

Stochastic Modeling and Simulation of the Colorado River Flows

T.S. Lee[1], J.D. Salas[2], J. Keedy[1], D. Frevert[3], and T. Fulp[4]

[1]Graduate Student, Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523; PH (970) 491-4302; FAX (970) 491-7727; emails: tae3lee@engr.colostate.edu and julia.keedy@gmail.com.
[2]Professor, Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523; PH (970) 491-6057; FAX (970) 491-7727; email: jsalas@engr.colostate.edu
[3]U.S. Bureau of Reclamation, Denver Federal Center, Mail Code D-8510, Denver, CO 80225; PH (303) 445-2473; FAX (303) 445-6351; email: dfrevert@do.usbr.gov
[4]Area Manager, Boulder Canyon Operations Office, U.S. Bureau of Reclamation, 400 Railroad Avenue, Boulder City, NV 89005; PH (702) 293-8414, FAX (702) 293-8156, email: tfulp@lc.usbr.gov.

Abstract

The Colorado River is one of the most important rivers systems in the western United States. Streamflows of this river represent a significant portion of the water supply in the region. Many river diversions and dams have been developed throughout the system in order to be able to deliver and meet the expected demands from a variety of water users. To plan and operate the river system planners and managers must be able to analyze and assess the effect of possible streamflow scenarios that may occur in the future. For this purpose stochastic simulation studies of the system have been undertaken. The spatial and temporal variability of streamflows in this river is quite complex not only because of its size but the effect of large-scale atmospheric and oceanic fluctuations. Monthly and annual data of naturalized flows at 29 sites have been analyzed to characterize the spatial and temporal variability in key statistics such as means, variances, covariances, skewness, and surplus and drought related statistics. The statistical relationships between sites are just as important as the statistical characteristics of each site. In principle a multivariate model could be applied directly to the 29-site system but such a model will require estimating a large number of parameters, in some cases the parameters may be difficult if not impossible to obtain, and for a river network system such a direct application of a model for the entire system may not be necessarily appropriate. Therefore, the system must be simplified so that it becomes statistically tractable and consistent with a river network. For this purpose alternative modeling strategies (schemes) have been developed in literature and appropriate software are available (e.g. SAMS and SPIGOT). In this study we have applied SAMS and compared three

major modeling schemes. We describe the assumptions behind them, the differences between them, and present results.

Introduction and data description

A stochastic simulation study has been conducted for analyzing the temporal and spatial variability of streamflows in the Colorado River system. It is essential to perform simulation studies in order to evaluate future potential streamflow scenarios that may occur in the system. In this study, the entire Colorado River System is modeled with the computer program SAMS. A multivariate model is needed for this system because it involves a large number of flow points (sites). However, a multivariate model applied directly to monthly flows at all sites will require estimating a very large number of parameters. Therefore, a disaggregation model is used to reduce the number of parameters. Three modeling schemes are utilized, and their results compared in order to choose the most appropriate scheme. The monthly and yearly key statistics for the entire Colorado River System are analyzed for both the intervening and accumulated streamflow volumes.

The Colorado River system consists of 29 selected gaging stations that characterize the river flows over the entire basin. The historical data are naturalized flows, and some of the sites have historical records beginning in 1906 but many other sites are short and incomplete. Thus part of the data has been extended so as to have a complete data set for the period 1906-2003 (Lee and Salas, 2006). Then, the naturalized data of the 29 stations for the years 1906-2003 are used to build a stochastic modeling scheme of the Colorado River system. It is important for the model to consider the geographical location and basin characteristics of each station. Consequently, the stations are grouped together according to their geographical location as well as their cross-correlations. Both intervening flows (IF) and accumulated flows (AF) are used in building the models depending on the site location. Note that intervening flow data at headwater sites (tributary sites) are equivalent to accumulated flows. Figure 1 shows the river network system where sites numbered 1-21 constitute the Upper Colorado (Upper Basin) and sites 22-29 the Lower Colorado (Lower Basin). Site 29'' represents the summation of the intervening flows for all sites in the Lower Basin plus the flows of site 21.

Given the large number of sites of the Colorado River system, there are many alternatives in which a stochastic model can be setup. Since the operational studies of the Colorado River system is commonly undertaken at the monthly time scale one alternative may be to model the monthly flows directly, and another alternative may be generating the annual flows first then disaggregate them into monthly flows. But in turn either alternative can be organized in multiple forms. In this study we select three modeling schemes based on the experience and the availability of software.

