WATER RESOURCES RESEARCH, VOL. 37, NO. 10, PAGES 2537–2549, OCTOBER 2001

A hybrid stochastic model for multiseason streamflow simulation

V. V. Srinivas and K. Srinivasan

Environmental and Water Resources Engineering Division, Department of Civil Engineering
Indian Institute of Technology, Madras, Chennai, India

Abstract. A hybrid model is presented for stochastic simulation of multiseason streamflows. This involves partial prewhitening of the streamflows using a parsimonious linear periodic parametric model, followed by resampling the resulting residuals using moving block bootstrap to obtain innovations and subsequently postblackening these innovations to generate synthetic replicates. This model is simple and is efficient in reproducing both linear and nonlinear dependence inherent in the observed streamflows. The first part of this paper demonstrates the hybrid character of the model through stochastic simulations performed using monthly streamflows of Weber River (Utah) that exhibit a complex dependence structure. In the latter part of the paper the hybrid model is shown to be efficient in simulating multiseason streamflows, through an example of the San Juan River (New Mexico). This model ensures annual-to-monthly consistency without the need for any adjustment procedures. Furthermore, the hybrid model is able to preserve both within-year and cross-year monthly serial correlations for multiple lags. Also, it is seen to be consistent in predicting the reservoir storage (validation) statistic at low as well as high demand levels.

1. Introduction

In hydrology, time series models are often used for stochastic simulation of streamflows. These simulations augment the description provided by the observed (historical) streamflow sequences (that are often limited in size) and are useful in the design of reservoirs, evaluation of alternate operation policies, assessment of risk and reliability of system operation, and analysis of critical droughts, to mention a few.

Since the work of *Box and Jenkins* [1970], tremendous effort has gone into developing and popularizing the conventional linear parametric time series models. *Salas et al*. [1980, 1985] and *Salas* [1993] provide reviews of parametric time series models that are used in water resources planning and management. In the context of stochastic modeling of streamflows, a major limitation of the widely used periodic parametric models is their inability to simultaneously reproduce summary statistics and dependence structure at different temporal levels. To circumvent this, linear disaggregation models were developed [*Harms and Campbell*, 1967; *Valencia and Schaake*, 1973; *Mejia and Rousselle*, 1976; *Lane*, 1979; *Grygier and Stedinger*, 1988, 1990; *Santos and Salas*, 1992]. However, these models are not parsimonious, and in addition they require empirical adjustments in order to restore summability of the disaggregated flows to the aggregate flows, in the event of normalizing transformations being applied. The increasing awareness of the need to model nonlinearity and nonstationarity in the geophysical time series has spurred the growth of nonparametric methods in several areas of hydrology in recent times [*Lall*, 1995; *Lall and Sharma*, 1996; *Vogel and Shallcross*, 1996; *Sharma et al*., 1997; *Tarboton et al*., 1998; *Rajagopalan and Lall*, 1999; *Kumar et al*., 2000]. This has gained from the development and use of nonparametric methods in more general time series analysis [*Efron*, 1979; *Künsch*, 1989; *Silverman*, 1986; *Scott*, 1992; *LePage and Billard*, 1992; *Efron and Tibshirani*, 1993; *Hjorth*, 1994; *Davison and Hinkley*, 1997]. More recently, *Tarboton et al*. [1998] have developed a nonparametric disaggregation (NPD) model. They have shown that a kernel density estimate of the joint distribution of disaggregate flow variables can form the basis for conditional simulation based on an input aggregate flow variable. Being data-driven and relatively automatic, it is able to model the nonlinearity inherent in the dependence structure of observed flows reasonably well as well as to provide a good amount of smoothing in synthetic simulations (unlike simple bootstrap methods). However, it is data and computationally intensive [*Tarboton et al*., 1998, p. 118].

Paper number 2000WR900383.
0043-1397/01/2000WR900383$09.00

An ideal single-site multiseason synthetic flow generation model should aim to reproduce the following statistics of observed streamflows: (1) the summary statistics (mean, standard deviation, and skewness) and marginal distribution of observed flows at periodic and annual timescales; (2) within-year and cross-year serial correlations; (3) autocorrelation structure at aggregated annual level; (4) month-to-year cross correlations; and (5) nonlinearity and nonstationarity in the underlying dependence structure. In addition, it should provide sufficient variety in the stochastic simulations with a reasonable degree of smoothing and extrapolation.

While the conventional parametric models require assumptions regarding the marginal distribution of flows and the order of dependence, the nonparametric methods are, in general, data-driven and can capture the linear and nonlinear dependence of observed flows without any prior assumptions [*Lall and Sharma*, 1996; *Vogel and Shallcross*, 1996; *Sharma et al*., 1997; *Tarboton et al*., 1998]. While parametric models provide considerable smoothing and extrapolation in the simulations, nonparametric bootstrap methods such as the moving block bootstrap [*Vogel and Shallcross*, 1996] and k-nearest-neighbor bootstrap [*Lall and Sharma*, 1996] cannot. They simply mimic

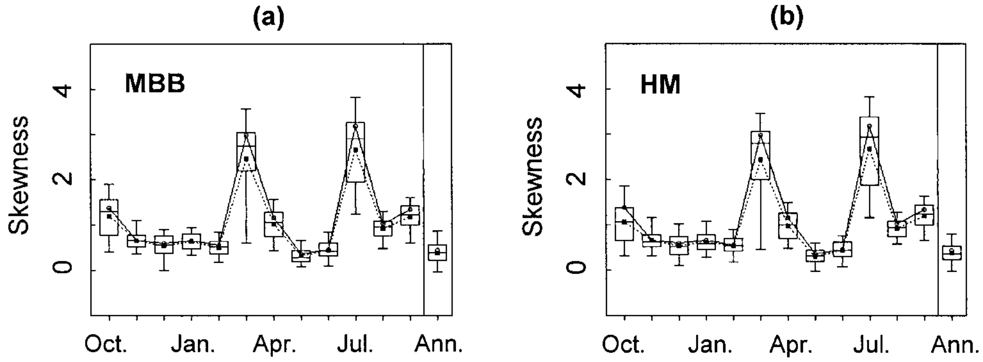

Figure 1. Preservation of skewness of Weber River streamflows at monthly and aggregated annual levels by the moving block bootstrap (MBB) model and the hybrid model (HM) graphed using box plots. A line in the middle of the box represents median. The historical statistic is represented by a circle, and the mean of the generated statistic over 100 replicates is represented by a solid square. The solid line that joins the circles indicates the historical trend, while the dotted line connecting the solid squares depicts the mean synthetic trend.

the marginal distribution of observed flows, because flow values are resampled from the historic data. Such parsing of the data defeats the purpose of synthetic streamflow simulation.

Considering the relative merits and demerits of both simple low-order linear periodic parametric models and the nonparametric bootstrap methods, we felt that simulations from a novel method that blends the merits of both parametric and nonparametric methods can represent the uncertainty in the historical streamflow record better. The Postblackening method suggested by *Davison and Hinkley* [1997] seems to be an appropriate one in this context. In this hybrid method the first step is to prewhiten the historical trace by fitting a linear 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 subjecting the residuals obtained from the linear parametric model to "moving block resampling," with a view to capturing the weak linear dependence (if any) present in the residuals and inherent nonlinearities. The innovation series is then "postblackened" by applying the estimated model to the resampled innovations. The effectiveness of the postblackening approach in modeling dependent annual streamflows has been brought out in an earlier paper by *Srinivas and Srinivasan* [2000], wherein this approach is shown to be efficient in modeling critical run characteristics of multiyear droughts.

In this paper, to start with, an attempt is made to gain some understanding of the roles played by the two constituents of the hybrid model (HM) (a simple linear parsimonious parametric model with no normalizing transformation (PAR(1)-NT) and the moving block bootstrap (MBB)) in enhancing the performance of HM over its constituents in the context of periodic streamflow modeling. For this purpose, synthetic simulations of Weber River (Utah) streamflows are used. Following this, a split sample validation test is performed on the Weber River monthly streamflows to show that the hybrid model is able to capture repeatable statistical structure present in the observed streamflows. In section 5, HM is used to simulate the historical monthly streamflow record of San Juan River, near Archuleta, New Mexico. The simulation results from HM are compared with those from the popular parametric disaggregation package SPIGOT [*Grygier and Stedinger*, 1990] and the nonparametric disaggregation model NPD [*Tarboton et al.*, 1998].

