A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow

Kenneth Nowak,[1,2] James Prairie,[3] Balaji Rajagopalan,[1,2,4] and Upmanu Lall[5]

Received 17 August 2009; revised 2 February 2010; accepted 31 March 2010; published 14 August 2010.

[1] Streamflow disaggregation techniques are used to distribute a single aggregate flow value to multiple sites in both space and time while preserving distributional statistics (i.e., mean, variance, skewness, and maximum and minimum values) from observed data. A number of techniques exist for accomplishing this task through a variety of parametric and nonparametric approaches. However, most of these methods do not perform well for disaggregation to daily time scales. This is generally due to a mismatch between the parametric distributions appropriate for daily flows versus monthly or annual flows, the high dimension of the disaggregation problem, compounded uncertainty in parameter estimation for multistage approaches, and the inability to maintain flow continuity across disaggregation time period boundaries. We present a method that directly simulates daily data at multiple locations from a single annual flow value via *K*-nearest neighbor (K-NN) resampling of daily flow proportion vectors. The procedure is simple and data driven and captures observed statistics quite well. Furthermore, the generated daily data are continuous and display lag correlation structure consistent with that of the observed data. The utility and effectiveness of this approach is demonstrated for selected sites in the San Juan River Basin, located in southwestern Colorado, and later compared with the disaggregation technique of Prairie et al. (2007) for several locations in the Colorado River Basin.

Citation: Nowak, K., J. Prairie, B. Rajagopalan, and U. Lall (2010), A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow, *Water Resour. Res.*, *46*, W08529, doi:10.1029/2009WR008530.

1. Introduction

[2] As the demand and use of water continues to increase, water management has become a complex task requiring detailed models and planning in order to effectively manage this valuable resource. Seasonal diversions, irrigation needs, instream flow requirements and hydropower all contribute to the challenge of present-day water administration. In order to maximize the utility of models with a high level of physical system detail, flow inputs need to be rich in variety and of an appropriate time step. These data should represent a range of hydrologic scenarios and be temporally consistent with the system obligations (e.g., daily hydropower demands, seasonal irrigation, etc.). Historic flows are not always available such that the data needs of a model are appropriately met. This highlights the utility of synthetic data techniques. Methods for generating synthetic streamflows provide the data needed for operations models. These methods are capable of producing extreme events (drought/surplus) that are greater in magnitude and/or duration compared to those of the observed record, thus further exploring operations beyond the historic streamflows.

[3] Methods to generate flows usually begin at the annual time step for a location of particular significance, (i.e., due to interstate compacts, physical location, etc.) in order to ensure statistical properties are effectively preserved at this location. However, as eluded earlier, single-site, annual flow values may not be sufficient for detailed models. Thus, the need to disaggregate flow values to multiple sites and finer time scales arises. The task of appropriately distributing flow in space and time requires that distributional statistics (i.e., mean, variance, skewness, and maximum and minimum values) are reproduced for all sites, while also preserving summability, cross correlations, and continuity. Parametric approaches [*Grygier and Stedinger*, 1988; *Stedinger and Vogel*, 1984; *Valencia and Schaake*, 1973] for addressing disaggregation have traditionally been framed on a linear approach, similar in structure to an auto-regressive (AR) model. This approach is represented by equation (1),

$$\mathbf{X}_t = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{v}_t, \quad (1)$$

where $\mathbf{X}_t$ is the disaggregate variables (e.g., monthly flows) at time t, $\mathbf{z}_t$ is the aggregate variable (e.g., annual flow), and $\mathbf{v}_t$ is a vector of random values from a normal distribution. The elements of $\mathbf{X}_t$ sum to $\mathbf{z}_t$, also known as the summability criteria, which is an important element for generating space-

[1]Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, Colorado, USA.
[2]Center for Advanced Decision Support for Water and Environmental Systems (CADSWES), Boulder, Colorado, USA.
[3]Bureau of Reclamation, University of Colorado, Boulder, Colorado, USA.
[4]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado, USA.
[5]Earth and Environmental Engineering, Columbia University, New York, New York, USA.

0043-1397/10/2009WR008530

time streamflows on a river network. Model parameters contained in matrix **A** and **B** are estimated such that the simulations preserve the cross correlation between and among the variables in addition to the summability property.

[4] Main drawbacks to this general class of models are that (1) the data are assumed to be normally distributed and (2) the **A** and **B** matrices used can quickly become unwieldy as the dimensions of the disaggregation increase (e.g., disaggregating to daily time scale and/or many spatial locations). Often the data at each time scale and location needs to be transformed to normality using a log or power transform. However, this does not ensure that summability or the observed statistics are preserved when the results are returned to the original space. Efforts to reduce the computational intensity associated with large **A** and **B** matrices suggest a stepwise disaggregation approach [*Santos and Salas*, 1992] (i.e., annual to seasonal to monthly, etc.) which results in additional challenges of continuity across the generated flows at the finer time scale. In summary, this class of methods is inadequate for disaggregation finer than the monthly time scale.

[5] Recently, several nonparametric techniques have been put forth that offer improvements, such as the ability to capture non-Gaussian distributions [*Prairie et al.*, 2007; *Tarboton et al.*, 1998]. The nonparametric methods cast the problem as simulating from the conditional probability density function (PDF) $f(\mathbf{X}_t|\mathbf{z}_t)$. Tarboton et al. employ a kernel density estimation-based approach to simulating from the conditional PDF. While this approach is effective at ameliorating the issues associated with non-normal distributions, it remains computationally intensive at even the annual to monthly scale disaggregation. Furthermore, kernel methods suffer from boundary biases that worsen with dimension. *Prairie et al.* [2007] circumvent this issue by replacing the kernel density estimation with a computationally faster *K*-nearest neighbor (K-NN)-based time resampling approach [*Lall and Sharma*, 1996]. This retains the ability to preserve non-Gaussian distributions while significantly reducing computing demands and also alleviates the boundary issue associated with kernel methods. However, both struggle with regard to basins with sites that vary significantly in flow magnitude and furthermore, these methods have only been shown effective for the disaggregation of annual to monthly flows.

[6] Very few techniques have proven useful for disaggregation to daily values. This is generally due to a mismatch between the parametric distributions appropriate for daily flows versus monthly or annual flows, the high dimension of the disaggregation problem, compounded uncertainty in parameter estimation for multistage approaches, and the inability to maintain flow continuity across disaggregation time period boundaries. In the realm of precipitation generation, there has been considerably more success in producing daily precipitation values from a wet spell total, which is largely attributed to the lack of a continuity requirement [*Bogardi et al.*, 1993; *Chang et al.*, 1984; *Chin*, 1977; *Lall et al.*, 1996; *Srikanthan and McMahon*, 2001]. *Kumar et al.* [2000] have put forth an effective approach for disaggregating monthly to daily streamflow by casting the task as a linear programming problem, which is then optimized in adherence to a number of constraints that preserve statistical relationships and continuity [*Kumar et al.*, 2000]. However, the method is very computationally intensive; for a five site, single month to daily disaggregation, there are more than 1500 decision variables which must be solved for in order to reach a solution. This poses obvious challenges for larger basins with many gauging locations.

[7] Here we propose a nonparametric disaggregation approach that resamples a historic vector of daily proportions, conditioned on an annual flow, via K-NN resampling. The disaggregated values are obtained by projecting the aggregate flows on to the proportion vector. This ensures summabililty, continuity across the daily time scale, and positive disaggregated values unlike all the method discussed above. Furthermore, it has the ability to capture all the distributional and cross-dependency properties. The method can be readily applied to disaggregation at multiple time and space scales. Independently, a similar approach has also been proposed by *Lee* [2008], which was applied to multisite, annual to monthly disaggregation. The method employs an adjusting procedure for resampled historic data with a genetic algorithm to increase variability [*Lee*, 2008]. To demonstrate the utility and effectiveness of our proposed method, it is used to generate daily flows at three sites on the San Juan River, a tributary of the Colorado River located in southwestern Colorado. Additionally, a comparison with the results of *Prairie et al.* [2007] for locations in the Colorado River Basin is provided.