Methodology

First of all, data transformation is applied for each year or season and each site if the data is not normally distributed. A multisite contemporaneous autoregressive order-1 model (CAR(1)) is employed to model first key or index sites of yearly flows. Then either a spatial-temporal or temporal-spatial disaggregation scheme is applied. Subsequently, spatial and temporal adjustments are made where applicable according to the condition of the modeling procedure and its requirements such as additivity and proportionality. Figure 2 shows an example of the model set up for scheme 2 (modeling schemes are defined below).

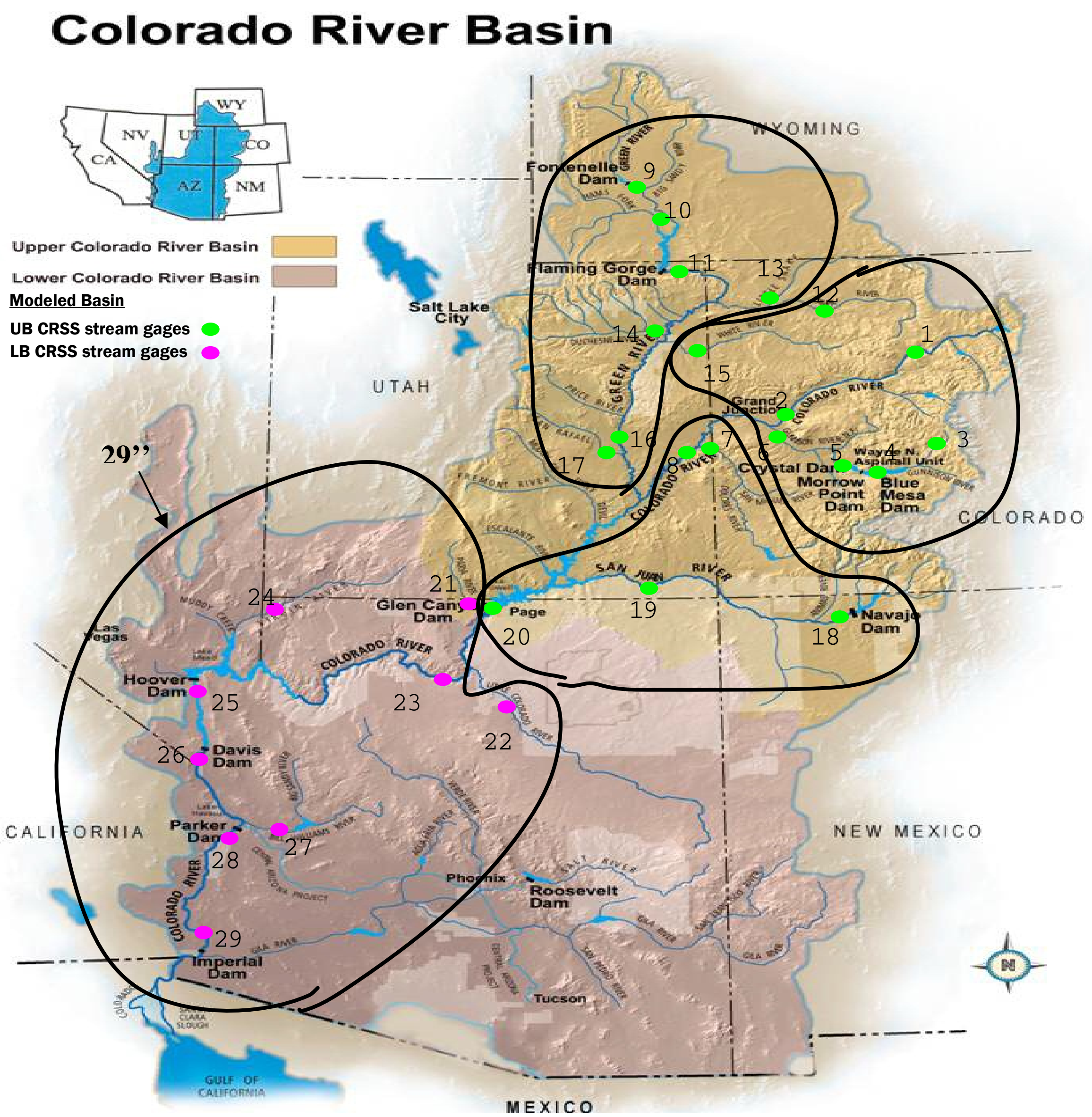

Figure 1. Colorado River System and Site Groupings for Schemes 2 and 3

Disaggregation models that are employed in this study are briefly described. Valencia and Schaake (1973) advocated a disaggregation model for synthetic hydrology employing the linear relation between the aggregate variable (e.g. yearly) and the disaggregate variable (e.g. monthly). Then Mejia and Rousselle (1976) suggested some modifications as

