Sl. no. 34

FINAL COMPLETION REPORT

on

RESERVOIR PERFORMANCE ANALYSIS USING STOCHASTIC STREAMFLOW MODELS

Sponsored by

Indian National Committee on Hydrology

(Ministry of Water Resources, New Delhi)

Submitted by

Dr. K. Srinivasan

Principal Investigator

and

Dr. B.S Thandaveswara

Co-Principal Investigator (Retired Professor)

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

CHENNAI – 600 036 (TN) INDIA

January 2010

FINAL COMPLETION REPORT

on

RESERVOIR PERFORMANCE ANALYSIS USING STOCHASTIC STREAMFLOW MODELS

Sponsored by

Indian National Committee on Hydrology

(Ministry of Water Resources, New Delhi)

Submitted by

Dr. K. Srinivasan

Principal Investigator

and

Dr. B.S Thandaveswara

Co-Principal Investigator (Retired Professor)

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

CHENNAI – 600 036 (TN) INDIA

January 2010

Department of Civil Engineering
Indian Institute of Technology Madras

1. NAME AND ADDRESS OF THE INSTITUTE:

Indian Institute of Technology Madras
Chennai – 600 036.

2. PI (NAME & ADDRESS) : Dr. K. Srinivasan
Professor,
Dept. of Civil Engineering
IIT Madras, Chennai – 600036
Tel.: 044-22574269
E-mail: ksrini@iitm.ac.in

Co-PI (NAME & ADDRESS) : Dr. B. S. Thandaveswara
Retired Professor (retired from service on 30-th June 2005)
Dept. of Civil Engineering
IIT Madras, Chennai – 600036

3. TITLE OF THE SCHEME : **Reservoir Performance Analysis Using Stochastic Streamflow Models.**

BROAD AREA OF RESEARCH : Water resources Planning & Management
SUB AREA : Reservoir Performance Analysis

4. FINANCIAL DETAIS - PROJECT DATA

Date of Start : 09/04/1997 Total cost of project : Rs. 6,30,000
(MOWR No. 15/1/96-R&D/229 dt. 07/02/1997)

Amount Released: Total: 4.14 lakhs in two instalments.

I instalment of the same, Rs.2.20 lakhs was released in Feb.1997.

II instalment of the grant for the above scheme has been recently effected by MOWR through a demand draft for an amount of Rs.1,94,000/- DD No.992919 dt. 24.11.2004

Unspent Balance at the end of the project period: Rs. 59159.00.
Return of the Unspent balance: A DD for Rs. 59159/- bearing number 023188 dated 01/09/2006 was returned to the MoWR from IIT Madras.

5. ORIGINAL OBJECTIVES OF PROPOSAL

I. Development of a versatile user-friendly software for periodic stochastic modeling of river flows, with built-in decision-aids at various stages of modeling.

II. Using the stochastic model fitted, large number of similar flow sequnces would be generated, which would be useful in : (a) fixing the required capacity of a reservoir being planned; (b) evaluating the performance of any existing reservoir system, for various existing and projected target demands and different operating policies.

III. Establishing trade-off relationships amongst the performance indicators such as reliability, resilience and vulnerability for a few existing systems, for standard operating policy and a few selected hedging policies.

IV. Construction of reservoir storage-performance-yield (S-P-Y) relationships and isoperformance lines for selected operating policies for reservoir systems.

V. Demonstration of the practical usefulness of these tradeoff relationships developed in making decisions regarding reservoir planning, design and operation.

6. No changes in the primary objectives during the operation of the scheme.

7-10 and 14: Have been dealt with in detail in the 208-page report enclosed.

11. No field tests were relevant to the project and hence not conducted.

12. The source programs developed for Periodic Stochastic Modeling of Streamflows, the executables, the sample input and output files required to run the programs, and the documentation regarding the above are provided in a CD (soft copy).

13. There may not be any possibility of patents/copyrights since the work is not product-oriented.

K. Srinivasan

Date : 31-01-2010

(K. Srinivasan)

Principal Investigator

Dr. K. SRINIVASAN
Professor
Water Resources Engg. Div.
Environmental & Water Resources Engg. Div.
Department of Civil Engineering
Indian Institute of Technology Madras
Chennai-600 036, India

Remarks of INC

Signature & Seal of the INC Member-Secretary

CHAPTER 1

INTRODUCTION

Traditionally, linear autoregressive moving average (ARMA) models of Box-Jenkins type have been used to model streamflows at single/multiple sites at the annual/periodic levels. The popularity of linear ARMA models for hydrologic time series analysis may be due to their simplicity and the availability of a well-developed modelling framework in the statistical literature for stationary processes and the availability of standard software packages. In case of parametric models, bias corrections done to unbias skewness and/or correlations, estimated from small size samples, often cause some other undesirable effects in the synthetic replicates generated. Transforming the non-normal historical flows to normal, may lead to distortion of the correlation structure in the generated flows. Moreover, the traditional parametric modeling framework directs the researcher's attention towards an efficient estimation of model parameters under a certain metric for the selected model form. Due to these limitations, researchers are attempting to propose alternative models that can perform better in modeling the streamflows.

Recently, Srinivas and Srinivasan (2000, 2001) have introduced the post-blackening approach (Davison and Hinkley, 1997) for stochastic modeling of streamflows that exhibit both linear and nonlinear dependence present in the streamflow data. Their model is basically a hybrid model that blends a linear parametric model with the moving block bootstrap (MBB), a non-parametric model. Although it is a certain improvement over the traditional ARMA models, it requires further improvement in capturing the non-linear effects. With a view to address this limitation, it may be worthwhile to go for completely

data-driven models. Accordingly, the Artificial Neural Network (ANN)-based hybrid model is being proposed in this study for modeling the kind of annual streamflow data that exhibit a complex dependence structure, which is a blend of two non-linear data-driven models, namely, ANN and Moving Block Bootstrap (MBB).

Seasonality of streamflow data adds a degree of complexity to the selection of an appropriate stochastic model to fit the data. However, if the autocorrelation structure of the observed data exhibits significant periodicity, then seasonal models that explicitly incorporate a periodic dependence structure must be used (Rasmussen et al., 1996). Varying degree of nonlinearity in the different periods/seasons, generation of streamflows owing to mixed precipitation mechanisms, add to the complexity further. Such characteristics of the geophysical time series make the modeling of multi-season streamflows a challenging task.

An ideal single site multi-season synthetic flow generation model should aim to reproduce: the summary statistics (mean, standard deviation and skewness) and marginal distribution of observed flows at periodic and annual time scales; autocorrelation structure of flows at aggregated annual level; within-year and cross-year serial correlations; month-to-year cross-correlations; and non-linearity stationarity in the underlying dependence structure. In addition, it should provide sufficient variety in the stochastic simulations with a reasonable degree of smoothing and extrapolation. The motivation for the method of periodic streamflow modeling presented in this research work comes from a desire to develop a potential nonparametric stochastic model that is

effective in reproducing summary statistics, dependence structure and the salient features of marginal distribution without compromising on smoothing, extrapolation and variety in simulations. Such an ideal model is expected to be effective in predicting storage capacity and critical run characteristics that are of interest to the investigator.

Based on storage capacity, inflow pattern and demand, the reservoir systems can be classified as over-year (or carry-over) and within-year systems. Within-year systems are sensitive to seasonal variations of both inflow and draft. Studies that model the within-year Storage-Performance-Yield (S-P-Y) relationships are more realistic. However, these relationships are difficult to generalize due to the large number of parameters associated with periodic stochastic streamflow models. Reservoirs, in which filling and emptying phases do not take place on an annual basis, but over a number of years, are known as over-year reservoirs, in which over-year storage effects predominate. Whenever severe, long-stretched deficits (shortages) in water supply are to be handled in a river system, carry-over storages become important and high storage capacities are provided for the reservoirs in such systems.

The operational performance of a water supply reservoir is usually expressed in terms of performance indicators that describe the failure characteristics, namely the frequency, the duration and the severity of failures. That is, reliability, resilience and vulnerability together characterize "risk" in the reservoir planning and operation context. The Storage-Performance-Yield (S-P-Y) relationships are useful in identifying the sensitive ranges of storage capacity of the over-year reservoirs, with regard to performance characteristics; and in selecting between capacity expansion and demand management options, in case of

deficit water supply systems. Such relationships need to be developed for over-year reservoir systems, in general and on a case by case basis for within-year reservoir systems.

During drought periods and even when drought is impending, effective demand management strategies must be devised to reduce the severity of shortage by distributing the deficits over longer periods. Hedging is one of the simple and common demand management strategies employed by water supply managers to reduce the severity of droughts. Hedging increases water stored in the reservoir by accepting small currents deficits to guard against unacceptable large deficits that may occur in future. Hedging rule decides the storage allocation of water across time to minimize the impact of the drought.

Although there have been a number of hedging rules proposed in the literature, such as the discrete and the continuous hedging rules proposed by Shih and Revelle (1994, 1995), supply and demand based hedging rules such as the one proposed by Srinivasan and Philipose (1998), a relative evaluation of these rules and their adequacy in terms of performance during water shortages have not been analyzed in detail. There is also a need to obtain optimal hedging policies using each of these rules and investigate them by detailed evaluation by simulation. Also, it may be worthwhile to propose a new hedging rule that will improve the overall performance of both over-year and within-year water supply reservoirs.

CHAPTER 2

NONLINEAR DATA-DRIVEN MODEL FOR ANNUAL STOCHASTIC STREAMFLOWS

2.1 INTRODUCTION

Traditionally, linear autoregressive moving average (ARMA) models of Box-Jenkins type have been used to model streamflows at single/multiple sites at the annual/periodic levels and the same have been described at length in texts on the subject (e.g., Salas et al., 1980; Salas, 1993; Loucks et al., 1981). The physical justification for the use of ARMA models in the context of modelling the annual streamflows is described in Salas et al. (1980). The annual streamflows are represented by mixed autoregressive and moving average processes. The flow recession during dry periods that may have significant persistence and small variation, can be represented by autoregressive (AR) processes. The high flows due to large rainfall and/or snowmelt could be represented through addition of the Moving Average (MA) component. The popularity of linear ARMA models for hydrologic time series analysis may be due to their simplicity and the availability of a well developed modelling framework in the statistical literature for stationary processes and the availability of standard software packages such as STATGRAPHICS (1984), IMSL (1984), SAS (1988), WASIM (McLeod and Hipel, 1978), LAST (Lane and Frevert, 1990), SPIGOT (Grygier and Stedinger, 1990), and CSUPAC1 (Salas et al., 1992). However, there are a number of drawbacks of the Box-Jenkins type of models as pointed out by Lall and Sharma (1996), Tarboton et.al (1998) and Srinivas and Srinivasan (2000). Furthermore, in case of parametric models, bias corrections are often applied to unbias skewness and/or correlations, estimated from small size samples. These corrections in turn give rise to some other undesirable effects in the

synthetic replicates generated (for instance, unbiasing skewness may result in increased variance of the same; transforming the non-normal historical flows to normal, may lead to distortion of the correlation structure in the generated flows). Moreover, the traditional parametric modeling framework directs the researcher's attention towards an efficient estimation of model parameters under a certain metric (e.g., least squares or maximum likelihood) for the selected model form. On the other hand, the performance metric of interest to the hydrologist or the water resources planner need not be optimal, for the same. This is one of the main reasons to consider the usage of flexible, adaptive, data exploratory methods, instead. Moreover, classical ARMA models are optimal only under squared error loss and only for linear operations on the variables. While, the risk/loss functions associated with hydrologic decisions are known to be asymmetric. In addition, incorporation of parameter uncertainty into the parametric time series models (Stedinger and Taylor, 1982; Grygier and Stedinger, 1990), is quite involved and not that simple to be understood or applied by practising hydrologists.

Despite making a large family, all Box-Jenkins models are essentially of short-memory type; that is, their autocorrelation structure decreases rapidly with the lag time. Hence, such models are inadequate in stochastic hydrology, if the long-term persistence of hydrologic (and other geophysical) processes is to be modelled. This property discussed by Hurst (1951), is related to the observed tendency of annual streamflows to stay above or below their mean value for long periods. As a result, these model are not able to predict critical run characteristics efficiently.

Other classes of models such as fractional Gaussian noise (FGN) models (Mandelbrot and Wallis, 1969a,b,c), fast fractional Gaussian noise models (Mandelbrot, 1971), broken line models (Mejia et al., 1972) are more appropriate to model long-term persistence (Bras and Rodriguez-Iturbe, 1985, pp.210-280). However, FGN and FFGN models have several weak points such as parameter estimation problems, narrow range of autocorrelation functions that they can preserve, and their inability to perform in multivariate problems (Koutsoyiannis, 2000).

While nonlinear models (Bendat and Piersol, 1986; Tong, 1990) can be used in place of the linear ARMA models, these nonlinear models require specifying the form of nonlinear dependence prior to the parameter estimation which may not be easy for the practitioner. Moreover, in a multi-site modeling context, this becomes a tedious exercise, especially if the model structure and the probability distributions that describe the flows at the different sites vary. From the practitioner's perspective, the key issues are reproducibility of the observed data characteristics, simplicity, dependability and robustness. Owing to the difficulties associated with the parametric methods in terms of parameter estimation, assumptions regarding the marginal probability distributions and the dependence structure of the variable of interest, nonparametric methods are becoming popular in stochastic hydrology in the last one decade. Readers are referred to Lall (1995) for an overview of nonparametric applications to hydrology.

The bootstrap (Efron, 1979) is the simplest nonparametric technique for simulating the probability distribution of any statistic. The use of bootstrap methods in time series analysis is receiving considerable attention in modern statistics, as documented by Lepage and Billard (1992), Efron and Tibshirani (1993), and Davison and Hinkley

(1997). Künsch (1989) proposed the moving block bootstrap (MBB) for resampling dependent data. This technique consists of dividing the data into blocks of observations and resampling the blocks randomly with replacement. In this method, though the original dependence structure is preserved within the blocks, it is lost at boundaries between blocks (Davison and Hinkley, 1997). This poses difficulty in preserving the dependence structure present in hydrologic records, as the available sample sizes are small. Lahiri (1993) addresses the limitation of MBB in statistical literature. In hydrologic literature, Srinivas and Srinivasan (2000, 2001) bring out the inefficiency of MBB in simulating streamflows at annual and periodic time scales.

In the last decade, Lall and Sharma (1996) proposed k-nearest neighbor (*k*-NN) bootstrap for resampling dependent hydrologic data. Multivariate nearest neighbor probability density estimation provides the basis for the resampling scheme. A discrete kernel is used to resample from the successors of k-nearest neighbors of the conditioning vector (Rajagopalan and Lall, 1999; Sharma and Lall, 1999; Kumar et al., 2000). The nearest neighbor bootstrap and its variations may be preferable if the data are plentiful, as in case of daily streamflow modeling (Lall and Sharma, 1996). The investigations by Srinivas and Srinivasan (2001) report that for historical time series with strong dependence, the *k*-NN model does not simulate the run characteristics satisfactorily (validation statistics as per Stedinger and Taylor, 1982), plausibly due to inadequate preservation of higher lag serial correlations.

A limitation of the aforementioned NP methods is that simulations from these resampling methods can neither fill in the gaps between the data points in the observed record nor extrapolate beyond the observed extrema. In the last decade, kernel-based nonparametric methods were proposed for streamflow simulation (Sharma et al.,

1997), streamflow disaggregation (Tarboton et al., 1998) and for generation of multivariate weather variables (Rajagopalan et al., 1997) with a view to alleviate the limitation of the bootstrap methods. However, these methods demand considerable computational effort for the estimation of bandwidth in higher dimensions. Moreover, the kernel methods suffer from severe boundary problems, especially in higher dimensions, that can bias the simulations (Prairie, 2002).

Recently, Srinivas and Srinivasan (2000, 2001) have introduced the post-blackening approach (Davison and Hinkley, 1997) for stochastic modeling of streamflows that exhibit complex dependence. Herein, complex dependence refers to both linear and nonlinear dependence present in the annual streamflow data. The different segments of the data based on thresholds (such as low flow, medium flow and high flow) may exhibit different dependence structures within the segments as well as between the segments. The model proposed by Srinivas and Srinivasan (2000) is basically a hybrid model that blends a linear parametric model with the moving block bootstrap (MBB), a non-parametric model. In this modeling approach, partial pre-whitening of the streamflow data is done by the linear parametric model, followed by block bootstrapping of the residuals extracted. This kind of blending was done with a view to: (i) capture both short- and long-term dependence characteristics in the observed streamflow data that are important for the prediction of storage and drought related characteristics; and (ii) introduce some smoothing/extrapolation value in the synthetic simulations. Srinivas and Srinivasan (2000) report a better preservation of the dependence structure and the critical drought characteristics compared with linear parametric models for annual streamflow data that have complex dependence structure.

Nevertheless, results presented by Srinivas and Srinivasan (2000) (*e.g.* figures 14, 16, 18, 20, 22 and 23) suggest scope for improvement in simulating the drought characteristics at various truncation levels (expressed as percent of mean annual flow). This is plausibly due to the complex dependence (defined in the previous paragraph) present in the streamflow data series not being modeled effectively by the linear parametric based hybrid model. We feel that such complex dependence present in the streamflow data can be more effectively captured by better modeling of the non-linearities inherent in the streamflow data, which may be accomplished by completely data-driven models. This prompted us to go for the Artificial Neural Network (ANN) based hybrid model, which is a blend of two non-linear data-driven models, namely, ANN and Moving Block Bootstrap (MBB) for modeling the kind of annual streamflow data that exhibit a complex dependence structure.

Artificial Neural Network (ANN) is a data-driven method of computation and information processing that takes advantage of mimicking the processes of biological neurons found in human brain. Over the last 2 decades, the ANNs are being successfully applied across a wide range of problem domains, as diverse as finance, medicine, engineering, geology, hydrology and physics (*e.g.* Buscema and Sacco, 2000; DeRoach, 1989; Gernoth et al., 1993). This data-driven technique is now widely accepted as a potentially useful way of modeling complex non-linear and dynamic systems, especially in situations where the underlying physical processes are not fully understood (Hsu et al., 1995) or where the nature of the event being modeled may display chaotic properties. The emergence of ANN technology has made a mark in the field of hydrology, wherein uncertainty rules. Some of the earliest applications of ANN to hydrology were reported by Daniel (1991), who also suggested further

possible scope. The approach has since then been used in various hydrological studies (Hsu et al., 1995; Dawson and Wilby, 1998; Campolo et al., 1999; Sudheer et al., 2002; 2003; Sudheer and Jain, 2004; Jain and Srinivasulu, 2004, to mention a few). For a comprehensive review of ANN applications to hydrology, the reader is referred to the reports of ASCE Task Committee (2000a, b), Maier and Dandy (2000) and Dawson and Wilby (2001).

It is observed that most of the previous works related to ANN reported in the hydrologic literature have addressed river flow forecasting problems, since this data-driven technique is very much suited for function approximation when the underlying physics of the process is unknown. Rivera et al. (2002), has reported the usage of ANN for stochastic streamflow generation, although not many works addresses this application. In their work, the hybrid model adopted for multi-site stochastic streamflow generation using ANN is very similar to the traditional modeling approach with the exception that the linear parametric model in the traditional modeling approach is replaced by an ANN. The results presented in Rivera et al. (2002) indicate that though the model offers marginal improvement over the linear parametric model (AR(2)) in predicting various long-term validation statistics such as drought and storage characteristics, it behaves in a very similar way as the linear parametric model. It is evident from the Fig. 9 presented in Rivera et al. (2002) that this multivariate ANN based hybrid model is not able to capture the jumps and variations in the drought characteristics with regard to various truncation levels (that define the streamflow drought) in case of both the examples considered. This observation indicates that the ANN alone may not be able to capture the entire dependence information present in the data, thus leaving some information (including non-linearities) hidden in the

residuals. We feel that this residual information can be effectively modeled by employing a nonparametric model such as the moving block bootstrap that is expected to capture the weak linear as well as the nonlinear dependence and any distributional information retained in the residuals. This leads to the focus of the current research work, namely, streamflow generation using a hybrid data-driven model. A later section of this chapter illustrates with the help of three unregulated annual streamflow records (that exhibit complex dependence structure) that the effective blend of the two data-driven models referred, enables efficient simulations of the long-term storage and drought-related characteristics.

2.2 HYBRID MODEL DEVELOPMENT

2.2.1 ANN Based Hybrid Model (ANNHM)

As discussed earlier, the basic idea behind the ANNHM model is to blend two nonlinear data-driven models (ANN and MBB). The choice of network type and its functioning as well as training are discussed in this section. The algorithm for blending the MBB with ANN is also presented.

2.2.1.1 *Choice of network type*

The most popular ANN architecture in hydrologic modeling is the multi-layer perceptron (MLP) trained with the back propagation (BP) algorithm (ASCE, 2000a). Although MLP can produce accurate results, it has several drawbacks, such as a long training time, and the BP algorithm being a gradient descent method may converge to a local minimum (Sudheer and Jain, 2003) resulting in a suboptimal solution to the problem. Moreover, when the data are limited, the BP algorithm may not lead to good generalization properties for the network. Another popular ANN architecture, radial

basis function (RBF) network can be trained in a shorter ime and has fewer parameters. In contrast to the MLP, the RBF network has the nonlinearity embedded in the transfer functions of its hidden-layer neurons, which makes the optimization of tunable parameters a linear search. An RBF network can offer approximation capabilities similar to those of the MLP (Chen et al. 1991), provided the hidden layer of the RBF network is fixed appropriately. Some studies (*e.g.*, Fernando and Jayawardena, 1998; Sudheer et. al, 2002) report that RBF predicts river flows more accurately than MLP.

2.2.1.2 *RBF network*

The RBF networks operate quite differently from the multi-layer perceptrons (MLP) that commonly use sigmoid type transfer function. While the structure of the RBF is identical to the MLP, the RBF simulates the unknown function using a network of radial functions in the hidden layer. The nonlinearity within an RBF network can be chosen from a few typical nonlinear functions, and the most common choice is the Gaussian function. The choice of nonlinearity of RBF nodes is not crucial for the performance of the method (Powell, 1987). The hidden layer in an RBF performs a fixed nonlinear transformation with no adjustable parameters and it maps the input space onto a new space. The output layer then implements a linear combination on this new space and the only adjustable parameters are the weights of this linear combiner. These parameters can therefore be approximated using the linear least squares method, which is a significant advantage of using this network (Sudheer et al., 2007).

Mathematically, in an RBF, for the pth input pattern X^p, the response (assuming Gaussian function) of the jth hidden node o_j is of the form

$$o_j = f\left\{\frac{\left\|X^p - U_j\right\|}{2\sigma_j^2}\right\} \qquad (2.1)$$

where $\|.\|$ denotes the standard Euclidian norm; U_j is the center of the j^{th} radial basis function $f(.)$; σ_j is the spread of the RBF that is indicative of the radial distance from the RBF center within which the function value is significantly different from zero. The network output is given by weighted summation of the hidden node responses at each node in the output layer. The output for k^{th} node on the output layer z_{pk} is computed as:

$$z_{pk} = \sum_{j=1}^{L} o_j w_{kj} \qquad (2.2)$$

where w_{kj} is the weight connection between hidden and output nodes and L is the number of radial basis functions.

2.2.1.3 *Training an RBF network*

Training an RBF network involves determining the radial basis functions on the hidden layer nodes and the output layer weights. Determining the RBF function involves finding suitable RBF centers and spreads. Ideally, the RBF networks require that there be as many RBF centers as data points, which is rarely practical in most of the applications as the number of data points is usually large. This also implies unnecessarily large networks and extremely long computation times. Hence many applications suggest optimizing the number of RBF centers that produce the output within reasonable tolerance. Consequently, a variety of techniques have been suggested to optimize the number of RBF centers (*e. g.* Moody and Darken, 1989; Chen et al., 1991). The orthogonal least squares (OLS) algorithm proposed by Chen et

al. (1991) has been used in the current research. This algorithm was implemented using the Neural Network Toolbox of MATLAB.

The OLS method involves the transformation of the set of regressors into a set of orthogonal basis vectors. This enables the calculation of the individual contributions to the desired output from each basis vector. It works based on the principle of maximizing the explained variance of the output by the regressor. This is achieved by progressively adding one new basis vector (regressor) during every cycle of the iteration. The algorithm is terminated when the residual error is within the chosen tolerance limit. This gives rise to a subset model containing significant regressors less than the number of patterns presented to the network. For a detailed description of OLS algorithm, readers are referred to Chen et al. (1991).

ANNHM Algorithm

In brief, the steps involved in the ANN based Hybrid Model are as follows (Sudheer et al., 2007)

(i) Consider y_t, t = 1, 2,...,n be the annual streamflow series. Scale the annual flow series to fall within the band [0,1] using any appropriate scaling function. The scaling function used in the current study is:

$$y_{s(t)} = \frac{y_t - \min(y_1, y_2, \ldots, y_n)}{\max(y_1, y_2, \ldots, y_n) - \min(y_1, y_2, \ldots, y_n)} \qquad (2.3)$$

in which $y_{s(t)}$ represents the scaled annual flow series.

(ii) Develop an RBF network to the scaled annual streamflow data to extract the dependence structure partially, adopting the procedure described earlier:

$$\hat{y}_{s(t)} = \{w\} y_{t-3(t)} \qquad (2.4)$$

in which {w} represents the RBF network parameter matrix in terms of weights (w_{kj}), centres (U_j) and spread (σ_j). Herein, $\hat{y}_{t(s)}$ is the RBF computed scaled annual streamflow. Herein it is to be noted that the idea of ANNHM is not to capture the entire dependence structure of the flow data series by RBF itself, but to employ RBF for partial extraction of the dependence structure. The remaining part of the structure (weak structure) present in the residuals is expected to be captured by MBB. Hence a rigourous validation of the RBF model is not warranted. However, a reasonable level of generalization property is expected at this stage.

(iii) Convert $\hat{y}_{t(s)}$ into actual flow units $\hat{y}_t$ by performing inverse transformation of the scaling function used in step (i).

(iv) Compute the residuals from the historic sequence,

$$\varepsilon_{t(s)} = y_t - \hat{y}_t \qquad (2.5)$$

(v) Obtain the simulated innovations $\varepsilon_1^*, \ldots, \varepsilon_n^*$ by bootstrapping $\varepsilon_{t(s)}$ using the moving block bootstrap (Künsch, 1989) method. The appropriate block size to be used for resampling the weakly dependent residuals is to be decided.

(vi) Considering the starting value of generated flow series, $y_{0(g)}$, to be the minimum of the original annual stream flow series (*i.e.* $y_{0(s)} = 0$), estimate the subsequent RBF computed scaled generated flow ($\hat{y}'_{t(s)}$) using the RBF network parameters.

i.e. $$\hat{y}'_{t(s)} = \{\mathbf{w}\} y_{0(s)} \qquad (2.6)$$

Note that the starting value of $y_{0(s)}$ can be any arbitrary value, provided sufficient 'warm-up' period is chosen to ensure the stationarity of the generated data set.

(vii) Convert $\hat{y}'_{t(s)}$ into actual flow units $\hat{y}_{t(s)}$ by performing inverse transformation of the scaling function used in step (i).

(viii) Corresponding value of the bootstrapped innovation series $\{\varepsilon_t^*\}$is then added back to the generated streamflow estimated by the ANN model, to obtain the synthetic streamflows $y_{t(s)}$. Thus,

$$y_{t(s)} = \hat{y}_{t(s)} + \varepsilon_t^* \qquad (2.7)$$

(ix) Repeat steps 6 through 8 until the desired length of the synthetic streamflow series is reached.

(x) Evaluate the performance of the ANNHM generated synthetic streamflow series using verification and validation statistics proposed by Stedinger and Taylor (1982).

(xi) If the performance of the model is not satisfactory, then go to step (v) and change the block size used for resampling the residuals, and repeat steps (v) through (x).

(xii) If the model performance is not satisfactory even after trying all the possible alternate block sizes, then go to step (ii) and tune the RBF structure and continue with steps (iii) through (x).

2.2.2 Linear Parametric based Hybrid Model (LPHM) Algorithm

The details of the model formulation of LPHM, which uses the post-blackening approach suggested by Davison and Hinkley (1997) can be found in Srinivas and Srinivasan (2000). However, a brief description of the algorithm for LPHM is presented in the following paragraph, assuming AR(1) to be the underlying parametric model (any other parametric model can be adopted in place).

(i) Fit a default AR(1) model to the centred historical streamflows $y_1, \ldots, y_n$

$$y_t = \hat{\alpha} y_{t-1} + \varepsilon_t \tag{2.8}$$

(ii) Estimate the residuals using the autoregressive parameter $\hat{\alpha}$ estimated from the historical sequence:

$$\varepsilon_t = y_t - \hat{\alpha} y_{t-1} \tag{2.9}$$

(iii) Define the residuals centred around their mean as

$$\varepsilon_{t(C)} = \varepsilon_t - \overline{\varepsilon}_t \tag{2.10}$$

where $\overline{\varepsilon}_t$ is the mean of the residuals ε_t

(iv) Obtain the simulated innovations $\varepsilon_0^*, \ldots, \varepsilon_N^*$ by bootstrapping $\varepsilon_{t(C)}$ using the MBB (Künsch, 1989) method. The appropriate block size to be used for resampling the weakly dependent, centred residuals is to be decided.

(v) The bootstrapped innovation series $\{\varepsilon_t^*\}$ is then post-blackened by applying the estimated model to the resampled innovations, to obtain the synthetic streamflows y_t

$$\hat{y}_{t(g)} = \hat{\alpha} \hat{y}_{t-1(g)} + \varepsilon_t^* \tag{2.11}$$

The starting value of $\hat{y}_{t(g)}$ is taken to be equal to ε_0^* itself (or it can be any arbitrary value). In this case also, the 'burn-in' or 'warm-up' period is chosen to be large enough to ensure that the subsequent values of the synthetic series are essentially stationary.

2.3 EVALUATION OF THE MODELS

When a stochastic stream flow model is to be used for water resources system planning and management, it is to be evaluated through a two-step process: (i) its ability to reproduce the summary statistics, marginal distribution, dependence structure of the

historical flows; (ii) its ability to predict the drought and the storage characteristics. These two steps are known as 'verification' and 'validation' of the stochastic model (Salas et al., 1980; Stedinger and Taylor, 1982). Accordingly in the current study, the means of the relevant statistics (mentioned above) have been computed over 1000 synthetic streamflow sequences generated from ANNHM and LPHM, and compared with their historical counterparts. The variability of the statistics among the generated sequences is evaluated by computing the dispersion/spread of the statistics obtained from the 1000 synthetic sequences. The results of the evaluation of the two models are discussed in detail in a subsequent section.

Herein, we wish to emphasize that as the synthetic series are not generated directly from the ANN models, their separate evaluation is not performed extensively. This kind of evaluation is required only when the ANN model is directly employed for forecasting purpose. Nevertheless, ensuring a reasonable level of generalization performance is desirable. This is accomplished by ensuring that the relative root mean square error (RRMSE) statistic of the training and the validation sets are in the same order. In case of the linear parametric based hybrid model (LPHM), it may be noted that only a partial pre-whitening of the historical flows is done using AR(1), which is followed by resampling of the residuals using moving block bootstrap technique. Hence, there is no need to separately evaluate the model performance at the pre-whitening stage.

2.4 CASE EXAMPLES

The proposed ANN based hybrid model has been applied to three typical unregulated streamflow records; two data records taken from Yevjevich (1967) and the third data

set obtained from USGS (http://nwis.waterdata.usgs.gov/nwis/annual/). The characteristics of the three data sets are summarized in Table 2.1. The example data sets chosen represent different geographic regions, possess different record lengths and varying sizes of the drainage area, magnitudes of flow and skewness of flows. The historical streamflow time series of the three flow records are shown in Figs. 2.1(a), 2.1(b), and 2.1(c), respectively.

