

A simple method for streamflow disaggregation

Anil Acharya¹ and Jae H. Ryu^{*2}

*Corresponding Author

Submitted to: Journal of Hydrologic Engineering, June 05, 2012

Revised September 25, 2012

Resubmitted to JHE, February, 5, 2013

Accepted Manuscript
Not Copyedited

¹ Postdoctoral Research Associate, Department of Biological and Agricultural Engineering, University of Idaho, 322 E. Front St., Boise, Idaho 83702. Currently, Department of Civil and Mechanical Engineering, Alabama A and M University, 4900 Meridian Street North, Huntsville, AL 35762

² Assistant Professor, Department of Biological and Agricultural Engineering, University of Idaho, 322 E. Front St., Boise, Idaho 83702. (Email/Ryu: jryu@uidaho.edu).

Abstract

Streamflow disaggregation from monthly to daily was performed using a relatively simple, flexible, and adaptive method. Only streamflow acts as decision variables in this disaggregation process. To disaggregate monthly to daily flow at the target station (TS), monthly counterparts at the source station (SS) were selected based on minimum error criteria which are calculated with respect to streamflow volume within three-month time window. Daily streamflow indexes at SS were then calculated to disaggregate monthly to daily streamflow at TS during the disaggregation process. The effectiveness of the proposed method has been demonstrated through its application at both regulated and unregulated waterways located in the northwest states, including Idaho and Wyoming. For both regulated and unregulated monthly streamflow, the proposed method well represents daily streamflow realizations similar to historical flows and preserves both mass balance and a series of statistical characteristics. However, the results also indicate that the quality of disaggregated daily streamflow varies for individual applications depending upon the selection of stations, their geographic information, and data availability. The disaggregation model used in this research is transparent, user friendly, less intensive, and less time consuming so that it can be utilized at any watershed without difficulty or much effort. Consequently, since development and availability of daily streamflow is important for water resources planning and management, including reservoir operation, water quality study, and environmental/ecological modeling, this research will help bridge the gap among interdisciplinary water research activities, especially for studies of the impacts of hydrologic events possibly driven by extreme weather variability and climate change.

Keywords: streamflow disaggregation; monthly to daily streamflow; flow index; water resources management; climate change.

1 Introduction

Streamflow disaggregation at shorter (e.g. daily) time scales is an important research avenue in water resources management and practices, including reservoir operations, water quality study, and biological modeling. Historically, water scientists commonly utilize monthly data to identify hydrologic variability and environmental impacts on many regulated and unregulated waterways. Understanding of river ecology, including ecological health of river system, level of biodiversity, biological response to instreamflow events, and ecology-hydrology nexus is mainly constrained by the lack of fine-scale temporal variability in hydrological datasets (Gippel 2001; Costelloe et al. 2005).

The availability of hydrological data at shorter time scales helps to characterize the ecological responses driven by consecutive flow events, which act as critical step functions limiting healthy fish habitat and survival. However, daily and shorter time steps in environmental/ecological modeling is still challenging in the sense that the data quality varies depending on geographical boundary conditions and techniques utilized by individual modelers during the disaggregation process. Numerous approaches have been applied for streamflow disaggregation at single and multiple sites, while a specific application varies depending upon hydrological characteristics embedded in the physical and computational constraints. For example, the target station (TS) streamflow, where streamflow needs to be generated or disaggregated, is dependent upon streamflow at the source station (SS). Thus, the statistical (e.g. mean, standard deviation, correlation coefficient, and skewness) and other dependent properties of the resulting streamflow should lay out characteristics of the observed historic streamflow (Sharma et al. 1997). Several disaggregation models (methods) have been developed and utilized in the past to produce streamflow realizations at finer temporal scales, such as daily time steps.

Two typical approaches, namely parametric and non-parametric models, have been widely used in literature to accomplish the goal of stochastic disaggregation and produce synthetic streamflow to mimic realizations of historic hydrologic behavior that can be statistically justified. Parametric methods impose linearity or distributional assumptions over historical data, while non-parametric methods can capture state dependency, non-linearity, and multi-modality within the historical datasets. Parametric approaches (Valencia and Schaake 1973; Stedinger and Vogel 1984; Grygier and Stedinger 1988) have been performed following a linear approach among which autoregressive moving average (ARMA) techniques were widely used by hydrologists for stochastic hydrologic modeling based on annual and/or monthly streamflow at single sites (Pegram et al. 1980; Salas et al. 1980; Stedinger and Vogel 1984; Bras and Rodriguez-Iturbe 1985). While preserving statistical properties, such as cross correlation and summability, this method assumes that the data is normally distributed. However, this method was suitable for only lower dimensions so that the summability criterion is no longer valid for other data transformation such as log or power transform (Nowak et al. 2010). Consequently, it is challenging to reproduce the nature of streamflow by any of the commonly used theoretical distributions (Sharma et al. 1997). The stepwise disaggregation approach proposed by Santos and Salas (1992) was also considered suitable for only coarser temporal scales (e.g. monthly, seasonal) because of additional challenges of coherency across the simulated flows between months or seasons during the transition time. Most of the methods discussed above are considered suitable to disaggregate data at shorter time scales not finer than monthly (Nowak et al. 2010).

Several non-parametric techniques (Sharma et al. 1997; Xu et al. 2001; Tarboton et al. 1998; Prairie et al. 2007; Lee et al. 2010) have been also developed to generate streamflow

realizations. These techniques are considered to be capable of reproducing linear and non-linear dependence (Sharma et al. 1997) and non-normal features, such as non-Gaussian distributions (Prairie et al. 2007; Tarboton et al. 1998). The non-parametric approaches utilized by earlier studies include Kernel density estimation approach (Tarboton et al. 1998) and K-nearest neighbor (KNN)-based time resampling approach (Prairie et al. 2007; Nowak et al. 2010). Although the kernel density methods address non-normality issues, they require intensive mathematical computations. They were considered inefficient due to the complexity of matrix processes (higher dimensions) to deal with statistical inference during parameter estimation, the difficulty of implementing in multivariate problems, and boundary conditions (Sharma and O'Neil 2002; Nowak et al. 2010). Meanwhile, the KNN approach proposed by Nowak et al (2010) was considered effective with respect to reproducing daily data at multiple sites, and preserving summability, continuity, and cross dependency amongst stations. However, this method does not capture the flow continuity in year-to-year transition and is more suitable at unregulated streamflow stations. The challenge to capture flow continuity between monthly or yearly transitions in non-parametric approaches is similar to the parametric approaches discussed earlier. Most of the studies discussed above were found effective only at coarser time scales (i.e., from annual or seasonal to monthly).

Some stochastic models were developed to disaggregate streamflow from monthly to daily but each method has pros and cons in terms of accuracy and limitations. Though it is considered an improvement, the disaggregation approach developed by Nagesh Kumar et al. (2000) was computationally intensive with a larger number of decision variables. Techniques, such as Markov autocorrelation pulse model (Xu et al. 2001) and cross-correlation pulse model (Xu et al. 2003) might be able to reproduce statistically similar characteristics of the historical

daily streamflow sequences. These models show better performances for streamflow disaggregation at some watersheds, but simulating extreme hydrologic events is still limited for broad applications.

Some deterministic approaches are also used to simulate daily streamflow along with limited applications. Green (1973) noted that daily streamflow generated by daily rainfall using a deterministic approach had some success in modeling streamflow, especially from small urban catchments. A series of statistic techniques, including logarithmic transformation of 5-day average flows and a stochastic error term associated with maximum and minimum daily flows, were incorporated to preserve the non-deterministic nature of daily flows (hydrograph shape) and long-term flows (statistically stabilized characteristics in a disaggregation model for daily streamflow) (Green 1973). However, it did not show a similar accuracy for all river segments nor it was feasible to simulate high flow realizations. Another disaggregation method proposed by Ganju et al. (2008) is based on the selection of a historical monthly hydrograph that is closest to the monthly hydrograph that will be subject to disaggregation. Selection of historical monthly hydrograph is done by considering least square error (metrics), and daily streamflow are then generated based on the selected monthly hydrograph. This approach presumed that the monthly hydrograph throughout the year also mimics the behavior of the daily unimpaired hydrograph after disaggregation processes. Sivakumar et al. (2004) developed a non-linear deterministic approach to generate sequences of daily streamflow by considering streamflow data series of successively doubled time resolutions between daily and 16 days (e.g., daily, 2-, 4-, 8-, and 16-day resolutions). Their method was performed between successive resolutions where best results were obtained for “low embedding dimensions (2 or 3)” and the accuracy was increased as time scales change from coarser to finer (16 days to 2 days) (Sivakumar et al. 2004).

Based on the literature review, we notice that many algorithms, techniques, and methodological approaches are available to generate streamflow at different time scales. The previous methods have been proposed at different times with certain modifications and improvements. Disaggregation of streamflow from monthly to daily time steps, however, is still challenging hydrologists face today. As Nowak et al. (2010) pointed out, the major constraints during the disaggregation processes from monthly to daily include intensive computational resources, high dimensionality of the disaggregation problem, degree of feasibility to meet targets (deterministic or stochastic, parametric or non-parametric), and the uncertainty embedded in estimating parameters. Although some of the latest methods are utilized in streamflow disaggregation at a shorter time scale, their conceptual difficulty and computational complexity still limit our ability to disaggregate streamflow from monthly to daily. Note that a direct comparison between previous methods may not be practical even if those are utilized in many water resources applications at different time scales. Some methods are too difficult to implement because of their complexity, while some others are still in the process of validating accuracy of daily streamflow at specific applications. Therefore, a relatively simple approach that could be implementable easily and quickly in many regulated and unregulated waterways is of great interest to water resources managers.

As such, the goal of this research is to develop a simple, transparent, user friendly and computationally less intensive statistical method that can be utilized without much difficulty to disaggregate streamflow from monthly to daily. This paper is organized as follows: a brief methodology for this analysis is provided in the next section followed by the description of the study area. Then, the summary of observed and simulated results along with case studies is discussed. The conclusions from this research and future research scope are finally described.

