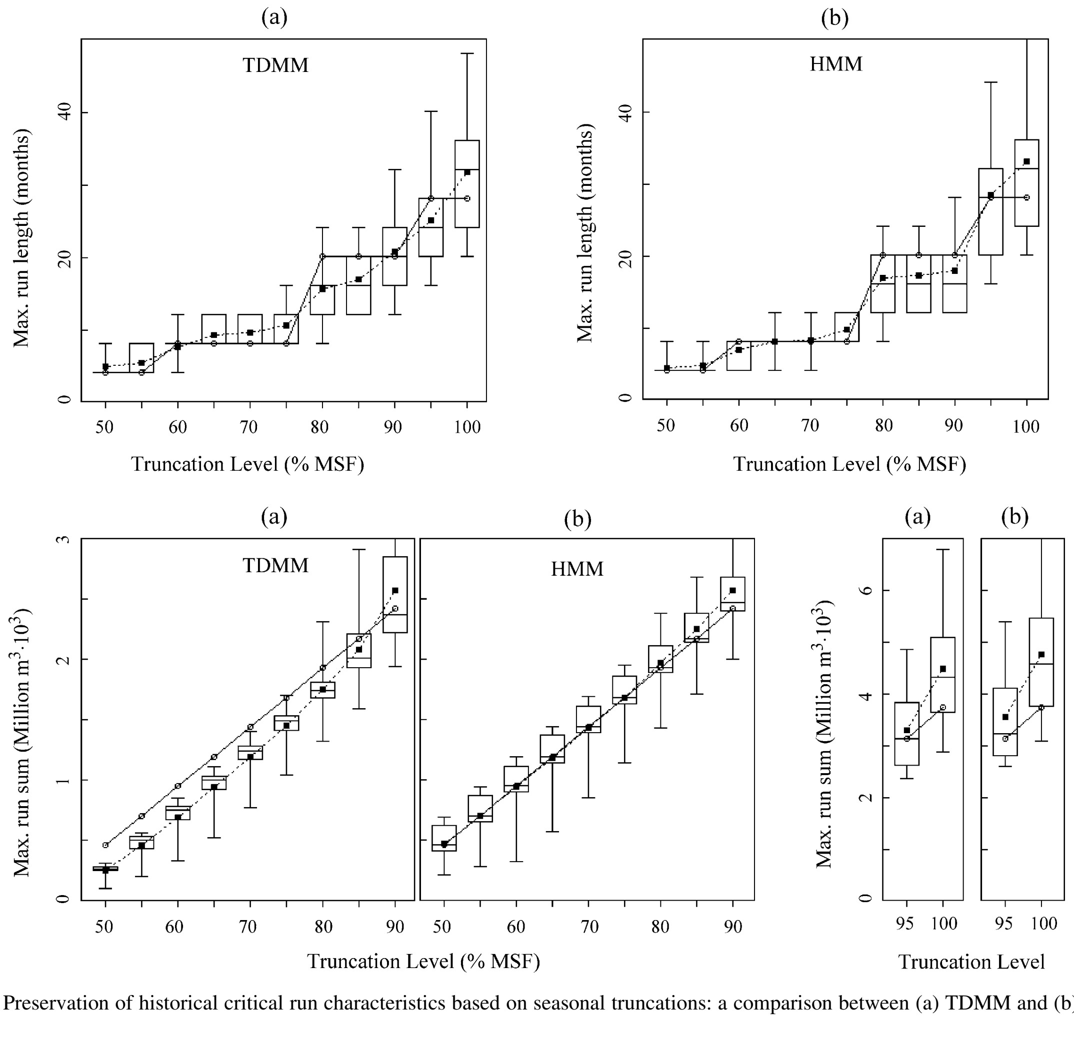

Fig. 15. Preservation of historical critical run characteristics based on seasonal truncations: a comparison between (a) TDMM and (b) HMM.