Table 2.1 Annual flow data characteristics of rivers selected for the current study.

Name of river	Name of station	State and/or country	Basin area (km^2)	Record duration	Mean discharge (m^3/s)	Coefficient of variation	Skewness
Lake Albert	Mongalla	Sudan	442890	1904 -1952	718.30	0.301	1.723
Red	Wahpeton	ND, USA	10381	1944-2001	18.00	1.733	0.973
Neva	Petrokrepost	USSR	271950	1859-1935	2588.20	0.163	0.437

In the present research work, the ANN models for all the data series are developed using the classical procedure of splitting the data into training set (first 75% of the total data length) and validation set (remaining 25% of the total data length). The Gaussian function has been used as the radial basis function in the current study. The RRMSE statistics are of similar order for both training and validation statistics (0.2097 for training set and 0.2029 for validation set for Lake Albert; 0.6967 for training set and 0.7023 for validation set for Red River; 1.9985 for training set and 2.1325 for validation set for Neva River). Thus error analysis shows reasonably good generalization property for the ANN models. The parameters of the ANN models are then frozen and used in developing the ANNHM models for the respective data series. The chosen final network architectures (in terms of input-hidden-output nodes) of the ANN models are 1-25-1, 1-32-1, 1-42-1 for Lake Albert, Red River, and Neva River respectively.

While developing ANNHM (and LPHM), the block size used for resampling of residuals extracted from the base model is varied from 1 to 15. Corresponding to each block size, one thousand replicates are generated. As mentioned earlier, the synthetic replicates from the ANNHM and the LPHM are evaluated based on the ability to reproduce the various statistics computed from historical flows such as summary statistics, marginal distribution, autocorrelation structure; and the preservation of critical and mean drought characteristics and storage characteristics. The results of the evaluation are discussed in the following section.

2.5 RESULTS AND DISCUSSIONS

2.5.1 Example I: Lake Albert

2.5.1.1 Reproduction of Summary Statistics

The results of the preservation of summary statistics of the historical streamflows by the ANN based hybrid model (ANNHM) and the linear parametric based hybrid model (LPHM) for the Lake Albert for various block sizes are presented in Fig 2.2 for comparison. It can be observed from Fig 2.2 that both the models reproduce the mean of the annual flow series, while there is some bias in reproduction of the standard deviation at lower block sizes. A more significant observation from Fig 2.2 is that skewness is better preserved by ANNHM at all block sizes compared to LPHM. This indicates that ANNHM exhibits a tendency to capture the nonlinearities more effectively utilizing the potential of ANN. This inference is confirmed by the fact that ANNHM is able to capture the skewness present in the historical flows to a reasonable level, even at a block size of 1 (which is in fact an ANN model with random resampling of residuals). On the other hand, LPHM is able to improve the preservation

of skewness at higher block sizes, evidently due to the blending with the nonlinear nonparametric bootstrap method, MBB.

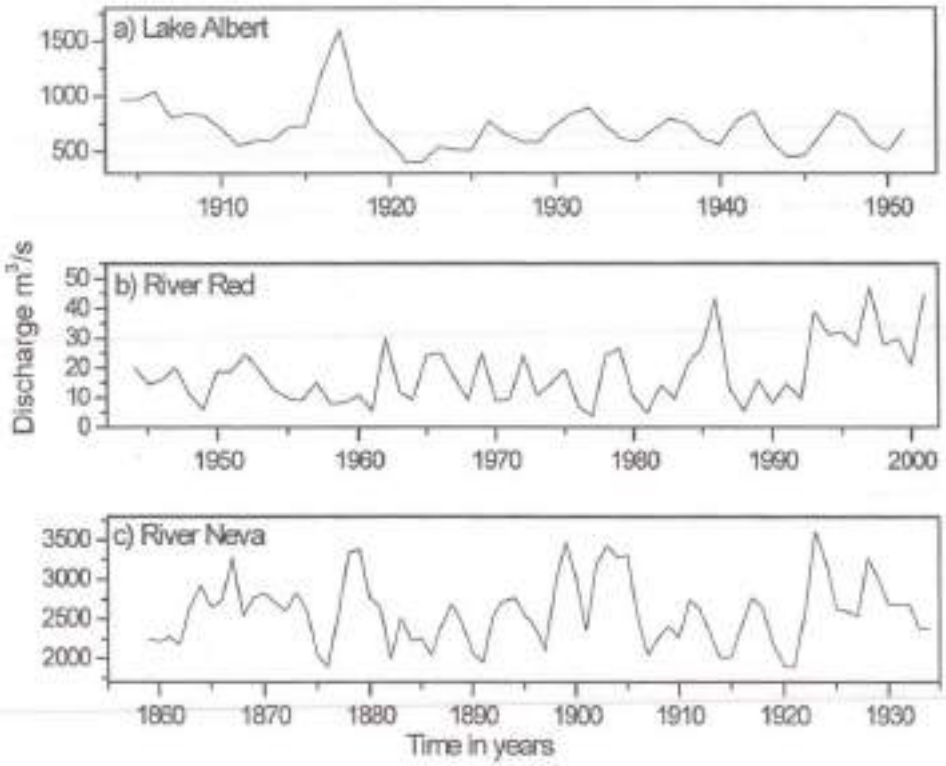


Fig. 2.1. Time series plots of annual streamflows of rivers a)Lake Albert b) River Red c) River Neva

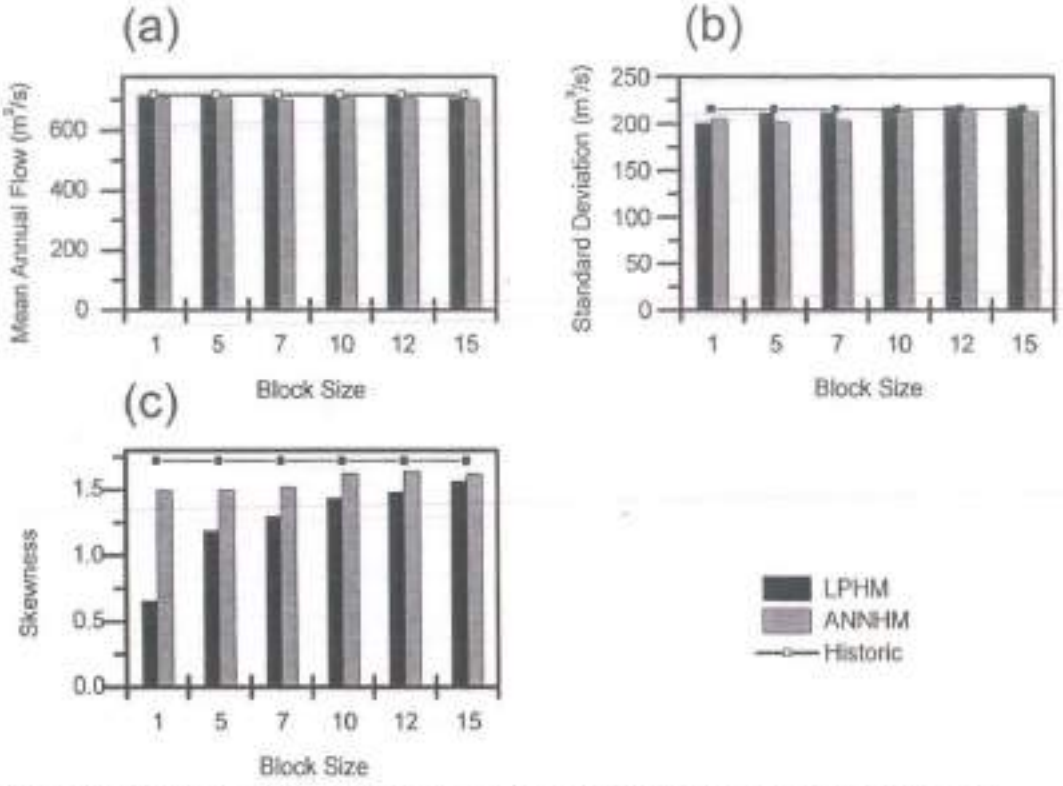


Fig 2.2 Comparison of the reproduction of summary statistics between LPHM and ANNHM: (a) Mean (b) Standard deviation (c) Skewness – River: Lake Albert.

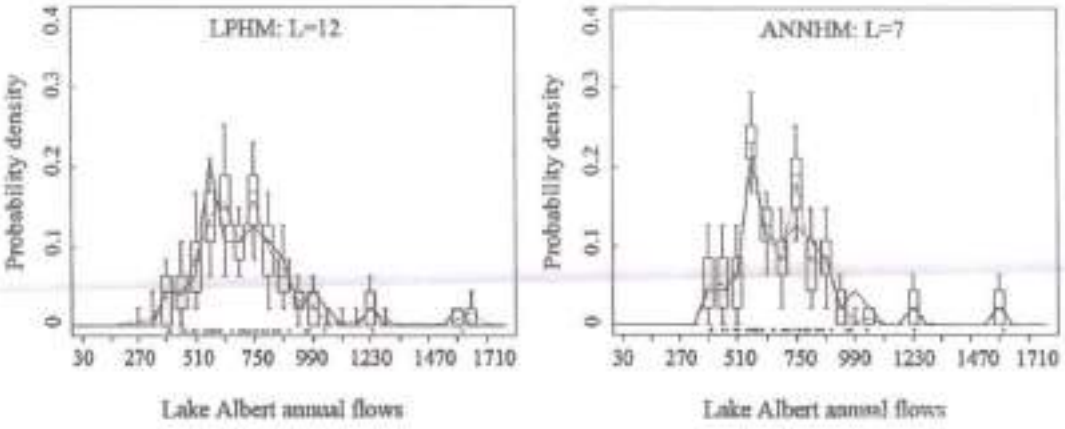


Fig. 2.3. Reproduction of marginal distribution of annual streamflows – A comparison between Linear Parametric Hybrid Model (LPHM) and ANN Hybrid Model (ANNHM). River: Lake Albert. *The dots below the box plots denote the observed flows. L denotes the block size used in MBB.*

2.5.1.2 Reproduction of Marginal Distribution

The reproduction of the marginal distribution of annual streamflows is presented in Fig. 2.3 for the selected block sizes 12 and 7 respectively for LPHM and ANNHM. These block sizes are selected for the respective models based on their performance in terms of prediction of drought and storage statistics (discussed in later sections). It may be observed that LPHM reproduces the features of the marginal distribution reasonably, while providing some smoothing as well as extrapolation value. On the other hand, ANNHM being a completely data-driven model reproduces the features of the marginal distribution more closely, but offers less smoothing and little extrapolation value Fig. 2.3. This limitation needs to be addressed in future research.

2.5.1.3 Preservation of Dependence (autocorrelation) Structure

The preservation of the autocorrelation structure of the historical flows by both the hybrid models is presented in Figures 2.4 and 2.5 respectively, for different block sizes. It can be observed from Fig 2.4 that LPHM with block size 1 exhibits a similar behavior as a linear parametric model of order 1. The influence of the linear parametric model seems to be dominant on LPHM up to block size 5. However, from block size 10 (used for resampling the residuals) onwards, the hybrid effect is noticed, indicated by the improvement in the preservation of the linear dependence structure. On the other hand, it may be seen from Fig 2.5 that ANNHM does not preserve the linear dependence structure at block size 1, which is effectively an ANN model with random resampling of residuals. However, with the increase in block size used for resampling of the residuals, ANNHM is found to improve significantly in terms of preserving the linear dependence structure. Nonetheless, LPHM exhibits a closer preservation of the linear dependence structure than ANNHM.

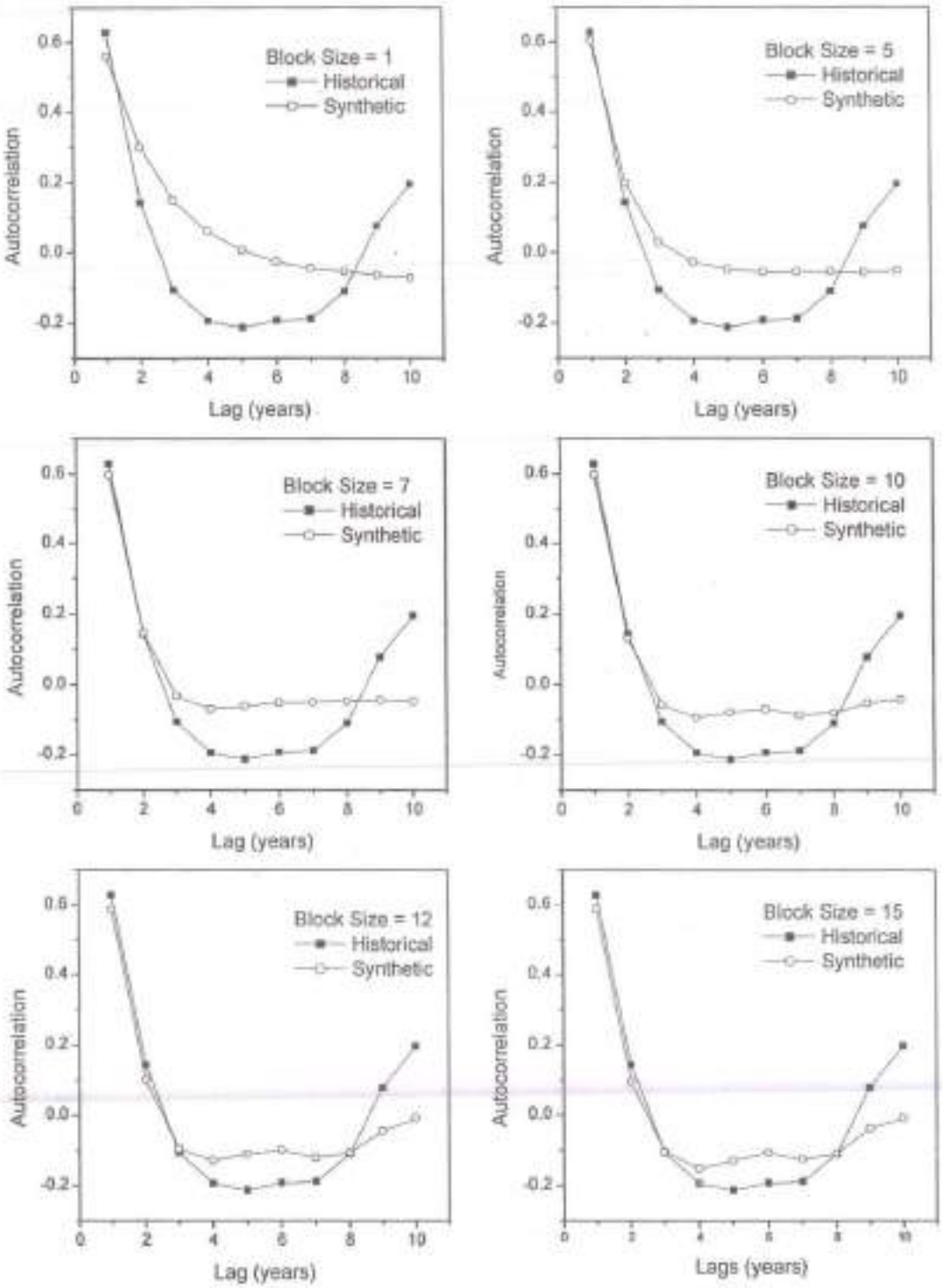


Fig 2.4 Preservation of autocorrelation structure by LPHM. River: Lake Albert

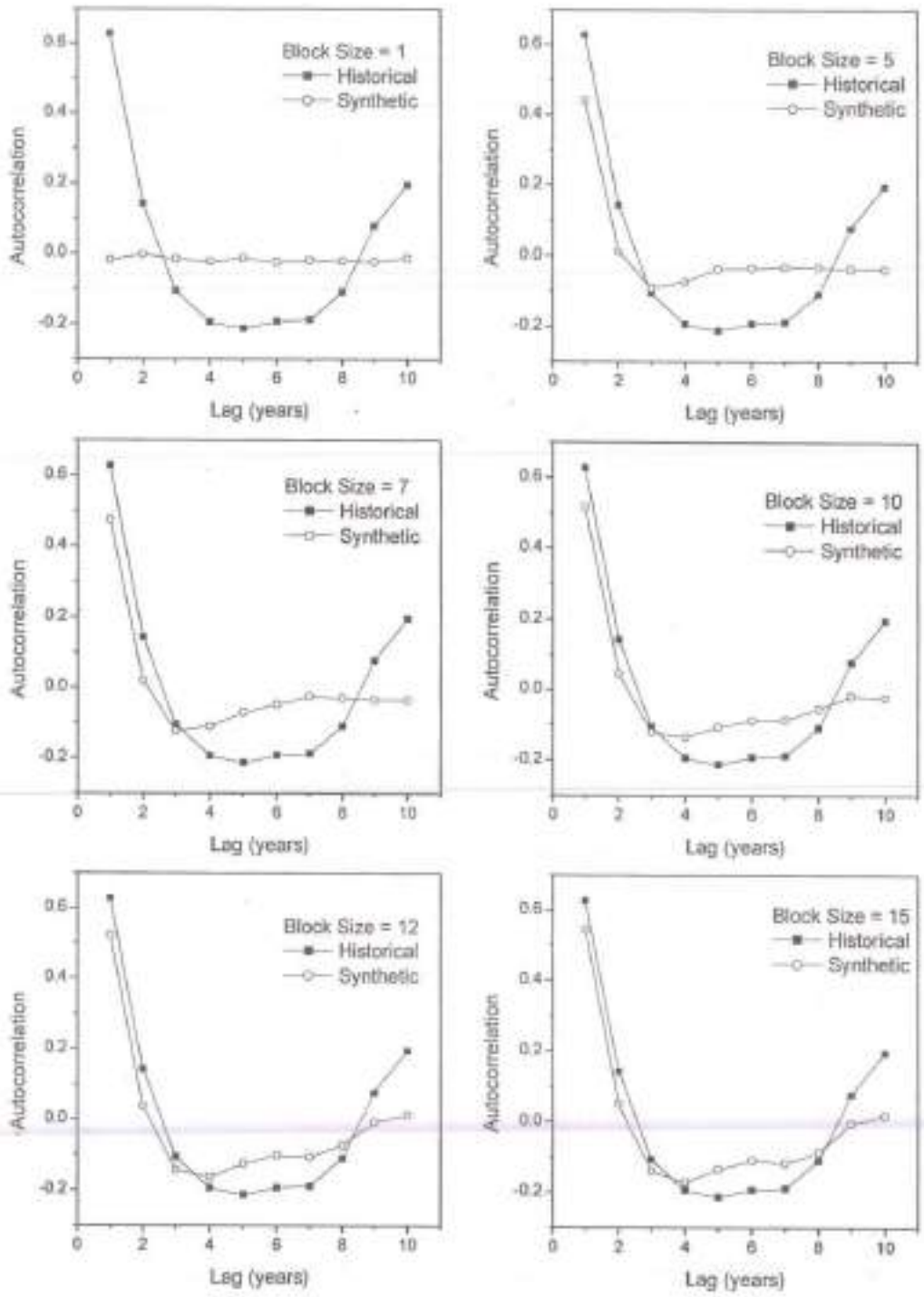


Fig 2.5 Preservation of autocorrelation structure by ANNHM. River: Lake Albert

2.5.1.4 Preservation of Drought Characteristics

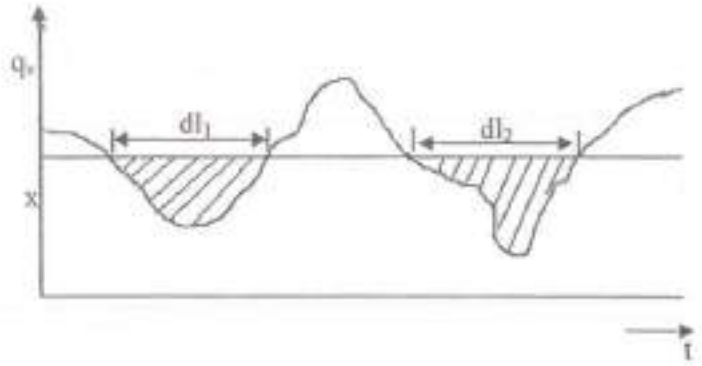
In order to analyze the capability of the two hybrid models in terms of the preservation of drought characteristics of the historical flows, the following definitions of drought characteristics given by Yevjevich (1967) and advocated by other researchers (Sen, 1991; Srinivas and Srinivasan, 2000) are adopted: A run is defined as an uninterrupted sequence of similar events (either surplus or deficit events with regard to a predefined truncation level), preceded and succeeded by different events. While an uninterrupted sequence of surplus is referred to as positive run length, an uninterrupted sequence of deficits is called negative run length. Truncation level is a level that separates the surplus and the deficit volumes of flow with regard to a pre-specified percent (or percentile) of the mean annual flow (MAF). In the current research, the truncation level is expressed as a percent of the mean annual flow (MAF).

Figure 2.6 provides the schematic of the basic definitions of the drought characteristics. Maximum run length (MARL) is taken to be the greatest negative run length in a given streamflow sequence, for a pre-specified truncation level. Maximum run sum (MARS) is defined as the largest volume deficit (negative run sum) encountered in a given streamflow sequence, for a pre-specified truncation level. Mean run length (MERL) is computed as the mean of all the negative run lengths identified in a given streamflow sequence, for a pre-specified truncation level. Mean run sum (MERS) is expressed as the mean of deficit volumes (negative run sums) computed from all drought occurrences, in a given streamflow sequence, for a pre-specified truncation level. The maximum run length and the maximum run sum are referred to as critical drought characteristics in terms of duration and volume respectively.

In this study, the truncation levels have been fixed at 50% MAF to 100% MAF at an incremental level of 5% MAF. Analyzing the run characteristics of the historic

streamflows at close intervals of truncation level, provides information regarding the variations (including jumps) in the run characteristics with regard to variations in the truncation level (Srinivas and Srinivasan, 2000). Ideally, a good stochastic stream flow model for use in drought planning and management is expected to preserve the critical and the mean drought characteristics defined above at all truncation levels specified. The average of the critical and the mean run characteristics corresponding to different truncation levels, computed over thousand replicates obtained from ANNHM and LPHM are presented in Figures 2.7-2.10, along with their historical counterparts for Lake Albert.

It is observed from Fig 2.7 that with regard to the truncation levels, there are two distinct jumps in the historical MARL values. The LPHM fails to capture these jumps of the critical drought duration (MARL) even upto a block size of 10. With further increase in block size, some improvement in prediction is noted. However, the critical drought duration at intermediate truncation levels (70% and 75% MAF) are overestimated. Whereas, ANNHM is able to capture these jumps to some extent, even at a block size of 1 (albeit significant bias), even though the linear dependence structure is not at all preserved by ANNHM at block size 1 (Fig 2.5). It can be seen that the preservation of MARL improves with further increase in block size and results in a good prediction at block size 7 for ANNHM (Fig 2.7). This demonstrates more effective blending of the two nonlinear models ANN and MBB, when compared with LPHM (which is a blend of linear parametric model and MBB).



where **x** is the truncation level (% mean annual flow);
dl_i is the length of i^{th} run; q_v is flow for year 'v'; **t** is time

Fig 2.6. Basic definitions of drought characteristics

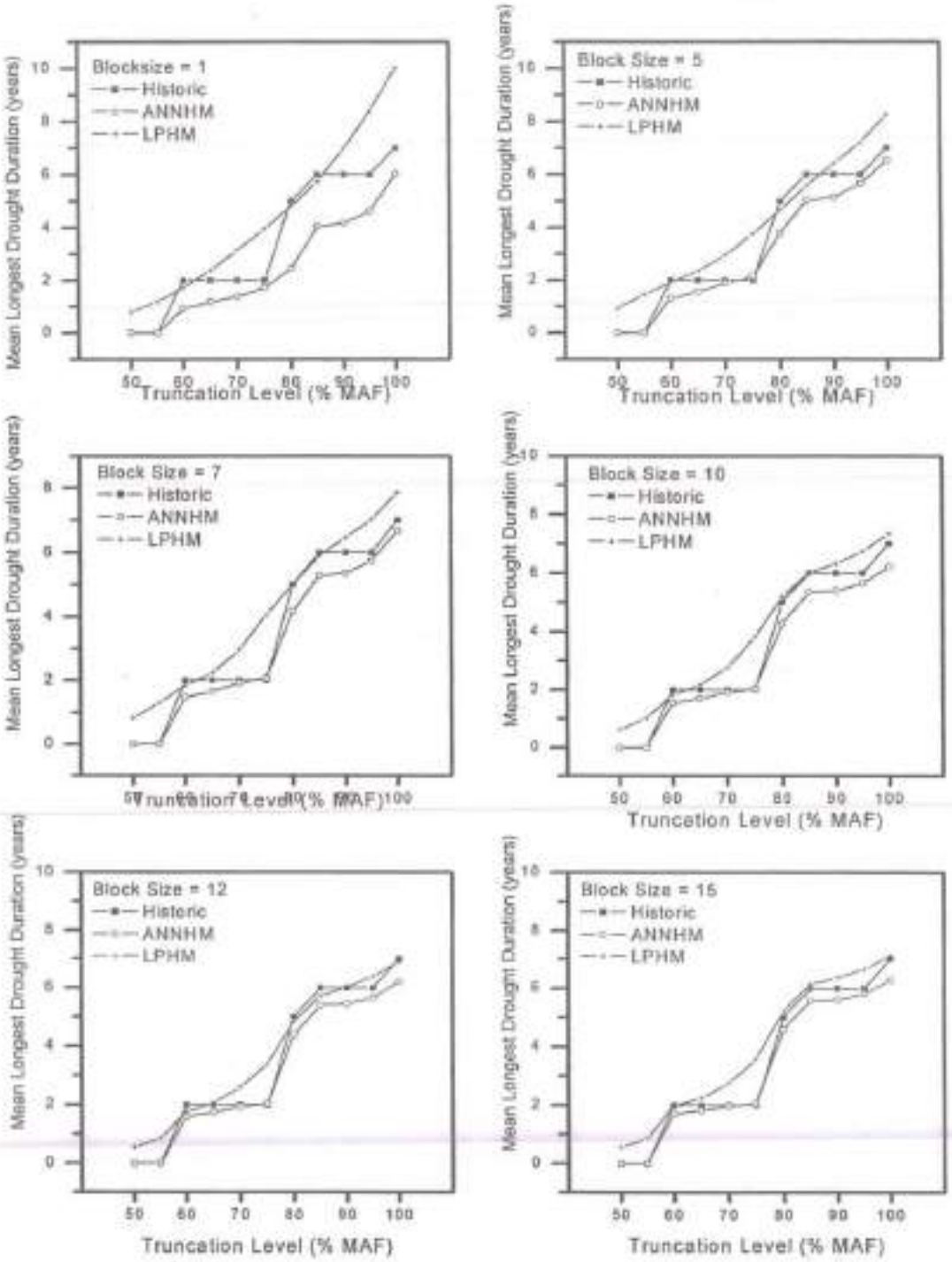


Fig 2.7 Comparison of the preservation of the maximum run length (MARL) between LPHM and ANNHM. River: Lake Albert.

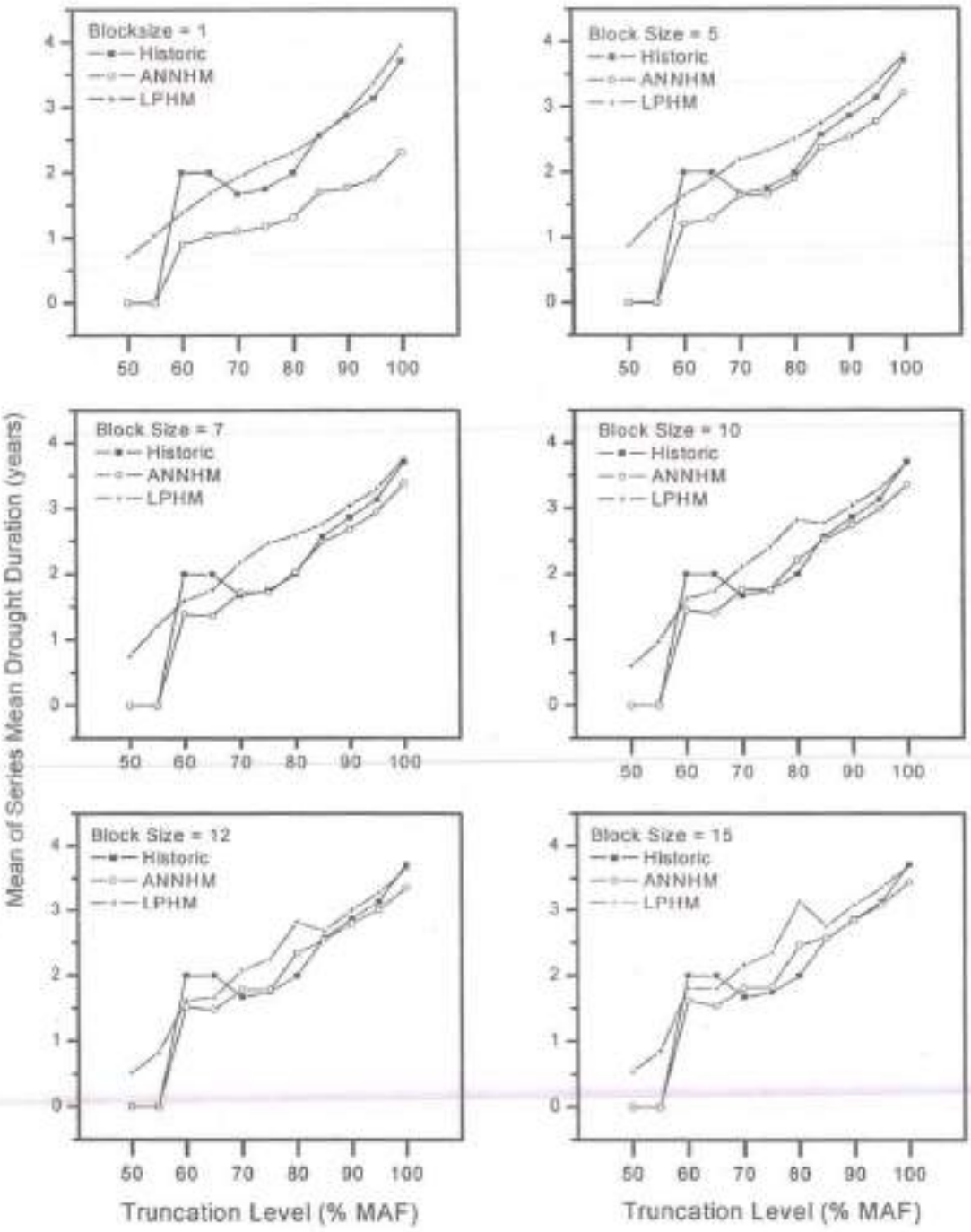


Fig 2.8 Comparison of the preservation of the mean run length (MERL) between LPHM and ANNHM. River: Lake Albert.

The mean run length (MERL) computed from both the models is presented in Fig 2.8, which indicates that LPHM is not able to model the variations of the mean drought durations with increasing truncation levels. This can be plausibly attributed to the dominance of linear parametric effect in the LPHM. On the contrary, ANNHM is able to reasonably model the variations in mean run length with truncation levels from block size 5 onwards (Fig 2.8). Again, this can be attributed to the effective blending of the two nonlinear models.

It is evident from Fig 2.9 (which presents the maximum run sum (MARS) of the historical streamflows), that both LPHM and ANNHM preserve the maximum run sum reasonably well from block size 7 onwards. Figure 2.10 depicts the preservation of the mean run sum (MERS) statistic by both the hybrid models. The inferior performance of LPHM is clearly seen at all truncation levels up to a block size of 5. Even at higher block sizes, the performance of LPHM is not satisfactory at a number of truncation levels. While, ANNHM is able to capture the variation of mean run sum statistic at block size 5, and shows further improvement in performance at block size 7. However, it is to be noted that at higher block sizes (greater than 10), ANNHM does not preserve the mean run sum statistic.