2 Methodology

2.1 Disaggregation Process and Statistics

The disaggregation process developed here is a simple linear deterministic technique to disaggregate monthly to daily streamflow at TS based on the selected monthly streamflow at SS using minimum error criteria. Thus, the minimum error is calculated based on a three-month centered time window to capture seasonal flow volume as adopted by Nagesh Kumar et al. (2000). The purpose of the 3-month window is to reduce uncertainty in the continuity of flows amongst embedded months. Since proportional adjustments are often needed to compute daily index during the disaggregation process, this approach is capable of generating extreme values that were previously not observed in the historical record.

Wood et al. (2004) also utilized similar techniques, but random sampling and temporal downscaling efforts has been additionally added to bias-corrected monthly precipitation and temperature data to daily. Unlike the use of random sampling and single month as adopted by Wood et al. (2004), a 3-month window in this study is considered for the month that needs disaggregation. Note that selection of historical monthly streamflow counterparts for the same window is based on relative error criteria, which is measured as Root Mean Square Error (RMSE) (see Appendix for equations). A complete streamflow disaggregation process is summarized in Figure 1.

The first step in the disaggregation process is the selection of Target Station (TS) and Source Station (SS). Since SS is utilized as a reference station, its selection is critical for streamflow disaggregation at TS because the ratio of daily streamflow to monthly flow at TS is solely derived from flow statistics at SS. Therefore, the SS located near TS and possesses similar watershed characteristics (e.g., vegetation, land use change, soil type) would be a good

choice. If multiple stations exist nearby, a K-nearest neighbor based resampling approach is recommended. Note that all unregulated gauge stations used in this study are located in snow-dominated watersheds (See Figure 2).

The SS includes both daily and monthly historical streamflow, while the TS has only monthly streamflow. Basically, the daily streamflow at TS is calculated based on monthly streamflow at TS and daily and monthly streamflow at SS. Typically, the SS can be determined based on the quality of data and no missing observation.

Once a month and year is selected for streamflow disaggregation at TS, a 3-month window centering the selected month at TS is chosen from historic observed monthly streamflow at SS. For example, streamflow disaggregation in April at TS requires information for a 3-month window, including March, April, and May over historic observed monthly streamflow at SS. Thus, the total monthly streamflow for this window at TS is compared with the same window for every year in the historical record available at SS. Next, the RMSE is computed to identify the best match year and months representing monthly volume between TS and SS during the predefined 3-months. Daily flow index, SI_{SS} is then computed as follows. SI_{SS} is defined as the ratio of daily streamflow (D_y) to monthly total streamflow (M_y) for the selected month at SS. Note that the data utilized in this study have no missing values. A detailed description of methods to reconcile missing observations at stations is beyond the scope of this paper and therefore not included here. The interested readers are referred to related literature (Little 1992; Allison 2001; Cohen et al. 2003; Howell 2008) for detailed explanations on gap filling techniques.

A cubic spline interpolation is applied as a post processing tool after performing the disaggregation process, but it is optional (see Figure 1). Thus, cubic spline interpolation

technique can be applied when streamflow difference between the last day of the month and the first day of the following month is greater than 10%, which is a user defined threshold based on engineering judgment. It appears that 10% threshold based on sampling is acceptable in the sense that total volume in the months is similar before and after cubic spline interpolation.

The major application of cubic spline interpolation in this analysis is to minimize hydrologic jumps in month-to-month transition. For this purpose, the disaggregated daily streamflow during the last week (7 days) of the selected month and first week (7 days) of the following month are subjected to cubic interpolation by considering disaggregated daily streamflow starting from the middle of the month to the middle of the following month (about 30 days' time horizon). For example, to minimize hydrologic jumps in between June and July 1967, a time window, June 24-30 and July 1-7, was selected to update streamflow by the interpolated streamflow from the spline. Since the length of interpolated streamflow is less than quarter percent of two-month time window (14 out of 60 days data records), the monthly flows (sum of daily flows) before and after smoothing were not significantly different, which means that the spline maintains the summability relatively well in this particular case. Perhaps, the cubic splines produce negative values during low flow conditions, but it doesn't apply to our cases. If that is the case, conditional spline techniques, such as conditional logspline (Mâsse and Truong, 1999) can be applied.

2.2 Statistics

Various statistical parameters, including mean (\bar{Y}), standard deviation (S_y), RMSE, and the bias (%) between observed and simulated hydrographs are computed to evaluate the effectiveness of the proposed disaggregation method. Additionally, the Pearson correlation

coefficient (r), coefficient of determination (R^2), and Nash-Sutcliffe efficiency (E) are also calculated for validation purposes. These statistics are calculated between observed and disaggregated daily streamflow over the study period and for selected peaks representing maximum, normal, and low (base) flow conditions during the same period. Ratio (R) is also calculated for each selected peak to determine the difference in observed and disaggregated volume. Mathematical justification of statistical parameters is available in Appendix.

A two-sample Kolmogorov-Smirnov (KS) non-parametric goodness of fit test is also applied to test if two streamflow distributions (observed vs. disaggregated) are significantly different from each other. This test calculates maximum cumulative distance in between two streamflow distributions (Stephens 1970; Acharya et al. 2012).

$$D = \max |D1 - D2| \quad (1)$$

Where, $D1$ and $D2$ represent data vectors for each streamflow distribution, respectively. The null hypothesis for this test is that the two data vectors are from same empirical distribution; the alternative hypothesis is that the two data vectors are from different distributions. This test rejects the null hypothesis, if the critical value based on sample size is less than the test value based on maximum cumulative distance.

3 Study Area

For this study, three pairs of streamflow gauge stations maintained by the United States Geological Survey (USGS) are selected to represent one regulated and two unregulated streamflow sites. As shown in Figure 2, the regulated (impaired) streamflow locations are selected at the Boise River, Idaho, while the unregulated (unimpaired) streamflow locations at the Salt River and Upper Snake River are situated near the border between Idaho (ID) and

Wyoming (WY). The gauge stations selected along the Boise River are near Parma (USGS 13213000, hereafter Sta A1) and at Glenwood Bridge near Boise (USGS 13206000, hereafter Sta A2). The Boise River is regulated by a series of dams and reservoirs upstream of the selected stations. The selected gauge stations on Snake River (USGS 13010065, hereafter Sta B1) and Salt River (USGS 13027500, hereafter Sta B2) in Wyoming are considered unimpaired because they are located upstream of Jackson Lakes Dam at Flagg Ranch and Palisades Dam near Etna, respectively. No diversions are made upstream of their locations. Another set of unimpaired gauge stations are located at Falls River near Squirrel (USGS 13047500, hereafter Sta C1) and Teton River above South Leigh Creek near Driggs (USGS 13052200, hereafter Sta C2) in Idaho.

As mentioned earlier, streamflow is a critical input requirement to evaluate system performance and/or additional hydrological (e.g., low flow, flood control, water management, river restoration) and ecological (e.g., temperature, water quality, biological index) modeling efforts. Both daily and monthly streamflow data for each selected stations are obtained from the USGS website (<http://waterdata.usgs.gov>). More detailed information for the selected stations, such as their locations and related flow characteristics is summarized in Table 1.

4 Case Study 1: Regulated Streamflow

Since the Boise River is regulated and the stations are located within the same watershed, characteristics of monthly flows at two regulated streamflow stations show similar patterns as shown in Figure 3. Both observed and disaggregated daily streamflow agree well in the sense that the nature of the disaggregated hydrograph, such as base streamflow and peak flows is well captured for the regulated waterway. It appears that the disaggregated peak flows are slightly higher than observed peak flows as shown in Figure 4 during 1996-2000. The calculated

statistics between Sta A1 and Sta A2 for the regulated case are also listed in Figure 4. A very low (negligible) positive bias (%) and higher correlation coefficient (r), coefficient of determination (R^2), and Nash-Sutcliffe efficiency (E) values are observed. Note that higher ' r ', ' R^2 ' and ' E ' represents a good match between observed and disaggregated streamflow volume, while positive bias indicate that disaggregated streamflow are higher than observed streamflow. Perhaps, a slightly higher disaggregated streamflow volume at TS is derived from higher magnitudes of streamflow characteristics at SS. Although the observed and disaggregated mean daily streamflow for each month coincide, the standard deviations during Jan-June are comparatively higher than that during July-Dec (Table 2). This is due to relatively higher daily streamflow observed during Jan-July in the study period (Figure 4).

Figure 5 further illustrates the comparison of observed and disaggregated daily streamflow for some selected peaks for regulated waterways in Boise River system during 1996-2000. Correlation coefficient, observed and disaggregated streamflow volume and their ratio, and RMSE are calculated for each selected peak for evaluations. Each of these cases shows a good correlation (0.55 to 0.95), except Peaks 1, 10, and 15, which show a correlation that may not be practically significant (< 0.40). The observed and disaggregated streamflow volume also matches quite well, while the ratio varies from 0.94 to 1.00. During this calculation, the RMSE varies from 0.5cms (18cfs) to 11cms (388cfs), higher RMSE is observed for high flows at Peaks 2 and 4, and lower RMSE for low flows at Peaks 13 and 14. A comparatively higher RMSE is observed for high flows, since a small difference in high flows may cause larger errors.

The range of observed and disaggregated daily streamflow is also compared in a box plot as shown in Figure 6. As expected, the range of maximum and minimum, interquartile as well as median streamflow are approximately well balanced for each month, except for Feb, where a

lower median and 3rd quartile range is observed for disaggregated daily streamflow. The outliers are also observed with similar magnitudes. During this analysis, both observed and disaggregated monthly total streamflow, which are calculated based on daily streamflow, are also well matched by verifying the mass balance for each month.

5 Case Study 2: Unregulated Streamflow

For unregulated streamflow at the Upper Snake River, the gauge stations, Sta B1 and Sta B2, are selected as shown in Table 1 and Figure 2. Sta B1 near Jackson Lake and Sta B2 near Palisades Reservoir are considered as SS and TS, respectively. Both of these stations are located in snow-dominated watersheds and show similar nature of hydrograph as shown in Figure 7. Similar to the regulated case study, the observed monthly streamflow volume at SS is higher than that at TS.