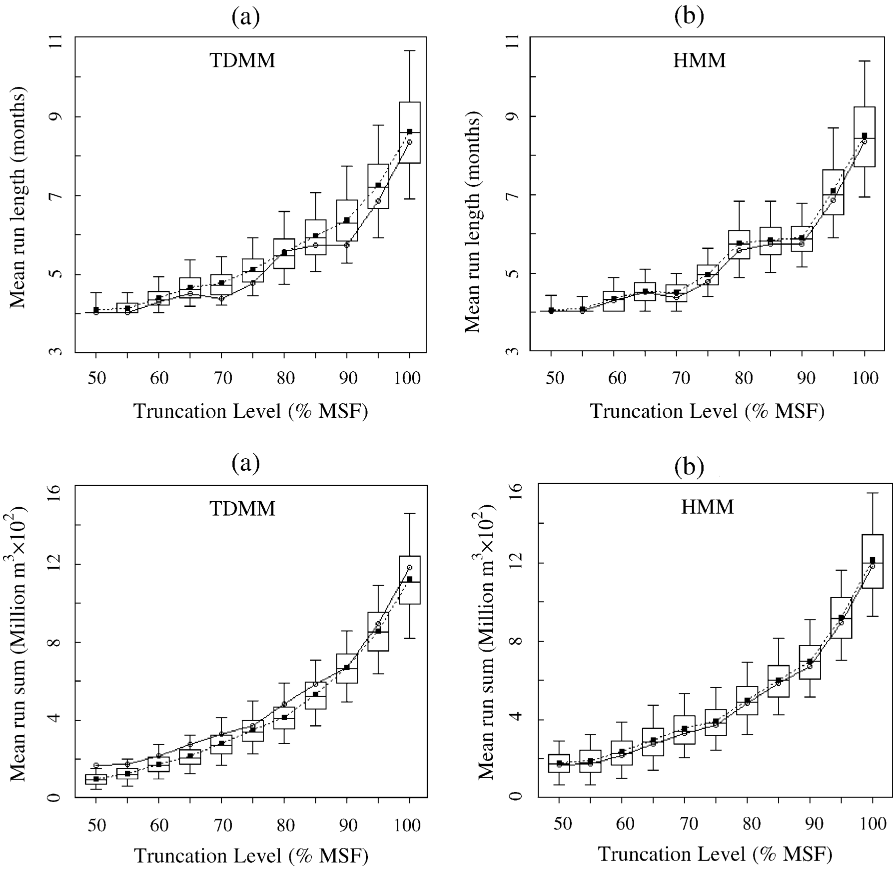

Fig. 16. Preservation of historical mean run characteristics based on seasonal truncations: a comparison between (a) TDMM and (b) HMM.

period is assumed to comprise of aggregated flow of last five dry months (January to May). Here too, TDMM is not able to model historical MARS (Fig. 18). Moreover, historical MERL and MERS are also not reproduced at most of the truncation levels (Fig. 19). In contrast, HMM exhibits a reasonable performance in reproducing the critical run characteristics (Fig. 18), and is able to reproduce the mean run characteristics very well (Fig. 19).

4. Summary and conclusions

A new hybrid stochastic model, termed Hybrid Moving block bootstrap Multi-site model (HMM), has been introduced for simulating multi-site multi-season streamflows. This model uses a parsimonious periodic parametric model without any normalizing transformation (PAR(1)NT) for partial pre-whitening of streamflows at each site. Then, the resulting residuals are resampled contemporaneously using MBB (NP method), with a view to reproduce site-to-site cross-correlations.

The performance of HMM in modeling the multi-site periodic streamflows is compared with that of Tasker and Dunne's multi-site model (TDMM) in reproducing a wide range of statistical attributes and multi-site drought characteristics. The hybrid model is seen to exhibit a better performance compared to TDMM in reproducing critical and mean multi-site drought characteristics apparently due to the better reproduction of summary statistics, marginal distribution features, cross-correlations and annual dependence.

By its very construct, the hybrid model is quite simple to implement and it offers enough flexibility to