2. Annual to Daily Disaggregation Method

[8] Nonparametric space-time streamflow disaggregation can be thought of as simulating from the conditional probability distribution function (PDF) $f(\mathbf{X}_t|\mathbf{z}_t)$, where $\mathbf{z}_t$ is a vector of annual aggregate flows to be disaggregated and $\mathbf{X}_t$ is a matrix of flows that sum to $\mathbf{z}_t$.

[9] Here we cast the problem as a conditional simulation of the proportion vector from the conditional PDF $f(\mathbf{X}_t|\mathbf{p}_t)$, where $\mathbf{p}_t$ is the vector of daily proportions, whose elements sum to unity by definition, and z is the aggregate flow. The simulation uses a K-NN resampling approach [*Lall and Sharma*, 1996; *Prairie et al.*, 2007] in that, *K*-nearest neighbors to z are identified from historical flows and one of them is resampled using a weight metric (see equation (2)) that gives more weight to those neighbors closest in magnitude to z. This is akin to constructing the conditional PDF locally in the neighborhood of z and simulating from it. The daily proportion vector corresponding to the selected neighbor (i.e., one of the historical years) is selected and the aggregate flow z is projected on to this proportion vector of the resampled historical year to obtain the disaggregated flows. This simple yet robust and local approach provides the ability to capture nonlinear and non-Normal features. Additionally, the method always generates positive values and the summability of the disaggregated flows to the aggregate flow is automatically preserved. The disaggregated values have a rich variety, including flows outside the range of observed values. Daily values at a single location are further disaggregated in space by projecting them on to the appropriate proportion vector. The implementation algorithm is described below followed by an example.

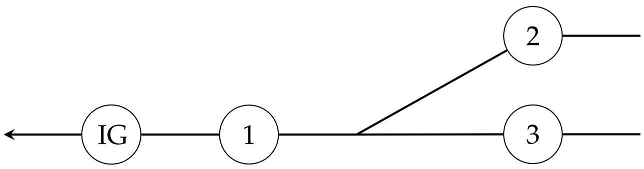

Figure 1. Simplified basin schematic with index gauge (IG): (1) San Juan River near Caraccas, CO; (2) San Juan River near Pagosa Springs, CO; and (3) Navajo River near Chromo, CO.

2.1. Implementation Algorithm

[10] For the purpose of describing the implementation of this technique, a single site, annual to daily disaggregation problem is considered. Extension of the method to a multisite network is straightforward and will be discussed later.

[11] 1. Observed daily streamflow values are converted to a proportion of the total annual flow of that particular year (i.e., daily flows in a given year are divided by the total annual flow of that year), giving a matrix **P**, with dimensions $n \times 365$ (n = number of years of observed data). By construction, each row of this matrix sums to unity.

[12] 2. Suppose Z is the annual flow that needs to be disaggregated to daily flows. First, K-nearest neighbors of Z are identified from the historical annual streamflow vector (**z**). The neighbors are computed based on the distance between Z and all the historical annual streamflow values. The number of nearest neighbors, $K = \sqrt{n}$, based on heuristics [*Lall and Sharma*, 1996] is found to be very effective in a variety of applications. Each of the K-nearest neighbors is assigned a weight based on the weight function proposed by *Lall and Sharma* [1996], given as,

$$W(i) = \left(\frac{1}{i}\right) \Big/ \left(\sum_{i=1}^{K} \frac{1}{i}\right), \tag{2}$$

where K is the number of nearest neighbors and i refers to the "neighbor index," with i = 1 being the closest of the nearest neighbors. Note the weights are normalized so that they sum to unity.

[13] 3. One of the K-nearest neighbors (i.e., one of the historical years) is chosen based on a weighted resampling (i.e., probability of picking a given year is determined by the corresponding weight from equation (2)). The proportion vector corresponding to the picked year, say, ($\mathbf{p}_y$) is selected. Next, the annual streamflow (Z) is projected on to this proportion vector to obtain the daily streamflow vector (**d**), which is given by

$$\mathbf{d} = Z \times \mathbf{p}_y. \tag{3}$$

Since the proportion vector sums to unity, the disaggregated daily streamflows sum to the aggregate annual flow.

[14] 4. Steps 2 and 3 are repeated to generate an ensemble of daily streamflows.

[15] To extend the method to both space and time, the process simply gains another dimension. The matrix **P** ($365 \times n$) will become an array of dimension $365 \times n \times s$, where s indicates the number of locations for disaggregation. Thus, the resampled proportion vector ($\mathbf{p}_y$) will now be a matrix of dimension $365 \times s$. For multisite disaggregation, the annual flow vector corresponds to an aggregate location downstream or an index gauge created by addition of streamflows at all the locations.

[16] This approach, as seen above, is simple to implement and can easily be applied to any space and time scales and obviates any need to adjust the postdisaggregation values, unlike methods developed previously. Also, it is equivalent to simulating from the conditional PDF $f(\mathbf{x}|z)$.

2.2. Numerical Example

[17] The following provides a simple numerical example of the technique, using limited data and a "4 day year."

$$\mathbf{p} = \begin{bmatrix} 1967 & .1 & .3 & .4 & .2 \\ 1968 & .15 & .25 & .35 & .25 \\ 1969 & .1 & .2 & .5 & .2 \\ 1970 & .05 & .15 & .65 & .15 \\ 1971 & .2 & .2 & .4 & .2 \\ 1972 & .1 & .2 & .4 & .3 \\ 1973 & .15 & .2 & .4 & .25 \\ 1974 & .05 & .1 & .8 & .05 \\ 1975 & .2 & .2 & .5 & .1 \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} 1967 & 35 \\ 1968 & 40 \\ 1969 & 33 \\ 1970 & 52 \\ 1971 & 43 \\ 1972 & 56 \\ 1973 & 38 \\ 1974 & 49 \\ 1975 & 32 \end{bmatrix}$$

$Z = 45$

On the basis of the weighted resampling of the "K" nearest neighbors, say year 1968 (y = 1968) is selected to be the nearest neighbor. Next, the simulated annual value (Z) is applied to the vector $\mathbf{p}_{1968}$ to produce the disaggregated values (**d**).

$$Z \times \mathbf{p}_y = \mathbf{d}$$

$$45 \times [.15 \quad .25 \quad .35 \quad .25] = [6.75 \quad 11.25 \quad 15.75 \quad 11.25]$$

$$\sum_{i=1}^{4} d_i = Z.$$

From the example, it can be seen that the technique is quite parsimonious and easy to implement. Section 3 discusses the ability of the method to reproduce observed statistics.

3. Model Evaluation

[18] The proportion based K-NN disaggregation method is evaluated by applying it to daily streamflow simulation at three gauges on the San Juan River. The selected sites are part of the U. S. Geological Survey (USGS) gauging network in the San Juan Basin, located in southwestern Colorado. Specifically, the gauging stations are San Juan River near Carracas, CO (09346400), San Juan River at Pagosa Springs (09342500), and Navajo River near Chromo, CO (09344400). Data for these sites were obtained from the USGS's National Water Information System: Web interface and span from 1972 to 1995. The gauge near Chromo, CO, is no longer maintained, thus truncating the common record at 1995. Daily flow values were converted from cubic feet per second (cfs) to acre feet per day. Figure 1 shows a schematic of these three gauges, the gauge IG is an index gauge constructed by aggregating the flows at the three locations.