The disaggregation tool is utilized to disaggregate five years (1996-2000) of monthly streamflow at Sta B2 (Target Station) based on 22 years (1984-2005) of daily and monthly streamflow at Sta B1 (Source Station). In addition to the disaggregation procedure described above for the regulated case study, a cubic spline interpolation is also applied as a post processing tool for unregulated flows due to 10% threshold criterion discussed earlier. Thus, it requires maintaining coherency of hydrographs between months. As shown in Figure 8, the comparison of observed and disaggregated daily streamflow shows similar hydrologic response throughout the study period, except some abnormal peaks identified (See the dotted box in Figure 8) during the periods of April-June in 1998. But, in most cases, the nature of hydrograph, base streamflow, and peak streamflow is well captured. It appears that the disaggregated peaks

are slightly higher than observed peaks at lower flow conditions because streamflow at SS often override streamflow at TS during the disaggregation process.

The calculated statistics based on observed and disaggregated daily streamflow during 1996-2000 for unregulated streamflow in the Upper Snake River are available in Figure 8. It is noted that a slightly lower disaggregated streamflow volume at TS is affected by sudden hydrological jumps observed during high or low flow conditions. While comparing mean and standard deviation of daily streamflow for each month, the mean daily streamflow are in the same range, except April, which is showing higher difference. In general, the standard deviations during April-Sep are higher in comparison with Oct-March over the study period. Note that higher standard deviation is observed during high flows conditions.

The computed statistics for an unregulated case show higher bias (%) and RMSE, and lower 'r', 'R²' and 'E' in comparison with the regulated case, perhaps due to different hydrologic characteristic related to local information of stream gauge stations, catchment scales, and observed streamflow. For example, unlike the regulated gauge stations in the Boise River, the unregulated gauge stations located in the Salt River and Snake River, which have different catchment characteristics, have resulted very high streamflow at SS in the Snake River as compared to the TS in the Salt River. This implies that the increased variation of streamflow at SS results in increased variation of streamflow indices, representing more frequent fluctuations in higher or lower daily streamflow at TS.

Boxplots in Figure 9 show the median, interquartile range, as well as maximum and minimum daily streamflow for each month at the unregulated waterways during 1996-2000. The interquartile range and median daily streamflow are approximately in the same range for all months except April, where median daily disaggregated streamflow is lower than observed

streamflow. Minimum and maximum streamflow also show similar range for each month except April, where disaggregated maximum streamflow along with outliers show higher range of streamflow as compared to observed streamflow. The higher disaggregated streamflow calculated for this month is due to the higher ratio (SI_{SS}) developed with respect to higher streamflow at SS. As similar to the regulated case, while comparing monthly total volume aggregated from daily streamflow, the disaggregated streamflow agree well with the observed streamflow, in terms of both magnitude and patterns. As noted earlier, this has become possible since the method follows proportional distribution, and the ratios (SI_{SS}) calculated from SS are applied to total monthly volume at TS, thus resulting in approximately equal monthly streamflow before and after disaggregation.

Additionally, some peaks representing high, normal, and low flows are selected from the disaggregated streamflow, and then statistics are calculated. The selected peak events, their time frame, and total volume for each event are summarized in Table 3. No significant correlation is observed for selected peaks based on high flows, while it is calculated in the range of 0.50-0.90 and 0.40-0.71 based on normal and low flows respectively. It is noted that the lower correlation calculated for selected peak flows are due to difficulty in exactly representing daily streamflow variations captured during SI_{SS} calculations. The variation in SI_{SS} associated with differences in watershed characteristics, station location, and streamflow pattern (wet/dry) at SS might be a dominant factor embedded in disaggregation logics. Nonetheless, the disaggregated daily streamflow over the study period generally provide good statistics as shown in Table 2 and Figure 8. The RMSE of the selected peaks for high flows and low flows are observed high values and low values, respectively. Although, RMSE is higher for some selected peaks for high

flows, the observed and disaggregated streamflow volume for each peak matches well for all cases, where the ratio varies from 0.90 to 1.14.

To further illustrate the disaggregation method at the unregulated gauge stations, an additional pair of unregulated streamflow stations located in Eastern Idaho is investigated. The selected gauge stations are namely Sta C1 in the Falls River near Squirrel and Sta C2 in the Teton River near Driggs (See Table 1 and Figure 2). Figure 10 shows observed vs disaggregated daily streamflow at Sta C2 during the period 1996-2000. As expected, the disaggregated daily streamflow follows the nature of hydrograph as well as streamflow magnitude throughout the study period, with underestimation/overestimation of some peaks. The calculated statistics over the study period can be considered good for this analysis. A negligible positive bias, higher correlation and efficiency ($r = 0.91$, $R^2 = 0.82$, $E = 0.81$) are obtained with a total RMSE of 5.5cms (195cfs) during the disaggregation process. However, after application of cubic spline interpolation, a negative bias (-1.8%), and increased correlation and efficiency ($r = 0.93$, $R^2 = 0.85$, $E = 0.84$) are obtained along with decreased error (RMSE 5cms/175cfs). It appears that the application of cubic spline interpolation results in a smooth hydrograph with reduction of peak flows (sudden jumps or drops) during the transition between months.

The K-S test and ranksum test are applied for each case study to determine if two streamflow distributions (observed vs disaggregated) are different with statistical significance. During the K-S test, the premise of our null hypothesis is that if two hydrographs from observed and disaggregated flows are substantially different from each other, we can infer, our disaggregated flows are not well matched with the observed flows at 5% level of significance. Thus, the null hypothesis indicates that cumulative distribution function (cdf) of the observed streamflow is larger than that of the disaggregated streamflow; the alternative hypothesis is that

the observed cdf is smaller than the disaggregated cdf. The null hypothesis is accepted (p value > 0.05) for the regulated case thus demonstrating observed streamflow distribution to be larger than disaggregated distribution, while rejecting null hypothesis (p value < 0.05), the case is reverse for unregulated cases.

The two sided ranksum test (Hollander and Wolfe 1999) is also applied between observed and disaggregated daily streamflow. For larger sample sizes, this test is based on the normal approximation, $[z = (T - \mu) / \sigma]$, with z statistics calculated based on test statistic (T), mean (μ) and standard deviation (σ) of T . For this test, the null hypothesis indicates that the two data vectors are from identical distributions and with equal medians; the alternative hypothesis indicates that the two data vectors do not have equal medians. This test is applied on observed and disaggregated daily streamflow for each month (Jan-Dec) to test statistical significance of median values during the study period. As expected, the null hypothesis is accepted ($p > 0.05$) for both regulated and unregulated cases for all months except April for unregulated cases. This demonstrates that observed and disaggregated streamflow are well matched over the study period.

6 Conclusions

This paper introduced a relatively simple statistical method to disaggregate streamflow from monthly to daily time scale, perhaps easily implementable in a user-friendly computational environment, such as Microsoft Excel or similar platform. A major constraint for this method, however, is the selection of SS and data availability of daily and monthly streamflow at SS. Also, there is a higher probability that results are slightly biased based on selection of SS and likely affected by missing observations. Missing observations can be easily seen in a basin due to anthropogenic impacts as well as measurement errors caused by instrumental modification,

man-made, physical constraints, and/or weather conditions. Several treatment methods from simple to complex (e.g. mean and median substitution, pairwise deletion, list-wise deletion, regression models, maximum likelihood, multiple imputation) have been proposed in the past so that the interested reader to deal with missing observations prior to disaggregation processes is referred to the literature (Little 1992; Allison 2001; Cohen et al. 2003; Howell 2008).

The suitability of this disaggregation method has been demonstrated through its application at regulated and unregulated waterways in the western watersheds. Visual inspections as well as statistical parameters were applied to evaluate the hydrographs of the disaggregated and observed streamflow. Overall, the magnitudes and patterns of the disaggregated streamflow correlate well with the observed streamflow for all cases over the study period. The statistics show that disaggregation for the regulated case better performed than that for the unregulated cases. It appears that the calculated RMSE varied depending upon streamflow at SS. Thus, the higher streamflow have resulted an increased RMSE. The K-S test also showed that cdf of the observed streamflow is wider than that of the disaggregated streamflow for the regulated case, while this pattern was shown in reverse for the unregulated case at 5% significant level. However, the ranksum test showed that both observed and disaggregated streamflow well matched based on median statistics applied to daily streamflow over the study period.

The goal of this research is to develop a relatively simple method for streamflow disaggregation from monthly to daily by reducing computational burden and time, and application to areas where very precise calculations may not be always required. As discussed earlier, some previously developed methods may provide a fairly better result (not tested here), but these methods are computationally intense. This tool, however, developed in an Excel

environment, is fairly simple, broadly applicable with less computational burden, and provides a solution to the disaggregation problem which is of interest to many state and local agencies at present. Thus, this tool can be applied directly at any site to conduct streamflow disaggregation from monthly to daily.

It is difficult to accurately represent all hydrological components and catchment response while developing a disaggregation model. Therefore, most of the proposed methods only utilize streamflow as a decision variable while performing disaggregation. Since the non-linear nature of precipitation over the watershed highly affects streamflow response, future work on streamflow disaggregation from monthly to daily timescale will be related to enhancement of this tool in user friendly software architecture, based on cross-relationship between hydrologic variability and weather forcing, such as precipitation and temperature as inputs.

Acknowledgement

This research is supported by the National Science Foundation (NSF) under award number EPS-0814387. The authors thank collaborators at U.S. Bureau of Reclamation (USBR), including Jennifer Johnson, Karl Tarbet, Jonathan Rocha at Pacific Northwest Region Office, Boise, Idaho. Their assistance and comments during the course of tool development are greatly helpful to implement the proposed method in Pisces system at <http://www.usbr.gov/pn/hydromet/pisces>. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NSF or USBR. Any queries should be directed to the corresponding author for the article. The software package used in this research is available at <http://water.cals.uidaho.edu/products/UIDisagg/>, as of March 1, 2013.

