

## A stochastic nonparametric technique for space-time disaggregation of streamflows

James Prairie,<sup>1</sup> Balaji Rajagopalan,<sup>2,3</sup> Upmanu Lall,<sup>4</sup> and Terrance Fulp<sup>5</sup>

Received 9 November 2005; revised 11 August 2006; accepted 25 October 2006; published 22 March 2007.

[1] Stochastic disaggregation models are used to simulate streamflows at multiple sites preserving their temporal and spatial dependencies. Traditional approaches to this problem involve transforming the streamflow data of each month and at every location to a Gaussian structure and subsequently fitting a linear model in the transformed space. The simulations are then back transformed to the original space. The main drawbacks of this approach are (1) transforming marginals to Gaussian need not lead to the correct multivariate distribution particularly if the dependence across variables is nonlinear, and (2) the number of parameters to be estimated for a traditional disaggregation model grows rapidly with an increase in space or time components. We present a K-nearest-neighbor approach to resample monthly flows conditioned on an annual value in a temporal disaggregation or multiple upstream locations conditioned on a downstream location for a spatial disaggregation. The method is parsimonious, as the only parameter to estimate is K (the number of nearest neighbors to be used in resampling). Simulating space-time flow scenarios conditioned upon large-scale climate information (e.g., El Niño–Southern Oscillation, etc.) can be easily achieved. We demonstrate the utility of this methodology by applying it for space-time disaggregation of streamflows in the Upper Colorado River basin. The method appropriately captures the distributional and spatial dependency properties at all the locations.

**Citation:** Prairie, J., B. Rajagopalan, U. Lall, and T. Fulp (2007), A stochastic nonparametric technique for space-time disaggregation of streamflows, *Water Resour. Res.*, 43, W03432, doi:10.1029/2005WR004721.

### 1. Introduction

[2] Synthetic simulation of streamflow sequences is used in a variety of applications including reservoir operation and for evaluating water supply reliability. Multiple reservoirs and stream sections are often considered in a system's operation plan. For this purpose, streamflows generated at different sites need to be consistent. This implies that the flow at a downstream gauge is the sum of tributary flows; the annual flow is the sum of monthly flows; the monthly fraction of flows in wet/dry years are representative; and the dependencies of flows between the sites have to be reproduced. To this end, the disaggregation problem can be thought of as simulation from the conditional probability density function (PDF)  $f(X|Z)$ , where  $X$  is a vector of disaggregated (e.g., monthly flows) flows and  $Z$  is the aggregate (e.g., annual) flows and other terms (e.g., the first month's correlation with the last month of the previous

year), subject to the condition that the disaggregated flows add up to the aggregate flows, which is the additivity property. Often a simpler approach has been used consisting of fitting a model of the form

$$X = AZ + B\varepsilon, \quad (1)$$

where  $Z$  is usually taken to be just the annual flow and  $A$  and  $B$  are matrices of the model parameters that are estimated to ensure the additivity property and  $\varepsilon$  is the stochastic term. Notice that the above form is that of a linear regression, which has a rich developmental history; consequently, the main assumption is that the stochastic term and hence the data ( $X$  and  $Z$ ) are assumed to be normally distributed. To achieve this, the data are typically transformed to a normal distribution by appropriate transforms before the model is fit. The simulation proceeds as follows: (1) An aggregate streamflow is generated from an appropriate linear or nonlinear model or equivalent data set. (2) The simulated aggregate flow is then disaggregated using the above model. The simulated flows are back transformed to the original space. This linear stochastic framework for streamflow disaggregation was first developed by *Valencia and Schaake* [1973] and subsequently modified and improved by several others [*Mejia and Rousselle*, 1976; *Lane*, 1979; *Salas et al.*, 1980; *Stedinger and Vogel*, 1984; *Stedinger et al.*, 1985; *Salas*, 1985; *Santos and Salas*, 1992].

[3] Since these models are fit in the transformed space, the additivity of the disaggregated flows to the aggregate

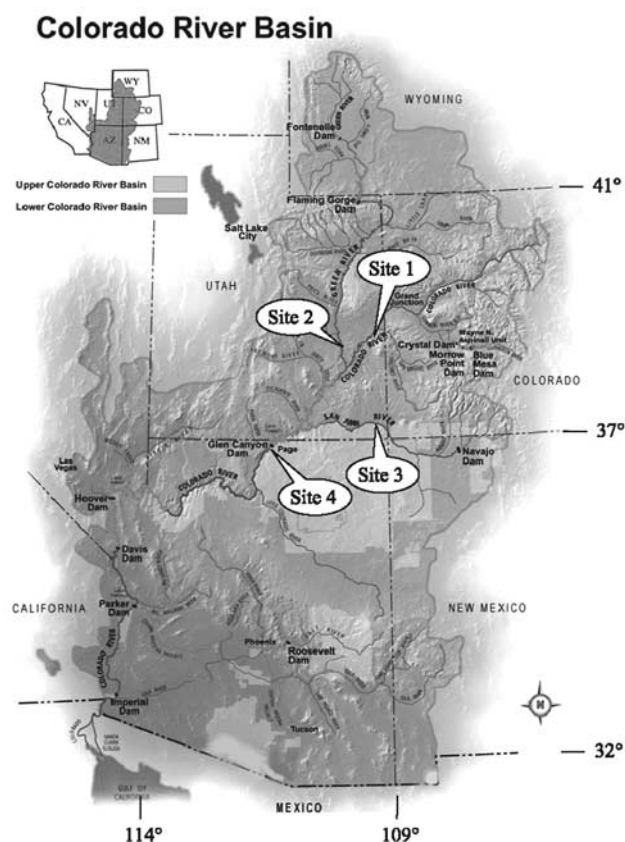
<sup>1</sup>Bureau of Reclamation, University of Colorado, Boulder, Colorado, USA.

<sup>2</sup>Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, Colorado, USA.

<sup>3</sup>Also at Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.

<sup>4</sup>Department of Earth and Environmental Engineering, Columbia University, New York, New York, USA.

<sup>5</sup>Bureau of Reclamation, Lower Colorado Region, Boulder City, Nevada, USA.



**Figure 1.** Streamflow locations within the Upper Colorado River basin.

flows in the original space after back transformation is not guaranteed. Hence several adjustments have to be made [e.g., Lane, 1982; Stedinger and Vogel, 1984; Grygier and Stedinger, 1988]. Furthermore, the model is designed to reproduce the statistics in the transformed space but reproduction is not guaranteed in the original space.