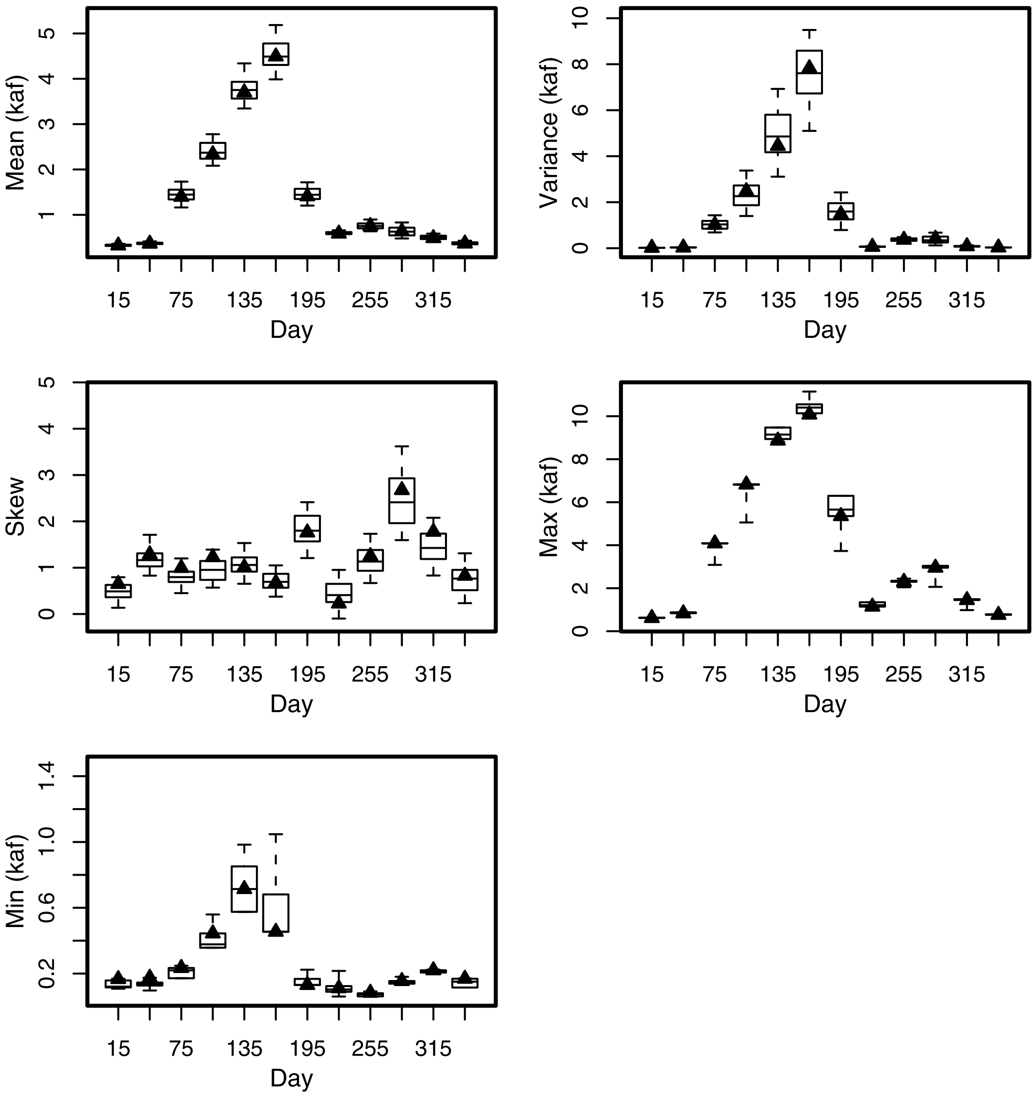

Figure 2. San Juan near Carracas, CO daily statistics; triangles represent values from observed data.

[19] For evaluation, 100 traces each of the 50 years in length for annual flow at the index gauge were generated using the K-NN lag-1 [*Lall and Sharma*, 1996] approach. Alternatively, a traditional AR(1) model could also have been applied, as the disaggregation method is independent of the annual simulation method. Thus, generated index gauge annual flows are disaggregated to daily streamflows at all the three locations using the methodology presented above. A suite of statistics are computed at monthly and daily time scales from the simulations and are presented as box plots along with the corresponding value from the historical data for comparison. The box of the box plot (see, e.g., Figure 2) represents the interquartile range (IQR) with the horizontal line being the median and the whiskers extending to the 5th and 95th percentiles. Performance on a given statistic is generally considered "good" when the historic value falls with the IQR.

3.1. Performance Statistics

[20] For the purpose of this work, the term "basic statistics" refers to mean, variance, coefficient of skew, maximum and minimum values, which are computed at the monthly and daily time scales for comparison. In addition, lag correlation, cross correlation between sites, and probability density functions (PDFs) are also used to evaluate the performance of the proposed method.

4. Results

[21] We describe the results of the daily statistics followed by the monthly. Then we implemented the proposed method for annual to monthly disaggregation at the three locations and compare them with the methodology proposed by *Prairie et al.* [2007].

4.1. Daily Results

[22] The daily statistics (mean, variance, skew, maximum and minimum values) are computed for a single day at the middle of each month and are presented. Box plots of all the basic distributional statistics for the San Juan River near Carracas, CO (Figure 2) indicate that the method effectively reproduces observed statistics at the daily time scale through all the months. The lag-1 autocorrelation is also well cap-

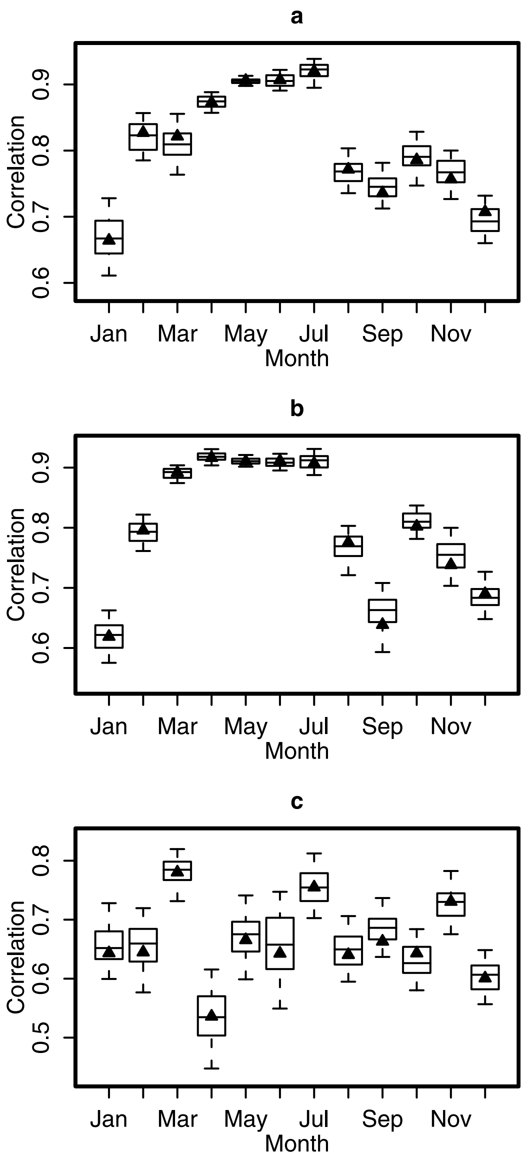

Figure 3. Daily lag-1 correlation by month for (a) San Juan River near Caraccas, CO; (b) San Juan River near Pagosa Springs, CO; and (c) Navajo River near Chromo, CO. Triangles represent values from observed data.

tured (Figure 3), indicating that the simulated flows have realistic continuity. The results were similar at other locations and days (figures not shown). Since the disaggregation is performed for each year separately, as to be expected, the correlation between the flow of the last day of a year and the first day of the following year is not well captured. This can be improved upon by including the streamflow of the last day from the previous year to the annual flow in computing the nearest neighbors. However, this can lead to deterioration of other statistics. Since the last and first months, December and January, are generally low flow months in this basin, we deem it not essential to explicitly preserve this autocorrelation so long as other key statistics are well preserved, especially during high flow months.

[23] The box plots of PDFs of daily flows from the simulations and the observed values for selected high flow month of May and low flow month of September are shown

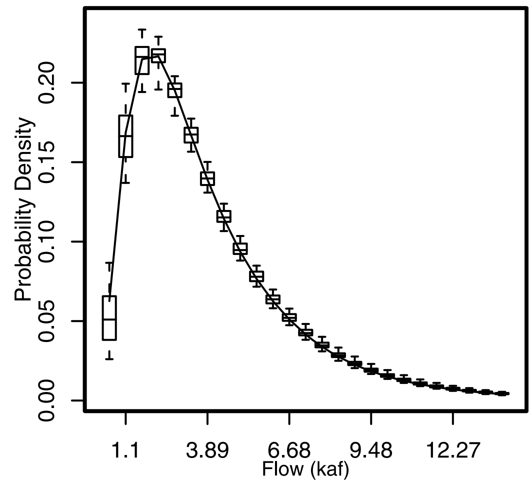

Figure 4. San Juan River near Carracas, CO, daily May flow PDF. Solid line is observed data, and box plots are simulated flow.

in Figures 4 and 5. This provides insights into the ability of the method to capture the entire distribution, arguably of greater importance than point statistics shown in previous figures. It can be seen that the historical PDFs of daily streamflows, which are highly non-normal, are very well captured by the model at different locations and months.