Appendix

Statistics used in this analysis

Major statistics used in this analysis are calculated as follows (Smith et al, 2004; Krause et al, 2005; Acharya et al, 2011):

a) Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{N}} \quad (a)$$

Where, i represents months in three months window, X_i and Y_i are monthly total streamflow for the target station (TS) and source station (SS) respectively, N is the number of observations (in this case, $N=3$).

b) Standard Mean (\bar{Y})

$$\bar{Y} = \frac{\sum_{i=1}^n O_i}{n} = \frac{\sum_{i=1}^n D_i}{n} \quad (b)$$

c) Standard Deviation (S_y)

$$S_y = \sqrt{\frac{\sum_{i=1}^n (O_i - \bar{O})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}} \quad (c)$$

d) Pearson Correlation Coefficient (r)

$$r = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{(n-1)S_o S_d} \quad (d)$$

e) Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} \quad (e)$$

f) Nash-Sutcliffe Efficiency (E)

$$E = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (f)$$

g) Bias (%)

$$Bias = \frac{\sum_{i=1}^n (P_i - O_i)}{\sum_{i=1}^n O_i} * 100\% \quad (g)$$

h) Ratio

$$R = \frac{\sum_{i=1}^n O_i}{\sum_{i=1}^n D_i} \quad (h)$$

Where, O_i and D_i represent observed and disaggregated streamflow respectively; S_o and S_d are sample standard deviations for observed and disaggregated streamflow; n is the number of observations; SS_{err} and SS_{tot} represent residual sum of squares and total sum of squares.

List of Tables

Table 1 Name, location and other characteristics of streamflow gauge stations used in this study

Table 2 Monthly mean and standard deviation of streamflow at Boise River (Case 1: Regulated River) and Upper Snake River (Case 2: UnRegulated River) based on daily observed and disaggregated streamflow during 1996-2000.

Table 3 Additional peak regions selected to calculate statistics for Case 2 (Unregulated Snake River).

Accepted Manuscript
Not Copyedited

List of Figures

Figure 1 Flowchart for streamflow disaggregation from monthly to daily time scale at the Target Station (TS) based on daily and monthly streamflow at the Source Station (SS).

Figure 2 Location of the selected streamflow gauge stations in the Snake River Basin. The sites, A1 and A2 represent the location of the regulated streamflow gauge stations along the Boise River near Parma (USGS 13213000) and at Glenwood Bridge near Boise (USGS 13206000) in Idaho, respectively. And, the unregulated streamflow gauge stations include pairs of B1 and B2, and C1 and C2, which are located in the Upper Snake River. B1 and B2 are located in Snake River above Jackson Lake at Flagg Ranch (USGS 13010065) in Wyoming and in Salt River above reservoir Palisades near Etna (USGS 13027500), respectively. Both C1 near Squirrel (USGS 13047500) and C2 above South Leigh Creek near Driggs (USGS 13052200) are located in Idaho.

Figure 3 Mean monthly streamflow (cubic meter per seconds, cms) pattern for the Boise River near Parma (Sta A1) and at Glenwood Bridge (Sta A2).

Figure 4 Disaggregated vs. Observed daily streamflow (cubic meter per seconds, cms) at Boise River at Glenwood Bridge (Sta A2) during the period of 1996 to 2000. Events 1 to 15 represents additional peak regions selected to calculate statistics for Boise River.

Figure 5 Observed vs disaggregated daily streamflow (cubic meter per seconds, cms) for the selected peaks (high, average and low), which is a subset of the peaks shown in Figure 4.

Figure 6 Box plots showing variation in observed (left) and disaggregated (right) daily streamflow for each month during 1996-2000 for the Boise River at Glenwood Bridge (Sta A2).

Figure 7 Mean monthly streamflow (cubic meter per seconds, cms) pattern for the Snake River near Jackson Lake at Flagg Ranch (Sta B1) and Salt River near Palisades Reservoir at Etna (Sta B2).

Figure 8 Disaggregated vs. observed daily streamflow at Sta B1 during the period of 1996 to 2000, after post processing using cubic spline interpolation. The rectangular dotted box in the year 1998 represents an example of abnormal disaggregated peak observed at Sta B1 as discussed in the text.

Figure 9 Boxplots show the daily variation of observed (left) and disaggregated (right) streamflow for each month during 1996-2000 in the Salt River near Etna (Sta B2).

Figure 10 Disaggregated vs. observed daily streamflow at the Teton River near Driggs (Sta C2) during the period of 1996 to 2000, after post processing using cubic spline interpolation.

Table 1 Name, location and other characteristics of streamflow gauge stations used in this study.

Station	Name	Gauge ID	Latitude	Longitude	Basin	State	Data
Sta A1*	Boise River NR Parma	13213000	43.7817	-116.973	Middle Snake	ID	1972-2011
Sta A2**	Boise River at Glennwood Bridge	13206000	43.6605	-116.279	Middle Snake	ID	1982-2011
Sta B1*	Snake River AB Jackson Lake AT Flagg Ranch	13010065	44.0989	-110.668	Upper Snake	WY	1984-2011
Sta B2**	Salt River AB Reservoir Palisades NR Etna	13027500	43.0797	-111.037	Upper Snake	WY	1954-2011
Sta C1*	Falls River NR Squirrel	13047500	44.0686	-111.241	Upper Snake	ID	1960-2011
Sta C2**	Teton River AB South Leigh Creek NR Driggs	13052200	43.7818	-111.209	Upper Snake	ID	1962-2011

Note: Sta A1 and A2 are regulated stations; other stations are unregulated, * and ** represent Source Station (SS) and Target Stations (TS), respectively.

Accepted Manuscript
Not Copyedited

Table 2 Monthly mean and standard deviation (std) of streamflow at the Boise River (Case 1: Regulated River) and the Upper Snake River (Case 2: UnRegulated River) based on daily observed and disaggregated streamflow during 1996-2000.

Months	Regulated				Unregulated			
	Observed (cms)		Disaggregated (cms)		Observed (cms)		Disaggregated (cms)	
	mean	std	mean	std	mean	std	mean	std
Jan	42.8	64.5	42.5	63.5	14.4	61.0	14.3	78.0
Feb	81.6	77.6	81.8	73.5	13.2	38.0	13.1	38.0
Mar	132.0	71.7	130.5	71.6	14.3	144.0	14.0	60.0
Apr	121.0	56.6	120.2	57.8	32.7	427.0	28.7	630.0
May	104.0	60.0	102.5	67.5	64.3	958.0	63.1	995.0
Jun	95.3	57.7	95.8	56.9	64.6	992.0	64.6	1090.0
Jul	37.9	9.8	37.5	11.5	28.3	397.0	28.2	504.0
Aug	35.7	4.4	35.5	4.7	19.3	189.0	19.4	227.0
Sep	23.0	3.7	23.6	2.8	19.0	126.0	19.1	132.0
Oct	13.0	4.1	12.8	3.6	18.0	96.0	18.0	106.0
Nov	6.9	1.0	7.2	1.0	17.0	80.0	16.8	94.0
Dec	9.0	8.1	9.1	5.4	15.0	77.0	14.0	72.0

Accepted Manuscript
 Not Copyedited

Table 3 Additional peak regions selected to calculate statistics for Case 2 (Unregulated Snake River).

High Flows			
Events	From	To	Total Volume (m³/s)
Peak 1	May 6, 1996	June 12, 1996	2835.93
Peak 2	May 14, 1997	June 18, 1997	4085.84
Peak 3	April 23, 1998	June 29, 1998	3778.32
Peak 4	May 23, 1999	June 2, 1999	968.15
Peak 5	April 18, 2000	June 5, 2000	1732.42
Normal Flows			
Peak 6	April 20, 1996	May 16, 1996	1263.78
Peak 7	April 26, 1997	May 2, 1997	441.18
Peak 8	April 29, 1999	May 13, 1999	726.89
Low Flows			
Peak 9	Nov. 9, 1996	March 7, 1997	1912.83
Peak 10	Dec. 10, 1997	March 16, 1998	1405.08
Peak 11	Dec. 10, 1998	March 12, 1999	1218.67
Peak 12	Dec. 4, 1999	March 15, 2000	1336.50

Accepted Manuscript
Not Copyedited

References

- Acharya, A., Piechota, T.C., Stephen, H., and Tootle, G. (2011). "Modeled streamflow response under cloud seeding in the North Platte River watershed." *Journal of Hydrology*, 409(1-2), 305-314.
- Acharya, A., Piechota, T.C., and Tootle, G. (2012). "Quantitative assessment of climate change impacts on the hydrology of the North Platte River watershed, Wyoming." *Journal of Hydrologic Engineering*, doi: 10.1061/(ASCE)HE.1943-5584.0000543.
- Allison, P.D. (2001). *Missing Data. Thousand Oaks.* CA: Sage Publications.
- Bras, R.L., and Rodriguez-Iturbe, I. (1985). *Random Functions and Hydrology.* Dover, Mineola, New York 1985.
- Cohen, J., Cohen, P., West, S.G., and Alken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd edition, Mahwah, New Jersey: Lawrence Erlbaum.
- Costelloe, J.F., Grayson, R.B., and McMahon, T.A. (2005). "Modeling streamflow for use in ecological studies in a large, arid zone river, central Australia." *Hydrological Processes*, 19, 1165-1183.
- Ganju, N.K., Knowles, N., and Schoellhamer, D.H. (2008). "Temporal downscaling of decadal sediment load estimates to a daily interval for use in hindcast simulations." *Journal of Hydrology*, 349, 512-523.
- Gippel, C.J. (2001). "Australia's Environmental Flow Initiative: filling some knowledge gaps and exposing others." *Water Science and Technology*, 43(9), 73-88.
- Green, N.M.D. (1973). "A synthetic model for daily streamflow." *Journal of Hydrology*, 20, 351-364.