2. Algorithm for the Hybrid Model (HM)

This section presents the algorithm for generating synthetic seasonal streamflows by the hybrid model proposed, which uses the postblackening approach suggested by *Davison and Hinkley* [1997]. It is to be noted that vectors will be represented by bold uppercase letters, and the elements of the vectors will be represented by lowercase letters.

Let the observed (historical) streamflows be represented by the vector $\mathbf{Q}_{\nu,\tau}$, where ν is the index for year ($\nu = 1, \ldots, N$) 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 $\mathbf{Q}_{\nu,\tau}$ as

$$y_{\nu,\tau} = \frac{q_{\nu,\tau} - \bar{q}_\tau}{s_\tau}, \tag{1}$$

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

2. Prewhiten the standardized historical streamflows, $\mathbf{Y}_{\nu,\tau}$, using a simple periodic autoregressive model of order one (PAR(1)) and extract the residuals $\boldsymbol{\varepsilon}_{\nu,\tau}$. Take $y_{1,0} = 0$:

$$\varepsilon_{\nu,\tau} = y_{\nu,\tau} - \phi_{1,\tau} y_{\nu,\tau-1}. \tag{2}$$

In (2), $\phi_{1,1}, \ldots, \phi_{1,\omega}$ are the periodic autoregressive parameters of order one. For the parameter estimation, a simple method of moments [*Salas et al.*, 1980] has been used. It is to be noted that the residuals $\boldsymbol{\varepsilon}_{\nu,\tau}$ may possess some weak dependence (since the parameters are estimated from a simple PAR(1) model). We wish to mention that bootstrap schemes like the moving block bootstrap (MBB) [*Künsch*, 1989] can serve as reliable tools for modeling the weak linear dependence, if any, in the residuals.

3. Obtain the simulated innovations $\boldsymbol{\varepsilon}^*_{\nu,\tau}$ by bootstrapping

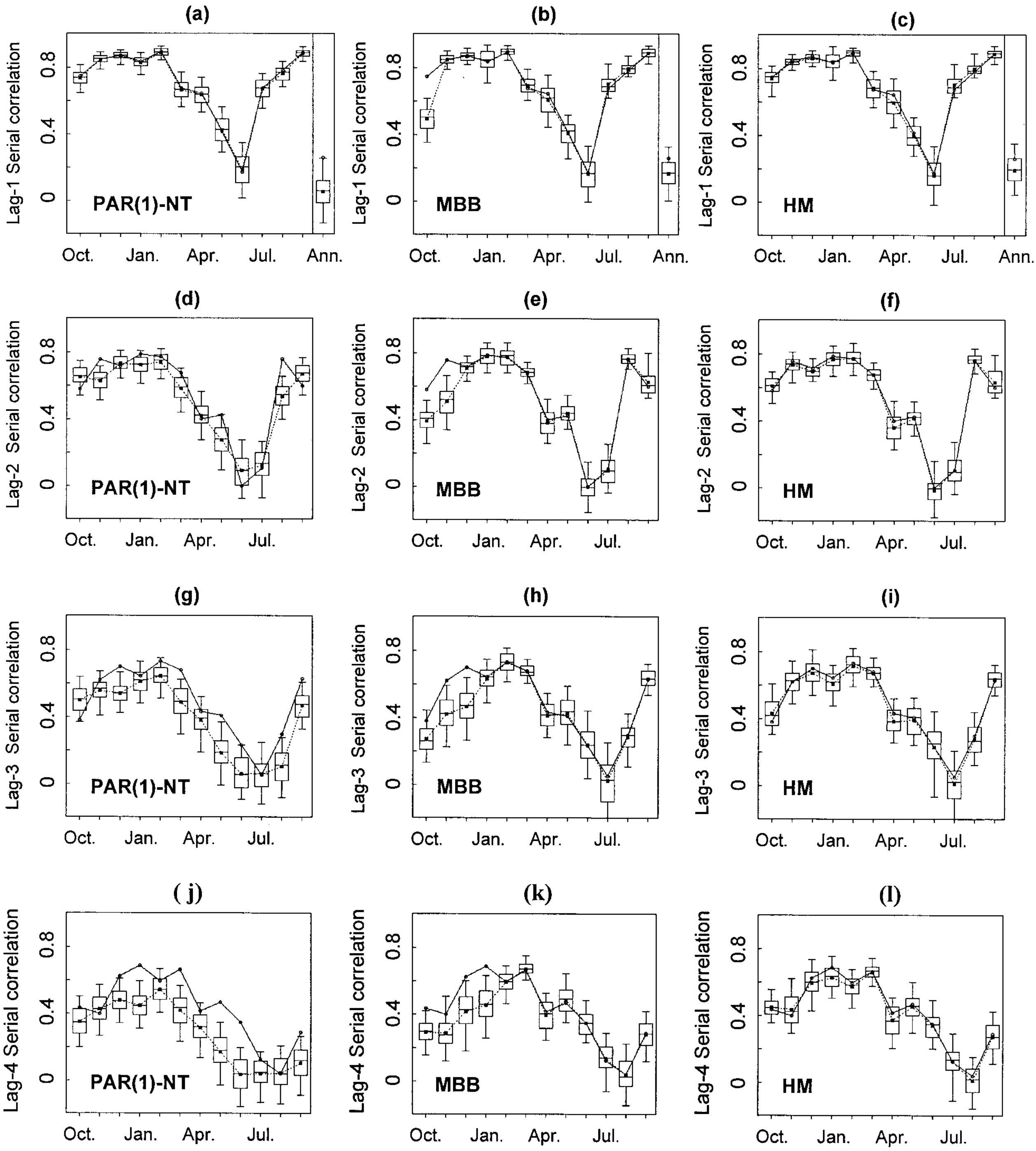

Figure 2. Preservation of serial correlations of Weber River streamflows at monthly and aggregated annual levels by periodic autoregressive model of order one with no normalizing transformation (PAR(1)-NT), moving block bootstrap (MBB), and hybrid model (HM).

