ICIWRM – 2000, Proceedings of International Conference on Integrated Water Resources
Management for Sustainable Development, 19 – 21 December, 2000, New Delhi, India

A nonparametric approach for preserving interannual dependence in synthetic flow sequences

ASHISH SHARMA
School of Civil and Environmental Engineering, The University of New South Wales, Sydney, Australia
ROBERT O'NEILL
Department of Land and Water Conservation, Sydney, Australia

Abstract
The estimation of risks associated with alternate plans and designs for water resources systems requires generation of synthetic streamflow sequences. The mathematical algorithms used to generate these sequences at monthly time scales are found lacking in two main respects: inability in preserving the dependence attributes particularly at large (seasonal to inter-annual) time lags; and, a poor representation of the observed distributional characteristics. Traditional approaches for representing such dependence consist mainly of stochastic disaggregation models. These models use generated annual streamflows that are disaggregated to monthly values while prescribing an assumed annual to monthly dependence structure. In this process, the dependence at the year boundaries and between years is not reproduced. These models are characterised based on conventional probability distributions that makes it difficult to represent “unusual” features such as asymmetry or multimodality.

Proposed here is an alternative to such conventional approaches that naturally incorporates both observed dependence and distributional attributes in the generated sequences. Use of a nonparametric framework provides a simple and effective method for reproducing the observed probability density characteristics. A careful selection of prior lags as conditioning variables imparts the appropriate short-term memory, while use of an “aggregate” flow variable defined as the aggregate flow during the past twelve months allows representation of interannual dependence in the generated sequences. The nonparametric simulation model is tested on two monthly streamflow datasets – the Beaver River near Beaver, Utah, USA, and the Burrendong dam inflows, New South Wales, Australia.

INTRODUCTION

An important goal of stochastic hydrology is to generate synthetic streamflow sequences that are statistically similar to the observed flow record. These sequences serve as inputs for Monte Carlo simulation of a reservoir system, to help identify plans and policies for efficient management of available water resources. A key requirement in stochastic streamflow simulation is that the generated sequences be similar to the observed flows. This implies that the distributional and dependence attributes of observed flows should be accurately reproduced in the simulations. Representation of seasonal to inter-annual dependencies commonly associated with sustained droughts or periods of large flows is of particular importance for reservoir system management. Absence of such dependencies in simulations can result in an inaccurate representation of the flows that are likely to occur. This can in-turn lead to biased reservoir operating policies causing both loss of revenue in reservoir operation, and a possible hazard for users downstream. This paper presents an approach for stochastic simulation of seasonal streamflow sequences that attempts to reproduce such longer term dependence characteristics and the observed distributional attributes in the generated flow sequences. The approach is developed within a nonparametric density estimation framework that ensures accurate representation of the distributional attributes present in the historical flow record. Use of an aggregate streamflow variable, details on which are presented in later sections, ensures an accurate characterisation of the seasonal to inter-annual dependencies in the model simulations.

Stochastic simulation of seasonal flows has traditionally been approached using two different perspectives. Autoregressive moving average (ARMA) models have been commonly used to model both seasonal and annual streamflow sequences. These models assume that the current flow is linearly related to previous observations. Many a times the actual flow values need to be transformed to an alternate variable that conforms well with the assumptions of linearity (or a Gaussian probability density) implicit in the model structure. Use of such a framework offers an accurate representation of the dependence between the current and a few past flow values, but does not necessarily ensure that longer-term (seasonal to interannual) dependencies are accurately reproduced.

An alternative to the ARMA models discussed above are stochastic disaggregation approaches. Here the stochastic simulation proceeds in two stages. First, an annual flow sequence is generated using an appropriately chosen model, using previous year flows as the basis to prescribe the observed annual dependence structure. Next, the generated annual or aggregate flow for each year is divided or disaggregated into the various seasonal components. This ensures that if the annual flow corresponds to a low flow year, the associated seasonal flows will also represent the same. While this offers a reasonable alternative to ARMA models, and also ensures that some measure of inter-annual dependence is translated to the seasonal flow simulations, the resulting flow sequences offer only an approximate representation of the processes observed in the historical flow record. Some of the disadvantages in the use of ARMA and stochastic disaggregation models for seasonal streamflow simulation are:

Representation of over-year dependence – A disaggregation approach is designed to reproduce the dependence structure between the aggregate annual flow and the seasonal components, as well as the dependence amongst the seasonal flow values. However, the dependence between the seasonal flows from one year to the next is not modelled. Of particular concern here is that the first season in each new year bears little resemblance to the flow in the preceding season. Hence, if the pattern at the end of the year indicates the development of a drought, this may be completely reversed in the flow values for the next year.