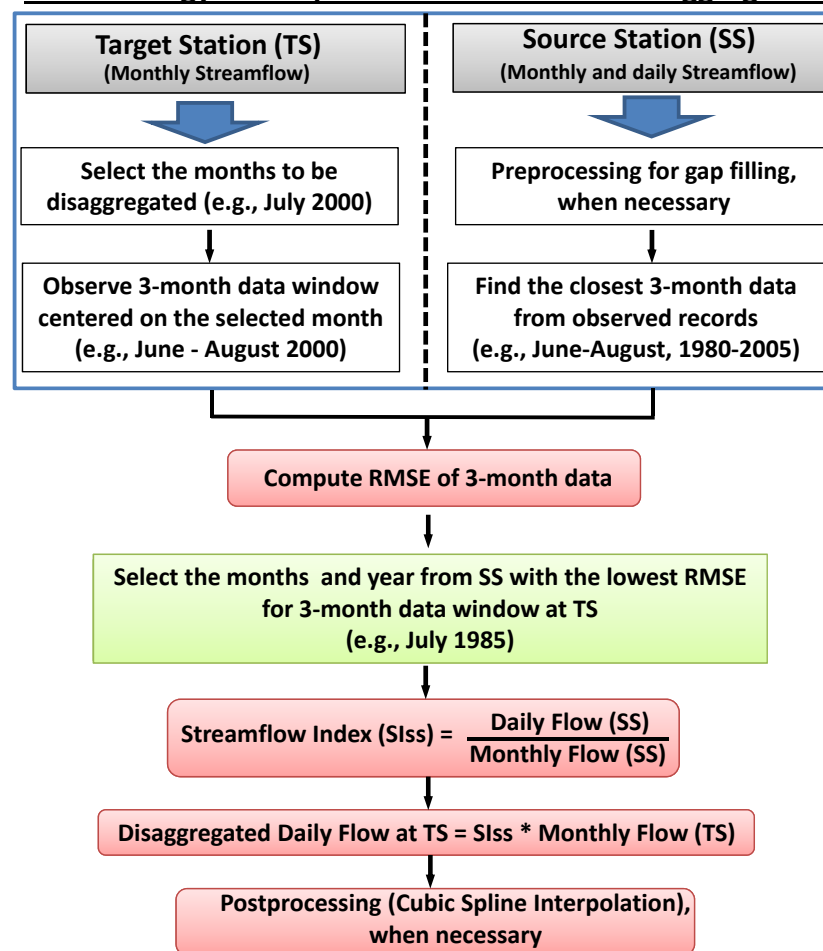
- Grygier, J.C. and Stedinger, J.R. (1988). "Condensed disaggregation procedure and conservation corrections for stochastic hydrology." *Water Resources Research*, 24, 1574-1584.
- Hollander, M. and Wolfe, D.A. (1999). "*Nonparametric Statistical Methods*". Hoboken, NJ: John Wiley & Sons Inc. 1999.
- Howell, D.C. (2008). "*The treatment of missing data*". In W. Outhwaite and S. Turner (Eds.) *Handbook of Social Science Methodology*, London: Sage.
http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html
- Krause, P., Boyle, D.P., and Base, F. (2005). "Comparison of different efficiency criteria for hydrological models." *Advances in Geosciences*, 5, 89-97.
- Lee, T., Sales, J.D., and Prairie, J. (2010). "An enhanced nonparametric streamflow disaggregation model with genetic algorithm." *Water Resources Research*, 46, W08545.
- Little, R.J.A. (1992). "Regression with missing X's: A review." *Journal of the American Statistical Association*, 87(420), 1227-1237.
- Mâsse, B. R. and Truong, Y. K. (1999). "Conditional log spline density estimation" *The Canadian Journal of Statistics*, 27(4), 819-832
- Nagesh Kumar, D., Lal, U., and Petersen, M.R. (2000). "Multisite disaggregation of monthly to daily streamflow." *Water Resources Research*, 36(7), 1823-1833.
- Nowak, K., Prairie, J., Rajagopalan, B., and Lall, U. (2010). "A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow." *Water Resources Research*, 46, W08529.
- Pegram, G.G., Salas, J.D., Boes, D.C., and Yevjevich, V. (1980). "Stochastic Properties of Water Storage." Colorado State University, Fort Collins, 1980.

- Prairie, J., Rajagopalan, B., Lall, U., and Fulp, T. (2007). "A stochastic nonparametric technique for space time disaggregation of streamflows." *Water Resources Research*, 43,W03432.
- Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L. (1980). "*Applied Modeling of Hydrologic Time Series*". Water Resources Publications, Littleton, Colorado, 1980.
- Santos, E.G. and Salas, J.D. (1992). "Stepwise disaggregation scheme for synthetic hydrology." *Journal of Hydraulic Engineering*, 118,765–784.
- Sharma, A., Tarboton, D.G., and Lall, U. (1997). "Streamflow simulation-a nonparametric approach." *Water Resources Research*, 33(2), 291-308.
- Sharma, A. and O'Neill, R. (2002). "A nonparametric approach for representing interannual dependence in monthly streamflow sequences." *Water Resources Research*, 38(7), doi: 10.1029/2001WR000953.
- Sivakumar, B., Wallender, W.W., Puente, C.E., and Islam, M.N. (2004). "Streamflow disaggregation: a nonlinear deterministic approach." *Nonlinear Processes in Geophysics*, 11, 383-392.
- Smith, M.B., Seo, D.J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S. (2004). "The distributed model intercomparison project (DMIP): motivation and experimental design." *Journal of Hydrology*, 298(1-4), 4-26.
- Stedinger, J.R. and Vogel, R.M. (1984). "Disaggregation procedures for generating serially correlated flow vectors." *Water Resources Research*, 20, 47-56.
- Stephens, M.A. (1970). "Use of the Kolmogorov–Smirnov, Cramer– von-Mises and related statistics without extensive tables." *Journal of Royal Statistical Society*, 32B, 115–122.

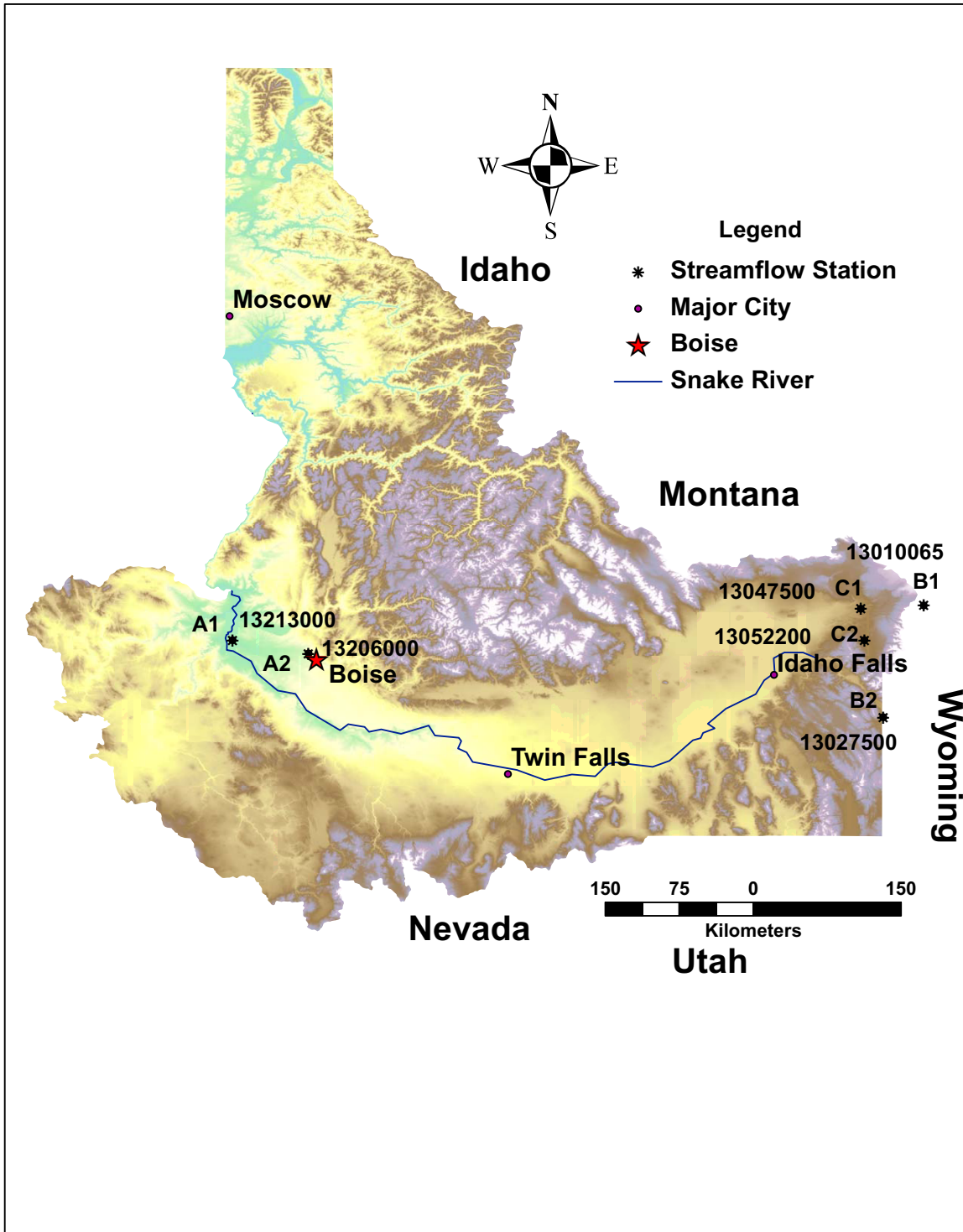
- Tarboton, D.G., Sharma, A., and Lall, U. (1998). "Disaggregation procedures for stochastic hydrology based on non-parametric density estimation." *Water Resources Research*, 34,107–119.
- Valencia, D. and Schaake, J.C. (1973). "Disaggregation processes in stochastic hydrology." *Water Resources Research*, 9, 580-585.
- Wood, A.W., Leung, L.R., Sridhar, V., and Lettenmaier, D.P. (2004). "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs." *Climatic Change*, 15, 189-216
- Xu, Z., Schumann, A., and Brass, C. (2001). "Markov autocorrelation pulse model for two sites daily streamflow." *Journal of Hydrologic Engineering*, 189-195.
- Xu, Z., Schumann, A., and Li, J. (2003). "Markov cross-correlation pulse model for daily streamflow generation at multiple sites." *Advances in Water Resources*, 26, 325-335.