With a view to compare and appreciate the spread of the critical and the mean drought statistics obtained from both the hybrid models, box plots of MARL, MERL, MARS and MERS are presented in Fig 2.11. The inter-quartile range is represented by the box. The limits of the upper and the lower arms indicate the 95 percentile and the 5 percentile points respectively. For the comparison, the block sizes selected for LPHM

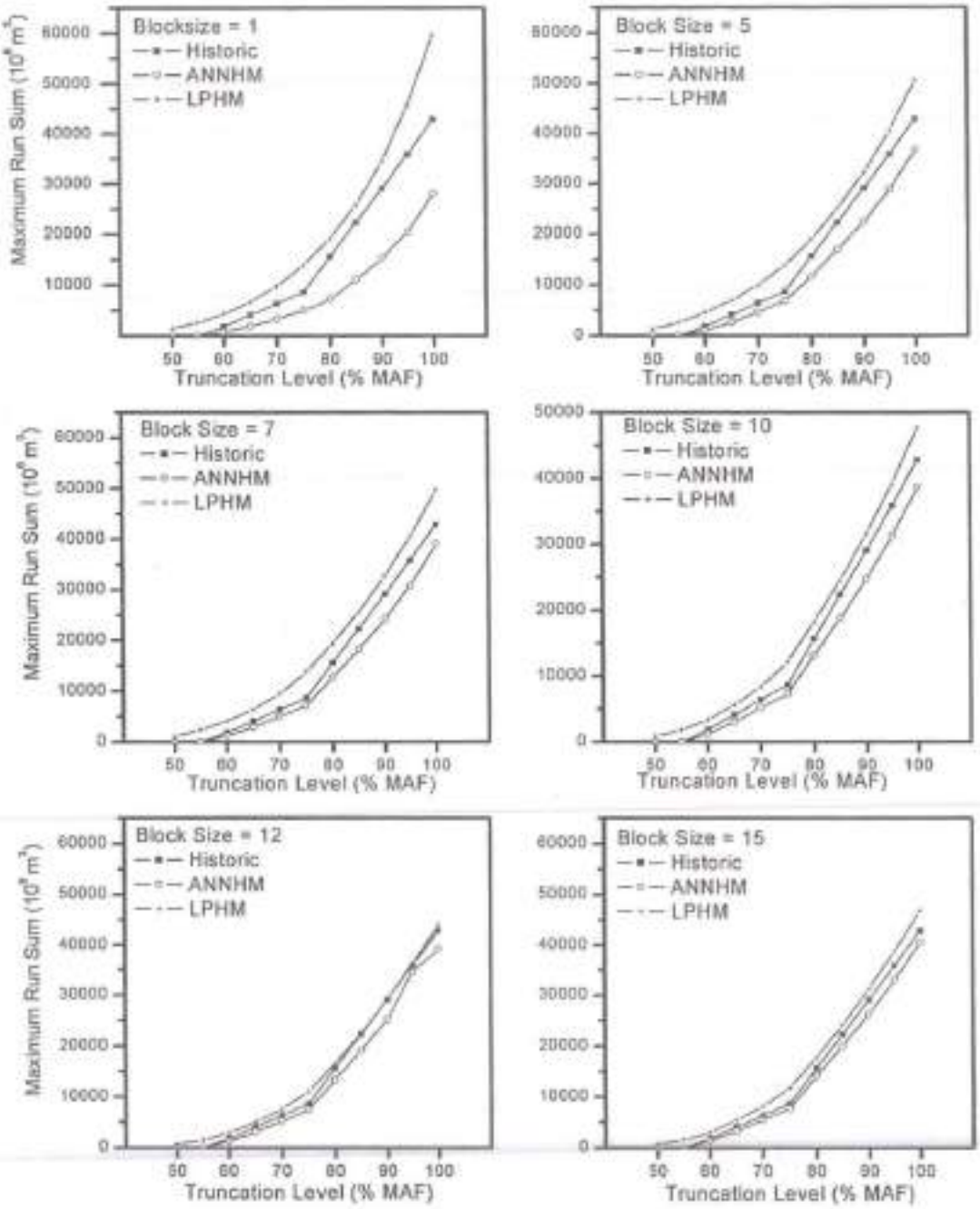


Fig 2.9 Comparison of the preservation of the maximum run sum (MARS) between LPHM and ANNHM. River: Lake Albert.

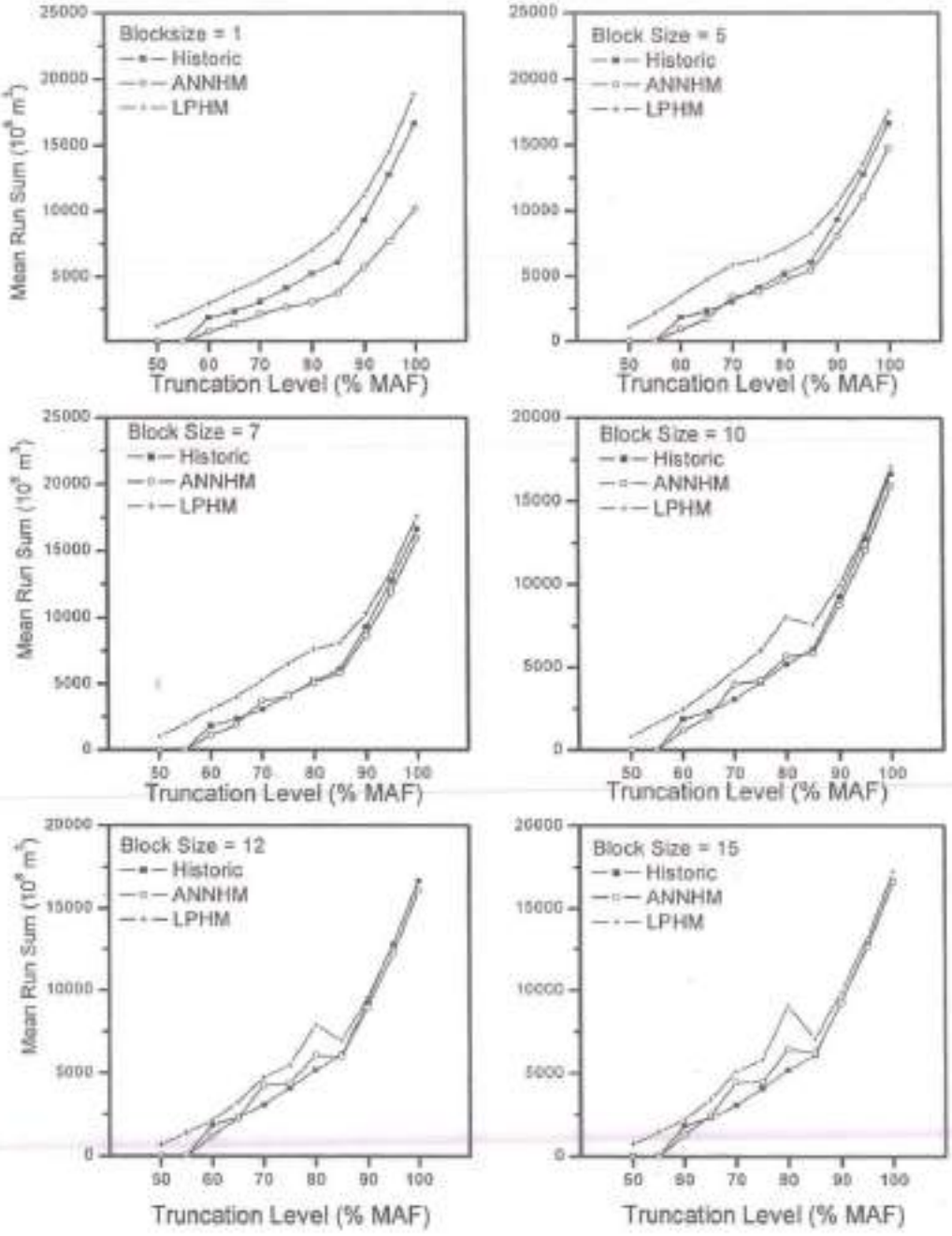


Fig 2.10 Comparison of the preservation of the mean run sum (MERS) between LPHM and ANNHM. River: Lake Albert.

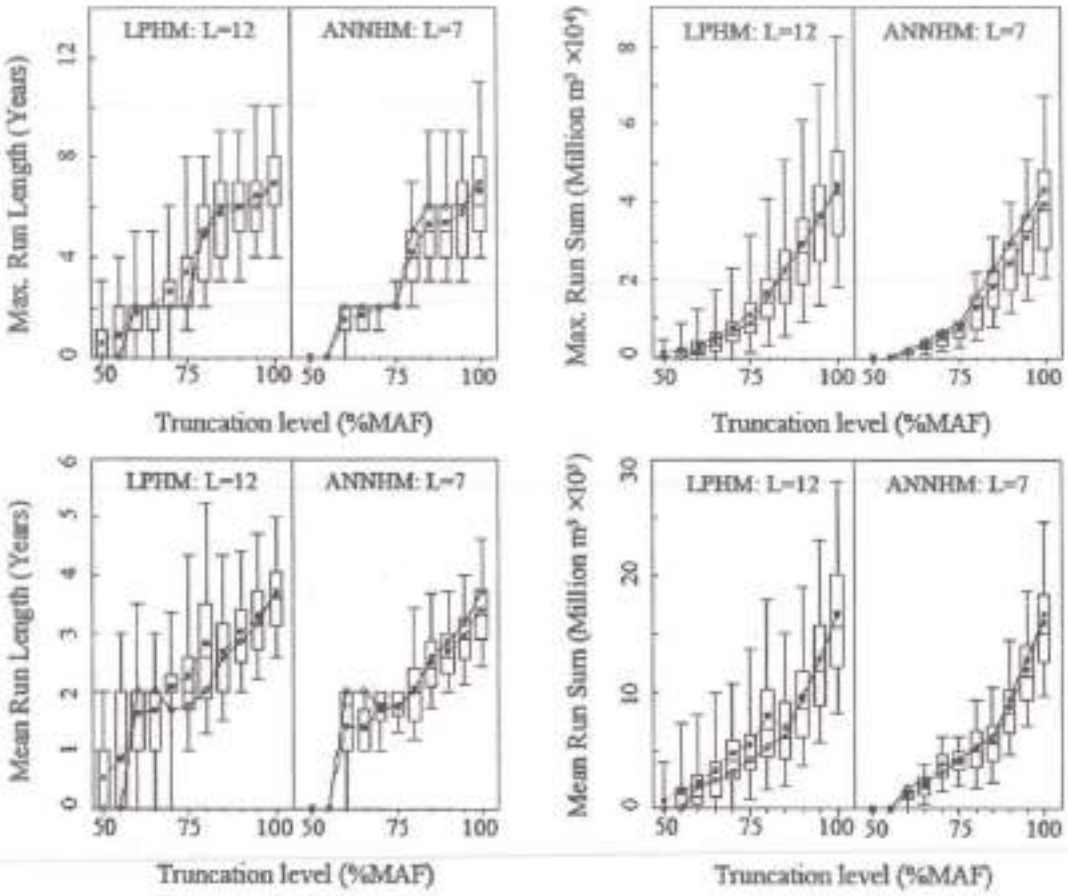


Fig 2.11. Spread of critical and mean run characteristics - A comparison between Linear Parametric Hybrid Model (LPHM: L=12) and ANN Hybrid Model (ANNHM: L=7). River: Lake Albert. *The circle denotes the historical value and the darkened square indicates average value of the statistic over 1000 replicates.*

and ANNHM are 12 and 7 respectively. This selection of block size was done considering the tradeoff between the bias and the spread of the above statistics.

It is seen from Fig 2.11 that LPHM overestimates MARL at a few of the lower and the intermediate truncation levels, while it predicts the same well at higher truncation levels. Whereas, in case of ANNHM, the spread of MARL is limited at two of the intermediate truncation levels. Although the maximum run sum (MARS) has been preserved well by both the models, LPHM shows a better prediction at higher truncation levels. With regard to the prediction of the mean drought characteristics (MERL and MERS), ANNHM outperforms LPHM. A significant observation from Fig. 2.11 is that the zero values of all the historical drought statistics at 50% and 55% truncation levels are reproduced by ANNHM, while the same are overestimated by LPHM. Also, the jump in the historical drought statistics from 55% to 60% truncation level is better captured by ANNHM. From the above, it may be inferred that ANNHM is able to capture the nonlinearities inherent in data better than LPHM.

2.5.1.5 Prediction of Reservoir Storage Capacity

The simulations from the two hybrid models are further validated by testing their ability to predict reservoir storage capacity. The reservoir storage capacities required to cater to yields of 50% Mean Annual Flow (MAF) to 90% MAF (at 5% MAF intervals), are computed using the sequent peak algorithm (Loucks et al., 1981, p.235). The results are presented in Table 2.2 and Fig 2.12. It is to be noted that the relative bias in predicting the storage statistics at lower demand levels is high in case of LPHM (Fig. 2.12) Moreover, the relative RMSE of storage capacity prediction is

phenomenally higher in case of LPHM compared to ANNHM (Table 2.2) because of the larger

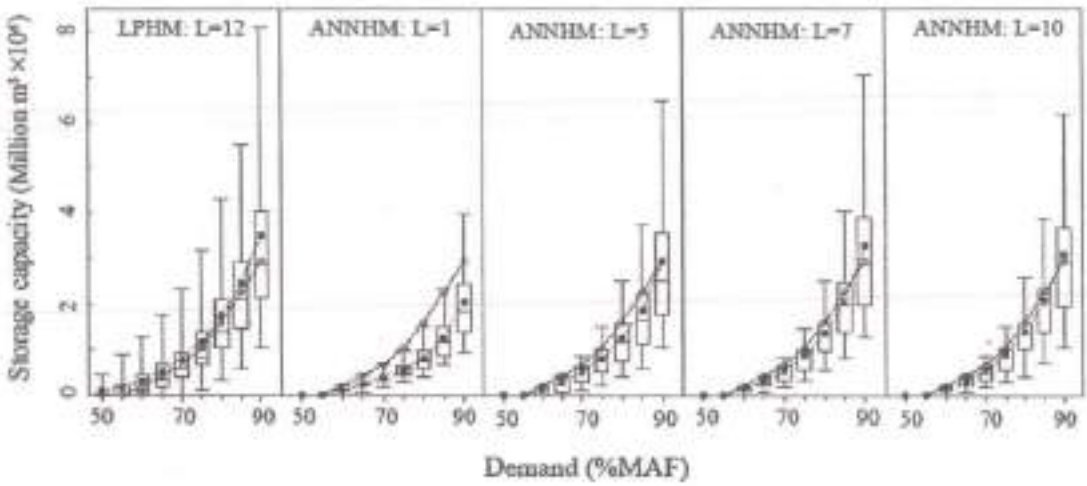


Fig. 2.12. Prediction of reservoir storage capacity - A comparison between Linear Parametric Hybrid Model (LPHM: L=12) and ANN Hybrid Model (ANNHM). River: Lake Albert. *The circle denotes the historical value and the darkened square indicates average value of the statistic over 1000 replicates.*

dispersion (Fig 2.12). It may also be noted from Fig 2.12 that the performance of ANNHM improves as the block size increases to 7. However, with further increase in block size, the spread of the storage statistic gets reduced, which is not desirable for design decision making.

Table 2.2 Comparison of predicted storage characteristics by LPHM and ANNHM for selected block sizes. River: Lake Albert.

	Demand (% MAF)	60	65	70	75	80	85	90
Relative Bias	LPHM, L=12	-0.593	-0.231	-0.213	-0.161	-0.103	-0.085	-0.194
	ANNHM, L=7	0.357	0.286	0.192	0.147	0.136	0.089	-0.110
Relative RMSE	LPHM, L=12	2.493	1.408	1.125	0.965	0.807	0.779	0.915
	ANNHM, L=7	0.516	0.467	0.384	0.384	0.415	0.454	0.715

L indicates the block size; MAF indicate mean annual flow

2.5.2 EXAMPLES 2 & 3: Red River (USA) and River Neva (USSR)

The mean and the standard deviation of the historical flows are well preserved by both the models for both Red River and River Neva, which are not presented herein for brevity. The ANNHM reproduces the skewness of the Red River flows (0.973, see Table 2.1) even at a block size of 1 (mean skewness over 1000 synthetic sequences = 0.924), while LPHM exhibits considerable bias (mean skewness over 1000 synthetic sequences = 0.808). Although the performance of LPHM improves as the block size increases (0.855 at L=6 and 0.869 at L=10), the model is not able to capture the skewness entirely. The better performance of ANNHM in preserving the skewness indicates that it has a tendency to capture the nonlinearities more effectively utilizing the potential of ANN, which confirms the earlier observations in the case of Lake Albert. In the case of River Neva, the historic flow series exhibits a low skewness of 0.438 (see Table 2.1) and hence ANNHM is able to show only a marginal improvement in reproducing the skewness. It is to be mentioned that with regard to

the preservation of linear dependence, both models exhibit a similar behaviour as in the case of Lake Albert flows, and hence the same are not presented herein for brevity.

2.5.2.1 Preservation of Drought Characteristics

The performances of ANNHM and LPHM in terms of their potential to preserve the drought characteristics when applied to the streamflow series of Red River and Neva River are presented in Tables 2.3 and 2.4 respectively. It is observed from Table 2.3 that in the case of Red River both the models underestimate the MARL at higher truncations (85-100% MAF). This is mainly because both the models are not able to simulate the jump in historical MARL (4 to 8 years) from 80% to 85% MAF. In the case of River Neva (see Table 2.4), the MARL at higher truncations (95%, 100% MAF) are significantly over estimated by LPHM. The preservation of MERL is seen to be better for ANNHM at lower truncation levels (50-60% MAF) in the case of Red River (Table 2.3). On the other hand, in the case of Neva River streamflows there is no significant difference (Table 2.4) between the two models in simulating MERL.

It is observed from Table 2.3 that in the case of Red River the MARS is overestimated by LPHM at lower truncation levels, while ANNHM underestimates the same at higher truncations. It may be noted that the standard deviation of the predicted MARS (over the replicates) is higher for LPHM when compared with ANNHM. In the case of Neva River (Table 2.4) it is observed that ANNHM performs well at higher truncation levels while it underestimates the MARS at intermediate truncation levels (75%, 80% MAF). On the contrary, LPHM is found to overestimate MARS at both intermediate and higher truncation levels in addition to high dispersion over the replicates. The

preservation of MERS is found to be better for ANNHM compared to LPHM in the case of both the rivers.

Table 2.3 Preservation of historical drought characteristics – Red River (values in the parenthesis denote the standard deviation over 1000 replicates, TL denotes truncation level (%))

TL	50	55	60	65	70	75	80	85	90	95	100
Maximum Run Length (years)											
Historical	2.00	2.00	4.00	4.00	4.00	4.00	4.00	8.00	8.00	8.00	8.00
LPHM	2.34	2.78	3.61	3.72	3.93	4.02	4.72	6.97	7.66	7.86	7.99
	(0.7)	(0.8)	(1.2)	(1.1)	(1.1)	(1.1)	(1.2)	(2.9)	(3.0)	(3.0)	(3.0)
ANNHM	2.02	2.30	3.55	3.71	3.83	3.98	4.44	6.51	7.58	7.63	7.63
	(0.5)	(0.6)	(1.2)	(1.1)	(1.1)	(1.0)	(1.1)	(2.5)	(2.7)	(2.7)	(2.7)
Mean Run Length (years)											
Historical	1.25	1.31	1.54	1.62	1.85	1.85	2.08	2.64	3.10	3.20	3.20
LPHM	1.35	1.46	1.62	1.71	1.83	1.85	2.11	2.68	3.00	3.14	3.20
	(0.2)	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	(0.4)	(0.6)	(0.7)	(0.7)	(0.8)
ANNHM	1.24	1.29	1.53	1.71	1.79	1.84	2.01	2.63	3.03	3.13	3.13
	(0.2)	(0.2)	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	(0.6)	(0.8)	(0.8)	(0.8)
Maximum Run Sum (in 10^6 m^3)											
Historical	231	288	345	446	559	673	786	1366	1593	1820	2047
LPHM	271	336	437	529	628	736	900	1327	1605	1839	2078
	(114)	(128)	(152)	(174)	(199)	(227)	(265)	(553)	(643)	(722)	(813)
ANNHM	197	250	348	445	546	649	791	1134	1448	1662	1875
	(62)	(76)	(90)	(113)	(140)	(170)	(204)	(410)	(519)	(593)	(670)
Mean Run Sum (in 10^6 m^3)											
Historical	95	87	127	173	221	274	354	456	586	676	767
LPHM	103	110	148	191	234	284	365	492	611	704	800
	(34)	(30)	(37)	(43)	(48)	(56)	(78)	(126)	(157)	(182)	(203)
ANNHM	90	91	124	172	221	272	341	466	592	678	767
	(23)	(22)	(26)	(31)	(37)	(44)	(59)	(98)	(149)	(171)	(192)

Table 2.4 Preservation of historical drought characteristics – Neva River (values in the parenthesis denote the standard deviation over 1000 replicates, TL denotes truncation level (%))

TL	50	55	60	65	70	75	80	85	90	95	100
Maximum Run Length (years)											
Historical	0	0	0	0	0	2.00	2.00	3.00	4.00	4.00	6.00
LPHM	0	0	0	0.03	0.31	1.78	3.25	3.06	3.75	5.52	6.82
				(0.2)	(0.6)	(0.6)	(0.7)	(0.9)	(1.2)	(1.7)	(1.7)
ANNHM	0	0	0	0	0	1.65	2.07	2.79	3.36	4.63	6.19
						(0.6)	(0.4)	(0.6)	(0.7)	(1.1)	(1.3)
Mean Run Length (Years)											
Historical	0	0	0	0	0	1.50	1.57	1.56	2.20	2.73	3.08
LPHM	0	0	0	0.03	0.30	1.40	1.54	1.66	2.01	2.64	3.17
				(0.2)	(0.6)	(0.4)	(0.2)	(0.3)	(0.3)	(0.5)	(0.6)
ANNHM	0	0	0	0	0	1.36	1.48	1.58	1.94	2.52	3.06
						(0.4)	(0.2)	(0.2)	(0.2)	(0.4)	(0.5)
Maximum Run Sum (in 10^6 m^3)											
Historical	0	0	0	0	0	2686	10848	19985	32228	44471	65677
LPHM	0	0	0	31	532	4472	12510	23336	37397	57570	85544
				(244)	(1371)	(3354)	(5343)	(8364)	(12508)	(18848)	(24950)
ANNHM	0	0	0	0	0	1529	8730	18027	29982	44923	70691
						(554)	(2268)	(4019)	(6786)	(10107)	(15242)
Mean Run Sum (in 10^6 m^3)											
Historical	0	0	0	0	0	2157	4133	8745	15425	23561	32763
LPHM	0	0	0	31	511	2810	5073	10256	16408	25417	34939
				(244)	(1324)	(1589)	(1589)	(2444)	(3532)	(5286)	(7891)
ANNHM	0	0	0	0	0	1224	4084	8796	14561	23002	33065
						(418)	(1122)	(1804)	(2378)	(3939)	(6376)

2.5.2.2 Prediction of Reservoir Storage Capacity

It can be seen from Table 2.5 (Red River) that the LPHM consistently overestimates the storage capacity for all demand levels (60% to 90% MAF) in addition to a high degree of dispersion resulting in high values of relative RMSE. On the contrary, the

ANNHM yields a good prediction of the storage capacity for demand levels 60% to 80% MAF. However, for higher demand levels (85% and 90% MAF) considerable overestimation is seen resulting in high relative RMSE. In the case of River Neva (Table 2.6), it may be noted that the performance of ANNHM is better than LPHM at demand levels of 85% and 90% MAF. However, both the models do not seem to perform well at a lower demand level of 75% MAF.

Table 2.5 Comparison of predicted storage characteristics by LPHM and ANNHM for selected block sizes: Red River

	Demand (% MAF)	60	65	70	75	80	85	90
Relative Bias	LPHM, L*= 6	-0.363	-0.297	-0.216	-0.213	-0.324	-0.568	-0.461
	ANNHM, L= 6	-0.056	-0.022	0.008	-0.020	-0.124	-0.361	-0.330
Relative RMSE	LPHM, L= 6	0.629	0.593	0.541	0.562	0.743	1.082	1.027
	ANNHM, L= 6	0.289	0.308	0.331	0.380	0.511	0.801	0.841

L indicates the block size; MAF indicate mean annual flow
Negative and positive values of relative bias indicate overestimation and underestimation of the historical storage capacity respectively.

Table 2.6 Comparison of predicted storage characteristics by LPHM and ANNHM for selected block sizes. River: Neva

	Demand (% MAF)	60#	65#	70#	75	80	85	90
Relative Bias	LPHM, L*=13	-	-	-	-0.667	-0.171	-0.198	-0.361
	ANNHM, L=13	-	-	-	0.431	0.188	0.084	-0.014
Relative RMSE	LPHM, L=13	-	-	-	1.415	0.537	0.477	0.618
	ANNHM, L=13	-	-	-	0.477	0.290	0.222	0.248

L indicates the block size; ; MAF indicate mean annual flow; # At the demand levels of 60%, 65% and 70% MAF, the storage capacity computed from historical flows is zero, hence the values in these columns can not be computed.
Negative and positive values of relative bias indicate overestimation and underestimation of the historical storage capacity respectively.

2.6 SUMMARY AND CONCLUSIONS

A hybrid model that blends the two non-linear data-driven models, ANN (deterministic) and MBB (stochastic) is proposed for modeling annual streamflows of rivers that exhibit complex dependence. First, a nonlinear deterministic model, ANN

(radial basis function network) is fitted to the historical annual streamflows, which captures the nonlinear trend in the data effectively. Then, the resulting residuals from the ANN model are resampled using a non-parametric resampling technique, moving block bootstrap with a view to capture the weak linear as well as the nonlinear dependence and any distributional information retained in the residuals. The proposed model has been applied to three annual streamflow data sets that exhibit complex dependence, drawn from different geographic regions with varying record lengths. The effective blending of the two data-driven models is shown to result in efficient simulations of the long-term storage and drought-related characteristics.

Skewness present in the streamflows is better preserved by the proposed ANN based hybrid model (ANNHM) compared to the linear parametric based hybrid model (LPHM) plausibly owing to the effective capturing of the nonlinearities. The ANNHM being a completely data-driven model, reproduces the features of the marginal distribution more closely compared to LPHM, but offers less smoothing and little extrapolation value. However, the linear dependence structure is better reproduced by LPHM than ANNHM.

Despite a better preservation of the linear dependence structure, LPHM is not able to effectively predict the variation of critical drought duration (including jumps) with respect to truncation level. On the contrary, ANNHM is able to model the variation of critical drought duration better, even though the preservation of linear dependence structure is inferior to LPHM. This is plausibly due to the effective blending of the two nonlinear models. Also, the mean drought characteristics are more efficiently modeled by ANNHM.

The relative bias in predicting the reservoir storage statistics at lower demand levels is found to be high in case of LPHM. Moreover, a large spread of the same is observed at all demand levels, thus increasing the relative RMSE significantly compared with ANNHM.

Future research should address the extension of the proposed ANN-based hybrid model to single-site and multi-site modeling of periodic stream flows.

CHAPTER 3

PERIODIC STOCHASTIC STREAMFLOW MODELS

3.1 INTRODUCTION

Seasonality of streamflow data adds a degree of complexity to the selection of an appropriate stochastic model to fit the data. It may be necessary to use a parametric model that has seasonally varying properties. If the seasonality of the flow data under consideration appears to be only in the mean and the variance, then such seasonality can be removed by simple seasonal standardization, and a stationary model can be fitted to the deseasonalized data. However, if the autocorrelation structure of the observed data exhibits significant periodicity, then seasonal models that explicitly incorporate a periodic dependence structure must be used (Rasmussen et al., 1996). Varying degree of nonlinearity in the different periods/seasons, generation of streamflows owing to mixed precipitation mechanisms, add to the complexity further. Such characteristics of the geophysical time series make the modeling of multi-season streamflows a challenging task.

An ideal single site multi-season synthetic flow generation model should aim to reproduce: the summary statistics (mean, standard deviation and skewness) and marginal distribution of observed flows at periodic and annual time scales; autocorrelation structure of flows at aggregated annual level; within-year and cross-year serial correlations; month-to-year cross-correlations; and non-linearity stationarity in the underlying dependence structure. In addition, it should provide sufficient variety in the stochastic simulations with a reasonable degree of smoothing and extrapolation.

3.2 LITERATURE REVIEW

3.2.1 Parametric Models

In operational hydrology, quite often synthetic seasonal streamflow sequences are generated using one of the following two approaches: (i) a direct approach using a seasonal model (whose parameters may change from season to season); (ii) disaggregation approach in which annual streamflows are generated first using a simple linear parametric model and the same are disaggregated into seasonal flows subsequently. Typical examples for the former approach include Hipel et al. (1977), Hirsch (1979), Salas et al. (1982), Vecchia et al. (1983), Haltiner and Salas (1988). The seasonal models are often used in modeling streamflows at a single site and are quite simple in structure and are parsimonious. However, these models may not preserve the dependence structure of historical streamflow sequences over periods of several months to one or more years and hence may not be preferable for long-term reservoir operation studies (Stedinger and Taylor, 1982).

In addition to the general drawbacks of AR/ARMA models (mentioned in the earlier subsection), a few more drawbacks/limitations surface in the context of periodic modeling of streamflows:

In certain cases, a different normalizing transformation may have to be applied for each period, for effective reduction of skewness close to zero and this may distort the correlation structure in the synthetic simulations upon inverse transformation to real space. If the streamflow data size is limited, then, a single transformation may have to be applied for all periods, which may not reduce the skewness to zero in all periods. This will result in modeling inaccuracies, as normality condition will not be satisfied for all periods. During

low flow periods, the standard deviation of flows may be equal or higher in magnitude than the mean of flows. This necessitates using certain transformations that have a lower bound, in order to avoid generation of negative flows. This may distort the marginal distribution of flows in such periods. Moreover, if each period apparently follows a different order model, then, the model may not be able to provide a good fit to the data, since exact statistical tests for identification as well as diagnostic checking of residuals, do not seem to exist for such cases. In case of periodic ARMA models, it becomes even complex, compared with periodic AR models. Additional complexities may arise due to stationarity conditions to be satisfied by the periodic AR parameters.

In the disaggregation approach, to start with, annual flows are generated using an appropriate annual streamflow model, and then the same are divided among the seasons (periods) within the year. These models describe the distributions of streamflows both at the aggregated (annual) and the disaggregated (seasonal) levels. Such disaggregation can proceed further down in the time scale up to even hourly flows, in stages. The pioneering work in this direction was that of Valencia and Schaake (1973) though Harms and Campbell (1967) predates it. The structure of the Valencia and Schaake (VS) model is designed to preserve the variance and covariance between the annual and the seasonal flows and to preserve the variance and covariance among the within-year flows (Salas et al., 1980). A major drawback of the VS model is that it fails to reproduce the cross-year serial correlations. In addition, the number of parameters to be estimated is quite large. Mejia and Rousselle (MR) (1976) extended the VS model by introducing an additional term with a view to preserve the correlations between the seasonal flows in the current year to those of

the previous year(s). In spite of increasing the model complexity through additional parameters, the MR model could not perform the intended function in all cases (Lane, 1982; Stedinger and Vogel, 1984). Driven by the concern of the complexity due to the large number of parameters required by the VS and the MR models, Lane (1979) developed a condensed disaggregation model. Even though the LAST computer package (Lane, 1979; Lane and Frevert, 1990) incorporates staging and the use of condensed models (thus significantly reducing the number of parameters), there are two major drawbacks: (i) it does not explicitly model the high-lag month-to-month serial correlations and (ii) it makes little effort to keep the sum of seasonal flows close to the specified annual total. Grygier and Stedinger (1988) mention that in case of Lane (1979)'s model, large adjustments are often required to make the generated flows at the seasonal level, add up to the specified annual value, resulting in distortion of the distribution of the generated flows.