$$\mathbf{Y}_{\nu} = \mathrm{A}\mathbf{X}_{\nu} + \mathrm{B}\mathbf{E}_{\nu} + C\mathbf{Y}_{\nu-1}$$

where in case of one site $\mathbf{Y}_{\nu} = \left[Y_{\nu,1} Y_{\nu,2} \cdots Y_{\nu,\omega}\right]^T$, $\mathbf{Y}_{\nu-1} = [Y_{\nu-1,\omega-1}, Y_{\nu-1,\omega}]^T$ (e.g. if two seasons of the previous year are included), ν denotes the year, τ denotes the season, ω is the number of seasons per year, and , $\mathbf{X}_{\nu}$ is the vector of the yearly values (scalar for one site), and $\mathbf{E}_{\nu}$ represents the stochastic vector $E_{\nu} = \left[E_{\nu,1} E_{\nu,2} \cdots E_{\nu,\omega}\right]^T$. A and B are parameter matrices. In all cases the means of the referred variables have been removed. This model can also be applied for spatial disaggregation by properly defining the variables in the models.

Lane (1990) and Greygier and Stedinger (1990) developed a parsimonious disaggregation model. It is expressed as

$$\mathbf{Y}_{\nu,\tau} = A_{\tau} \mathbf{X}_{\nu} + C_{\tau} \mathbf{E}_{\nu,\tau} + B_{\tau} \mathbf{Y}_{\nu,\tau-1}$$

where $\mathbf{Y}_{\nu,\tau}$ and $\mathbf{Y}_{\nu,\tau-1}$ represents the monthly streamflow for the month τ and $\tau-1$, respectively for year ν, and $\mathbf{E}_{\nu,\tau}$ represents the stochastic term. Accordingly, different parameters (A_{τ}, B_{τ}, and C_{τ}) are used for each month. Lane's model does not meet the additivity condition. Furthermore, modeling with transformed data does not guarantee that data in the real domain will meet the additivity condition even if the disaggregation model meets the additivity condition in the transformed domain. Therefore, temporal adjustments are required in order to meet the additivity condition. However, these adjustments will bias the monthly statistics. Greygier and Stedinger (1990) developed a model includes an additional term to Lane's model in order to minimize the magnitude of the adjustment

$$\mathbf{Y}_{\nu,\tau} = A_{\tau}\mathbf{X}_{\nu} + C_{\tau}\mathbf{E}_{\nu,\tau} + B_{\tau}\mathbf{Y}_{\nu,\tau-1} + D_{\tau}\left[\sum_{u=1}^{\tau-1} W_u \mathbf{X}_{\nu,u}\right]$$

where W_u are weights defined according to the transformation of the monthly streamflows, D_{τ} is a parameter matrix, and $\mathbf{X}_{\nu,u}$ is the monthly streamflow for the additional months considered. Furthermore, spatial adjustments are applied to preserve the proportionality condition for the spatial disaggregation.

In this study we compare three major modeling schemes using SAMS 2005 program. The following is a brief description of the three different disaggregation model schemes.

- Scheme 1: Temporal-Spatial Disaggregation of AF in the Upper Basin and IF in the Lower Basin
  (1) Sites 8, 16, 20, and 29'' are modeled at the annual time scale using a contemporaneous autoregressive model of order 1 (CAR(1)).

(2) The annual flows at 8, 16, 20 and 29'' are disaggregated into monthly flows using Lane's model.
(3) The monthly flows obtained from (2) are disaggregated spatially using Mejia-Rousselle's model.

- Scheme 2: Spatial-Temporal Disaggregation of IF in the Upper and Lower Basins
  (1) Sites 20 and 29'' are modeled at the annual time frame using CAR(1) model.
  (2) The annual flows are disaggregated spatially using Mejia-Rousselle's model (refer to Figure2).
  (3) The annual flows obtained from (2) are disaggregated into monthly flows using Lane's model.