Mis-representation of inter-annual dependence – Disaggregation models use the water year as the basis for simulating an aggregate annual flow value. While such an approach is essential in the disaggregation modelling framework, the assumption that seasonal flows in the current year are dependent on the previous water year's flow may not always be realistic.An improved alternative could be to model the dependence between each seasonal flow component and the aggregate flow in the 12 months that precede the season.

Representation of nonlinear dependence and non-standard probability density functional forms – Traditional approached for stochastic simulation are often based on rigid assumptions about the form of dependence between the various flow variables, or the underlying joint or marginal probability density functions. Such assumptions may not always be valid, as illustrated in [*Sharma et al.*, 1997]. An alternative that effectively removes the above mentioned problem is suggested in [*Lall and Sharma*, 1996; *Sharma et al.*, 1997; *Tarboton et al.*, 1998]. These approaches are nonparametric and make no prior assumptions about the form of dependence or probability distribution. Use of the data-based framework ensures that resulting simulations have similar dependence and distributional attributes as observed in the historical record.

Proposed here is a seasonal streamflow generation approach that is free from the disadvantages noted above. Generated sequences reproduce both the short term as well as interannual dependence present in the historical flows. Use of the nonparametric framework ensures that dependence and distributional attributes in generated flows are similar to those in the historical record. What follows is a brief background on nonparametric methods, their applications in hydrology and water resources, and how they can be used to formulate conditional streamflow simulation models. Next, methodological and algorithmic details on the nonparametric streamflow simulation model proposed here are presented. The model is next applied to two streamflow datasets - 84 years (1914-1998) of monthly streamflows in the Beaver River near Beaver, Utah, USA, and to 105 years (1890 to 1994) of Burrendong dam inflows on the Maquarie River in eastern NSW, Australia. We conclude with a discussion of the approach, its pros and cons, and mention some of the work that lies ahead.

NONPARAMETRIC APPLICATIONS FOR STOCHASTIC STREAMFLOW GENERATION

The past few years have seen a surge in applications of nonparametric methods for probability density and regression function estimation to a range of hydrologic problems. Interested readers may refer to [*Lall*, 1995] for a review. Some of the applications related to the present work are – a synthetic streamflow resampling approach using nearest neighbour density estimation principles [*Lall and Sharma*, 1996]; a nonparametric alternative to the Autoregressive order p model (the NPp or the nonparametric order p streamflow simulation model) [*Sharma et al.*, 1997]; and, a nonparametric alternative to traditional disaggregation approaches (the NPD or the nonparametric disaggregation model) [*Tarboton et al.*, 1998]. Streamflow simulation is an exercise in conditional probability distributions [*Bras and Rodriguez-Iturbe*, 1985]. Simulation of flow X_t conditional to p prior flows (X_{t-1}, X_{t-2}, …, X_{t-p}) involves estimation of the conditional probability density function $f(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_{t-p})$. Similarly, disaggregation of an aggregate flow $Z = X_1 + X_2 + \ldots + X_d$ into the d seasonal components (X_1, X_2, …, X_d) requires estimation of the conditional multivariate probability density $f(X_1, X_2, \ldots, X_d \mid Z)$. Conventional approaches assume certain distributional forms for the joint and marginal probability densities of the flow variables, from which the above conditional probability density functions are derived. These conditional densities are then expressed using parameters such as the mean, variance and skewness, and measures of dependence such as correlation. As these methods rely solely on parameters (mean, variance, skewness, correlation) of the data to characterise the assumed probability density functions, they are termed *parametric*. Such methods are useful only if the assumptions about the underlying distributional forms are accurate. One often comes across streamflow records that are not easily characterisable by the commonly used probability distributions (see examples [*Lall and Sharma*, 1996; *Sharma et al.*, 1997; *Tarboton et al.*, 1998]).

