ELSEVIER

Journal of Hydrology 241 (2001) 221–269

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Post-blackening approach for modeling periodic streamflows

V.V. Srinivas, K. Srinivasan*

Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600 036, India

Received 1 December 1999; revised 28 August 2000; accepted 5 October 2000

Abstract

The post-blackening (PB) approach introduced by the authors for modeling annual streamflows in an earlier work is extended to model periodic streamflows. This is basically a semi-parametric approach that blends a simple low-order, linear periodic parametric model with the moving block resampling scheme. The first part of the paper demonstrates the hybrid character of the PB model through Monte-Carlo simulations performed on hypothetical data sets drawn from a known population. Following this, the PB model is used for stochastic simulation of periodic streamflows of Beaver and Weber rivers in the US. The results show that the PB model is more consistent in reproducing a wide variety of statistics of periodic streamflows, compared to low-order linear periodic parametric models (Box–Jenkins type) and the periodic *k*-nearest-neighbor bootstrap (nonparametric) method. In addition, the PB model is able to preserve cross-year serial correlations as well as the month-to-year cross-correlations. This hybrid approach seems to offer considerable scope for improvement in hydrologic time series modeling.

Keywords: Streamflow modeling; Nonparametric; Bootstrap; Semi-parametric; Hybrid model

1. Introduction

Water resources planning studies essentially require modeling of uncertainty inherent in hydrologic inputs such as streamflow, precipitation and evaporation. This requires generation of synthetic sequences that are similar to the observed historical time series, in terms of summary statistics, marginal distribution of flows and dependence properties, and are able to preserve the storage/drought related statistics, consequently. For more than two decades, linear autoregressive moving average (ARMA) models (Box and Jenkins, 1976) have been used to model streamflows at single and multiple sites at the annual as well as the periodic levels and the same have been described at length in texts on the subject (e.g. Salas et al., 1980; Salas, 1993; Loucks et al., 1981; Bras and Rodriguez-Iturbe, 1985). In this approach, to account for the fact that the real streamflows may not be Gaussian, the flows are often first transformed to Gaussian and then the transformed flows are modeled using the ARMA(p,q) models (Stedinger, 1981; Stedinger and Taylor, 1982; Stedinger et al., 1985). The popularity of linear ARMA models for hydrologic time series analysis may be due to their simplicity, the availability of a well-developed modeling framework in the statistical literature for stationary processes and the availability of standard software packages such as SAS (1988); STATGRAPHICS (1984); IMSL (1984); SPIGOT (Grygier and Stedinger, 1990); CSUPAC1 (Salas et al., 1992); and LAST (Lane, 1979). However, there are a number of drawbacks of the Box–Jenkins type of models (Lall and Sharma,

* Corresponding author.
E-mail address: srini@civil.iitm.ernet.in (K. Srinivasan).

0022-1694/01/$ - see front matter
PII: S0022-1694(00)00363-2

1996; Srinivas and Srinivasan, 2000). The problems encountered in the case of fitting higher-order periodic models have been discussed by Rasmussen et al. (1996).

Furthermore, the incorporation of the parametric uncertainty into the parametric time series models (Stedinger and Taylor, 1982; Grygier and Stedinger, 1990) is quite involved and not that simple to be understood or applied by practicing hydrologists. Even though parametric nonlinear models (Bendat and Piersol, 1986; Tong, 1990) can be used in place of linear ARMA models to model nonlinear time series, it is essential to specify the form of nonlinear dependence, which may not be easy for the practitioner. From the practitioner's perspective, the key issues are: reproducibility of the observed data characteristics, simplicity, dependability and robustness. The recognition of nonlinearity of the underlying dynamics of geophysical processes, gain in computational ability and the availability of large data sets, in addition to the inherent simplicity, dependability and the ability to reproduce the characteristics of the historical data sets, have triggered the exploration of flexible, adaptive, data-driven nonparametric methods for hydrologic applications. Silverman (1986) and Scott (1992) provide introductory material on nonparametric methods. Readers are referred to Helsel and Hirsch (1992) and Lall (1995) for an overview of nonparametric applications to hydrology.

The bootstrap is a simple nonparametric technique for simulating the probability distribution of any statistic. Herein, the central idea is to resample from the original data, either directly or through a fitted model to create replicate data sets, from which the empirical probability distribution of the statistic of interest can be found (Davison and Hinkley, 1997). Indeed, the bootstrap offers the potential for highly accurate inferencing and can do away with the need to assume or impose a convenient model that may not have a rational basis. However, modeling dependent data poses a significant challenge that calls for new methods of resampling that would ensure the preservation of the temporal and the spatial covariance structure of the original time series. With the advent of powerful computers, bootstrap techniques are gaining importance in modern statistical analysis, especially in the field of time series analysis, as documented by Lepage and Billard (1992), Efron and Tibshirani (1993) Hjorth (1994) and Davison and Hinkley (1997).

Efron (1979) introduced the classical bootstrap resampling scheme which prescribes a data resampling strategy using the random mechanism that generated the data. Efron's bootstrap technique provides good approximation to the distribution of many commonly used statistics when the random variables are independent and identically distributed (i.i.d.), while it fails to model dependent random variables (Lahiri, 1995). Random resampling methods have been used in hydrology by Tasker (1987), Zucchini and Adamson (1988, 1989), Woo (1989) and Moss and Tasker (1991). Some of the popular bootstrapping techniques used to model dependent hydrologic time series data are: model based resampling (MBR); moving block bootstrap (MBB) and *k*-nearest-neighbor (*k*-NN) bootstrap.

Model based resampling for time series has been discussed by Freedman (1984), Freedman and Peters (1984), Efron and Tibshirani (1986, 1993) and Bose (1988), among others. In this method, to start with, a model structure is assumed, its parameters and residuals are estimated. Then the estimated model residuals are recentered around their mean and are resampled with replacement considering them as i.i.d.. Finally, these bootstrapped residuals are used to synthesize a time series. This approach preserves the empirical density function of the original time series. It is simple to apply and leads to good theoretical behavior, provided the fitted model is correct. The major drawback with this resampling scheme is that, in practice, the model structure is to be correctly identified and its parameters are to be estimated from the data. Otherwise, the resampled series will be generated from a wrong model, and hence they will not have the same statistical properties as that of the original data (Davison and Hinkley, 1997). In the context of hydrology, Pereira et al. (1984) and Oliveira et al. (1988) randomly resampled residuals from a multisite disaggregation model of lag-one annual streamflows. Tasker and Dunne (1997) have fitted a periodic autoregressive moving average model with log transformation (PARMA(1,1)-LT) to monthly stream flows and have used a nonoverlapping block bootstrap with a block size equal to 12 months (equal to 1 year in annual context) for resampling monthly residuals resulting from the periodic

model. The 12-month-long traces of monthly runoff generated have been used in position analysis model for a water supply storage and delivery system in central New Jersey, USA. This model can be viewed as "model based resampling applied to periodic data".

The other popular approach to resample in the time domain, known as the block bootstrap scheme, treats blocks of consecutive observations as exchangeable. This was developed by Hall (1985) and Carlstein (1986). Subsequently, Künsch (1989) and Liu and Singh (1992) independently proposed the MBB approach for time series analysis. The MBB provides "synthetic" time series that preserve the empirical probability distribution of the original observations. In this method, the key issue is to select the appropriate block size that generates synthetic replicates that are statistically indistinguishable from the historical trace. The idea that underlies the block resampling scheme is that if the blocks are long enough, the original dependence will be reasonably preserved in the resampled series. Clearly, this approximation is best if the dependence is weak and the blocks are as long as possible. Hausman (1990) and Vogel and Shallcross (1996) have used the MBB approach in the context of water resources planning studies. For the monthly streamflow modeling by the MBB method, the block size may be chosen to be an integral multiple of 12 months, in order to account for the within-year effects. During the process of resampling, the overlapping blocks that are picked at random are pasted end-to-end to form a replicate of the historical trace. In this method, the original dependence structure is maintained within the blocks, but is destroyed at boundaries between blocks. In other words, the within-year monthly serial correlations are well preserved, while the serial correlations between months of one water year to the previous (cross-year serial correlations), are not preserved by the MBB method. We find that even with higher block sizes, this basic drawback of the MBB method could not be overcome. Moreover, the simulations from MBB cannot fill in the gaps between the data points in the historical record. Also, they fail to generate extrema more severe than those found in the historical record. Furthermore, the value addition through simulations is limited. Due to the above drawbacks, the moving block bootstrap is not considered for modeling monthly streamflows in this study.

More recently, Lall and Sharma (1996) have introduced the *k*-NN bootstrap technique for modeling streamflows. Here, the dependence is preserved in a probabilistic sense. Multivariate nearest neighbor probability density estimation provides the basis for this method. This method involves searching the observed record to find the historical nearest neighbors and subsequently resampling their successors with a view to preserve the empirical dependence of the flow trace. This method uses only the historical trace (without any perturbations) in constructing the replicates, and as a result, the simulations from the *k*-NN method do not produce values that have not been observed in the historical data. This is a major limitation if extreme values outside the available record are of interest. The subsequent work of Sharma et al. (1997) avoids this limitation by resampling the historical data with perturbations. The perturbations serve to smooth over the gaps between the data points in the density estimate and provide alternate streamflow realizations that are different, but are statistically similar to the historical record (Sharma et al., 1997). However, since the streamflows are bounded, there is the possibility of leakage of probabilities across boundaries when the perturbation is added, and this may result in bias in the simulated density in the neighborhood of the boundary. In order to minimize this bias, appropriate kernel functions and/or bandwidths are to be chosen (Sharma et al., 1997), which may be a demanding task for any practicing hydrologist. The nearest-neighbor bootstrap technique and its variations are preferable if the data are plentiful, as in case of daily streamflow modeling (Lall and Sharma, 1996).

The generalized cross-validation (GCV) score function (Craven and Wahba, 1979) can be used to choose the number of nearest neighbors and the order of the *k*-NN model. This is somewhat similar to the use of Akaike information criteria (AIC) for model selection in the traditional ARMA modeling framework. A GCV-based choice of the model order and the number of neighbors may be suboptimal for the particular water resources planning study under consideration, since it only considers the performance of the model with respect to conditional mean and variance (Rajagopalan and Lall, 1999). This necessitates

further tuning to arrive at the appropriate combination of the model order and the number of neighbors for the study of interest.

Tarboton et al. (1998) have extended the use of nonparametric methods to disaggregation models. It is shown therein that a kernel density estimate of the joint distribution of disaggregation flow variables can form the basis for conditional simulation based on an input aggregate flow variable. The preservation of a variety of statistical attributes using this conditional simulation procedure has been demonstrated through applications to synthetic data and streamflows from the San Juan River in New Mexico, USA. Possibly due to the smoothing of the kernel density estimate, some amount of bias is observed in the standard deviations and skewnesses of the disaggregated monthly flows from this nonparametric disaggregation model (see Figs. 8 and 9, Tarboton et al., 1998). Even though the marginal distributions and the state-dependent correlations of observed flows are reported to be better preserved compared to SPIGOT (Fig. 11, Tarboton et al., 1998), further improvement is desirable. The drawbacks of this method as given by Tarboton et al. (1998) are:

(1) it is data and computationally intensive;

(2) estimating an optimal bandwidth to use is a computationally demanding task. As with the method of moments and the method of maximum likelihood in the parametric case, different optimality criteria can lead to quite different bandwidths being selected;

(3) the choice of kernel function is not crucial, but the parameterization of the bandwidth matrix in the multivariate case may affect the results dramatically;

(4) the sample size required increases, as the complexity of the underlying density function increases, thus reducing the advantage of the NPD approach for heterogeneous functions; and

(5) no simple equation for the model is available to report.

More recently, Rajagopalan and Lall (1999) have suggested a *k*-NN based multivariate nonparametric time series simulation method to generate random sequences of daily weather variables. Sharma and Lall (1999) have used *k*-NN for daily rainfall simulation and Kumar et al. (2000) have extended *k*-NN to multisite disaggregation of monthly to daily streamflows.

Understanding the relative merits as well as the limitations of both parametric and nonparametric methods, it is felt that a proper blend of the two methods might result in the generation of synthetic replicates that would represent better the observed hydrologic time series compared with the replicates generated from either of the methods. The *post-blackening* (PB) approach suggested by Davison and Hinkley (1997) is one such method that seems to blend the two methods effectively, and the authors have found that this would be very much suited for the hydrologic time series modeling. The first step in the PB approach is to "prewhiten" the historical trace by fitting a simple *parametric model* that is intended to remove much of the dependence present in the observations of the historical sequence. A series of innovations is then generated by *block resampling of the residuals* obtained from the fitted model, with a view to capture the weak dependence (if any) present in the residuals. The innovation series is then "post-blackened" by applying the estimated model to the resampled innovations. However, it is to be mentioned that the properties of this PB scheme have not yet been established theoretically (Davison, personal communication).

We wish to mention that this method has been introduced by the authors in an earlier paper (Srinivas and Srinivasan, 2000) for modeling annual streamflows with complex dependence. In this article, we extend this scheme to periodic streamflow modeling. Our interest is to investigate the *performance* of the PB approach in terms of preservation of the following characteristics of historical periodic streamflows:

(1) summary statistics;

(2) within-year and cross-year serial correlations;

(3) marginal distributions; and

(4) state-dependent correlations, and compare the same with the low-order linear periodic parametric models and the *k*-NN bootstrap method of Lall and Sharma (1996). We also intend to examine the performance of the PB model in terms of preservation of cross-correlations between various months within the year as well as the month-to-year cross-correlations. For the purpose of the synthetic simulation, the monthly streamflow records of: (i) *Weber River* near Oakley, UT, located at 40°44′10″N latitude, 111°14′45″W longitude and at an elevation of 2012 m above mean sea level; and (ii) *Beaver River* near Beaver, UT, located at 38°16′50″N latitude, 112°34′25″W longitude and at an elevation of 1890 m above mean sea level, have been used. The data have been extracted from the United

States Geological Survey (USGS) hydro climate data network (HCDN) CD-ROM. Table 3 presents the salient features of these two rivers. Even though a number of monthly streamflow data sets have been tested by the authors, the modeling results are presented and discussed herein only for the Weber and the Beaver river streamflows. The reasons for selecting these two rivers are: (i) they display a complex dependence structure extending over a number of lags; (ii) the month-to-annual cross-correlations are also significant in a number of months; (iii) there is a variety of monthly marginal distribution patterns; (iv) the dry and the wet periods within the year are well defined; (v) reasonable lengths of streamflow records are available; and (vi) these two data sets have been used recently in literature for nonparametric modeling (Lall and Sharma, 1996; Sharma et al., 1997).

For the sake of comparison:

1. For the parametric approach, the best of the commonly used low-order periodic parametric models of Box-Jenkins type (PAR(1), PAR(2), PARMA(1,1)), is considered. Natural logarithm and Wilson–Hilferty transformations (WHT) are considered for modeling.
2. For the k-NN method, the number of nearest neighbors (k) is taken to be equal to the square root of the sample size and the model order (d) adopted is equal to 1 (as given in Lall and Sharma, 1996). This was adopted after investigating/testing the performance of the k-NN method over different combinations of k and d for the two rivers selected. Our investigations have shown that increase in the model order ($d > 1$) results in further underestimation of the standard deviation of the historical flows. In addition, it has also been observed by the authors that the first few lag serial correlations get distorted in an effort to preserve higher lag serial correlations.
3. For the PB approach, a simple PAR(1) model (a default option/model commonly adopted by stochastic hydrologist to model the periodic streamflows) is used as the underlying parametric model. The block size of MBB adopted for residual resampling is decided based on visual inspection of box plots (that indicate the spread) of various statistics.

It is to be noted that for the k-NN and the PB models, no normalizing transformation is applied to the observed streamflow data.

2. Model structure

This section describes the model structure of the three models considered. In these descriptions, vectors will be represented by bold upper case letters.

2.1. Periodic autoregressive moving average (PARMA) models

Low-order periodic autoregressive moving average (PARMA(p,q)) models are extensively used for modeling periodic hydrologic time series (Tao and Delleur, 1976; Hirsch, 1979; Salas et al., 1980).

The general structure of the PARMA(p,q) model is given by:

$$Y_{v,\tau} = \sum_{j=1}^{p} \varphi_{j,\tau} Y_{v,\tau-j} + \epsilon_{v,\tau} - \sum_{k=1}^{q} \theta_{k,\tau} \epsilon_{v,\tau-k} \tag{1}$$

where $v = 1, \ldots, N$ and $\tau = 1, \ldots, \omega$; where v is the index for year ($v = 1, \ldots, N$) and τ denotes the index for month within the year ($\tau = 1, \ldots, \omega$), N refers to the number of years of historical record and ω represents number of months within the year; $Y_{v,\tau}$ is the time series already suitably transformed (to satisfy the normality assumption) and standardized (deseasonalized). $\varphi_{1,\tau}, \ldots, \varphi_{p,\tau}$ are autoregressive parameters and $\theta_{1,\tau}, \ldots, \theta_{q,\tau}$ are moving average parameters; $\{\epsilon_{v,\tau}\}$ is the error or noise term assumed to be uncorrelated and has zero mean and variance $\sigma_\tau^2(\epsilon)$.

The key steps involved in the generation of synthetic streamflows by periodic streamflow modeling are (Salas et al., 1980):

1. Transformation of the observed (historical) streamflows to satisfy normality assumption followed by standardization (deseasonalization).
2. Identification of the appropriate model form using autocorrelation function (ACF) and partial autocorrelation function (PACF).
3. Estimation of the model parameters (method of moments is used in this study).
4. Diagnostic checking of residuals: test for independence; test for normality; skewness test for individual months.

5. Generation of synthetic replicates (in the transformed domain) each of size equal to that of the observed sample.
6. Inverse standardization of the replicates generated, followed by inverse transformation, to the original flow domain.

For parameter estimation and diagnostic checking of residuals of the low-order periodic models considered in this study (PAR(1), PAR(2), PARMA(1,1)), the CSU001 program (Salas et al., 1992) has been used.

2.2. The k-NN resampling algorithm

The *k*-NN bootstrap method for resampling hydrologic time series was proposed by Lall and Sharma (1996). This method has been developed for dependent data and it preserves the dependence in a probabilistic sense. Multivariate nearest-neighbor probability density estimation provides the basis for this method.

Let the time series of historical streamflows be denoted by $\mathbf{Q}_{v,\tau}$, where v is the index for year ($v = 1,\ldots,N$) and τ denotes the index for month within the year ($\tau = 1,\ldots,\omega$), N refers to the number of years of historical record and ω represents number of months ($=12$) within the year.

Let us say that the hydrological water year starts with the month of October of a calendar year and ends with the month of September of the subsequent calendar year. Now, the first value to be simulated will be October's flow. For this, one has to pick randomly any one of the N October flows from the historical data. Let it be denoted as $q_{i,1}$, where i is the water year to which the flow value belongs.

Following are the sequential steps involved in the synthetic simulation of historical streamflow data using the method, as given by Lall and Sharma (1996).

1. Define the composition of the "feature vector" of dimension d. For example, for order of dependence equal to two ($d = 2$), initial feature vector for simulating October's flow will be the conditioning set $\{q_{i-1,\omega}, q_{i-1,\omega-1}\}$. This represents the dependence of the October flow to be simulated on two prior monthly flows (i.e August and September flows of the previous water year). The historical state vectors $\mathbf{D}_\tau$ for any month τ, are the feature vectors of all $q_{v,\tau}$ in the historical record. For example, for simulating October's flow, the historical state vectors will be: $\{q_{1,\omega}, q_{1,\omega-1}\},\ldots,\{q_{N,\omega}, q_{N,\omega-1}\}$.
2. Denote the current feature vector as $\mathbf{D}_i$ and determine its k-nearest-neighbors from among the historical state vectors for that month $\mathbf{D}_\tau$, using the weighted Euclidean distance r_{iv}.

$$r_{iv} = \left(\sum_{j=1}^{d} w_j (\mathrm{v}_{ij} - \mathrm{v}_{\tau j})^2 \right)^{1/2} \tag{2}$$

   In Equation (2), r_{iv} is the weighted Euclidean distance from the current feature vector to the vth historical state vector among the historical state vectors $\mathbf{D}_\tau$ for the month τ; v_{ij} is the jth component of the current feature vector; and v_{vj} stands for the jth component of the vth historical state vector. The weights w_j are chosen a priori as inverse of some measure of scale such as standard deviation or range of $\mathbf{V}_j$. Where, $\mathbf{V}_j$ is a set comprising of the jth components of the historical state vectors $\mathbf{D}_\tau$ for the month τ. The number of neighbors k is a smoothing parameter. It may be chosen using any appropriate order selection strategy such as GCV (Craven and Wahba, 1979). Lall and Sharma (1996) suggest using k equal to square root of the sample size as a rule of thumb.
3. Denote the ordered set of nearest-neighbor indices by $J_{i,u}$, where $u = 1,\ldots,k$. An element $j(i)$ of this set records the time v associated with the jth closest historical state vector $\mathbf{D}_i$ to $\mathbf{D}_\tau$. Denote $x^{s}_{j(i)}$ as the successor to $\mathbf{D}_{j(i)}$. If the data are highly quantized, it is possible that a number of observations may be at the same distance from the conditioning point, in which case a permuting may help.
4. Define a discrete kernel $K(j(i))$ for resampling one of the $x^{s}_{j(i)}$ as follows:

$$K(j(i)) = \frac{1/j}{\sum_{j=1}^{k} 1/j} \tag{3}$$

   where $K(j(i))$ is the probability with which $x^{s}_{j(i)}$ is resampled. It is to be noted that this resampling kernel is the same for any i, and can be computed

and stored prior to the start of simulation. Lall and Sharma (1996) develop this kernel through a local Poisson approximation of the probability density function of state space neighbors.

5. Using the discrete probability mass function $K(j(i))$, resample an $x^{s}_{j(i)}$ and update the current feature vector. Proceed to step 2 if additional simulated values are required to be generated. For a more detailed discussion on the k-NN algorithm, the reader is referred to Lall and Sharma (1996) and Rajagopalan and Lall (1999).

2.3. PB approach

This section presents a new algorithm for generating synthetic monthly streamflows, using the PB approach suggested by Davison and Hinkley (1997).

Let the observed (historical) streamflows be represented by the vector $\mathbf{Q}_{v,\tau}$, where 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 refers to the number of years of historical record and ω represents the number of periods within the year. The modeling steps involved are as follows:

1. Standardize the elements of the vector $\mathbf{Q}_{v,\tau}$ as:

$$y_{v,\tau} = \frac{q_{v,\tau} - \bar{q}_{\tau}}{s_{\tau}} \tag{4}$$

where $\bar{q}_{\tau}$ and s_{τ} are, respectively, the mean and the standard deviation of the observed streamflows in period τ. Note that the historical streamflows are not transformed to remove skewness.

2. Prewhiten the standardized historical streamflows, $\mathbf{Y}_{v,\tau}$, using a simple periodic autoregressive model of order one (PAR(1)), and extract the residuals $\epsilon_{v,\tau}$.

$$\epsilon_{v,\tau} = y_{v,\tau} - \varphi_{1,\tau} y_{v,\tau-1} \tag{5}$$

In Eq. (5), $\varphi_{1,1}, \ldots, \varphi_{1,\varpi}$ are the periodic autoregressive parameters of order one. Herein, for the parameter estimation, a simple method of moments (Salas et al., 1980) has been used. It is to be noted that the residuals $\epsilon_{v,\tau}$ may possess some weak dependence (since the parameters are estimated from a simple PAR(1) model). Herein, we wish to mention that bootstrap schemes like MBB (Künsch, 1989) can serve as a reliable tool for modeling the weak linear dependence, if any, in the residuals. Moreover, this scheme being data-driven can be expected to capture the marginal distribution features, and to a certain extent may be able to preserve the nonlinear dependence inherent in the observed record, possibly with some trade-off with regard to smoothing and generation of extrema, compared to historical record. The gains due to hybrid effect are discussed in Section 3.