[4] Alternate approaches to disaggregation [Tao and Delleur, 1976; Todini, 1980; Koutsoyiannis, 1992; Koutsoyiannis and Manetas, 1996; Koutsoyiannis, 2001] allow representation of non-Gaussian data directly in the disaggregation scheme to avoid the need for data transformation. These techniques can incorporate the skewness from the historic data into the stochastic term [Tao and Delleur, 1976; Todini, 1980; Koutsoyiannis, 1999]. Koutsoyiannis [2001] provides a stepwise disaggregation scheme that incorporates an adjustment procedure that preserves the additivity property and certain higher-order statistics. These methods are iterative in nature and thus computationally intensive besides requiring assumptions of linearity.

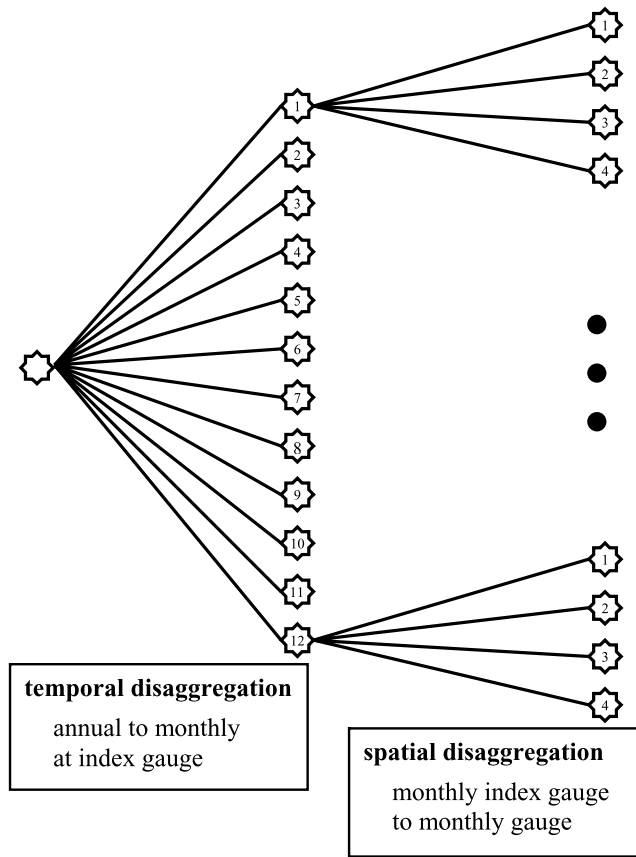
[5] Recent advances in nonparametric methods (see Lall [1995] for an overview of nonparametric methods and their applications to hydroclimatic data) provide an attractive alternative to linear parametric methods. Unlike the linear approach where a single linear model is fit to the entire data, the nonparametric methods involve “local” functional fitting. The function is fit to a small number of neighbors at each point. This approach has the ability to capture any arbitrary features (nonlinearities, non-normal, etc.) exhibited

by the data. Nonparametric methods have been applied to a variety of hydroclimate modeling questions including stochastic daily weather generation [Rajagopalan and Lall, 1999; Yates et al., 2003], streamflow simulation [Lall and Sharma, 1996; Sharma et al., 1997; Prairie et al., 2006], streamflow forecasting [Grantz et al., 2005; Singhrattna et al., 2005], and flood frequency estimation [Moon and Lall, 1994] to mention a few.

[6] Kernel estimator based nonparametric streamflow simulation at a single site was developed by Sharma et al. [1997] where they also demonstrate its advantage over traditional linear models. Sharma and O’Neil [2002] improved on this to capture the interannual dependence. However, kernel methods can be inefficient in higher dimensions (e.g., space-time disaggregation), as noted by Sharma and O’Neil [2002] and as such, difficult to implement in multivariate problems such as space-time disaggregation in a network. Lall and Sharma [1996] developed a K-nearest-neighbor (K-NN) bootstrap approach to time series modeling and applied it to streamflow simulation. Being a bootstrap method, values not observed in the historic data will not be generated in the simulations. To address this, a modified version of the K-NN bootstrap was developed by Prairie et al. [2005, 2006], and this was further used in streamflow forecasting [Grantz et al., 2005; Singhrattna et al., 2005]. Semiparametric approaches and bootstrap methods for streamflow simulation have also been developed [Souza Filho and Lall, 2003; Srinivas and Srinivasan, 2001].

[7] Tarboton et al. [1998] developed a kernel-based approach (an extension of their single site methodology by Sharma et al. [1997]) for temporal (i.e., annual to monthly) streamflow disaggregation. Kumar et al. [2000] adopted K-NN bootstrap techniques in conjunction with an optimization scheme for spatial and temporal disaggregation of monthly streamflows to daily flows. They indicate that disaggregating monthly flow to daily involves a higher-dimensional problem that cannot always be well represented by traditional parametric disaggregation techniques. Additionally, daily flows typically display nonlinear flow dynamics that are not adequately modeled with traditional techniques. The optimization framework allows for increased flexibility in specifying the functional relationships the disaggregation scheme needs to preserve but at a great computational cost. Srinivas and Srinivasan [2005] developed a semiparametric disaggregation method for a multi-site model they termed as hybrid moving block bootstrap multisite model (HMM). In this approach a parametric model (such as a linear autoregressive model) is fit to the data and the residuals from this model are resampled by block bootstrapping (the nonparametric component). This method is able to incorporate the strengths of both parametric and nonparametric models but still requires multiple steps.

[8] In practical terms, there is a need for a robust, simple, and parsimonious approach for space-time streamflow disaggregation that can capture the features exhibited by the data. To this end, here we develop a K-NN based disaggregation framework. The proposed framework and the algorithm are first described, followed by its application to four streamflow sites on the Upper Colorado River basin, con-



**Figure 2.** Schematic of space-time disaggregation.

cluding with a summary and discussion of applications and the future direction for this research.