Nonparametric methods offer an efficient alternative to traditional parametric approaches. A nonparametric kernel probability density estimate is obtained by considering the cumulative effect of smooth functions called kernels placed over each sample data point. Using a Gaussian kernel function, the multivariate kernel probability density $\hat{f}_{\mathbf{X}}(\mathbf{x})$ of a d-dimensional variable set **X** at coordinate location **x** is estimated as:

Error! Bookmark not defined.
$$\hat{f}_{\mathrm{X}}(\mathrm{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(2\pi)^{d/2}\lambda^{d}\det(\mathrm{S})^{1/2}}\exp\left(-\frac{(\mathrm{x}-\mathrm{x}_i)^T\mathrm{S}^{-1}(\mathrm{x}-\mathrm{x}_i)}{2\lambda^2}\right) \quad (1)$$

where
$\mathbf{x}_i$ is the i'th multivariate data point, for a sample of size n,
S is the sample covariance of the variable set **X**, and,
λ is a smoothing parameter, known as the "bandwidth" of the kernel density estimate.

The bandwidth, λ, is the key to an accurate estimate of the probability density. A large value of λ results in an oversmoothed probability density, with subdued modes and over-enhanced tails. A low value, on the other hand, can lead to density estimates overly influenced by individual data points, with noticeable bumps in the tails of the probability density. Several operational rules for choosing optimal values of the bandwidth λ are available in the literature. This study uses the Least Squares Cross Validation approach, details on which are given in [*Sharma et al.*, 1998].

PROPOSED APPROACH

This approach is aimed at accurately representing interannual dependence in simulated flows. Consider the flow at time t to be X_t, where t could represent annual, seasonal or monthly time steps. For example, for monthly flows, X_1, X_2, …, X_{12} would be the flows for the first 12 months, X_{13}, …, X_{24} the flows for the next 12 months, and so on. The aggregate flow variable Z_t can then be defined as:

Error! Bookmark not defined.
$$Z_t = \sum_{j=1}^{m} X_{t-j} \quad (2)$$

where *m* is the number of prior flows included in the aggregate variable. This study uses monthly flows (*m*=12) to formulate the simulation model. The variable Z_t thus represents the annual flow during the past 12 months for the month being simulated, and its use as a conditioning variable enables proper representation of interannual dependence features. Simulation proceeds from the following conditional probability density:

Error! Bookmark not defined.

$$f\left(X_t \mid X_{t-1}, X_{t-2},..., X_{t-p}, Z_t\right) = \frac{f\left(X_t, X_{t-1}, X_{t-2},..., X_{t-p}, Z_t\right)}{\int f\left(X_t, X_{t-1}, X_{t-2},..., X_{t-p}, Z_t\right) dX_t} = \frac{f\left(X_t, X_{t-1}, X_{t-2},..., X_{t-p}, Z_t\right)}{f_m\left(X_{t-1}, X_{t-2},..., X_{t-p}, Z_t\right)} \quad (3)$$

where $f_m(.)$ represent the marginal probability density of the variable set. Note that the above conditional probability density is a function of (p+1) variables: Z_t and (X_{t-1}, X_{t-2}, …, X_{t-p}). While use of the variables (X_{t-1}, X_{t-2}, …, X_{t-p}) enforces a short term (till lag p) dependence structure in the simulated flow value, the aggregate variable Z_t ensures that the annual dependence pattern is correctly represented. Also note that the conditional probability density in (3) has been specified as a function of p prior lags of X_t. One needs to estimate the appropriate value for p in case of a real application using an order selection scheme such as the Akaike Information Criterion (AIC) [*Akaike*, 1974] or Generalised Cross Validation (GCV) [*Craven and Wahba*, 1979]. The authors recommend the use of GCV for estimation of the optimal model lag. The present application assumes p equal to 1 for the sake of simplicity. The conditional density used for simulation then becomes:

Error! Bookmark not defined.

$$f\left(X_t \mid X_{t-1}, Z_t\right) = \frac{f\left(X_t, X_{t-1}, Z_t\right)}{f_m\left(X_{t-1}, Z_t\right)} \quad (4)$$

Using the kernel density estimator in (1), the conditional density in (4) is estimated as:

Error! Bookmark not defined.

$$\hat{f}\left(X_t \mid X_{t-1}, Z_t\right) = \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\lambda^2 S'}} w_i \exp\left(-\frac{(X_t - b_i)^2}{2\lambda^2 S'}\right) \quad (5)$$

where:

$\hat{f}\left(X_t \mid X_{t-1}, Z_t\right)$ is the conditional probability density estimate;

S’ is a measure of spread of the conditional probability density, expressed as:

Error! Bookmark not defined. $S' = S_{11} - \begin{bmatrix} S_{12} \\ S_{1z} \end{bmatrix}^T \begin{bmatrix} S_{22} & S_{2z} \\ S_{2z} & S_{zz} \end{bmatrix}^{-1} \begin{bmatrix} S_{12} \\ S_{1z} \end{bmatrix}$

where the covariance matrix of the variable set (X_t, X_{t-1}, Z_t) is written as:

$$Cov(X_t, X_{t-1}, Z_t) = \begin{bmatrix} S_{11} & S_{12} & S_{1z} \\ S_{12} & S_{22} & S_{2z} \\ S_{1z} & S_{2z} & S_{zz} \end{bmatrix};$$

w_i is the weight associated with each kernel that constitutes the conditional probability density:

Error! Bookmark not defined. $$w_i = \frac{\exp\left(-\frac{1}{2\lambda^2}\begin{bmatrix}(X_{t-1}-x_{i-1})\\(Z_t-z_i)\end{bmatrix}^T\begin{bmatrix}S_{22} & S_{2z}\\S_{2z} & S_{zz}\end{bmatrix}^{-1}\begin{bmatrix}(X_{t-1}-x_{i-1})\\(Z_t-z_i)\end{bmatrix}\right)}{\sum_{j=1}^{n}\exp\left(-\frac{1}{2\lambda^2}\begin{bmatrix}(X_{t-1}-x_{i-1})\\(Z_t-z_i)\end{bmatrix}^T\begin{bmatrix}S_{22} & S_{2z}\\S_{2z} & S_{zz}\end{bmatrix}^{-1}\begin{bmatrix}(X_{t-1}-x_{i-1})\\(Z_t-z_i)\end{bmatrix}\right)};$$

b_i is the conditional mean associated with each kernel:

$$b_i = x_i - \begin{bmatrix}S_{12}\\S_{1z}\end{bmatrix}^T\begin{bmatrix}S_{22} & S_{2z}\\S_{2z} & S_{zz}\end{bmatrix}^{-1}\begin{bmatrix}(X_{t-1}-x_{i-1})\\(Z_t-z_i)\end{bmatrix};$$

x_i and z_i represent observations, z_i being estimated using the 12 prior flows as expressed in equation (2).

The conditional probability density estimate in (5) can be viewed as consisting of n kernels having relative areas equal to weight w_i, centered at b_i, and having a spread proportional to S'. Each of these are slices of the trivariate kernels that constitute the joint probability density of (X_t, X_{t-1}, Z_t), along the conditioning plane specified by (X_{t-1}, Z_t). The weight w_i depends directly on how far the kernel is from the conditioning plane. A smaller weight implies that the kernel is far from the conditioning plane and does not make up a significant proportion of the conditional density estimate. On the other hand, a large w_i implies that kernel i is close to the conditioning plane and constitutes a significant portion of the conditional density estimate.

Synthetic streamflow generation from the conditional density in (4) as follows:

Estimate the bandwidth λ and the covariances S_{11}, S_{12}, S_{1z}, S_{22}, S_{2z}, S_{zz}.

Start the simulation by arbitrarily assigning values to X_{t-1} and Z_t

Given X_{t-1} and Z_t, estimate the weight w_i associated with each kernel

Pick a data point i with probability w_i

The new value of X_t can now be obtained as $X_t = b_i + \lambda(S')^{1/2}W_t$ where W_t is a Gaussian random variate with zero mean and unit standard deviation

Increment time step t, X_{t-1} and Z_t

Repeat steps 3 to 6 as many times as required.

In practice the first few values simulated are discarded to reduce the effect of the arbitrary initialisation used. In all results reported next, the number of values discarded was set to 120 (the first 10 years of the simulation).
Because of the smooth and non-varying nature of the kernel function used, if an excessive percentage of observed data points lie on or close to the zero flow boundary, it can result in a significant amount of probability to be associated with negative (hence infeasible) values of flow. To get around this problem, a "variable kernel" [*Scott*, 1992] has been used for data points close to the boundary. The bandwidth or spread of the conditioned kernel slice is reduced depending on how far its center is from the left boundary. In addition to the variable kernel approach, a special provision has been made to cater for any zero flows that may be present in the data. The simulated value is set equal to zero when the selected kernel slice is associated with a zero-valued data point. Our rationale for using this approach is that the zero-flows represent a “state” in the flow system for a particular river, which cannot be simulated if any continuous probability density function is used to represent the flows. Obviously, this condition is never used if the data contains no zero flows, and is used only for streams of an ephemeral nature.