3. Obtain the simulated innovations $\epsilon^{*}_{v,\tau}$. by bootstrapping $\epsilon_{v,\tau}$. using the MBB (Künsch, 1989) method. Herein, the monthly residuals resulting from the PAR(1) model are divided into q number of (possibly) overlapping blocks B_i with block size L taken as an integral multiple of the number of periods (ω) within the year. It is to be noted that each of the overlapping blocks starts with the first period of a hydrological water year. This is done with a view to capture the within-year correlations for significant number of lags. For example, the block sizes of residuals in monthly streamflow modeling context would be 12, 24, 36, and so on (abbreviated as $L = \omega$, $L = 2\omega$, $L = 3\omega$, and so on). Note that when blocklength L is n years long, the overlap is $(n - 1)$ years, indicating that when the block size is 1 year long, there is no overlap.

In general, the ith block with size $L = m\omega$ (m is a positive integer, such that, $m = 1, \ldots, N$), may be written as: $B_i = (\epsilon_{i,1}, \ldots, \epsilon_{i+m-1,\omega})$ where $i = 1, \ldots, b$ and $b = N - m + 1$.

For example, if $L = 3\omega$ and $\omega = 12$, the fourth block is written as: $B_4 = (\epsilon_{4,1}, \ldots, \epsilon_{6,12})$.

The block size L, to be selected for resampling the residuals, would primarily depend on the amount of unextracted weak dependence present in the residuals. Innovations $\epsilon^{*}_{v,\tau}$, are generated by resampling the overlapping blocks B_i, at random, with replacement from the set $(B_1, \ldots, B_b)$ and pasting them end-to-end. It is to be noted that each of the (possibly) overlapping blocks has equal probability $(1/b)$ of being resampled.

4. The innovation series $\epsilon^{*}_{v,\tau}$ is then post-blackened by reversing Eq. (5), to obtain the sequence $\mathbf{Z}_{v,\tau}$ (Eq. (6)).

$$z_{v,\tau} = \varphi_{1,\tau}\, z_{v,\tau-1} + \epsilon^{*}_{v,\tau} \tag{6}$$

The synthetic generation process starts with $z_{1,0} = 0$. The "warm-up" period is chosen to be large enough to remove any initial bias. The values of $\mathbf{Z}_{v,\tau}$ are then inverse standardized (using Eq. (7)) to obtain the synthetic streamflow replicate $\mathbf{X}_{v,\tau}$.

$$x_{v,\tau} = (z_{v,\tau} s_\tau) + \bar{q}_\tau \tag{7}$$

It is to be noted that no normalizing transformation is applied in case of the PB model. Herein, we wish to mention that when the number of data points in the historical record is limited (as in case of annual streamflow modeling), the mean of residuals recovered from the prewhitening stage need not be necessarily equal to zero. In such a case, the residuals are to be recentered to zero before proceeding with resampling them for generating the innovation series (See Davison and Hinkley, 1997, p 397). However, when the data points are relatively plentiful (as in case of periodic streamflow modeling), we find that the sum of residuals recovered from partial prewhitening stage tends to zero and hence the residuals need not be recentered.

3. Hybrid effect

In this section of the article, we attempt to gain some understanding into the roles played by the two constituents of the PB model, namely, a simple low-order linear periodic parametric model (PAR(1)NT) and the MBB, in enhancing its performance. It is also of interest to examine if the PB model can recover statistical attributes and dependence structure from small samples generated from a known population. To effect this, a two-level Monte-Carlo simulation experiment is designed using hypothetical data sets generated from a known self-exciting seasonal threshold ARMA model. In the first level, 100 samples, each of length N years × four seasons, are generated from a known self-exciting seasonal threshold ARMA model. These 100 samples are referred to as level-1 samples. The second level involves generating 100 replicates, each of size N years × 4 seasons, for each of the 100 level-1 samples, using each of the three models namely: (i) a simple periodic autoregressive model of order one, with no normalizing transformation (PAR(1)NT); (ii) MBB; and (iii) PB. The 10,000 replicates resulting from the Monte-Carlo simulations at the second level are referred to as level-2 replicates. These simulation experiments are done for three typical sample sizes ($N = 40$, 60 and 80), with a view to appreciate the effect of sample size on the performance of the PB model.

The self-exciting seasonal threshold ARMA model used for generating level-1 samples is:

First season:

$$x_t = 0.2x_{t-1} + 0.35x_{t-2} + 0.6W_t \quad \text{if } x_{t-1} \le 0$$
$$x_t = 0.9x_{t-1} - 0.16W_{t-1} + 0.6W_t \quad \text{otherwise}$$

Second season:

$$x_t = 0.5x_{t-1} - 0.12W_{t-1} + 0.7W_t \quad \text{if } x_{t-1} \le 0$$
$$x_t = 0.9x_{t-1} + 0.2x_{t-2} + 0.7W_t \quad \text{otherwise}$$

Third season:

$$x_t = 0.45x_{t-1} - 0.245W_{t-1} + 0.5W_t \quad \text{if } x_{t-1} \le 0$$
$$x_t = 0.15x_{t-1} + 0.3x_{t-2} + 0.5W_t \quad \text{otherwise}$$

Fourth season:

$$x_t = -1.0 - 0.5x_{t-1} - 0.1W_{t-1} + 0.8W_t \quad \text{if } x_{t-1} \le 0$$
$$x_t = 0.8x_{t-1} + 0.2x_{t-2} + 0.8W_t \quad \text{otherwise}$$

where W_t is a Gaussian random variate with zero mean and unit standard deviation.

The performance of the PB model is examined in terms of reproduction of the following statistics: (i) summary statistics (mean, standard deviation and skewness) at both seasonal and annual levels; (ii) marginal distributions at seasonal and annual levels; (iii) serial correlations (both within-year and cross-year) at the seasonal level; (iv) autocorrelations at the aggregated annual level; and (v) state-dependent correlations (Sharma et al., 1997) at both seasonal and annual levels (used as a measure of nonlinear dependence). To enable better appreciation of the hybrid character of PB model, the results from this model are presented alongside those from its constituent models (PAR(1)-NT and MBB) in Tables

Table 1
Preservation of average seasonal statistical attributes over 100 level-1 samples by 10,000 level-2 replicates from the AR(1)NT, MBB and PB models (L, block size used for resampling; SC, serial correlation; AMF, above median and forward correlation; BMF, below median and forward correlation; AMB, above median and backward correlation; BMB, below median and backward correlation)

Statistic	Season	$N = 40$				$N = 60$				$N = 80$			
		Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$
Mean	1	0.305	0.302[a]	0.319	0.317	0.294	0.303	0.295	0.293	0.289	0.288	0.294	0.294
			(0.276)[b]	(0.291)	(0.295)		(0.206)	(0.225)	(0.227)		(0.179)	(0.190)	(0.193)
	2	0.457	0.449	0.470	0.469	0.445	0.445	0.443	0.442	0.440	0.435	0.447	0.447
			(0.317)	(0.333)	(0.335)		(0.251)	(0.268)	(0.270)		(0.215)	(0.226)	(0.228)
	3	0.187	0.178	0.196	0.195	0.181	0.170	0.179	0.179	0.180	0.176	0.181	0.181
			(0.184)	(0.187)	(0.189)		(0.145)	(0.152)	(0.154)		(0.124)	(0.130)	(0.131)
	4	0.107	0.100	0.117	0.116	0.086	0.094	0.080	0.080	0.086	0.092	0.087	0.088
			(0.315)	(0.333)	(0.338)		(0.263)	(0.273)	(0.267)		(0.228)	(0.242)	(0.245)
Std. Dev.	1	0.870	0.871	0.846	0.843	0.869	0.851	0.852	0.849	0.870	0.862	0.859	0.856
			(0.176)	(0.176)	(0.173)		(0.142)	(0.141)	(0.138)		(0.113)	(0.122)	(0.118)
	2	1.048	1.051	1.021	1.008	1.043	1.041	1.021	1.007	1.035	1.029	1.023	1.010
			(0.205)	(0.216)	(0.209)		(0.161)	(0.169)	(0.165)		(0.131)	(0.143)	(0.138)
	3	0.600	0.592	0.585	0.579	0.594	0.583	0.583	0.577	0.600	0.589	0.592	0.586
			(0.103)	(0.105)	(0.103)		(0.080)	(0.082)	(0.081)		(0.069)	(0.073)	(0.071)
	4	1.095	1.096	1.071	1.063	1.085	1.066	1.067	1.060	1.084	1.070	1.074	1.067
			(0.200)	(0.188)	(0.185)		(0.161)	(0.156)	(0.154)		(0.130)	(0.135)	(0.134)
Skew	1	0.317	−0.007	0.270	0.210	0.332	−0.003	0.290	0.226	0.356	0.009	0.326	0.257
			(0.411)	(0.482)	(0.453)		(0.298)	(0.396)	(0.377)		(0.278)	(0.358)	(0.331)
	2	0.434	−0.040	0.361	0.296	0.420	−0.019	0.377	0.312	0.432	−0.018	0.395	0.326
			(0.365)	(0.500)	(0.481)		(0.291)	(0.401)	(0.383)		(0.266)	(0.378)	(0.356)
	3	0.168	0.034	0.128	0.106	0.141	0.017	0.117	0.100	0.160	0.038	0.153	0.133
			(0.381)	(0.490)	(0.478)		(0.320)	(0.414)	(0.411)		(0.260)	(0.378)	(0.368)
	4	0.054	0.008	0.025	0.012	0.047	−0.009	0.027	0.017	0.053	0.013	0.048	0.037
			(0.387)	(0.469)	(0.466)		(0.303)	(0.376)	(0.381)		(0.251)	(0.329)	(0.329)
Lag-1 SC	1	0.675	0.674	**.539**	0.662	0.673	0.664	**.535**	0.660	0.680	0.677	**.544**	0.672
			(0.143)	(**0.158**)	(0.147)		(0.116)	(**0.128**)	(0.123)		(0.089)	(**0.106**)	(0.097)
	2	0.693	0.696	0.677	0.669	0.694	0.688	0.680	0.672	0.697	0.691	0.690	0.682
			(0.131)	(0.150)	(0.150)		(0.105)	(0.118)	(0.120)		(0.089)	(0.097)	(0.097)
	3	0.470	0.453	0.455	0.439	0.471	0.450	0.454	0.439	0.474	0.456	0.467	0.451
			(0.204)	(0.205)	(0.206)		(0.164)	(0.160)	(0.161)		(0.136)	(0.141)	(0.143)
	4	0.583	0.585	0.573	0.567	0.579	0.574	0.567	0.561	0.579	0.572	0.575	0.568
			(0.153)	(0.157)	(0.156)		(0.125)	(0.124)	(0.126)		(0.103)	(0.105)	(0.105)
Lag-2 SC	1	0.535	0.416	**.418**	0.487	0.532	0.397	**.417**	0.485	0.538	0.399	**.423**	0.494
			(0.186)	(**0.178**)	(0.179)		(0.148)	(**0.144**)	(0.147)		(0.124)	(**0.125**)	(0.125)
	2	0.575	0.475	**.454**	0.543	0.569	0.459	**.447**	0.538	0.572	0.466	**.456**	0.549
			(0.171)	(**0.176**)	(0.175)		(0.137)	(**0.141**)	(0.143)		(0.118)	(**0.122**)	(0.119)
	3	0.517	0.314	0.503	0.481	0.514	0.304	0.500	0.478	0.523	0.318	0.513	0.490

Table 1 (*continued*)

Statistic	Season	$N = 40$				$N = 60$				$N = 80$			
		Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$
			(0.191)	(0.187)	(0.185)		(0.171)	(0.153)	(0.152)		(0.130)	(0.136)	(0.136)
	4	0.394	0.266	0.373	0.358	0.393	0.263	0.374	0.358	0.392	0.263	0.382	0.366
			(0.193)	(0.219)	(0.215)		(0.152)	(0.179)	(0.179)		(0.135)	(0.156)	(0.153)
Lag-3 SC	1	0.351	0.185	**.255**	0.293	0.351	0.193	**.258**	0.294	0.350	0.177	**.269**	0.306
			(0.173)	(**0.199**)	(0.205)		(0.140)	(**0.163**)	(0.171)		(0.123)	(**0.145**)	(0.150)
	2	0.441	0.298	**.339**	0.388	0.435	0.282	**.335**	0.383	0.438	0.281	**.343**	0.393
			(0.185)	(**0.193**)	(0.196)		(0.146)	(**0.156**)	(0.161)		(0.127)	(**0.139**)	(0.141)
	3	0.389	0.204	**.297**	0.340	0.387	0.209	**.296**	0.340	0.394	0.217	**.307**	0.351
			(0.195)	(**0.198**)	(0.204)		(0.168)	(**0.159**)	(0.165)		(0.127)	(**0.139**)	(0.143)
	4	0.387	0.184	0.368	0.345	0.388	0.174	0.370	0.345	0.391	0.187	**.380**	0.355
			(0.198)	(0.228)	(0.221)		(0.163)	(0.186)	(0.182)		(0.133)	(**0.155**)	(0.151)
Lag-4 SC	1	0.338	0.123	**.239**	0.269	0.337	0.132	**.239**	0.267	0.345	0.120	**.258**	0.286
			(0.172)	(**0.191**)	(0.195)		(0.133)	(**0.159**)	(0.166)		(0.112)	(**0.140**)	(0.143)
	2	0.299	0.135	**.199**	0.222	0.297	0.139	**.207**	0.229	0.296	0.120	**.218**	0.243
			(0.163)	(**0.200**)	(0.205)		(0.139)	(**0.162**)	(0.168)		(0.114)	(**0.142**)	(0.145)
	3	0.295	0.113	**.205**	0.228	0.301	0.142	**.218**	0.240	0.314	0.138	**.232**	0.255
			(0.184)	(**0.194**)	(0.198)		(0.148)	(**0.160**)	(0.165)		(0.113)	(**0.136**)	(0.138)
	4	0.275	0.100	**.190**	0.216	0.285	0.109	**.202**	0.224	0.290	0.121	**.212**	0.237
			(0.183)	(**0.202**)	(0.208)		(0.148)	(**0.161**)	(0.167)		(0.120)	(**0.148**)	(0.153)
AMF	1	0.683	0.515	0.651	0.619	0.684	0.513	0.658	0.624	0.684	0.506	0.671	0.637
			(0.206)	(0.211)	(0.206)		(0.164)	(0.176)	(0.176)		(0.143)	(0.137)	(0.138)
	2	0.430	0.302	0.407	0.382	0.438	0.280	0.421	0.397	0.451	0.286	0.436	0.409
			(0.228)	(0.270)	(0.261)		(0.200)	(0.217)	(0.211)		(0.165)	(0.192)	(0.187)
	3	0.436	0.441	0.418	0.416	0.425	0.433	0.424	0.426	0.438	0.423	0.431	0.429
			(0.216)	(0.292)	(0.281)		(0.191)	(0.235)	(0.232)		(0.162)	(0.213)	(0.208)
	4	0.634	0.486	**.466**	0.577	0.638	0.467	**.476**	0.589	0.650	0.493	**.490**	0.605
			(0.222)	(**0.247**)	(0.219)		(0.158)	(**0.197**)	(0.171)		(0.141)	(**0.167**)	(0.140)
BMF	1	0.241	0.493	0.242	0.293	0.233	0.488	0.227	0.279	0.231	0.491	0.233	0.285
			(0.215)	(0.303)	(0.277)		(0.166)	(0.250)	(0.226)		(0.134)	(0.225)	(0.201)
	2	0.140	0.319	0.156	0.170	0.131	0.295	0.136	0.151	0.145	0.309	0.147	0.160
			(0.246)	(0.308)	(0.291)		(0.206)	(0.262)	(0.247)		(0.163)	(0.217)	(0.204)
	3	0.065	0.379	0.069	0.097	0.009	0.385	0.017	0.046	0.032	0.387	0.030	0.054
			(0.209)	(0.377)	(0.360)		(0.170)	(0.324)	(0.313)		(0.141)	(0.292)	(0.280)
	4	0.273	0.478	**.203**	0.326	0.273	0.480	**.194**	0.316	0.272	0.479	**.193**	0.315
			(0.212)	(**0.277**)	(0.272)		(0.161)	(**0.218**)	(0.219)		(0.139)	(**0.193**)	(0.193)
AMB	1	0.610	0.497	**.445**	0.557	0.605	0.487	**.447**	0.565	0.613	0.498	**.458**	0.575
			(0.216)	(**0.260**)	(0.233)		(0.157)	(**0.207**)	(0.187)		(0.148)	(**0.170**)	(0.150)
	2	0.639	0.517	0.614	0.592	0.643	0.503	0.622	0.595	0.646	0.503	0.635	0.607
			(0.198)	(0.219)	(0.211)		(0.173)	(0.173)	(0.175)		(0.153)	(0.148)	(0.147)
	3	0.379	0.306	0.357	0.339	0.388	0.281	0.365	0.346	0.400	0.283	0.386	0.366

Table 1 (*continued*)

Statistic	Season	$N = 40$				$N = 60$				$N = 80$			
		Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$
			(0.222)	(0.287)	(0.277)		(0.178)	(0.227)	(0.225)		(0.152)	(0.207)	(0.200)
	4	0.444	0.403	0.430	0.425	0.452	0.408	0.430	0.424	0.436	0.394	0.433	0.424
			(0.214)	(0.271)	(0.263)		(0.173)	(0.220)	(0.217)		(0.153)	(0.194)	(0.189)
BMB	1	0.346	0.486	**.258**	0.380	0.336	0.486	**.254**	0.368	0.325	0.488	**.241**	0.359
			(0.211)	(**0.266**)	(0.251)		(0.162)	(**0.216**)	(0.209)		(0.141)	(**0.194**)	(0.188)
	2	0.288	0.498	0.292	0.334	0.284	0.498	0.285	0.323	0.300	0.493	0.300	0.337
			(0.212)	(0.289)	(0.264)		(0.165)	(0.247)	(0.226)		(0.130)	(0.208)	(0.191)
	3	0.168	0.295	0.171	0.183	0.186	0.293	0.189	0.200	0.194	0.301	0.190	0.197
			(0.251)	(0.293)	(0.284)		(0.192)	(0.244)	(0.237)		(0.159)	(0.214)	(0.204)
	4	0.212	0.388	0.217	0.232	0.216	0.388	0.213	0.222	0.220	0.393	0.224	0.233
			(0.226)	(0.301)	(0.289)		(0.180)	(0.236)	(0.232)		(0.136)	(0.203)	(0.198)

[a] Mean value of statistic over 10,000 level-2 replicates.

[b] ( ) Standard deviation of statistic over 10,000 level-2 replicates.

Table 2
Preservation of average annual statistical attributes over 100 level-1 samples by 10,000 level-2 replicates from the AR(1)NT, MBB and PB models (L, block size used for resampling; AC, autocorrelation; AMF, above median and forward correlation; BMF, below median and forward correlation; AMB, above median and backward correlation; BMB, below median and backward correlation)

Statistic	$N = 40$				$N = 60$				$N = 80$			
	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$	Historical	PAR(1)NT	MBB $L = 5\omega$	PB $L = 5\omega$
Mean	0.264	0.257[a]	0.276	0.274	0.252	0.253	0.249	0.249	0.249	0.248	0.253	0.252
		(0.251)[b]	(0.267)	(0.271)		(0.198)	(0.212)	(0.215)		(0.172)	(0.182)	(0.185)
Std. dev.	0.720	0.680	0.698	0.685	0.714	0.663	0.695	0.683	0.714	0.665	0.704	0.691
		(0.148)	(0.164)	(0.158)		(0.120)	(0.131)	(0.127)		(0.096)	(0.114)	(0.109)
Skew	0.548	0.005	0.474	0.402	0.559	0.000	0.509	0.442	0.571	−0.005	0.533	0.459
		(0.378)	(0.470)	(0.449)		(0.308)	(0.388)	(0.377)		(0.261)	(0.364)	(0.347)
Lag-1 AC	0.509	0.313	0.372	0.435	0.514	0.332	0.383	0.445	0.527	0.332	0.402	0.468
		(0.159)	(0.167)	(0.165)		(0.128)	(0.140)	(0.141)		(0.107)	(0.123)	(0.121)
Lag-2 AC	0.163	−0.006	0.048	0.073	0.174	0.029	0.067	0.092	0.189	0.028	0.088	0.117
		(0.168)	(0.184)	(0.193)		(0.148)	(0.160)	(0.170)		(0.130)	(0.142)	(0.151)
AMF	0.400	0.197	0.267	0.312	0.413	0.218	0.294	0.336	0.411	0.221	0.302	0.348
		(0.226)	(0.264)	(0.243)		(0.175)	(0.215)	(0.201)		(0.150)	(0.197)	(0.183)
BMF	0.198	0.199	0.129	0.185	0.174	0.207	0.123	0.177	0.183	0.197	0.137	0.190
		(0.211)	(0.273)	(0.256)		(0.164)	(0.223)	(0.212)		(0.153)	(0.192)	(0.181)
AMB	0.433	0.203	0.298	0.335	0.442	0.230	0.318	0.353	0.459	0.230	0.337	0.375
		(0.210)	(0.261)	(0.245)		(0.172)	(0.210)	(0.201)		(0.147)	(0.179)	(0.168)
BMB	0.174	0.204	0.124	0.194	0.195	0.197	0.141	0.210	0.212	0.200	0.146	0.218
		(0.215)	(0.281)	(0.260)		(0.182)	(0.218)	(0.206)		(0.163)	(0.191)	(0.183)

[a] Mean value of statistic over 10,000 level-2 replicates.
[b] ( ) Standard deviation of statistic over 10,000 level-2 replicates.

Table 3
Salient features of the rivers selected for the study

Name of river	USGS station number	State code	County	Basin name	Drainage area (km^2)	Record duration	Mean annual discharge (m^3/s)
Beaver	10234500	49	Beaver	Beaver bottoms-upper Beaver	236	1914–1992 (79 years)	1.476
Weber	10128500	49	Summit	Upper Weber	420	1905–1988 (83 years)	6.316

1 and 2. It may be noted that the results of PB and MBB are presented for a typical block size of $L = 5\omega$ ($5 \times 4 = 20$ seasons), and the blocks start with the first season of a water year. The issue of block size selection is discussed in a later section of this article.

Tables 1 and 2 show that mean and standard deviation over 100 level-1 samples (referred to as population statistics) are well preserved by the 10,000 replicates generated from all the three models, at both seasonal and annual levels, for all three sample sizes considered.