## 2. K-Nearest-Neighbor Based Disaggregation Framework

[9] The framework follows the work of *Tarboton et al.* [1998] except that the kernel-based density estimation is replaced with a K-nearest-neighbor approach. We describe the framework and the implementation algorithm for the case of a temporal (annual to monthly) disaggregation, and the same steps follow for spatial disaggregation. As an example, consider  $X$  to be a  $d = 12$  dimensional monthly flow vector where  $Z$  is the annual flow. As mentioned earlier, the disaggregation problem amounts to simulation from the conditional PDF  $f(X|Z)$  with the constraint that the disaggregated flows sum up to the aggregate flow. The conditional PDF can be written as

$$f(X|Z) = f(X, Z) / \int f(X, Z) dX. \quad (2)$$

The numerator in the above equation requires the estimation of a  $d + 1$  dimensional joint density function  $f(X, Z)$ . However, because of the additivity requirement, all the mass of this joint PDF is situated on the  $d$ -dimensional hyperplane defined by

$$X_1 + X_2 + \dots + X_d = Z. \quad (3)$$

Thus, for a particular value of  $Z$  (the aggregate annual flow) the conditional PDF can be visualized geometrically as the probability density on a  $d - 1$  dimensional hyperplane slice through the  $d$ -dimensional density  $f(X)$ . The conditional PDF can be specified through a rotation of the vector  $X$  into a new vector  $Y$  whose last coordinate is aligned perpendicular to the hyperplane defined by (3). *Tarboton et al.* [1998] describe this in detail and illustrate this point very well in their Figure 1. The conditional PDF is constructed in the rotated space ( $f(Y|Z)$ ), and the simulation is also done in this rotated space before back rotation. In the *Tarboton et al.* [1998] framework, kernel density estimators are used to construct this conditional PDF and subsequently for simulation. As mentioned earlier, the kernel methods are known to be inefficient and cumbersome to implement in higher dimensions. This limits their ability to extend the approach to space and time disaggregation.

[10] We depart from the *Tarboton et al.* [1998] framework here and instead develop a K-NN based bootstrap approach to construct and simulate from the conditional PDF ( $f(Y|Z)$ ). The methodology is described in the algorithm below.

### 2.1. The Algorithm

[11] The steps involved in the algorithm are as follows:

[12] 1. The historic data of monthly flows are oriented in  $X$  such that seasons are across rows and years are across columns.  $X$  is rotated into  $Y$  by the rotation matrix  $R$  where,

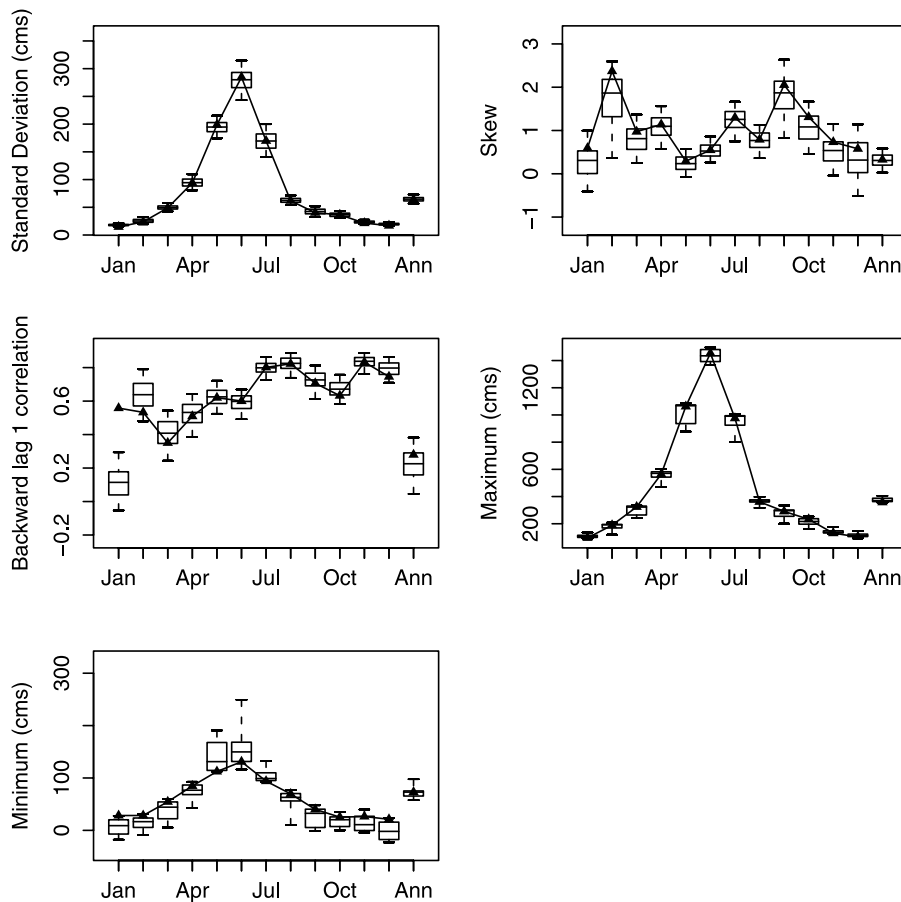
$$Y = RX. \quad (4)$$

The procedure for obtaining the rotation matrix is described in detail in the appendix of *Tarboton et al.* [1998], here we summarize from their description. The rotation matrix is developed from a standard basis (basis vectors aligned with the coordinate axes) which is orthonormal but does not have a basis vector perpendicular to the conditioning plane defined by (3). One of the standard basis vectors is replaced by a vector perpendicular to the conditioning plane. Operationally, this entails starting with an identity matrix and replacing the last row with  $1/\sqrt{d}$ . The basis set is then no longer orthonormal. The Gram Schmidt orthonormalization procedure is applied to the remaining  $d - 1$  standard basis vectors to obtain an orthonormal basis that now includes a vector perpendicular to the conditioning plane. The resulting  $R$  matrix is orthonormal and has the property  $R^T = R^{-1}$ . Further, note that  $R$  is only a function of the dimension  $d$ .

[13] The last row of the matrix  $Y$  is  $Y_d = Z/\sqrt{d} = Z'$ . The first  $d - 1$  components of the vector  $Y$  can be denoted as  $U$  and the last component is  $Z'$ , i.e.,  $Y = (U, Z')$ . Hence the simulation involves resampling from the conditional PDF ( $f(U|Z')$ ).

[14] 2. An aggregate flow (i.e., annual flow)  $z^*$  is generated from an appropriate model fitted to the annual flow data. This could be a traditional autoregressive model [Salas, 1985] or a K-NN bootstrap approach [Lall and Sharma, 1996] or a kernel density estimator based method [Sharma et al., 1997] or a modified K-NN bootstrap [Prairie et al., 2005, 2006] or a block bootstrap resampling [Vogel and Shallcross, 1996]. Here we used the modified K-NN [Prairie et al., 2006].