  The scheme 2 is shows schematically in Figure 2. Note that the notation IF23_yr means the yearly intervening streamflow data for site 23 (similar to other sites).

- Scheme 3: Temporal-Spatial Disaggregation of IF in the Upper and Lower Basins
  (1) Annual flows for sites 20 and 29'' are modeled using CAR(1) model.
  (2) The annual flows are disaggregated into monthly flows using Greygier and Stedinger's model.
  (3) The monthly flows obtained from (2) are disaggregated spatially using Mejia-Rousselle's model (Figure1).

Results

In order to evaluate the performance of each scheme, 100 traces, 98 years long of monthly streamflows are generated from each scheme. The generated yearly and monthly statistics (mean, standard deviation, skewness, lag-1 auto-correlation, and cross-correlation) are estimated and compared with corresponding historical statistics. The data generation gives 100 estimates of each statistic, which are compared with the historical value. The statistics are compared in tabular and graphical forms. For example, Figure 3 shows the cross-correlations coefficients between the annual intervening flows for all the sites obtained from the generated (for the three schemes) and historical data. Each value shown for the generated data is the mean value of the 100 generated cross-correlations. Also box plots are used to illustrate the variability of the generated statistics (mean, standard deviation, skewness, and lag-1 correlations) in comparison with the historical statistics. The box extends from the upper quartile to the lower quartile with a line in the middle specifying the median value. The whiskers extend from the 95 percent quantile to the 5 percent quantile. The 'x' markers and the dotted line indicate the historical values of the monthly statistics (e.g. Figure 4). The results for each scheme are summarized below.