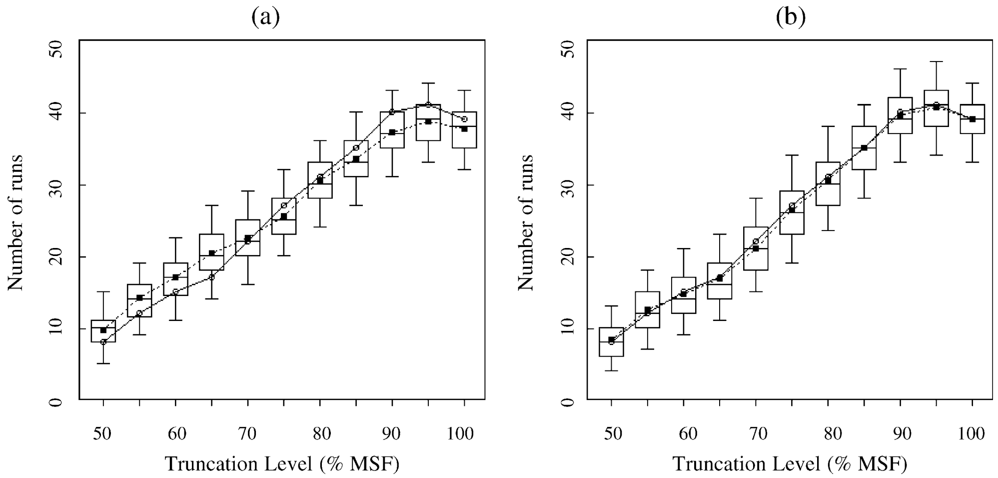

Fig. 17. Preservation of historical number of runs based on seasonal truncations: a comparison between (a) TDMM and (b) HMM.

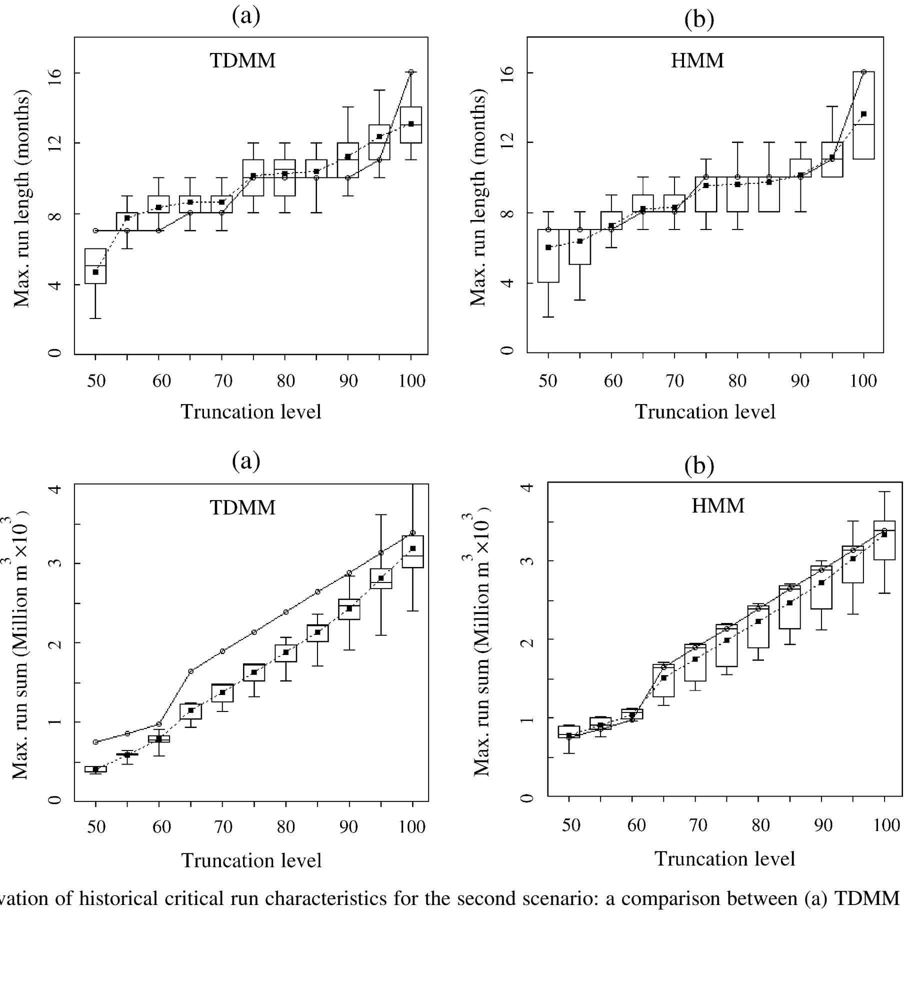

Fig. 18. Preservation of historical critical run characteristics for the second scenario: a comparison between (a) TDMM and (b) HMM.