[15] If a simple K-NN based approach is applied, the annual flows will be resampled from the historic data only generating values seen in the historic record. To generate



**Figure 3.** Box plots of monthly and annual statistics for flows at Green River at Green River, Utah. The box represents the interquartile range, and whiskers extend to the 5th and 95th percentile of the simulations. The statistics of the historic data are represented as a triangle.

annual values not seen before, either the kernel density estimator, the modified K-NN, or a traditional parametric model can be implemented.

[16] 3. K-nearest neighbors (corresponding to K historic years) of the generated  $z_{sim}' = z^*/\sqrt{d}$  are identified. The nearest neighbors are obtained by computing the distance between the generated  $z_{sim}'$  and the historic  $Z'$ . The neighbors are assigned weights based on the function

$$W(k) = \frac{1}{k \sum_{i=1}^k \frac{1}{i}} \quad k = 1, 2, \dots, K. \quad (5)$$

This weight function gives more weight to the nearest neighbors and less to the farthest neighbors. For further discussion on the choice of the weight function, readers are referred to *Lall and Sharma* [1996].

[17] The number of nearest neighbors K is based on the heuristic scheme  $K = \sqrt{N}$  where N equals the sample size [Lall and Sharma, 1996], following the asymptotic arguments of *Fukunaga* [1990]. Objective criteria such as generalized cross validation (GCV) can also be used. The above heuristic scheme has performed well in a variety of applications [Lall and Sharma, 1996; Rajagopalan and Lall, 1999; Yates et al., 2003].

[18] Using these weights as a probability metric, one of the neighbors is resampled. Suppose the selected neighbor corresponds to year j of the historic record.

[19] 4. The corresponding vector  $Y^*$  is created as

$$Y^* = (U_j, z'_{sim}) \quad (6)$$

[20] 5. The final step is the back rotation to the original space,

$$x^* = R^T Y^* \quad (7)$$

where  $x^*$  is the vector of disaggregated (i.e., monthly) flows that will sum to  $z^*$ .

[21] Steps 2–5 are repeated to generate ensembles of monthly streamflows. The same steps can be used for spatial disaggregation, in which case the matrix X represents the spatial streamflows and Z represents the spatial aggregate flow.

[22] Even though we resample historic data, steps 4 and 5 enable the simulation of monthly values not seen in the historic record and can also generate negative values. However, in our application here the negative values simulated were extremely small in number; less than 0.4% of the simulated values for all gauges were negative.

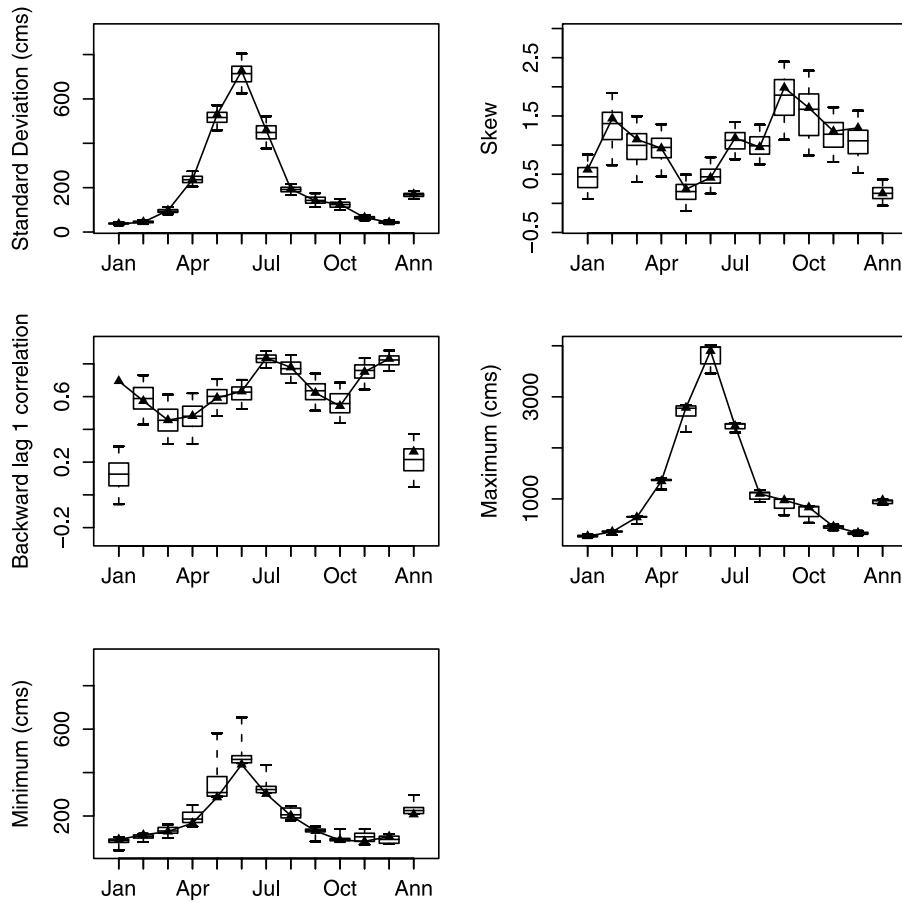


Figure 4. Same as Figure 3 but for flows at Colorado River at Lees Ferry, Arizona.

2.2. Numerical Example

[23] To further explain the algorithm described above, a simple numerical example is presented. In this example we assume two variables (say, two seasons) and they sum to the aggregate flows. The  $\mathbf{X}$  matrix (seasonal flows) and the vector  $Z$  (aggregate flows) are given as

$$\mathbf{X} = \begin{bmatrix} 222 & 585 \\ 232 & 1206 \end{bmatrix}, Z = \begin{bmatrix} 454 \\ 1791 \end{bmatrix} \quad (8)$$

The rotation matrix  $\mathbf{R}$  is obtained as described in step 1 of algorithm resulting in

$$\mathbf{R} = \begin{bmatrix} 0.7071068 & -0.7071068 \\ 0.7071068 & 0.7071068 \end{bmatrix}. \quad (9)$$

The rotated matrix  $\mathbf{Y}$  is computed as

$$\mathbf{Y} = \mathbf{R}\mathbf{X} = \begin{bmatrix} -7 & -439 \\ 321 & 1266 \end{bmatrix}. \quad (10)$$