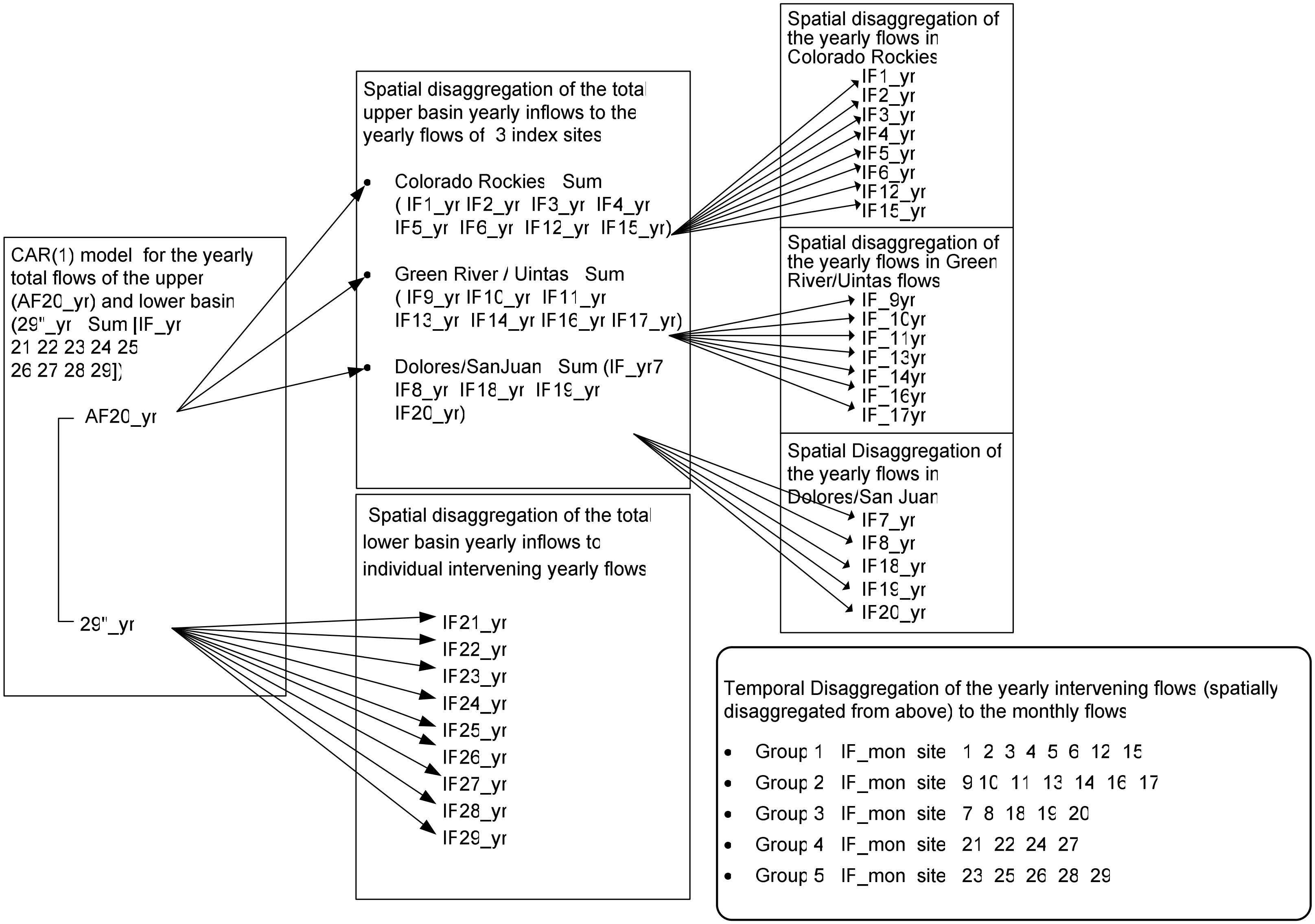

Figure 2. Spatial-Seasonal Disaggregation Model in the Upper and Lower Basins : Scheme 2

- Scheme 1: Temporal-Spatial Disaggregation of AF in the Upper Basin and IF in the Lower Basin

The means of the yearly accumulated flows are well preserved in this scheme, except for sites 22 and 27. The standard deviations of the yearly accumulated flows for sites 9, 10, 22, and 27 are underestimated while for sites 13, 14, and 15 are overestimated. The skewness, ACF(1), and cross-correlations of the yearly accumulated flows are relatively well preserved. The standard deviations of the yearly intervening flows are not well preserved for sites 6, 8, 13, 16, 20, 26, and 27. The other statistics are reproduced relatively well for the yearly intervening flows (cross-correlation shows in Figure 3(b)). The monthly statistics of the accumulated flows are well preserved except for the correlation between the first month of a particular year and the last month of the previous year. The standard deviations of the monthly intervening flows for sites 8, 16, and 20 are not preserved properly. Overall, this scheme is not very good for the reproduction of the yearly statistics. However, it does reproduce the monthly statistics relatively well with some exceptions noted above. These results are expected due to the nature of this scheme (modeled with monthly statistics except AF of site 20). This generation scheme models directly the annual flows only for the key sites 8, 16, 20, and 29" but after the temporal disaggregation for the key sites (8, 16, 20, and 29''), the model does not take into account the yearly statistics of the substation sites except indirectly through the monthly flows.