The level-2 replicates generated from PAR(1)NT model are seen to exhibit near-zero skewness at both seasonal and annual levels. This is because no normalizing transformation is applied in model fitting. In contrast, the level-2 replicates from the MBB model are seen to be good at reproducing the skewness at both seasonal and annual levels because MBB is a data-driven model. Whereas, in case of PB model, the skewness exhibited by level-2 replicates is seen to be quite close to that of the MBB model. The reason for this behavior of PB model can be understood, if we look into the construction of the PB model. In the case of the PB model, level-1 samples are prewhitened using a simple PAR(1) model without any normalizing transformation. As a result, skewness of level-1

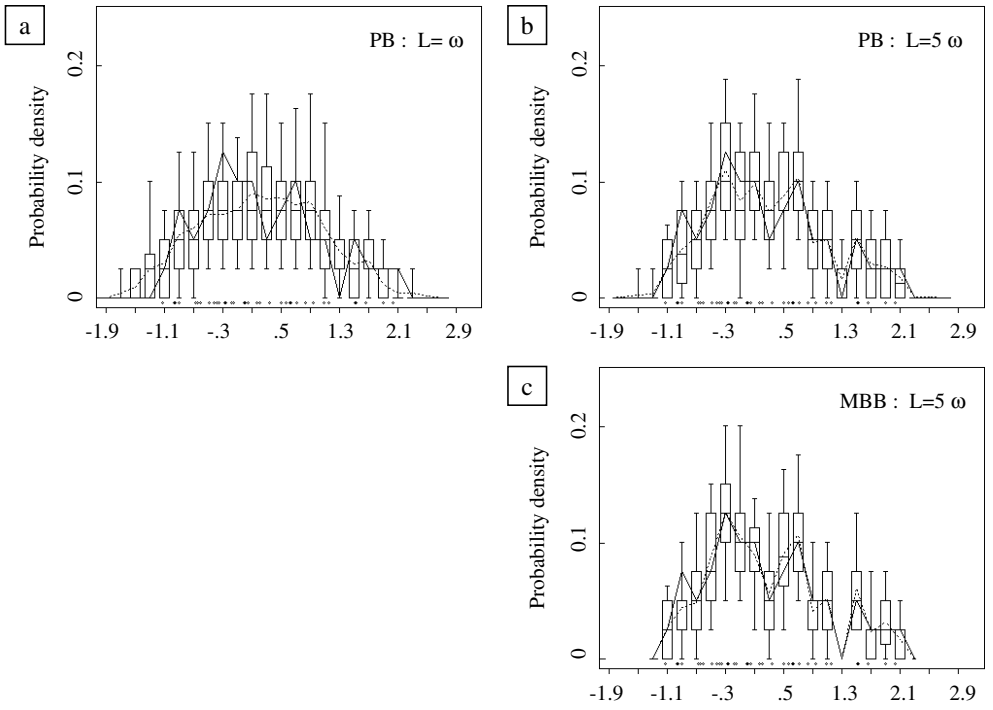

Fig. 1. Preservation of marginal distribution of the first season data points in a typical level-1 sample of size 40 years by 100 level-2 realizations. A comparison between the PB and the MBB models.

samples is apparently retained in the residuals extracted from the prewhitening stage. Bootstrapping these residuals using MBB, with a reasonable block size enables residual skewness to be preserved in the resampled innovations. Post blackening of these innovations, in turn, generates level-2 replicates that exhibit nearly the same behavior as MBB with regard to preservation of skewness. Furthermore, it may be noted that with further increase in block size of MBB used for bootstrapping the residuals, the skewness exhibited by level-2 replicates from PB model tends more towards the skewness exhibited by level-2 replicates from MBB model. Thus, we infer that PB model gains from its nonparametric constituent, MBB, in simulating skewness.

Figs. 1a–c and 2a–c show the respective preservation of marginal distribution corresponding to the first two seasons of a typical level-1 sample of size 40, by the level-2 simulations from: (i) PB model with $L = \omega$; (ii) PB model with $L = 5\omega$; and (iii) MBB model with $L = 5\omega$. Herein, it is to be mentioned that in the MBB model, level-1 sample record of length N is divided into $(N - L + 1)$ overlapping blocks each of equal length L. These overlapping blocks are then resampled with replacement to produce level-2 replicates. As a result, the level-2 replicates from MBB model are able to capture the salient features of the marginal distribution, namely, multimodality, asymmetry and peakedness (Figs. 1c and 2c), but they cannot either fill in the gaps within the data points of the level-1 sample or extrapolate the data beyond the extrema of level-1 sample. In contrast, it is well known that the parametric models offer appreciable amount of smoothing in modeling marginal distribution and can generate some values beyond extrema. The PB model is seen to gain this smoothing characteristic from its parametric constituent (PAR(1)NT) (see Figs. 1a and b and 2a and b). On the other hand, the nonparametric constituent of PB model (MBB) is responsible for the preservation of important features of the marginal distribution (multimodality, asymmetry and peakedness) like level-2 replicates from MBB (see Figs. 1c and 2c). A comparison between Fig. 1a and b (or Fig. 2a and b) reveal

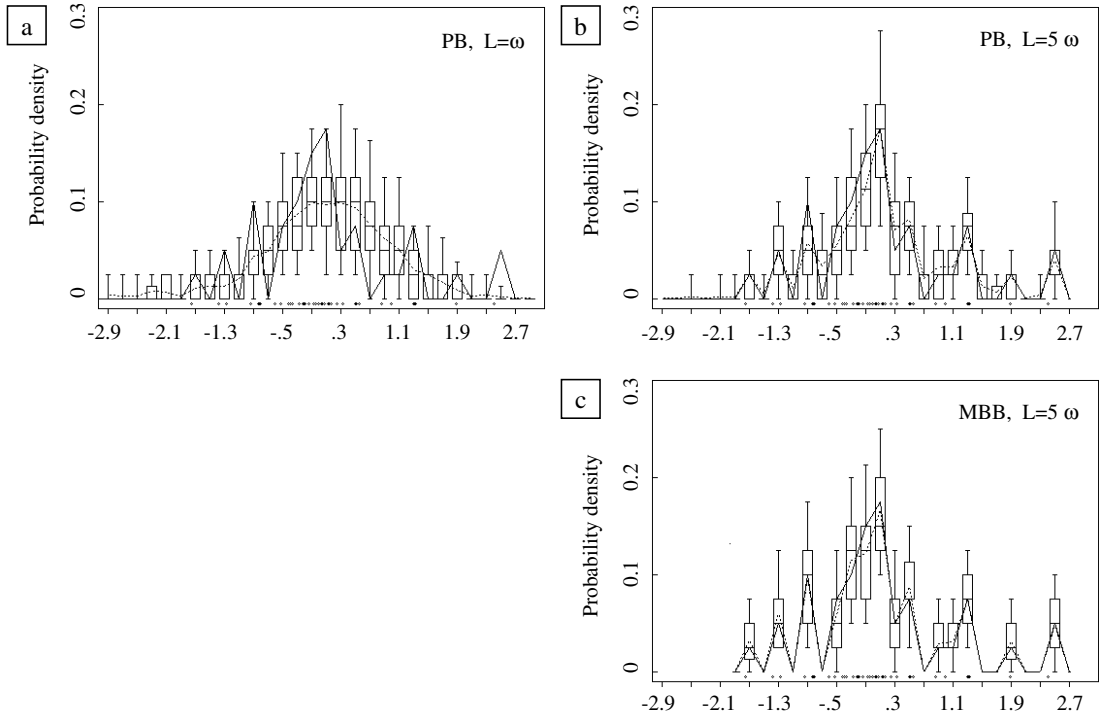

Fig. 2. Preservation of marginal distribution of the second season data points in a typical level-1 sample of size 40 years by 100 level-2 realizations. A comparison between the PB and the MBB models.

that when low block size ($L = \omega$) is adopted for resampling in the PB model, significant amount of smoothing and extrapolation beyond the extrema result (as in case of parametric models), and in the process, preservation of peakedness and multimodality of the distribution get affected. On the other hand, a longer block size ($L = 5\omega$, Figs. 1b and 2b) aids in preservation of these distributional features better, but smoothing and extrapolation beyond the extrema get reduced. Thus, the level-2 replicates from the PB model gain from both the constituent models (hybrid effect) and as a result they are able to not only reproduce the features of the marginal distribution of level-1 samples, adequately, but also provide some smoothing and extrapolation.

The theoretical structure of PAR(1)-NT model enables it to preserve the lag-1 serial correlations over all seasons (see Table 1), but not thereafter. In addition, being a periodic model, it is not able to capture the autocorrelations at the aggregated annual level. In the case of MBB model, the dependence of level-1 sample gets destroyed at the boundaries between adjoining blocks, though the dependence is preserved within the block. This enables all the within-year serial correlations to be well reproduced by the MBB model (see Table 1), since the block size is taken in multiples of years. However, the serial correlations between seasons of adjoining years (shown in bold font) may not be preserved by MBB owing to the discontinuities between the blocks. Due to the same reason, the autocorrelation structure at the annual level is also not preserved (see Table 2). In contrast, the PB model is able to preserve both within-year and cross-year serial correlations at seasonal level and autocorrelations at the aggregated annual level. This is because, considerable portion of dependence in level-1 sample is extracted by the parametric constituent (PAR(1)-NT) of the PB model during the prewhitening process and the uncaptured weak structure remaining in the residuals is modeled by the nonparametric constituent (MBB).

Now, we proceed to illustrate the hybrid effect on the preservation of state-dependent correlations. State-dependent correlation is a statistic that serves as a measure to quantify nonlinear dependence in flow data (Sharma et al., 1997). It quantifies the dependence of correlation on the magnitude of flow. These are computed as the correlations between flows above or below the median flow value in a month with the preceding or succeeding month flows. For a linear Gaussian process, the pair of above- and below-median correlations should be the same in either the forward or the backward direction. Differences between above-median and below-median correlations indicate nonlinearity or state dependence in the correlation structure. Sharma et al. (1997) defines the following four state-dependent correlations: (a) forward above-median correlation is defined as the correlation between above-median flows and flows in the subsequent time step; (b) forward below-median correlation is the correlation between all below-median flows and the flows in the subsequent time step; (c) backward above-median correlation is the correlation between above-median flows and the preceding time step's flows; and (d) backward below-median correlation is the correlation between below-median flows and the preceding time step's flows.

Table 1 shows that the pairs of above- and below-median state-dependent correlations for each season is nearly the same for PAR(1)-NT model simulations in both forward and backward directions, which is a typical characteristic of the linear Gaussian process. In other words, the level-2 simulations from PAR(1)-NT fail to capture the nonlinearity inherent in the level-1 sample. In contrast, MBB is good at reproducing the within-year state-dependent correlations because it resamples overlapping blocks of level-1 sample with block sizes in multiples of years (see Tables 1 and 2). However, MBB is not able to preserve the cross-year state-dependent correlations (i.e. forward and above-median, forward and below-median correlations for the last season of a year; backward and above-median, backward and below-median correlations for the first season of a year) due to the discontinuities between adjoining blocks. For the same reason, it is not able to reproduce any of the state-dependent correlations at the aggregated annual level. Note that the cross-year correlations are shown in bold italics font in Table 1. Some improvement can be seen in the preservation of the statistic by MBB with higher block size. But, this implies repeating large chunks of level-1 record as such in level-2 simulations and this is against the goal/spirit of stochastic simulation since this fails to offer any variety in simulations. On the other hand, the PB model is able to overcome the aforementioned

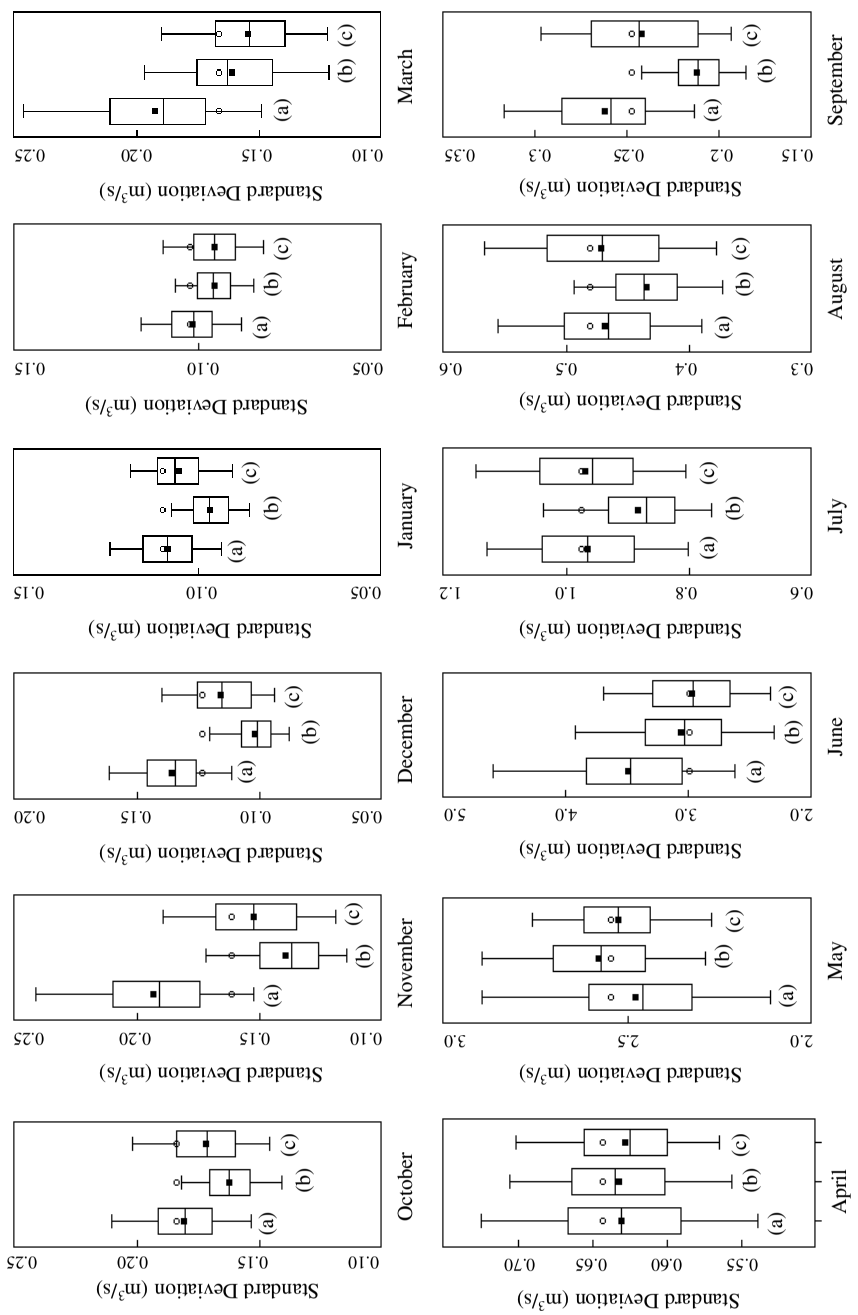

Fig. 3. Preservation of standard deviation (SD) of monthly streamflows. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River. The circle denotes the historical value and the darkened square denotes the average statistic from 500 synthetic replicates.

shortcoming of MBB owing to the hybrid effect (see Tables 1 and 2).

In summary,

1. The PB model is able to perform well even for small sample sizes and the performance improves with increase in sample size.
2. With regard to the preservation of skewness, within-year serial correlations and within-year state-dependent correlations, the PB model seems to gain from its nonparametric constituent (MBB) significantly.
3. The hybrid character of the PB model enables it to overcome the ill-effects resulting from block discontinuities found in the MBB model.
4. The PB model exhibits some smoothing in simulations owing to its parametric constituent and is able to gain in terms of modeling multimodality, peakedness and asymmetry, from its nonparametric constituent (MBB).

4. Results and discussion

The results from the application of the PB approach to modeling the periodic streamflows are compared with those resulting from the best low-order parametric model (out of the ones considered) and the *k*-NN bootstrap method. Five hundred replicates are generated from each one of the above three models, for both Beaver and Weber rivers. A detailed comparison of the preservation of a wide variety of statistics considered for the study is presented using Box-plots and tables. The box-plot referred to consists of a box that ranges from the upper to the lower quartile of the quantity being plotted and "whiskers" that extend from the box to 5 and 95% quantiles on the lower and the upper sides of the box, respectively. A line in the middle of the box represents the sample median. The span of the box represents the interquartile range of the statistic of concern. The historical statistic is represented by a circle and the mean of the generated statistic over 500 replicates is represented by a darkened square. If the historical statistic falls within the range of the box, then, differences between model and data can be ascribed to sampling variability. If the historical statistic falls outside the range of the box, then this indicates that the model does not reproduce the statistic.

4.1. Preservation of summary statistics

For the two rivers considered, the preservation of summary statistics of historical flows at both monthly and annuals levels, is discussed herein for parametric, *k*-NN and PB models. For the Beaver River, the summary statistics are presented in the form of box plots, while for the Weber river, most of the results are reported in tables.

For the Beaver River, the historical mean monthly flows are well reproduced by all the three models. However, the replicates from the PB model have been found to exhibit more variation of the mean monthly flows, compared to the other two models. This is not presented herein due to brevity. It may be noted from Fig. 3 that the PB model is able to model the standard deviation of monthly streamflows reasonably well, followed by the parametric model. In contrast, in the case of the *k*-NN model, considerable underestimation of the standard deviation is observed in eight out of 12 months (Fig. 3). The preservation of skewness of the monthly streamflows is shown in Fig. 4. It is seen that the parametric model (wherein a WHT is adopted) is able to reproduce the skewness of the monthly streamflows well, followed by the PB model. In contrast, the *k*-NN model considerably underestimates the statistic in five out of 12 months.

At the aggregated annual level, the historical mean annual flow is well reproduced by all the three models being studied. It may be noted from Table 4 that the parametric model overestimates the standard deviation, while the *k*-NN model slightly deflates the statistic. In contrast, the PB model is able to reproduce the standard deviation better. The parametric model highly overestimates the skewness at the aggregated annual level with a high standard deviation of the same. In contrast, the *k*-NN model is able to preserve the same better, though with some deflation, while the PB model is seen to exhibit a better performance (Table 4 and Fig. 4).

In case of the Weber River, it may be observed from Table 5 that the trend of preservation of the historical mean monthly flows is similar to that noted for the Beaver River flows. Furthermore, the parametric model is able to reproduce the standard

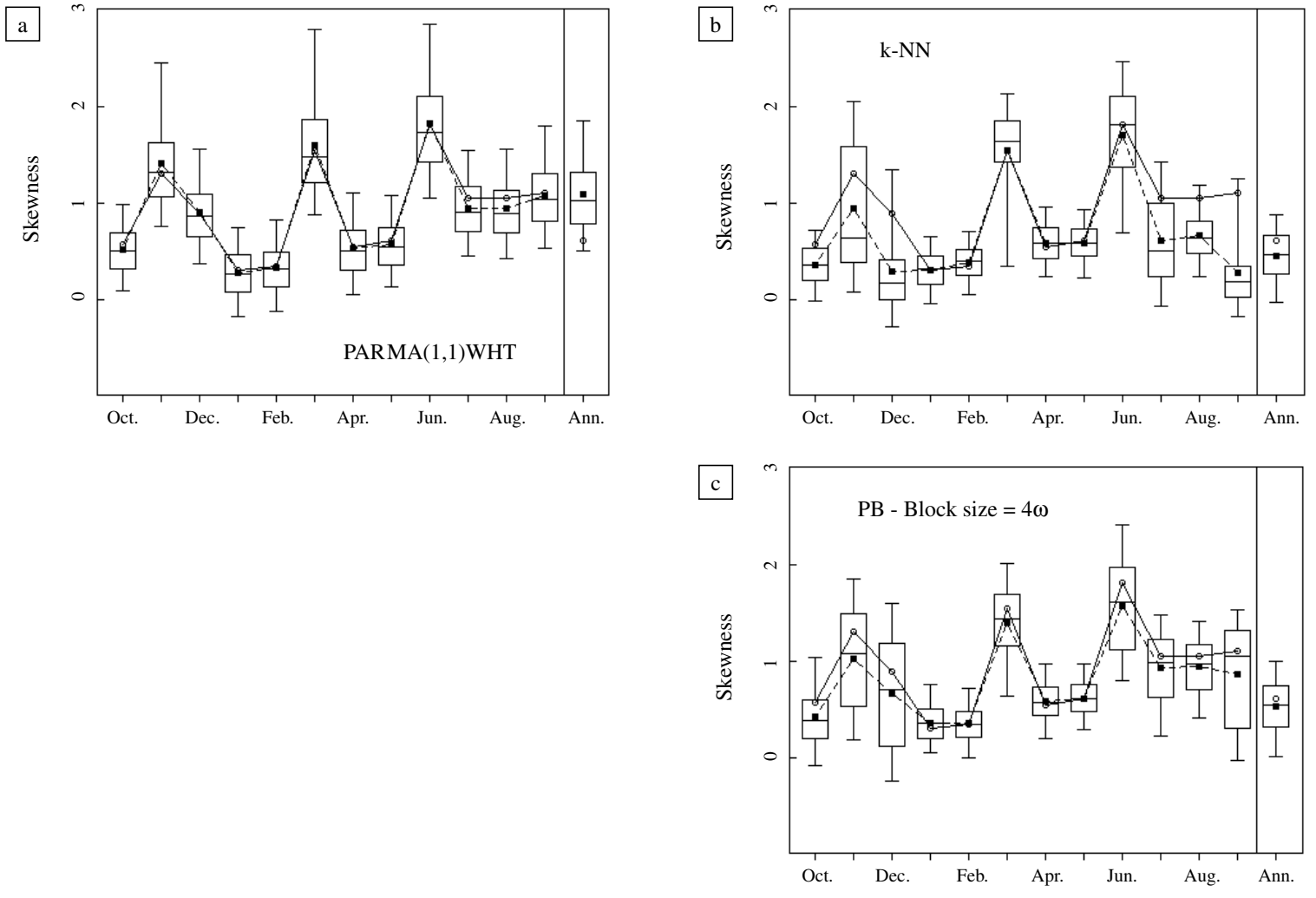

Fig. 4. Preservation of skewness of streamflows at monthly level and aggregated annual (Ann.) level. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models, for the Beaver River. The circle denotes skewness of historical flows and the darkened square denotes mean skewness computed from 500 synthetic replicates.

Table 4
Preservation of aggregated annual flow statistics: Beaver River (figures in parentheses denote the standard deviation over 500 synthetic replicates)

Model	Summary statistics			Lag-1
	Mean (m^3/s)	SD (m^3/s)	Skew	Autocorrelation
Hist.	1.476	0.564	0.59	0.362
PAR	1.483	0.604	1.07	0.154
	(0.085)	(0.076)	(0.45)	(0.109)
k-NN	1.486	0.542	0.44	0.143
	(0.067)	(0.044)	(0.27)	(0.105)
PB	1.471	0.555	0.51	0.275
$L = 4\omega$	(0.087)	(0.054)	(0.30)	(0.110)
PB	1.477	0.558	0.51	0.293
$L = 5\omega$	(0.087)	(0.061)	(0.29)	(0.102)

deviation of monthly streamflows reasonably well in most of the months. However, considerable overestimation of the statistic is observed in the months of October and September. On the other hand, the *k*-NN model consistently underestimates the standard deviation of monthly streamflows. In contrast, the PB model is able to preserve the same reasonably well, even though a slight deflation is observed. The parametric model (PAR(2) with WHT) is able to reproduce the skewness of monthly flows reasonably well, except in the months of March and July wherein skewness of the flows is high. The performance of the *k*-NN and the PB models in terms of skewness preservation, is quite comparable to the parametric model, even though no normalizing transformation is applied.

At the aggregated annual level, the mean flow is well reproduced by all the three models (Table 6). Furthermore, the replicates from the PB model exhibit more variation of the statistic compared to the other two models. It may be noted from Table 6 that the parametric and the PB models reproduce the standard deviation of streamflows at the aggregated annual level reasonably well, though with a small deflation. On the other hand, the *k*-NN model underestimates the standard deviation considerably, with low variation of the same compared to the other two models. Moreover, it can be seen that the parametric model (PAR(2) WHT) highly overestimates the skewness with a high standard deviation of the same. In contrast, both the *k*-NN and the PB models are better in reproducing the skewness.

4.2. Preservation of the empirical marginal distribution

The empirical marginal distributions for a few selected months obtained from the historical as well as the 500 synthetic streamflow replicates generated from parametric, *k*-NN and PB models are presented in Figs. 5–9. The observed historical flow values are marked as individual dots below the box plots. The box plot drawn for each class interval represents the spread of the synthetic flows (over 500 replicates) falling in that class interval. The dashed line represents the distribution of synthetic flows averaged over 500 replicates.

For the Beaver River, the preservation of the marginal distribution is presented typically for the months of June and July (Figs. 5 and 6). Both *k*-NN and PB models are seen to capture the complex pattern of the marginal distribution of the historical flows of both months. While, the parametric model (PARMA(1,1) with WHT) fails to capture the marginal distributional patterns, even though the skewness of the monthly streamflows is well preserved.

For the Weber River, the marginal distributions plotted for the months of October, June and July are shown in Figs. 7–9. It can be seen from these figures that the parametric model (PAR(2) with WHT) is poor in reproducing marginal distribution of the historical flows. In contrast, the PB model is able to capture the asymmetry, peakedness, multimodality and the tail behavior, closely followed by the *k*-NN model. Herein, it is to be mentioned that the *k*-NN model cannot generate flow values other than the ones found in the historical record. On the other hand, the parametric model offers considerable smoothing, while the PB model being a hybrid one, is able to offer some smoothing (Figs. 5–9).