Note that the last row of  $\mathbf{Y}$  is equal to  $Z/\sqrt{d}$  (here  $d = 2$ ). Suppose the simulated aggregate flow is  $z_{sim} = 736$ ; then

$$z'_{sim} = \frac{736}{\sqrt{2}} = 520.$$

On the basis of the resampling method described in step 3 of the algorithm above, suppose that we chose the second year; then the vector

$$Y^* = (u_{2,1}, z'_{sim}) = [-439 \quad 520]. \quad (11)$$

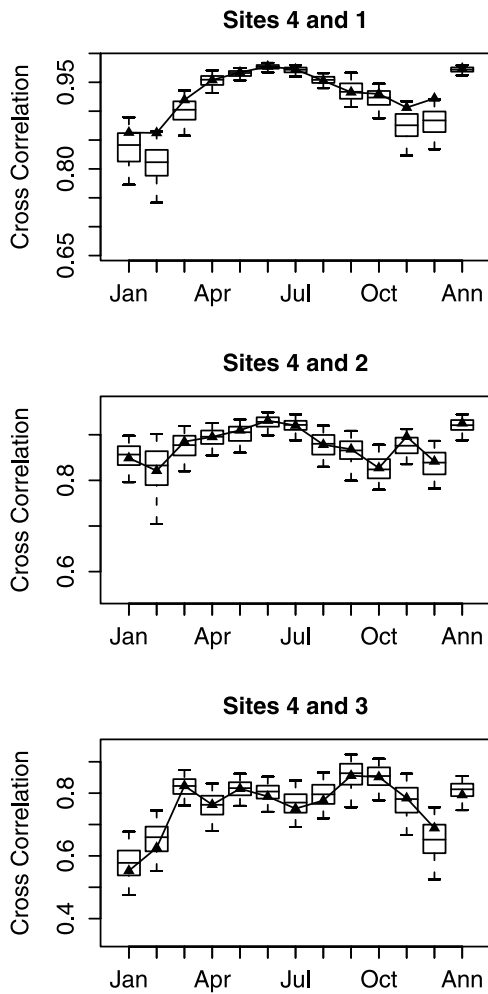
The disaggregated vector  $x_{sim}$  is obtained as

$$x_{sim} = \mathbf{R}^T \mathbf{Y}^* = \begin{bmatrix} 0.7071068 & 0.7071068 \\ -0.7071068 & 0.7071068 \end{bmatrix}^T \begin{bmatrix} -439 \\ 520 \end{bmatrix} = \begin{bmatrix} 57 \\ 679 \end{bmatrix}. \quad (12)$$

Note that the additivity property  $\sum x_{sim} = z_{sim}$  is satisfied.

3. Model Evaluation

[24] The performance of the K-NN space-time disaggregation approach is evaluated by applying it to four streamflow locations on the Upper Colorado River basin shown in Figure 1. These gauges are Colorado River near Cisco, Utah (site 1); Green River at Green River, Utah (site 2); San Juan River near Bluff, Utah (site 3); and Colorado River at Lees Ferry, Arizona (site 4). Monthly natural streamflows at these locations are available for the 98-year period spanning 1906–2003. Naturalized streamflows are computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use, etc.) from the recorded historic flows. (The natural flow data and additional reports describing



**Figure 5.** Box plots of monthly and annual cross correlation between the streamflows at the four locations.

these data are available at <http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html>.)

[25] The disaggregation schematic is shown in Figure 2. In this, we begin with an annual streamflow at an “index” gauge which is temporally disaggregated to 12 monthly flows. The monthly flows are then disaggregated to flows at the spatial locations. Thus the disaggregation algorithm is applied twice, first for the temporal and second for the spatial disaggregation. The “index” gauge is an imaginary

gauge whose monthly flows are created as the sum of the monthly flows at all the four locations. The annual flow at the index gauge was generated from the modified K-NN lag-1 approach [Prairie et al., 2005, 2006]. Using the space-time disaggregation approach, we made 500 simulations each of 98 years length. The following statistics are calculated from the simulations and compared with those from the historic data to evaluate the performance of the proposed approach.

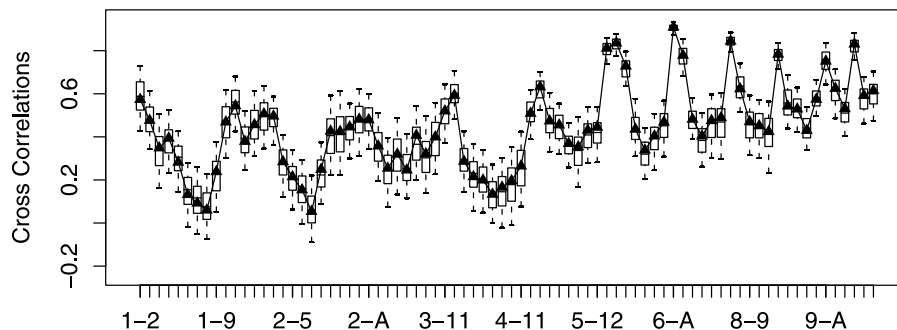
**3.1. Performance Statistics**

[26] These performance statistics include monthly and annual (1) mean, (2) standard deviation, (3) coefficient of skew, (4) maximum, (5) minimum, (6) backward lag-1 autocorrelation of the flows at the four locations, (7) probability density functions (PDF), (8) correlation of flows between the locations, and (9) surplus and drought statistics. Comparisons with a standard parametric alternative [Salas et al., 1980] are also provided.

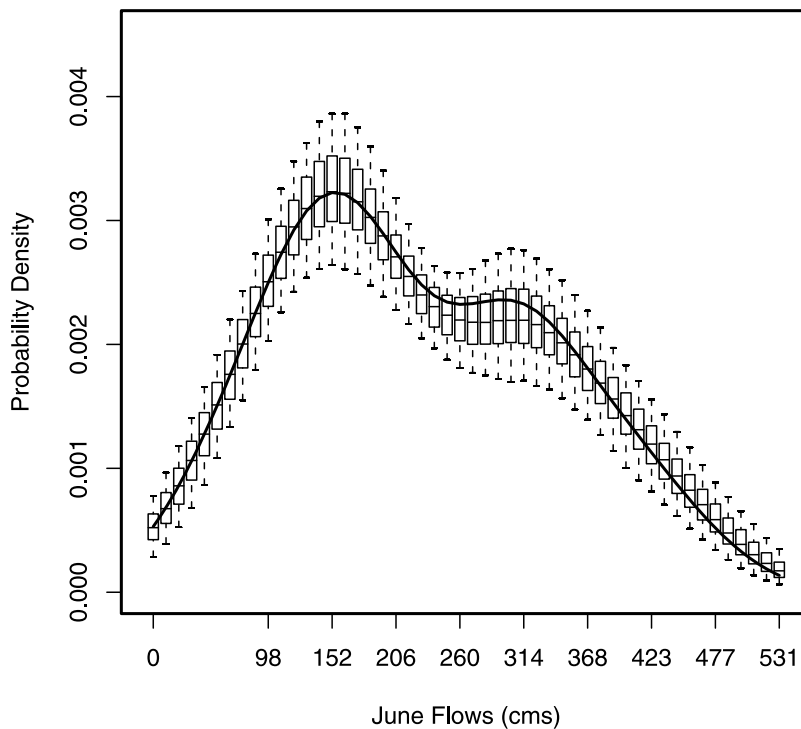
**4. Results**

[27] The results are displayed in Figures 3–10 as box plots where the box represents the interquartile range and whiskers extend to the 5th and 95th percentile of the simulations (note this is different from the standard box plot definition). The statistics of the historic data are represented as a triangle connected by a solid line. Performance on a given statistic is judged as good when the historic value falls within the interquartile range of the box plots, while increased variability is indicated by a wider box plot.