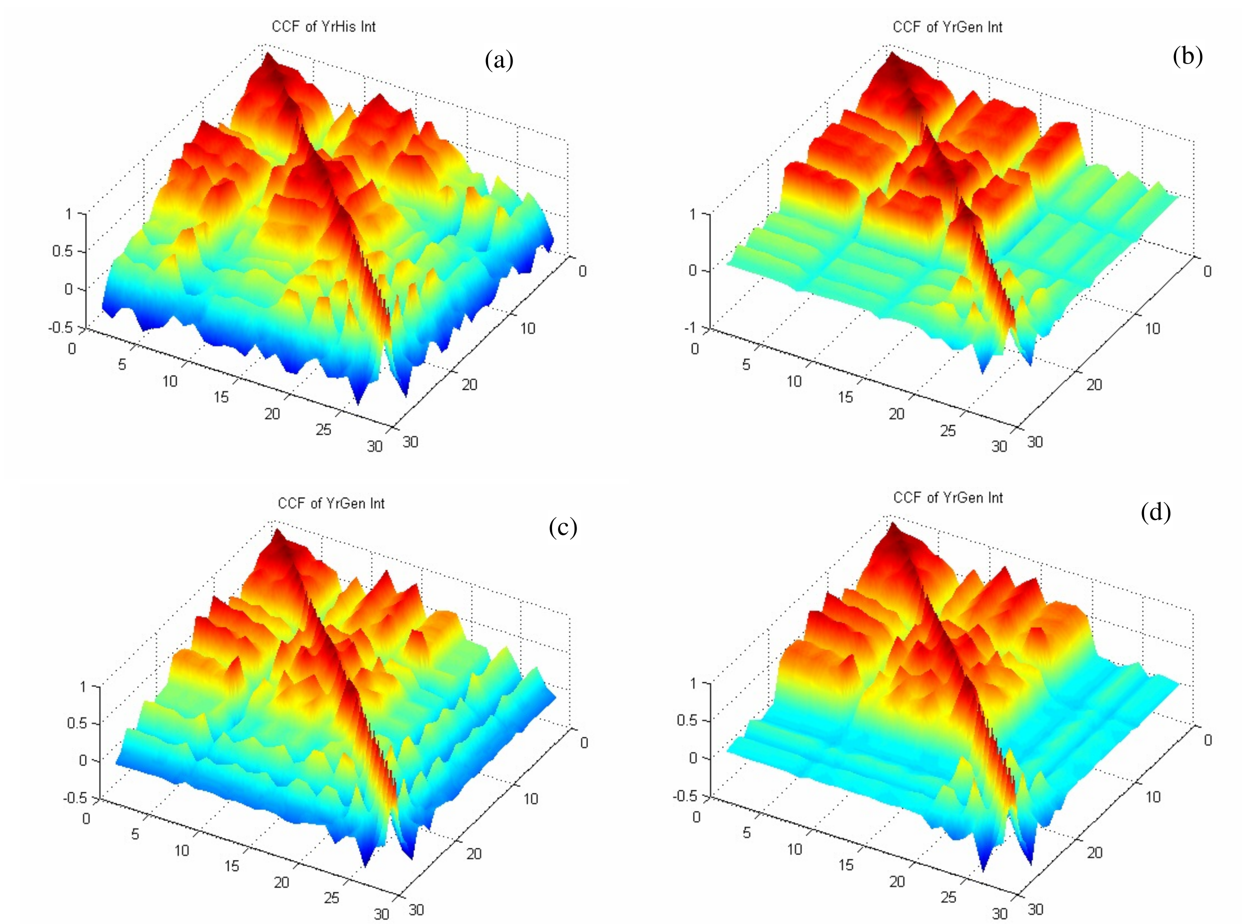

Figure 3 Cross-correlation coefficient between intervening yearly streamflows for (a) historical, (b) generated with scheme 1, (c) generated with scheme 2, and (d) generated with scheme 3

- Scheme 2: Spatial-Temporal Disaggregation of IF in the Upper and Lower Basins

The yearly statistics of the accumulated and intervening flows for this scheme are generally well preserved (e.g. Figure 3(c) shows the cross-correlation of intervening flows). The means and standard deviations of the monthly intervening and accumulated flows are well preserved (Figure4 standard deviations of the monthly AF for sites 13-24). The skewness of the monthly intervening flows are generally well reproduced, except sites 8 and 17 in which some underestimations are observed. The month-to-month correlation of the intervening and accumulated flows are well preserved, except for the serial correlations between the first month of a particular year and the last month of the previous year (e.g. refer to the month-to-month correlation of the accumulated streamflow of the Colorado River system site 13-24 in Figure 5). The cross-correlations are fairly well preserved for the monthly accumulated and intervening flows. However, minor reductions in the cross-correlations between monthly disaggregation groups are observed.