The condensed disaggregation models (Lane (1979, 1982); Stedinger and Pei, 1982; Stedinger et al., 1985; Grygier and Stedinger, 1988) attempt to reproduce only a selected subset of the correlation statistics with a view to reduce the number of parameters required and hence the model size. The condensed temporal disaggregation model of Stedinger et al. (1985) reproduces explicitly only the correlations between monthly flows and annual flows, and between consecutive monthly flows. The SPIGOT (Grygier and Stedinger, 1990) stochastic streamflow package uses univariate and multivariate generalizations of the temporal disaggregation model developed in Stedinger et al. (1985). Furthermore, empirical adjustment procedures suggested by Grygier and Stedinger (1988) have been incorporated into SPIGOT, in order to restore summability of

the disaggregate flows to the aggregate flows, in the event of normalizing transformations being used. Santos and Salas (1992) have presented a stepwise disaggregation scheme that would preserve means, variances and a specified covariance structure, also maintaining the additivity property. However, this scheme is known to be limited to reproducing these properties only in the normalized flow domain.

In the last decade, further advancements have been made in the parametric modeling front by Koutsoyiannis (1992; 1999; 2000) and Koutsoyiannis and Manetas (1996). Koutsoyiannis (1992) developed a parsimonious nonlinear multi-variate dynamic disaggregation model (DDM) that follows a stepwise approach for simulation of hydrologic series. This involved two parts (i) a linear step-by-step moments determination and (ii) an independent non-linear partitioning. This model was shown to treat the skewness of the lower level variables explicitly, without loss of additive property. Koutsoyiannis and Manetas (1996) proposed another simpler multivariate disaggregation method, that retained the parsimony in model parameters for lower level variables as in DDM (Koutsoyiannis, 1992), and implemented accurate adjusting procedures to allocate the error in the additive property, followed by repetitive sampling to improve the approximations of the statistics that are not explicitly preserved by the adjustment procedures. More recently, a generalized mathematical framework for stochastic simulation and forecasting problems in hydrology has been proposed by Koutsoyiannis (2000). A generalized autocovariance function is introduced and is implemented in a generalized moving average generating scheme that yields a new time-symmetric (backward-forward) representation. A notable highlight of this model

framework is that unlike in the traditional stochastic models, the number of model parameters, the type of generation scheme and the type of autocovariance function can be decided separately by the modeler. This framework is shown to be appropriate for stochastic processes with either short-term or long-term memory. Koutsoyiannis (2001) also proposed a methodology for coupling stochastic models of hydrologic processes applying to different time scales.

3.2.2 Nonparametric Models

Nonparametric procedures offer significant advantages over their parametric counterparts. Nonparametric procedures generally reproduce the empirical structure of multivariate data sets, yet, they do not require assumptions about data or model structure, or complexities associated with parameter estimation. As a result, simple data resampling schemes such as the bootstrap and jackknife have gained acceptance by hydrologists as conceptually simple (yet computationally intensive) alternatives to more complex parametric alternatives. Helsel and Hirsch (1992) and Lall (1995) provide a comprehensive review on applications of nonparametric techniques to a wide class of water and environmental applications.

While parametric methods of time series modeling require assumptions regarding the marginal probability distributions and the correlation structure of the variables of interest, nonparametric methods are, in general, data-driven and simply retain the empirical structure of the observed variables. Parametric methods require estimates of a number of model parameters, which the nonparametric methods can either minimize or avoid

altogether, depending on the method adopted (Vogel and Shallcross, 1996). Moreover, parametric uncertainty is considered implicitly in the nonparametric approach, since a broad class of models is approximated. It is to be noted that a parametric probability density function is one that is fully defined by a finite set of parameters, while a nonparametric probability density estimate is based on the entire sample (rather than a few sample moments). In case of the conventional linear parametric models, the data near the modes of the marginal distribution of observed streamflows dominate the model fit, and the tails can be viewed as extrapolation of that behavior. In the process, the parametric models generate synthetic streamflows beyond the extrema and in between the observed streamflows (including in between large discontinuities (if any)), inflicting considerable smoothing to the histogram of observed streamflows. It is to be mentioned that in certain cases, it may be rather unwise to over-smooth the wide discontinuities seen in the histogram of observed flow trace, since it may lead to some amount of misrepresentation of the real hydrologic behavior. In contrast, synthetic replicates from simple bootstrapping techniques, such as MBB, mimic the multimodality, peakedness and asymmetry seen in the marginal distribution of observed flows. However, this kind of bootstrapping fails to generate flow values other than those seen in the historical flow trace. In other words, it has a tendency to parse the data (no smoothing effect is seen) and thus defeats the purpose of synthetic streamflow simulation. Hence, there is need for a model that is not only good at reproducing the salient features of the marginal distribution of observed streamflows, but also flexible enough to provide reasonable degree of smoothing and extrapolation in the tails. A nonparametric density estimator is consistent, whereas a mis-specified parametric PDF has a bias that does not reduce with increasing

sample size (Sharma et al., 1998). Silverman (1986) and Scott (1992) provide introductory material on nonparametric methods. The increasing awareness of the need to model nonlinearity and nonstationarity (such as jumps and periodicities) in the underlying dynamics of geophysical processes coupled with the availability of reasonable length of historical records and fast and inexpensive computing facilities has spurred the use of nonparametric methods in several areas of hydrology, in recent times (Lall, 1995; Lall and Sharma, 1996; Lall et al., 1996).

A variety of non-parametric "smoothers" are available in the statistical literature. They differ in their estimation efficiency, in their computational demands, in their applicability, and in their mathematical form. However, they share the goal of approximating (with asymptotically vanishing error) an arbitrary, unknown function of the data, and the notion that each estimate be local (i.e., influenced only by a nearby data). Smoothers are interpretable as weighted moving averages (kernel estimators) of some function of the data. Localization is achieved by weights that vanish with distance from the point of estimate.

For trend analysis and investigating bivariate dependence, locally weighted estimation or LOESS has emerged as the nonparametric method of choice. Helsel and Hirsch (1992) and Hirsch et al. (1991;1993) formalize procedures for using LOESS for trend analysis of hydrologic and environmental data, and to remove systematic variations in the environmental variable of interest. Sangoyomi and Lall (1993) used kernel density estimate (k.d.e.) to investigate the number of modes in the p.d.f. of several hydroclimate

time series in the Great Salt Lake basin, with the aim to identify distinct regimes in long term climate, and thereby improve the predictability of the Great Salt Lake volume variations. Lall and Bosworth (1993) have developed a multivariate kernel density estimator that employs a set partitioning strategy to define local bandwidth matrices proportional to subset covariance, and explore multivariate dependence between precipitation, evaporation, net precipitation and annual inflow into the Great Salt Lake. An interesting interplay between precipitation and evaporation in generating inflows is seen. Serial dependence issues are not properly dealt with. The sensitivity of k.d.e. to bandwidth variation is examined, but optimal bandwidth selection is not attempted.

Tong (1990) provides motivation for nonlinear time series analysis methodology and for nonparametric modelling and visualization of time series. He uses a daily river flow example to illustrate that such data with sudden jumps, time irreversibility, asymmetric joint distributions, persistence, lots of high level crossings, and state dependent correlation between lagged flows do not support the assumptions inherent in classical linear ARMA modelling.

Yakowitz (1987, 1993), Yakowitz and Karlsson (1987), Karlsson and Yakowitz (1987a, 1987b) motivate and provide theoretical basis of nearest neighbour (NN) regression for prediction of time series and specifically for rainfall-runoff modelling. Galeati (1990) shows that this simple NN predictor provides lower mean inflow to an Italian reservoir relative to an autoregressive model with exogenous inputs that was coupled to physically based, calibrated, rainfall-runoff and snow cover evolution models.

Smith (1991) and Smith et al. (1992) present some interesting applications of Yakowitz's ideas that expose the flexibility of nonparametric methods for seeking relationships between arbitrary functions of possibly linked data sets. Kember et al. (1993) connect the NN predictor to state space reconstruction methods used to reconstruct nonlinear dynamics (Farmer and Sidorowich, 1987) from time series.

Lall et al. (1994) use Multivariate Adaptive Regression Splines due to Friedman (1991), to recover the map of the dynamical system, based on the time series of biweekly volume of the Great Salt Lake. Parameters including model order, delay, and spline parameters are chosen using generalized cross validation (GCV). Blind predictions up to 4 years ahead using only prior data are seen to be dramatically superior as the forecast horizon increases, compared to those from the best fit autoregressive (AR) model. In fact, the unprecedented, and dramatic 4-year rise and fall of the Great Salt Lake in the 1980's, could be predicted.

Rajagopalan et al. (1993, 1994) and Lall et al. (1993) develop a seasonal nonparametric renewal model (NPR) for simulating daily precipitation, where successive dry and wet spell lengths may be dependent or independent. All requisite probability density functions are estimated by kernel methods. Monte-Carlo results with real data show that spell characteristics as well as other statistics are well reproduced. Tarboton et al. (1993) have developed a multivarite kernel density estimate (k.d.e.) with local bandwidths proportional to local covariance based on k nearest neighbours (similar in spirit to Lall and Bosworth (1993)), as well as requisite conditional k.d.e.'s for simulation of

streamflow time series. Tarboton (1994) evaluates the performance of the Colorado river annual streamflows simulated by SPIGOT (Grygier and Stedinger, 1990) through visual inspection of plots of k.d.e.'s of marginal p.d.f. of recorded and simulated traces.

The bootstrap is a simple non-parametric technique for simulating the probability distribution of a statistic or a specific feature of the distribution. The key idea is to resample from the original data, either directly or through a fitted model to create replicate data sets, from which the empirical probability distribution of the statistic of interest can be found (Davison and Hinkley, 1997). It is a good example of a new class of nonparametric statistical methods that substitute computer intensive computations for complex mathematical (parametric) models. Indeed, the bootstrap offers the potential for highly accurate inferencing and can eliminate the need to assume or impose a convenient model that does not have a strong scientific basis. At the same time, the basic idea motivating the bootstrap approach is conceptually simple. Resampling methods are applicable quite generally, and their implementation is usually automatic (Leger et al., 1992). However, while dealing with dependent data, it is a challenge to resample the records in such a way to ensure the preservation of the temporal and the spatial covariance structure of the original time series. With the advent of powerful computers, bootstrap resampling methods are emerging as potential techniques in modern statistical analysis for formulating inferential procedures such as constructing confidence regions, finding standard errors of estimates, carrying out tests of hypothesis. In addition, there is hope that the bootstrap can address complicated issues that arise in model selection. The use of bootstrap methods in time series analysis is receiving considerable attention in

modern statistics, as documented by Lepage and Billard (1992), Efron and Tibshirani (1993), Hjorth (1994), and Davison and Hinkley(1997).

The classical bootstrap resampling scheme was introduced by Efron (1979). This technique prescribes a data resampling strategy using the random mechanism that generated the data. In other words, it resamples with replacement from the empirical distribution function of independent and identically distributed (i.i.d.) data. Its applications for estimating confidence intervals and parameter uncertainty are well known (Tasker, 1987; Hardle and Bowman, 1988; Zucchini and Adamson, 1989). The procedure of synthetic streamflow generation addressed by Maass et al. (1967) is analogous to the classical bootstrap approach. Random resampling methods have been used in hydrology for comparing statistical methods, estimating parameter uncertainty, and comparing network design techniques. A few such works are those of Tasker (1987), Zucchini and Adamson (1988; 1989), Woo (1989) and Moss and Tasker (1991). When the random variables are i.i.d., this procedure provides very good approximation to the distribution of many commonly used statistics. However, for dependent random variables, the Efron's bootstrap fails (Lahiri, 1995).

To model dependent time series data, there are two popular bootstrapping approaches. One is known as the Model Based Resampling (MBR) approach and the other is the Moving Block Bootstrap (MBB) approach. Model based resampling for time series has been discussed by Freedman (1984), Freedman and Peters (1984) and Efron and Tibshirani (1986; 1993), Bose (1988), among others. In this method, to start with, a model structure is assumed, its parameters and residuals are estimated. Then, the

estimated model residuals are recentred around their mean and are resampled with replacement considering them as independent and identically distributed (i.i.d). Finally, these bootstrapped residuals are used to synthesize a time series. It is simple to apply and leads to good theoretical behavior, provided the fitted model is correct. The major drawback with this resampling scheme is that, in practice, the model structure is to be correctly identified and its parameters are to be accurately estimated from the data. If the chosen structure is incorrect, the generated series will be from a wrong model, and hence they will not have the same statistical properties as that of the original data (Davison and Hinkley, 1997).

The other popular approach to resampling in the time domain, known as the Block Bootstrap scheme, resamples blocks of consecutive observations. According to this, the data (of size N) is divided into b non-overlapping blocks, each of length L. Then, the synthetic replicates are constructed by resampling the blocks at random, with replacement. In this method, the original dependence structure is maintained within the blocks, but is destroyed at boundaries between blocks. This was developed by Hall (1985) and Carlstein (1986). Subsequently, Künsch (1989) and Liu and Singh (1992) independently proposed the Moving Block Bootstrap (MBB) approach for time series analysis. In this method, the observations from a univariate dependent sequence $x_1,\ldots,x_N$ are divided into blocks B_i of l consecutive observations starting with x_i; i.e., $B_i = (x_i,\ldots,x_{i+l-1})$, where i =1,...,b and b=N-l+1. Replicates are generated by resampling the overlapping blocks with replacement from the set (B_1,..., B_b), wherein each of the overlapping blocks B_i has equal probability (1/b) of being resampled. The

overlapping blocks picked at random are then pasted end-to-end to form a replicate of the historical sample.

The MBB provides "synthetic" time series that preserve the empirical probability distribution of the original observations. The MBB does not require one to select a model and the only parameter required is the block length (Vogel and Shallcross, 1996). The idea that underlies this block resampling scheme is that if the blocks are long enough, much of the original dependence will be preserved in the resampled series. This approximation is best if the dependence is weak (Davison and Hinkley, 1997). The number of blocks available for resampling should be large enough to ensure a good estimate of the distribution of the statistic. Use of long block sizes for resampling results in fewer blocks available for resampling as a result the synthetic replicates lack variety. Thus, unless the record length is considerable to accommodate longer and more number of blocks, the preservation of the correlation structure of the original series may not be possible, especially in cases of complex, long-range dependence structure. In such cases, the block resampling schemes tend to generate synthetic series that are less dependent than the original data. In some circumstances, this leads to very poor resampling approximations. Since this kind of bootstrap will never generate an observation either larger or smaller than the maximum or minimum historical observation, this technique is not useful for examining the probability distribution of the largest or the smallest observation, unless the sample size is greater than the planning horizon.

More recently, Lall and Sharma (1996) have used a nearest neighbour bootstrap technique for modelling dependent streamflows. Here, the dependence is preserved in a

probabilistic sense. This method involves searching the historical record to find the historical nearest neighbours and subsequently resampling their successors with a view to preserve the empirical dependence of the flow trace. Both MBB and k-NN resample the observed streamflow values for generating synthetic sequences and hence the simulations from these models cannot fill in the gaps between the data points in historical record. Also, they fail to provide extrema more severe than what is found in the historical record. Since the minimum extremes are of interest in modelling critical inter-annual low flow sequences, particularly in arid regions, such bootstrap methods may not be suitable for the prediction of critical droughts. The subsequent work of Sharma et al. (1997) avoids this limitation by resampling from the historic data with perturbations. The perturbations serve to smooth over the gaps between data points in the density estimate and provide alternate streamflow realizations that are different, but are statistically similar to the historical record (Sharma et al., 1997). However, since the streamflows are bounded, there is a possibility of leakage of probabilities across boundaries, when the perturbation is added and this may result in a bias in the simulated density in the neighbourhood of the boundary. In order to minimize this bias, appropriate kernel functions and/or bandwidths are to be chosen (Sharma et al., 1997), which may be a demanding task for any practising hydrologist. The nearest neighbour bootstrap technique and its variations are preferable if the data are plentiful, as in case of daily streamflow modeling (Lall and Sharma, 1996). The limitations of this technique for modelling monthly streamflows, have been brought out by Srinivas and Srinivasan (2001). They have reported that the standard deviation of historical flows during low flow months are underestimated by this technique. With increase in the model order (d >1), greater underestimation of the same is found to occur. In addition, the first few lag serial correlations are reported to get distorted in an effort to preserve higher lag serial correlations, when higher order models are tried.

The Generalized Cross Validation (GCV) score function (Craven and Wahba, 1979) can be used to choose the number of nearest neighbours and the order of the k-NN model. This is somewhat similar to the use of Akaike information criteria (AIC) for model selection in the traditional parametric modelling framework. A GCV based choice of the model order and the number of neighbours may be suboptimal for the particular water resources planning study under consideration, since it only considers the performance of the model with respect to conditional mean and variance (Rajagopalan and Lall, 1999). This necessitates further tuning to arrive at the appropriate combination of the model order and the number of neighbours for the study of interest.

Tasker and Dunne (1997) generate periodic streamflow traces based on a stochastic streamflow model (periodic autoregressive moving average model with log-transformation [PARMA(1,1)-LT]) using bootstrap resampling of residuals. While generating sequences for a single site, each month of bootstrap sequence of residuals is selected by randomly selecting a year with replacement and choosing the residual for the month for that year. This means that the selected residuals for each month in a bootstrap sequence may be from a different year. They mention that the residuals extracted from the fitted stochastic streamflow model will be approximately independent (since appropriate model is fitted at pre-whitening stage), but not identically distributed in time and that the seasonal differences of the residual distributions can be accommodated by bootstrapping the residuals for specific months. This model can be viewed as "model based resampling applied to periodic data". Srinivas (2001) shows the inadequacy of this

method in modelling the serial-dependence structure and the nonlinear dependence (estimated using state-dependent correlations) present in the streamflows.

Lall et al. (1996) present a stochastic model for resampling daily precipitation where the probability distribution functions (pdf's) of alternating wet and dry spells and of rainfall amount are estimated nonparametrically using kernel density estimators. This is equivalent to a bootstrap or sampling with replacement of the observed data sequence of spell lengths and precipitation amounts. Herein, smoothed empirical distribution functions are used for resampling, and sequential attributes of spells may be preserved. Necessary calibration parameters are chosen automatically from the data set using measures aimed at providing a good fit to the unknown underlying pdf. The application was to model the precipitatiion in the western USA, wherein during winter, it is in the form of snow due to orographic and frontal mechanisms, and convective rainfall processes are dominant in other seasons. Marked differences in the storm tracks and moisture sources over the seasons have been reported. A mixture of markedly different mechanisms (some related to the El Nino-Southern Oscillation) leads to the precipitation process in the western United States (Cayan and Riddle, 1992). It is unlikely that a robust parametric framework for model specification and selection can be devised for uniform application given the likely heterogeneity in precipitation generation mechanisms (Lall et al., 1996). The primary differences when compared with the traditional parametric wet/dry spell models are the following: (1) the relevant probability functions are estimated without recourse to prior assumptions as to the parametric form of the model, and (2) a more general conditional dependence structure is admitted.

According to Silverman (1986), sampling from the kernel density estimate (k.d.e.) can lead to reduced variance of the Monte-Carlo design. In addition, it also avoids the usual problem with the bootstrap "a number of the historical values being repeated in a generated sample", and provides an ability to fill in and extrapolate to a limited extent beyond the observed values. Moreover, arbitrary, finite component mixtures are readily admitted, without any hypothesization or formal identification. This provides a more direct and parsimonious representation of such structure if present in the data.

Recently, Tarboton et al. (1998) have developed a nonparametric temporal disaggregation model for single site periodic streamflow modelling. They have shown that a kernel density estimate of the joint distribution of the disaggregate flow variables can form the basis for conditional simulation based on an input aggregate flow variable. Being data-driven and relatively automatic, this method is able to model the nonlinearity in the dependence structure of the historical flows to a reasonable extent. The preservation of a variety of statistical attributes using this conditional simulation procedure has been demonstrated through applications to synthetic data and periodic streamflows from the San Juan river in New Mexico, USA. Possibly due to the smoothing of the kernel density estimate, some amount of bias is observed in the monthly standard deviations and skewnesses of the disaggregated flows from this nonparametric disaggregation model (see Figs. 8 and 9, Tarboton et al. 1998). Even though the marginal distributions and the state-dependent correlations of observed flows are reported to be better preserved compared to SPIGOT (Fig.11, Tarboton et al., 1998), further improvement is desirable. The drawbacks of this method as given by Tarboton et al. (1998) are: (1) it is data and

computationally intensive; (2) estimating an optimal bandwidth to use is a computationally demanding task. As with the method of moments and the method of maximum likelihood in the parametric case, different optimality criteria can lead to quite different bandwidths being selected; (3) the choice of kernel function is not crucial, but the parameterization of the bandwidth matrix in the multivariate case may affect the results dramatically; (4) the sample size required increases, as the complexity of the underlying density function increases, thus reducing the advantage of the NPD approach for heterogeneous functions; (5) no simple equation for the model is available to report. Disaggregation of annual flows to monthly flows at a single site using this nonparametric disaggregation (NPD) model requires the estimation of a complex, 13-dimensional density function. Tarboton et al. (1998) report that the data points in a monthly historical record of length 80 years are inadequate for the accurate estimation of the 13-dimensional complex joint probability density function in terms of statistical efficiency criteria.

3.2.3 Hybrid Moving Block Bootstrap Method

More recently, hybrid moving block bootstrap (HMBB) method has been introduced by *Srinivas and Srinivasan* (2000; 2001a; 2001b) for stochastic modeling of periodic streamflows that exhibit complex dependence, based on the post-blackening approach of *Davison and Hinkley* (1997)..This approach suggests using a parsimonious linear parametric model for partial pre-whitening of the observed streamflows. The structure present in the residuals extracted from the partial pre-whitening stage is simulated by MBB. The resulting innovations are post-blackened to synthesize the replicates of the observed flows. Hereafter, this model will be referred to as Hybrid Moving Block

Bootstrap (HMBB). In order to preserve the autocorrelation structure at the annual level and the cross-year serial correlation structure (that are essential for the efficient prediction of reservoir storage statistic and modeling critical run characteristics), HMBB requires resampling long blocks of residuals, particularly when the cross-year dependence is strong. This is because HMBB uses MBB for resampling the residuals, the limitations of which have already been mentioned. As a result, variety and smoothing in the simulations get reduced, which in turn, reduces the variability in the simulated critical run characteristics and reservoir storage capacity (validation statistics of direct interest to investigator), thus affecting the design decisions.

3.3 MOTIVATION and PRESENT STUDY

The motivation for the method of periodic streamflow modeling presented in this research work comes from a desire to develop a potential nonparametric stochastic model that is effective in reproducing summary statistics, dependence structure and the salient features of marginal distribution (multimodality, peakedness and asymmetry) without compromising on smoothing, extrapolation and variety in simulations. Such an ideal model is expected to be effective in reproducing validation statistics [*Stedinger and Taylor*, 1982] such as storage capacity and critical/mean run characteristics that are of interest to the investigator. In this regard, the matched block bootstrap method presented in recent works [*Hesterberg*, 1997; *Carlstein et al.*, 1998] seems to be useful. These works suggest improving the performance of MBB in modeling data with strong and/or long range dependence through matching rules for resampling moving blocks. Out of a few matching rules recommended by *Carlstein et al.* [1998], the rank matching rule was

found to be more accurate and generally satisfactory [*Hesterberg*, 1997]. In a rank matching procedure, the blocks are matched using a single value at the beginning or the end of a block. It is well suited for Markov processes, in which the last value in a block is assumed to contain all the information in the block for predicting future observations.

In this research work, a new nonparametric simulation model is proposed to synthesize multi-season (periodic) streamflows, based on the rank matching idea of *Carlstein et al.* [1998]. In the proposed method, non-overlapping within-year blocks (formed from the observed time series) are first resampled using rank matching rule of *Carlstein et al.* [1998] and these resampled blocks are subsequently perturbed using a simple weighted smoothing strategy to achieve smoothing and extrapolation in simulations.

The following sections will deal with: i) the description of the algorithm of the perturbed matched block bootstrap (PMABB) method that is proposed in this research work; ii) the evaluation of the performance of PMABB through application to synthetic data from a known self-exciting seasonal threshold autoregressive moving average model; and iii) the comparison of the performance of the proposed model with the periodic k-nearest neighbour bootstrap technique (Lall and Sharma, 1996; Srinivas and Srinivasan, 2001a) (referred as Pk-NN) and the hybrid stochastic streamflow model introduced by *Srinivas and Srinivasan* (2001b) (referred as HMBB), using the monthly streamflows measured at two sites, one on the river Narmada and the other on the river Hemavathi (a major tributary of Cauvery). Finally a set of conclusions is drawn.

3.4 ALGORITHMS

This section describes the model structure and the algorithms of: (i) the Periodic k-Nearest Neighbour (Pk-NN) resampling technique; (ii) the Hybrid moving block bootstrap (HMBB) method and (iii) the Perturbed Matched Block Bootstrap (PMABB) method..In these descriptions of the model algorithms, bold upper case letters will represent vectors.

3.4.1 The Periodic k-Nearest Neighbour (Pk-NN) Resampling Algorithm

The k-NN bootstrap method for resampling hydrologic time series was proposed by Lall and Sharma (1996). This method has been developed for modelling dependent data and it preserves the dependence in a probabilistic sense. Multivariate nearest neighbour probability density estimation provides the basis for this method. Herein, the k-NN resampling algorithm of Lall and Sharma (1996) is adapted to model periodic streamflows and hence the same is referred to as Periodic k-Nearest Neighbour (Pk-NN) resampling algorithm.

Let the time series of observed periodic streamflows be denoted by $Q_{\nu,\tau}$, where ν is the index for year (ν =1,..., N) and τ denotes the index for month within the year (τ = 1, ..., ω), N refers to the number of years of historical record and ω represents number of months (=12) within the year.

Let the hydrological water year start with the month of June of a calendar year and end with the month of May of the subsequent calendar year. Now, the first value to be

simulated will be June month flow. For this, one has to pick randomly any one of the N June month flows from the historical data. Let it be denoted as $q_{i,j}$, where i is the water year to which the flow value belongs.

Following are the sequential steps involved in the synthetic simulation of observed periodic streamflow data using this method:

1. Define the composition of the "feature vector" of dimension d.

   For example, for order of dependence equal to two (d=2), initial feature vector for simulating the June month flow will be the conditioning set $\{q_{i-1,\omega}, q_{i-1,\omega-1}\}$. This represents the dependence of the June flow to be simulated on two prior monthly flows (i.e, May and April flows of the previous water year respectively).

   The historical state vectors D_τ for any month τ, are the feature vectors of all $q_{i,\tau}$ in the historical record. For example, for simulating June month flow, the historical state vectors will be: $\{q_{1,\omega}, q_{1,\omega-1}\}, \ldots, \{q_{N,\omega}, q_{N,\omega-1}\}$

2. Denote the current feature vector as D_i and determine its k nearest neighbours from among the historical state vectors for that month D_τ, using the weighted Euclidean distance r_{iv}.

$$r_{iv} = \left(\sum_{j=1}^{d} w_j (v_{ij} - v_{vj})^2 \right)^{1/2} \qquad (3.1)$$

In eq.(4.5), r_{iv} is the weighted Euclidean distance from the current feature vector to the "v th" historical state vector among the historical state vectors D_τ for the month τ; "v_{ij}" is the "j th" component of the current feature vector; and "v_{vj}" stands for the "j th" component of the "v th" historical state vector. The weights w_j are chosen a priori as inverse of some measure of scale such as standard deviation or range of $\mathbf{V}_j$ that comprises of the j th components of the historical state vectors D_τ for the month τ. The number of neighbours "k" is a smoothing parameter. It may be chosen using any appropriate order selection strategy such as generalised cross-validation (GCV) (Craven and Wahba, 1979). Lall and Sharma (1996) suggest using k equal to square root of the sample size as a rule of thumb.

3. Denote the ordered set of nearest neighbour indices by $J_{i,u}$, where u=1,...,k. An element j(i) of this set records the time v associated with the j-th closest historical state vector to D_j among D_τ. Denote $x^S_{j(i)}$ as the successor to $D_{j(i)}$. If the data are highly quantized, it is possible that a number of observations may be at the same distance from the conditioning point, in which case a permuting may help.

4. Define a discrete kernel K(j(i)) for resampling one of the $x^S_{j(i)}$ as follows :

$$K(j(i)) = \frac{1/j}{\sum_{j=1}^{k} 1/j} \qquad (3.2)$$

where $K(j(i))$ is the probability with which $x^S_{j(i)}$ is resampled. It is to be noted that this resampling kernel is the same for any i, and can be computed and stored prior to the start of simulation. Lall and Sharma (1996) develop this kernel through a local Poisson approximation of the probability density function of state space neighbours.

5. Using the discrete probability mass function $K(j(i))$, resample an $x^S_{j(i)}$ and update the current feature vector. Proceed to step 2, if additional simulated values are required to be generated. For a more detailed discussion on the k-NN algorithm, the reader is referred to Lall and Sharma (1996) and Rajagopalan and Lall (1999).

3.4.2 Periodic Hybrid Moving Block Bootstrap (HMBB)

This section presents a new algorithm for generating periodic synthetic streamflows that extends the Post-blackening approach suggested by Davison and Hinkley (1997).

Let the observed (historical) streamflows be represented by the vector $Q_{v,\tau}$, where v is the index for year (v=1,...,N) and τ denotes the index for season (period) within the year (τ =1,...,ω), N refers to the number of years of historical record and ω represents the number of periods within the year. The modeling steps involved are as follows:

1. Standardize the elements of the vector $Q_{v,\tau}$ as:

$$y_{v,\tau} = \frac{q_{v,\tau} - \bar{q}_\tau}{s_\tau} \qquad (3.3)$$

where $\bar{q}_\tau$ and s_τ are respectively the mean and the standard deviation of the observed streamflows in period τ. Note that the historical streamflows are not transformed to remove skewness.

2. Pre-whiten the standardized historical streamflows, $Y_{v,\tau}$, using a simple periodic autoregressive model of order one (PAR(1)), and extract the residuals $\varepsilon_{v,\tau}$.

$$\varepsilon_{v,\tau} = y_{v,\tau} - \varphi_{1,\tau}\, y_{v,\tau-1} \qquad (3.4)$$

In eq. (3.4), $\varphi_{1,1}, \ldots, \varphi_{1,\omega}$ are the periodic autoregressive parameters of order one. Herein, for the parameter estimation, a simple method of moments (Salas et al., 1980) has been used. It is to be noted that the residuals $\varepsilon_{v,\tau}$ may possess some weak dependence (since the parameters are estimated from a simple PAR(1) model). Herein, it is to be mentioned that bootstrap schemes like moving block bootstrap (MBB) (Künsch, 1989) can serve as a reliable tool for modeling the weak linear dependence, if any, in the residuals. Moreover, this scheme being data-driven, can be expected to capture the marginal distribution features, and to a certain extent may be able to preserve the non-linear dependence inherent in the observed record, possibly with some trade-off with regard to smoothing and generation of extrema, when compared to parametric models.

3. Obtain the simulated innovations $\varepsilon^{*}_{v,\tau}$ by bootstrapping $\varepsilon_{v,\tau}$ using the Moving block bootstrap (MBB) (Künsch, 1989) method. Herein, the monthly residuals resulting from the PAR(1) model are divided into q number of (possibly) overlapping

blocks B_i with block size "l" taken as an integral multiple of the number of periods (ω) within the year. It is to be noted that each of the overlapping blocks starts with the first period of a hydrological water year. This is done with a view to capture the within-year correlations for significant number of lags. For example, the block sizes of residuals in monthly streamflow modelling context would be 12, 24, 36, and so on (abbreviated as $l = \omega$, $l = 2\omega$, $l = 3\omega$, and so on). Note that when block length l is "n" years long, the overlap is (n-1) years, indicating that when the block size is one year long, there is no overlap.