4.3. Preservation of dependence structure

A comparison of the results from the investigation regarding the preservation of the serial correlations, cross-correlations and state-dependent correlations of the historical monthly streamflows, and the autocorrelations of the aggregated annual streamflows, between the three models considered, is presented in this section.

Table 5
Preservation of summary statistics of monthly flows. A comparison between: (a) parametric (PAR); (b) *k*-nearest-neighbor (*k*-NN); and (c) PB models. River: Weber (figures in parentheses denote the standard deviation over 500 synthetic replicates)

Month	Model	Summary statistics			Serial correlations						
		Mean (m^3/s)	SD (m^3/s)	Skew	Cor1	Cor2	Cor3	Cor4	Cor5	Cor6	Cor7
Oct.	Hist.	2.305	0.777	1.38	0.741	0.573	0.375	0.428	0.248	0.149	0.324
	PAR	2.324	0.885	1.42	0.826	0.636	0.481	0.477	0.124	0.016	0.035
		(0.096)	(0.118)	(0.48)	(0.045)	(0.072)	(0.107)	(0.087)	(0.111)	(0.110)	(0.109)
	k-NN	2.287	0.695	1.22	0.682	0.541	0.420	0.394	0.083	0.011	0.017
		(0.080)	(0.091)	(0.56)	(0.060)	(0.083)	(0.099)	(0.083)	(0.107)	(0.104)	(0.109)
	PB	2.290	0.758	1.06	0.731	0.590	0.408	0.432	0.211	0.083	0.234
		(0.114)	(0.112)	(0.48)	(0.051)	(0.061)	(0.091)	(0.061)	(0.104)	(0.095)	(0.102)
Nov.	Hist.	2.022	0.514	0.65	0.843	0.751	0.615	0.395	0.517	0.289	0.110
	PAR	2.019	0.506	0.61	0.820	0.726	0.585	0.426	0.449	0.117	0.014
		(0.055)	(0.047)	(0.32)	(0.040)	(0.058)	(0.075)	(0.097)	(0.089)	(0.113)	(0.111)
	k-NN	2.005	0.472	0.66	0.822	0.606	0.496	0.384	0.369	0.081	0.014
		(0.054)	(0.039)	(0.21)	(0.039)	(0.077)	(0.091)	(0.103)	(0.093)	(0.105)	(0.105)
	PB	2.011	0.492	0.68	0.835	0.720	0.598	0.410	0.476	0.232	0.047
		(0.076)	(0.050)	(0.23)	(0.030)	(0.056)	(0.078)	(0.104)	(0.076)	(0.100)	(0.103)
Dec.	Hist.	1.756	0.406	0.60	0.866	0.709	0.693	0.616	0.456	0.541	0.242
	PAR	1.753	0.407	0.57	0.877	0.666	0.598	0.488	0.354	0.376	0.097
		(0.046)	(0.040)	(0.33)	(0.027)	(0.065)	(0.073)	(0.086)	(0.105)	(0.099)	(0.113)
	k-NN	1.750	0.369	0.62	0.847	0.693	0.520	0.429	0.332	0.322	0.074
		(0.041)	(0.032)	(0.28)	(0.033)	(0.059)	(0.087)	(0.095)	(0.104)	(0.100)	(0.109)
	PB	1.748	0.384	0.57	0.853	0.687	0.655	0.578	0.438	0.482	0.200
		(0.059)	(0.039)	(0.29)	(0.029)	(0.047)	(0.082)	(0.091)	(0.091)	(0.087)	(0.102)
Jan.	Hist.	1.623	0.331	0.66	0.831	0.782	0.637	0.681	0.601	0.403	0.503
	PAR	1.621	0.326	0.60	0.848	0.784	0.605	0.540	0.440	0.318	0.339
		(0.036)	(0.031)	(0.33)	(0.034)	(0.046)	(0.075)	(0.082)	(0.091)	(0.107)	(0.101)
	k-NN	1.619	0.298	0.59	0.819	0.714	0.581	0.440	0.364	0.283	0.275
		(0.032)	(0.024)	(0.24)	(0.071)	(0.069)	(0.081)	(0.095)	(0.103)	(0.107)	(0.107)
	PB	1.617	0.312	0.62	0.818	0.751	0.591	0.616	0.540	0.382	0.433
		(0.049)	(0.028)	(0.24)	(0.073)	(0.056)	(0.085)	(0.082)	(0.087)	(0.096)	(0.085)
Feb.	Hist.	1.621	0.318	0.56	0.889	0.768	0.726	0.590	0.561	0.450	0.270
	PAR	1.618	0.315	0.49	0.886	0.788	0.725	0.557	0.497	0.404	0.292
		(0.035)	(0.028)	(0.31)	(0.027)	(0.045)	(0.059)	(0.083)	(0.088)	(0.093)	(0.109)
	k-NN	1.625	0.288	0.53	0.858	0.717	0.627	0.516	0.390	0.320	0.249
		(0.031)	(0.026)	(0.24)	(0.031)	(0.073)	(0.080)	(0.090)	(0.100)	(0.106)	(0.106)
	PB	1.617	0.303	0.58	0.880	0.750	0.700	0.559	0.517	0.422	0.279
		(0.049)	(0.029)	(0.24)	(0.028)	(0.057)	(0.063)	(0.072)	(0.084)	(0.098)	(0.103)
Mar.	Hist.	1.904	0.656	2.97	0.671	0.671	0.674	0.658	0.438	0.485	0.379
	PAR	1.939	0.640	2.39	0.661	0.660	0.570	0.528	0.423	0.372	0.290
		(0.074)	(0.135)	(0.81)	(0.064)	(0.068)	(0.083)	(0.090)	(0.116)	(0.120)	(0.114)
	k-NN	1.925	0.630	2.64	0.660	0.570	0.486	0.427	0.355	0.267	0.220
		(0.073)	(0.141)	(0.88)	(0.052)	(0.068)	(0.083)	(0.097)	(0.108)	(0.119)	(0.119)
	PB	1.889	0.593	2.43	0.664	0.669	0.658	0.650	0.424	0.456	0.350
		(0.075)	(0.127)	(0.93)	(0.072)	(0.050)	(0.058)	(0.059)	(0.086)	(0.100)	(0.113)
Apr.	Hist.	5.124	2.457	1.15	0.637	0.393	0.426	0.411	0.399	0.213	0.392
	PAR	5.113	2.371	1.03	0.521	0.386	0.391	0.339	0.313	0.237	0.215
		(0.271)	(0.253)	(0.37)	(0.095)	(0.097)	(0.095)	(0.101)	(0.104)	(0.110)	(0.109)
	k-NN	5.077	2.291	0.99	0.505	0.356	0.316	0.280	0.247	0.205	0.157

Table 5 (*continued*)

Month	Model	Summary statistics			Serial correlations						
		Mean (m^3/s)	SD (m^3/s)	Skew	Cor1	Cor2	Cor3	Cor4	Cor5	Cor6	Cor7
		(0.252)	(0.226)	(0.34)	(0.116)	(0.097)	(0.100)	(0.102)	(0.107)	(0.113)	(0.112)
	PB	5.089	2.380	1.01	0.597	0.348	0.382	0.360	0.341	0.157	0.320
		(0.237)	(0.240)	(0.36)	(0.107)	(0.091)	(0.094)	(0.099)	(0.104)	(0.100)	(0.114)
May	Hist.	19.752	6.440	0.35	0.410	0.419	0.404	0.462	0.419	0.443	0.296
	PAR	19.716	6.460	0.36	0.418	0.405	0.333	0.334	0.291	0.270	0.202
		(0.723)	(0.511)	(0.28)	(0.095)	(0.096)	(0.103)	(0.103)	(0.105)	(0.108)	(0.111)
	k-NN	19.858	6.356	0.37	0.367	0.214	0.169	0.145	0.129	0.114	0.096
		(0.725)	(0.522)	(0.21)	(0.083)	(0.103)	(0.108)	(0.110)	(0.108)	(0.106)	(0.106)
	PB	19.776	6.319	0.33	0.388	0.401	0.366	0.432	0.373	0.393	0.229
		(0.726)	(0.509)	(0.21)	(0.075)	(0.071)	(0.100)	(0.091)	(0.099)	(0.087)	(0.094)
June	Hist.	26.512	12.477	0.45	0.165	−0.007	0.225	0.342	0.304	0.309	0.271
	PAR	26.483	12.373	0.43	0.185	0.006	0.043	0.033	0.034	0.030	0.021
		(1.368)	1.072	(0.30)	(0.110)	(0.111)	(0.110)	(0.112)	(0.109)	(0.109)	(0.112)
	k-NN	26.477	12.006	0.41	0.131	0.007	0.046	0.045	0.039	0.034	0.029
		(1.338)	0.995	(0.25)	(0.104)	(0.109)	(0.114)	(0.115)	(0.113)	(0.115)	(116)
	PB	26.516	12.273	0.40	0.138	−0.023	0.202	0.331	0.303	0.294	0.254
		(1.442)	(1.028)	(0.24)	(0.105)	(0.097)	(0.140)	(0.089)	(0.080)	(0.093)	(0.107)
July	Hist.	7.629	5.821	3.17	0.672	0.100	0.044	0.117	0.219	0.189	0.243
	PAR	7.825	5.655	2.57	0.740	0.191	0.034	0.055	0.043	0.042	0.037
		0.624	(1.216)	(0.81)	(0.048)	(0.111)	(0.114)	(0.117)	(0.115)	(0.114)	(0.111)
	k-NN	7.603	5.411	2.58	0.696	0.097	0.006	0.029	0.030	0.029	0.025
		(0.596)	(1.177)	(0.92)	(0.061)	(0.097)	(0.110)	(0.109)	(0.107)	(0.109)	(0.108)
	PB	7.676	5.747	2.72	0.692	0.089	0.021	0.113	0.219	0.188	0.231
		(0.703)	(1.260)	(0.84)	(0.063)	(0.093)	(0.147)	(0.110)	(0.082)	(0.087)	(0.105)
Aug.	Hist.	3.205	1.211	1.01	0.777	0.754	0.288	0.034	0.130	0.310	0.355
	PAR	3.204	1.241	0.97	0.795	0.775	0.194	0.027	0.054	0.046	0.044
		(0.138)	(0.125)	(0.38)	(0.047)	(0.047)	(0.110)	(0.109)	(0.105)	(0.110)	(0.109)
	k-NN	3.208	1.130	1.00	0.766	0.709	0.153	0.027	0.040	0.040	0.036
		(0.123)	(0.117)	(0.31)	(0.078)	(0.053)	(0.109)	(0.107)	(0.107)	(0.108)	(0.109)
	PB	3.212	1.194	0.92	0.791	0.759	0.262	0.012	0.104	0.274	0.315
		(0.155)	(0.130)	(0.26)	(0.037)	(0.040)	(0.100)	(0.112)	(0.115)	(0.099)	(0.100)
Sep.	Hist.	2.341	0.834	1.33	0.881	0.593	0.624	0.284	0.031	0.177	0.357
	PAR	2.364	0.951	1.38	0.853	0.664	0.640	0.161	0.023	0.045	0.040
		(0.104)	(0.118)	(0.46)	(0.035)	(0.083)	(0.070)	(0.110)	(0.113)	(0.107)	(0.111)
	k-NN	2.326	0.753	1.17	0.845	0.661	0.602	0.122	0.016	0.029	0.026
		(0.084)	(0.086)	(0.36)	(0.042)	(0.087)	(0.067)	(0.109)	(0.108)	(0.110)	(0.108)
	PB	2.341	0.816	1.19	0.873	0.612	0.622	0.259	0.006	0.145	0.321
		(0.109)	(0.111)	(0.33)	(0.034)	(0.081)	(0.058)	(0.096)	(0.089)	(0.107)	(0.087)

4.3.1. Preservation of monthly serial correlations

In the case of the Beaver River, it may be noted from Fig. 10 that lag-1 serial correlations are reasonably preserved by the parametric model, while a consistent underestimation is observed in seven out of the 12 months in case of the *k*-NN model. In contrast, the PB model is able to preserve the statistic better. The lag-2 serial correlations are reasonably preserved by the parametric model (Fig. 11). However, there is a considerable inflation of the statistic in the months of June–August. On the other hand, the *k*-NN model does not preserve the lag-2 serial correlations, since its structure is designed to preserve only lag-1 serial correlation ($d = 1$). Furthermore, it is

Table 6
Preservation of aggregated annual flow statistics: Weber River (figures in parentheses denote the standard deviation over 500 synthetic replicates)

Model	Summary statistics			Autocorrelation	
	Mean (m^3/s)	SD (m^3/s)	Skew	Lag-1	Lag-2
Hist.	6.316	1.874	0.44	0.253	0.180
PAR	6.331	1.844	0.77	0.098	−0.015
	(0.225)	(0.192)	(0.41)	(0.104)	(0.108)
k-NN	6.313	1.701	0.39	0.063	−0.021
	(0.199)	(0.153)	(0.28)	(0.105)	(0.108)
PB	6.315	1.820	0.38	0.190	0.038
$L = 3\omega$	(0.251)	(0.162)	(0.26)	(0.089)	(0.114)
PB	6.296	1.779	0.30	0.231	0.104
$L = 5\omega$	(0.267)	(0.163)	(0.27)	(0.085)	(0.116)

to be mentioned that adopting the higher-order *k*-NN models distorts the preservation of lag-1 serial correlations. In contrast, the PB model is able to capture lag-2 serial correlations (Fig. 11). It is to be mentioned that neither the parametric nor the *k*-NN model captures serial correlations for lag-3 (Fig. 12) and higher lags. In contrast, the PB model exhibits a much better performance in terms of preservation of monthly serial correlations upto several lags (Figs. 12 and 13).

4.3.2. Preservation of annual autocorrelations

The preservation of autocorrelation of flows at the aggregated annual level is examined for both the rivers for all the three models considered. Both the parametric and the *k*-NN models are not able to reproduce the autocorrelations of the Beaver and Weber rivers (Tables 4 and 6, respectively). In contrast, the PB model ($L = 4\omega$ for Beaver and $L = 3\omega$ for Weber) is able to preserve the same reasonably. For the PB model, the results are presented for an alternative block size ($L = 5\omega$) that enables the appreciation of the flexibility associated with the method. In the case of the Beaver River, there is a significant improvement in the preservation of lag-1 autocorrelation of flows (Table 4). A similar trend may be noted in the preservation of lag-1 and lag-2 autocorrelations for the Weber River (Table 6).

4.3.3. Preservation of monthly state-dependent correlations

For the Beaver River, the preservation of "above and forward", "above and backward", "below and forward" and "below and backward" state-dependent correlations are presented in Figs. 14–17, respectively, for the three models being compared. It may be noted from the figures that the parametric model (PARMA(1,1) WHT) overestimates "above and forward" and "above and backward" correlations considerably, while it underestimates "below and forward" and "below and backward" state-dependent correlations. On the other hand, the *k*-NN model deflates "above and forward" and "above and backward" correlations significantly for most of the months. In contrast, the PB model exhibits better performance in modeling the same. It is to be noted that the "below and forward" and the "below and backward" correlations are reasonably well preserved by both *k*-NN and PB models.

In the case of the Weber River, Figs. 18–21 show the preservation of "above and forward", "above and backward", "below and forward" and "below and backward" state-dependent correlations, respectively, by the three models considered. It may be noted that the parametric model shows high amount of bias for all the four state-dependent correlations, whereas, the *k*-NN model is able to perform reasonably well in most months. However, considerable underestimation is noted in a few of the months. In contrast, the PB model is able to capture the typical historical trend of all the four state-dependent correlations.

4.3.4. Preservation of cross-correlations

An additional attraction of the PB model is that the within-year cross-correlations between different month pairs and the month-to-annual cross-correlations are reasonably preserved. This may be appreciated from the results presented in Tables 7 and 8 for month-to-month cross-correlations between the various month pairs and Figs. 22 and 23 for the month-to-annual cross-correlations for the Beaver and the Weber rivers. On the other hand, the other two models are not able to preserve the same, since they are not designed for preserving these cross-correlations (not shown for brevity).

4.4. Preservation of storage capacity statistics

Synthetic streamflow sequences are often used for estimation of reservoir storage capacity for a

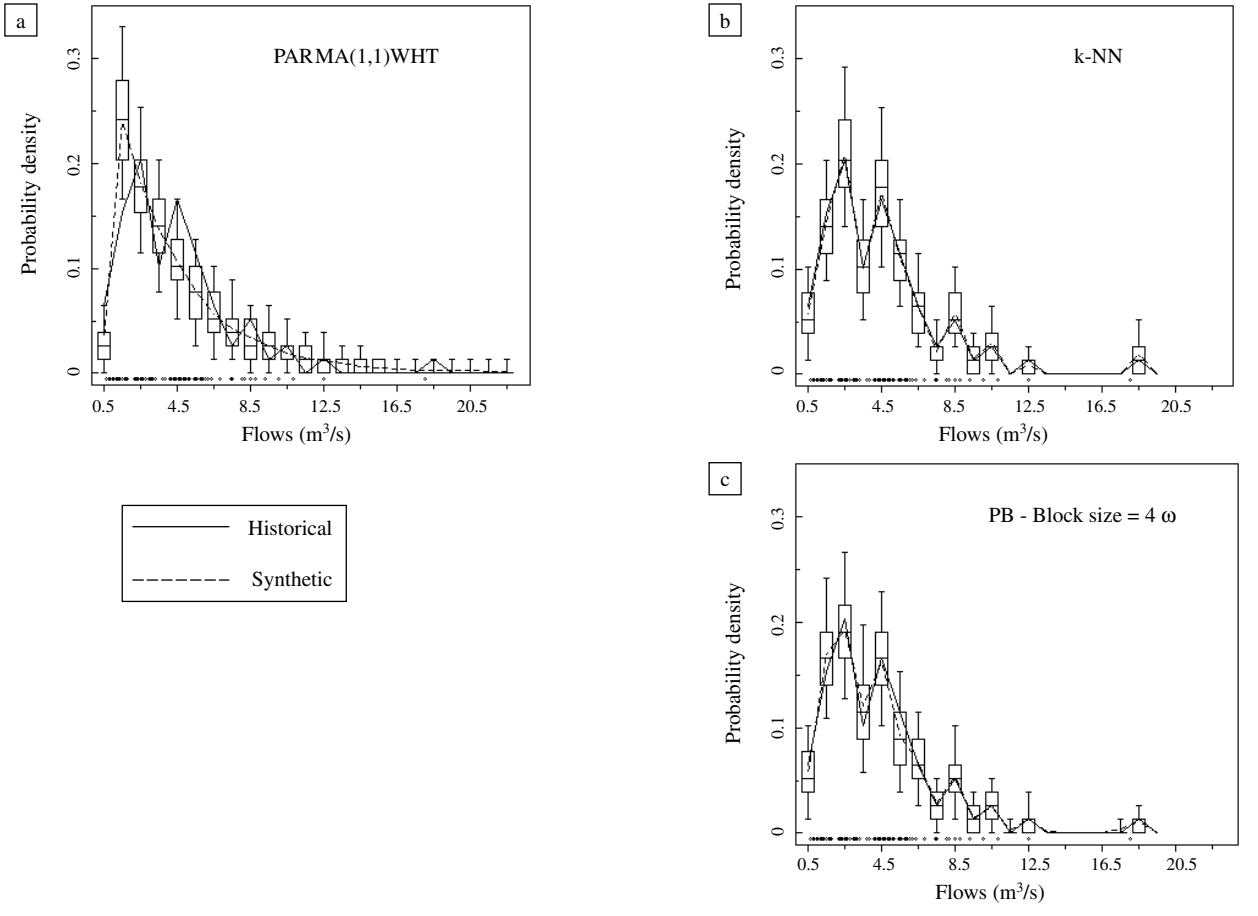

Fig. 5. Preservation of marginal distribution for June month flows. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models, for the Beaver River. The dots below the box plots denote the observed flow values.

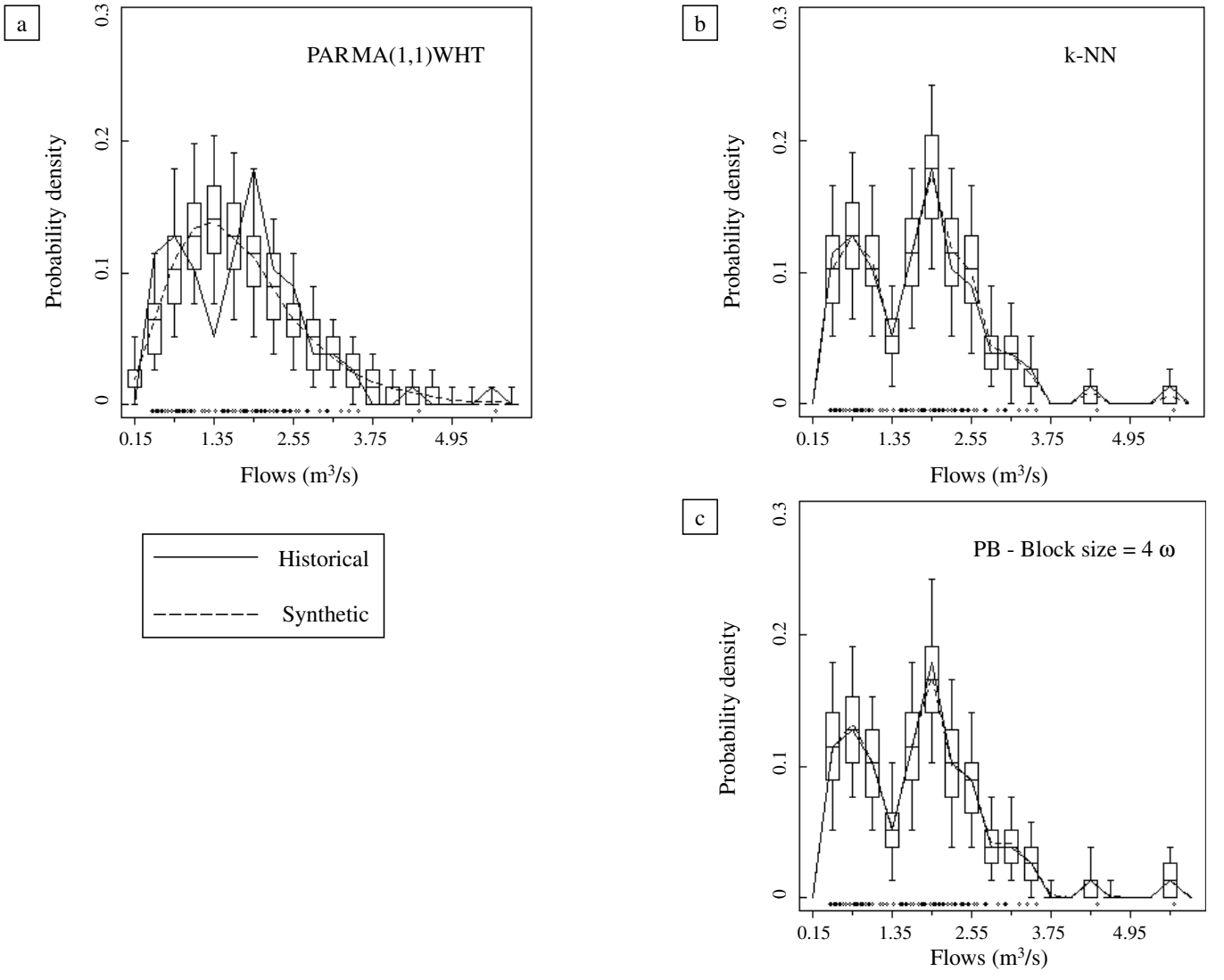

Fig. 6. Preservation of marginal distribution for July month flows. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River. The dots below the box plots denote the observed flow values.