[28] The mean was well reproduced at all sites and therefore not included in Figures 3 and 4. Performance statistics of Green River at Green River, Utah are shown in Figure 3. The standard deviation and skews are well preserved for most all the months and at the annual time step. The low flow months January and February skews are slightly underrepresented. The lag-1 autocorrelations are also well simulated though February is slightly over correlated. However, the correlation between the first month of a year and the last month of the preceding year is not preserved. At the index gauge the temporal disaggregation does not incorporate this dependence; therefore it is not captured in the simulations at the spatial locations. In this basin the flows are largely snowmelt driven, and thus the first (January) and last (December) months of the calendar



**Figure 6.** Temporal cross correlation pairs for streamflows at Colorado River at Lees Ferry, Arizona. The x-axis sequence is 1–2, 1–3, . . . , 1–12, 1–A, 2–3, 2–4, . . . , 2–12, 2–A, 3–4, . . . Months are numbered according to calendar year, and letter A represents annual.



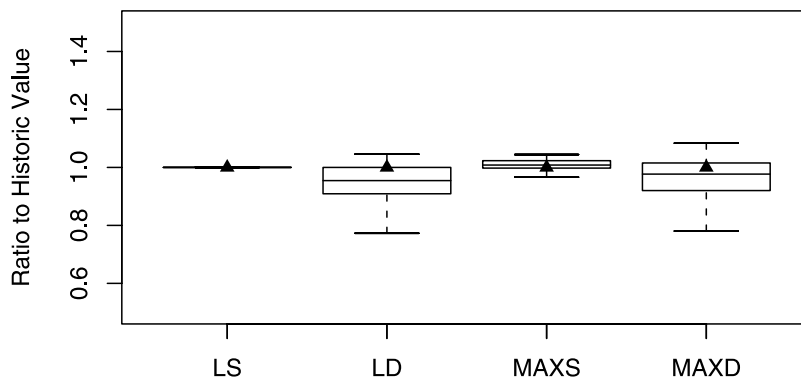
**Figure 7.** Box plots of PDF of June flows from San Juan River near Bluff, Utah.

year are part of the low-flow season (accounting for about 4% on the annual flow); hence capturing their correlation was not essential. However, we deemed it important to capture the correlations in the remaining months especially during the high-flow months which are well preserved. *Kumar et al.* [2000] resolved this issue with their optimization framework but at a computation cost. Linear adjustment procedures have also been developed to capture the first month’s correlation with the last month of the preceding year [*Grygier and Stedinger, 1988; Lane and Frevert, 1990; Koutsoyiannis and Manetas, 1996; Koutsoyiannis, 2001*]. They all, though, involve estimating several additional parameters and can affect reproduction of other statistics.

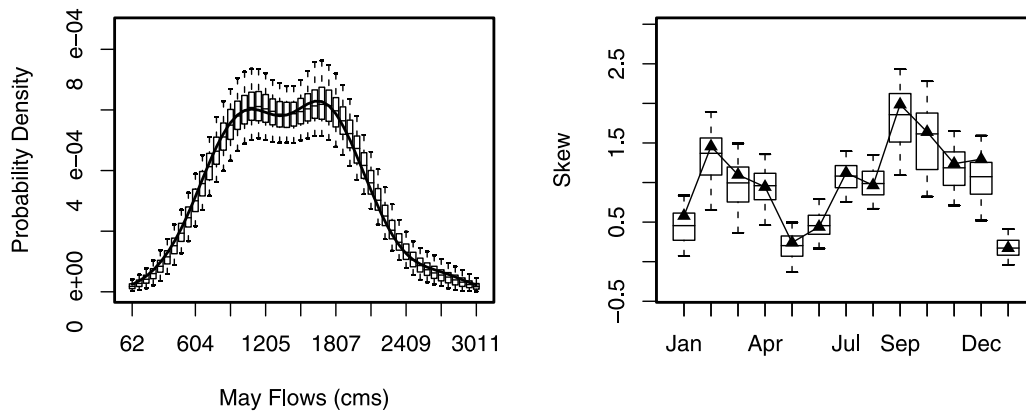
[29] The maximum and minimum flow statistics are also reasonably well simulated for most of the months. Extrapolation beyond the maximum historic flow occurs more extensively for some months (January, February, August–

September, November), while other months (March–July, October, December) display limited to no extrapolation. A very small number (0.4%) of negative numbers were generated, mostly in low-flow months, and had no significant impact on statistics. Similar results were obtained for the flows at Colorado River near Lees Ferry, Arizona (Figure 4) and also at the remaining two locations (figures not shown).

[30] The spatial cross correlation between the monthly flows at Colorado River at Lees Ferry, Arizona (site 4 the downstream location) and the other three gauges are shown in Figure 5. The cross correlations are very well captured during the spring months (the high-flow season) and also during other months. There is a slight undersimulation of the cross correlations during the low-flow months of January–March and November–December. Figure 6 displays the temporal cross correlation of the monthly and annual flows at several lags for the Colorado River at Lees Ferry, Arizona (site 4). These statistics are also very well simulated.



**Figure 8.** Box plots of drought statistics for Colorado River at Lees Ferry, Arizona.



**Figure 9.** Box plots from nonparametric disaggregation model of (left) PDF of May flows and (right) monthly and annual coefficient of skew from Colorado River at Lees Ferry, Arizona.