4.2. Monthly Results

[24] The disaggregated daily streamflows are aggregated to obtain monthly values and the suite of basic statistics is computed. Box plots of these statistics at the monthly time step are shown for the San Juan River near Carracas, CO (Figure 6). It can be seen that they are all well preserved. Also, the maximum and minimum values generated are beyond the range of the observed data, indicating that even though a resampling approach is used, it is possible to generate extremes outside the range of the data. Furthermore, all values are positive. Monthly cross-correlations at the same location are displayed in Figure 7, which are also

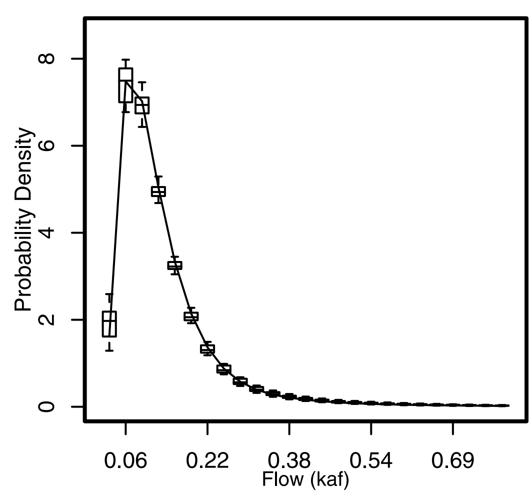

Figure 5. Daily September flow PDF for Navajo River near Chromo, CO. Solid line is observed data, and box plots are simulated flow.

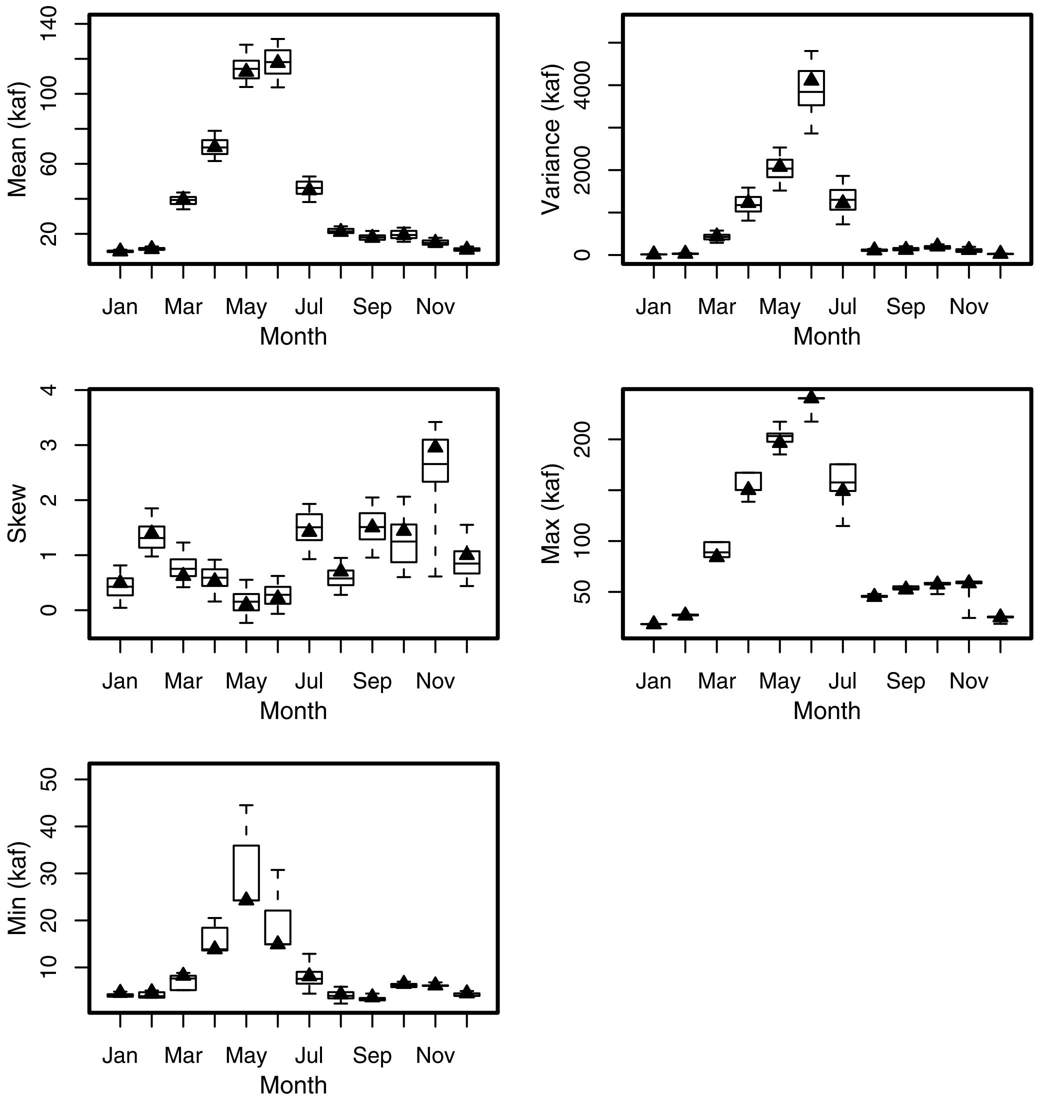

Figure 6. San Juan River near Carracas monthly statistics; triangles represent values from observed data.

well captured. Similar results were seen at the other locations (not shown). Cross correlation of monthly flows between Pagosa Springs and Chromo sites are shown in Figure 8. From the plot, it can be seen that the high correlation between monthly flows is reflected in the disaggregated simulations. The lower correlation in April may be due to spatial variability associated with the start of snowmelt and high flows. Similar monthly cross-correlation results were seen for other gauge combinations. Additionally, the simulated January flow PDFs for the San Juan near Carracas are presented as box plots with the observed PDF overlaid (Figure 9). This demonstrates the ability to capture the entire distribution, which is non-normal. These results indicate that the statistics at scales different from the scale of disaggregation are very well reproduced by this methodology.

4.3. Comparison With Methodology of *Prairie et al.* [2007]

[25] To demonstrate the wide applicability of the proposed technique to disaggregate flows at different temporal locations and basins, a brief comparison with the results of *Prairie et al.* [2007] on the Colorado River Basin is provided. As mentioned earlier, *Prairie et al.* [2007] present an effective nonparametric annual to monthly disaggregation which builds upon the earlier work of *Tarboton et al.* [1998]. Four streamflow gauges (Colorado River near Cisco, UT, Green River at Green River, UT, San Juan River near Bluff, UT and Colorado River at Lee's Ferry, AZ; see their Figure 1) in the Colorado River Basin were considered in their paper. We apply our methodology to these gauges in the same manner (i.e., annual to monthly disaggregation) for direct comparison. Figures 10 and 11 show box plots of monthly and annual basic statistics at the Lees Ferry, AZ gauge, from the methodology proposed here and their work, respectively. It can be seen that both the methods capture the basic statistics very well. However, the proportional disaggregation approach has the ability to produce values beyond the observed data range without generating any negative values. Furthermore, the methods have the ability to capture non-normal distributions very effectively, as can be seen from Figures 12 and 13, showing the June flow PDF for the San Juan River near Bluff, UT.