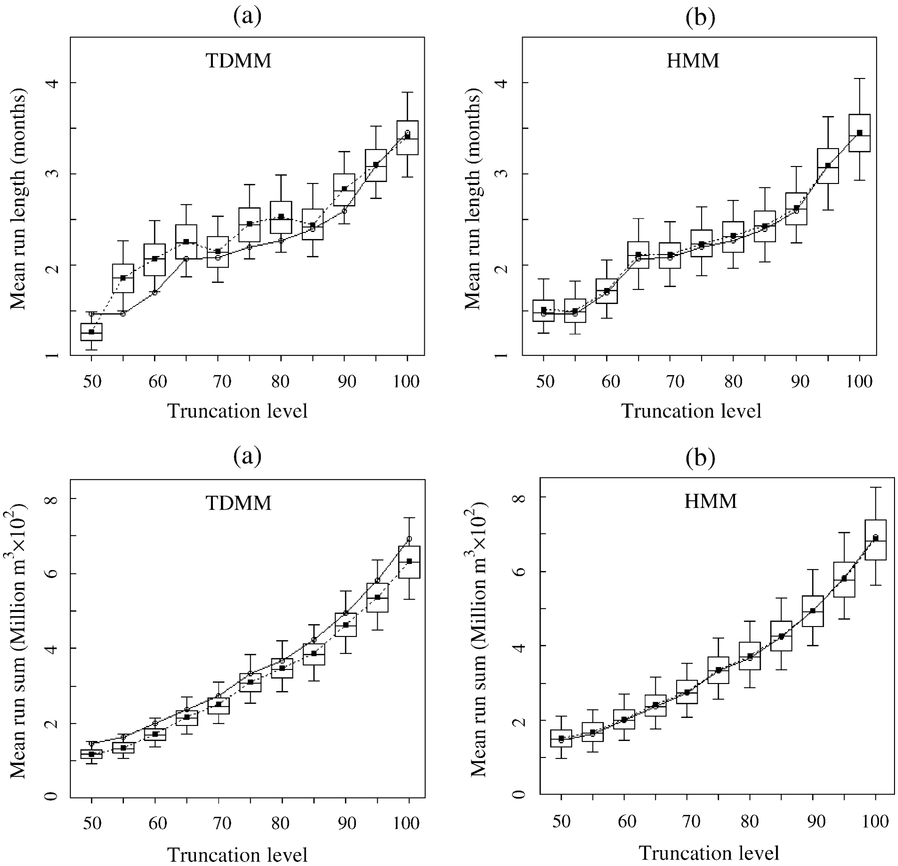

Fig. 19. Preservation of historical mean run characteristics for the second scenario: a comparison between (a) TDMM and (b) HMM.

the modeler for decision making in water resources studies. Several avenues should be explored further to refine this attempt to streamflow modeling. Although the hybrid model presented in this paper uses a simple PAR(1) model for partial pre-whitening of streamflows at each site and MBB for contemporaneous resampling of the residuals, other hybrid variants may also be tried. The HMM considers the block size of residuals for contemporaneous resampling to be an integral multiple of the number of periods in a hydrological water year. Especially, if long block sizes are used for resampling the residuals, variety and smoothing in the simulations may get reduced, which in turn, lessens the spread in the simulated run characteristics, thus affecting the design decisions. The modeler needs to exercise caution in this regard. The kinds of box plots used herein, would be helpful in deciding the block size. However, for resampling the residuals, if a better NP method can be identified, the hybrid model may offer better simulations especially from variety, smoothing and extrapolation points of view. Research in this direction is in progress.

Acknowledgements

The authors thank IISc Bangalore and IIT Madras, India, for providing the necessary facilities to carry out this research work. The authors express their gratitude to the Associate Editor Prof. Koutsoyiannis for his constructive comments and suggestions on the earlier draft of this manuscript, which were very helpful in improving the presentation and the organization of the article. Special thanks are also due to the Editor and an anonymous reviewer for their suggestions on the earlier version of this manuscript.

References

Bartolini, P., Salas, J.D., Obeysekera, J.T.B., 1988. Multivariate periodic ARMA(1,1) processes. Water Resources Research 24 (8), 1237–1246.