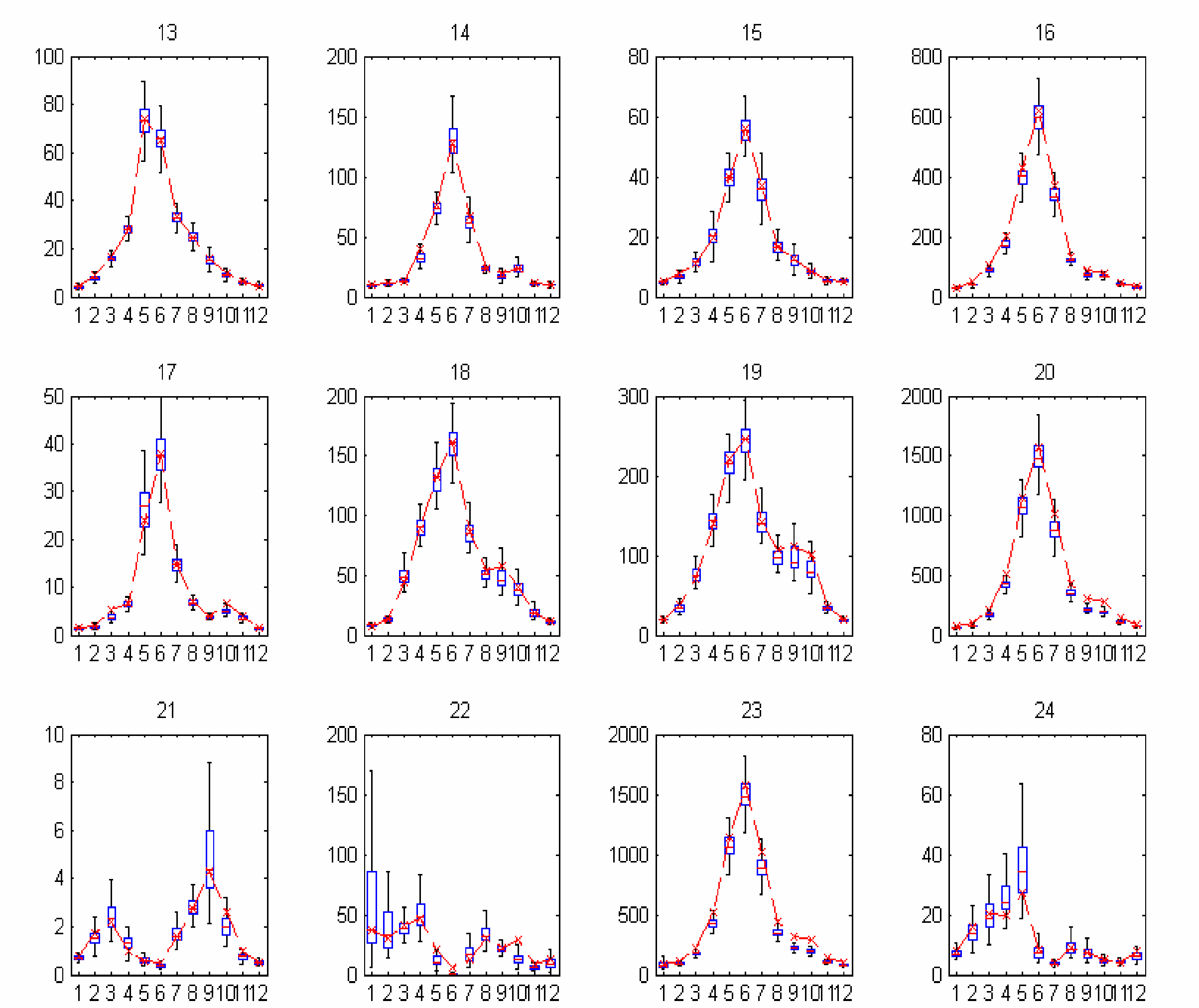

Figure 4 Box plots of the standard deviation of monthly AF for the Colorado River system for sites 13-24 obtained from the generated samples base on Scheme2. The historical statistics are also shown (symbol x).

- Scheme 3: Temporal-Spatial Disaggregation of IF in the Upper and Lower Basins

The statistics of the accumulated yearly flows are generally well preserved in this scheme, except for the overestimation of skewness. However, the standard deviations of the yearly intervening flows are not well preserved for sites 8, 20, and 26. Furthermore, the skewness values are overestimated while the serial correlations are underestimated for the yearly intervening flows. The cross-correlations of intervening flows shows are shown in Figure 3(d). The monthly statistics of the accumulated and intervening flows are well preserved, except for the correlation between the first month of a particular year and the last month of the previous year. This scheme is best among the three schemes for the reproduction of the monthly statistics for both the accumulated monthly flows and for the intervening monthly flows. However, the yearly statistics are not well preserved, especially skewness (for IF and AF) and lag-1 serial correlation (for IF).