Readers should note that the model proposed here is similar to the NP1 model of [*Sharma et al.*, 1997], except that the proposed model uses an aggregate flow variable in addition to the previous month’s flow as the two model predictors. The use of the aggregate flow variable is to impose a longer term dependence in the simulated flows. Such dependence is missing in the NP1 or any other Markov order 1 dependence models. To distinguish between the NP1 model of [*Sharma et al.*, 1997] and the nonparametric simulation model proposed here, the following convention will be used: the NP1 model of [*Sharma et al.*, 1997] with no long term dependence will be denoted ($NP1_{no\_longterm}$) whereas the nonparametric model proposed here will be denoted ($NP1_{longterm}$) in the discussions that follow.

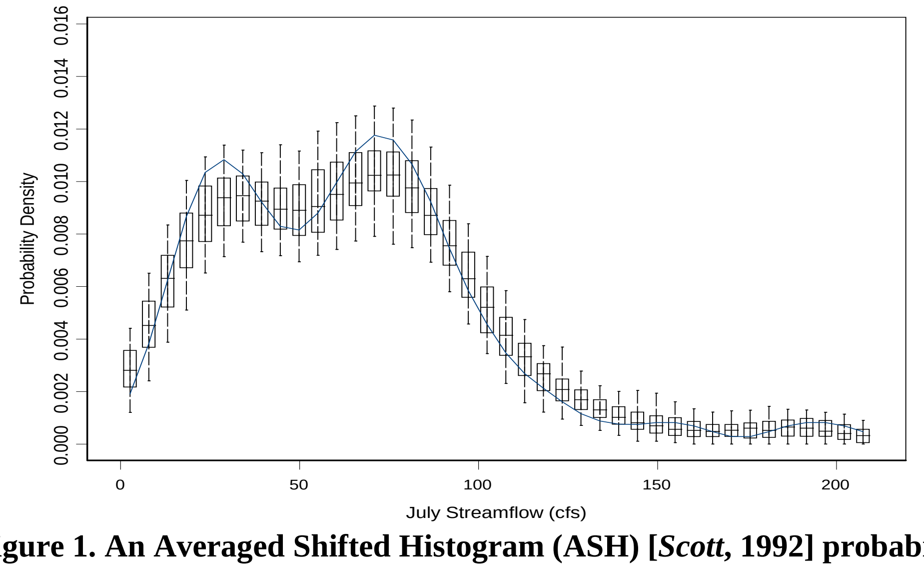

Figure 1. An Averaged Shifted Histogram (ASH) [*Scott*, 1992] probability density estimate of the July month flows in Beaver River near

Beaver, Utah, USA. Boxplots represent the PDF estimates for 100 NP1$_{longterm}$ model simulations.

APPLICATION TO MONTHLY STREAMFLOW FOR BEAVER RIVER NEAR BEAVER, UTAH, USA

Eighty-four (84) years (October 1914 to September 1998) of monthly streamflow data from the Beaver River near Beaver, Utah, USA (USGS station number 10234500) was used to test the applicability of the NP1$_{longterm}$ simulation model. This station is at 6200 feet above MSL and represents a total catchment area of 91 square miles. This data has been used in earlier studies [*Sharma et al.*, 1997; *Tarboton et al.*, 1998] illustrating the use and applicability of nonparametric techniques, for reasons evident in the probability density functions for the observed and simulated data sets for the month of July shown in figure 1. The July month streamflow has a clearly bimodal probability density function, which is difficult to model using conventional stochastic techniques. The boxplots in figure 1 represent the variability in the probability density function of 100 flow sequences, each of an 84 year length, simulated using the nonparametric simulation model described earlier. As one can infer from the figure, the nonparametric model is able to represent the bimodality is a reasonably accurate manner. This would not have been possible through the use of conventional parametric approaches.

The monthly mean, standard deviation, coefficient of skewness, and lag 1 correlations of the simulated sequences were estimated and found to compare well with that of the observed flows. It was interesting to note that the annual flow standard deviations and coefficients of skewness was also reproduced rather well by the NP1$_{longterm}$ model simulations. These results are not reproduced here for lack of space, but are available from the authors on request.