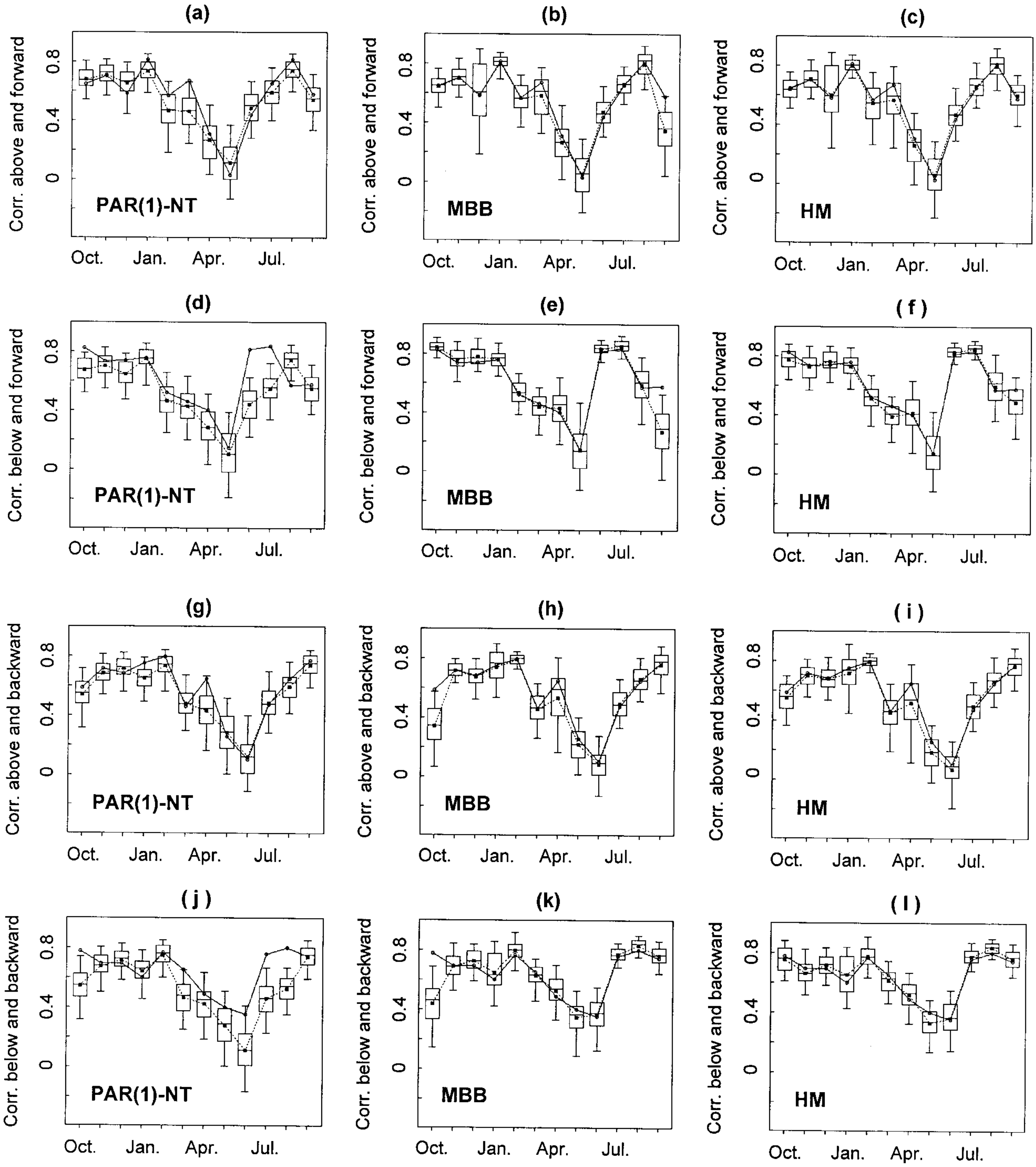

Figure 3. Preservation of state-dependent correlations of monthly streamflows of Weber River by periodic autoregressive model of order one with no normalizing transformation (PAR(1)-NT), moving block bootstrap (MBB), and hybrid model (HM).

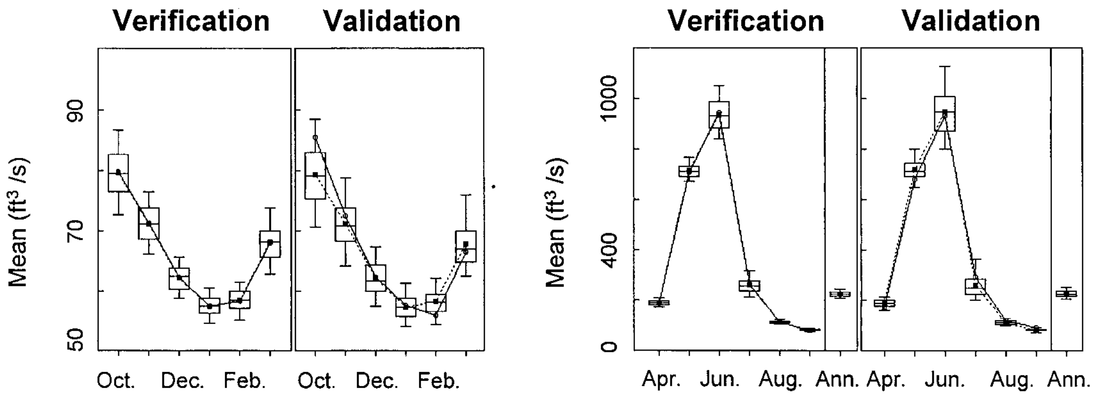

Figure 4. Results from split sample validation of hybrid model. Shown is the preservation of historical mean of monthly and annual streamflows, Weber River.

$\boldsymbol{\varepsilon}_{\nu,\tau}$ using the moving block bootstrap (MBB) method. The monthly residuals resulting from the PAR(1) model are divided into (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 in a hydrological water year. This is done with a view to capturing the within-year correlations for a 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 the block length L is n years long, the overlap is $(n - 1)$ years, so that when it is 1 year long there is no overlap. In general, the ith block with size $L = m\omega$, may be written as

$$B_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i+m-1,\omega}), \tag{3}$$

where $i = 1, \ldots, q$ and $q = N - m + 1$. For example, if $L = 3\omega$ and $\omega = 12$, the fourth block is written as $B_4 = (\varepsilon_{4,1}, \ldots, \varepsilon_{6,12})$. 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. Bootstrapped innovations $\boldsymbol{\varepsilon}^*_{\nu,\tau}$ are generated by resampling the overlapping blocks B_i at random, with replacement from the set $(B_1, \ldots, B_q)$ and pasting them end-to-end. It is to be noted that each of the (possibly) overlapping blocks has equal probability $(1/q)$ of being resampled.

4. The bootstrapped innovation series $\boldsymbol{\varepsilon}^*_{\nu,\tau}$ is then post-blackened by reversing (2) to obtain the sequence $\mathbf{Z}_{\nu,\tau}$:

$$z_{\nu,\tau} = \phi_{1,\tau} z_{\nu,\tau-1} + \varepsilon^*_{\nu,\tau}. \tag{4}$$

The synthetic generation process is started with $z_{1,0} = 0$. The "burn-in" or "warm-up" period is chosen to be large enough to remove any initial bias. The values of $\mathbf{Z}_{\nu,\tau}$ are then inverse standardized (using equation (5)) to obtain the synthetic streamflow replicate $\mathbf{X}_{\nu,\tau}$:

$$x_{\nu,\tau} = (z_{\nu,\tau} \times s_\tau) + \bar{q}_\tau. \tag{5}$$

It is to be noted that no normalizing transformation is applied in the case of the hybrid model. In this context 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 partial 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 the partial prewhitening stage tends to zero, and hence one need not recenter the residuals.

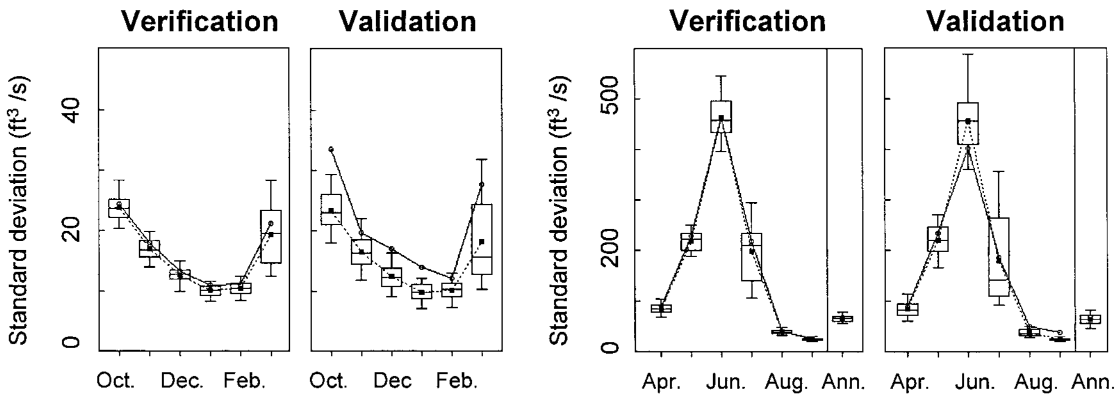

Figure 5. Results from split sample validation of hybrid model. Shown is the preservation of standard deviation of monthly and annual streamflows, Weber River.