In general, the i th block of size $l = m\omega$ (m is a positive integer, such that, m=1,...,N), may be written as : $B_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i+m-1,\omega})$

where $i = 1, \ldots, b$ and $b = N-m+1$.

For example, if $l = 3\omega$ and $\omega = 12$, the 4-th block is written as: $B_4 = (\varepsilon_{4,1}, \ldots, \varepsilon_{6,12})$.

The block size l, to be selected for resampling the residuals, would primarily depend on the amount of unextracted weak dependence present in the residuals. Innovations $\varepsilon^*_{\nu,\tau}$, are generated by resampling the overlapping blocks B_i, at random, with replacement from the set ($B_1, \ldots, B_b$) and pasting them end-to-end. It is to be noted that each of the (possibly) overlapping blocks has equal probability (1/b) of being resampled.

4. The innovation series $\varepsilon^*_{\nu,\tau}$ is then "post-blackened" by using eq.(3.5) to obtain the sequence $Z_{\nu,\tau}$.

$$z_{y,\tau} = \varphi_{1,\tau}\, z_{y,\tau-1} + \varepsilon^{*}_{y,\tau} \qquad (3.5)$$

The synthetic generation process starts with $z_{1,0} = 0$. The "warm-up" period is chosen to be large enough to remove any initial bias. The values of $Z_{y,\tau}$ are then inverse standardized using eq. (3.6), to obtain the synthetic streamflow replicate $X_{y,\tau}$.

$$x_{y,\tau} = (z_{y,\tau} \cdot s_{\tau}) + \bar{q}_{\tau} \qquad (3.6)$$

It is to be noted that no normalizing transformation is applied in case of the hybrid model. Herein, it is to be mentioned that, when the number of data points in the historical record is limited (as in case of annual streamflow modeling), the mean of residuals recovered from the pre-whitening stage need not be necessarily equal to zero. In such a case, the residuals are to be re-centered to zero before proceeding with resampling them for generating the innovation series (Davison and Hinkley, 1997; p.397). However, when the data points are relatively plentiful (as in case of periodic streamflow modeling), we find that the sum of residuals recovered from partial pre-whitening stage tends to zero and hence the residuals need not be re-centered.

3.4.3 Perturbed Matched Block Bootstrap (PMABB)

In this section, we present a new nonparametric method "Perturbed Matched Block Bootstrap (PMABB)" for simulation of multi-season hydrologic time series, based on the idea of rank matching proposed by *Carlstein et al.* [1998]. The rank matching method suggests constructing a Markov chain by resampling the blocks formed from the given data sample by aligning with higher likelihood those blocks that match at their ends. Here, we propose the algorithm, in general, for resampling within-year blocks of either

equal or unequal lengths. In cases where the identification of the different seasons within the water year is clear, the block lengths can be chosen accordingly, so that the complex dependence structure present in the underlying hydrologic process can be well recovered. In the description to follow, vectors will be represented by bold upper case letters.

Let the observed (historical) periodic streamflows be represented by the vector $\mathbf{Q}_{v,\tau}$, where v is the index for year ($v = 1,\ldots,N$) and τ denotes the index for period within the year ($\tau = 1,\ldots,\omega$), N refers to the length of observed record (in years), and ω represents the number of periods within the year.

For each year of the observed record, prepare "n" number of non-overlapping within-year blocks $B^{w}_{v,1},\ldots,B^{w}_{v,n}$ with the respective lengths being $L^{w}_{1},\ldots,L^{w}_{n}$, such that the lengths of all the within-year blocks sum to ω. i.e., $\sum_{j=1}^{n} L^{w}_{j} = \omega$. Herein, $B^{w}_{v,i}$ denotes the "i" th within-year block in the year "v" of the observed streamflow record.

$$B^{w}_{v,i} = \{q_{v,S+1},\ldots,q_{v,S+L^{w}_{i}}\} \quad \text{where} \quad S = \begin{cases} 0 & \text{if } i = 1 \\ \sum_{j=1}^{i-1} L^{w}_{j} & \text{otherwise} \end{cases} \tag{3.7}$$

Let $e^{w}_{v,i}$ denote the last flow value (end element) of $B^{w}_{v,i}$.

$$e^{w}_{v,i} = q_{v,S+L^{w}_{i}} \quad \text{where} \quad S = \begin{cases} 0 & \text{if } i = 1 \\ \sum_{j=1}^{i-1} L^{w}_{j} & \text{otherwise} \end{cases} \tag{3.8}$$

Form the sets $\mathbf{E}^{w}_{i}$: $\mathbf{E}^{w}_{j} = \{e^{w}_{1,i},\ldots,e^{w}_{N,i}\} \quad 1 \le i \le n$

Arrange the elements of $\mathbf{E}_i^w$ in ascending (or descending) order of their magnitude and assign ranks. Let $r_{v,i}^e$ denote the rank of $e_{v,i}^w$ for $1 \le i \le n$ and $1 \le v \le N$. The algorithm is initialized by randomly selecting one of the "N" first within-year blocks $B_{1,1}^w, \ldots, B_{N,1}^w$. This block is referred to as *current within-year block* at the start of simulation.

The key steps in the resampling algorithm are as follows:

(i) Identify the rank of the end element of the current within-year block. Let it be denoted by R_c.

(ii) Select the nearest neighbor to the current within-year block. For this, randomly select one of the "2m+1" neighboring blocks to the current block, whose end elements have ranks between $(R_c - m)$ and $(R_c + m)$, where m is a small positive integer. This requires generating a uniform random number "U" in the range of integers $(R_c - m)$ and $(R_c + m)$. Whenever $(R_c - m)$ becomes less than 1 or $(R_c + m)$ becomes greater than "b", one may end up with a value of U, that is either less than 1 or greater than b. Such a rank does not correspond to any block. This artifact can be overcome by folding back (or reflecting) U to U^*. Hereafter, this procedure would be referred to as method of reflection for rank. Neighboring block is the one that corresponds to the rank U^* in the given record.

$$U^* = \begin{cases} 2b+1-U & \text{if} \quad U > b \\ 1-U & \text{if} \quad U < 1 \end{cases} \tag{3.9}$$

The neighborhood "2m+1" to a current within year block is referred to as band width.

(iii) Obtain the within-year block that follows the selected neighboring within-year block in the given record and append it to the current within-year block. Note that, if the current within-year block is the last within-year block in a year, then the block to be appended will be the first within-year block in the next year. However, if the selected within-year block is $B^{w}_{N,n}$, then it does not have any block following it. One option to overcome this problem is to create a circular time series such that the block following $B^{w}_{N,n}$ would be $B^{w}_{1,1}$. Another option is to reflect the within-year blocks of the last year of the given record using $B^{w}_{v,i} = B^{w}_{2N+1-v,i}$ for v = N+1, so that the block following $B^{w}_{N,n}$ (i.e., $B^{w}_{N+1,1}$th block) would be $B^{w}_{N,1}$. Herein, the latter option referred to as method of reflection of blocks (MRB) is used for coping with this problem.

(iv) The recently appended block becomes the new current within-year block. Proceed to step (i), if additional simulated values are required to be generated.

One of the options for synthesizing replicates could be to break the long sequence of simulated values (appended blocks) into non-overlapping blocks of length equal to that of the observed record. Other option would be to initialize simulation of each replicate through random selection of current block. The latter option would enhance the chance of selection of the first block $B^{w}_{1,1}$ when MRB method is used in resampling the blocks. Otherwise the first block in the data set would never be selected in the simulation (except upon initialization), because it does not follow any block with end element in vector $\mathbf{E}^{w}_{i}$.

Smoothing the resampled blocks

Let the sequence resulting from concatenating the blocks resampled using rank matching rule be denoted by the vector $\mathbf{X}_{\nu,\tau}$. One can smooth $\mathbf{X}_{\nu,\tau}$ by using an appropriate perturbation strategy, such as the weighted smoothing adopted herein. However, it has to be ensured that the perturbation does not have any adverse effect on the reproduction of the statistical attributes of $\mathbf{X}_{\nu,\tau}$. In literature, smoothing plans have been developed for many kinds of data [*Velleman et al.*, 1981; *Bowman and Azzalini*, 1997]. Herein, we use weighted smoothing with a window size of 12 months (equal to the number of seasons in a water year). Random weights "W_ν" are generated in the range p of real numbers $\Re$ [1-δ, 1+δ] for each year ν of the resampled series, where δ is a small positive fraction. The simulated time series $\mathbf{Y}_{\nu,\tau}$ is

given by:
$$Y_{\nu,\tau} = W_\nu \times X_{\nu,\tau} \quad \text{for } \nu = 1,\ldots,N;\ \tau = 1,\ldots,\omega \tag{3.10}$$

Typically, for finite samples of size 50 to 200 years, a choice of δ ranging from 0.05 to 0.15 has been found appropriate. Even though the choice of higher value of δ enhances smoothing and extrapolation value in simulations, it affects the performance of the model in simulating the correlation structure and other statistics of concern. Consequently, the validation performance of the model, measured in terms of prediction of storage and run characteristics, drops. Thus, caution should be exercised in opting for a higher δ value.

The issue of choosing the number of neighbours w (= $2m$ +1) to be used in rank matching, has been discussed in detail by *Hesterberg* [1997] and *Carlstein et al.* [1998]. They have provided suggestions for the number of neighbors based on minimizing

asymptotic mean square error of the logarithm of the variance of the sample mean considering sample sizes ranging from 200 to 5000. However, while dealing with short hydrologic records, it would be appropriate to experiment with different possible choices of "m" and select the effective value to be adopted, after analyzing the resulting behavior (*E. Carlstein*, Personal communication, 2000). The information required to run the source programs developed for PMABB (periodic streamflow model) is given in Appendix – 1.

3.5 APPLICATION STUDY

In this section, the ability of the proposed nonparametric method in recovering the statistical attributes from a known population is tested. The proposed method is then applied to simulate monthly streamflows of the Narmada and the Hemavathi rivers. The performance of the model is verified in terms of preservation of the following statistical attributes: (i) summary statistics (mean, standard deviation and skewness); (ii) serial correlations (both within-year and cross-year); (iii) autocorrelations at the aggregated annual level; and (iv) state-dependent correlations (introduced by Sharma et al. [1997] as a measure of nonlinear dependence).

3.5.1 Test with Synthetic Data

To test the proposed multi-season PMABB method for its ability to recover the statistical attributes including the dependence structure (both linear and non-linear) from a known population, we use the self-exciting seasonal threshold ARMA (SESTARMA) model described by *Srinivas and Srinivasan* [2001a]. A two-level Monte-Carlo experiment is designed. In the first level, 100 samples, each of size 4N (N years × 4 seasons), are generated from SESTARMA model (known population). These 100

samples are referred to as level-1 samples. The second level involves generating 100 replicates, each of size 4N (N years × 4 seasons), for each of the 100 level-1 samples, using the proposed PMABB method. The 10,000 replicates resulting from the Monte-Carlo simulations at the second level are referred to as level-2 replicates.

The SESTARMA model used for generating level-1 samples is:

First season:

$$x_t = 0.2x_{t-1} + 0.35x_{t-2} + 0.6W_t \quad \text{if } x_{t-1} \le 0$$

$$x_t = 0.9x_{t-1} - 0.16W_{t-1} + 0.6W_t \quad \text{otherwise}$$

Second season:

$$x_t = 0.5x_{t-1} - 0.12W_{t-1} + 0.7W_t \quad \text{if } x_{t-1} \le 0$$

$$x_t = 0.9x_{t-1} + 0.2x_{t-2} + 0.7W_t \quad \text{otherwise}$$

Third season:

$$x_t = 0.45x_{t-1} - 0.245W_{t-1} + 0.5W_t \quad \text{if } x_{t-1} \le 0$$

$$x_t = 0.15x_{t-1} + 0.3x_{t-2} + 0.5W_t \quad \text{otherwise}$$

Fourth season:

$$x_t = -1.0 - 0.5x_{t-1} - 0.1W_{t-1} + 0.8W_t \quad \text{if } x_{t-1} \le 0$$

$$x_t = 0.8x_{t-1} + 0.2x_{t-2} + 0.8W_t \quad \text{otherwise} \qquad (3.11)$$

where W_t is a Gaussian random variate with zero mean and unit standard deviation.

Preservation of the various average statistical attributes over the 100 level-1 samples (indicative of the population statistics), by the 10000 level-2 replicates generated from PMABB would indicate that the candidate model is able to recover the structure contained in the population effectively.

A comparison of the population statistics and the simulated statistics from 10,000 replicates, is presented in Figure 3.1 for a typical sample size N = 40 (for brevity) using box-plots. In these box-plots, the span of the box (inter-quartile range) and the whiskers together indicate the sampling variability associated with each statistic. It is evident from the figure that the PMABB model is able to reproduce the summary statistics and the dependence structure (both linear and non-linear) fairly well. These results suggest that the multi-season PMABB model can recover the statistical attributes including the dependence structure from a known population (Fig. 3.1).

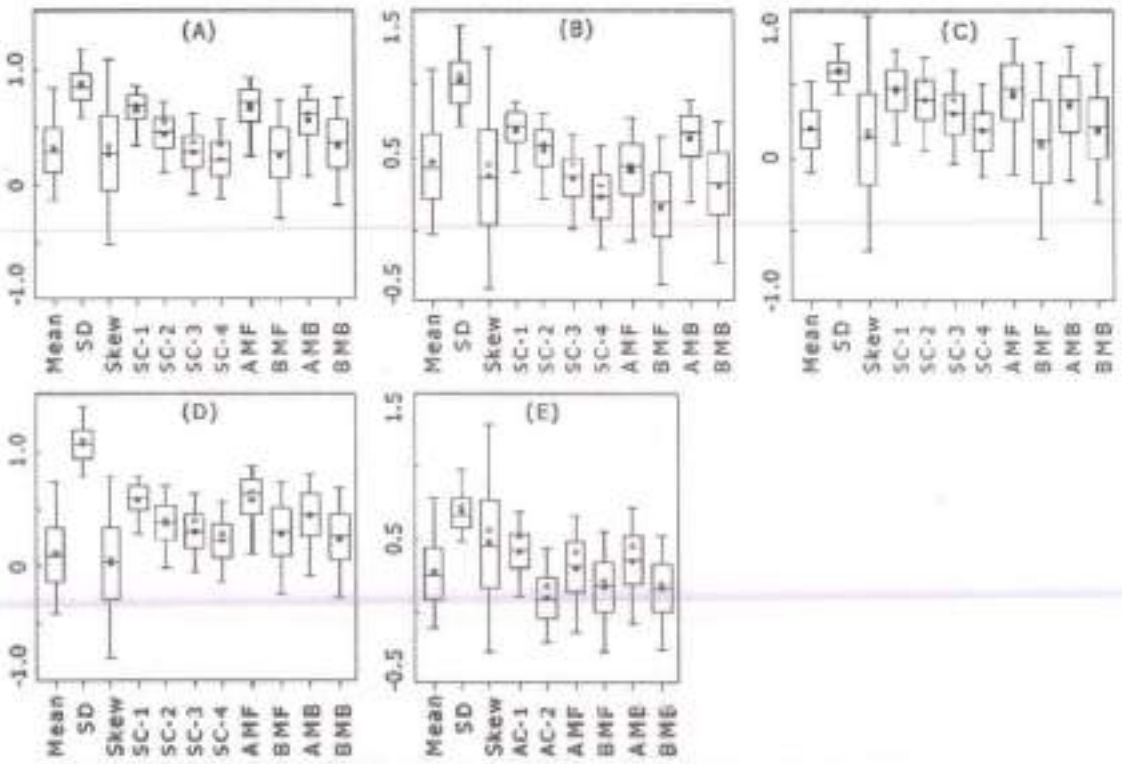


Figure 3.1: Preservation of the summary statistics and the dependence structure of level-1 samples by level-2 replicates from PMABB (L= 2; w = 3; p = 0.90-1.10) model for sample size 40. (A) Season-1;(B) Season-2; (C) Season-3; (D) Season-4 and (E) Annual. The circle denotes the average value of statistic over 100 level-1 samples and the darkened square indicates the mean value of statistic over 10,000 level-2 replicates. A line in the middle of the box represents median. SC-1, SC-2, SC-3, and SC-4 denote serial correlations for lag-1, lag-2, lag-3, and lag-4 respectively. AC-1 and AC-2 refer to autocorrelations for lag-1 and lag-2 respectively.
AMF: forward and above median correlation; BMF: forward and below median correlation;
AMB: backward and above median correlation; BMB: backward and below median correlation.

3.5.2. Test with Real Data Sets

In this section, the proposed multi-season PMABB model is applied to the monthly streamflow records measured at (i) Jamtara gauge site (just downstream of the Bargi dam site) on the river Narmada and (ii) Akkihebbal gauge site located on the river Hemavathi, one of the major tributaries of the river Cauvery. To demonstrate effectiveness of the proposed method, the results from the application of the PMABB model is compared with those from (i) the periodic k-NN model (Lall and Sharma, 1996; Srinivas and Srinivasan, 2001); and (ii) the Hybrid periodic stochastic model, HMBB (Srinivas and Srinivasan, 2001a,b).

The Narmada river rises in the Mikel range in Shahdol district near Amarkantak at an elevation of 1050 m. It passes through the states of Madhya Pradesh and Gujarat and flows westwards to the Arabian sea. The unregulated streamflows measured at the gauge site Jamtara (located at 16 km downstream of Bargi dam site) have been used. Systematic gauging has been done at this site since 1949 (NIH, 1996). The average annual flow at the dam site is 7197 Mm^3. Most part of the streamflows are received during the southwest monsoon months, namely, July to October. The Bargi dam is a multi-purpose project that caters to water supply (for domestic and industrial consumption), irrigation and hydropower. The index map (Fig. 3.2) shows the dam and the river gauging site on the river Narmada. The monthly inflow data at the Bargi reservoir for the period 1951-1990 (Table 3.1) as provided in the technical report TR(BR) 143 of the National Institute of Hydrology (NIH, 1996) has been used as the first case example for the periodic stochastic streamflow modeling in this study.

Table 3.1: Monthly Inflow series at the Bargi reservoir for the period 1951-1990

YEAR	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APL	MAY
1951	35.63	98.66	2307.1	1164.3	355.6	51.7	31.6	14.86	9.93	6.43	3.1	0.81
1952	116.3	2023	3745.1	2068.2	224.45	63.76	34.61	34.12	21.74	7.77	2.15	0.91
1953	0.4	1688.2	2524.9	1060.3	213.65	65.56	36.22	21.61	9.48	3.89	2.79	1.56
1954	12.59	1137.1	1967.5	2398.3	307.12	79.43	28.92	24.1	18.23	8.25	3.35	1.18
1955	447.16	1227.3	3179.1	3490.5	1257.6	216.72	84.45	40.49	22.35	10.69	6.07	3.57
1956	158.69	3650.3	5869.4	1433	688.17	281.78	92.77	97.89	36.09	52.5	31.08	8.11
1957	13.83	1739.2	3552.8	1315.3	195.33	59.71	31.98	18.82	13.5	25.72	7.63	1.48
1958	4.4	2022.8	1865.1	1620.2	1146.3	187.1	66.67	51.68	40.11	11.7	8.13	3.37
1959	2.95	2694.9	4152.7	2863.3	536.59	124.44	61.72	68.46	31.73	21.06	18.36	5.51
1960	85.27	1564.9	3714.4	853.95	737.19	128.97	65.74	48.53	45.37	17.89	5.92	1.59
1961	241.28	4981.9	4355.6	4547.7	949.99	218.43	131.4	72.34	42.36	31.23	18.89	7.13
1962	77.21	763.68	1974	1638.2	260.48	84.62	302.04	55.73	22.05	14.83	8.87	8.76
1963	118.06	963.5	2636.3	2679.9	248.68	110.62	53.67	111.05	24.04	15.6	5.38	2.21
1964	66.73	2712.8	4626.1	2682.1	483.77	114.09	62.24	40.63	24.41	14.86	25.04	3.34
1965	103.82	427.64	572.82	1059.4	109.65	35.97	20.62	18.68	10.77	3.92	1.54	1.15
1966	234.88	1084.8	2189.5	290.08	56.62	23.56	22.42	13.59	8.17	33.94	28.35	4.74
1967	101.37	2610.6	5149.1	2616.8	352.06	91.05	132.43	191.24	72	42.48	11.28	6.45
1968	17.69	1087.8	3079.2	610.52	205.2	65.56	38.93	32.39	10.78	6.56	3.52	1.5
1969	1.83	1490.1	4679.4	1258.1	224.03	89.12	42.13	37.87	18.44	46.3	9.09	7.39
1970	239.99	1240.5	3131.7	5231.1	361.04	179.74	64.77	41.88	35.33	28.9	15.51	14.69
1971	1093.1	3736.7	2477.9	2163.5	783.91	116.77	59.32	46.77	31.9	16.74	9.39	4.04
1972	6.44	283.75	4001.1	1743.5	196.83	93.74	74.29	37.2	86.41	14.69	6.81	1.5
1973	9.52	1945.8	4011.9	2719.1	1122.6	153.09	84.21	56.75	35.26	16.94	8.37	3.24
1974	29.37	454.58	3178.6	204.49	159.44	41.14	33.13	24.97	16.67	13.22	6.24	1.3
1975	87.97	1817.9	4624.6	1654.2	954.54	156.84	83.36	54.05	34.07	16.25	11.53	5.68
1976	7.61	555	2096	870.45	51.46	36.63	31.11	18.84	13.53	10.46	3.4	2.16
1977	873.99	1939.1	3603.2	1470.2	386.85	171.29	65.5	65.86	99.78	55.43	17.78	6.41
1978	189.8	1602.9	2681	776.59	123.71	55.42	119.28	65.96	96.3	27.31	7.77	4.79
1979	15.91	378.64	1377	130.37	38.22	17.9	13.67	11.3	5.7	3.72	1.98	0.61
1980	451.76	2535.9	4186.5	2557.8	240.95	88.07	55.64	49.05	25.26	17.66	7.52	3.58
1981	119.21	1293.7	1359.2	478.73	278.6	79.65	50.1	50.45	49.83	18.8	5.95	4.41
1982	26.24	185.86	2654.7	1034.7	145.66	87.64	40.53	32.25	33.09	12.08	4.21	1.55
1983	11.8	716.17	2084.9	2497.3	719.62	93.42	71.68	110.68	94.52	30.91	12.32	8.61
1984	30.37	360.84	4742.9	1068	138.51	72.32	41.25	113.55	129.33	19.31	8.17	4.47
1985	12.29	853.82	3521.4	887.65	596.99	188.32	47.36	33.63	125.97	54.42	14.19	5.69
1986	179.71	2156	1680.5	622.72	498.53	138.32	66.45	79.16	58.58	56.29	10.23	3.64
1987	5.94	50.9	733.4	22.96	415.4	177.52	98.23	71.86	31.45	25.36	4.16	3.19
1988	293.69	2127.7	4840.7	920.38	255.13	258.94	190.38	50.96	52.15	40.07	19.16	4.18
1989	100.98	417.35	490.96	221.96	328.43	329.31	369.19	211.67	65.13	121.6	111.73	130.83
1990	35.63	98.66	2307.1	1164.3	355.6	51.7	31.6	14.86	9.93	6.43	3.1	0.81

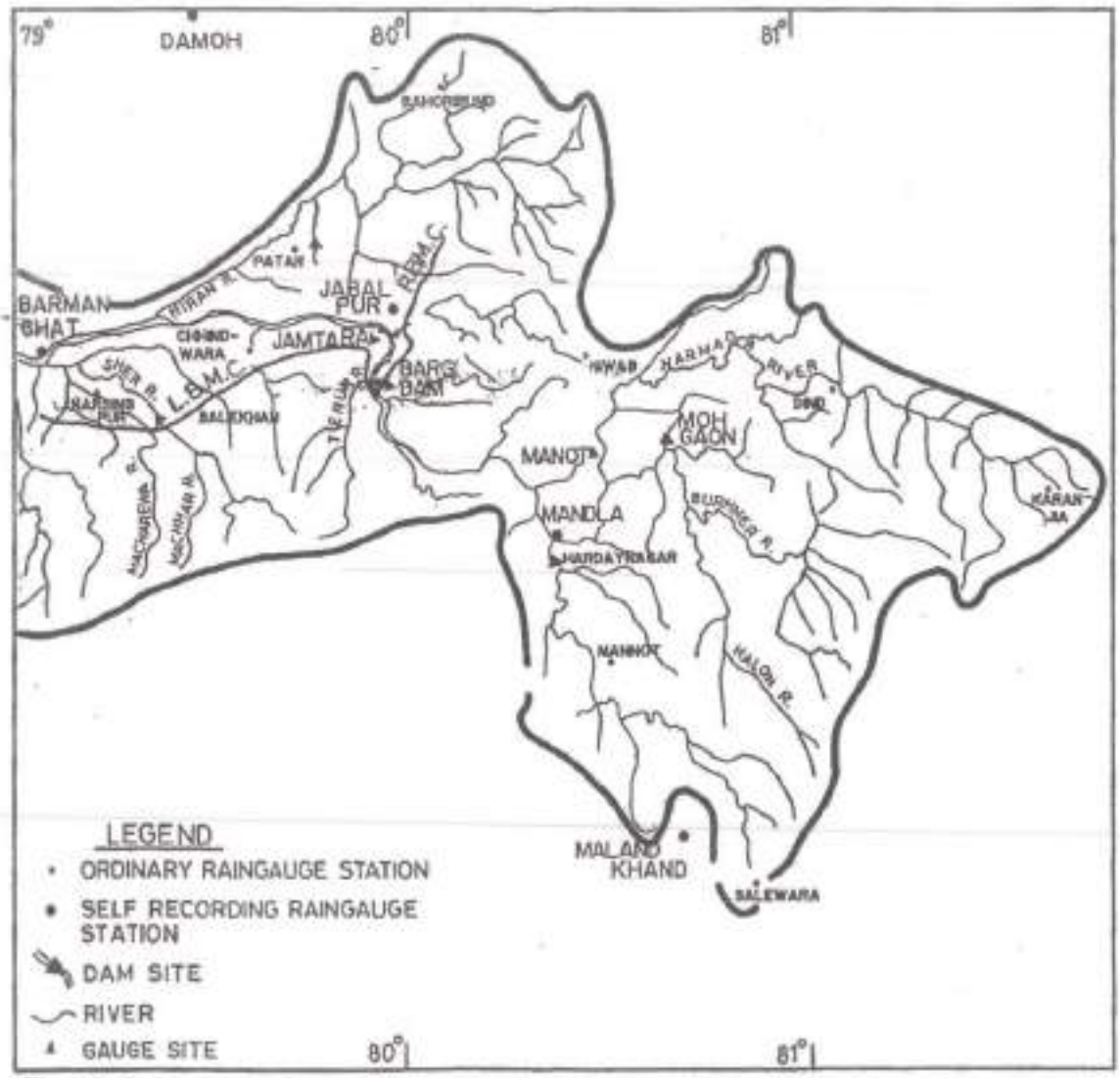


Fig 3.2: Index Map of Bargi Dam Site (Source: Report No. TR(BR) 143, NIH, 1996)

The Hemavathi river is one of the main tributaries of the river Cauvery (Fig. 3.3). Its source is in the Mudigere taluk of Chickmagalur district of Karnataka state in Southern India. The river travels a distance of 193 km through the districts of Hassan and Mandya before joining the river Cauvery in the water spread of Krishnarajasagar (KRS) reservoir in Mandya district. The total drainage area of the river is 5910 km^2. The mean monthly

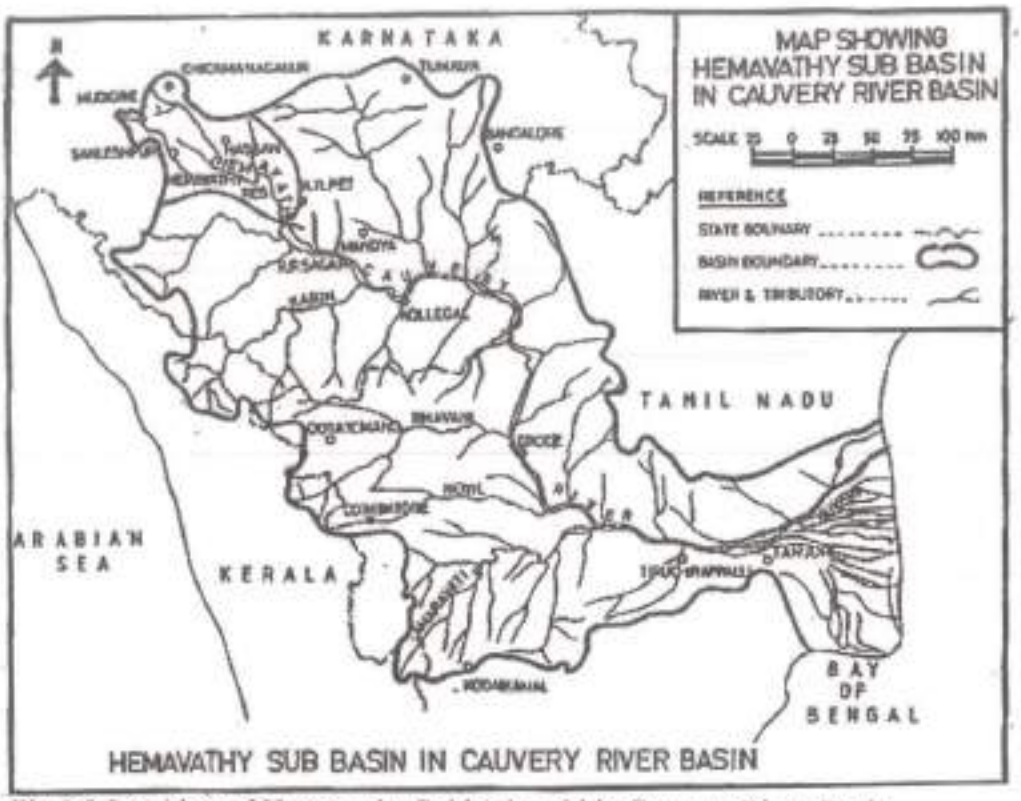


Fig 3.3 Location of Hemavathy Subbasin within Cauvery River Basin
(Source: Report No. TR-53, NIH, 1986)

temperatures vary in the range 18 to 32 degrees Celsius and the predominant soils are red loam and red loamy sand. The rainfall is received from the Southwest monsoon and accordingly the flows are significant in the months June to October. The unregulated flow data at the gauging station Akkihebbal (shown in Fig. 3.4) has been used as the second case example for the periodic stochastic modeling of streamflows in this study. The runoff data used are for the period 1916-1974 (58 years) (WRDO, 1976).