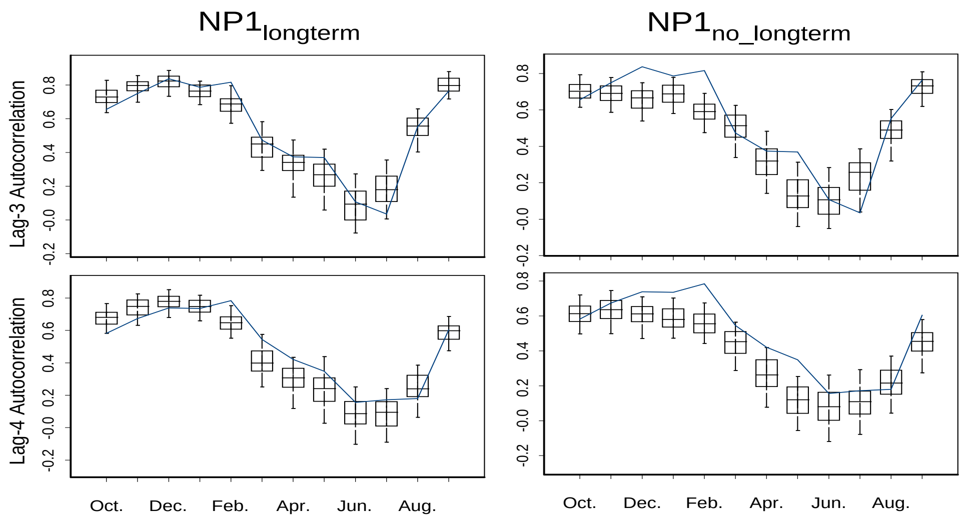

Figure 2. Lag 3 and 4 Auto-Correlations for Beaver River monthly flows.

Figure 2 illustrates the lag 3 and lag 4 correlations of the observed and simulated datasets. Also shown are lag 2 and lag 3 correlations of the $NP1_{no\_longterm}$ model. The use of the aggregate flow variable (the running sum of the last 12 months of flow) in $NP1_{longterm}$, enables the higher lag correlations to be represented well. This is an important result, as the model was not designed to ensure accurate reproduction of these higher lag correlations. One would need to have a higher order Markov dependence structure in a conventional stochastic simulation model (an Auto-regressive lag 4 (AR4) or its non-parametric equivalent, the NP4 model) to achieve the same results.

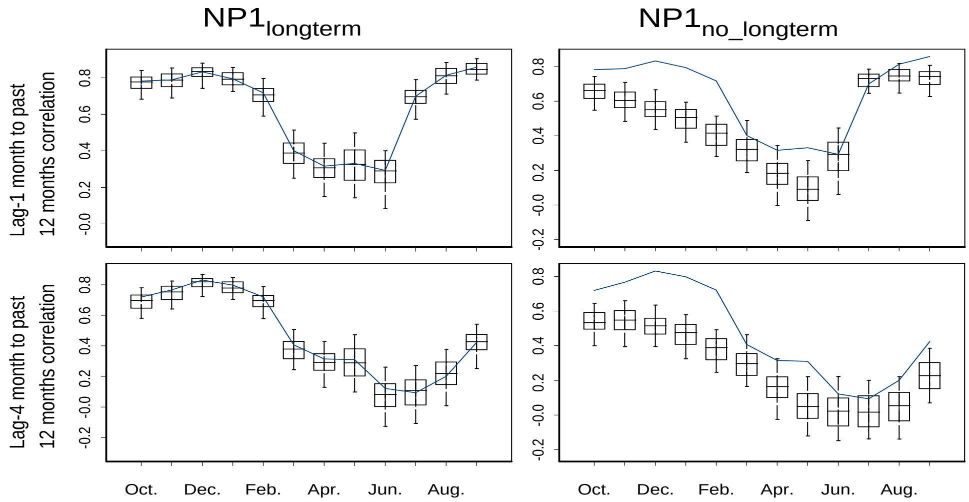

Figure 3. Lag 1 and 4 month to 12 month aggregate flow correlations.

Figure 3 illustrates the correlation between one month's flow and the sum of the previous 12 months flows, for lags of 1 and 3. Similar results are also shown for simulations using the $NP1_{no\_longterm}$ model where the longer term dependence is not explicitly modelled. As would be expected, the proposed model reproduces well the correlations for a lag of 1. Higher lag correlations are also well reproduced. The $NP1_{no\_longterm}$ model results are not as encouraging as would be expected for any Markov order 1 dependence model. This is an important result as one could now expect the simulations be perform better at reproducing drought or storage related statistics where longer term dependence matters most.

To understand better how well the model reproduces dependence statistics at an annual level, annual flows were estimated by adding the monthly simulated flow values for each water year. Annual flow lag-1 autocorrelations for the historical and simulated flow values are presented in table 1. Note how well the annual lag-1 correlations are simulated by the $NP1_{longterm}$ model. This is an important result as it represents how capable the monthly simulation model is at modelling longer (annual) term dependence.