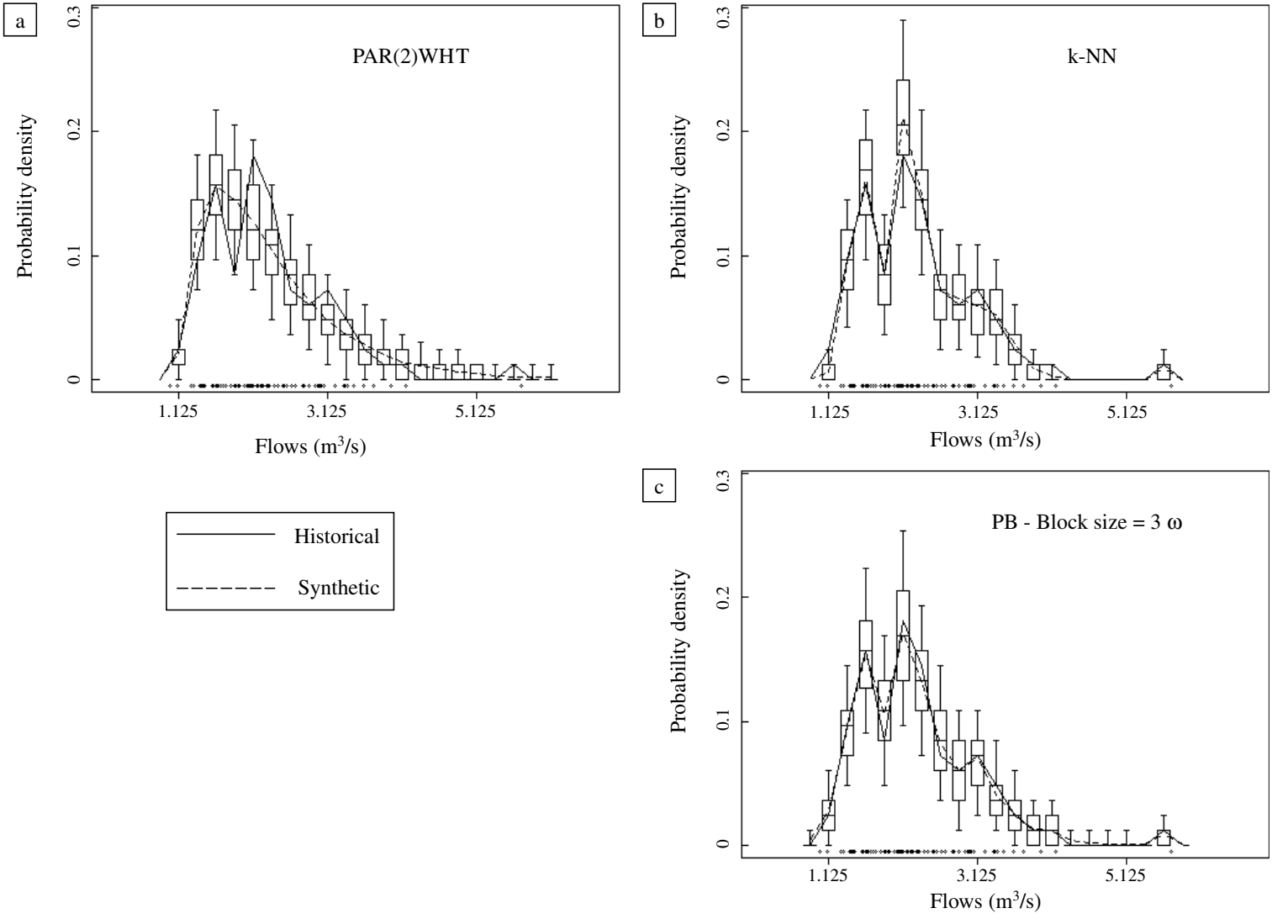

Fig. 7. Preservation of marginal distribution for October month flows. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Weber River. The dots below the box plots denote the observed flow values.

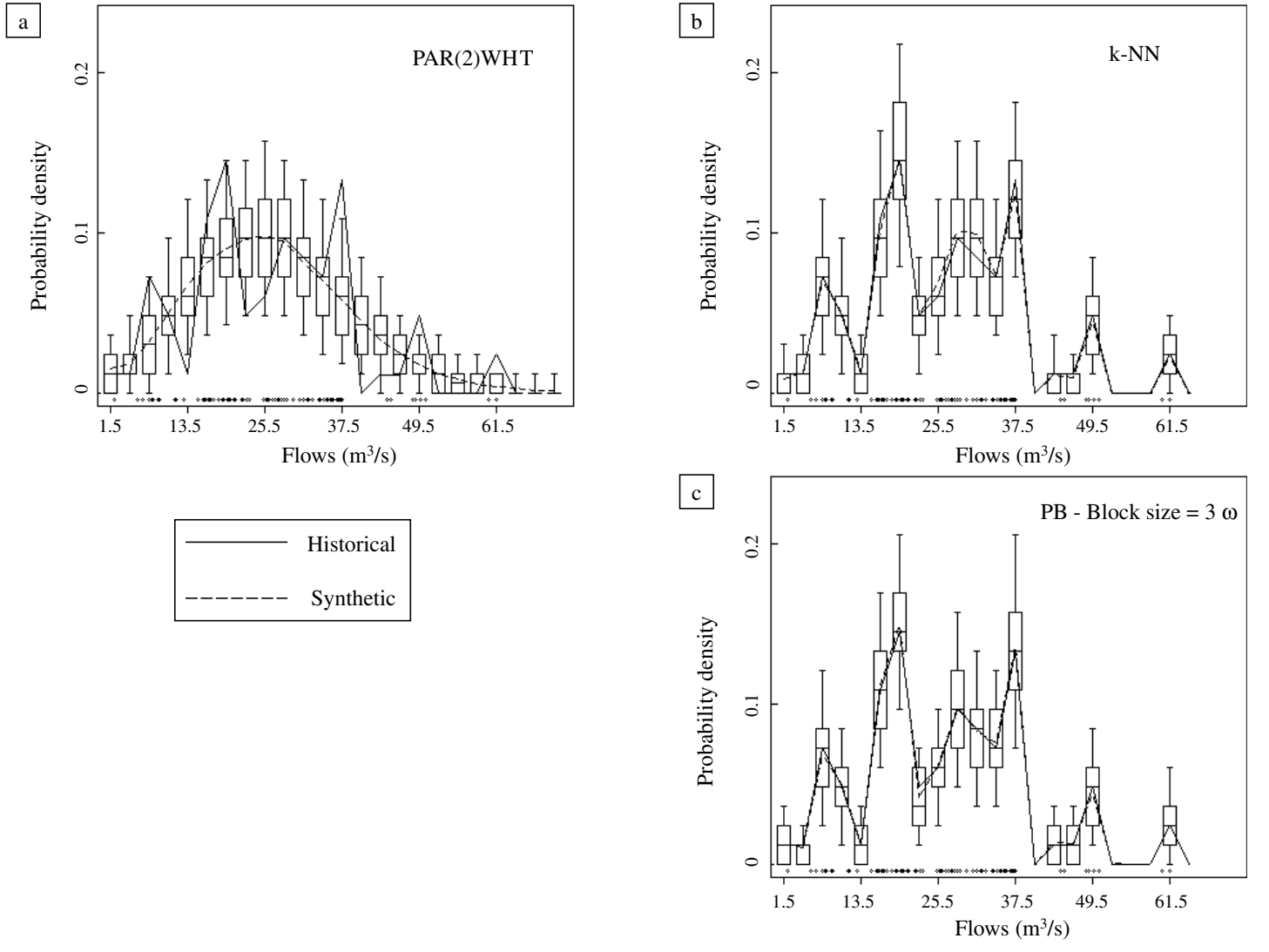

Fig. 8. Preservation of marginal distribution for June month flows. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Weber River. The dots below the box plots denote the observed flow values.

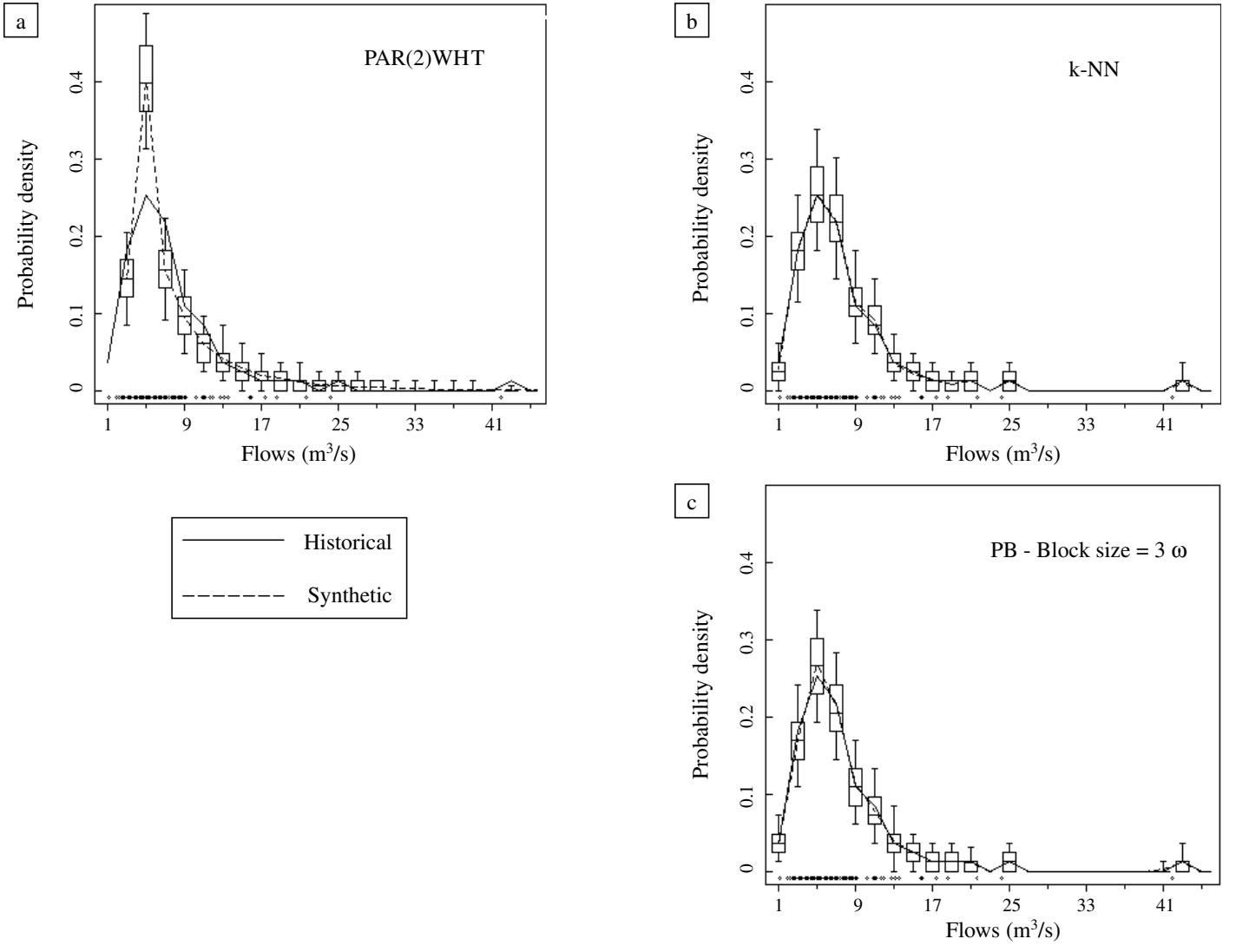

Fig. 9. Preservation of marginal distribution for July month flows. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Weber River. The dots below the box plots denote the observed flow values.

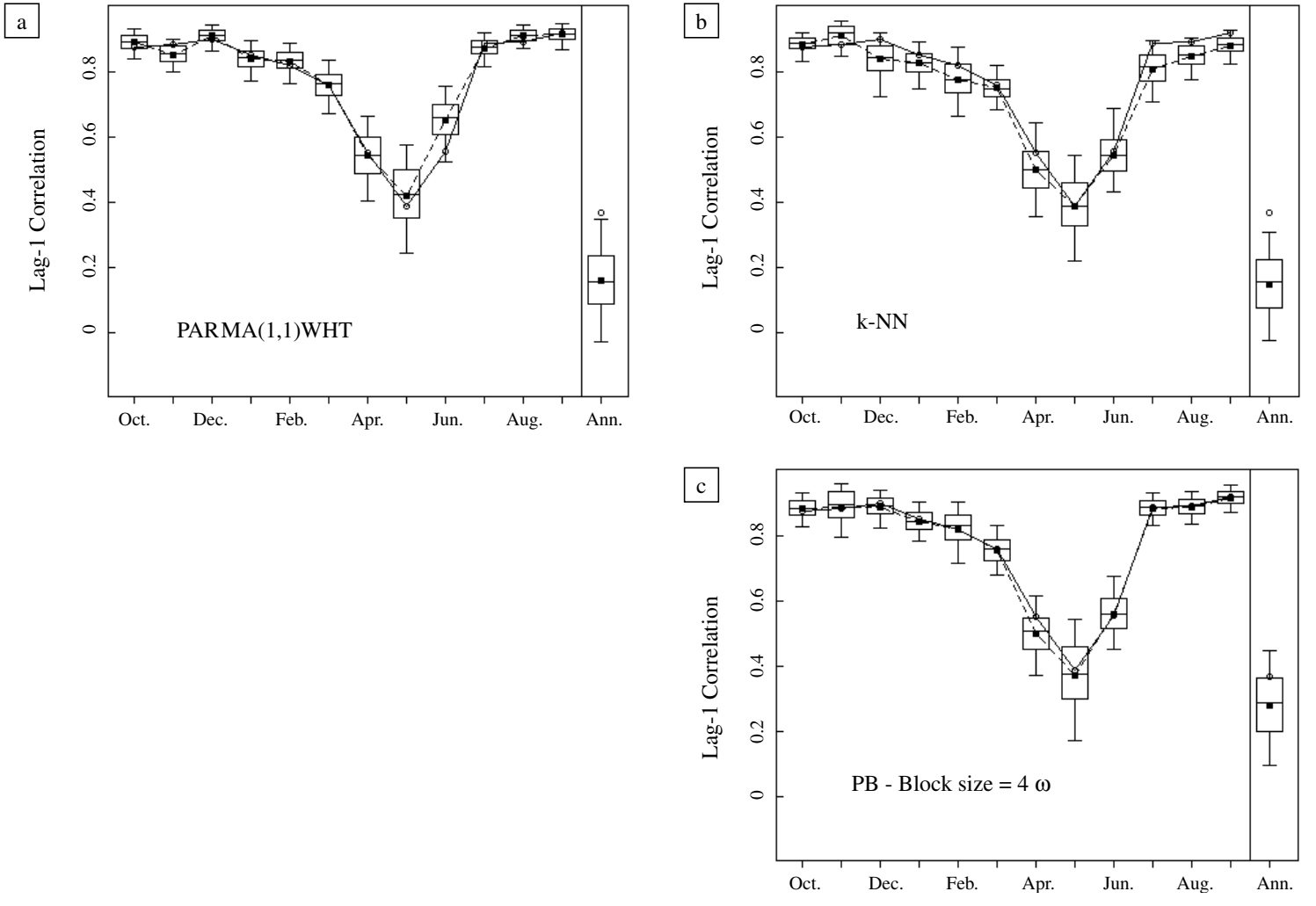

Fig. 10. Preservation of lag-1 serial correlation of streamflows at monthly level and lag-1 autocorrelation at the aggregated annual (Ann.) level-A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River. The circle denotes historical value of correlation and the darkened square denotes mean correlation computed from the 500 synthetic replicates.

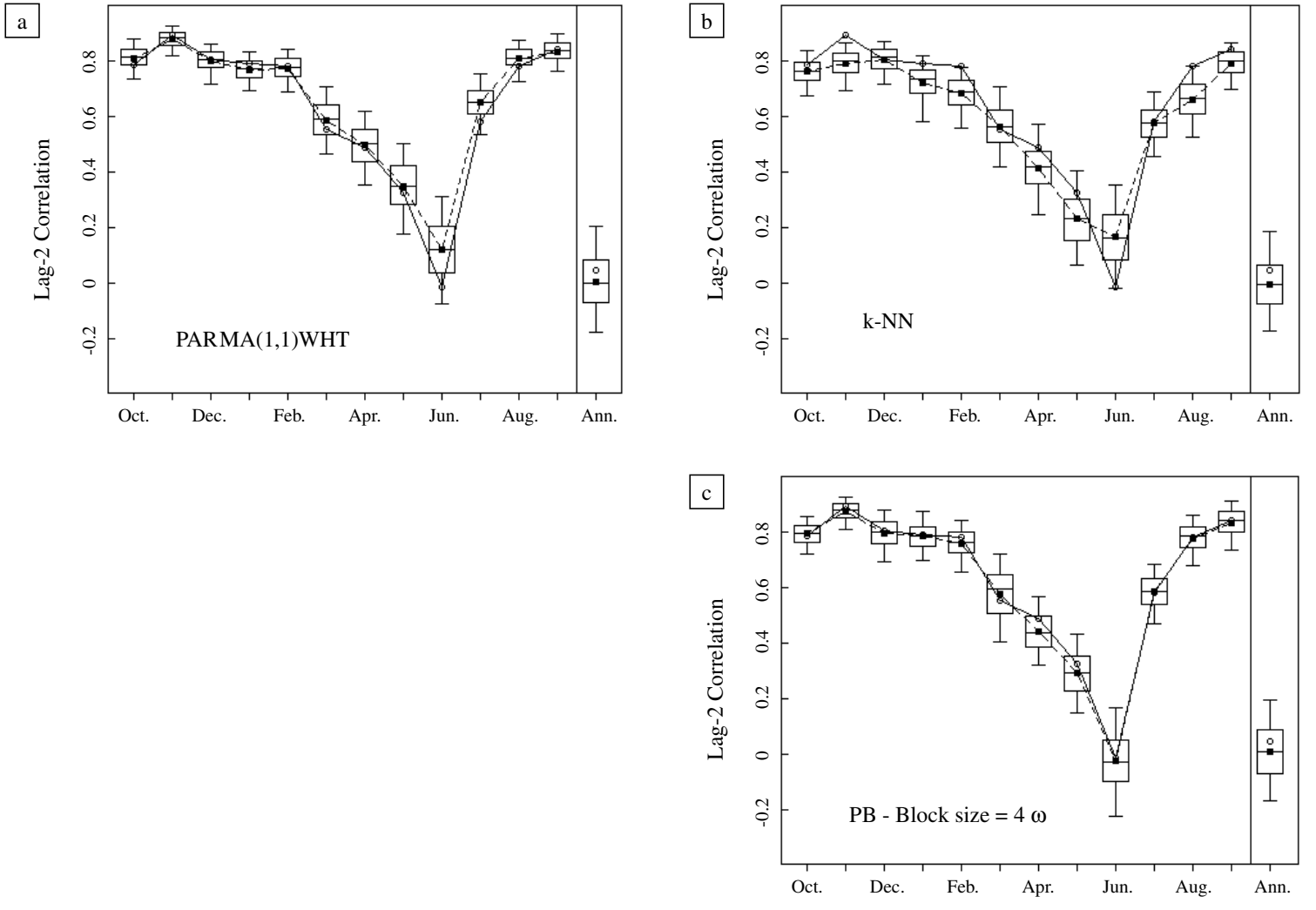

Fig. 11. Preservation of lag-2 serial correlation of streamflows at monthly level and lag-2 autocorrelation at the aggregated annual (Ann.) level-A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River. The circle denotes historical value of correlation and the darkened square denotes mean correlation computed from the 500 synthetic replicates.

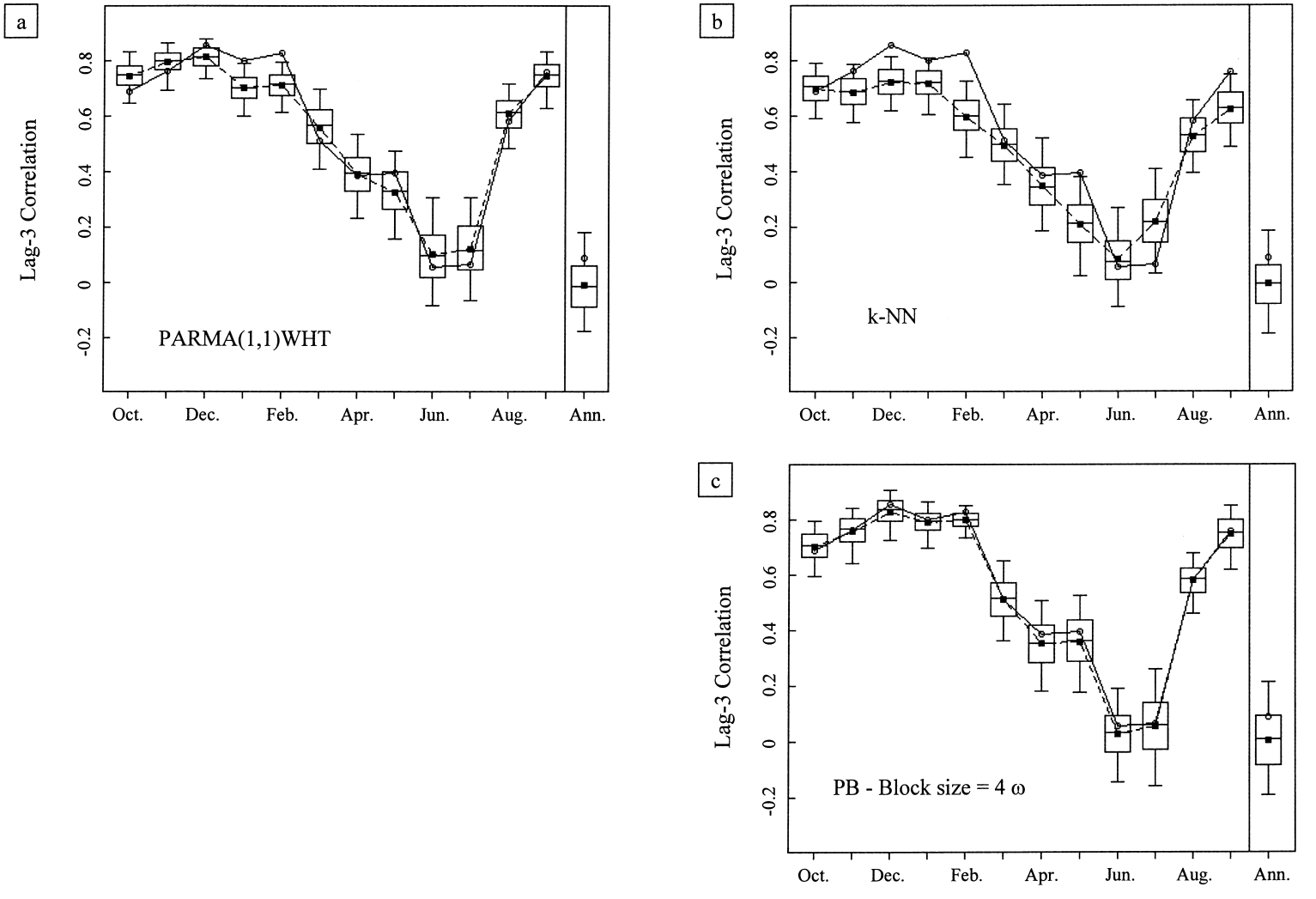

Fig. 12. Preservation of lag-3 serial correlation of streamflows at monthly level and lag-3 autocorrelation at the aggregated annual (Ann.) level-A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River. The circle denotes historical value of correlation and the darkened square denotes mean correlation computed from the 500 synthetic replicates.