Accepted Manuscript
Not Copyedited

Methodology development for streamflow disaggregation

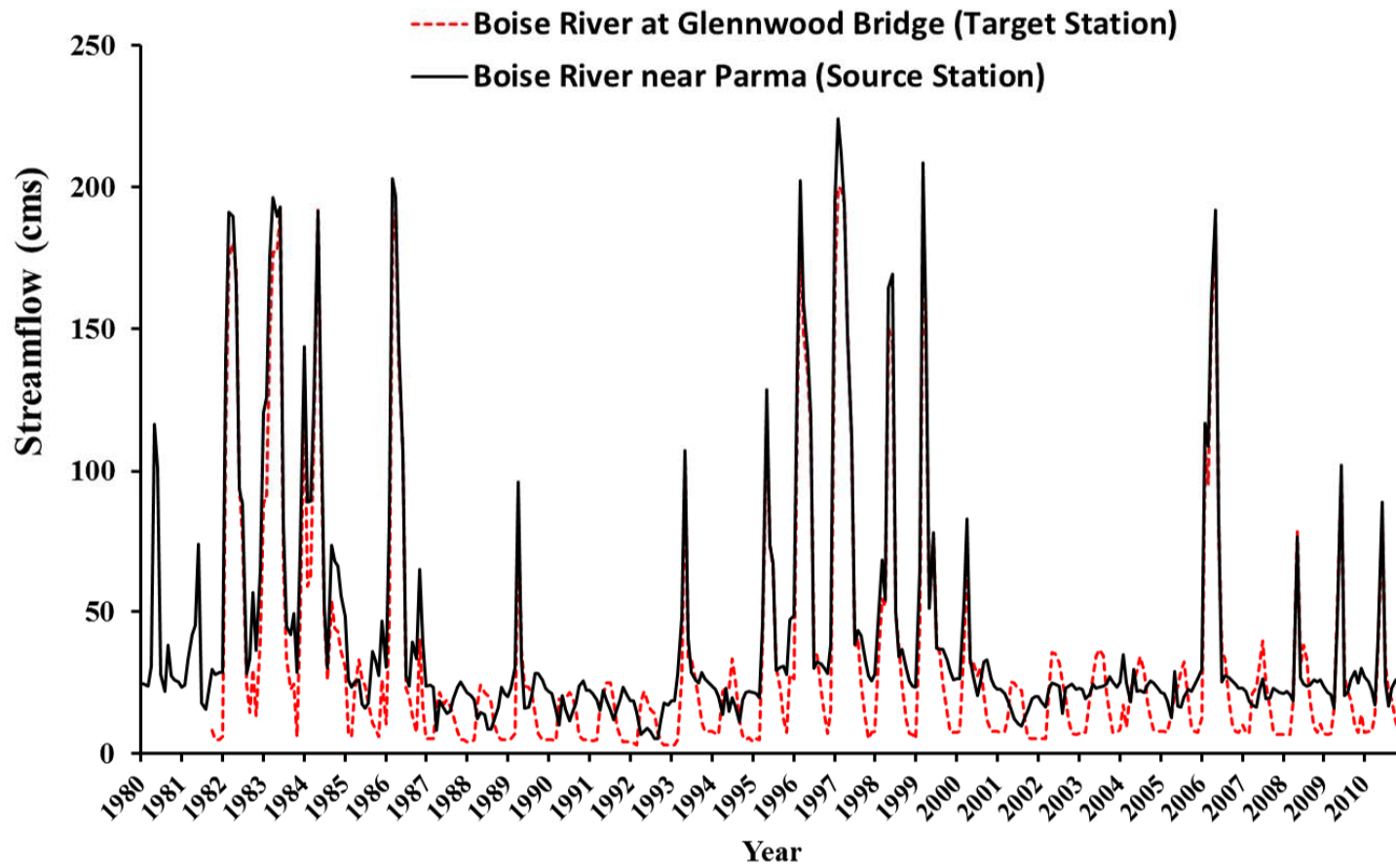


Accepted Manuscript
Not Copied



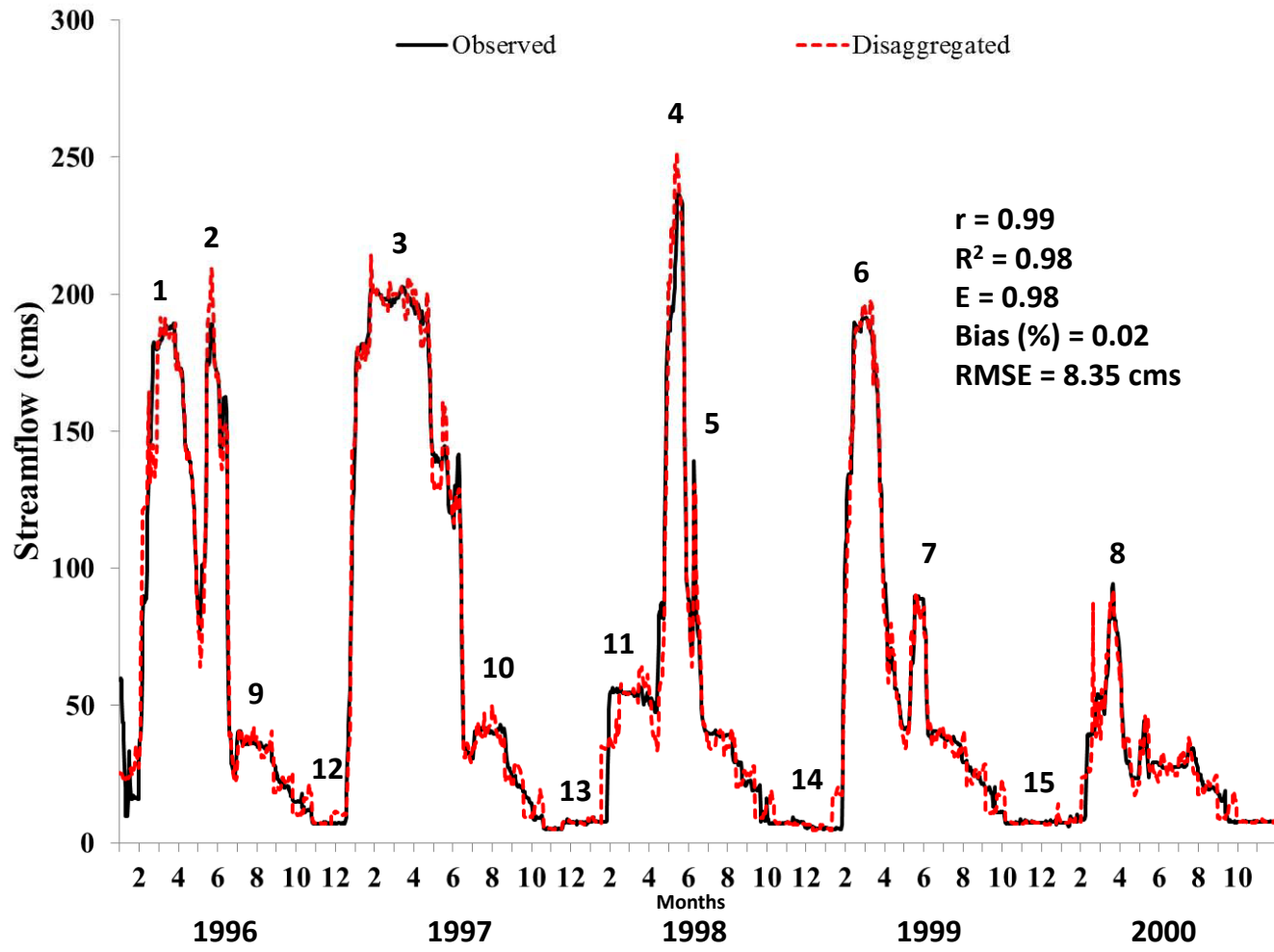
Accepted Manuscript
Not Copyedited

Downloaded from ascelibrary.org by Idaho, University Of on 03/19/13. Copyright ASCE. For personal use only; all rights reserved.