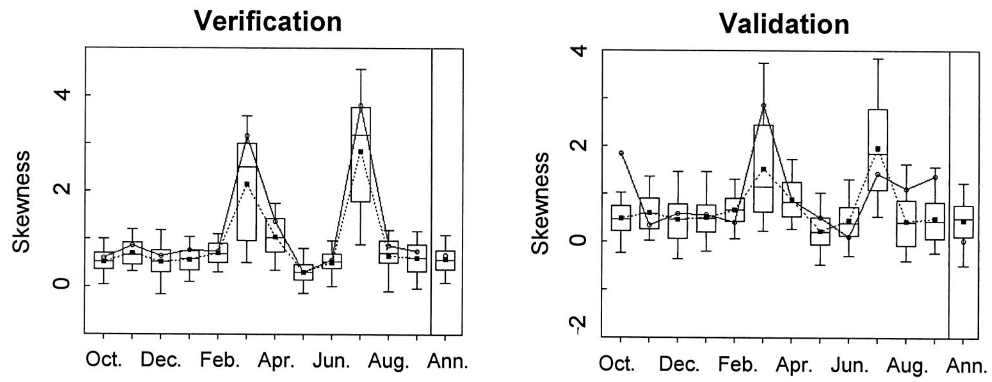

Figure 6. Results from split sample validation of hybrid model. Shown is the preservation of skewness of monthly and annual streamflows, Weber River.

3. Hybrid Effect

The results presented in this section aim to bring out the efficacy of the hybrid model (HM) in effectively blending the two constituents, namely, a simple periodic autoregressive model of order one with no normalizing transformation (PAR(1)-NT) and moving block bootstrap (MBB), so that the resulting simulations are statistically indistinguishable from the historical streamflows. The data set chosen for the illustration is the 83-year (1905–1988) record of observed monthly streamflows of the Weber River, near Oakley, Utah (U.S. Geological Survey station number 10128500). This streamflow data set has been chosen because it displays a complex linear dependence structure extending over a number of lags and a reasonable length of reliable streamflow record is available. Recently, this data set has been used by *Lall and Sharma* [1996] for modeling periodic streamflows. To illustrate the hybrid effect, a reasonable block size of $L = 3\omega$ is used for both HM and MBB. In order to enable the appreciation of the hybrid effect, the results of performance of the hybrid model are presented alongside those of its own constituents, namely, PAR(1)-NT and MBB, in the form of box plots.

The mean and the standard deviation of observed streamflows are well reproduced by the hybrid model and its constituents (PAR(1)-NT and MBB) at both monthly and aggregated annual levels. The same is not presented herein, for brevity. Being a data-driven model, MBB reproduces skewness of flows at monthly and aggregated annual levels (see Figure 1a). In the case of HM, no normalizing transformation is applied to the historical data, and hence skewness of historical streamflows is apparently retained in the residuals that are extracted from the partial prewhitening stage. The skewness contained in these residuals is well reproduced in the bootstrapped innovations. Postblackening these innovations, in turn, synthesizes replicates that exhibit nearly the same behavior as MBB with regard to the preservation of skewness of monthly flows (Figure 1b). The hybrid model is found to inherit the characteristic of capturing the salient features of the marginal distribution (asymmetry, peakedness, and multimodality) of observed flows from its nonparametric constituent (MBB) and is able to provide some smoothing and limited extrapolation, owing to its parametric constituent. More details on these results are available from the authors and can be obtained on request.

As expected, the first-order serial correlations are well reproduced by PAR(1)-NT (the simple parametric model used as a component in HM), but not the significant higher-lag serial correlations (Figure 2). In contrast, MBB (the nonparametric component in HM) is able to preserve all the within-year serial correlations well. This is because MBB resamples blocks of observed streamflows, with block size taken in multiples of the number of periods in a water year (12 for monthly streamflow modeling) and that each block begins with the first month of a water year.

Modeling monthly serial correlations across water years (abbreviated SCAWY) is important for the efficient simulation of the critical water use (validation) statistics (especially when such correlations are significant). It is noted from Figure 2 that MBB does not preserve SCAWY. This is because year-to-year dependence gets destroyed at the boundaries between the adjoining blocks of streamflows. If this performance is to be improved, a much longer block size must be chosen for resampling the flow data. However, this reduces the variety in the simulations, which is undesirable. On the other hand, it is observed from Figure 2 that the hybrid model is able to preserve the serial correlations within the water year as well as those between adjoining water years satisfactorily, owing to the hybrid effect. With regard to preservation of lower-lag SCAWY (lag-1 serial correlation of October), it may be noted that HM is gaining from its parametric constituent almost entirely, with only a very minor supplementation by MBB, while in the case of higher-lag SCAWY (for instance, lag-4 serial correlations of flows of October–January months), a considerable portion of the dependence is extracted by the PAR(1)-NT model itself during the prewhitening process (though not entirely, since it is only a first-order model), and the weak dependence remaining in the residuals from the PAR(1)-NT model is captured reasonably well by the innovations during the resampling process using MBB. Eventually, when the postblackening of the innovations is performed, the amount of dependence present in the observed streamflows is well preserved by the resulting synthetic replicates. One more interesting point to note in Figure 2 is that whenever the parametric model (at the prewhitening stage) provides a high amount of overfitting or underfitting of any of the serial correlations referred, the same effect is transmitted to HM also, but to a lesser degree. For example, see lag-3 serial correlation of October month flows for the overfitting effect (Figures 2g and 2i) and lag-4 serial correlation of January month flows for the underfitting effect (Figures 2j and 2l). This moderation is provided by the resampling of residuals using MBB.

The hybrid effect with regard to preservation of nonlinear

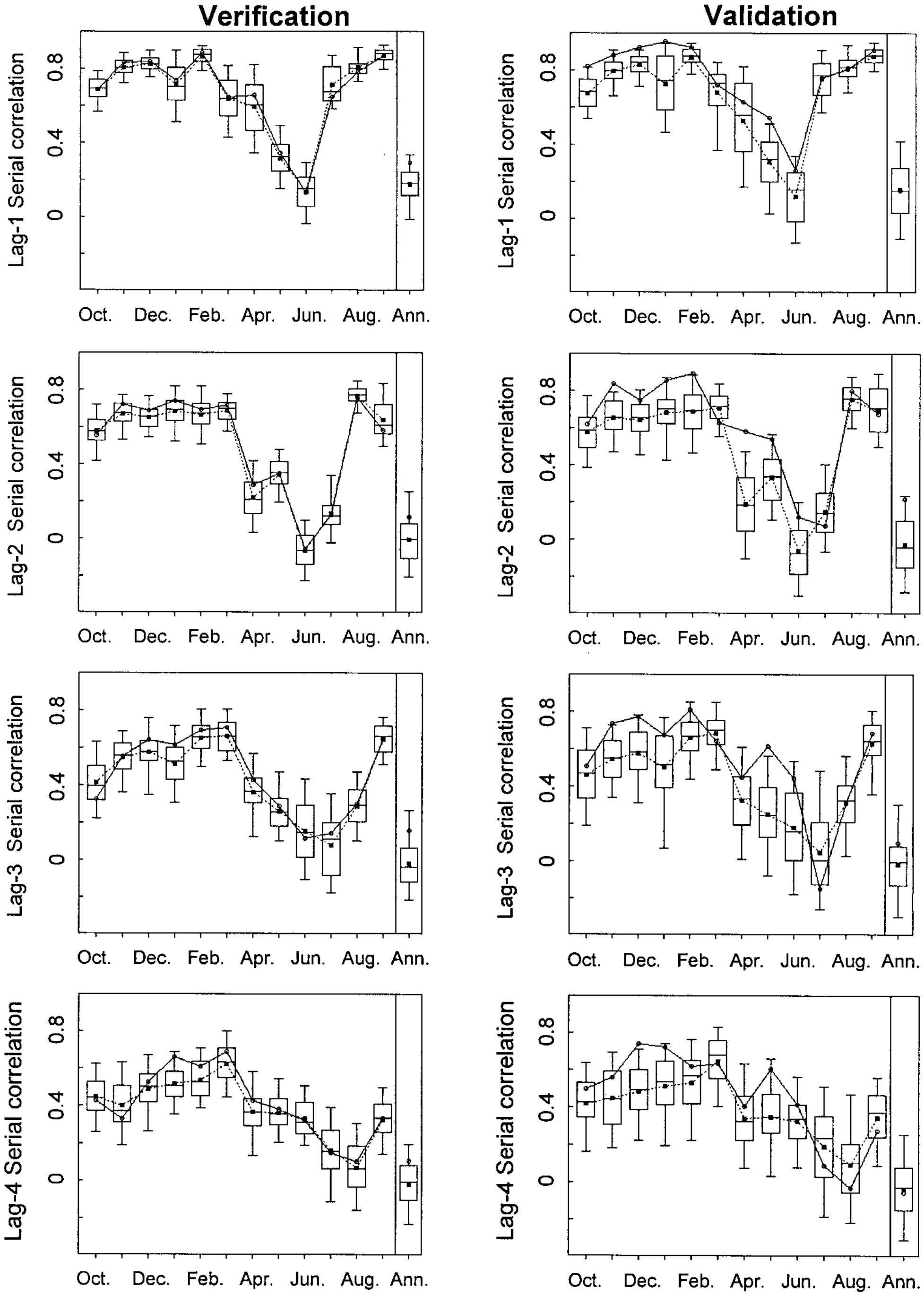

Figure 7. Results from split sample validation of hybrid model. Shown is the preservation of serial correlations of monthly streamflows and autocorrelation of annual streamflows, Weber River.