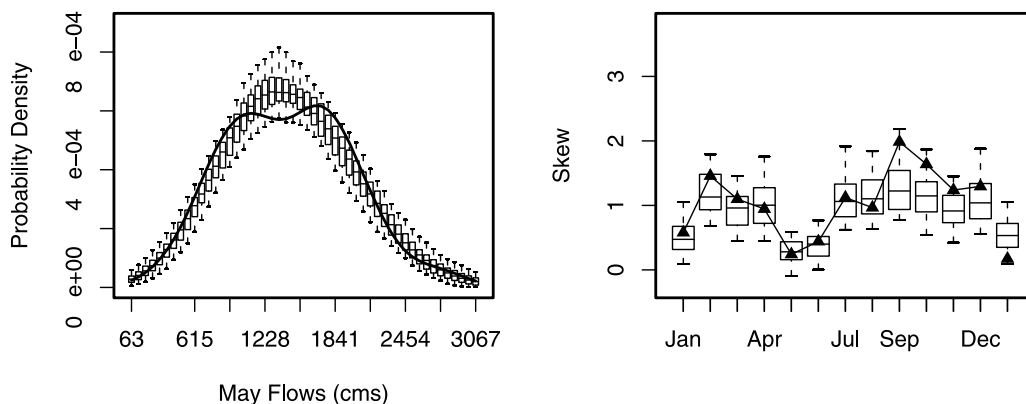
[31] As described earlier, one of the advantages of non-parametric methods is the ability to capture any arbitrary PDF structure. To test this we estimated the PDF from the simulations and compared them with those of the observed data. Figure 7 presents the PDF for June flows at San Juan River near Bluff, Utah. The PDF of the historic data is shown by the solid line, and the boxes and whiskers are those of the simulations. The simulations capture the non-normal feature of the historic PDF very well. Nonparametric kernel density estimators are used to compute the PDF [Bowman and Azzalini, 1997]. Similar performance was seen with PDFs from other months and locations.

[32] To evaluate the performance of the disaggregation approach in capturing longer temporal properties, we calculated surplus and drought statistics which include the longest surplus (LS), the longest drought (LD), the maximum surplus (MS), and the maximum deficit (MD) based on the long-term mean as the threshold for drought. These statistics for the flows at Colorado River at Lees Ferry are shown in Figure 8. The LS statistics exactly reproduces the historic data. The LD statistic is captured within the interquartile range of the simulations though it tends to be underrepresented. The MS is again captured within the interquartile range of the simulations and is well represented. While the MD statistic shows the greatest variability of the all these statistics, though captured within the interquartile range, the MD tends to be underrepresented.

The simulations generate droughts that are longer in length and greater in magnitude than those in the observed record, though these are only generated for less than 25% of the simulations.

**4.1. Comparison With a Parametric Model**

[33] We compared the simulations from the K-NN space-time disaggregation approach developed in this research to a traditional parametric model (of the form in equation (1)) developed by Mejia and Rousselle [1976] and Salas et al. [1980]. The parametric disaggregation models are designed to capture all the basic statistics, but as described earlier, they have difficulty in capturing nonnormal PDF structure and also the coefficient of skewness. These two statistics depend upon the appropriate transformations used to transform the data to a normal distribution. The transformations applied for the parametric model all passed a skewness test for normality, i.e., the transformed data had a coefficient of skew close to zero. Figures 9 and 10 display the PDF of May flows and the monthly and annual coefficient of skew for the Colorado River at Lees Ferry, Arizona. The historic PDF displays a clear bimodal feature which is extremely well preserved by the K-NN disaggregation model (Figure 9), while the parametric model (Figure 10) is unable to capture this feature, instead reproducing a normal structure. Similar results are seen with the coefficient of skewness that is well represented by the nonparametric model but not by



**Figure 10.** Box plots from parametric disaggregation model of (left) PDF of May flows and (right) monthly and annual coefficient of skew from Colorado River at Lees Ferry, Arizona.



the parametric model. It should be noted that even though the transformations were effective, the parametric model was fitted to the transformed data and hence does not guarantee the reproduction of the statistics in the original space.

## 5. Summary and Discussion

[34] We have presented a simple, robust, and parsimonious framework for space-time simulation of streamflows on a large river network. We adapted the *Tarboton et al.* [1998] framework but used a K-NN approach to construct and simulate from the conditional PDF. The model captures all the distributional properties and the spatial dependence of the flows at all the locations. Simulating space-time flow scenarios conditioned upon large-scale climate information (e.g., El Niño–Southern Oscillation, etc.) for seasonal forecasts can be easily achieved.

[35] A few limitations exist in the proposed nonparametric disaggregation approach, which should be considered. An obvious limitation was the inability to capture the correlation between the first month of one year and the last month of the previous year. As presented, the proposed approach only solved for the monthly flows conditioned on the dependence structure for the current year. Incorporating the last month's flow in the conditional function could be explored to remedy this limitation in scenarios where preserving this correlation is essential. Another limitation in the proposed approach can arise if extreme values are of interest. Extrapolation beyond observed monthly flows is limited in comparison with parametric counterparts. This limitation needs to be considered on a case-by-case basis. With the proposed approach this limitation can be addressed with the choice of a proper annual flow model. An annual simulation model that generates more extreme annual flows will in turn generate more extreme monthly values.

[36] The proposed approach involved a two-step process in which the temporal disaggregation was first performed followed by the spatial disaggregation. The work of *Kumar et al.* [2000] considers a simultaneous space-time disaggregation based on the K-NN method in an optimization framework, albeit with significant computational effort. For the annual to monthly disaggregation, an approach that blends elements of the method presented here with the optimization approach of *Kumar et al.* [2000] may provide a means to perform a simultaneous disaggregation.

[37] Efforts are under way to integrate this framework with a basin-wide salinity model [*Prairie et al.*, 2005] to generate salinity ensembles. Additionally, tree ring reconstructions of past annual streamflows at Lees Ferry will be incorporated into this approach to simulate (i.e., reconstruct) monthly streamflows at all the locations in the Upper Colorado River basin that may include extreme events addressing the second aforementioned limitation. Together these projects will enable the evaluation of various policy strategies in the basin.

[38] **Acknowledgments.** Funding for this research by the Bureau of Reclamation's Lower Colorado regional office via grant 04PG303326 is gratefully acknowledged. Continued support by Kib Jacobson of the Bureau of Reclamation's Upper Colorado regional office is appreciated. Insightful comments from David Tarboton, Demetris Koutsoyiannis, one anonymous reviewer, and associate editor Tim Cohn are thankfully acknowledged. Thanks are also due the Center for Advanced Decision Support in Water

and Environmental Systems (CADSWES) at the University of Colorado, Boulder, for use of its facilities and computational support.