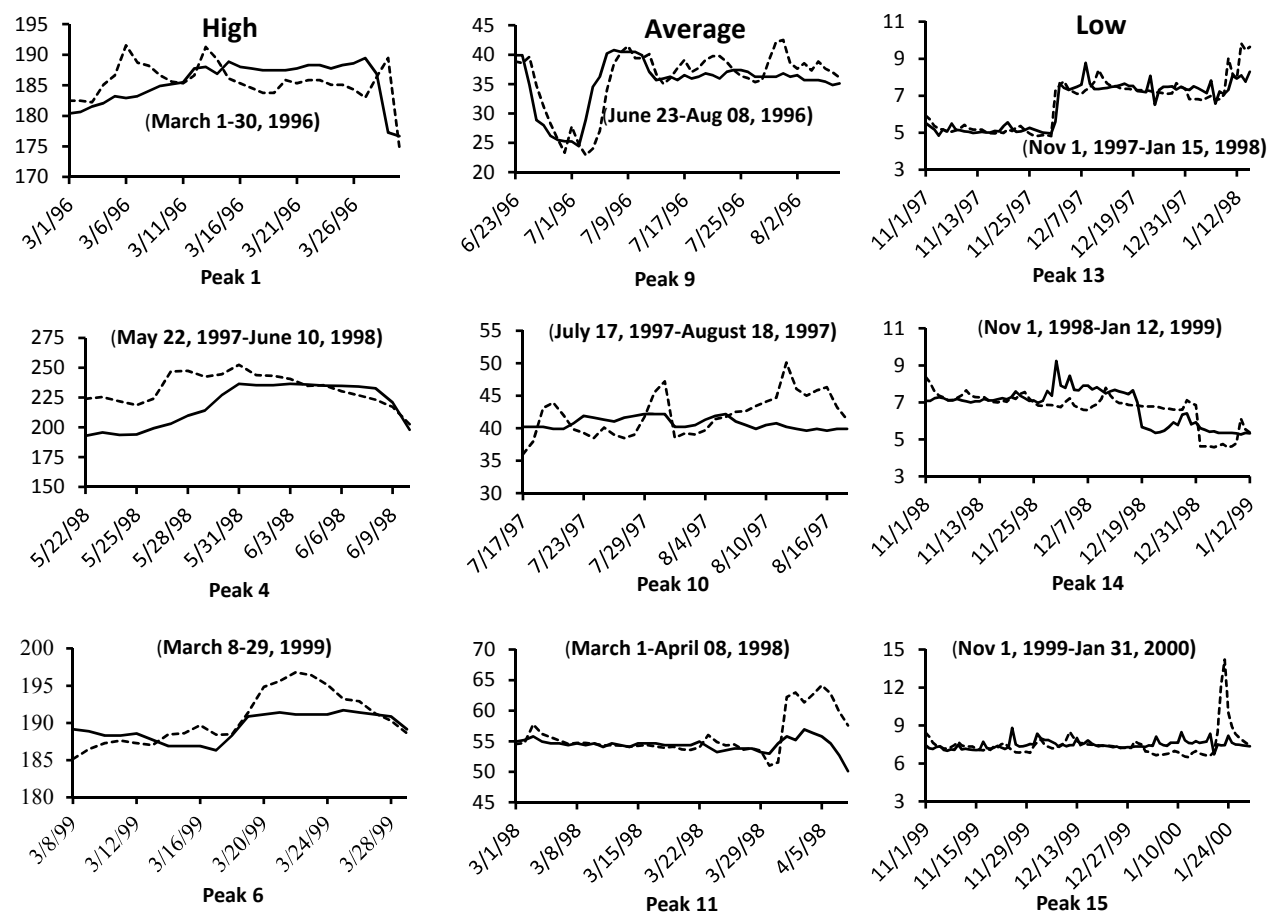
Accepted Manuscript
Not Copyedited

Downloaded from ascelibrary.org by Idaho, University Of on 03/19/13. Copyright ASCE. For personal use only; all rights reserved.



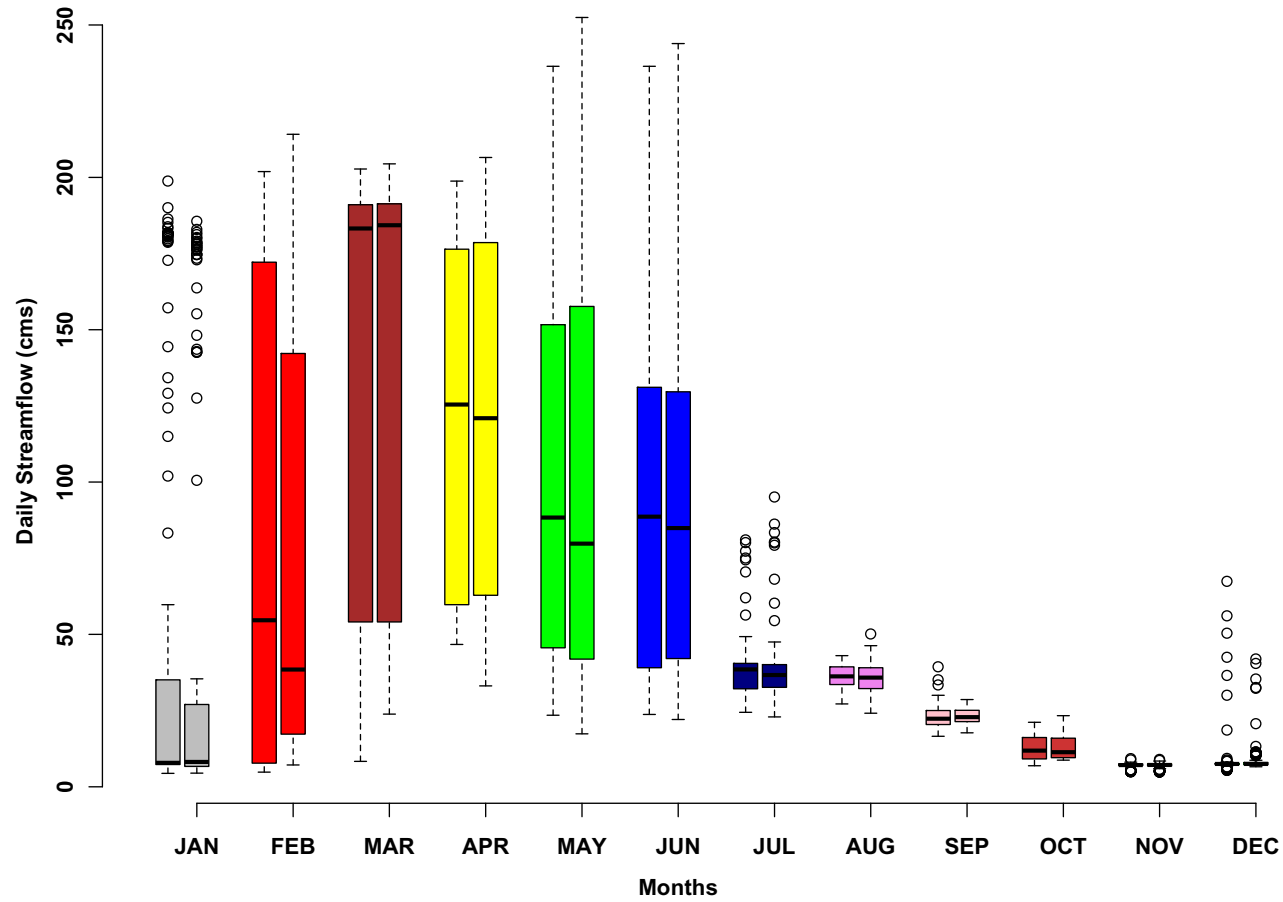
Accepted Manuscript
Not Copyedited

Downloaded from ascelibrary.org by Idaho, University Of on 03/19/13. Copyright ASCE. For personal use only; all rights reserved.