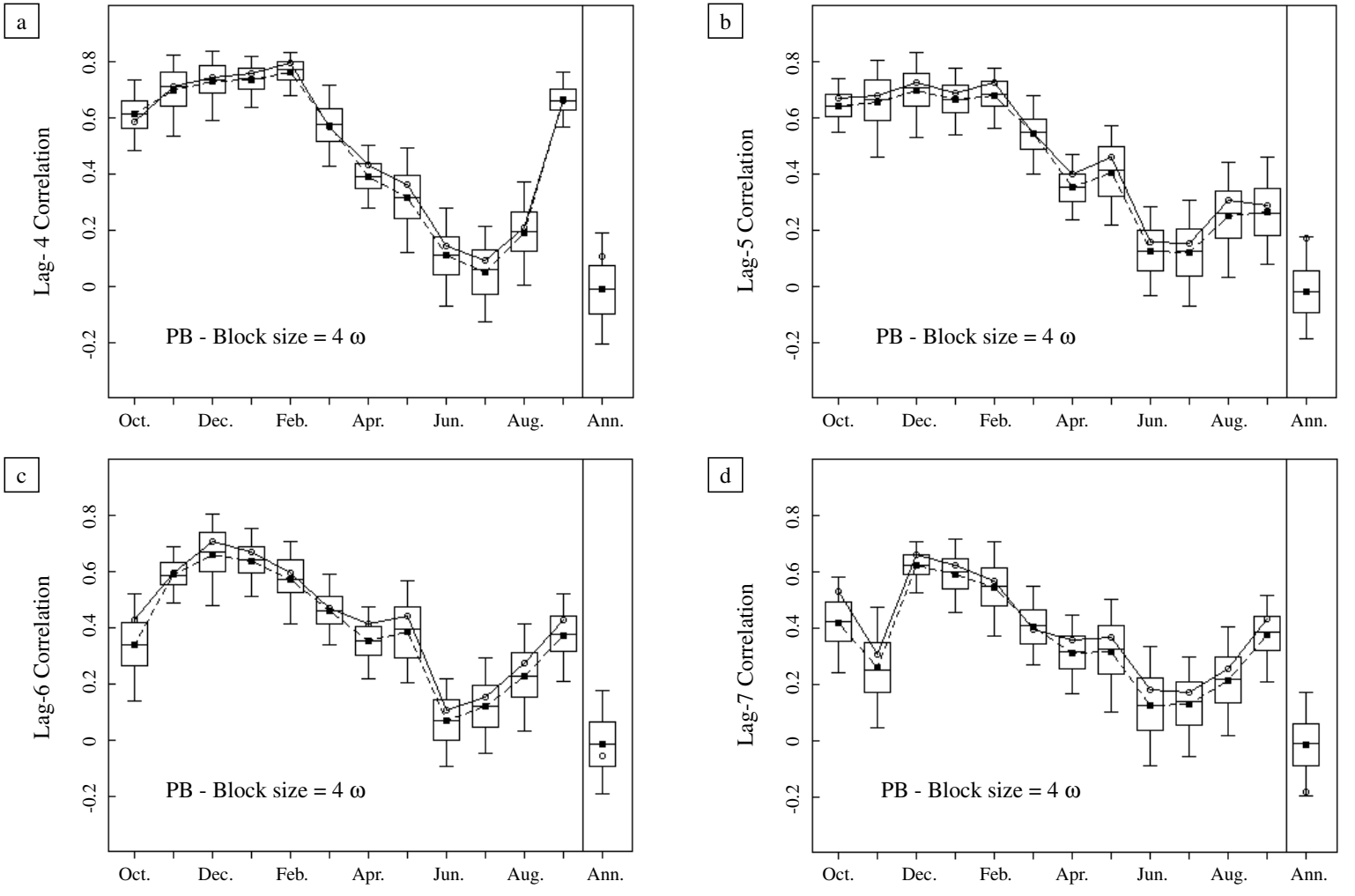

Fig. 13. Preservation of higher lag monthly serial correlations by the PB model for the Beaver River.

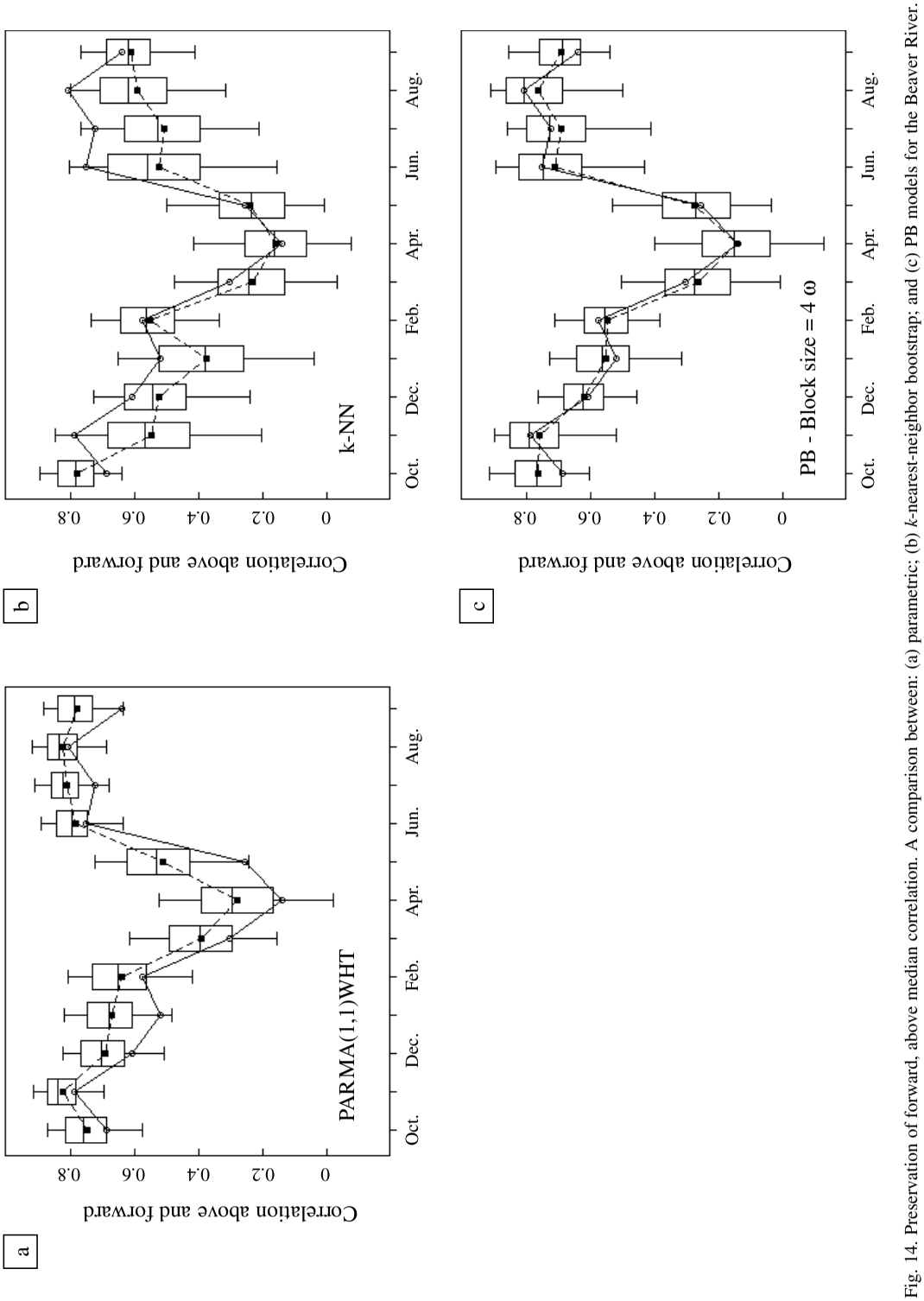

Fig. 14. Preservation of forward, above median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River.

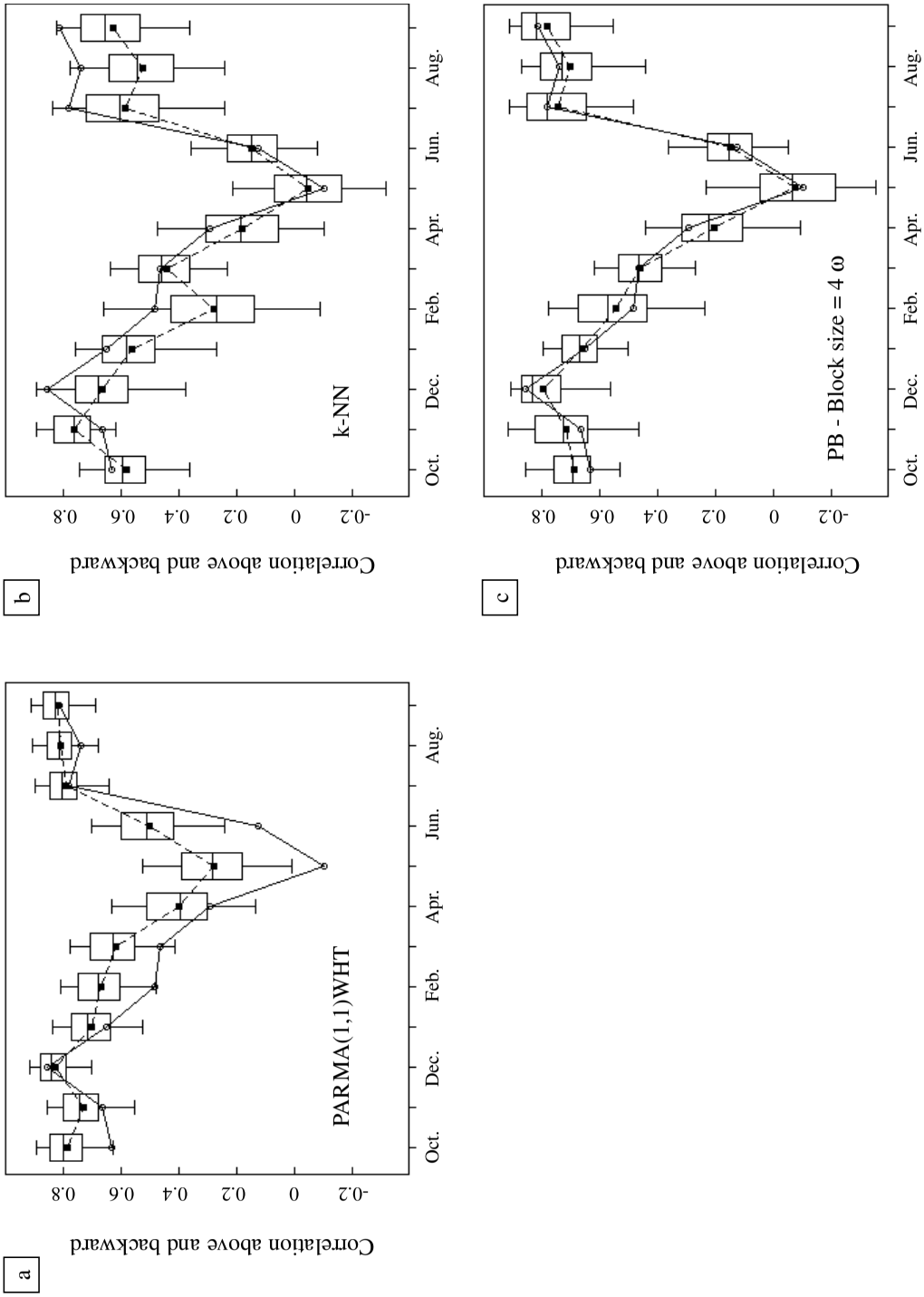

Fig. 15. Preservation of backward, above median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River.

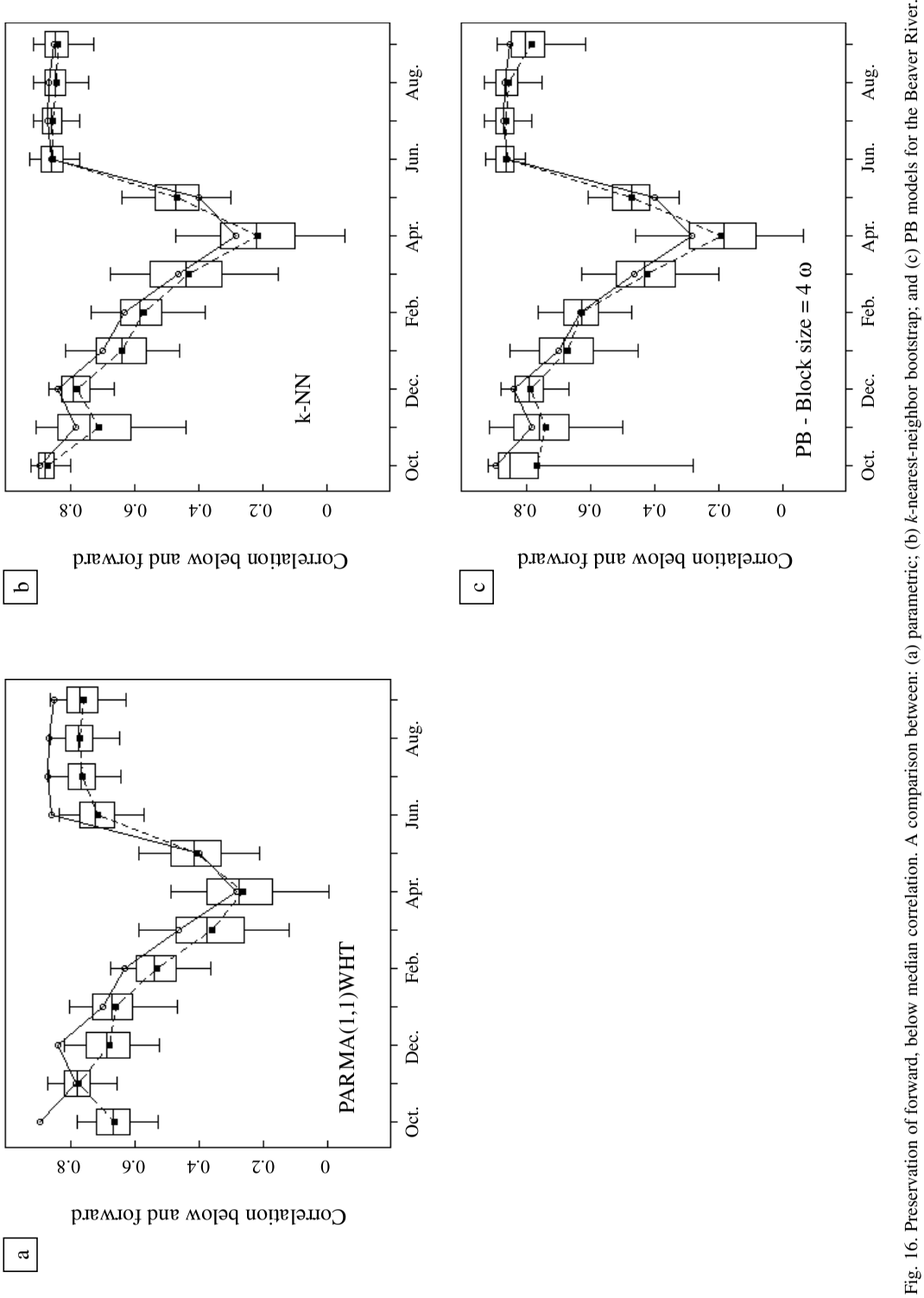

Fig. 16. Preservation of forward, below median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River.

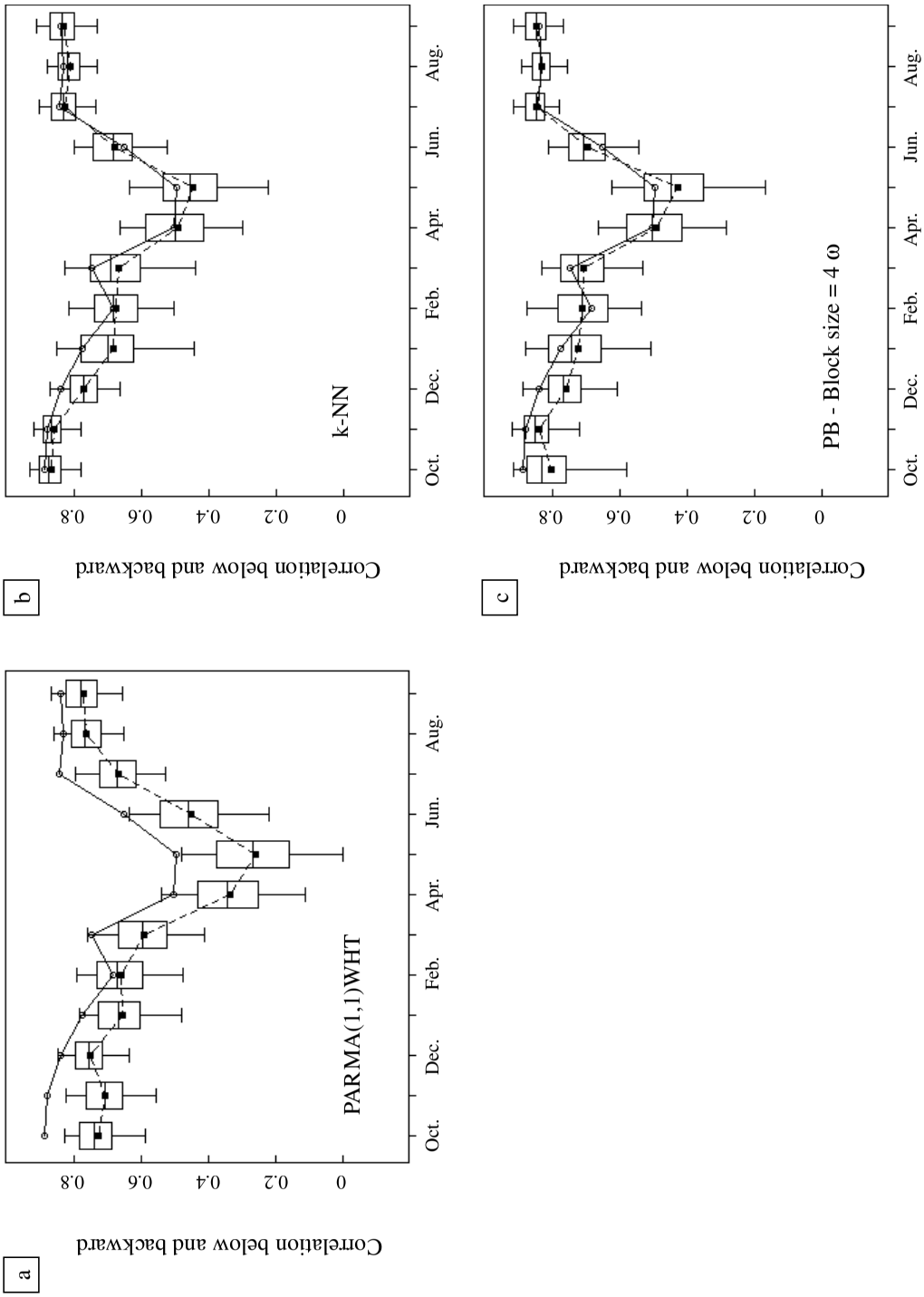

Fig. 17. Preservation of backward, below median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Beaver River.

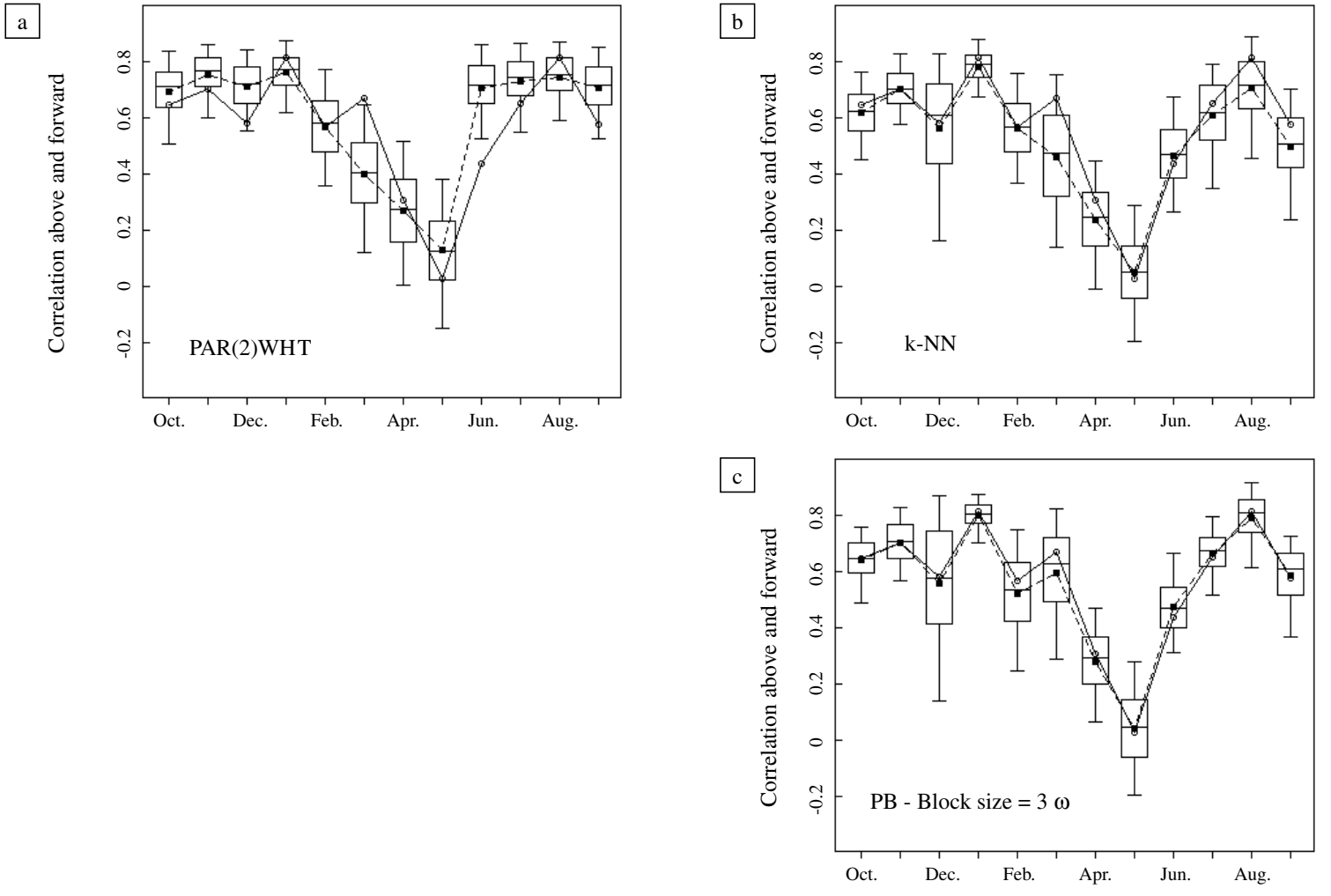

Fig. 18. Preservation of forward, above median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Weber River.