dependence (defined in terms of state-dependent correlations) [*Sharma et al.*, 1997] is presented in Figure 3. A significant difference between the above- and below-median pairs of historical correlations in either the forward or backward direction indicates the presence of nonlinearity. Simulations from PAR(1)-NT are not able to preserve state-dependent correlations for months with significant nonlinearity (Figures 3a, 3d, 3g, and 3j). In contrast, MBB is good at reproducing the state-dependent correlations except for the ones between months of adjoining water years (Figures 3b, 3e, 3h, and 3k). Here again, the reason for MBB not being able to preserve the state-dependent correlations between months of adjoining water years is the loss of dependence due to the discontinuities between the moving blocks. As discussed earlier, HM is able to overcome the aforementioned shortcoming of MBB owing to the hybrid effect (Figures 3c, 3f, 3i, and 3l).

4. Split Sample Test

In this section we intend to investigate whether the hybrid model presented herein is able to capture repeatable statistical structure present in the observed streamflows, through a split

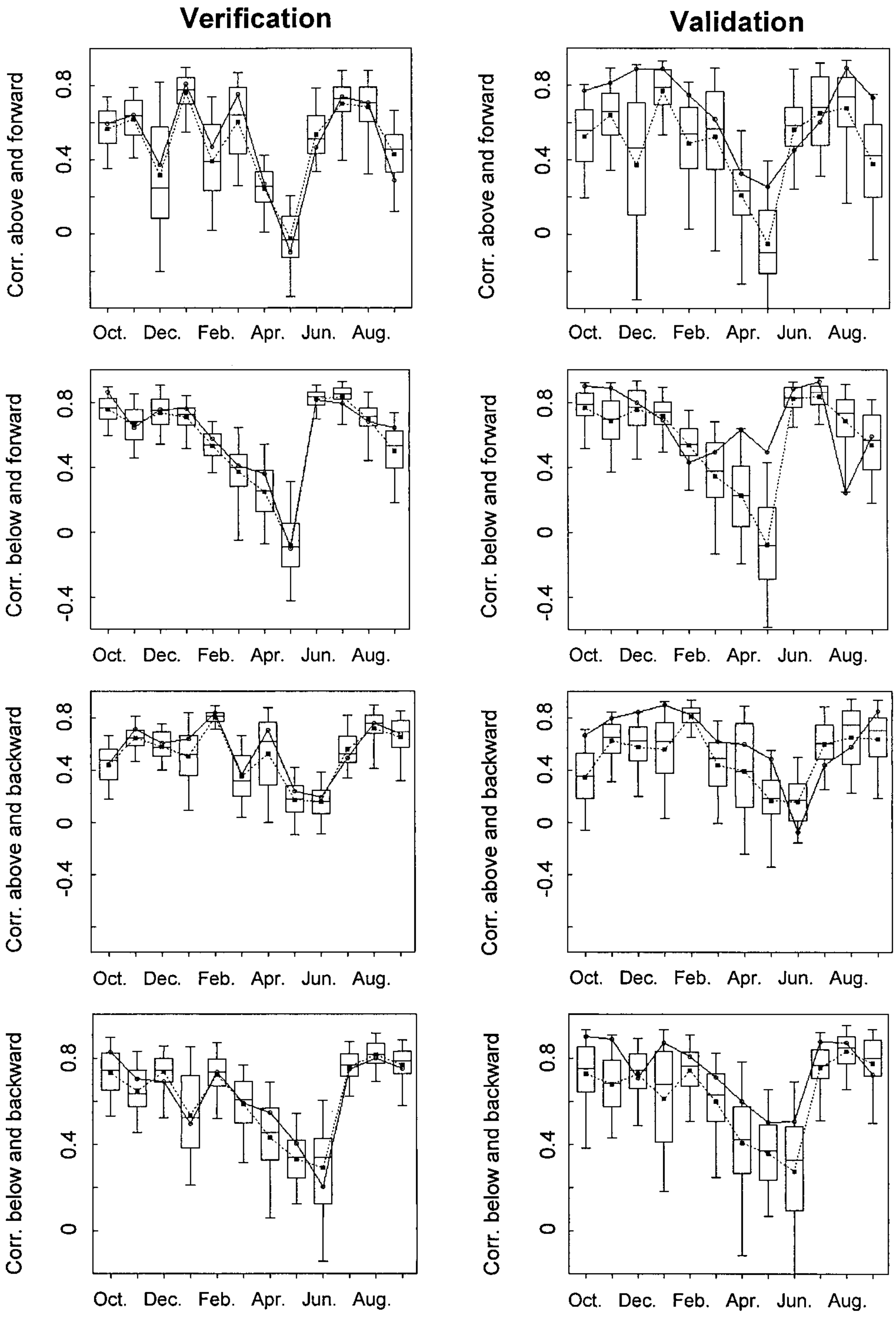

Figure 8. Results from split sample validation of hybrid model. Shown is the preservation of state-dependent correlations of historical monthly streamflows, Weber River.

sample validation test, performed on monthly streamflows of Weber River (Utah). The split sample test was carried out in two phases, namely, calibration and validation.

4.1. Calibration Phase

The hybrid model was calibrated with the first 55-year record (1905–1960) of Weber River's historical streamflows (equal to about two thirds of the historical record length of 83 years). The calibration process involves estimation of (1) periodic means; (2) periodic standard deviations; (3) periodic autoregressive parameters of the linear parametric model (PAR(1)-NT) used for partial prewhitening of the historical streamflows; and (4) the block size to be adopted for resampling the resulting residuals using MBB. One hundred synthetic replicates, each of size 55 years, are generated, as per the procedure described in section 2. The block size of residuals for which the historical statistics of interest are well reproduced in the synthetic simulations is selected. In the case of the 55-year Weber

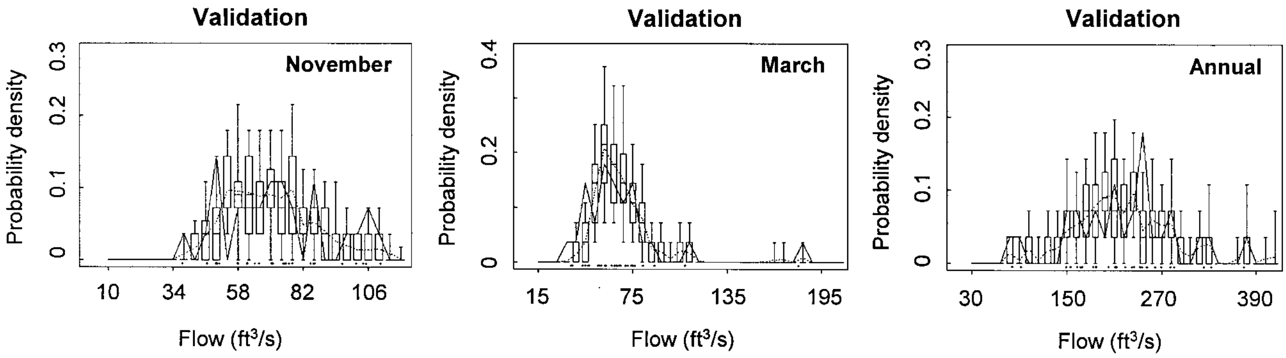

Figure 9. Results from split sample validation of hybrid model. Shown is the preservation of marginal distributions of monthly streamflows of November and March and annual streamflows, Weber River.