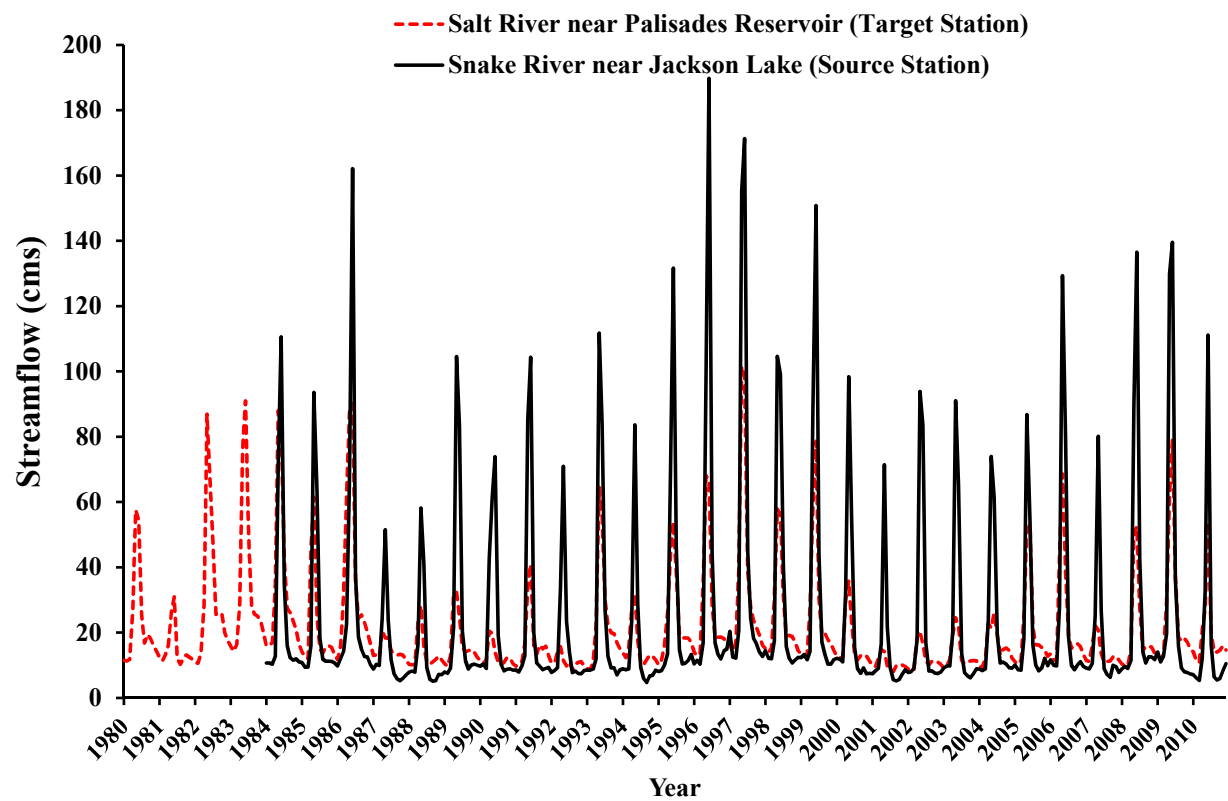


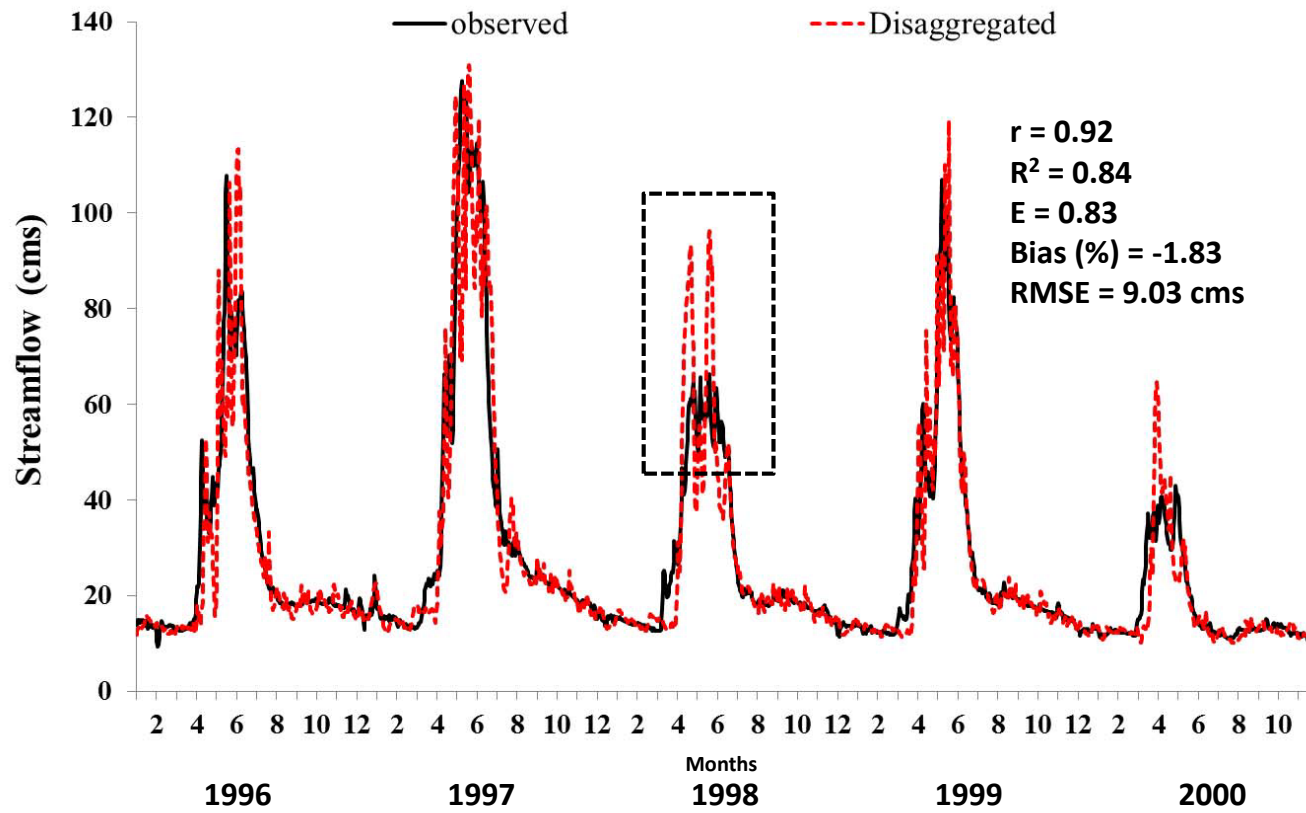
Downloaded from ascelibrary.org by Idaho, University Of on 03/19/13. Copyright ASCE. For personal use only; all rights reserved.

Accepted Manuscript
Not Copyedited

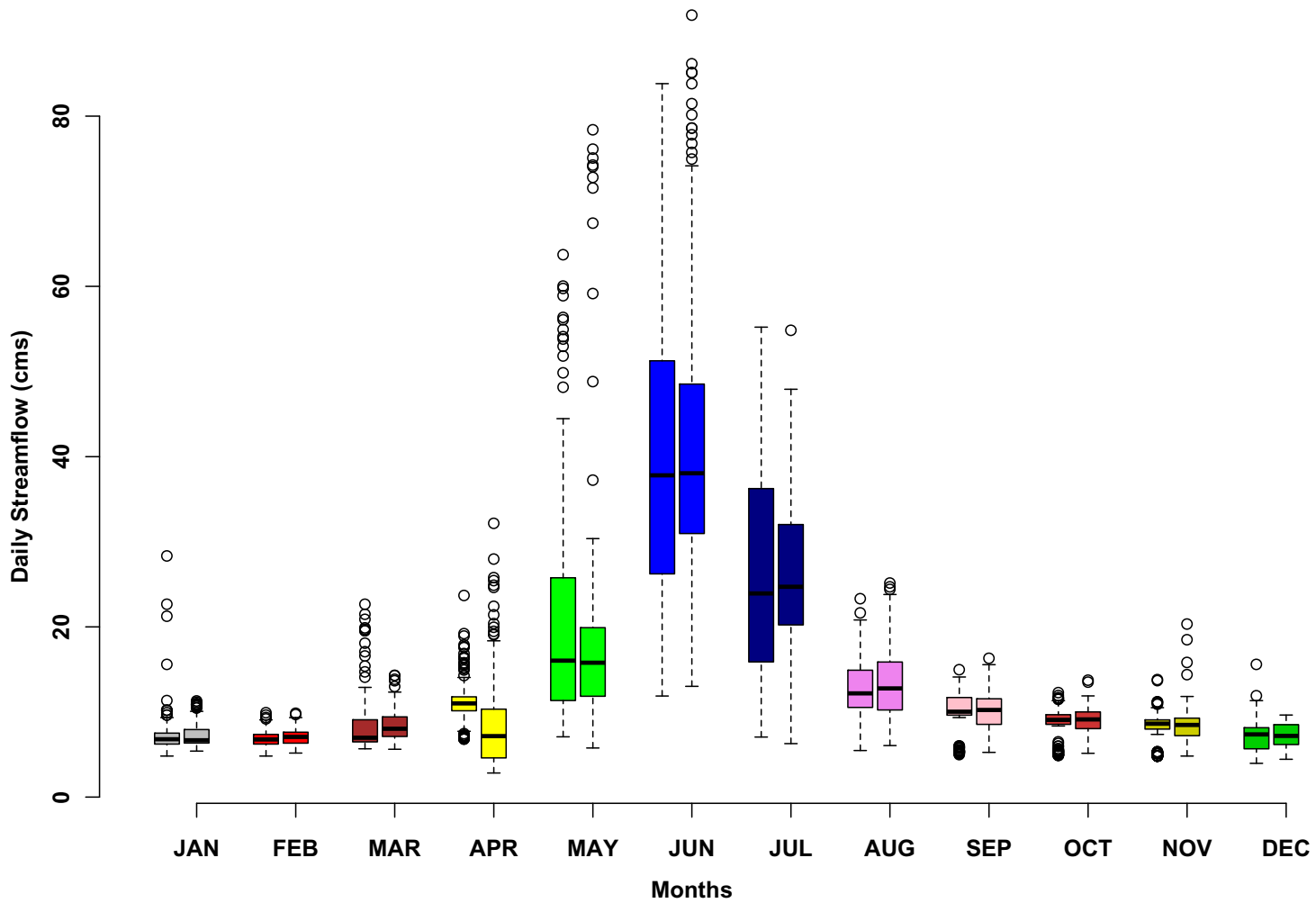


Accepted Manuscript
Not Copyedited

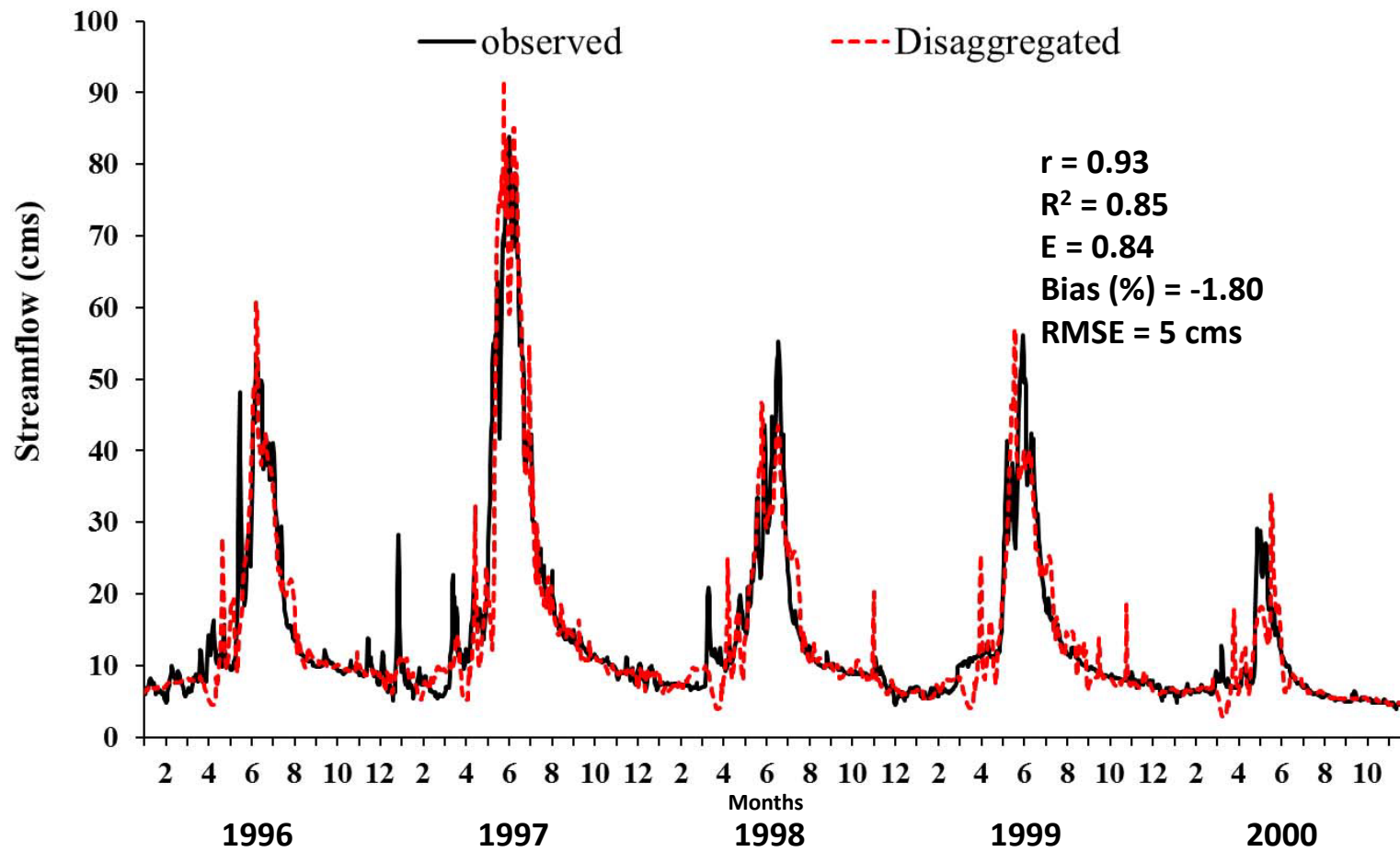




Accepted Manuscript
Not Copyedited



Accepted Manuscript
Not Copyedited



Accepted Manuscript
Not Copyedited

Downloaded from ascelibrary.org by Idaho, University Of on 03/19/13. Copyright ASCE. For personal use only; all rights reserved.