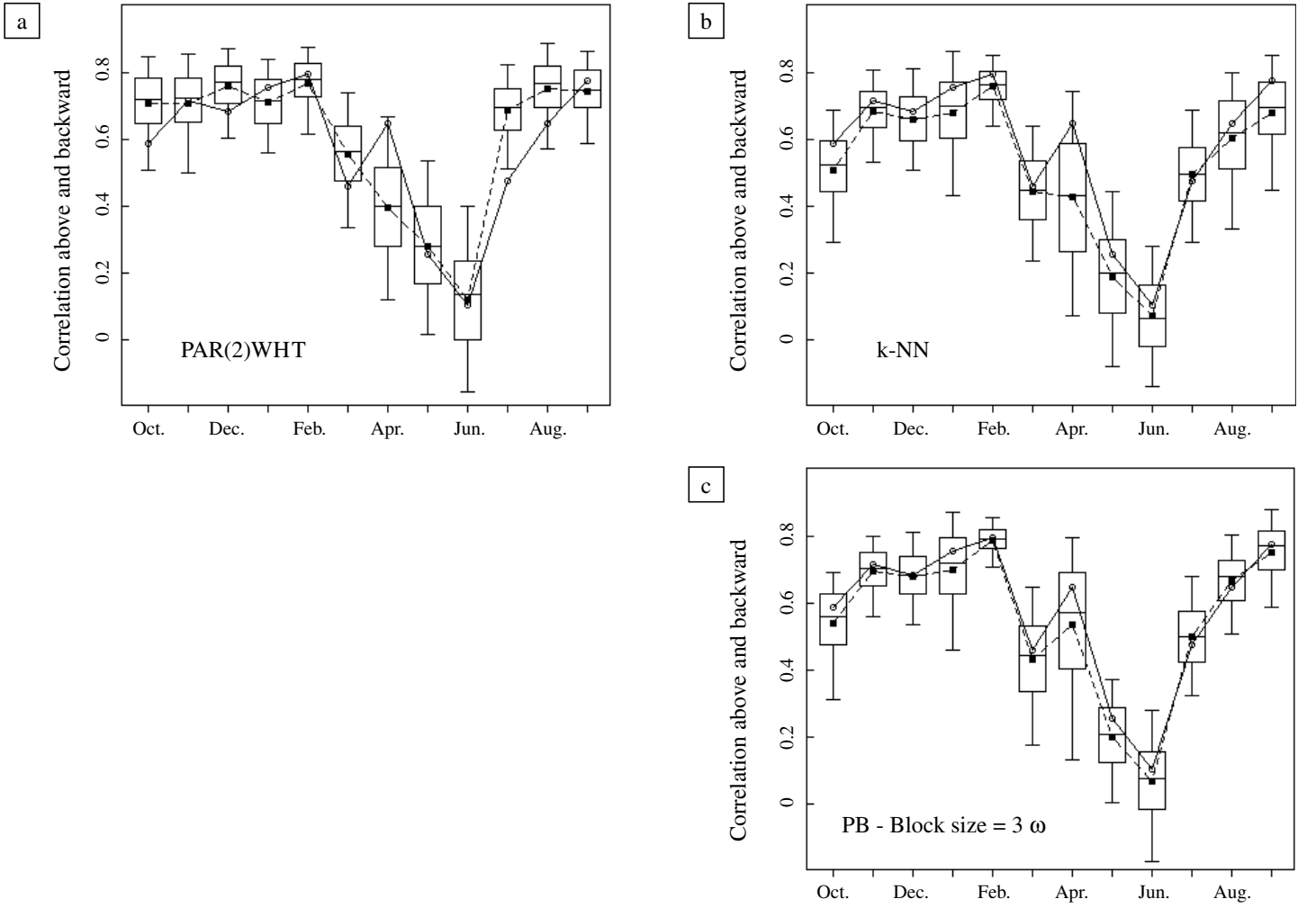

Fig. 19. Preservation of backward, above median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Weber River.

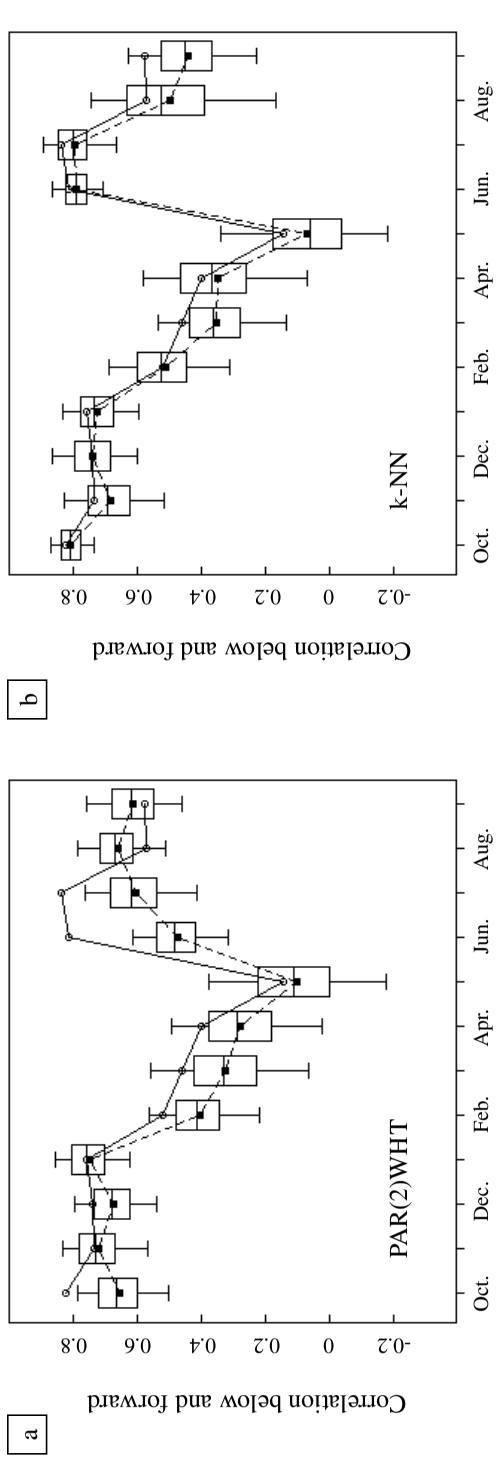

Fig. 20. Preservation of forward, below median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Weber River.

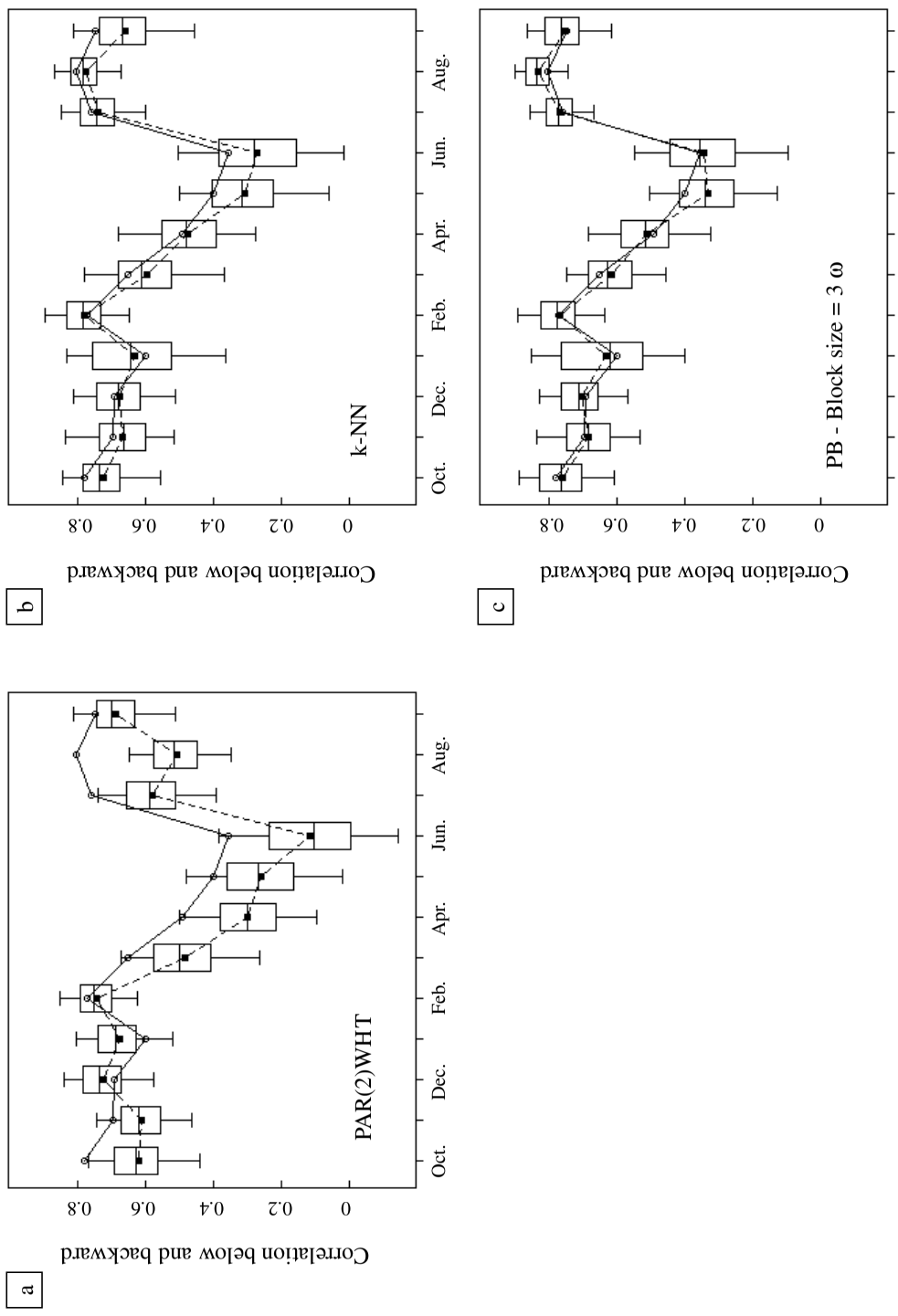

Fig. 21. Preservation of backward, below median correlation. A comparison between: (a) parametric; (b) *k*-nearest-neighbor bootstrap; and (c) PB models for the Weber River.

Table 7
Preservation of the month-to-month cross-correlations by the PB model. River: Beaver (figures in parentheses denote the standard deviation over 500 synthetic replicates)

Month		Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Oct.	Hist.	1.000	0.876	0.801	0.799	0.794	0.539	0.407	0.361	0.151	0.144	0.199	0.359
	PB	1.000	0.883	0.794	0.786	0.762	0.540	0.347	0.314	0.101	0.135	0.264	0.326
		(0.000)	(0.054)	(0.057)	(0.051)	(0.048)	(0.082)	(0.077)	(0.119)	(0.129)	(0.136)	(0.119)	(0.117)
Nov.	Hist.		1.000	0.893	0.788	0.823	0.565	0.396	0.439	0.174	0.194	0.303	0.429
	PB		1.000	0.883	0.781	0.796	0.569	0.348	0.382	0.120	0.140	0.227	0.336
			(0.000)	(0.036)	(0.053)	(0.036)	(0.091)	(0.072)	(0.116)	(0.130)	(0.137)	(0.151)	(0.115)
Dec.	Hist.			1.000	0.846	0.777	0.506	0.429	0.454	0.103	0.168	0.257	0.380
	PB			1.000	0.839	0.755	0.506	0.386	0.402	0.063	0.127	0.195	0.307
				(0.000)	(0.037)	(0.057)	(0.091)	(0.070)	(0.112)	(0.096)	(0.106)	(0.122)	(0.088)
Jan.	Hist.				1.000	0.813	0.548	0.380	0.357	0.151	0.147	0.250	0.365
	PB				1.000	0.815	0.572	0.347	0.309	0.121	0.115	0.211	0.315
					(0.000)	(0.061)	(0.101)	(0.098)	(0.113)	(0.097)	(0.104)	(0.113)	(0.093)
Feb.	Hist.					1.000	0.755	0.484	0.392	0.137	0.148	0.268	0.427
	PB					1.000	0.750	0.436	0.354	0.105	0.115	0.225	0.374
						(0.000)	(0.045)	(0.078)	(0.107)	(0.103)	(0.115)	(0.119)	(0.092)
Mar.	Hist.						1.000	0.546	0.322	0.053	0.087	0.304	0.424
	PB						1.000	0.495	0.287	0.023	0.046	0.248	0.370
							(0.000)	(0.073)	(0.088)	(0.102)	(0.108)	(0.123)	(0.096)
Apr.	Hist.							1.000	0.382	−0.018	0.058	0.206	0.282
	PB							1.000	0.366	−0.029	0.049	0.187	0.258
								(0.000)	(0.114)	(0.115)	(0.122)	(0.109)	(0.119)
May	Hist.								1.000	0.550	0.579	0.580	0.656
	PB								1.000	0.555	0.582	0.575	0.661
									(0.000)	(0.069)	(0.065)	(0.065)	(0.058)
Jun.	Hist.									1.000	0.882	0.778	0.757
	PB									1.000	0.879	0.776	0.744
										(0.000)	(0.031)	(0.055)	(0.072)
Jul.	Hist.										1.000	0.888	0.840
	PB										1.000	0.884	0.831
											(0.000)	(0.030)	(0.055)
Aug.	Hist.											1.000	0.915
	PB											1.000	0.911
												(0.000)	(0.027)

prespecified demand. In this section, the reservoir storage capacities required to meet various prespecified demand levels (ranging from 50 to 95% mean annual flow (MAF), distributed uniformly over 12 months of the year) are computed, using the sequent peak algorithm (Loucks et al., 1981, p. 235), for the historical streamflows as well as the 500 synthetic replicates generated from each of the three models considered (parametric, *k*-NN and PB). A relative comparison of the errors in the prediction of the reservoir storage capacity statistics between the three models is presented in Tables 9 and 10 for the rivers Beaver and Weber, respectively, in terms of the performance measures relative bias (R-bias) and relative root mean square error (R-RMSE), given below.

$$\text{R-bias} = \frac{\left(S_h - \frac{1}{n_r}\sum_{i=1}^{n_r} S_{si}\right)}{S_h} \tag{8}$$

$$\text{R-RMSE} = \frac{\left(\frac{1}{n_r}\sum_{i=1}^{n_r}(S_h - S_{si})^2\right)^{1/2}}{S_h} \tag{9}$$

where S_h denotes the storage capacity estimated from the observed (historical) flows, S_{si} is the storage capacity estimated from the ith synthetic replicate and n_r denotes the number of synthetic replicates.

Table 8
Preservation of the month-to-month cross-correlations by the PB model. River: Weber (figures in parentheses denote the standard deviation over 500 synthetic replicates)

Month		Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
	Hist.	1.000	0.843	0.709	0.637	0.590	0.438	0.213	0.296	0.249	0.195	0.399	0.478
Oct.	PB	1.000	0.835	0.687	0.591	0.559	0.424	0.157	0.229	0.221	0.163	0.284	0.361
		(0.000)	(0.030)	(0.047)	(0.085)	(0.072)	(0.086)	(0.100)	(0.094)	(0.109)	(0.121)	(0.156)	(0.163)
	Hist.		1.000	0.866	0.782	0.726	0.658	0.399	0.443	0.271	0.158	0.345	0.432
Nov.	PB		1.000	0.853	0.751	0.700	0.650	0.341	0.393	0.254	0.143	0.272	0.364
			(0.000)	(0.029)	(0.056)	(0.063)	(0.059)	(0.104)	(0.087)	(0.107)	(0.120)	(0.108)	(0.109)
	Hist.			1.000	0.831	0.768	0.674	0.411	0.419	0.309	0.243	0.393	0.432
Dec.	PB			1.000	0.818	0.750	0.658	0.360	0.373	0.294	0.231	0.337	0.377
				(0.000)	(0.073)	(0.057)	(0.058)	(0.099)	(0.099)	(0.093)	(0.105)	(0.093)	(0.090)
	Hist.				1.000	0.889	0.671	0.426	0.462	0.304	0.189	0.355	0.391
Jan.	PB				1.000	0.880	0.669	0.382	0.432	0.303	0.188	0.315	0.349
					(0.000)	(0.028)	(0.050)	(0.094)	(0.091)	(0.080)	(0.087)	(0.100)	(0.091)
	Hist.					1.000	0.672	0.393	0.404	0.342	0.219	0.310	0.357
Feb.	PB					1.000	0.664	0.348	0.366	0.331	0.219	0.274	0.321
						(0.000)	(0.072)	(0.091)	(0.100)	(0.089)	(0.082)	(0.099)	(0.087)
	Hist.						1.000	0.637	0.419	0.225	0.117	0.130	0.177
Mar.	PB						1.000	0.597	0.401	0.202	0.113	0.104	0.145
							(0.000)	(0.107)	(0.071)	(0.140)	(0.110)	(0.115)	(0.107)
	Hist.							1.000	0.410	−0.007	0.044	0.034	0.031
Apr.	PB							1.000	0.388	−0.023	0.021	0.012	0.006
								(0.000)	(0.075)	(0.097)	(0.147)	(0.112)	(0.089)
	Hist.								1.000	0.165	0.101	0.288	0.284
May	PB								1.000	0.138	0.089	0.262	0.259
									(0.000)	(0.105)	(0.093)	(0.100)	(0.096)
	Hist.									1.000	0.672	0.754	0.624
Jun.	PB									1.000	0.692	0.759	0.622
										(0.000)	(0.063)	(0.040)	(0.058)
	Hist.										1.000	0.777	0.593
Jul.	PB										1.000	0.791	0.612
											(0.000)	(0.037)	(0.081)
	Hist.											1.000	0.881
Aug.	PB											1.000	0.873
												(0.000)	(0.034)

For the Beaver River, it is seen from Table 9 that the *k*-NN model highly underestimates the storage capacity at low (50–55% MAF) and high (85–95% MAF) demand levels. Even though the parametric model (PARMA(1,1)WHT) also exhibits considerably high R-bias at both low and high demand levels, it is significantly lower that that of the *k*-NN model. However, at the intermediate demand levels (60–80% MAF) the performance in terms of both R-bias and R-RMSE for either of these models is comparable. In contrast, the PB model (with $L = 4\omega$) exhibits much lower R-bias compared to the other two models, at low as well as high demand levels specified. Though the performance of the three models in terms of R-RMSE is comparable for the demands 50–55% MAF and 80–90% MAF, the bias in the reservoir storage is considerably higher for simulations from the parametric and *k*-NN models than those from the PB model, indicating inadequacy of both the parametric and *k*-NN models in predicting the reservoir storage capacity for this river's flow data. The better performance of the PB model at the higher demand levels, may be attributed to the better preservation of higher-order dependence by the model (Fig. 13).

For the Weber River, it is seen from Table 10 that both the parametric and the *k*-NN models highly underestimate the storage capacity at low (50–60% MAF) as well as high (90–95% MAF)

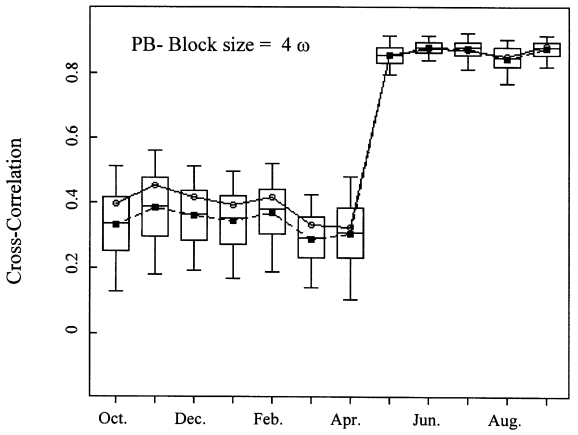

Fig. 22. Preservation of month-to-annual cross-correlation by the PB model for the Beaver River. The circle denotes correlation of historical flows and the darkened square denotes the average correlation computed from the 500 synthetic replicates.

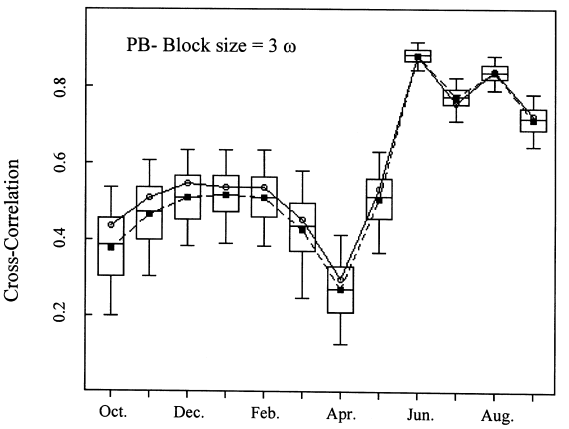

Fig. 23. Preservation of month-to-annual cross-correlation by the PB model for the Weber River. The circle denotes correlation of historical flows and the darkened square denotes the average correlation computed from the 500 synthetic replicates.

demand levels. However, for the lower demand levels, the *k*-NN model exhibits less R-bias and R-RMSE compared to the parametric model, whereas at higher demand levels, both the models exhibit nearly the same level of performance. In contrast, the PB model (with $L = 3\omega$) shows a consistently better performance at both lower and higher demand levels mentioned (Table 10) in terms of both R-bias and R-RMSE. This may be attributed to the better preservation of the correlation structure by the PB model at both periodic and annual levels. For the PB model, an alternative block choice ($L = 5\omega$) is also presented that enables appreciation of the flexibility offered by the model. However, for $L = 5\omega$, the R-bias of the storage capacity estimate at higher demand levels (90–95% MAF) reduces considerably. Herein, we

Table 9
Preservation of storage capacity statistics. A comparison between: (a) parametric; (b) *k*-nearest-neighbor (*k*-NN) bootstrap; and (c) PB models. River: Beaver

Demand level	Hist. storage capacity (million m^3)	Model							
		Parametric		*k*-NN		PB			
		PARMA(1,1)WHT				$L = 4\omega$		$L = 5\omega$	
		R-Bias	R-RMSE	R-Bias	R-RMSE	R-Bias	R-RMSE	R-Bias	R-RMSE
50	16.90	0.184	0.310	0.291	0.338	0.163	0.299	0.156	0.273
55	22.21	0.166	0.300	0.220	0.293	0.123	0.265	0.113	0.233
60	28.41	0.142	0.291	0.157	0.269	0.080	0.248	0.072	0.213
65	34.62	0.087	0.284	0.081	0.265	0.013	0.249	0.013	0.219
70	40.95	0.006	0.302	−0.006	0.287	−0.085	0.302	−0.068	0.268
75	57.97	0.101	0.305	0.093	0.284	0.001	0.288	0.019	0.261
80	78.34	0.144	0.325	0.152	0.306	0.043	0.306	0.054	0.282
85	113.25	0.219	0.372	0.255	0.356	0.134	0.338	0.140	0.318
90	166.99	0.276	0.432	0.345	0.421	0.198	0.417	0.201	0.391
95	226.54	0.191	0.506	0.310	0.468	0.096	0.547	0.112	0.510

Table 10
Preservation of storage capacity statistics. A comparison between: (a) parametric; (b) *k*-nearest-neighbor (*k*-NN) bootstrap; and (c) PB models. River: Weber

Demand level	Hist. storage capacity (million m^3)	Model							
		Parametric		*k*-NN		PB			
		PAR(2)WHT				$L = 3\omega$		$L = 5\omega$	
		R-Bias	R-RMSE	R-Bias	R-RMSE	R-Bias	R-RMSE	R-Bias	R-RMSE
50	68.13	0.294	0.321	0.219	0.271	0.061	0.149	0.060	0.140
55	84.73	0.284	0.320	0.180	0.245	0.048	0.141	0.048	0.132
60	101.85	0.249	0.296	0.138	0.221	0.033	0.132	0.035	0.121
65	119.28	0.197	0.260	0.096	0.208	0.015	0.134	0.018	0.120
70	136.71	0.132	0.228	0.046	0.211	−0.015	0.160	−0.005	0.131
75	158.53	0.073	0.219	0.009	0.232	−0.053	0.205	−0.046	0.163
80	195.87	0.061	0.230	0.028	0.244	−0.051	0.241	−0.055	0.208
85	235.87	−0.002	0.256	−0.003	0.261	−0.120	0.319	−0.159	0.339
90	397.06	0.200	0.317	0.227	0.317	0.088	0.316	−0.002	0.374
95	776.78	0.374	0.472	0.424	0.486	0.278	0.463	0.168	0.485

wish to mention that caution has to be exercised in opting for such a high block size, since the smoothing effect gets drastically reduced. Although the R-RMSE of the storage capacity estimate of the three models are quite comparable at higher demand levels, it is to be noted that the parametric and the *k*-NN models exhibit high amount of bias in addition to poor variation, which are undesirable.

4.5. Selection of block size

In the PB method applied to monthly streamflow modeling, the block sizes considered for resampling the residuals obtained from the periodic parametric model fitted are in multiples of 12. Just like in the case of the MBB technique, the selection of the appropriate block size is subjective. It depends on the particular statistic being modeled for the application of interest. We have investigated the effect of block size on the preservation of a wide variety of statistics from historical flows. For want of space, only a few of these results are presented herein.