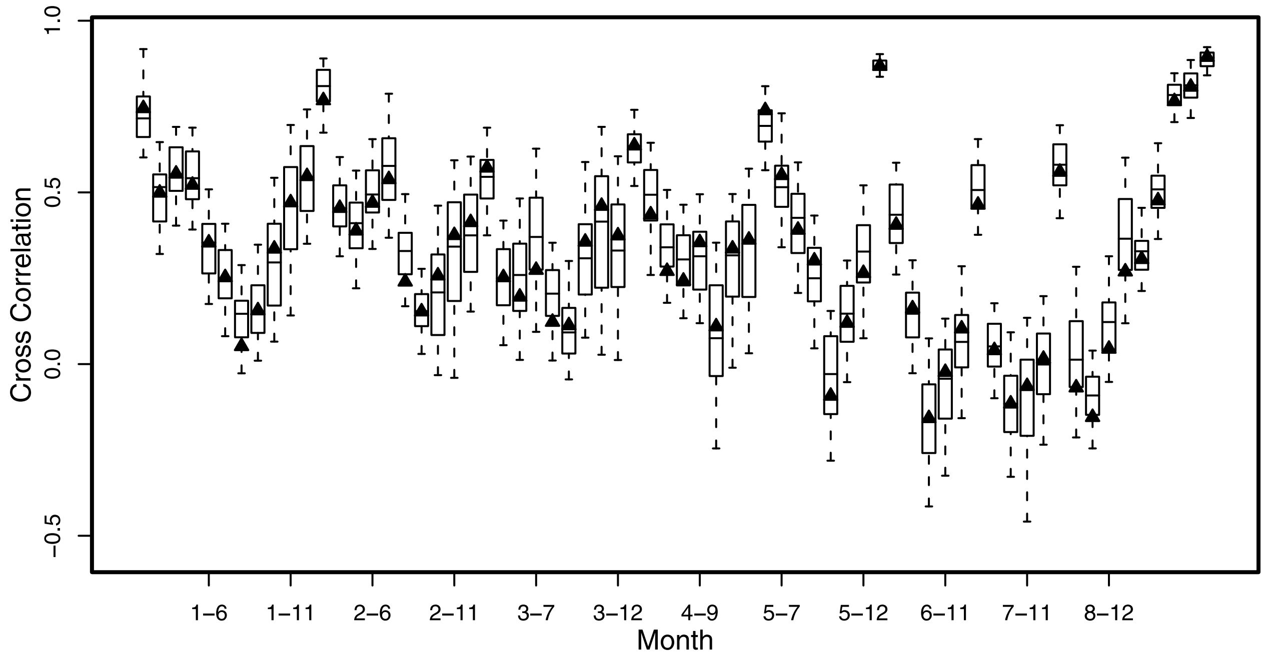

Figure 7. Monthly cross correlations for San Juan near Pagosa Springs, CO. Triangles represent values from observed data. Labels refer to months for which the cross correlation is computed (e.g., 1–6 is the cross correlation between January and June).

[26] Regarding the utility of the Prairie et al. method for daily disaggregation, this technique has a tendency to produce negative flows at the location with the smallest flow magnitude, especially during the low flow season. This issue was also seen in the work of *Lee* [2008] and further complicated when attempting to apply the genetic algorithm to daily data. However, at the monthly time scale, results were quite good and similar to those of *Prairie et al.* [2007].

4.4. Lag Structure and Additional Daily Results

[27] To examine the persistence structure, we provide daily autocorrelation function (ACF) plots for wet (May–July) and dry (January–March) seasons at the Navajo River near Chromo, CO site. For each disaggregated simulation, an average ACF is computed for the two seasons, the range of which is shown by the gray region (Figure 14). Similarly, average ACFs are computed based on the observed data and are shown as solid lines. The observed curve falls within the range of the disaggregated results for both seasons. Additionally, daily statistics for a 14 day period of the wet season are provided for the San Juan near Caraccas, CO and Navajo near Chromo, CO sites (Figures 15 and 16). The mean, variance, and skew of the observed data are well reproduced

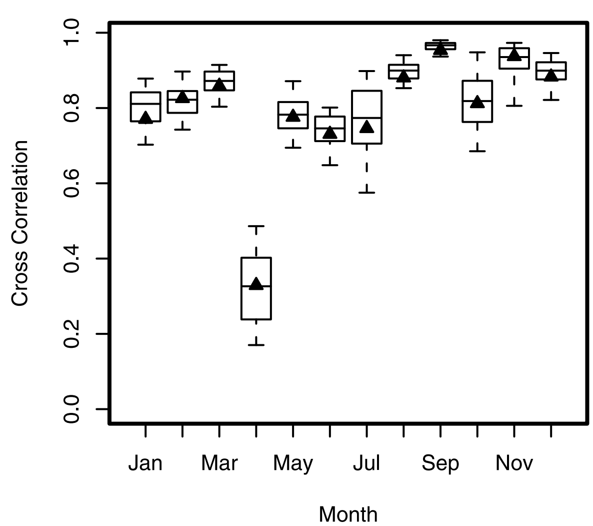

Figure 8. Monthly correlation between San Juan River near Carracas, CO, and Navajo River near Chromo, CO. Triangles represent values from observed data.

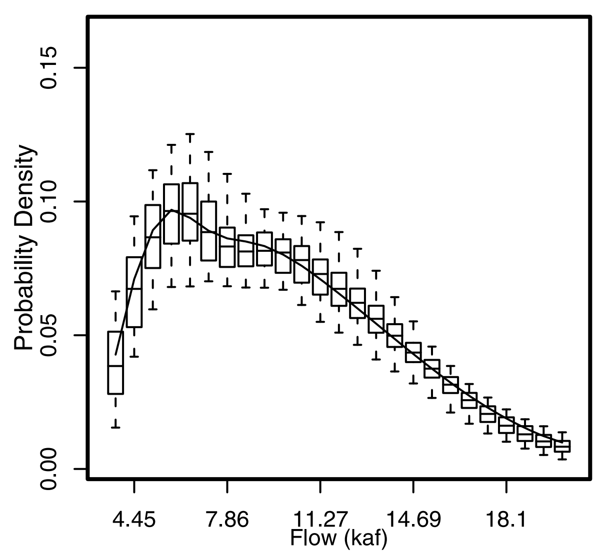

Figure 9. January Flow PDF for San Juan River near Carracas, CO. Solid line is observed data, and box plots are simulated flow.

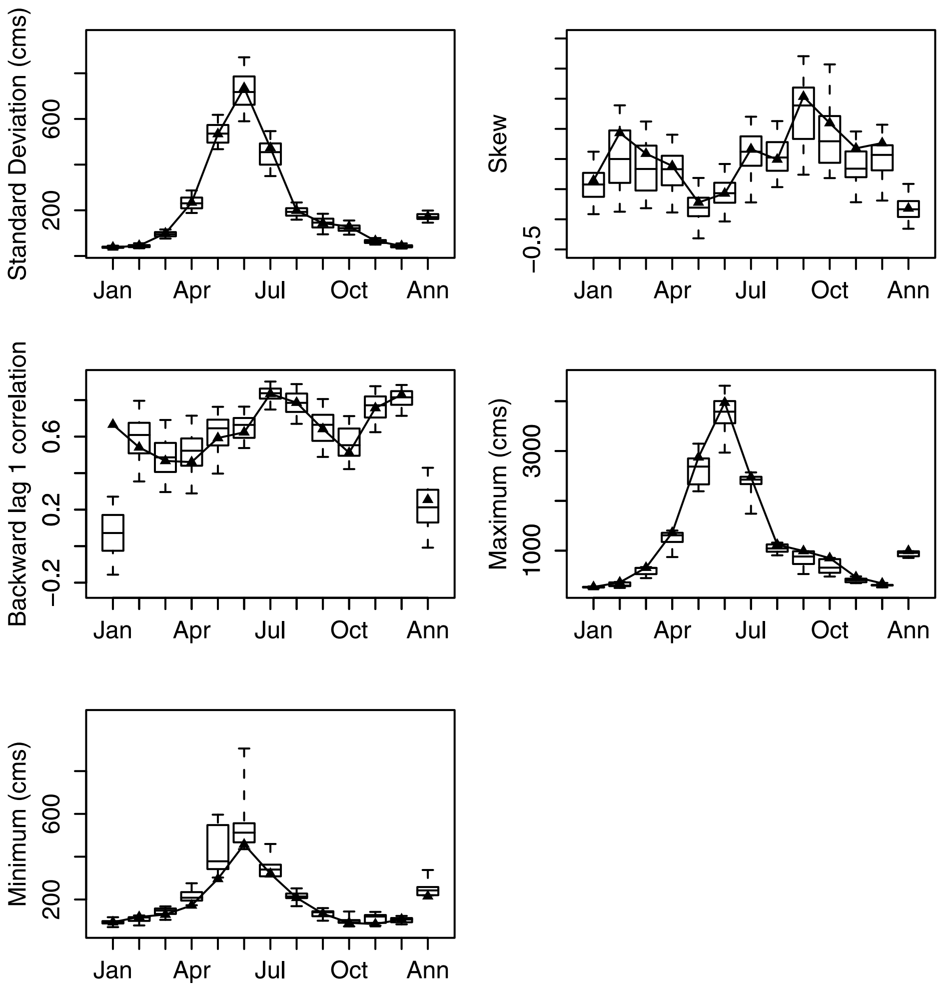

Figure 10. Colorado River at Lee's Ferry, AZ distributional statistics (12 months and annual, based on proportional disaggregation method). Triangles represent values from observed data.