## References

- Bowman, A. W., and A. Azzalini (1997), *Applied Smoothing Techniques for Data Analysis*, Clarendon, Oxford, U. K.
- Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition*, Elsevier, New York.
- Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona (2005), A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts, *Water Resour. Res.*, *41*, W10410, doi:10.1029/2004WR003467.
- Grygier, J. C., and J. R. Stedinger (1988), Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resour. Res.*, *24*(10), 1574–1584.
- Koutsoyiannis, D. (1992), A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series, *Water Resour. Res.*, *28*(12), 3175–3191.
- Koutsoyiannis, D. (1999), Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology, *Water Resour. Res.*, *35*(4), 1219–1229.
- Koutsoyiannis, D. (2001), Coupling stochastic models of different time-scales, *Water Resour. Res.*, *37*(2), 379–391.
- Koutsoyiannis, D., and A. Manetas (1996), Simple disaggregation by accurate adjusting procedures, *Water Resour. Res.*, *32*(7), 2105–2117.
- Kumar, D. N., U. Lall, and M. R. Peterson (2000), Multisite disaggregation of monthly to daily streamflow, *Water Resour. Res.*, *36*(7), 1823–1833.
- Lall, U. (1995), Recent advances in nonparametric function estimation: Hydraulic applications, *U. S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994, Rev. Geophys.*, *33*, 1093–1102.
- Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resour. Res.*, *32*(3), 679–693.
- Lane, W. L. (1979), *Applied Stochastic Techniques, Users Manual*, Eng. and Res. Cent., Bur. of Reclam., Denver, Colo.
- Lane, W. L. (1982), Corrected parameter estimates for disaggregation schemes, in *Statistical Analysis of Rainfall and Runoff*, edited by V. P. Singh, Water Resour. Publ., Highlands Ranch, Colo.
- Lane, W. L., and D. K. Frevert (1990), *Applied Stochastic Techniques, Personal Computer Version 5.2, Users Manual*, Earth Sci. Div., Bur. of Reclam., Denver, Colo.
- Mejia, J. M., and J. Rousselle (1976), Disaggregation models in hydrology revisited, *Water Resour. Res.*, *12*(2), 185–186.
- Moon, Y.-I., and U. Lall (1994), Kernel function estimator for flood frequency analysis, *Water Resour. Res.*, *30*(11), 3095–3103.
- Prairie, J. R., B. Rajagopalan, T. J. Fulp, and E. A. Zagona (2005), Statistical nonparametric model for natural salt estimation, *J. Environ. Eng.*, *131*(1), 130–138.
- Prairie, J. R., B. Rajagopalan, T. J. Fulp, and E. A. Zagona (2006), Modified K-NN model for stochastic streamflow simulation, *J. Hydrol. Eng.*, *11*(4), 371–378.
- Rajagopalan, B., and U. Lall (1999), A k-nearest-neighbor simulator for daily precipitation and other weather variables, *Water Resour. Res.*, *35*(10), 3089–3101.
- Salas, J. D. (1985), 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.
- Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane (1980), *Applied Modeling of Hydrologic Time Series*, 484 pp., Water Resour. Publ., Highlands Ranch, Colo.
- Santos, E. G., and J. D. Salas (1992), Stepwise disaggregation scheme for synthetic hydrology, *J. Hydraul. Eng.*, *118*(5), 765–784.
- Sharma, A., and R. O'Neill (2002), A nonparametric approach for representing interannual dependence in monthly streamflow sequences, *Water Resour. Res.*, *38*(7), 1100, doi:10.1029/2001WR000953.
- Sharma, A., D. G. Tarboton, and U. Lall (1997), Streamflow simulation: A nonparametric approach, *Water Resour. Res.*, *33*(2), 291–308.
- Singhtrattna, N., B. Rajagopalan, M. Clark, and K. Krishna Kumar (2005), Forecasting Thailand summer monsoon rainfall, *Int. J. Climatol.*, *25*(5), 649–664.
- Souza Filho, F. A., and U. Lall (2003), Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm, *Water Resour. Res.*, *39*(11), 1307, doi:10.1029/2002WR001373.
- Srinivas, V. V., and K. Srinivasan (2001), Post-blackening approach for modeling periodic streamflows, *J. Hydrol.*, *241*, 221–269.

- Srinivas, V. V., and K. Srinivasan (2005), Hybrid moving block bootstrap for stochastic simulation of multi-site multi-season streamflows, *J. Hydrol.*, 302, 307–330.
- Stedinger, J. R., and R. M. Vogel (1984), Disaggregation procedures for generating serially correlated flow vectors, *Water Resour. Res.*, 20(1), 47–56.
- Stedinger, J. R., D. Pei, and T. A. Cohn (1985), A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations, *Water Resour. Res.*, 21(5), 665–675.
- Tao, P. C., and J. W. Delleur (1976), Multistation, multiyear synthesis of hydrologic time series by disaggregation, *Water Resour. Res.*, 12(6), 1303–1312.
- Tarboton, D. G., A. Sharma, and U. Lall (1998), Disaggregation procedures for stochastic hydrology based on nonparametric density estimation, *Water Resour. Res.*, 34(1), 107–119.
- Todini, E. (1980), The preservation of skewness in linear disaggregation schemes, *J. Hydrol.*, 47, 199–214.
- Valencia, D. R., and J. C. Schaake (1973), Disaggregation processes in stochastic hydrology, *Water Resour. Res.*, 9(3), 580–585.
- Vogel, R. M., and A. L. Shallcross (1996), The moving blocks bootstrap versus parametric time series models, *Water Resour. Res.*, 32(6), 1875–1882.
- Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003), A technique for generating regional climate scenarios using a nearest neighbor bootstrap, *Water Resour. Res.*, 39(7), 1199, doi:10.1029/2002WR001769.
- 
- T. Fulp, Bureau of Reclamation, Boulder Canyon Operations Office, Lower Colorado Region, P.O. Box 61470, Boulder City, NV 89006-1470, USA. (tfulp@lc.usbr.gov)
- U. Lall, Department of Earth and Environmental Engineering, Columbia University, 500 West 120th Street, New York, NY 10027, USA. (ula2@columbia.edu)
- J. Prairie, Bureau of Reclamation, University of Colorado, 421-UCB, Boulder, CO 80309, USA. (prairie@colorado.edu)
- B. Rajagopalan, Department of Civil, Environmental, and Architectural Engineering, University of Colorado, 428-UCB, Boulder, CO 80309-0428, USA. (balajir@colorado.edu)