Stochastic simulation models with non-parametric approaches: Case study for the Colorado River basin

Taesam Lee1,JoseD.Salas2, James R. Prairie3, Donald Frevert4, and Terry Fulp5

이태삼, 호세 살라스, 제임스 프레리, 도널드 프리버트, 테리 플립

Abstract

Stochastic simulation of hydrologic data has been widely developed for several decades. However, despite the several advances made in literature still a number of limitations and problems remain. In the current study, some stochastic simulation approaches tackling some of the existing problems are discussed. The presented models are based on nonparametric techniques such as block bootstrapping, and K-nearest neighbor resampling (KNNR), and kernel density estimate (KDE). Three different types of the presented stochastic simulation models are (1) Pilot Gamma Kernel estimate with KNNR (a single site case) and (2) Enhanced Nonparametric Disaggregation with Genetic Algorithm (a disaggregation case). We applied these models to one of the most challenging and critical river basins in USA, the Colorado River. These models are embedded into the hydrological software package, Pros and cons of the models compared with existing models are presented through basic statistics and drought and storage-related statistics.

Key words: Stochastic simulation, Nonparametric, Colorado River, streamflow, disaggregation

1. Introduction

The Colorado River is one of the main sources of water for several states in the western United States. Water management is an important issue in the Colorado River system. Generally, some water is delivered from places where plenty of water exists to the places where water is scarce. Planning the storage, diversion, and delivery of water must consider current and future conditions of the available water resources. Estimating the future availability of water resources is not be easy task. Stochastic simulation have been suggested to create possible streamflow scenarios that may arise in the future. The simulated data allow water managers checking many possible options. Many models for simulating streamflows at monthly and yearly time scales have been developed and applied in water resources management area.

1 정회원, Postdoctoral Fellow, INRS-ÉTÉ490, de la Couronne, Québec (Québec) G1K 9A9, CANADA, Email: tae_sam.lee@ete.inrs.ca

2 정회원, Professor, Dep. of Civil and Env. Engr. B208, Engr. Bldg.Colorado State University,Fort Collins, Colorado 80523-1372, USA

3 비회원, Ph.D.Hydraulic Engineer, Reclamation, CADSWES, Univ. of Colorado, 421-UCBBoulder CO, 80305, USA

4 비회원, Ph.D., Associate Editor of J. Hydro. Engr.

5 비회원, Ph.D., Deputy Regional Director of Reclamation's Lower Colorado Region

The time series simulation models that are typical in hydrology and water resources include Autoregressive Moving Average (ARMA), periodic ARMA (PARMA), multisite ARMA (MARMA), contemporaneous ARMA (CARMA), and disaggregation models (Salas, 1980, Loucks et al., 1981). These models are linear and assume normal distribution. Since hydrologic data such as rainfall and streamflow are not normally distributed, data transformation is unavoidable. The data transformation might induce bias on key statistics such as the mean and standard deviation of the original variable in real domain even if there are no biases in the transformed (normal) variable.

In the last two decades, nonparametric simulation techniques such as bootstrapping, k-nearest neighbors resampling, conditional kernel density estimate, nonparametric disaggregation, and more have been developed to provide alternatives and get around some of the shortcomings of parametric models. In this study, the current nonparametric simulation techniques for streamflow data are investigated and drawbacks of the techniques are revealed such as generating only historical values, and the repetition of seasonal and spatial patterns. In order to tackle the drawbacks of the current nonparametric models, a number of modifications are proposed such as Gamma KDE perturbation, the inclusion of aggregate or pilot variable, Genetic Algorithm mixture, and combination of nonparametric disaggregation and accurate adjusting. Furthermore, the proposed modifications will be useful for simulating intermittent and non-intermittent streamflows jointly at several sites.

The general objective of this research is developing nonparametric simulation techniques that are applicable to hydrologic data such as streamflow and rainfall. For streamflow data, nonparametric models are mainly focused considering the long-term variability and the joint modeling of intermittent and non-intermittent data.

2. Applied conceptions

2.1 Gamma Kernel Density Estimate

The following gamma kernel distribution is defined as:

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} K_{x^2/h^2, h^2/x}(X_i) \tag{1}$$

$$K_{x^2/h^2, h^2/x}(t) = \frac{t^{x^2/h^2-1} e^{-t/(h^2/x)}}{(h^2/x)^{x^2/h^2}\,\Gamma(x^2/h^2)} \tag{2}$$

$$E\{f(Z)\} \approx f(a) + \left.\frac{\partial f(Z)}{\partial Z}\right|_{Z=a} E[Z-a] + \left.\frac{\partial^2 f(Z)}{\partial Z^2}\right|_{Z=a} E[Z-a]^2$$
$$= f(x) + \frac{1}{2} f''(x)\sigma^2(Z) = f(x) + \frac{1}{2} f''(x) h^2$$

In simulation the basic model is to use the Eqs (1) and (2) as well as the pilot variable combined with k-nearest neighbor resampling referred to Salas and Lee (2010) and Figure 1..