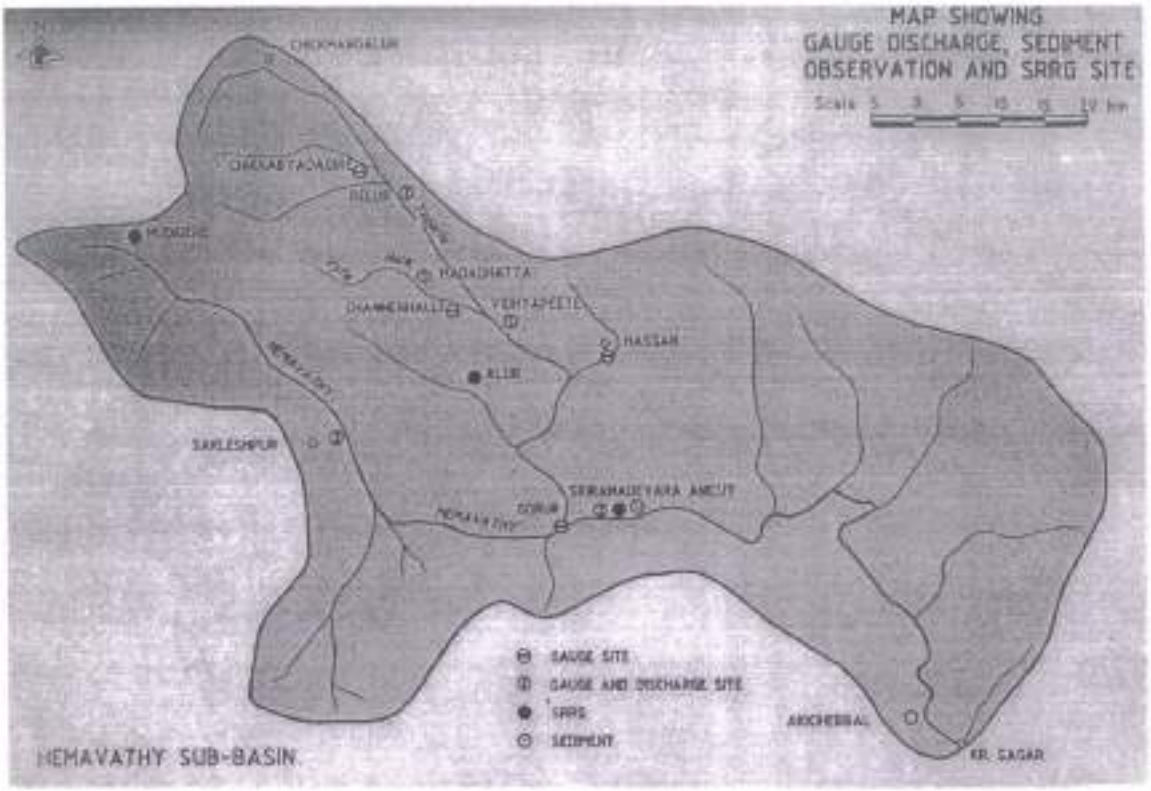


Fig 3.4 Location of River Gauge Stations in the Hemavathy Subbasin
(Source: Report No. TR-53, NIH, 1986)

As mentioned earlier, the following three periodic stochastic models were used to model the monthly streamflows at Bargi dam site and at Akkihebbal: (i) k-nearest neighbour bootstrap (k-NN), a non-parametric model; (ii) hybrid moving block bootstrap, a hybrid of low-order linear parametric model and moving block bootstrap (HMBB); (iii) perturbed matched block bootstrap (PMABB), a non-parametric method proposed in this study. For the k-NN model, the model order (d) and the number of neighbours (k) were arrived at based on the guidelines provided by Lall and Sharma (1996). For the HMBB model, the structure of the linear periodic parametric model (used for partial pre-whitening) and the size of the moving block to be used for bootstrapping the residuals

were decided together due to the hybrid character of the model. For the PMABB model, the size of the within-year matched block size to be used for bootstrapping the flows, the number of adjoining elements (band width) to be considered in the rank matching procedure, and the weighting parameter to be used in smoothing had to be decided together through a detailed trial and error process. In case of both the sites and for all the three methods adopted, the selection of the appropriate combination of the models and the parameters was done through a detailed verification followed by rigorous validation process (based on the preservation of storage and drought related statistics at various demand levels/truncation levels). The verification was based on the reproducibility of the summary statistics, the serial correlation structure for four lags, the one-lag state-dependent correlations of the historical monthly streamflows. It is to be mentioned that the ability of the models to simulate non-linear dependence is portrayed using state-dependent correlations suggested by Sharma et al. (1997).

3.5.2.1 Case Example – 1: Jamtara (Bargi dam site) on river Narmada

For all the three models attempted, the alternative model structure choices considered and the parameters of the selected model structure are presented in Table 3.2 A comparative analysis of the efficacy of the three types of models in modeling the monthly streamflows at Bargi dam site, is presented in the following few paragraphs.

Table 3.2: Parameters of the Selected Models – Bargi dam site.

k-Nearest Neighbour Model:	k (number of neighbours for resampling) = 8; d (model order) =1.
HMBB Model:	PAR(1) model with no transformation Non-overlapping block size L = 24 months (for resampling)
PMABB Model:	Matched Block Size: L = 4 Number of elements taken for resampling: w = 5 Smoothing Parameter: p = 0.9 – 1.1

Reproduction of summary statistics:

The reproduction of the summary statistics of the monthly flows is presented in Table 3.3.The means of the monthly flows are underestimated by k-NN in 6 out of 12 months; HMBB underestimates the same in 4 out of 12 months and over estimates the same in 2 out of 12 months, while the proposed PMABB model is able to reproduce the mean monthly flows well in all the 12 months. The standard deviations of the monthly flows are underestimated by both k-NN and HMBB in 6 out of 12 months; while in general, PMABB exhibits much less bias in reproducing the standard deviation of the monthly flows (Table 3.3). It may be observed from Table 3.3 that the skewness coefficients of the monthly flows are preserved well in general by all the 3 models considered, except in three of the low flow months. Even in that case, the proposed PMABB model seems to perform better than the k-NN and the HMBB models. Thus, the PMABB model reproduces the summary statistics of the historical flows of the Narmada river at Bargi dam site better than the other two models considered, namely, k-NN and HMBB.

Table 3.3: Reproduction of Summary Statistics: Monthly Streamflows – Bargi Dam site

Month	Model	Mean	Std. Deviation	Skewness
June	Historical	144.25	229.73	2.89
	k-NN	119.82	175.26	2.62
	HMBB	154.24	230.05	2.48
	PMABB	145.69	218.83	2.63
July	Historical	1503.1	1095.2	1.08
	k-NN	1481.2	1074.3	1.18
	HMBB	1556.5	1081.4	0.92
	PMABB	1523.2	1078.1	1.01
August	Historical	3072.3	1336.7	-0.06
	k-NN	3002	1325.3	-0.08
	HMBB	3111.2	1303.8	-0.04
	PMABB	3040.3	1386.5	-0.07
September	Historical	1613.5	1173.6	1.14
	k-NN	1600.8	1132.9	1.06
	HMBB	1645	1155.3	1
	PMABB	1593	1167.2	0.99
October	Historical	419.2	324.18	1.15
	k-NN	414.95	314.93	1.12
	HMBB	421.42	323.39	1.1
	PMABB	414.72	318.99	1.15
November	Historical	118.67	73.19	1.04
	k-NN	117.45	70.14	1.08
	HMBB	116.55	67.93	0.83
	PMABB	118.57	73.26	0.97
December	Historical	76.92	70.88	2.86
	k-NN	73.43	60.43	2.57
	HMBB	73.63	59.62	2.42
	PMABB	76.97	68.76	2.59
January	Historical	56.95	43.38	2.07
	k-NN	55.68	38.76	1.81
	HMBB	54.96	38.37	1.78
	PMABB	56.89	42.54	1.89
February	Historical	41.07	32.5	1.34
	k-NN	41.64	32.31	1.31
	HMBB	40.54	31.19	1.3
	PMABB	40.83	31.83	1.33
March	Historical	25.02	21.96	2.5
	k-NN	23.41	18.01	1.61
	HMBB	23.66	18.11	1.57
	PMABB	24.95	21.26	1.93
April	Historical	12.49	17.83	4.79
	k-NN	10.34	9.23	2.02
	HMBB	11.08	11.96	2.56
	PMABB	12.56	16.21	3.14
May	Historical	7.32	20.5	6.05
	k-NN	4.32	4.12	1.55
	HMBB	5.41	10.59	2.88
	PMABB	7.39	17.01	3.69

Preservation of Serial Correlations

The preservation of serial correlations for lags 1-4 are shown in Fig. 3.5a. From Fig. 3.5a it is observed that k-NN model is not able to preserve the serial correlations for 5 out of 12 months for lags 1-4. Similarly it is observed (Fig 3.5a) that HMBB models are not able to preserve the serial correlations in: 5 out of 12 months for both lags 1 and 2; 2 out of 12 months for lag-3 serial correlation and 1 out of 12 months for lag-4 serial correlation. In case of PMABB model, it is observed (Fig. 3.5a) that serial correlations for lag-1 are preserved well for all the months; whereas, lag-2, lag-3 and lag-4 serial correlations display some bias in a few months. The fact that PMABB model could simulate the linear correlations fairly well by resampling short within-year blocks (L = 4), is quite striking. Here, the PMABB model is primarily gaining from the conditional resampling based on matching the end elements of the within-year blocks.

Preservation of First order state-dependent correlations

The preservation of the first order state-dependent correlations namely, above median backward (AMB), below median backward (BMB), above median forward (AMF), and below median forward (BMF) proposed by Sharma et al. (1997), has been presented in Fig. 3.6. From the Fig 3.6 it may be observed that all the three models have performed reasonably well in preserving BMB and BMF correlations. However, it may be seen that k-NN and HMBB models underestimate the above median correlations AMB and AMF in 4 out of 12 months, compared with PMABB.

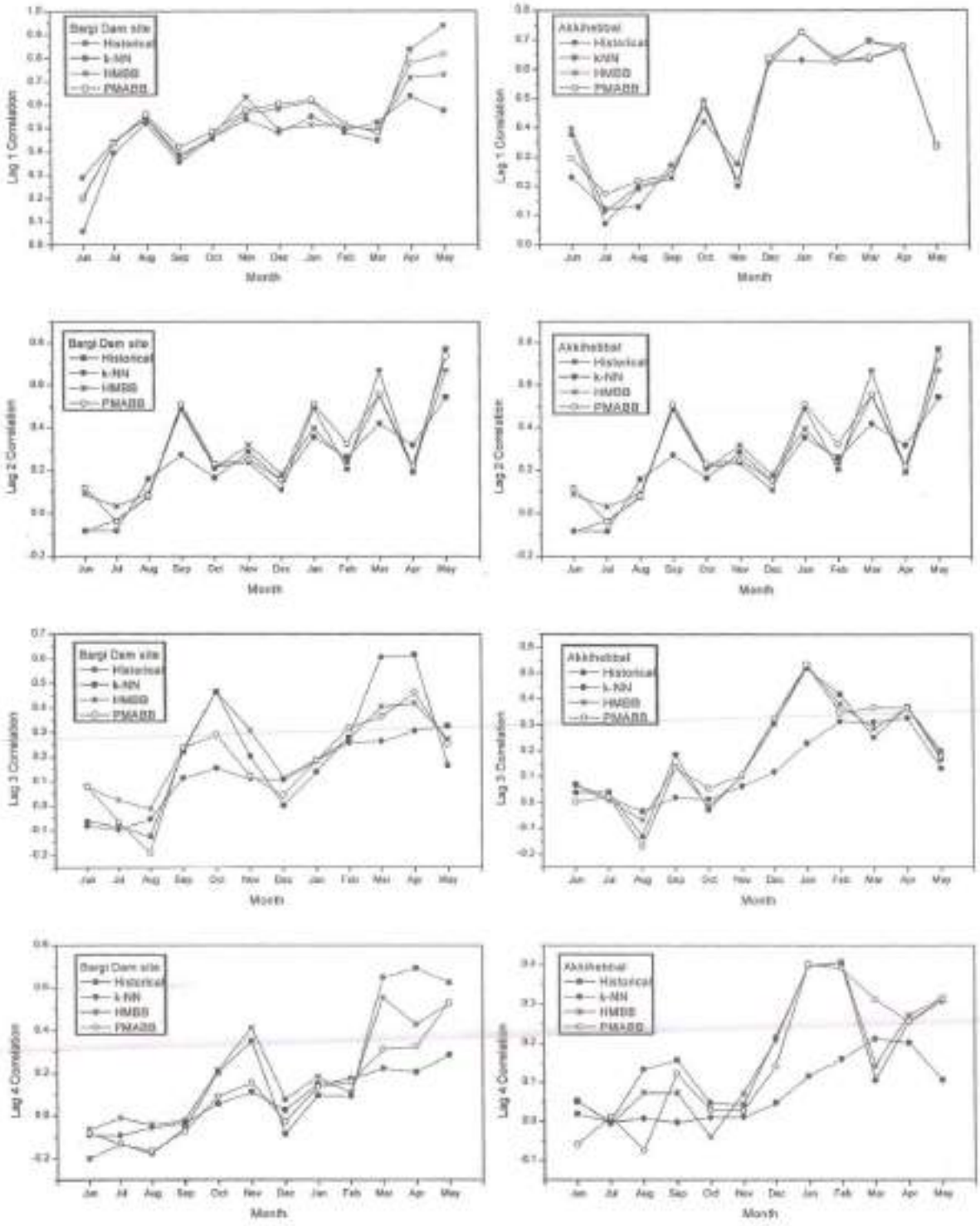


Fig 3.5: Preservation of Serial Correlation of Monthly Streamflows – (a) Bargi Dam site; (b) Akkihebbal

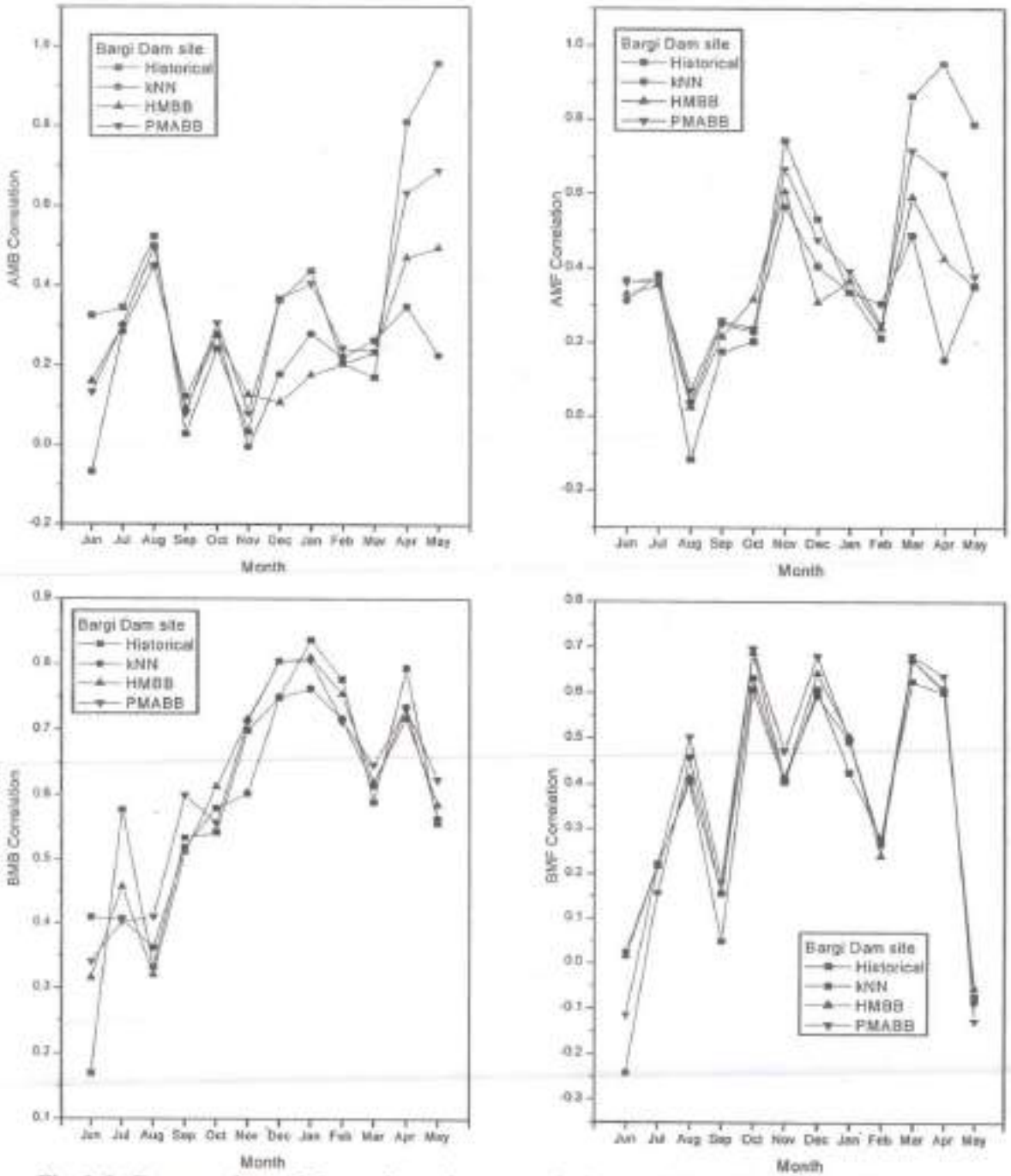


Fig 3.6: Preservation of State-dependent correlations of Monthly Streamflows – Bargi Dam site

3.5.2.2 Case Example – 2: River Hemavathi (station: Akkihebbal)

For all the three models attempted, the alternative model structure choices considered and the parameters of the selected models are presented in Table 3.4 A comparative analysis of the efficacy of the three types of models in modeling the monthly streamflows for river Hemavathi (measured at Akkihebbal), is presented in the following few paragraphs.

Table 3.4 Parameters of the Selected Models – Akkihebbal.

k-Nearest Neighbour Model:	k (number of neighbours for resampling) = 8; d (model order) =1.
HMBB Model:	PAR(1) model with no transformation Non-overlapping block size L = 24 months (for resampling of the residuals)
PMABB Model:	Matched Block Size: L = 4 Number of elements taken for resampling: w = 5 Smoothing Parameter: p = 0.9 – 1.1

Reproduction of summary statistics:

The reproduction of the summary statistics of the monthly flows is presented in Table 3.5. The means of monthly flows are generally well reproduced by all the three models excepting some bias in 1 out of 12 months. The standard deviations of the monthly flows is underestimated by k-NN model in 4 out of 12 months and in 2 out of 12 months by HMBB model; while, PMABB model underestimates the same in 1 out of 12 months. The skewness coefficients of the monthly flows are well preserved, in general, by all the three models considered (Table 3.5).

Table 3.5: Reproduction of Summary Statistics: Monthly Streamflows – Akkihebbal

Month	Model	Mean	Std. Deviation	Skewness
June	Historical	149.56	125.23	1.62
	k-NN	145.79	123.15	1.59
	HMBB	142.9	105.32	0.92
	PMABB	144.85	109.07	1.02
July	Historical	856.11	519.59	2.06
	k-NN	878.2	514.88	1.8
	HMBB	863.92	509.8	1.91
	PMABB	868.63	500.32	1.87
August	Historical	665.29	351.74	1.09
	k-NN	667.07	351.04	1.09
	HMBB	664.27	348.96	1.06
	PMABB	660.98	350.82	1.13
September	Historical	297.82	148.24	1.49
	k-NN	291.95	141.84	1.43
	HMBB	298.26	146.1	1.36
	PMABB	295.61	147.81	1.42
October	Historical	285.34	184.45	1.5
	k-NN	281.42	174.32	1.43
	HMBB	286.88	180.15	1.41
	PMABB	283.95	184.05	1.45
November	Historical	127.16	96.13	1.35
	k-NN	128.92	95.91	1.38
	HMBB	126.09	92.33	1.35
	PMABB	125.08	94.02	1.37
December	Historical	54.67	32.26	2.36
	k-NN	54.77	31.79	1.76
	HMBB	54.67	31.26	1.77
	PMABB	54.45	31.39	1.83
January	Historical	29.94	10.39	0.35
	k-NN	29.51	9.65	0.24
	HMBB	29.99	10.39	0.33
	PMABB	29.81	10.45	0.38
February	Historical	18.25	6.81	-0.13
	k-NN	18.05	6.54	-0.05
	HMBB	18.11	6.61	-0.2
	PMABB	18.19	6.82	-0.04
March	Historical	13.68	6.65	-0.16
	k-NN	13.76	6.62	-0.18
	HMBB	13.69	6.52	-0.17
	PMABB	13.63	6.68	-0.1
April	Historical	14.24	7.49	-0.05
	k-NN	14.22	7.32	-0.01
	HMBB	14.33	7.43	-0.06
	PMABB	14.22	7.48	-0.01
May	Historical	35.57	28.27	1.64
	k-NN	34.95	26.54	1.52
	HMBB	36.29	27.86	1.54
	PMABB	35.16	27.27	1.51

Preservation of Serial Correlations:

From Fig. 3.5b it is observed that k-NN model is not able to preserve: the lag-1 serial correlations in 5 out of 12 months; lag-2 serial correlations in 6 out of 12 months; and lag-3 and lag-4 serial correlations in 7 out of 12 months. While, in case of HMBB model, the lag-1 to lag-4 serial correlations are not preserved in only one out of 12 months. In case of PMABB model, it is observed (Fig. 3.5b) that the lag-1 serial correlations are preserved well, while, the lag-2 and the lag-3 serial correlations are not preserved satisfactorily in 2 out of 12 months and the lag-4 serial correlations are not preserved well in 1 out of 12 months.

Preservation of First order state-dependent correlations:

The preservation of the first order state-dependent correlations have been presented in Fig. 3.7. From the Fig 3.7, it may be observed that all the three models have performed reasonably well in preserving BMB and BMF correlations. However, it may be seen that the k-NN model exhibits more relative bias in the above median correlations AMB and AMF in a few months, when compared with the HMBB and the PMABB models.

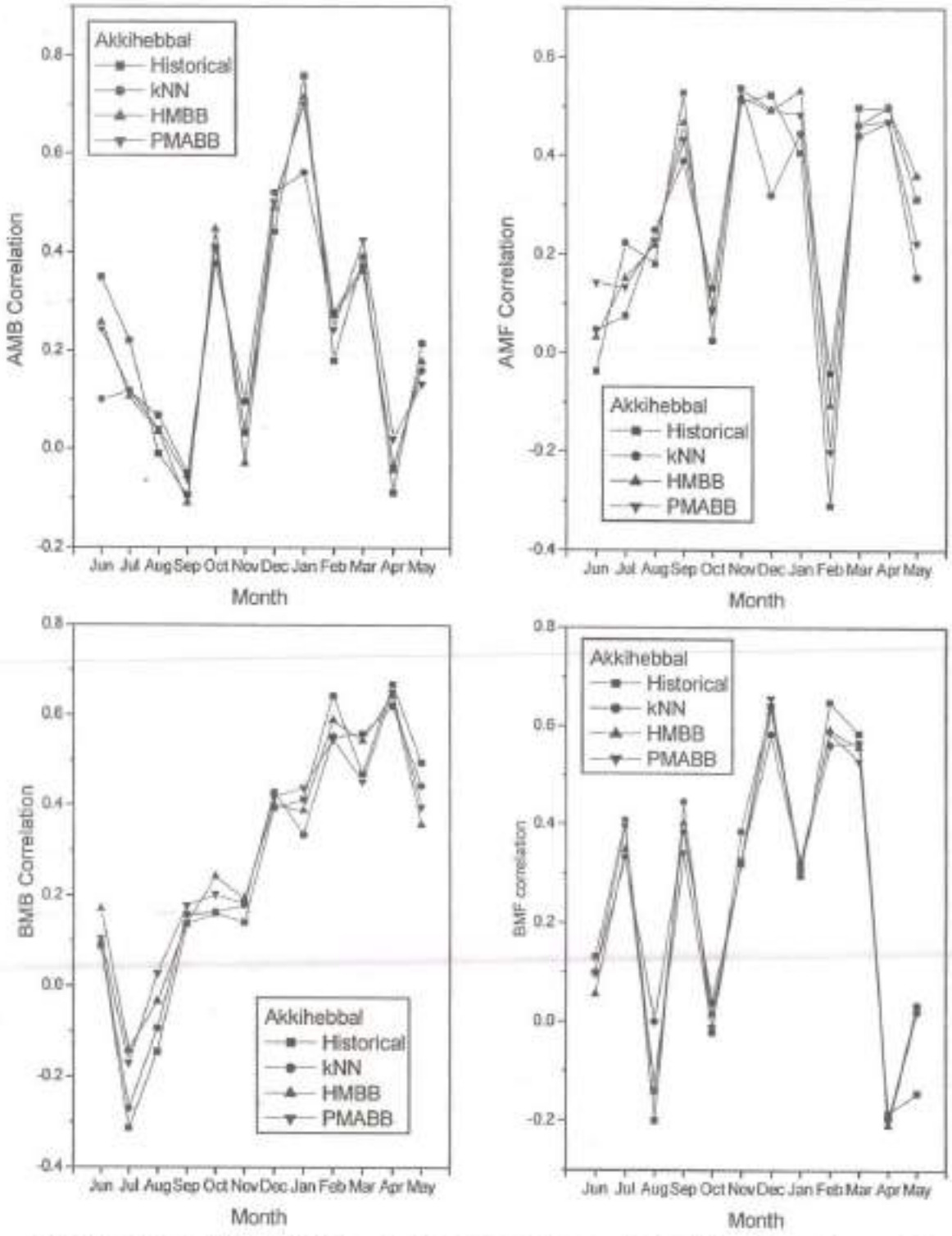


Fig 3.7: Preservation of State-dependent correlations of Monthly Streamflows – Bargi Dam site

3.5.3 Performance Validation of the Stochastic Streamflow Models

The simulations from the models are validated by examining their ability to (i) predict reservoir storage capacity and (ii) preserve critical run characteristics (validation statistics according to *Stedinger and Taylor*, 1982). Herein, the reservoir storage capacities required to cater to yields of 50% Mean Annual Flow (MAF) to 90% MAF (at 5% MAF intervals) are computed using the sequent peak algorithm (*Loucks et al.*; 1981, p.235), assuming the demand to be fixed and uniform over the twelve months of the water year. Drought characteristics are quantified using the theory of runs (*Yevjevich*, 1967), that is based on a threshold level, referred to as truncation. According to this theory, drought is viewed as a negative run that denotes an uninterrupted sequence of streamflow values that lies below the specified truncation level. The run characteristics considered for the evaluation of the relative performance of the three nonparametric models are: (i) Maximum Run Length (MARL), (ii) Maximum Run Sum (MARS), (iii) Mean Run Length (MERL), and (iv) Mean Run Sum (MERS). Herein, the truncation levels have been chosen as percentages of the historical mean monthly flows (MMF) (50% to 100% MMF at intervals of 5% MMF).

Let "dl_i" denote the length and s_i denote the volume of water below a specified truncation level (i.e., deficit volume in Million m^3) for the i th negative run. Then, the aforementioned run characteristics can be expresses as:

$$\text{MARL} = \max [dl_1, \ldots, dl_{NR}]; \tag{3.12}$$

$$\text{MARS} = \max [s_1, \ldots, s_{NR}]; \tag{3.13}$$

$$\text{MERL} = \frac{\sum_{i=1}^{NR} dl_i}{NR} \tag{3.14}$$

$$\text{MERS} = \frac{\sum_{i=1}^{NR} s_i}{NR} \tag{3.15}$$

where NR denotes the total number of runs in the flow sequence (historical/synthetic).

3.5.3.1 Prediction of Reservoir Storage Capacity – Bargi dam site

Table 3.6: shows that both k-NN and PMABB models overestimate the reservoir storage capacity in the range of demand levels 50% to 75% MAF with low variance as well. On the other hand, the HMBB model predicts the reservoir storage capacity better in the range of above demand levels, while it underestimates the same at higher demand levels.

Table 3.6: Comparison of Prediction of Reservoir Storage Capacity –Bargi Dam site

Demand	Hist.	Mean Synthetic Capacity			Relative Bias			Relative RMSE		
(%MAF)	Capacity	k-NN	HMBB	PMABB	k-NN	HMBB	PMABB	k-NN	HMBB	PMABB
50	3562.3	4338.6	3822.4	4475	-0.218	-0.073	-0.256	0.333	0.168	0.333
55	4153.1	5120.8	4549.9	5237.5	-0.233	-0.096	-0.261	0.353	0.194	0.344
60	4744.1	5984.9	5394.6	6065.7	-0.262	-0.137	-0.279	0.386	0.245	0.369
65	5548.1	6948.7	6351.4	6997.4	-0.252	-0.145	-0.261	0.389	0.274	0.366
70	6493.6	8072.7	7384.9	8084.9	-0.243	-0.137	-0.245	0.393	0.29	0.369
75	7439	9429.1	8548.2	9395.7	-0.268	-0.149	-0.263	0.426	0.318	0.407
80	9584	11176	10014	11077	-0.166	-0.045	-0.156	0.372	0.287	0.352
85	13395	13620	11888	13362	-0.017	0.113	0.002	0.338	0.295	0.311
90	18292	17656	14444	16803	0.035	0.21	0.081	0.423	0.359	0.361
95	23846	24642	18692	22954	-0.033	0.216	0.037	0.567	0.431	0.471

3.5.3.2 Prediction of Reservoir Storage Capacity - Akkihebbal

Table 3.7 shows that all three models (k-NN, HMBB and PMABB) perform well in predicting the reservoir storage capacity. However, k-NN underestimates the reservoir

storage capacity at high demand levels, whereas, on the other hand PMABB model over predicts the reservoir storage capacity at lower demand levels.

Table 3.7: Comparison of Prediction of Reservoir Storage Capacity – Akkihebbal

Demand	Hist.	Mean Synthetic Capacity			Relative Bias			Relative RMSE		
(%MAF)	Capacity	k-NN	HMBB	PMABB	k-NN	HMBB	PMABB	k-NN	HMBB	PMABB
50	736.5	716.19	727.07	737.19	0.028	0.013	-0.001	0.069	0.032	0.089
55	826.75	837.99	820.47	856.52	-0.014	0.008	-0.036	0.101	0.034	0.129
60	952.35	1004.3	944.76	1019.8	-0.055	0.008	-0.071	0.145	0.066	0.164
65	1164.7	1213.9	1154.7	1239.4	-0.042	0.009	-0.064	0.152	0.093	0.167
70	1385.8	1468.3	1420.3	1526.5	-0.06	-0.025	-0.102	0.175	0.124	0.197
75	1711.7	1779.8	1742.4	1900.3	-0.04	-0.018	-0.11	0.192	0.148	0.226
80	2049.5	2199.8	2182.6	2390.3	-0.073	-0.065	-0.166	0.238	0.203	0.291
85	2855.5	2811.2	2886.2	3075	0.016	-0.011	-0.077	0.256	0.239	0.268
90	4343.3	3775.8	4044.1	4181.1	0.131	0.069	0.037	0.315	0.305	0.299
95	6455.7	5633.2	6315.3	6297.4	0.127	0.022	0.025	0.445	0.49	0.454

3.5.3.3 Prediction of Drought Characteristics – Bargi dam site:

A discussion on the comparative analysis of the prediction of the critical and the mean drought characteristics is provided in the following few paragraphs, based on the results presented in Table 3.8.