Camacho, F., McLeod, A.I., Hipel, K.W., 1983. The Use and Abuse of Multivariate Time Series in Hydrology, Technical Report 83-12, Department of Statistics and Actuarial Sciences, The University of Western Ontario, Canada.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and their Application. Cambridge University Press, Cambridge, MA.

Fiering, M.B., 1964. Multivariate technique for synthetic hydrology. Journal of Hydraulics Division ASCE 90 (HY5), 43–60.

Grygier, J.C., Stedinger, J.R., 1988. Condensed disaggregation procedures and conservation corrections for stochastic hydrology. Water Resources Research 24 (10), 1574–1584.

Grygier, J.C., Stedinger, J.R., 1990. SPIGOT, A Synthetic Streamflow Generation Software Package. Technical Description, Version 2.5, Cornell University, Ithaca, New York.

Haltiner, J.P., 1985. Stochastic Modeling of Seasonal and Daily Streamflow, PhD Thesis, Colorado State University, Fort Collins, CO.

Haltiner, J.P., Salas, J.D., 1988. Development and testing of a multivariate seasonal ARMA(1,1) model. Journal of Hydrology 104, 247–272.

Herbst, P.H., Bredenkamp, D.B., Barker, H.M.G., 1966. A technique for the evaluation of drought from rainfall data. Journal of Hydrology 4, 264–272.

Izzeldin, M., Murphy, A., 2000. Bootstrapping the small sample critical values of the rescaled range statistic. The Economic and Social Review 31 (4), 351–359.

Kendall, D.R., Dracup, J.A., 1991. A comparison of index-sequential and AR(1) generated hydrologic sequences. Journal of Hydrology 122 (1–4), 335–352.

Koutsoyiannis, D., 1992. A nonlinear disaggregation model with a reduced parameter set for simulation of hydrologic series. Water Resources Research 28 (12), 3175–3191.

Koutsoyiannis, D., 1999. Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology. Water Resources Research 35 (4), 1219–1229.

Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. Water Resources Research 36 (6), 1519–1533.

Koutsoyiannis, D., 2001. Coupling stochastic models of different time scales. Water Resources Research 37 (2), 379–391.

Koutsoyiannis, D., Manetas, A., 1996. Simple disaggregation by accurate adjusting procedures. Water Resources Research 32 (7), 2105–2117.

Kumar, D.N., Lall, U., Peterson, M.R., 2000. Multisite disaggregation of monthly to daily streamflow. Water Resources Research 36 (7), 1823–1833.

Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17 (3), 1217–1241.

Lall, U., 1995. Recent advances in nonparametric function estimation: hydrologic applications, US National Report to International Union of Geodesy and Geophysics 1991–1994. Reviews of Geophysics 33, 1093–1102.

Lall, U., Sharma, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. Water Resources Research 32 (3), 679–693.

Lane, W.L., 1979. Applied Stochastic Techniques (LAST Computer Package): User Manual. Division of Planning Technical Services, USBR, Denver, CO.

Lane, W.L., 1982. Corrected parameter estimates for disaggregation schemes, in: Singh, V.P. (Ed.), Statistical Analysis of Rainfall and Runoff. Water Resources Publications, Littleton, CO.

Lane, W.L., Frevert, D.K., 1990. Applied Stochastic Techniques, User's Manual, Personal Computer Version 5.2. Earth Sciences Division, USBR, Denver, CO.

Mejia, J.M., Rousselle, J., 1976. Disaggregation models in hydrology revisited. Water Resources Research 12 (2), 185–186.

Oliveira, G.C., Kelman, J., Pereira, M.V.F., Stedinger, J.R., 1988. A representation of spatial correlations in large stochastic seasonal streamflow models. Water Resources Research 24 (5), 781–785.

Ouarda, T.B.M.J., Labadie, J.W., Fontane, D.G., 1997. Indexed sequential hydrologic modeling for hydropower capacity estimation. Journal of the American Water Resources Association 33 (6), 1337–1349.

Pereira, M.V.F., Oliveira, G.C., Costa, C.C.G., Kelman, J., 1984. Stochastic streamflow models for hydroelectric systems. Water Resources Research 20 (3), 379–390.

Prairie, J.R., 2002. Long-Term Salinity Prediction with Uncertainty Analysis: Application for Colorado River above Glenwood Springs, Colorado, MS Thesis, University of Colorado, Boulder, CO.