River flows, HM is seen to satisfactorily reproduce summary statistics, monthly serial correlations, autocorrelations at annual level, and state-dependent correlations, when a block size of 2ω is used for resampling the residuals (see verification in Figures 4–8).

4.2. Validation Phase

The residuals extracted from the calibration phase are bootstrapped in blocks of size 2ω (block size selected in calibration phase) to obtain innovations. These innovations are then post-blackened using the periodic autoregressive parameters (estimated in calibration phase), followed by inverse standardization using periodic means and periodic standard deviations (estimated in the calibration phase) to obtain 100 synthetic replicates each 28 years long. These synthetic replicates are tested for their ability to reproduce a wide variety of statistics of the remaining 28-year (1960–1988) observed streamflow record of Weber River (not used in calibration phase). Figures 4–9 show that simulations from the validation phase reproduce various statistics of interest fairly well. One may note considerable deflation in the preservation of the monthly standard deviation in the low-flow months (validation in Figure 5). This

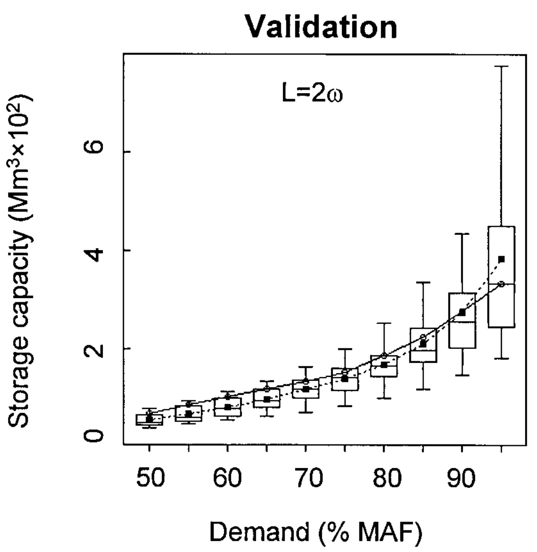

Figure 10. Results from split sample validation of hybrid model. Shown is the preservation of reservoir storage capacity of historical monthly streamflows for the test data set, Weber River (1×10^6 m^3 = 0.8112×10^3 acre-feet).

is possibly due to some amount of nonstationarity inherent in the historical flow data (as can be observed from the differences in the historical monthly standard deviation values for the calibration and the test data sets (Figure 5)) that the stationary hybrid model cannot capture. Further, it may be noted from Figure 10 that the reservoir storage statistic (known as storage validation statistic according to *Stedinger and Taylor* [1982]) is well reproduced for the 28-year test data set. The reservoir storage capacities required to cater to yields of 50% mean annual flow (MAF) to 95% MAF (at 5% MAF intervals) are computed using the sequent peak algorithm [*Loucks et al.*, 1981, p. 235] assuming the demand to be fixed and uniform over the 12 months of the water year.

5. Performance Comparison With SPIGOT and NPD

In this section we compare the performance of the hybrid model with the popular SPIGOT [*Grygier and Stedinger*, 1990] and nonparametric disaggregation (NPD) [*Tarboton et al.*, 1998] models in simulating the 80-year (1906–1985) observed monthly streamflow record of the San Juan River (station number AF3555 from U.S. Bureau of Reclamation Colorado River simulation system). This station is located near Archuleta, New Mexico, at 36°48′05″N latitude and 107°41′51″W longitude and at an elevation of 1724 m (5655 feet) above mean sea level. This streamflow data set has been chosen because it contains appreciable nonlinear dependence and the record length available is reasonable. Moreover, it has been recently used for temporal disaggregation modeling by *Tarboton et al.* [1998], and the results reported therein enable the performance comparison. The performance comparison is presented in terms of preservation of (1) summary statistics at monthly and annual levels; (2) linear dependence structure (expressed in terms of serial correlations and autocorrelations); (3) nonlinear dependence (expressed in terms of state-dependent correlations [*Sharma et al.*, 1997]; and (4) reservoir storage statistic.

Historical mean monthly streamflows and the historical mean annual streamflows are well reproduced by the hybrid model (not presented herein for brevity). In reference to the preservation of standard deviation, the SPIGOT model shows some deflation in the October, August, and September months, whereas the NPD model is seen to inflate standard deviations at both monthly and annual levels. In contrast, HM is able to reproduce standard deviations at both periodic and annual levels (Figure 11). Likewise, one may note from Figure 12 that HM is good at reproducing the skewness of observed

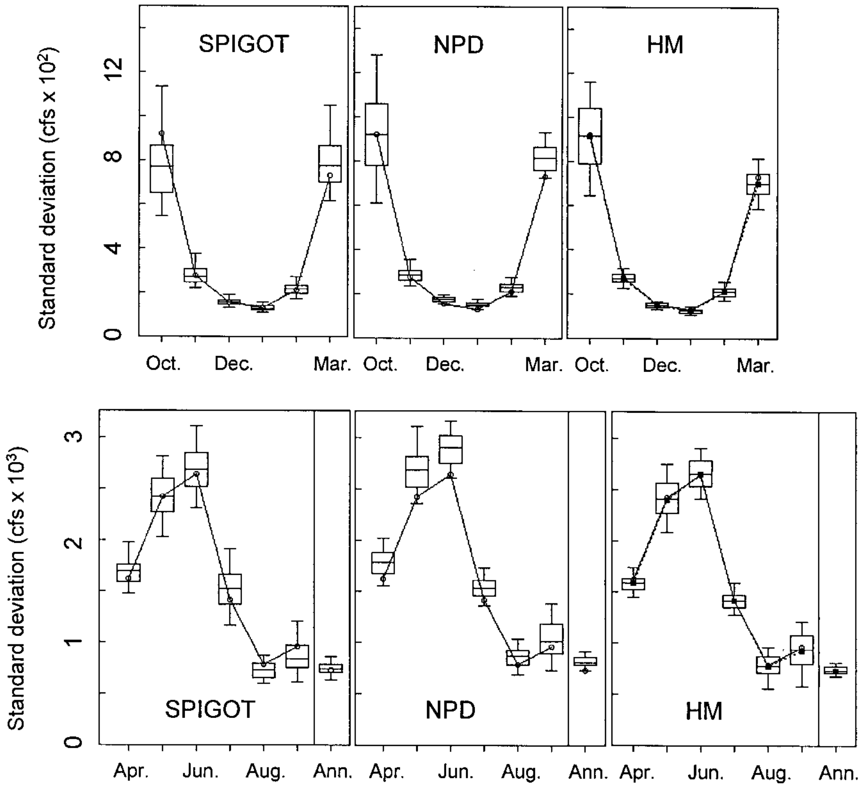

Figure 11. Preservation of standard deviation (SD) of San Juan streamflows at monthly and aggregated annual levels. Shown is a comparison between (a) SPIGOT, (b) nonparametric disaggregation (NPD) model, and (c) hybrid model (HM).

flows at both monthly and annual levels compared with the SPIGOT and NPD models.

From the results presented by *Tarboton et al.* [1998, Figure 10], it may be noted that although the NPD model exhibits a better performance compared with SPIGOT, in terms of reproducing month-to-month cross correlations, bias is seen for some of the higher-lag within-year cross correlations (such as 2-11, 2-12, 3-11, and 3-12). Furthermore, the NPD model is seen to inflate a few of the month-to-annual cross correlations. In contrast, it can be seen from Figure 13 that HM is better at reproducing the month-to-month and month-to-annual cross correlations for the 80-year streamflow record of San Juan River.