Table 3.8: Comparison of Prediction of critical Drought Characteristics – Bargi Dam site

TL	Hist	Model			TL	Hist	Model		
		k-NN k - 8	HMBB L-24	PMABB W5 – NWB 3			k-NN k - 8	HMBB L-24	PMABB W5 – NWB 3
Maximum Run Length (years)					Maximum Run Sum (million m^3)				
50	12	11.51	11.87	11.67	50	2353.4	2169.8	2197.9	2357.9
		(2.37)	(2.95)	(1.63)			(410.2)	(437.9)	(301.1)
		[28.2]	[42.2]	[27.2]			[27.6]	[62.2]	[54.6]
55	12	11.87	12.5	12.03	55	2670.4	2516.1	2539	2702.9
		(2.37)	(2.85)	(1.71)			(468.9)	(486.5)	(316.6)
		[32.2]	[44.8]	[36.2]			[31.6]	[63.4]	[59.4]
60	12	12.39	12.88	12.28	60	2987.4	2863.8	2877.4	3049.3
		(2.71)	(2.86)	(1.75)			(514)	(537.7)	(330)
		[40.2]	[48.8]	[41.2]			[38.6]	[64.8]	[62.6]
65	12	13.14	13.81	12.7	65	3304.5	3242.8	3239.8	3398.1
		(2.95)	(3.29)	(1.88)			(591.4)	(607.6)	(343)
		[50.6]	[60.4]	[49.8]			[44.6]	[65.2]	[66.8]
70	14	14.95	16.08	13.57	70	3621.5	3710.1	3706.5	3755
		(3.83)	(3.87)	(2.07)			(814.5)	(731.8)	(360.5)
		[45.8]	[57.4]	[29]			[51.2]	[69.2]	[71.2]
75	14	16.22	16.74	14.61	75	3938.5	4168.1	4101.9	4123.3
		(3.97)	(3.8)	(2.19)			(949.4)	(819.9)	(380.1)
		[63.8]	[73.6]	[49.6]			[60.4]	[70.4]	[78]
80	14	17.48	16.98	15.61	80	4255.5	4679.1	4523.2	4519.7
		(4.36)	(3.73)	(2.38)			(1106.9)	(913.2)	(465.8)
		[78]	[75.6]	[67.4]			[67.4]	[74.2]	[81.8]

85	14	18.79	17.41	16.38	85	4572.9	5313.2	4998	4937.9
		4(.63)	(3.65)	(2.52)			(1354.5)	(1020.2)	(571.7)
		[86]	[79.8]	[77.2]			[73.6]	[76.6]	[86.6]
90	18	19.91	20.02	17.23	90	4890.6	5889.5	5673.5	5391.7
		(4.94)	(4.16)	(2.68)			(1572.8)	(1203.1)	(723.5)
		[52.6]	[53]	[29.4]			[77.4]	[77.6]	[89.8]
95	18	20.25	20.65	18.08	95	5208.2	6467.2	6057.8	5919
		(4.99)	(4.22)	(2.99)			(1766.8)	(1345.5)	(967.4)
		[57.4]	[59.2]	[43.2]			[79.6]	[79.8]	[91.4]
100	18	20.93	21.85	18.54	100	5529.7	7096.4	6593.3	6420.8
		(5.13)	(4.37)	(2.93)			(1939.3)	(1492.4)	(1112)
		[63.2]	[71]	[49.6]			[82.8]	[81]	[92.8]
Mean Run Length (years)					Mean Run Sum (million m^3)				
50	3.02	3	3.04	3.06	50	301.8	300.6	285.4	328.3
		(0.39)	(0.39)	(0.41)			(72.7)	(74)	(86.9)
		[45.8]	[49.6]	[50.4]			[48.6]	[38.8]	[57.4]
55	2.98	2.86	3.06	3.15	55	318.9	308.2	304.9	363
		(0.37)	(0.42)	(0.4)			(75.1)	(75.9)	(91.3)
		[35.6]	[55.8]	[63.2]			[43]	[40.4]	[66]
60	3.05	3.02	3.18	3.24	60	352.6	350.1	342.1	396.9
		(0.38)	(0.42)	(0.37)			(79.5)	(79.3)	(92.3)
		[43.8]	[59.2]	[67.6]			[47.6]	[43]	[66.2]
65	3.11	3.11	3.2	3.39	65	386.9	386.9	370.9	446.9
		(0.36)	(0.36)	(0.36)			(84.2)	(80.7)	(96.6)
		[48.8]	[58.2]	[75.8]			[48.2]	[40.4]	[71.8]
70	3.51	3.45	3.66	3.66	70	476.6	474.3	470.9	521.6
		(0.41)	(0.43)	(0.38)			(100)	(101.2)	(108.8)
		[44]	[62.8]	[64]			[47]	[45.6]	[65.2]
75	3.88	3.84	4.02	3.94	75	571.7	569.9	558.6	607.6
		(0.45)	(0.48)	(0.41)			(118.1)	(116.9)	(124.5)
		[43.2]	[57.2]	[52.2]			[47.2]	[42]	[60.8]
80	4.05	4.04	3.99	4.1	80	644.9	649	604.3	681.5
		(0.48)	(0.53)	(0.44)			(130.7)	(123.5)	(138.1)
		[46.6]	[41.6]	[54.2]			[48.6]	[33.2]	[59.6]
85	4.27	4.18	4.02	4.17	85	732.5	724.7	657.2	746.1
		(0.48)	(0.5)	(0.47)			(139)	(130.3)	(145.4)
		[41.4]	[29.4]	[38.8]			[47]	[26]	[52.4]
90	4.17	4.2	4.1	4.27	90	767.9	786.7	726.7	827.7
		(0.48)	(0.6)	(0.5)			(145.3)	(151)	(158.6)
		[50.6]	[40.4]	[55.4]			[53.2]	[36.2]	[62.8]
95	4.35	4.39	4.25	4.44	95	884.3	905.1	834	929.6
		(0.49)	(0.63)	(0.52)			(161.8)	(167.4)	(174.8)
		[48.4]	[37.4]	[52]			[53.4]	[33.8]	[57.8]
100	4.57	4.54	4.58	4.63	100	1005.2	1012.9	968.5	1053.3
		(0.5)	(0.66)	(0.54)			(175.3)	(184.6)	(192)
		[45]	[47.6]	[50.8]			[49.6]	[37.6]	[57.6]

Maximum Run Length (MARL)

At higher truncations levels (90%-100% MAF), it is observed that both k-NN and HMBB overestimate the maximum run length (MARL), whereas PMABB is able to give good prediction. Similarly for intermediate truncation levels (70%-85% MAF), both k-NN and HMBB overestimate the MARL, while PMABB provides reasonably good predictions.

At lower truncation levels (50%-65%), PMABB predicts MARL better than the other models k-NN and HMBB.

Maximum Run Sum (MARS)

At higher truncations levels (90%-100% MAF), it is observed that k-NN highly overestimates the maximum run sum (MARS), whereas both HMBB and PMABB models slightly overestimates the MARS. All the models reasonably predict well at intermediate truncation levels (70%-80%). At lower truncation levels (50%-65%), both k-NN and HMBB underestimates MARS, whereas PMABB model gives very good prediction.

Mean Run Length (MERL)

All the models are able to preserve the mean run length (MERL) at all truncation levels.

Mean Run Sum (MERS)

At all the truncation levels, the mean run sum (MERS) is very well preserved by k-NN model. On the other hand, HMBB slightly underestimates the MERS, while PMABB slightly overestimates the same.

3.5.3.4 Preservation of Drought Characteristics – Akkihebbal:

A discussion on the comparative analysis of the prediction of the critical and the mean drought characteristics is provided in the following few paragraphs, based on the results presented in Table 3.9.

Table 3.9: Comparison of Prediction of critical Drought Characteristics - Akkihebbal

TL	Hist	Model			TL	Hist	Model		
		k-NN k = 8	HMBB L=24	PMABB (W=5,NWB =3)			k-NN k = 8	HMBB L=24	PMABB (W=5,NWB =3)
Maximum Run Length (years)					Maximum Run Sum (million m^3)				
50	8	5.33	6.66	5.9	50	297.3	301.9	302.2	305.9
		(1.3)	(1.85)	(1.44)			(50)	(28.1)	(71.9)
		[2.2]	[2]	[0]			[33.2]	[72.6]	[51.4]
55	8	6.07	6.74	6.37	55	490.1	380.7	443.3	411.6
		(1.42)	(1.81)	(1.18)			(71.8)	(73.7)	(111.4)
		[5.2]	[5.2]	[0]			[7.8]	[30.8]	[28.2]
60	8	7.42	7.53	7.05	60	595.3	460	534.1	513.9
		(1.55)	(0.97)	(1.39)			(88.4)	(97.8)	(130.2)
		[21.6]	[6.4]	[12.4]			[7.2]	[30.8]	[32.8]
65	12	8.4	10.26	8.21	65	700.5	552.1	630.1	607.5
		(1.82)	(2.58)	(1.63)			(103.5)	(123.2)	(149)
		[2.4]	[7]	[0.4]			[8.4]	[31.4]	[32.8]
70	12	9.44	10.33	9.66	70	805.7	651.4	730.7	711.1
		(1.87)	(2.53)	(1.52)			(118.4)	(141.1)	(158.2)
		[5]	[7.4]	[1.8]			[10.4]	[31.4]	[33]
75	12	11.23	11.21	10.78	75	911.0	774.3	840.3	818.8
		(2.33)	(1.79)	(1.31)			(133.7)	(152)	(165.9)
		[21.4]	[11.6]	[7.8]			[13.2]	[31.4]	[33]
80	12	12.8	11.99	12.46	80	1016.2	908.1	953.9	942.3
		(2.71)	(1.82)	(2.13)			(161.2)	(162. 1)	(175.3)
		[43.8]	[17.2]	[39.4]			[19]	[33.8]	[36]
85	14	13.53	13.47	14.57	85	1121.4	1039.9	1092	1087.2
		(2.91)	(1.81)	(3.16)			(182.6)	(177)	(201.5)
		[27.2]	[12.2]	[40.2]			[24.6]	[38.4]	[42]
90	18	15.72	16.92	16.42	90	1226.7	1226.5	1254.6	1262.4
		(3.67)	(3.77)	(3.69)			(232.2)	(206.7)	(232.8)
		[19.2]	[36]	[25.2]			[40.4]	[43.8]	[52.8]
95	23	17.01	20.93	18.22	95	1390.4	1410.3	1472.3	1471.7
		(3.83)	(5.19)	(4.1)			(272.8)	(268.7)	(264.1)
		[5]	[34.6]	[10.8]			[41.4]	[44]	[56.2]
100	23	18.78	21.47	19.58	100	1608.7	1665.2	1689.1	1705.6
		(4.45)	(4.91)	(4.3)			(369.1)	(315.4)	(336.3)
		[12.4]	[35.6]	[16.8]			[43.8]	[40.2]	[57.2]
Mean Run Length (years)					Mean Run Sum (million m^3)				
50	1.59	1.59	1.61	1.63	50	39	37.8	39.5	38.3
		(0.11)	(0.1)	(0.11)			(6.2)	(7.8)	(6.8)
		[49.4]	[56.6]	[65]			[39.8]	[48.6]	[44.6]
55	1.75	1.73	1.75	1.75	55	48.9	46.7	50.1	48.9
		(0.11)	(0.11)	(0.11)			(7.2)	(9.4)	(7.6
		[38.8]	[46.4]	[49]			[34.4]	[51.8]	[47]
60	1.86	1.85	1.88	1.88	60	63.1	60.8	63.4	62
		(0.12)	(0.12)	(0.12)			(8.3)	(10.8)	(8.7)
		[42.2]	[54.4]	[53.2]			[38.2]	[48.2]	[43.8]
65	1.97	1.97	2.01	2.03	65	79	76.6	79.6	78.2
		(0.14)	(0.15)	(0.14)			(9.6)	(12.3)	(10)
		[45.6]	[59.6]	[64.2]			[38.4]	[49.8]	[45.4]
70	2.19	2.19	2.09	2.23	70	99.2	96.7	94.1	97
		(0.15)	(0.15)	(0.17)			(11.2)	(13.3)	(11.8)
		[45.4]	[24]	[56.6]			[42.4]	[32]	[41]
75	2.45	2.47	2.45	2.48	75	120.8	118.4	119.5	119.8
		(0.18)	(0.19)	(0.19)			(13.4)	(15.2)	(13.5)
		[51.6]	[49.6]	[57.6]			[42.4]	[46.8]	[45]
80	2.76	2.78	2.78	2.8	80	149.1	146.5	146.7	149.6
		(0.22)	(0.24)	(0.23)			(16)	(17)	(16.4)
		[52.6]	[48.2]	[54.2]			[41.2]	[44.6]	[50.8]
85	3.02	3	3.04	3.09	85	183.5	178.4	182.1	184
		(0.23)	(0.28)	(0.27)			(18.4)	(19.7)	(19.7)
		[42.4]	[52]	[56.8]			[35.4]	[47.8]	[50.8]

90	3.32	3.29	3.32	3.34	90	227.1	221.1	225	223
		(0.27)	(0.31)	(0.31)			(23.1)	(24.8)	(24)
		[45.2]	[50.4]	[49.2]			[36.4]	[47]	[43.4]
95	3.5	3.53	3.53	3.58	95	266.5	264.3	265	265.2
		(0.3)	(0.37)	(0.34)			(27.8)	(30)	(28.6)
		[51.8]	[51.2]	[56.4]			[43.8]	[46.4]	[47.6]
100	3.77	3.78	3.79	3.79	100	310.8	307.4	308.9	307.8
		(0.33)	(0.37)	(0.35)			(33.3)	(33.3)	(33)
		[50.8]	[50]	[50.2]			[44]	[45.8]	[45.2]

Maximum Run Length (MARL)

At all truncation levels, it is observed that a considerable bias is exhibited by all the three models. The HMBB model (with L=24) is seen to be somewhat better than the other two models in the prediction of MARL.

Maximum Run Sum (MARS)

At higher truncations levels (90%-100% MAF), it is observed that k-NN exhibits less bias compared with the other two models. However, at lower and intermediate truncation levels (55%-85%MAF), the k-NN model under predicts MARS much more than HMBB and PMABB.

Mean Run Length (MERL)

All the models are able to preserve the mean run length (MERL) well at all truncation levels.

Mean Run Sum (MERS)

All the models are able to preserve the mean run sum (MERS) well at all truncation levels.

It is to be mentioned that in case of both the streamflow examples, the HMBB model requires a block size of 24 months to be used for resampling the residuals in order to preserve the serial correlations, and to predict the reservoir storage statistics and the critical drought characteristics, while the PMABB model shows a better preservation/prediction of these statistics even by using a block size of 4. This, in turn, gives rise to a better variety of simulations for the PMABB model compared with the HMBB model (with a block size of 24 months). On the other hand, in order to achieve sufficient variety in simulations, if a within-year block size (L < 12 months) is used for resampling the residuals in case of the HMBB model, significant bias is exhibited in the preservations of the serial correlations, the state-dependent correlations and the prediction of the reservoir storage statistics and the critical drought characteristics, thus resulting in a poor performance of the model. This is not shown here for brevity.

In summary, these results strongly suggest that the PMABB model proposed in this work is a potential nonparametric method, which is effective in simulating the multi-season streamflows.

3.6 SELECTION OF MODEL PARAMETERS

The PMABB method has been found to be effective in providing acceptable simulations for various alternate combinations of block size (L), bandwidth (w) and perturbation range (p). Choice of inappropriate parameters could result in either underestimation of the critical dry spells or overestimation of the wet spells at certain truncations levels. If a smaller value of L is selected, the PMABB would be ineffective in reproducing higher lag within-year serial correlations and cross-year serial correlations. Consequently, the

model cannot capture the historical trend of critical run length. Preservation of these correlations would be of interest to the investigator especially when the dependence structure underlying the hydrologic process is strong. From the various streamflow examples tried out in this research work (although only two are presented in this report), it is found that a reasonable block size of L = 4 or L = 6 would be sufficient to reproduce the within-year and the cross-year dependence so that the critical drought characteristics and the storage statistics are modeled efficiently. It is worth mentioning that a much longer block size (more than 12 months), may result in repeating the observed patterns, thus reducing the chance of simulating innovative patterns. Hence, caution has to be exercised in opting for a long block size.

The bandwidth w has a direct effect on the variety obtained in simulations. A large value for w implies resampling from a larger domain of nearest neighbors, which would enhance the possibility of simulating innovative patterns. However, with a larger w, there is a chance of resampling more distant neighboring blocks and as a result the historical dependence structure may not be well preserved in simulations. Consequently, the model would not be able to capture complex trends and jumps in the critical run length. Thus, selecting the combination of L and w together is important from the point of reducing the bias in simulations, while ensuring sufficient variety as well as variability in the synthetic simulations.

Increase in the range of perturbation, $p[1-\delta, 1+\delta]$, while enhancing smoothing and extrapolation in simulations,may also increase the bias in simulating the historical dependence structure and other important statistics of concern. Consequently, the

validation performance of the model, measured in terms of prediction of storage and run characteristics, drops (not shown herein for brevity). It is suggested that with a sample size N of 50 to 200 years, the choice of δ could be varied from 0 to 0.15. Visual comparison of simulated attributes (summary statistics, dependence structure etc.) with historical sample attributes could be tried for various combinations of the parameters L, w and p to choose the near-optimal set of parameters that would result in the most acceptable stochastic simulation of the observed streamflow data for the desired practical use. Although this research study considers only equally sized within-year blocks, it is possible to adopt unequal (variable) within-year block sizes. Future research can focus towards formulating appropriate optimization models that can help in automating the selection of these parameters.

3.7 SUMMARY AND CONCLUSIONS

A new nonparametric method of conditional bootstrap is presented for simulating multiseason hydrologic time series. It resamples non-overlapping within-year blocks of hydrologic data (formed from the observed time series) using the rank matching rule of *Carlstein et al.* [1998]. This algorithm searches the historical record to find neighbouring blocks whose ends closely match the end element of the current block and subsequently resamples their successor blocks. The resampled blocks are perturbed using a weighted smoothing strategy with a window size of 12 months to achieve smoothing and extrapolation in simulations.

The proposed method, termed perturbed matched-block bootstrap (PMABB), is shown to be efficient in reproducing a wide variety of statistical attributes for both hypothetical and

real data sets. The verification and validation results presented here support PMABB as a plausibly better alternative to the non-parametric method "k-nearest neighbour bootstrap of Lall and Sharma (1996)" and the more recently proposed hybrid periodic model HMBB *of Srinivas and Srinivasan* [2001a,b] in simulating periodic streamflows. It is believed that PMABB can provide a rather flexible and adaptive method for simulating time series at finer time scales (e.g., weekly, daily and hourly), where there is progressively more structure to exploit.

The method provides simulations that are efficient in reproducing summary statistics, dependence structure and the salient features of the marginal distribution, without compromising on smoothing, extrapolation and variety in simulations, so that better prediction of storage capacity and critical run characteristics for water resources design is achieved.

CHAPTER 4

PERFORMANCE OF WATER SUPPLY RESERVOIR SYSTEMS

4.1 INTRODUCTION

The storage-yield relationship has been the conventional tool used by water resources engineers to determine the required storage capacity of a reservoir to deliver a specified target yield. The most commonly used sequential procedure for determining the storage-yield relationship has been the mass curve method proposed by Rippl (1883), which assumes that both inflows and demand are known functions of time. In this method, the minimum storage that is required to provide the target yield with absolutely no water shortages over the historical period would be determined graphically. An automated version of the mass curve method is the sequent peak algorithm (Thomas and Burden, 1963). In this procedure, double cycling takes care of the case when critical sequence of flows occurs at the end of the streamflow record. Fiering (1967) has identified the principal shortcomings of using historical flow record in conjunction with the mass curve or the sequent peak algorithm. The classical "safe yield estimate" is simply a single estimate of the yield that could be sustained by the system during the worst drought on record. Almost certainly, a more severe drought will occur, in future, but the traditional safe yield analysis does not provide any estimate of risk, thus the estimate cannot be really considered "safe yield". Thus, any single value of the safe yield should always be accompanied by a clear account of its statistical significance to avoid potential surprises. Vogel (1985) and Vogel and Stedinger (1988) have shown that by using stochastic streamflow models, the precision of storage capacity estimates can be improved

drastically, compared to single historical flow based estimates. They have shown this to be valid, even if the correct model is not identified.

4.2 OVER-YEAR AND WITHIN-YEAR RESERVOIR SYSTEMS

Based on storage capacity, inflow pattern and demand, the reservoir systems can be classified as over-year (or carry-over) and within-year systems. Within-year systems are sensitive to seasonal variations of both inflow and draft. Studies that model the within-year Storage-Reliability-Yield (S-R-Y) relationships are more realistic. However, these relationships are difficult to generalize due to the large number of parameters associated with periodic stochastic streamflow models. Hence, a case-wise study is required to obtain S-R-Y relationships for within-year reservoirs. The failure duration of a within-year system is generally less than a year whereas for over-year reservoir systems, it is more than a year. Reservoirs, in which filling and emptying phases do not take place on an annual basis, but over a number of years, are known as over-year reservoirs, in which over-year storage effects predominate. For the purpose of planning such over-year reservoirs, stochastically generated annual streamflows are to be considered and not periodic streamflows, since the periodic stochastic models will not be able to preserve the year to year historical correlations, when aggregated. Whenever severe, long-stretched deficits (shortages) in water supply are to be handled in a river system, carry-over storages become important and high storage capacities are provided for the reservoirs in such systems. Most reservoir systems exhibit combination of both, over-year and within-year behaviours.

A non-dimensional parameter known as "standardized net inflow" (m) introduced by Hazen (1914) and used by Hurst (1951, 1965) is very useful in identifying if the over-year effects are predominant for a given reservoir. This parameter is also called as resilience index. The same was further adopted in the analytic investigations of Gomide (1975), Troutman (1978) and Pegram et al. (1980) and in the Monte-Carlo investigations of the S-R-Y relationships of Perrens and Howell (1972), Bayazit (1982), Vogel and Stedinger (1987) and Vogel et al. (1995). The expression for m is given by:

$$m = (1 - D)\frac{\mu}{\sigma} = \frac{1 - D}{C_v} \qquad (4.1)$$

wherein C_v is the coefficient of variation of inflows, D is the target yield (expressed as % of Mean Annual Flow (MAF)), μ is the mean and σ is the standard deviation of annual streamflows. Here, m is referred to as the standardized net inflow, since the mean net inflow, (μ - μ D) is standardized by the scale parameter σ of the inflows (Vogel and Stedinger, 1987). Vogel and Bolognese (1995) termed 'm' as the resilience index since it indicates the potential of the system to refill once emptied. For over-year storage systems, usually, m lies between 0 and 1, indicating low resilience (Vogel and Stedinger, 1987). Subsequently, Montaseri, Adeloye (1999) attempted to incorporate other characteristics in equation (1) for discriminating between within-year and over-year reservoir systems.

A higher value of C_v requires higher capacity to meet a given target yield. For instance, for a river with C_v of annual streamflows in the order of 0.10, even a high target yield of about 90%-95% MAF, will not be critical and hence may not require large capacities to

be built. But for a river with C_v of annual streamflows in the order of 0.5 - 0.7, the same target yield would be highly critical and will require large storage capacities to be built, if long and/or severe deficits and their number of occurrences are to be minimized. However, the upper limit of the storage capacity of such reservoirs would very much depend on the incremental improvement in the desired performance and consequently the marginal increase in the benefit-cost ratio of the project.

4.3 STORAGE-PERFORMANCE-YIELD (S-P-Y) RELATIONSHIPS

The operational performance of a water supply reservoir is usually expressed in terms of performance indicators that describe the failure characteristics, namely the frequency, the duration and the severity of failures. That is, reliability, resilience and vulnerability together characterize "risk" in the reservoir planning and operation context. A clear understanding of how unpleasant the periods of unsatisfactory performance may be, will aid in better planning decisions (Hashimoto et al., 1982). Even though reliability is the most commonly used measure of performance in reservoir planning and operation, it is only indicative of the frequency of the deficit (shortfall) and not the continuity of deficits or the consequence. Resilience is nothing but the ability of a system to recover from failure and get back to normalcy within a specified interval of time. It is to be noted that the question of resilience does not arise as far as failure is not experienced. Further, even when the probability of failure is small, the possible consequences of the failure are to be taken care of. When the system is able to perform to a reasonable level of reliability, it will be wiser to reduce the severity of failure (vulnerability) rather than attempting to marginally increase the reliability. The definition of vulnerability proposed by Hashimoto

et al. (1982) states vulnerability as the average of the maximum deficits that occur in each run of failures within an operating horizon. If the loss function is convex in nature, this definition may mislead (Datta and Burges, 1984). Moy et al. (1986) in their work defined vulnerability as the magnitude of the largest deficit during the period of operation, and the same is used in this study also. Srinivasan et al. (1999) improved the resiliency indicator used by Moy et al. (1986).

All over the globe, while more and more surface water supply sites are being pressed into service, target yields have been continuously increasing, at both the existing and the proposed sites. Quite often, increased demands are being met by efficient management and utilization of existing reservoir systems, rather than by adding on new systems (or facilities). Thus, whether new facilities are envisaged or the existing reservoir system is to be operated more efficiently, it is essential to construct the S-R-Y relationships for any reservoir system (Vogel, 1987), considering the uncertainty in the natural inflows into the reservoir. Often, this is done either by the no-failure capacity approach that uses a sequent peak algorithm coupled with a stochastic streamflow model (Vogel, 1985; Vogel and Stedinger, 1987), or by the behavior analysis method (Pretto et al., 1997). The sequent peak algorithm does not allow failures. This method is used with large number of stochastically generated inflow sequences and the resulting storages are ranked and the reliability is estimated. On the other hand, for finding the reliability of a single purpose, single reservoir, sequent peak algorithm may be a convenient tool, but it is not possible to use the same for finding the recovery potential or the vulnerability of even a single purpose, single reservoir. This is because, by using this algorithm, only non-failure capacity estimates can be obtained, for each streamflow sequence routed through the

reservoir. Further, for use in multireservoir and/or multipurpose systems, this technique may not be suitable. All these flaws are overcome by using a behaviour analysis based on stochastic reservoir simulation model (Pretto et al., 1997), in which the streamflow sequences generated from an appropriate stochastic model are routed through the reservoir using an operating policy (such as a standard operating policy), and the complete information regarding failure characteristics is obtained. In this method, the generated sequences will be long enough to ensure steady state performance. Hence, for the descriptive assessment of storage performance on the storage-yield plane of the reservoir, it would be more appropriate to use the behaviour analysis method, so that more comprehensive information regarding the dynamic performance of a reservoir system can be obtained.

The Storage-Performance-Yield (S-P-Y) relationships are useful in: (i) gaining an understanding of the variation of reservoir performance indicators namely, reliability, resilience, and vulnerability on the storage-yield plane; (ii) identifying the sensitive ranges of storage capacity of the over-year reservoirs, with regard to performance characteristics; and (iii) selecting between capacity expansion and demand management options, in case of deficit water supply systems.

McMahon and Mein (1986), Klemes (1987), Vogel and Stedinger (1987), Phatarford (1989), Vogel and Bolognese (1995) and Vogel and McMahon (1996) provide reviews of literature relating to the development of general S-R-Y relationships. In the last few decades, quite a number of researchers have used stochastic streamflow models in

conjunction with sequent peak algorithm (either single-cycling or double-cycling) to obtain the S-R-Y relationships. Fiering (1963, 1965, 1967) generated 200 synthetic sequences of gamma and normally distributed annual streamflows from a first order autoregressive (AR(1)) model and analyzed the same by the double-cycling sequent peak algorithm. Burges and Linsley (1971) generated the complete probability density function of over-year storage capacity, using single-cycling sequent peak algorithm, assuming the annual streamflows to be normally distributed and to follow a AR(1) model. They suggested that 1000 streamflow sequences would be required to specify the probability distribution function of storage capacity.

Perrens and Howell (1972) developed generalized S-R-Y relationships in graphical form when annual streamflows are assumed to be normally distributed and to follow a AR(1) model. They used an algorithm, which allows failures and computes the reliability based on the number of times failure did not occur. Gomide (1975) derived the probability distribution function of storage capacity and its mean and presented the results graphically for full regulation, and planning period ranging from 0 to 100 years, using the single-cycling sequent peak algorithm, for the case when annual streamflows follow a AR(1) model. He also presented the probability distribution function of storage capacity and its mean and standard deviation for different partial regulations of the reservoir for planning periods, which range from 0 to 50 years. Troutman (1978) derived the mean and the variance of the asymptotic distribution of storage capacity for the case of full regulation, when inflows are described by an AR(1) lognormal model.

Bayazit (1982) provided diagrams to determine the mean and standard deviation of the deficit for the cases of full and partial regulation. Vogel (1985) developed approximate but general S-R-Y relationships using both single and double-cycling sequent peak algorithm for normally distributed annual inflows and subsequently, Vogel and Stedinger (1987) developed the relationships for annual inflows characterized by a two-parameter lognormal distribution and first-order Markov process. They used Monte-Carlo simulation and double-cycling sequent peak algorithm. Klemes (1969) employed a s-state (s number of discrete states) Markov chain model in an effort to describe the complex structure of sequences of reservoir surpluses and failures. Vogel (1987) found that a two-state Markov model gave a satisfactory representation of the complex structure of sequences of within-year surpluses and failures and later Vogel and Bolognese (1995) showed that a two-state Markov model can accurately represent over-year reservoir systems also.

Buchberger and Maidment (1989) defined the index P, analogous to the Peclet number used to measure the relative importance of convection and diffusion processes, as

$$P = \frac{\mu C}{2\sigma^2} \tag{4.2}$$

for the purpose of determining when a storage reservoir of finite capacity behaves as one with a semi-infinite capacity. They show that finite reservoirs with $P < -1$ or $P > 1$ behave as if they have no top or bottom respectively. In this equation, μ is mean annual inflow, C is the capacity of the reservoir and σ^2 is the variance of annual inflow. Based on Markov diffusion process, they presented an analytic method for approximating the equlibirium probability distribution of storage in a finite reservoir. Their analysis was

subject to inflows and outflows which during a unit time interval produce potential storage displacements that are independent and identically distributed with a Gaussian distribution. Vogel and Bolognese (1995) show that

$$p = (1 - q)\left[1 - \frac{rq}{1 - q}\right]^{N-1} \qquad (4.3)$$

where p = probability for failure-free operation over an N-year planning period; q = steady-state probability of a failure; and r = system resilience estimated. Equation (3), which is based on two-state Markov model of reservoir system states, provides a very good approximation to the relationship between p and q as long as the resilience index (equation 1), m, is greater than 0.2. Vogel et al. (1995) combined analytic storage model with the regional model of annual streamflows, resulting in general relations among storage, reliability, resilience and yield (S-R-Rs-Y) in the Northeastern United States. Vogel and McMahon (1996) derived approximate S-R-Rs-Y relationships for over-year water supply systems fed by autoregressive lag-one Gamma and normal inflows. They have shown that the resilience of an over-year water supply system is generally independent of its steady-state reliability.

Pretto et. al. (1997) have shown that a sequence length of 1000 years or more is required in order to get a stationary value of storage estimate for a given reliability and demand. They have also mentioned that the S-R-Y relationships are only benchmarks for planning that allow comparison of alternative plans and systems. However, where planners are interested in operation over short planning horizons, the length of planning horizon of

interest will dictate the length of inflow sequence which should be used in conjunction with the appropriate initial reservoir volume. Srinivasan and Philipose (1998) have investigated the effect of single phase hedging on the performance of over-year reservoirs using behaviour analysis and have constructed trade-off relationships among the performance indicators, reliability, resilience, vulnerability and average deficit. Philipose and Srinivasan (1997) have constructed Storage-Performance-Yield (S-P-Y) relationships in the form of isolines for a within-year reservoir system in southern India.

Weeraratne et al. (1986) employed reliability, resilience and vulnerability measures to evaluate reservoir release policies for low flow augmentation. Moy et al. (1986) investigated the trade-off between reliability, resilience and vulnerability for a water supply reservoir using multiobjective programming model. Burn et al. (1991) formulated a multiobjective compromise-programming model for real-time reservoir operation representing reliability, resilience and vulnerability as performance criteria. Simonovic (1992) formulated a simulation-optimization model with reliability and vulnerability constraints for finding the minimum required capacity. The formulation of Moy et al. (1986) was improved for more complete representation of resilience by Srinivasan et al. (1999). Introducing constraints, which describe reservoir performance explicitly into the optimization, would have to deal with a large number of integer variables (depending on the number of time periods considered). For such problems, the formulation is of mixed integer type, and this would require enormous computer time. Building resilience into the model requires tracking of the number of crossovers from failure to success, and this makes the formulation much more complex (Srinivasan et al., 1999). Though many

attempts have been done to generalize the S-R-Y relationships of over-year systems, except the works of Vogel and Bolognese (1995), Vogel and McMahon (1996), there has not been significant attempts to generalize Storage-Performance-Yield (S-P-Y) relationships of over-year systems, which consider also resilience and the vulnerability.