2.3 Genetic algorithm in simulation modeling (Lee and Salas, 2010)

The genetic algorithm (GA) is utilized in order to produce the unprecedented mixing sets of disaggregated variables such as monthly streamflow from annual data. Schematic representation of GA in disaggregation model is shown in Figure 2.

3. Results: Colorado River

The representative 29 stations are illustrated in Figure 3. These stations are selected and managed by Bureau of Reclamation for water resources management and drought analysis. The key station is Site 20 in which the Colorado river system is divided into upper and lower system. The results with Gamma kernel and pilot variable based on k-nearest neighboer resampling (KNNR), called KGKP are presented in Figure 4 and Figure 5. It shows that KGKP model reproduces the statistical characteristics of Site 20.

The representative results of nonparametric disaggregation model asssisted with GA (NPDGA) are presented. It leads that the NPDGA model well reprodeces the drought and storage characteristics.

4. Conclusions

For streamflow data nonparametric modeling techniques are focused in the current study. From the meticulous investigations over the existing models, new models are proposed and some existing models are enhanced. The proposed and enhanced model in this study eliminates the limitations and drawbacks that the existing nonparametric models have. Those developed models are tested with various streamflow data, mainly in the Colorado River system. The results show that they are reliable and useful models to simulate streamflows of a single site and a large river basin even with intermittent and non-intermittent sites jointly.

References

Lee, T., Salas, J.D., Prarie, J., 2010. An Enhanced Nonparametric Streamflow Disaggregation Model with Genetic Algorithm. Water Resources Research, Accepted, DOI: 2009WR007761.

Salas, J.D., Lee, T., 2010. Non-Parametric Simulation of Single Site Seasonal Streamflows. Journal of Hydrologic Engineering, 15(4): (in press.)

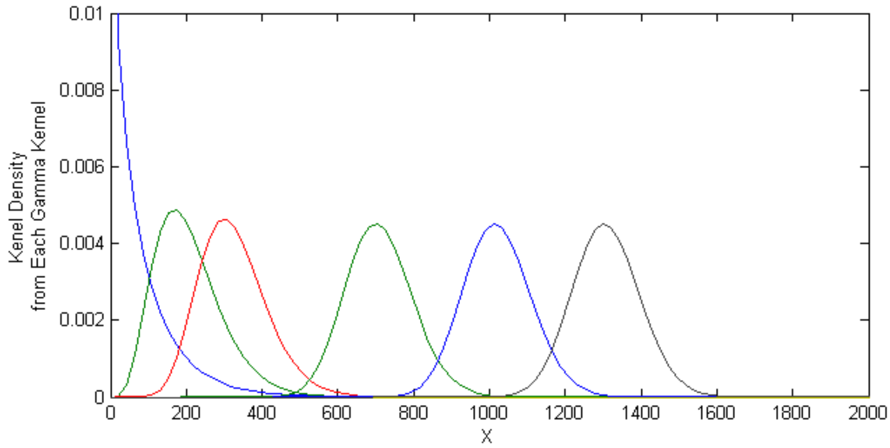


Figure 1 Example of Gamma Kernel Density Estimate with different locations of kernels