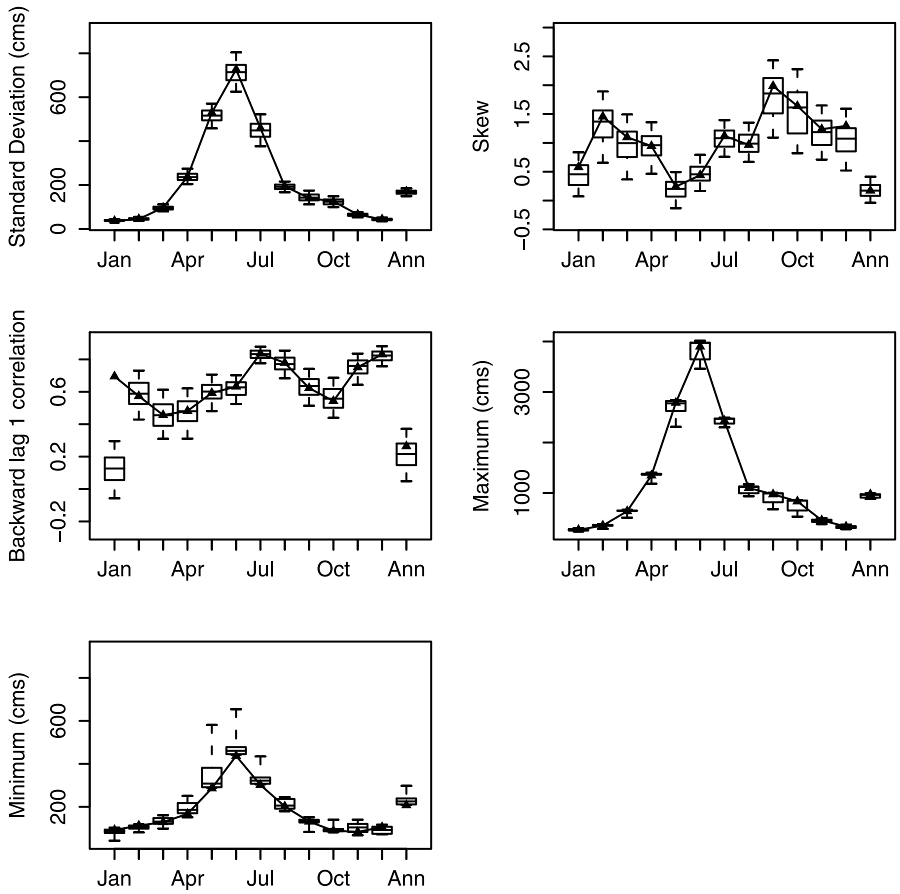

Figure 11. Colorado River at Lee's Ferry, AZ distributional statistics (12 months and annual, based on *Prairie et al.* [2007] method). Triangles represent values from observed data.

by the simulations. However, simulated maximum and minimum distributions do not capture the respective observed values.

[28] For the San Juan near Caraccas, few values are simulated below the observed minimum, but a variety extend beyond the observed maximum (Figure 15). In the case of the Navajo near Chromo (Figure 16), the opposite is seen; few values are larger than the observed maximum, yet a wealth of new minimum flows are simulated. These results were determined to be an artifact of the observed data and index gauge simulation method. For Figure 15, when new minimum flows are not produced, this is due to the smallest proportion values coming from the same year as the index gauge annual minimum flow. While this is not often the case, it can occur. This prohibits any new minimum flows from being generated when the index simulation is a resampling of the observed record (e.g., index sequential method, K-NN). However, this can easily be remedied by employing a technique to generate annual values at the index location which extend beyond the observed extremes (e.g., AR-1, etc.).

[29] Figure 17 shows the daily range of observed values for all three locations as well as the median. In the case of the Navajo River, the observed minimum and median are quite close, indicating that at least 50% of the values for a given day are similar to the value corresponding to the historic low flow value. These consistently similar magni-

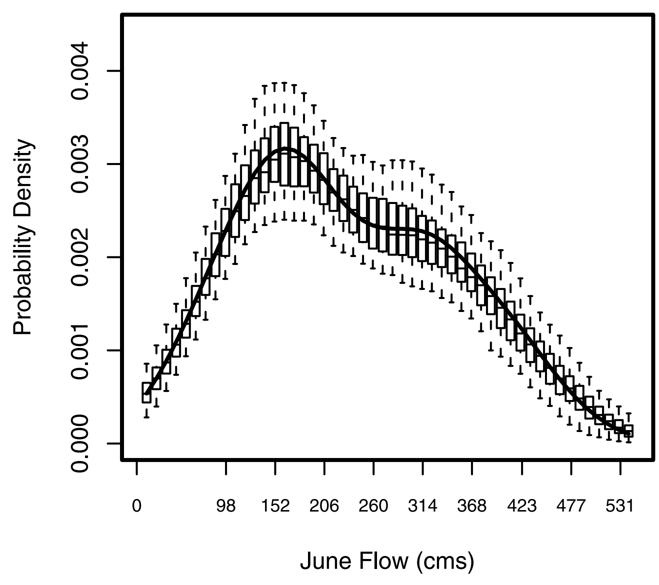

Figure 12. June flow PDF for the San Juan River near Bluff, UT (based on proportional disaggregation method). Solid line is observed data, and box plots are simulated flow.

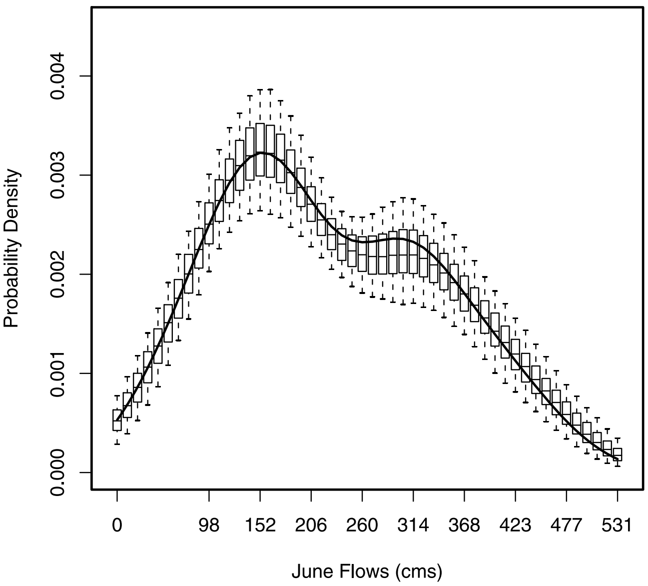

Figure 13. June flow PDF for the San Juan River near Bluff, UT (based on *Prairie et al.* [2007] method). Solid line is observed data, and box plots are simulated flow.

tudes translate to substantial variability in the proportion space; the local flow is fairly constant while the index flow changes considerably year to year (i.e., for a fixed release, the smallest proportion values are linked to the largest index flow). Therefore, there is a high propensity to resample a proportion vector that, when coupled with a below average annual flow, will generate unseen values.

[30] The ability to generate a wide range of flow magnitudes is generally considered a strength in stochastic simulation. However, a caveat exists; if there is a minimum flow requirement, this method has the potential to simulate flow values that violate such regulations. This is highlighted in Figure 16. Therefore, it is essential to emphasize the importance of understanding the basin in which flows are being disaggregated and any special needs or requirements that may exist.

5. Summary and Discussion

[31] A simple and highly adaptable disaggregation technique has been presented. The approach resamples historic flow proportion vectors conditioned on flow at an index location to produce disaggregated values that are continuous and are guaranteed to sum to the original value being disaggregated. The most novel application of the method is to produce daily flows from a simulated annual value. However, it has been demonstrated that the method is useful and effective at producing values at a variety of time scales. Distributional statistics of the historic data are reproduced in almost any space and time domain combination. Furthermore, the proportional disaggregation has the ability to generate extreme values previously unseen in the historic record, albeit with less frequency than some parametric approaches.