Table 1. Lag-1 autocorrelations for observed and simulated annual flow values.

Historical Flows		$NP1_{longterm}$	$NP1_{no\_longterm}$

0.307	25^{th} %ile	0.204	0.030
	Median	0.306	0.081
	75^{th} %ile	0.363	0.189

The reservoir storage required to sustain constant monthly demands equal to 0.5, 0.6, 0.7, 0.8 and 0.9 times the observed mean annual flow were estimated using the sequent peak algorithm. These storages are illustrated in figure 4. While both models perform reasonably well, the bias and variance of the storages calculated for the $NP1_{longterm}$ model simulations is marginally smaller than that for the $NP1_{no\_longterm}$ model simulations.

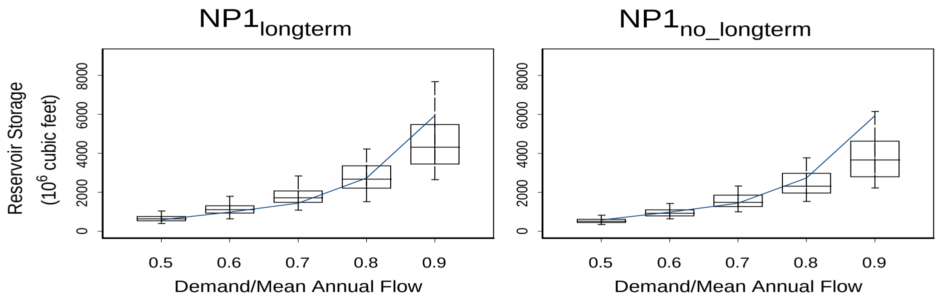

Figure 4. Reservoir storage volume required to meet constant monthly demands.

APPLICATION TO BURRENDONG DAM INFLOWS

The nonparametric simulation model was next applied to 105 years (1890 to 1994) of reservoir inflows to the Burrendong dam in eastern NSW. The Burrendong dam is located on the Maquarie River and has an approximate catchment area of 7500 km^2. While flow data has been measured since the opening of the dam in 1967, the earlier periods of record have been estimated by the Department of Land and Water Conservation using the observed rainfall record and a Sacramento rainfall-runoff model. This streamflow data poses a few problems to the stochastic modeller. Firstly, the there are several instances where the flow has stayed at fairly low levels for 6-10 months at a stretch. Secondly, there are several 'zeroes' in the flow record, which always pose a few challenges when prescribing a continuous probability density function. And lastly, these flow records are known to be highly related to long-term climatic anomalies such as the El Niño Southern Oscillation, hence containing periods of significantly low flows at frequencies of 3-5 years. Hence, whether using a single aggregate variable (sum of past 12 month flows) in the $NP1_{longterm}$ approach will be sufficient for modelling this record is a difficult question to answer.

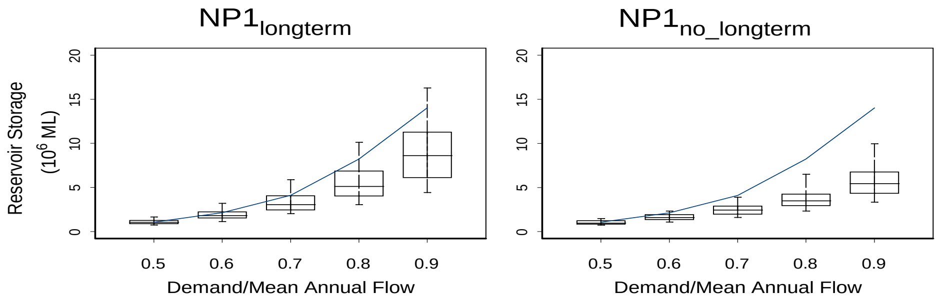

Figure 5. Reservoir storage volume required to meet constant monthly demands.

One hundred realisations each 105 years long were simulated using the two nonparametric stochastic streamflow generation models. All the statistical comparisons presented in the earlier section were found to produce similar results in case of the Burrendong dam inflows data. Hence these results are not being reproduced here but are available on request.