For the Weber River, the improvement obtained in preservation of skewness, lag-3 and lag-5 serial correlations and month-to-annual cross-correlations, is shown in Figs. 24–27. In general, it may be noted from these figures that the preservation of the statistics of historical flows improves with increase in block size. However, beyond a particular block size (in this case, $L = 3\omega$), no substantial improvement is seen in the preservation of the various statistics considered, besides a drastic reduction in smoothing in the simulated replicates. Hence, the block size of $L = 3\omega$ is selected for the Weber River.

5. Summary and conclusions

In this article, the PB approach is extended to modeling periodic streamflows. In the first part of the paper, the hybrid character of the PB model is demonstrated through Monte-Carlo simulations performed on hypothetical data sets drawn from a known self-exciting seasonal threshold ARMA model. Following this, the performance of PB model is compared with that of the low-order periodic parametric models (commonly used in hydrology) and the recently introduced *k*-NN bootstrap method.

For the PB approach, only a simple PAR(1) model (a default option/model commonly used by stochastic hydrologist to model monthly streamflows) is used as the underlying parametric model. For the *k*-NN model, the model order d and number of nearest neighbors k are adopted as 1, and square root of the sample size, respectively (as recommended by Lall and Sharma, 1996). On the other hand, for the

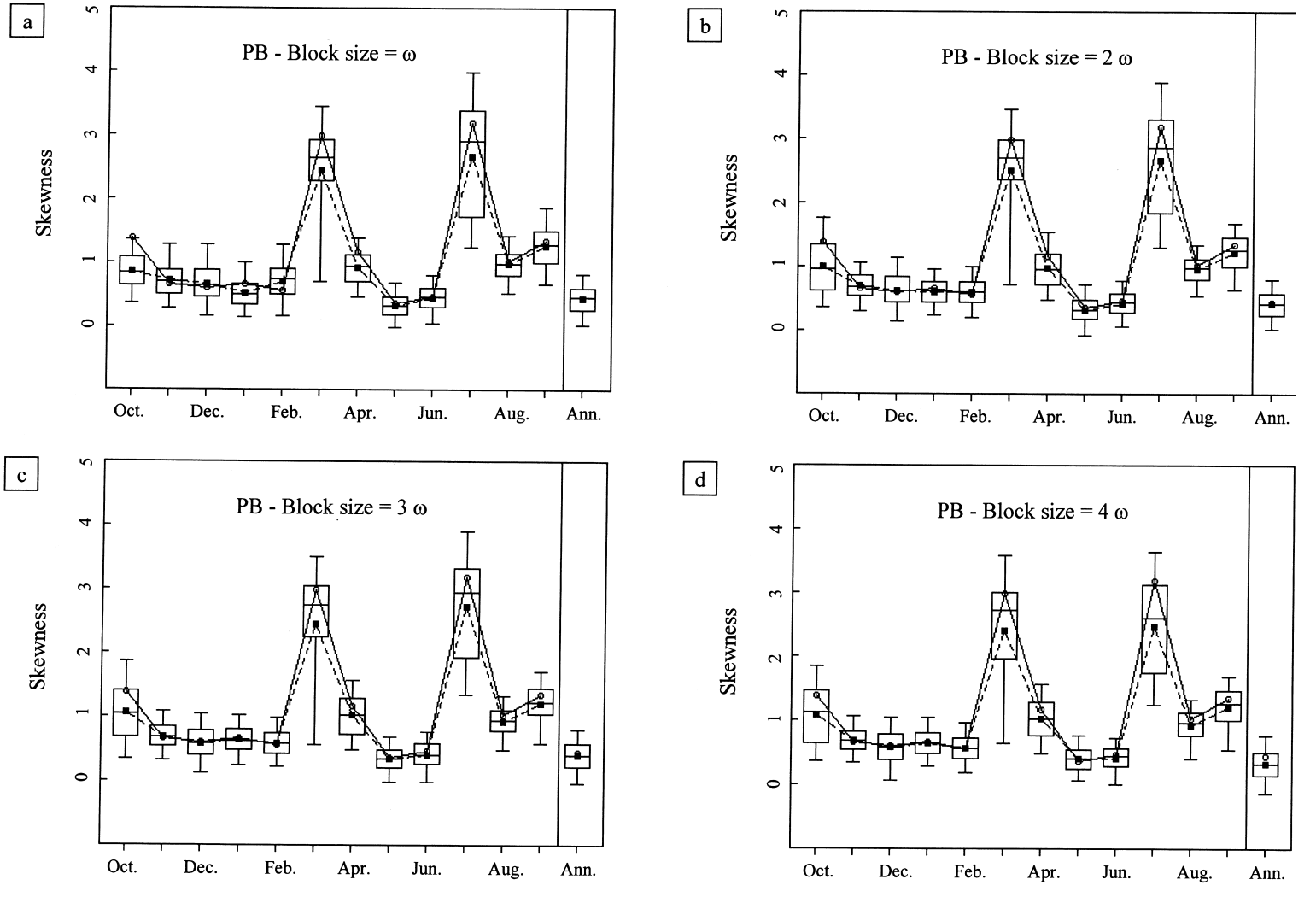

Fig. 24. Effect of block size on the preservation of skewness by the PB model for the Weber River. The circle denotes the historical value and the darkened square denotes the average statistic from 500 synthetic replicates.

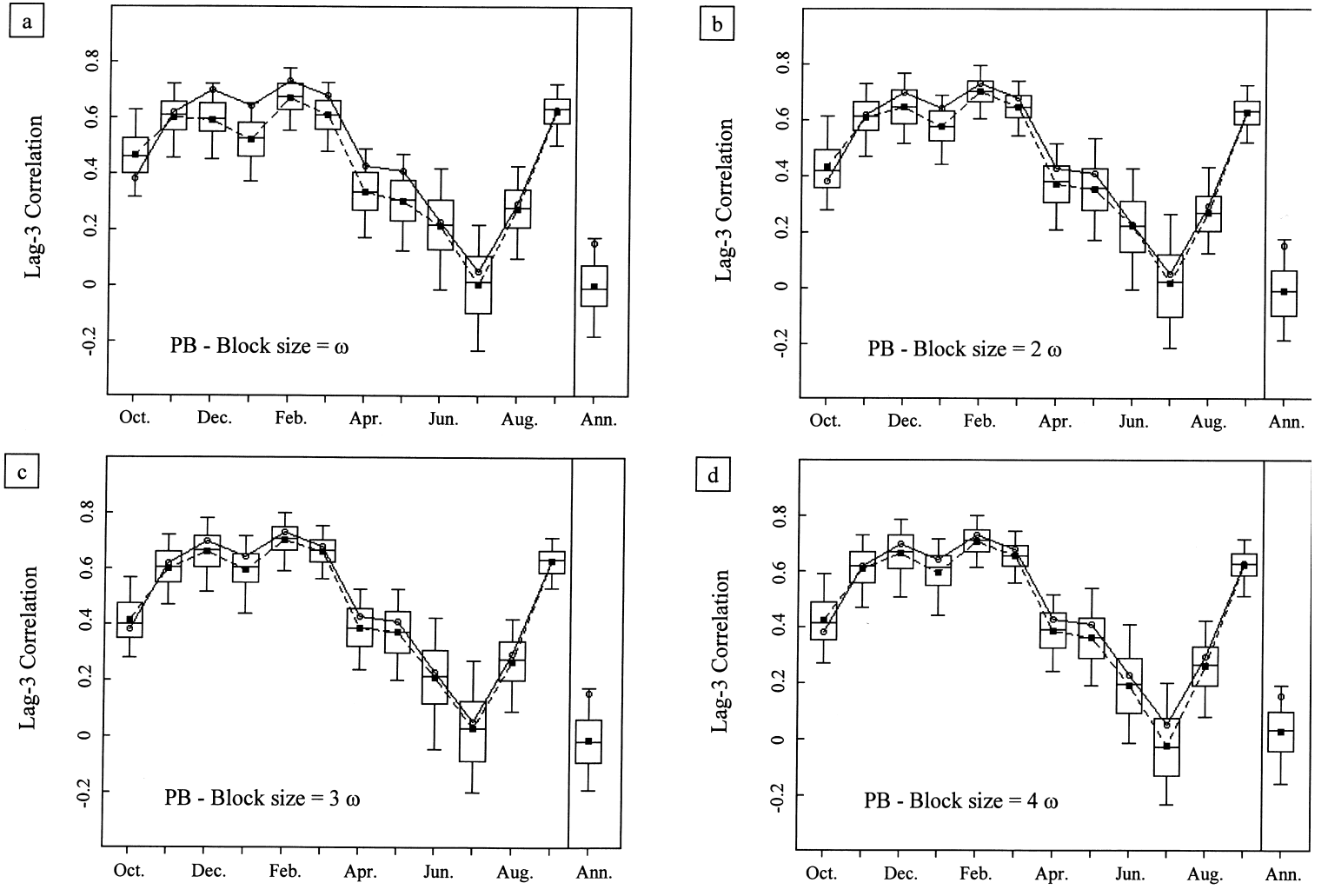

Fig. 25. Effect of block size on the preservation of lag-3 monthly serial correlation by the PB model for the Weber River.

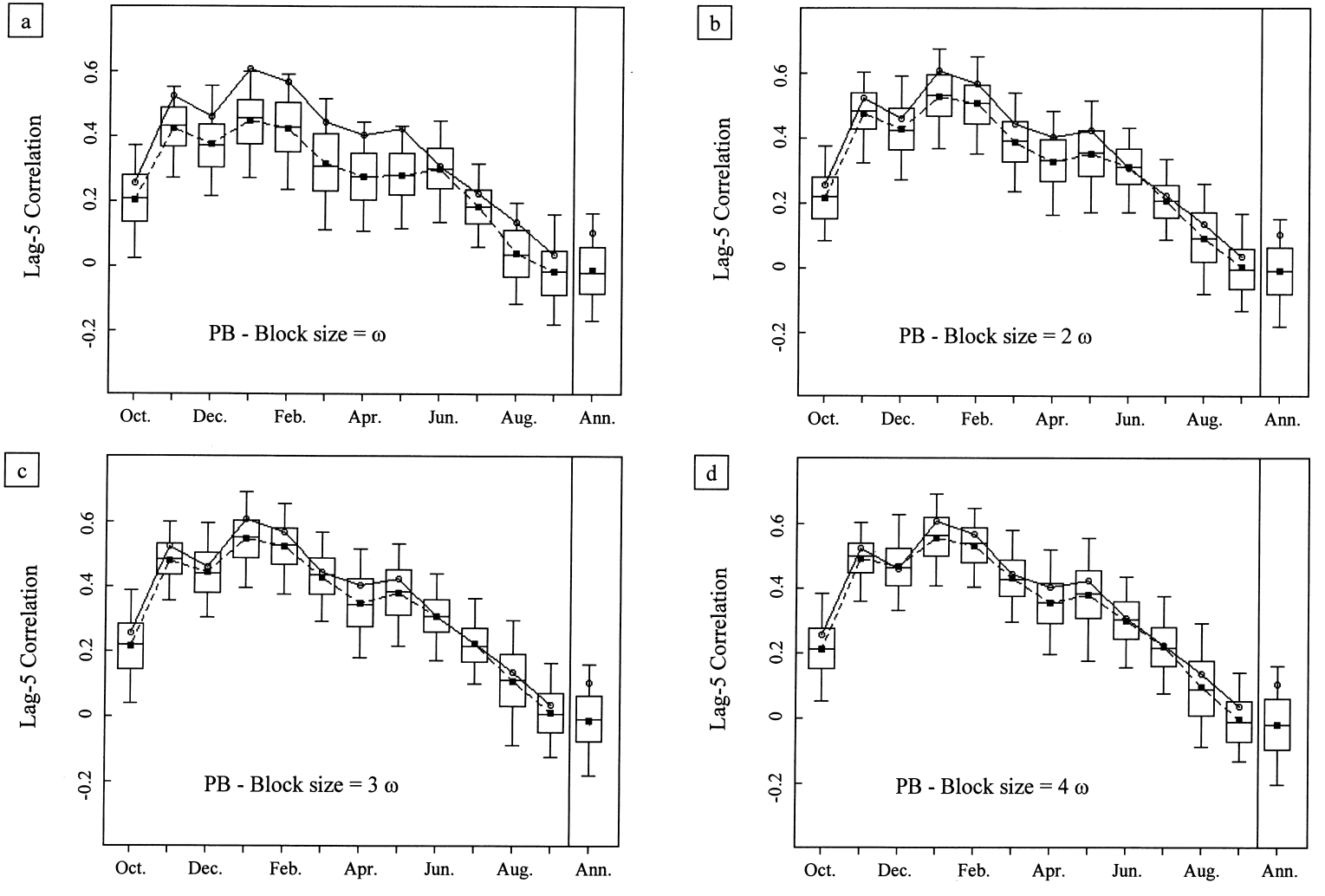

Fig. 26. Effect of block size on the preservation of lag-5 monthly serial correlation by the PB model for the Weber River.

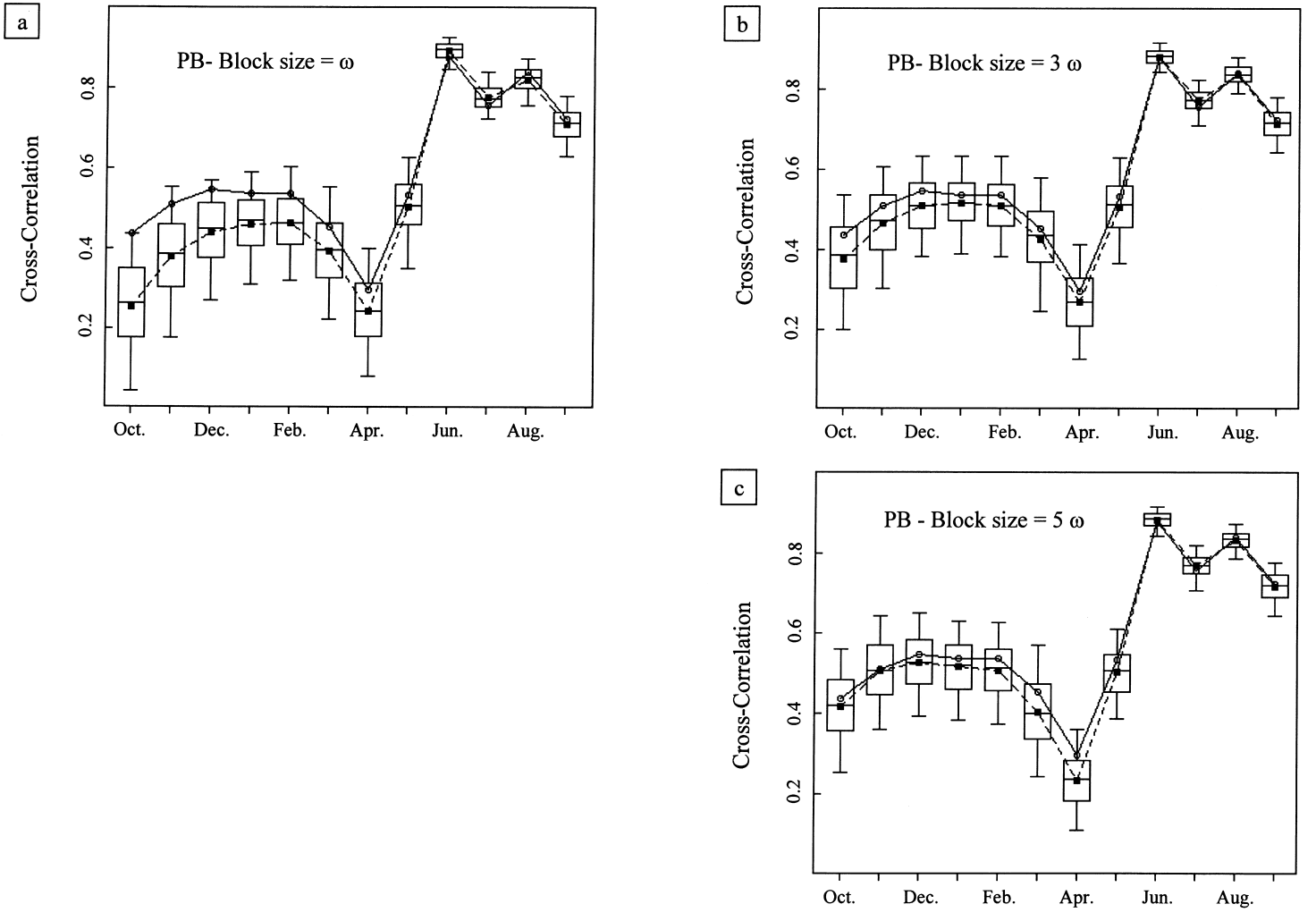

Fig. 27. Effect of block size on the preservation of month-to-annual cross-correlation by the PB model for the Weber River. The circle denotes correlation of historical flows and the darkened square denotes the mean correlation computed from the 500 synthetic replicates.

parametric approach, the best low-order linear stochastic model from among PAR(1), PAR(2) and PARMA(1,1), has been considered along with natural logarithm and WHT options. In the case of the *k*-NN and the PB models, no normalizing transformation has been applied.

A wide variety of statistics were considered for the comparison. Further, the three models were tested for their precision in predicting the reservoir storage capacity statistic, for various prespecified demand levels (ranging from 50 to 95% of the mean annual flow) for both the Beaver and the Weber rivers in USA.

The PB approach seems to gain considerably by utilizing the merits of both the parametric model and the MBB (nonparametric model adopted for resampling residuals). The PB model is found to exhibit a consistently better performance compared to the parametric and the *k*-NN models, in terms of the preservation of basic summary statistics, marginal distributions, dependence structure (monthly serial correlations, monthly state-dependent correlations and lag-1 autocorrelations at the aggregated annual level) and reservoir storage capacity statistics, in spite of no normalizing transformation being applied. In addition, it is seen that the month-to-annual and month-to-month cross-correlations are well preserved.

There seems to be considerable scope for extending the PB approach to disaggregation/multivariate/multisite modeling of geophysical data at different time levels. Furthermore, the PB approach offers considerable flexibility to the practicing hydrologist for decision making in water resources planning studies. It is also easy to implement on a personal computer.

Acknowledgements

The authors thank the Indian Institute of Technology Madras, Chennai, India, for providing the necessary facilities to carry out this research work. The authors express their gratitude to Prof. A.C. Davison, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, for all the encouragement provided and the useful discussions during the course of this research work. The authors gratefully acknowledge Prof. Upmanu Lall, Utah State University, USA, and the other anonymous reviewer for the valuable comments and suggestions on the first draft of the manuscript, which were useful in improving the quality of the paper. Special thanks are due to Dr. Ashish Sharma, University of New South Wales, Australia and Dr. Balaji Rajagopalan of Lamont-Doherty Earth Observatory, Columbia University, USA, for the fruitful discussions and clarifications regarding *k*-nearest-neighbor bootstrap.

References

Bendat, J.S., Piersol, A.G., 1986. Random Data: Analysis and measurement procedures. 2nd ed. Wiley, New York.

Bose, A., 1988. Edgeworth corrections by bootstrap in autoregressions. Ann. Statist. 16, 1709–1722.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, CA.

Bras, R.L., Rodriguez-Iturbe, I., 1985. Random Functions and Hydrology. Addison-Wesley, Reading, MA.

Carlstein, E., 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. Ann. Statist. 14, 1171–1179.

Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions. Numer. Math. 31, 377–403.

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

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1–26.

Efron, B., Tibshirani, R.J., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). Statist. Sci. 1, 54–96.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Freedman, D.A., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist. 12, 827–842.

Freedman, D.A., Peters, S.C., 1984. Bootstrapping a regression equation: some empirical results. J. Am. Stat. Assoc. 79, 97–106.

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

Hall, P., 1985. Resampling a coverage pattern. Stoch. Proc. Appl. 20, 231–246.

Hausman, E.D., 1990. Analysis of Water Supply and Demand in Eastern Massachusetts, Masters thesis, Tufts University, Medford, MA.

Helsel, D.R., Hirsch, R.M., 1992. Statistical Methods in Water Resources. Elsevier, New York.

Hirsch, R.M., 1979. Synthetic hydrology and water supply reliability. Water Resour. Res. 15 (6), 1603–1615.

Hjorth, J.S.U., 1994. Computer Intensive Statistical Methods — Validation Model Selection and Bootstrap. Chapman and Hall, New York.

IMSL, 1984. IMSL User's Manual. IMSL, Inc. Houston, TX.

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

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

Lahiri, S.N., 1995. On the asymptotic behaviour of the moving block bootstrap for normalized sums of heavy-tail random variables. Ann. Statist. 23 (4), 1331–1349.

Lall, U., 1995. Recent advances in nonparametric function estimation: Hydrologic applications, U.S. Natl. Rep. Int. Union Geod. Geophys., 1991–1994. Rev. Geophys. 33, 1093–1102.

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

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

LePage, R., Billard, L., 1992. Exploring the Limits of Bootstrap. Wiley, New York (525 pp.).

Liu, R.Y., Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: Lepage, R., Billard, L. (Eds.). Exploring the Limits of Bootstrap. Wiley, New York, pp. 225–248.

Loucks, D.P., Stedinger, J.R., Haith, D.A., 1981. Water Resource Systems Planning and Analysis. Prentice-Hall, Englewood Cliffs, NJ (559 pp.).

Moss, M.E., Tasker, G.D., 1991. An intercomparison of hydrological network-design technologies. Hydrol. Sci. J. 36 (3), 209–221.

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 Resour. Res. 24 (5), 781–785.

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

Rajagopalan, B., Lall, U., 1999. A *k*-nearest neighbor simulator for daily precipitation and other weather variables. Water Resour. Res. 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 Resour. Res. 32 (10), 3151–3160.

Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1980. Applied modelling of hydrologic time series. Water Resources Publications, Fort Collins, CO.

Salas, J.D., Smith, R.A., Markus, M., 1992. Modeling and generation of univariate seasonal hydrologic data (Programs CSU001 and CSU002). Technical Report 2, Colorado State University, Fort Collins, CO.

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

SAS/ETS, 1988. SAS/ETS User's Guide, version 6, 1st ed., SAS Inc., Cary, NC.

Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York.

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

Sharma, A., Lall, U., 1999. A nonparametric approach for daily rainfall simulation. Math. Comput. Simul. 48, 361–371.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York (175 pp.).

Srinivas, V.V., Srinivasan, K., 2000. Post-blackening approach for modeling dependent annual streamflows. J. Hydrol. 230, 86–126.

Statgraphics, 1984. Statgraphics User Manual. Statistical Graphics Corp., Rockville, MD.

Stedinger, J.R., 1981. Estimating correlations in mutivariate streamflow models. Water Resour. Res. 17 (1), 200–208.

Stedinger, J.R., Taylor, M.R., 1982. Synthetic streamflow generation-1. Model verification and validation. Water Resour. Res. 18 (4), 909–918.

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

Tao, P.C., Delleur, J.W., 1976. Seasonal and nonseasonal ARMA models in hydrology. ASCE J. Hydrol. Div. 102 (HY10), 1541–1559.

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

Tasker, G.D., 1987. Comparison of methods for estimating low flow characteristics of streams. Water Resour. Bull. 23 (6), 1077–1083.

Tasker, G.D., Dunne, P., 1997. Bootstrap position analysis for forecasting low flow frequency. J. Water Resour. Planning Mgmt. ASCE 123 (6), 359–367.

Tong, H., 1990. Nonlinear Time Series Analysis: A Dynamical Systems Perspective. Academic Press, San Diego, CA.

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

Woo, M.K., 1989. Confidence intervals of optimal risk-based hydraulic design parameters. Can. Water Resour. J. 14 (2), 10–16.

Zucchini, W., Adamson, P.T., 1988. On the application of the bootstrap to assess the risk of deficient annual inflows to a reservoir. Water Resour. Mgmt. 2, 245–254.

Zucchini, W., Adamson, P.T., 1989. Bootstrap confidence intervals for design storms from exceedence series. Hydrol. Sci. J. 34 (1), 41–48.