[32] The main drawback to this technique is one that frequently plagues nonparametric disaggregation schemes: flow continuity between the end of one year and the start of a subsequent year. *Lee* [2008] has proposed a nonparametric disaggregation resampling approach that addresses this issue. However, it is mainly effective at the monthly time scale and for sites that dominate the index gauge. Proper lag relationships and continuity are not guaranteed for upstream locations that minimally contribute to the index gauge. Additionally, at the daily disaggregation resolution, the method of *Lee* [2008] offers little aid in preserving flow continuity.

[33] Additionally, as discussed earlier, results suffer in highly regulated reaches. As such, the method is best suited for locations with limited anthropogenic impact (e.g., USGS Hydro-Climatic Data Network) or completely naturalized flows. Once disaggregated, these data can be used as inputs to a water decision support tool, enabling the proper modeling of diversions, environmental flows, hydropower, reservoir levels, etc., without having to make such considerations in stochastic methods.

[34] *Kumar et al.* [2000] successfully developed a daily disaggregation of monthly values through a kernel density estimation of the conditional PDF from which to resample, coupled with optimization to preserve continuity. The approach was demonstrated on some of the same locations within the San Juan Basin as utilized in this work. Results indicate that both have the ability to effectively disaggregate to daily values; however, a major advantage of this work is the simplicity and limited computational requirements. Thus, given the parsimony and versatility of this approach, we believe it is a significant contribution to the field of

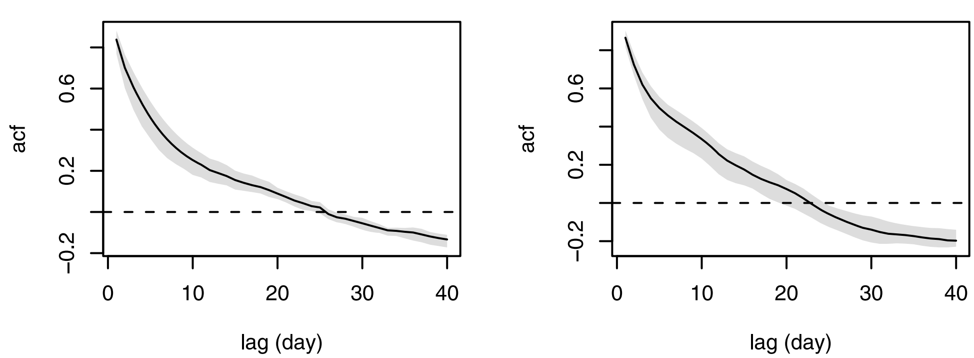

Figure 14. Navajo River near Chromo, CO, average daily autocorrelation function for (left) dry season (January–March) and (right) wet season (May–July). Gray region is range for disaggregated data (averaged over each simulation), and solid line is from observed data (averaged over period of record).

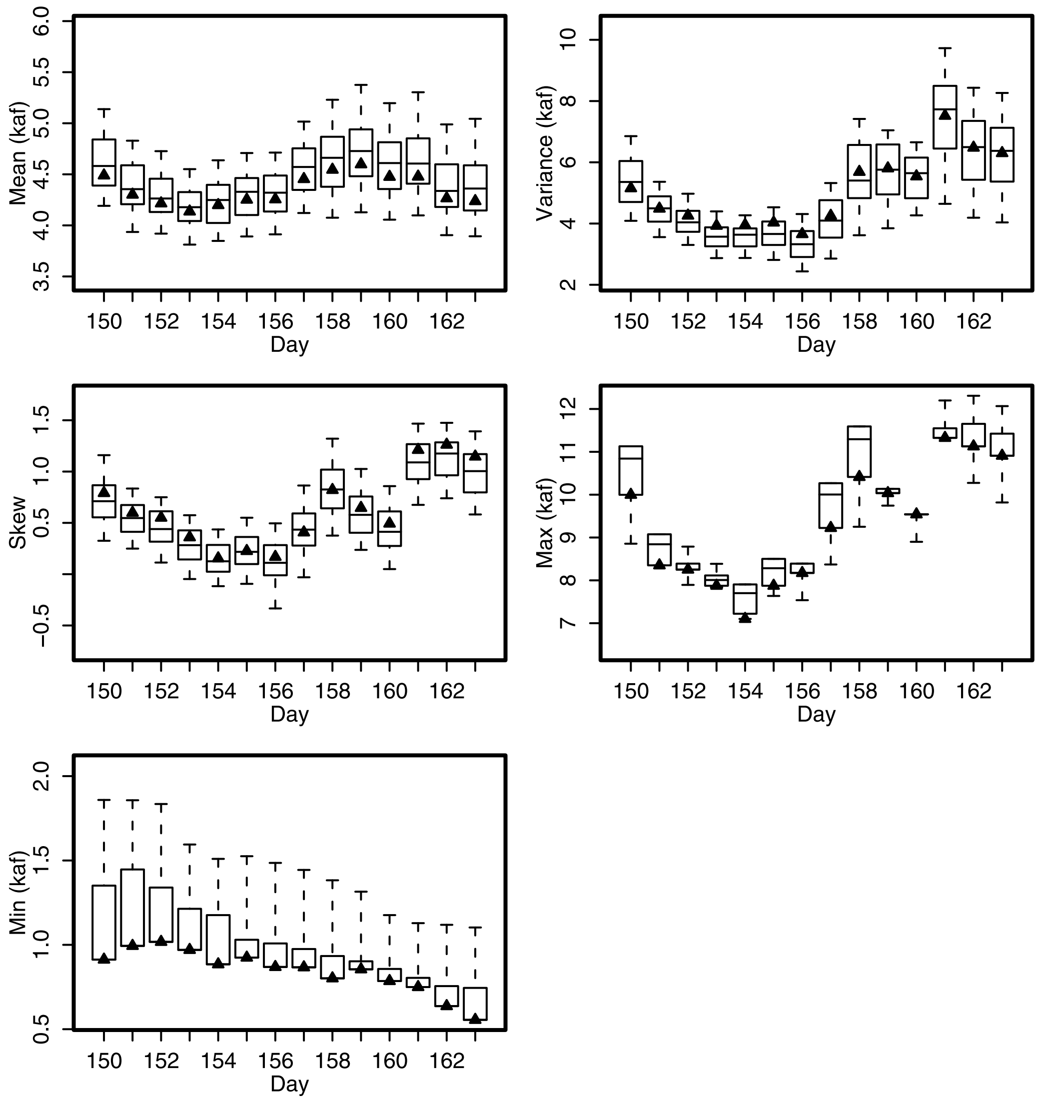

Figure 15. San Juan River near Caraccas, CO, daily statistics for the period of 30 May to 12 June. Triangles represent values from observed data.