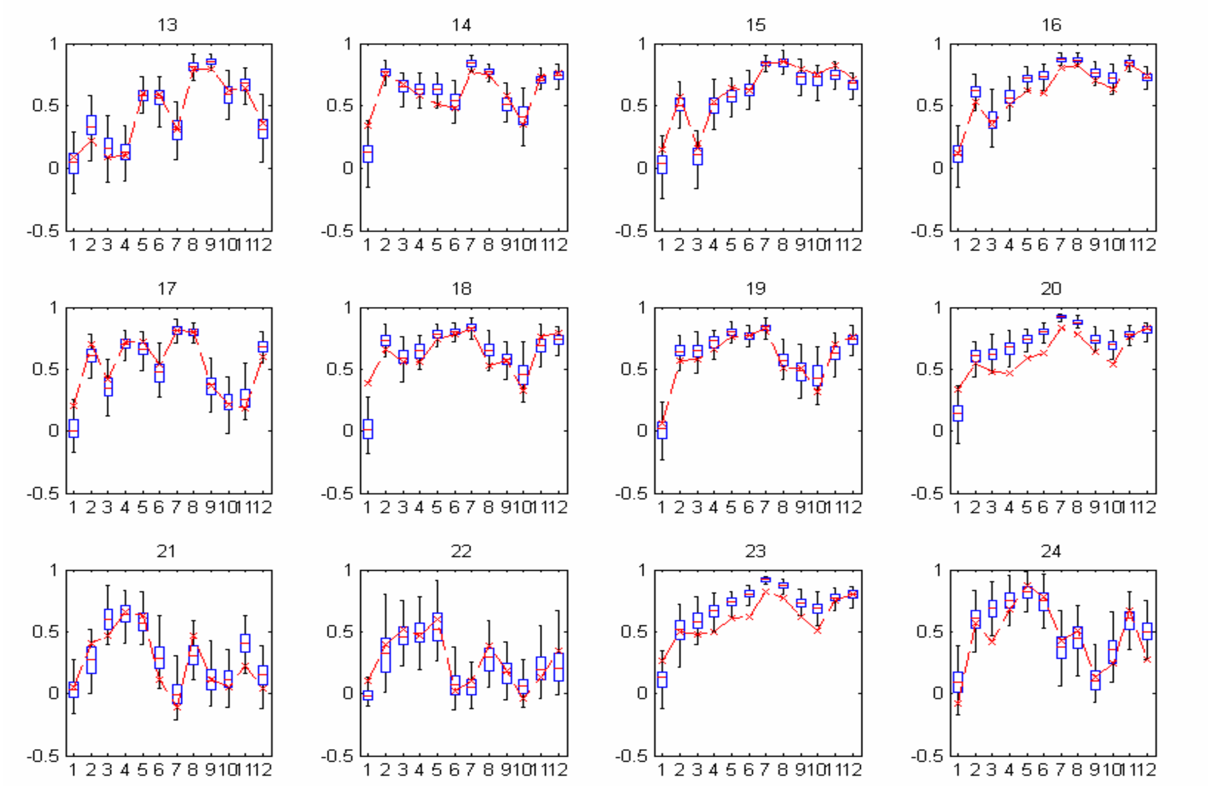

Figure 5 Box plots of month-to-month correlation of monthly AF for the Colorado River system for sites 13-24 obtained from the generated samples base on Scheme2. The historical statistics are also shown (symbol x).

Final Remarks

Three different stochastic modeling schemes are compared for streamflow data generation of the entire Colorado River system. One hundred traces of the same record length as the historical record are generated for each scheme. Various test statistics are calculated over the historical and generated data. A comparison of these statistics reveals the performance of each scheme. From the comparison, it is concluded that scheme 2 well preserves the monthly and yearly statistics and scheme 3 is the best for the preservation of monthly statistics. Overall, scheme 2 is chosen as the most appropriate scheme for the preservation of the statistics on both time scales.

Acknowledgement. This study has been undertaken in connection to the project entitled "Development of Stochastic Hydrology for the Colorado River System", agreement between the U.S. Bureau of Reclamation, Lower Colorado Region and Colorado State University.

References

Grygier JC and Stedinger JR, 1990, SPIGOT, A Synthetic Streamflow Generation Software Package, Technical Description, Version 2.6, School of Civil and Environmental Engineering, Cornell University, Ithaca, New York

Lane WL et al., 1990, APPLIED STOCHASTIC TECHNIQUES (Personal Computer Version), User Manual

Lee TS and Salas JD, 2006, Record Extension of Monthly Flows for the Colorado River System, Technical Report, Bereau of Reclamation

Mejia JM, Rousselle J, 1976, Disaggregation Models in Hydrology Revisited, Water Resources Research, Vol.12(2) pp. 185-186

Valencia, RD and Schaake JC, 1973, Disaggregation Processes in Stochastic Hydrology, Water Resources Research Vol.20(1) pp.580-585