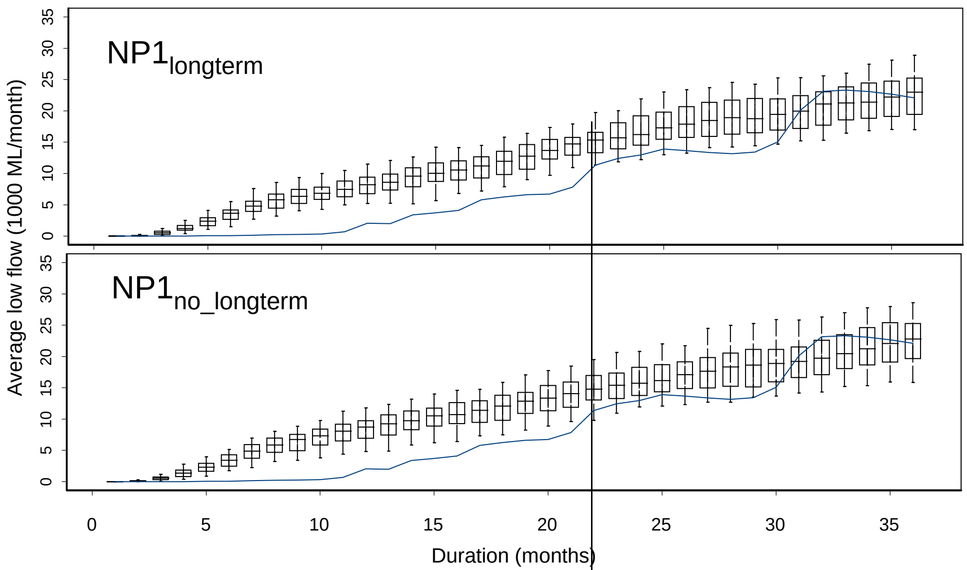

Figure 6. Variation of observed and simulated low flows with duration.

Figures 5 and 6 illustrate reservoir storage related aspects of the two simulation models. Figure 5 shows the reservoir storages estimated based on the sequent peak algorithm, while figure 6 presents the monthly lowest flows in a sequence (averaged to maintain the same units for all durations considered) as a function of duration. Only durations shorter than 3 years are shown as the results for longer durations were found to be satisfactory for both the models. The $NP1_{longterm}$ model simulations lead to reservoir storage volumes that are more representative of the observed data, as compared to the case of the $NP1_{no\_longterm}$ model simulations. However, none of the models is able to properly simulate the smallest low-flow sequence (up to a duration of 21 months) in the observed record. It is likely that a separate aggregate flow variable representing a longer period of aggregation is needed to properly model this highly complex flow record. It is also possible, that more than one aggregate flow variables, or, more than one short term variables might be required to properly model this data. Future studies will investigate the effect of such choices on the model results.

SUMMARY

A synthetic streamflow generation model capable to modelling both short term and inter-annual dependencies, as well as non-standard probability density functional forms, was presented. This model was tested on two monthly streamflow data sets representing very different climatological and topographical regimes. The results indicated that the proposed nonparametric model is able to represent longer-term dependence features in a much better way as compared to models that do not take longer-term memory into account. These results were encouraging as the model used arbitrary choices for the lag of the short term dependence (1 month) and the period of aggregation for the longer-term dependence (12 months) variable. It is likely that results will be even better if these choices are based on a careful analysis of the dependence features present in the data.

References

Akaike, H., A new look at the statistical model identification, IEEE Transactions on Automatic Control, AS-19 (6), 716-723, 1974.

Bras, R.L., and I. Rodriguez-Iturbe, Random functions and hydrology, 559 pp., Dover Publications, Inc., New York, 1985.

Craven, P., and G. Wahba, Smoothing noisy data with spline functions, Numerical Mathematics, 31, 377-403, 1979.

Lall, U., Recent Advances in Nonparametric Function Estimation, in Reviews of Geophysics, pp. 1093-1102, 1995.

Lall, U., and A. Sharma, A nearest neighbor bootstrap for time series resampling, Water Resources Research, 32 (3), 679-693, 1996.

Scott, D.W., Multivariate Density Estimation: Theory, Practice and Visualisation, 317 pp., John Wiley & Sons Inc., New York, 1992.

Sharma, A., U. Lall, and D.G. Tarboton, Kernel bandwidth selection for a first order nonparametric streamflow simulation model, Stochastic Hydrology and Hydraulics, 12, 33-52, 1998.

Sharma, A., D.G. Tarboton, and U. Lall, Streamflow Simulation : A Nonparametric Approach, Water Resources Research, 33 (2), 291-308, 1997.

Tarboton, D.G., A. Sharma, and U. Lall, Disaggregation Procedures For Stochastic Hydrology Based On Nonparametric Density Estimation, Water Resources Research, 34 (1), 107-119, 1998.