In reference to state-dependent correlations, it is seen from *Tarboton et al.* [1998, Figure 14] that the above-median and forward correlations are poorly preserved by the SPIGOT model, while the NPD model shows a reasonable preservation. Furthermore, both the SPIGOT and NPD models are not able to capture the historical trend of the below-median and forward correlations, as can be noted from Figure 12 of *Tarboton et al.* [1998, p. 116]. In contrast, HM is able to exhibit a reasonable preservation of all the four state-dependent correlations (Figure 14). Moreover, it may be seen from Figure 15 that the hybrid model is able to preserve serial correlations across water years (SCAWY) owing to the hybrid effect.

To enable comparison of the performance of HM with SPIGOT and NPD models in predicting reservoir storage statistics of the San Juan River, results are presented in terms of relative bias (R-bias) (equation (6)) and relative root-mean-

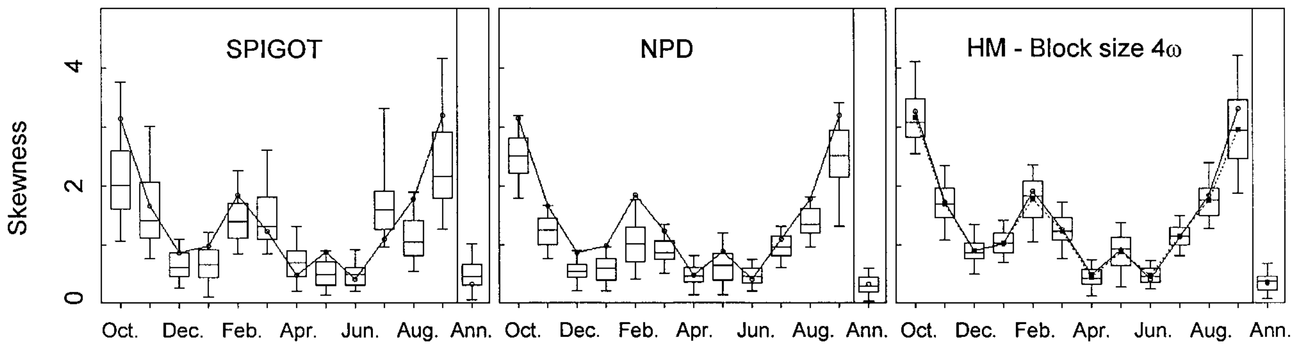

Figure 12. Preservation of skewness of San Juan streamflows at monthly and aggregated annual levels. Shown is a comparison between (a) SPIGOT, (b) nonparametric disaggregation (NPD) model, and (c) hybrid model (HM).

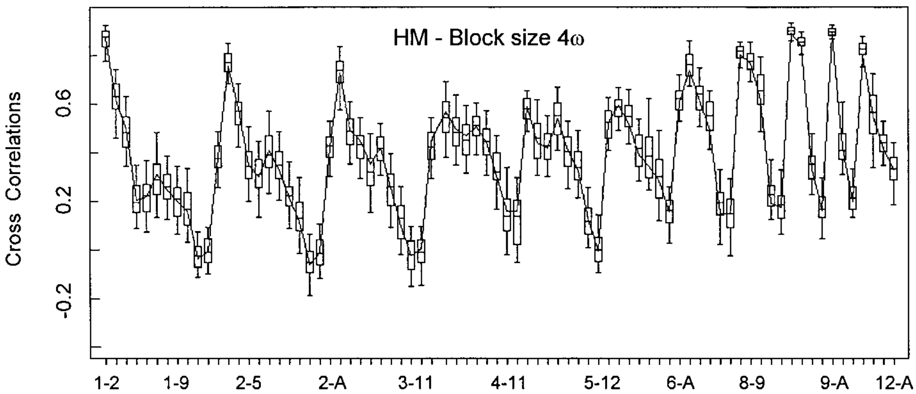

Figure 13. Simulated and observed cross-correlation pairs using HM. The sequence along the x axis is 1-2, 1-3, . . . , 1-12, 1-A, 2-3, 2-4, . . . , 2-12, 2-A, 3-4, and so on. Here (1,2) indicates cross correlation between months 1 and 2, (1,A) indicates cross correlation between month 1 and annual aggregate. Months are numbered according to the water year (1, October; 2, November; 4, January; and so on), San Juan River.

square error (R-RMSE) (equation (7)) computed over 100 synthetic replicates.

$$\text{R-bias} = \left[K_{\text{hist}} - \frac{1}{N} \sum_{i=1}^{N} K_i \right] \Big/ K_{\text{hist}} \tag{6}$$

$$\text{R-RMSE} = \left[\frac{1}{N} \sum_{i=1}^{N} (K_{\text{hist}} - K_i)^2 \right]^{1/2} \Big/ K_{\text{hist}}, \tag{7}$$

where K_{hist} denotes the storage capacity estimated from observed (historical) flows; K_i is the storage capacity estimated from the ith synthetic replicate, and N denotes the number of synthetic replicates. It is to be noted that for the SPIGOT and NPD models, the results extracted from *Tarboton et al.* [1998, p. 116] (wherein R-bias and R-RMSE are reported for only 50% and 90% MAF demand levels) are presented in Table 1 alongside the results of the hybrid model, for the sake of comparison. However, for the hybrid model, the preservation of reservoir storage statistics is presented in Table 2 for the intermediate demand levels (from 55% to 85% MAF at 5% MAF intervals).

The simulations from the SPIGOT and NPD models overestimate the storage capacity at a low demand level of 50% MAF and underestimate the same at the higher demand level of 90% MAF. It may be noted from Table 1 that at 90% demand level, the SPIGOT model, in addition to highly underestimating the storage capacity, is not able to show sufficient variation, indicating a poor preservation of the statistic.

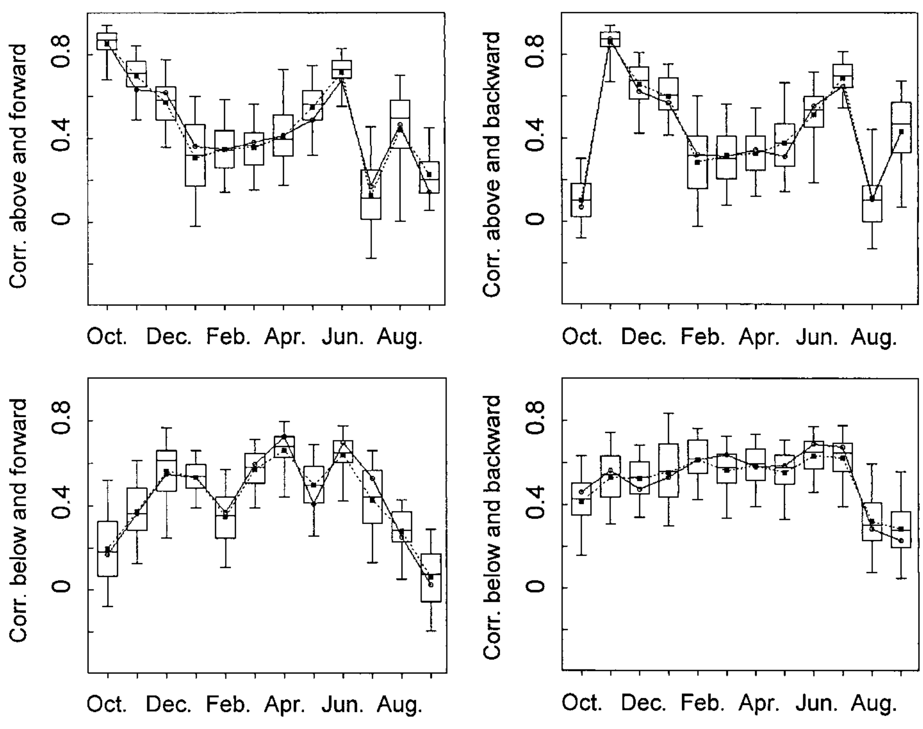

Figure 14. Preservation of state-dependent correlations of monthly streamflows of San Juan River by the hybrid model (HM) for a block size of 4ω.