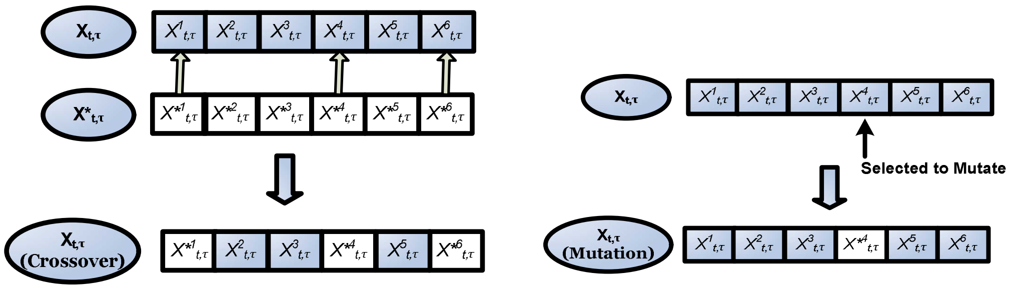
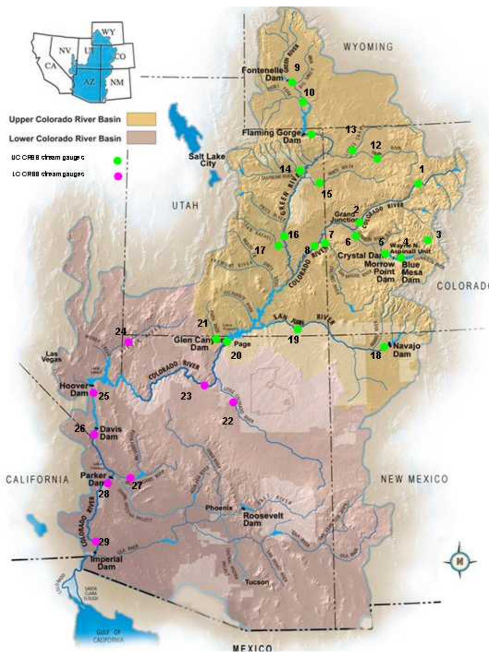


Figure 2 Sketch of the crossover and mutation process in simulation



Figur 3 Map of Colorado River System with twenty nine stations; the system is divided into two as the upper Colorado River basin(1-21) and the lower Colorado River basin (22-29); the map is obtained from Bureau of Reclamation (2007)

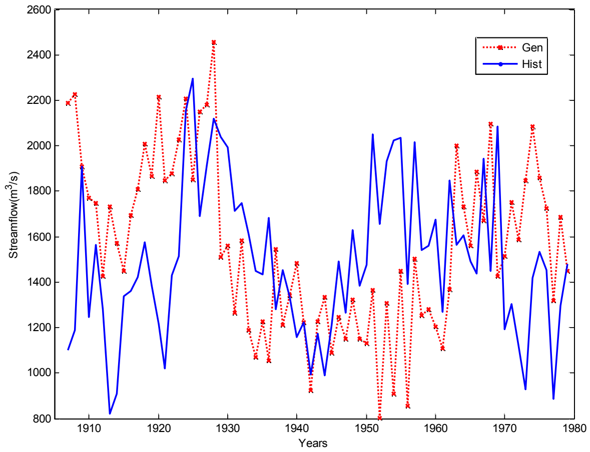


Figure 5 Yearly time series of historical data (segment line) and one set (dotted line with ‘x’) of the summation of generated monthly data (KGKP)

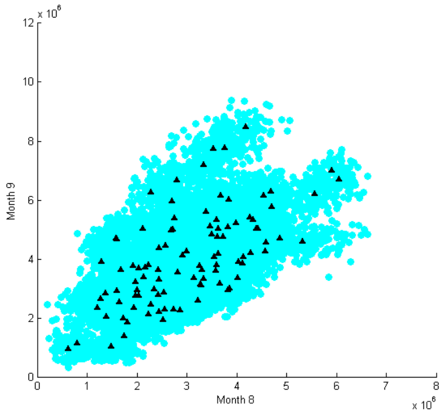


Figure 4 Scatter plot of monthly streamflow data with month 8 (x-axis) and month 9 (y-axis) for historical (filled triangle) and 50 sets of the generated data from KGKA model (grey circle) for Colorado River at Lees Ferry (Acre-feet)

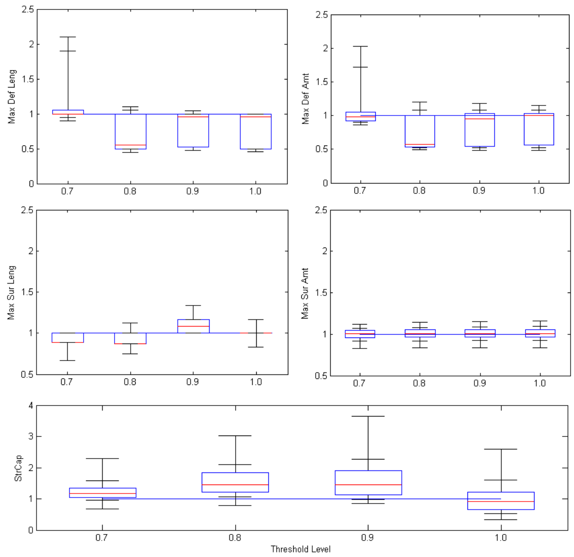


Figure 6 Monthly drought statistics at different threshold levels of Historical (dot line) and KPAG simulations (boxplot) for Site 20 of the ColoradoRivermonthlystreamflow