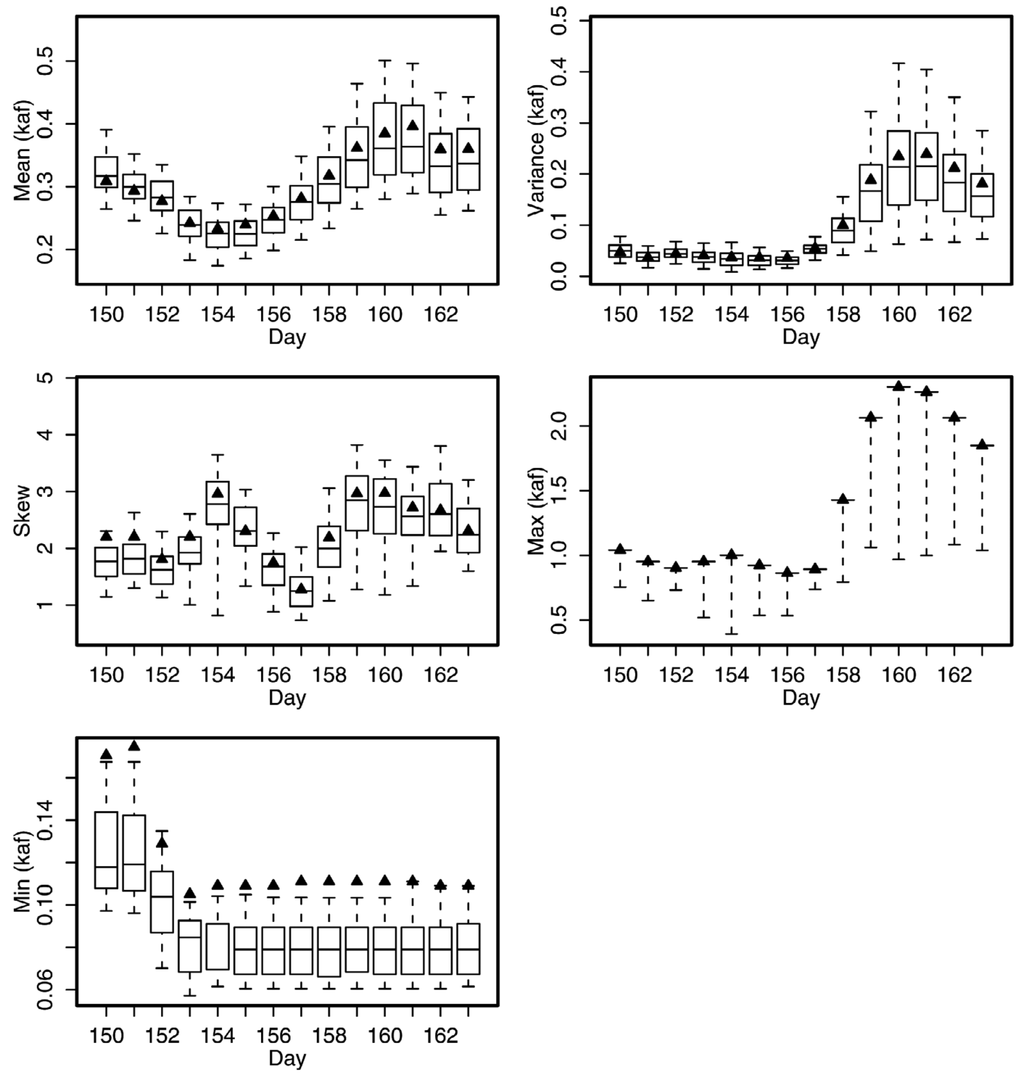

Figure 16. Navajo River near Chromo, CO, daily statistics for the period of 30 May to 12 June. Triangles represent values from observed data.

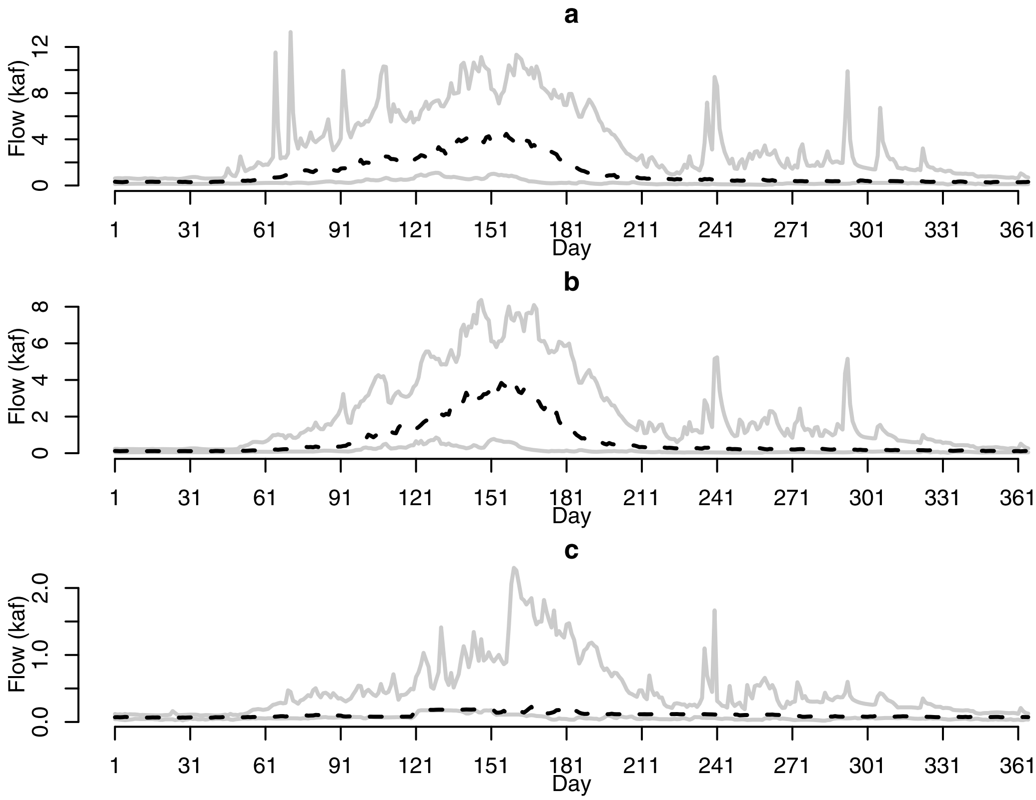

Figure 17. Daily flow ranges (solid gray) based on observed data with median values shown as dashed line for (a) San Juan River near Caraccas, CO; (b) San Juan River near Pagosa Springs, CO; and (c) Navajo River near Chromo, CO.

stochastic flow simulation and will prove quite useful as a tool in water management and planning.

[35] **Acknowledgments.** Funding for this research by Bureau of Reclamation and Western Water Assessment is gratefully acknowledged. Thanks are also due to 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

Bogardi, I., et al. (1993), Application of a space-time stochastic-model for daily precipitation using atmospheric circulation patterns, *J. Geophys. Res.*, *98*, 16,653–16,667, doi:10.1029/93JD00919.

Chang, T. J., et al. (1984), Daily precipitation modeling by discrete autoregressive moving average processes, *Water Resour. Res.*, *20*, 565–580, doi:10.1029/WR020i005p00565.

Chin, E. H. (1977), Modeling daily precipitation occurrence process with Markov-Chain, *Water Resour. Res.*, *13*, 949–956, doi:10.1029/WR013i006p00949.

Grygier, J. C., and J. R. Stedinger (1988), Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resour. Res.*, *24*, 1574–1584, doi:10.1029/WR024i010p01574.

Kumar, D. N., et al. (2000), Multisite disaggregation of monthly to daily streamflow, *Water Resour. Res.*, *36*, 1823–1833, doi:10.1029/2000WR900049.

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resour. Res.*, *32*, 679–693, doi:10.1029/95WR02966.

Lall, U., et al. (1996), A nonparametric wet/dry spell model for resampling daily precipitation, *Water Resour. Res.*, *32*, 2803–2823, doi:10.1029/96WR00565.

Lee, T. (2008), *Stochastic Simulation of Hydrologic Data Based on Nonparametric Approaches*, Colorado State University, Fort Collins.

Prairie, J., et al. (2007), A stochastic nonparametric technique for space-time disaggregation of streamflows, *Water Resour. Res.*, *43*, W03432, doi:10.1029/2005WR004721.

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

Srikanthan, R., and T. A. McMahon (2001), Stochastic generation of annual, monthly and daily climate data: A review, *Hydrol. Earth Syst. Sci.*, *5*, 653–670.

Stedinger, J. R., and R. M. Vogel (1984), Disaggregation procedures for generating serially correlated flow vectors, *Water Resour. Res.*, *20*, 47–56, doi:10.1029/WR020i001p00047.

Tarboton, D. G., et al. (1998), Disaggregation procedures for stochastic hydrology based on nonparametric density estimation, *Water Resour. Res.*, *34*, 107–119, doi:10.1029/97WR02429.

Valencia, D., and J. C. Schaake (1973), Disaggregation processes in stochastic hydrology, *Water Resour. Res.*, *9*, 580–585, doi:10.1029/WR009i003p00580.

U. Lall, Earth and Environmental Engineering, Columbia University, 918 S. W. Mudd Hall, Mail Code 4711, 500 West 120th St., New York, NY 10027, USA.

K. Nowak and B. Rajagopalan, Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO 80309-0421, USA. (kenneth.nowak@colorado.edu)

J. Prairie, Bureau of Reclamation, Upper Colorado Region, University of Colorado, 421 UCB, Boulder 80309-0421, CO, USA.