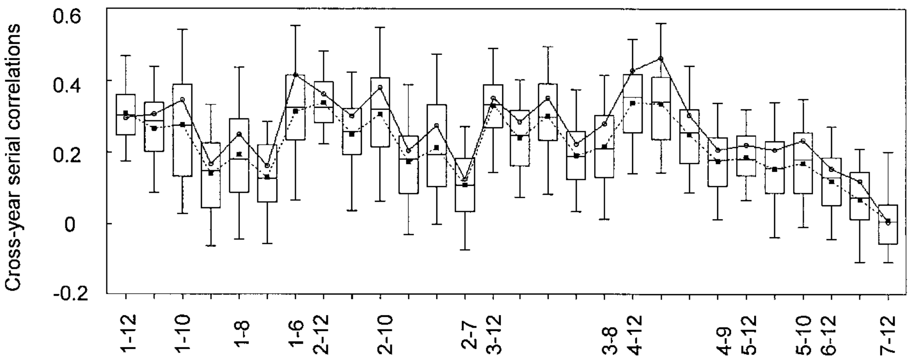

Figure 15. Preservation of cross-year serial correlations of monthly streamflows of the San Juan River by the hybrid model for a block size of 4ω. Notation 1–12 on the X axis indicates serial correlation between month 1 of current hydrological water year and month 12 of previous water year. Months are numbered according to the water year (1, October; 2, November; 4, January; and so on).

This may be because the SPIGOT model is not designed to preserve the higher-lag serial correlations. Moreover, it is not able to preserve the skewness that well. Furthermore, the preservation of the marginal distribution also seems wanting (see Figure 11 of *Tarboton et al.* [1998, p. 115]). On the other hand, the hybrid model (with $L = 4\omega$) is able to provide a reasonable prediction of the storage capacity at the lower demand level of 50% MAF. However, even the hybrid model deflates the storage capacity at higher demand level (90% MAF). This is due primarily to the inability to reproduce some of the higher-lag annual autocorrelations that are significant in the case of San Juan River flows. In addition, it may be observed from Table 2 that HM is able to predict the reservoir storage capacity from 55% to 85% MAF fairly well. This may be attributed to the better preservation of skewness, marginal distribution, and the seasonal dependence structure including SCAWY.

6. Summary and Conclusions

A new hybrid stochastic model that effectively blends the merits of the parsimonious parametric model (PAR(1)NT) and simple moving block bootstrap (nonparametric) model has been presented for simulating multiseason streamflows. The first part of the paper demonstrates the hybrid character of the model through stochastic simulations performed using monthly streamflows of Weber River (Utah) that exhibit a complex dependence structure. Following this, a split sample validation is performed on the Weber River monthly streamflows to show that the hybrid model is able to capture repeatable statistical structure present in the observed streamflows. The latter part of the paper presents a performance comparison between SPIGOT [*Grygier and Stedinger*, 1990], NPD [*Tarboton et al.*, 1998], and HM in simulating historical monthly streamflows of San Juan River (New Mexico).

This hybrid model is shown to offer better simulations than its own constituents, by acquiring certain properties that are characteristic of either of these models. The efficiency of HM with regard to preservation of skewness and salient features of the marginal distributions is attributed primarily to the nonparametric component MBB, while the parametric component aids in achieving some smoothing. The preservation of multiple-lag cross-year serial correlations is due to the hybrid effect. The hybrid model ensures annual-to-monthly consistency, thus averting the adjustments to monthly or annual flows and the associated problems that surface in the case of linear parametric disaggregation models.

For the appropriate block size chosen, the hybrid model is seen to perform reasonably well in predicting the reservoir storage (validation) statistic. Compared with SPIGOT (parametric) and NPD (nonparametric) models, HM is seen to be better at reproducing a wide variety of statistics for the San Juan River.

Although the hybrid model presented here uses a simple PAR(1) model for partial prewhitening and MBB for bootstrapping the residuals, one can try other hybrid variants too. The extension of this hybrid model to multisite, multiseason hydrologic modeling requires devising the residual resampling strategy in such a way as to maintain the contemporal relationships between the residuals of different sites considered. Research in this direction is under way. Further theoretical and computational efforts should focus on exploring methods that can combine the advantages of the parsimonious parametric models with the wealth of nonparametric methods to effect better streamflow synthesis, which is important in operational hydrology.

Table 1. Comparison of Predictions of Reservoir Storage Capacity Statistic for the 80-Year Monthly Streamflow Record of San Juan River

Demand Level, % Mean Annual Flow	SPIGOT		NPD		HM	
	R-Bias	R-RMSE	R-Bias	R-RMSE	R-Bias	R-RMSE
50	−0.192	0.457	−0.387	0.520	0.097	0.216
90	0.412	0.457	0.284	0.395	0.337	0.453

Table 2. Prediction of Reservoir Storage Capacity Statistic by the Hybrid Model for the 80-Year Monthly Streamflow Record of San Juan River

	Demand Level, % Mean Annual Flow						
	55	60	65	70	75	80	85
R-Bias	0.054	−0.047	−0.017	0.007	−0.006	0.027	0.207
R-RMSE	0.194	0.220	0.234	0.244	0.272	0.293	0.337

Acknowledgments. The authors thank the Indian Institute of Technology, Madras, Chennai, India, for providing the necessary facilities to carry out this research work. We are grateful to A. C. Davison, Department of Mathematics, Swiss Federal Institute of Technology, for useful discussions. We express our gratitude to D. G. Tarboton, Utah State University, for providing us the San Juan River data and for his valuable comments and suggestions on the earlier drafts of this manuscript, which were useful in improving the quality of the paper. We are thankful for comments made by David Goldman of USACE and two more anonymous reviewers on the earlier versions of this manuscript, which have helped in improving the quality of presentation and the organization of the paper. We duly acknowledge Upmanu Lall, Utah State University, for all the discussions and the encouragement.

References

Box, G. E. P., and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day, Boca Raton, Fla., 1970.

Davison, A. C., and D. V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge Univ. Press, New York, 1997.

Efron, B., Bootstrap methods: Another look at the jackknife, *Ann. Stat*., *7*, 1–26, 1979.

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

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

Grygier, J. C., and J. R. Stedinger, SPIGOT: A synthetic streamflow generation package, technical description, version 2.5, School of Civ. and Environ. Eng., Cornell Univ., Ithaca, N. Y., 1990.

Harms, A. A., and T. H. Campbell, An extension to the Thomas-Fiering model for the sequential generation of streamflow, *Water Resour. Res*., *3*(3), 653–661, 1967.

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

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

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

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

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

Lane, W. L., *Applied Stochastic Techniques*, *User's Manual*, Bur. of Reclam., Eng. and Res. Cent., Denver, Colo., 1979.

LePage, R., and L. Billard, *Exploring the Limits of Bootstrap*, John Wiley, New York, 1992.

Loucks, D. P., J. R. Stedinger, and D. A. Haith, *Water Resources Systems Planning and Analysis*, 559 pp., Prentice-Hall, Englewood Cliffs, N. J., 1981.

Mejia, J. M., and J. Rousselle, Disaggregation models in hydrology revisited, *Water Resour. Res*., *12*(2), 185–186, 1976.

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

Salas, J. D., Analysis and modeling of hydrologic time series, in *Handbook of Hydrology*, edited by D. R. Maidment, pp. 19.1–19.72 McGraw-Hill, New York, 1993.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane, *Applied Modelling of Hydrologic Time Series*, Water Resour. Publ., Littleton, Colo., 1980.

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

Santos, E. G., and J. D. Salas, Stepwise disaggregation scheme for synthetic hydrology, *J. Hydraul. Eng*., *118*(5), 765–784, 1992.

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

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

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1986.

Srinivas, V. V., and K. Srinivasan, Post-blackening approach for modeling dependent annual streamflows, *J. Hydrol*., *230*(1–2), 86–126, 2000.

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

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

Valencia, D. R., and J. C. Schaake Jr., Disaggregation processes in stochastic hydrology, *Water Resour. Res*., *9*(3), 580–585, 1973.

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

V. V. Srinivas and K. Srinivasan, Environmental and Water Resources Engineering Division, Department of Civil Engineering, Indian Institute of Technology, Madras, Chennai 600 036, India. (srini@civil.iitm.ernet.in)

(Received May 17, 2000; revised November 7, 2000;
accepted November 22, 2000.)