Rajagopalan, B., Lall, U., 1999. A *k*-nearest neighbor simulator for daily precipitation and other weather variables. Water Resources Research 35 (10), 3089–3101.

Rasmussen, P.F., Salas, J.D., Fagherazzi, L., Rassam, J.-C., Bobee, B., 1996. Estimation and validation of contemporaneous PARMA models for streamflow simulation. Water Resources Research 32 (10), 3151–3160.

Salas, J.D., 1993. Analysis and modeling of hydrologic time series, in: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York (Chapter 19).

Salas, J.D., Pegram, G.G.S., 1977. A seasonal multivariate multilag auto-regressive model in hydrology, Proceedings of the Third International Symposium on Theoretical and Applied Hydrology. Colorado State University, Fort Collins, CO.

Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1980. Applied Modelling of Hydrologic Time Series. Water Resources Publications, Littleton, CO.

Salas, J.D., Boes, D.C., Smith, R.A., 1982. Estimation of ARMA models with seasonal parameters. Water Resources Research 18 (4), 1006–1010.

Salas, J.D., Tabios III., G.Q., Bartolini, P., 1985. Approaches to multivariate modeling of water resources time series. Water Resources Bulletin 21 (4), 683–708.

Santos, E.G., Salas, J.D., 1992. Stepwise disaggregation scheme for synthetic hydrology. Journal of Hydraulic Engineering, ASCE 118 (5), 765–784.

Sharma, A., Tarboton, D.G., Lall, U., 1997. Streamflow simulation: a nonparametric approach. Water Resources Research 33 (2), 291–308.

Srinivasan, K., Thandaveswara, B.S., Naik, P.C., 1992. Monthly streamflow data generation–a case study. Journal of Indian Water Resources Society, 12 (1 and 2) 29– 38.

Srinivas, V.V., Srinivasan, K., 2000. Post-blackening approach for modeling dependent annual streamflows. Journal of Hydrology 230 (1–2), 86–126.

Srinivas, V.V., Srinivasan, K., 2001a. Post-blackening approach for modeling periodic streamflows. Journal of Hydrology 241, 221–269.

Srinivas, V.V., Srinivasan, K., 2001b. A hybrid stochastic model for multiseason streamflow simulation. Water Resources Research 37 (10), 2537–2549.

Stedinger, J.R., Pei, D., 1982. An annual-monthly streamflow model for incorporating parameter uncertainty into reservoir simulation, in: El-shaarawi, A.H., Esterby, S.R. (Eds.), Time Series Methods in Hydrosciences, Development in Water Science, vol. 17. Elsevier, New York, pp. 520–529.

Stedinger, J.R., Vogel, R.M., 1984. Disaggregation procedures for generating serially correlated flow vectors. Water Resources Research 20 (1), 47–56.

Stedinger, J.R., Lattenmaier, D.P., Vogel, R.M., 1985. Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. Water Resources Research 21 (4), 497–509.

Tallaksen, L.M., Madsen, H., Clausen, B., 1997. On the definition and modelling of streamflow drought duration and deficit volume. Hydrological Sciences Journal 42 (1), 15–33.

Tarboton, D.G., Sharma, A., Lall, U., 1998. Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. Water Resources Research 34 (1), 107–119.

Tasker, G.D., Dunne, P., 1997. Bootstrap position analysis for forecasting low flow frequency. Journal of Water Resources Planning and Management ASCE 123 (6), 359–367.

Valencia, D.R., Schaake, J.C., 1973. Disaggregation processes in stochastic hydrology. Water Resources Research 9 (3), 580–585.

Vecchia, A.V., 1985a. Maximum likelihood estimation for periodic autoregressive moving average models. Technometrics 27 (4), 375–384.

Vecchia, A.V., 1985b. Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. Water Resources Bulletin 21 (5), 721–730.

Vogel, R.M., Shallcross, A.L., 1996. The moving blocks bootstrap versus parametric time series models. Water Resources Research 32 (6), 1875–1882.

WRDO, 1976. Master Plan for Equitable Use of Water Resources in the Cauvery River Basin—An Outline. Water Resources Development Organisation, Government of Karnataka, India, 13 vols.

Yevjevich, V., 1972. Stochastic Processes in Hydrology. Water Resources Publications, Fort Collins, CO.