4.4 PROPOSED STUDY

In this chapter, it is proposed to investigate the S-P-Y relationships of both over-year and within-year reservoir systems. While the first part of the study deals with the construction of general S-P-Y relationships and a S-P-Y database for use in over-year water supply reservoir planning and design applications, the second part involves the construction of S-P-Y relationships for a specific within-year reservoir, namely, Dharoi reservoir on Sabramathi river system in India.

The over-year system generalization considers the commonly used performance indicators, namely, reliability, resilience and vulnerability. Modularized annual streamflows generated from AR(1) model are used for the evaluation of the performance indicators. In most cases, for modeling annual streamflows, the assumption of AR(1) model with lognormal distribution would be sufficient. Either models with higher order dependence or more complex 3-parameter lognormal or gamma distributions may not be necessary to describe annual streamflows. The approach followed in this study is "behaviour analysis based on stochastic simulation", which allows failures and explicitly tracks the failure characteristics during the period of the long-run reservoir simulation. That is, the system reliability is expressed as a steady state probability. The S-P-Y

database constructed would serve as useful screening level reference information for reservoir planners and decision-makers dealing with over-year water supply reservoirs.

The within-year reservoir performance study is carried out for the Dharoi reservoir on the river Sabramathi in India, considering the performance indicators, namely, volume reliability, occurrence reliability, resilience, period vulnerability, event vulnerability, mean period deficit and mean event deficit. These seven indicators are defined in a later section of this chapter. The reservoir storage performance indicators are evaluated using a long synthetic sequence generated, which is similar to the historical flows recorded at the site of interest. The three alternative periodic stochastic models (described in Chapter 3), namely, k-NN, HMBB and PMABB have been considered for the purpose of generating the long sequence of monthly streamflows. A series of S-P-Y plots are proposed to be developed for the Dharoi reservoir using all the three stochastic models and a comparison would be presented. Also, a brief discussion of the usefulness of these plots would be provided.

4.5 S-P-Y RELATIONSHIPS FOR OVER-YEAR RESERVOIR SYSTEMS

4.5.1 MONTE-CARLO SIMULATION EXPERIMENTS

In this study, detailed information regarding reservoir performance on the storage-yield plane is obtained based on exhaustive number of Monte-Carlo simulation experiments, assuming the modular annual streamflows to be lognormally distributed (LN-2) and having a AR(1) dependence structure. For the Monte-Carlo simulation experiments, 13 cases of coefficient of variation (C_v = 0.1, 0.15, 0.2, 0.25, ..., 0.7) and 11 cases of lag-one

auto-correlation coefficient (ρ_1 = 0.0, 0.05, 0.1, 0.15, 0.2,...,0.5) of flows are considered. This selection of inflow parameters is based on the real world ranges of these parameters of rivers, at a global level. A study of annual streamflow data from 140 gauging stations around the world (Yevjevich, 1964), with records of at least 37 years indicates that ρ_1 values for most of the rivers are found to be less than 0.40. Similarly, from a study of 106 basins in New England, Vicens et al. (1975) found the mean and the standard deviation of estimates of ρ_1 to be 0.22 and 0.14 respectively. Hence, a reasonable range of ρ_1 for these Monte-Carlo experiments has been taken as 0.0 to 0.50, which includes most cases of practical interest. Values of m (standardized net inflow) range from 0.1 to 1.0, which include most over-year storage problems of interest, which correspond to demand levels in the range of 99-30% for values of C_v from 0.10 to 0.70. The storage-yield plane is characterized by the combinations of storage capacities in the range 0.3-5.0 MAF (at an interval of 0.1 MAF) and a number of cases of target yield (covering the over-year range of m = 0.1, 0.2,...,1.0). Thus, in all, the total number of combinations of independent variables amounts to 80784 (Table 1). The stepwise procedure followed for obtaining the S-P-Y information is given in the following section.

Table 4.1. Combinations of Parameters Used for Monte-Carlo Simulation

Coefficient of Variation (C_v) (1)	Correlation Coefficient (ρ_1) (2)	Yield (D) (MAF) (3)	Storage Capacity (K) (MAF) (4)	Number of Combinations (5)
0.10	0.0 - 0.50 (0.05)	0.90 - 1.00 (0.010)	0.30 - 5.00 (0.01)	5808
0.15	0.0 - 0.50 (0.05)	0.85 - 1.00 (0.015)	0.30 - 5.00 (0.01)	5808
0.20	0.0 - 0.50 (0.05)	0.80 - 1.00 (0.020)	0.30 - 5.00 (0.01)	5808
0.25	0.0 - 0.50 (0.05)	0.75 - 1.00 (0.025)	0.30 - 5.00 (0.01)	5808
0.30	0.0 - 0.50 (0.05)	0.70 - 1.00 (0.030)	0.30 - 5.00 (0.01)	5808
0.35	0.0 - 0.50 (0.05)	0.65 - 1.00 (0.035)	0.30 - 5.00 (0.01)	5808
0.40	0.0 - 0.50 (0.05)	0.60 - 1.00 (0.040)	0.30 - 5.00 (0.01)	5808
0.45	0.0 - 0.50 (0.05)	0.55 - 1.00 (0.045)	0.30 - 5.00 (0.01)	5808
0.50	0.0 - 0.50 (0.05)	0.50 - 1.00 (0.050)	0.30 - 5.00 (0.01)	5808
0.55	0.0 - 0.50 (0.05)	0.45 - 1.00 (0.050)	0.30 - 5.00 (0.01)	6336
0.60	0.0 - 0.50 (0.05)	0.40 - 1.00 (0.050)	0.30 - 5.00 (0.01)	6864
0.65	0.0 - 0.50 (0.05)	0.35 - 1.00 (0.050)	0.30 - 5.00 (0.01)	7392
0.70	0.0 - 0.50 (0.05)	0.30 - 1.00 (0.050)	0.30 - 5.00 (0.01)	7920
Total number of combinations				**80784**

Note: Values within parentheses indicate the interval in the respective range of the parameter.

4.5.1.1 Stepwise Procedure for Obtaining the S-P-Y Information

Following are the sequential steps carried out in this study to obtain the storage performance-yield (S-P-Y) information for over-year reservoir systems:

1. For a particular combination of coefficient of variation (C_v) and lag-one autocorrelation coefficient (ρ_1) of annual streamflows, lognormally distributed

modular annual streamflows with AR(1) dependence structure of length = (N*1000) are generated.

2. The modular annual streamflows generated are divided into 1000 sequences, each N year in length, in sequential order. It is to be noted that the length N is taken such that the average performance indicators computed over 1000 sequences are not affected by the initial storage condition at the beginning of the simulation, and the average performance indicators reach a steady state.
3. The generated modular annual streamflows are routed through the reservoir using standard operating policy for different combinations of active storage capacity (K) and target yield (D) considered. The combinations of K and D describe the reservoir storage-yield plane of interest.
4. The mean reliability, the mean resilience and the mean vulnerability along with their respective standard deviations (over 1000 sequences), are computed for all combinations of active storage capacity and target yield considered. Now the S-P-Y information is available for the particular combination of C_v and ρ_1 considered.
5. The steps 1 to 4 are repeated to obtain the performance information for all the combinations of C_v and ρ_1 considered.

This entire S-P-Y information obtained is stored in a database, which will be useful in construction of performance relationships and performance analysis on the storage yield-plane. For the long-run Monte-Carlo simulation referred, a simulation length (N) of 1000 years was found to be sufficient for the mean performance indicators, namely, reliability and resilience to reach a steady state, while a simulation length of approximately 10000

years was required to obtain the mean vulnerability over 1000 sequences. Hence, N=10000 years is adopted for the long-run stochastic reservoir simulation in this paper.

4.5.1.2 Generation of Modular Annual Streamflows

The annual streamflows are often lognormally distributed and have a single order of dependence. This means that a AR(1) model would be sufficient to model the annual streamflows. If the annual streamflows measured are given by Q_i, then the modular annual streamflows will be given as $Y_i = Q_i / \overline{Q}$, where $\overline{Q}$ is the mean annual flow (MAF). This modular form of the annual streamflows is used in the Monte-Carlo simulation for performance study. Once C_v and ρ_1 of Y_i are assumed, the corresponding modular flows can be generated. The log-transformed modular annual streamflows (X_i) following a AR(1) model may be written as:

$$X_{i+1} = \mu_x + \rho_1(x)(X_i - \mu_x) + \sigma_x\left(\sqrt{1-\rho_1^2(x)}\right)\varepsilon_i \qquad (4.4)$$

in which ε_i are independent normal disturbances with mean zero and variance unity, and μ_X, σ^2_X and $\rho_1(x)$ are the mean, the variance and the lag-one correlation of the log-transformed modular streamflows (X_i). The relationships between the statistics of the historical data ($\mu_y, \sigma_y, \rho_1(y)$), and the statistics of the transformed sequence (μ_X,σ_X, $\rho_1(x)$), as given by Matalas (1967) have been used in this study to estimate μ_X, σ_X, $\rho_1(x)$. Once these transformed parameters are obtained, they can be substituted into equation

(4.4) to obtain the generated values (X_j). Then, these generated values (X_j) are exponentiated to get the generated modular streamflows (Y_j).

4.5.2 RESULTS AND DISCUSSION

4.5.2.1 Storage-Performance-Yield (S-P-Y) Relationships

It is to be noted that the Storage-Performance-Yield (S-P-Y) relationships are obtained from the results of the Monte-Carlo simulation, assuming the Standard Operating Policy (SOP). In this section, two of the typical sets of S-P-Y relationships are presented in Figures 4.1 - 4.6. The Figures 4.1 - 4.3 correspond to the S-P-Y relationships for ρ_1 = 0.30 and C_v = 0.50; and the Figures 4.4 – 4.6 correspond to the same for ρ_1 = 0.30 and C_v = 0.70. For the purpose of constructing the storage-yield plane shown in Figures 4.1 - 4.6, a wide range of storage capacities (0.3-5.0 MAF) and target yields are considered.

4.5.2.2 Storage-Reliability-Yield (S-R-Y) Relationships

S-R-Y relationships are the most commonly discussed performance relationships in reservoir problems. These relationships provide the primary information regarding the reservoir performance. It is seen from Figures 4.1 and 4.4 that reliability drops significantly with increase in yield for a given storage capacity, C_v and ρ_1 of the inflows. This drop is found to be more for flows with higher C_v (Figures 4.1 and 4.4), which is only expected. The change in the slope of the S-R-Y relationship provides valuable information for reservoir planners and designers, in terms of identifying the sensitive ranges of storage capacity. It may be seen from Figures 4.1 and 4.4 that as the target yield increases, the storage capacity at which the S-R-Y curve becomes horizontal, also

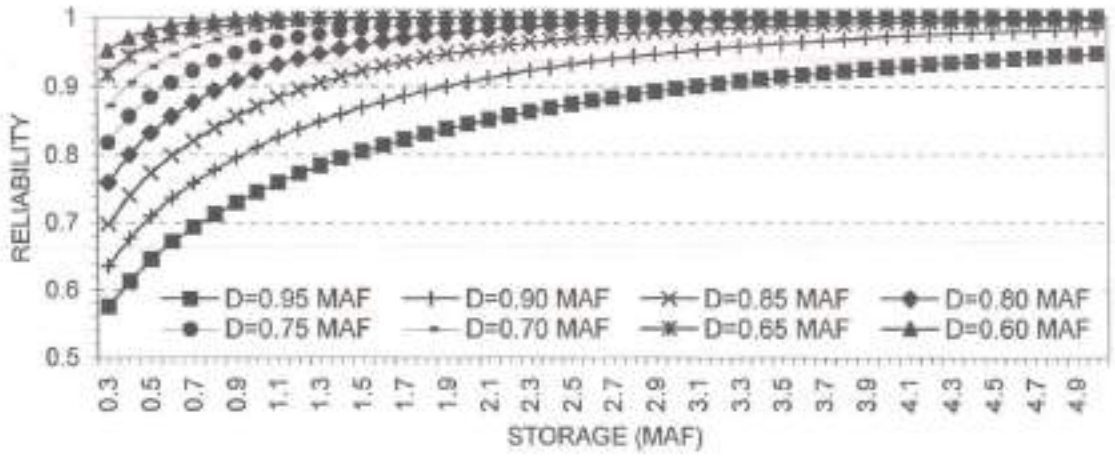


Fig. 4.1 STORAGE-RELIABILITY-YIELD RELATIONSHIPS
(rho = 0.3; Cv = 0.5)

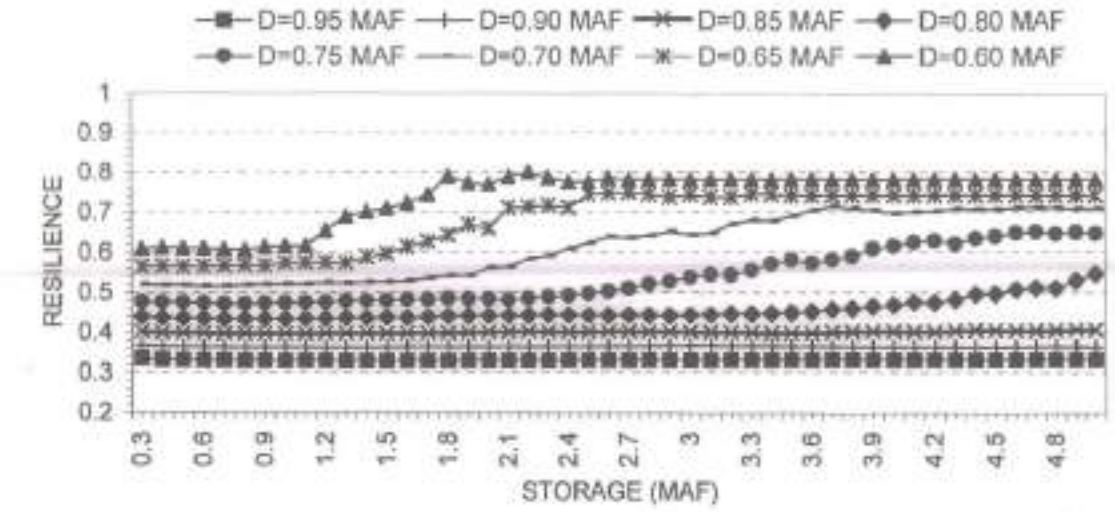


Fig. 4.2 STORAGE-RESILIENCE-YIELD RELATIONSHIPS
(rho = 0.3; Cv = 0.5)

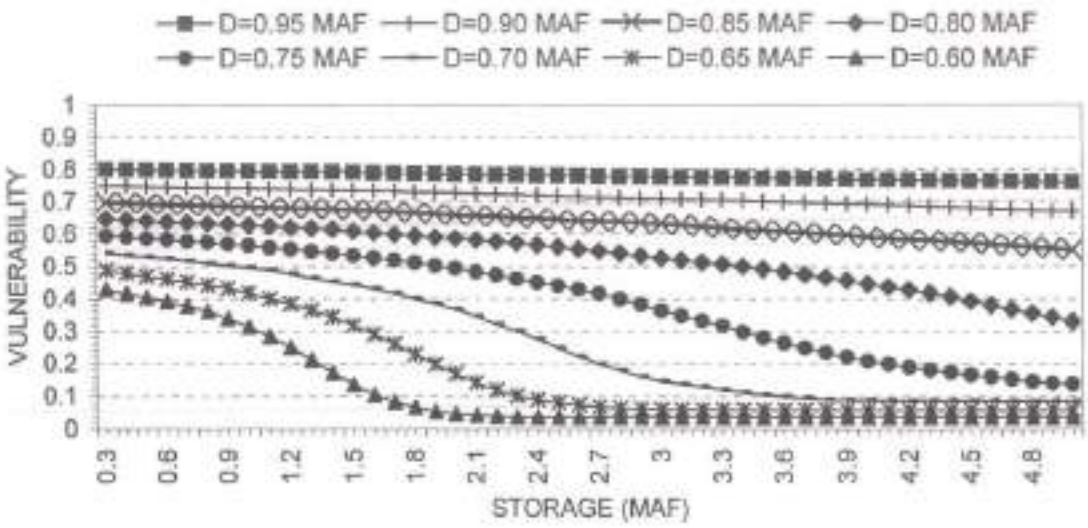


Fig. 4.3 STORAGE-VULNERABILITY-YIELD RELATIONSHIPS
(rho = 0.3; Cv = 0.5)

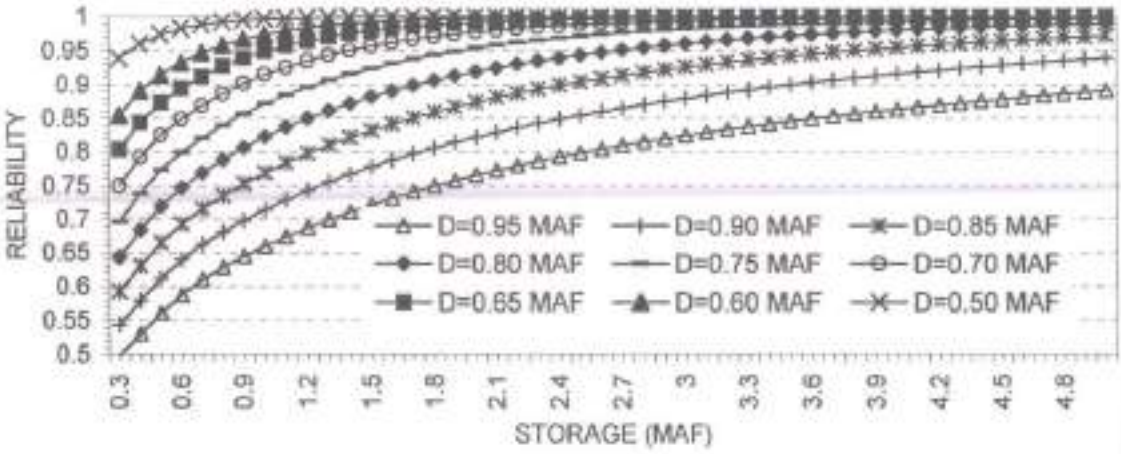


Fig. 4.4 STORAGE-RELIABILITY-YIELD RELATIONSHIPS
(rho = 0.3; Cv = 0.7)

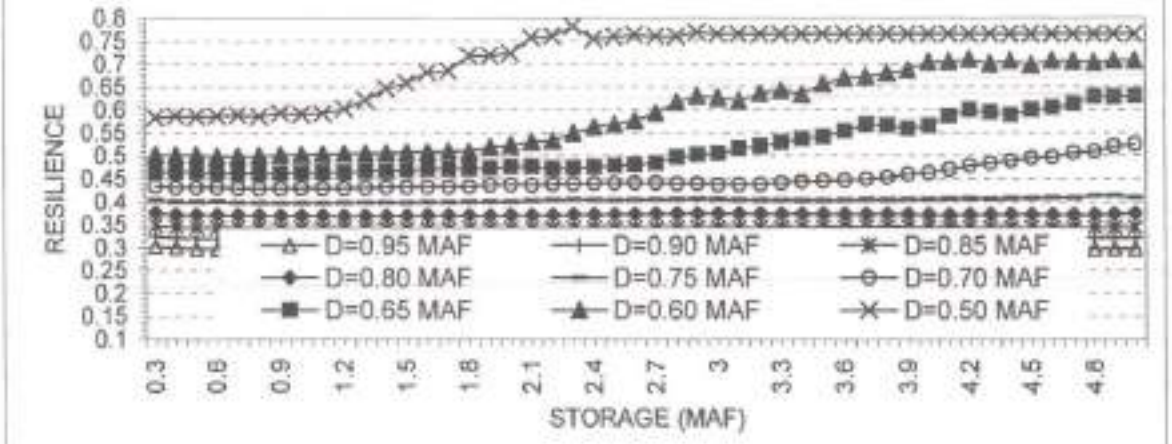


Fig. 4.5 STORAGE-RESILIENCE-YIELD RELATIONSHIPS
(rho = 0.3; Cv = 0.7)

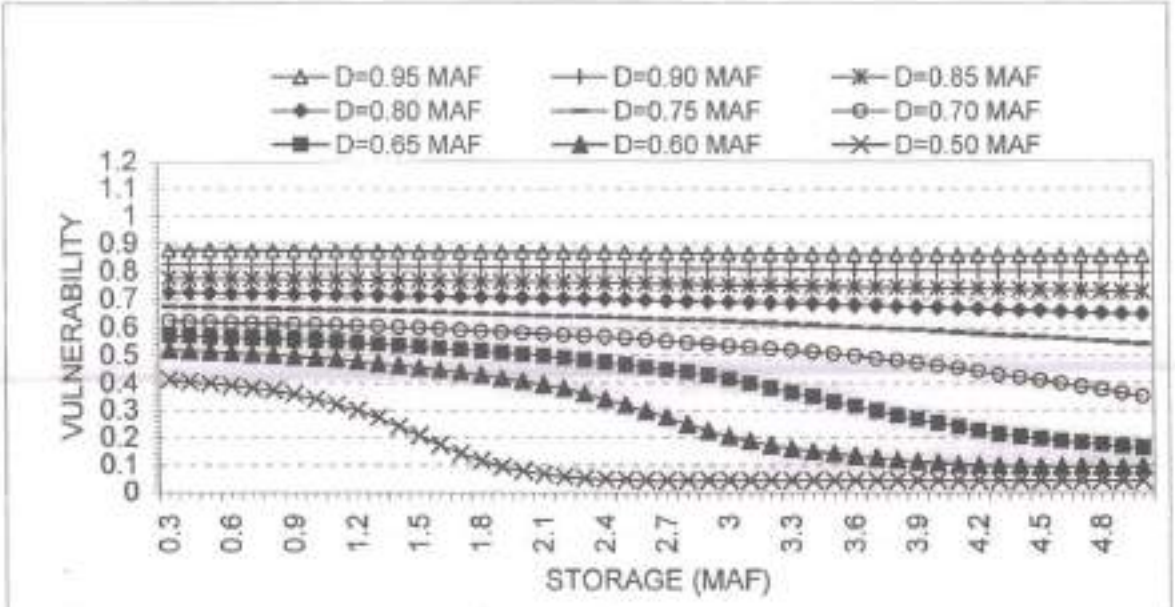


Fig. 4.6 STORAGE-VULNERABILITY-YIELD RELATIONSHIPS
(rho = 0.3; Cv = 0.7)

increases. However, for higher target yields (>=0.90 MAF), the S-R-Y curve does not become horizontal even at a storage capacity of 5.0 MAF.

For water supply reservoirs, the desirable range of reliability is 0.95-0.99 (Vogel and McMahon, 1996). The ranges of minimum capacity required to achieve this range of reliability are extracted from the S-P-Y database and are presented in Tables 4.2 - 4.7 for various yields and six different combinations of ρ_1 and C_v of the inflows. The ranges of the other two performance indicators corresponding to the ranges of the storage capacities referred above are also given in Tables 4.2 – 4.7. It may be noted from Tables 4.2 – 4.7 that in case of low target yields, if the reliability obtained corresponding to the minimum storage considered (0.3 MAF) itself would be greater than 0.95, then the actual value of reliability is entered. Likewise, in the case of high target yields, if the reliability obtained for the upper limit of the storage capacity considered (5.0 MAF) is less than 0.99, then the actual value of reliability is reported. The performance information presented in Tables 4.2 – 4.7 will be quite useful at the planning level for water resources decision making.

Table 4.2. Capacity requirement and performances in good reliability range for $\rho_1 = 0.3$ and $C_v = 0.3$

Yield D (MAF)	Reliability	Min Capacity K	Resilience	Vulnerability (MAF)	Ave. Deficit (MAF)
1.00	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.97	0.9490	3.0	0.3617	0.6050	0.1890
	0.9792	5.0	0.3581	0.5717	0.1897
0.94	0.9524	1.9	0.3937	0.5734	0.1764
	0.9894	3.7	0.3956	0.5099	0.1764
0.91	0.9522	1.3	0.4229	0.5470	0.1627
	0.9901	2.5	0.4246	0.4811	0.1626
0.88	0.9489	0.9	0.4561	0.5251	0.1498
	0.9905	1.8	0.4632	0.4591	0.1503
0.85	0.9539	0.7	0.4915	0.4945	0.1385
	0.9897	1.3	0.4994	0.4382	0.1380
0.82	0.9506	0.5	0.5269	0.4695	0.1268
	0.9902	1.0	0.5339	0.4144	0.1279
0.79	0.9561	0.4	0.5667	0.4398	0.1158
	0.9912	0.8	0.5673	0.3886	0.1164

0.76	0.9563	0.3	0.6021	0.4138	0.1044
	0.9840	0.5	0.6019	0.3837	0.1057
0.73	0.9735	0.3	0.6418	0.3736	0.0953
	0.9921	0.5	0.6358	0.3355	0.0954
0.70	0.9852	0.3	0.6783	0.3319	0.0868

Table 4.3. Capacity requirement and performances in good reliability range for $\rho_1 = 0.3$ and $C_v = 0.5$

Yield D (MAF)	Reliability	Min Capacity K	Resilience	Vulnerability (MAF)	Ave. Deficit (MAF)
1.00	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.95	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.90	0.9496	3.0	0.3664	0.7087	0.2508
	0.9834	5.0	0.3632	0.6700	0.2505
0.85	0.9518	2.0	0.4004	0.6591	0.2261
	0.9895	3.7	0.4004	0.6041	0.2257
0.80	0.9476	1.3	0.4354	0.6170	0.2019
	0.9895	2.5	0.4419	0.5597	0.2016
0.75	0.9480	0.9	0.4738	0.5698	0.1795
	0.9906	1.8	0.4852	0.5122	0.1787
0.70	0.9449	0.6	0.5161	0.5246	0.1575
	0.9911	1.3	0.5236	0.4644	0.1590
0.65	0.9438	0.4	0.5660	0.4794	0.1377
	0.9902	0.9	0.5669	0.4321	0.1382
0.60	0.9522	0.3	0.6096	0.4285	0.1163
	0.9882	0.6	0.6101	0.3923	0.1179
0.55	0.9763	0.3	0.6605	0.3662	0.1000
	0.9873	0.4	0.6609	0.3493	0.1012
0.50	0.99	<0.3	-	-	-

Table 4.4. Capacity requirement and performances in a good reliability range for $\rho_1 = 0.3$ and $C_v = 0.7$

Yield D (MAF)	Reliability	Min Capacity K	Resilience	Vulnerability (MAF)	Ave. Deficit (MAF)
1.00	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.95	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.90	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.85	0.9509	3.8	0.3454	0.7443	0.2950
	0.9724	5.0	0.3432	0.7300	0.2946
0.80	0.9515	2.7	0.3728	0.6942	0.2679
	0.9898	5.0	0.3738	0.6500	0.2667
0.75	0.9493	1.9	0.3994	0.6480	0.2418
	0.9901	3.6	0.4025	0.6015	0.2406
0.70	0.9508	1.4	0.4310	0.6003	0.2166
	0.9900	2.6	0.4402	0.5566	0.2159
0.65	0.9496	1.0	0.4627	0.5528	0.1923
	0.9899	1.9	0.4734	0.5107	0.1918
0.60	0.9476	0.7	0.4989	0.5060	0.1687
	0.9900	1.4	0.5062	0.4642	0.1694
0.55	0.9439	0.5	0.5412	0.4576	0.1459
	0.9894	1.0	0.5463	0.4255	0.1478
0.50	0.9607	0.4	0.5870	0.4051	0.1263
	0.9885	0.7	0.5882	0.3814	0.1258
0.45	0.9677	0.3	0.6301	0.3553	0.1057
	0.9890	0.5	0.6257	0.3341	0.1042
0.40	0.9860	0.3	0.6795	0.2934	0.0884
0.35	0.99000	<0.3	-	-	-

Table 4.5. Capacity requirement and performances in good reliability range for $\rho_1 = 0.5$ and $C_v = 0.3$

Yield D (MAF)	Reliability	Min Capacity K	Resilience	Vulnerability (MAF)	Ave. Deficit (MAF)
1.00	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.97	0.9497	4.6	0.2758	0.6115	0.2039
	0.9555	5.0	0.2744	0.6068	0.2042
0.94	0.9496	2.8	0.2986	0.5841	0.1965
	0.9842	5.0	0.2949	0.5373	0.1906
0.91	0.9498	1.9	0.3247	0.5587	0.1782
	0.9899	3.9	0.3257	0.4936	0.1775
0.88	0.9526	1.4	0.3493	0.5316	0.1656
	0.9892	2.7	0.3471	0.4758	0.1650
0.85	0.9515	1.0	0.3760	0.5109	0.1537
	0.9895	2.0	0.3783	0.4522	0.1530
0.82	0.9488	0.7	0.4041	0.4857	0.1423
	0.9896	1.5	0.4084	0.4311	0.1424
0.79	0.9481	0.5	0.4342	0.4608	0.1308
	0.9911	1.2	0.4357	0.3993	0.1325
0.76	0.9552	0.4	0.4676	0.4313	0.1197
	0.9874	0.8	0.4652	0.3957	0.1223
0.73	0.9586	0.3	0.5052	0.4038	0.1092
	0.9913	0.7	0.4995	0.3608	0.1132
0.70	0.9743	0.3	0.5385	0.3683	0.1003
	0.9900	0.5	0.5328	0.3435	0.1024

Table 4.6. Capacity requirement and performances in good reliability range for $\rho_1 = 0.5$ and $C_v = 0.5$

Yield D (MAF)	Reliability	Min Capacity K	Resilience	Vulnerability (MAF)	Ave. Deficit (MAF)
1.00	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.95	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.90	0.9493	4.6	0.2737	0.7147	0.2714
	0.9561	5.0	0.2741	0.7092	0.2711
0.85	0.9503	3.0	0.3024	0.6679	0.2462
	0.9834	5.0	0.3001	0.6311	0.2442
0.80	0.9494	2.0	0.3299	0.6228	0.2223
	0.9899	4.0	0.3302	0.5671	0.2216
0.75	0.9519	1.4	0.3598	0.5786	0.1998
	0.9903	2.8	0.3575	0.5240	0.1987
0.70	0.9481	0.9	0.3918	0.5343	0.1776
	0.9894	1.9	0.4003	0.4859	0.1771
0.65	0.9491	0.6	0.4291	0.4887	0.1561
	0.9909	1.4	0.4317	0.4389	0.1572
0.60	0.9522	0.4	0.4728	0.4415	0.1352
	0.9888	0.9	0.4691	0.4072	0.1377
0.55	0.9633	0.3	0.5215	0.3916	0.1163
	0.9887	0.6	0.5160	0.3677	0.1189
0.50	0.9824	0.3	0.5658	0.3340	0.0999
	0.9895	0.4	0.5612	0.3225	0.1006

Table 4.7. Capacity requirement and performances in good reliability range for $\rho_1 = 0.5$ and $C_v = 0.7$

Yield D (MAF)	Reliability	Min Capacity K	Resilience	Vulnerability (MAF)	Ave. Deficit (MAF)
1.00	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.95	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.90	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.85	0.9500	>5.0	-	-	-
	0.9900	>5.0	-	-	-
0.80	0.9497	4.1	0.2764	0.7010	0.2915
	0.9657	5.0	0.2747	0.6914	0.2911
0.75	0.9486	2.9	0.2994	0.6527	0.2649
	0.9849	5.0	0.2978	0.6228	0.2642
0.70	0.9532	2.2	0.3227	0.6040	0.2391
	0.9902	4.2	0.3219	0.5642	0.2376
0.65	0.9492	1.5	0.3462	0.5594	0.2140
	0.9896	3.0	0.3448	0.5209	0.2127
0.60	0.9526	1.1	0.3747	0.5121	0.1898
	0.9900	2.2	0.3798	0.4741	0.1878
0.55	0.9566	0.8	0.4060	0.4630	0.1665
	0.9903	1.6	0.4089	0.4291	0.1662
0.50	0.9527	0.5	0.4409	0.4175	0.1428
	0.9897	1.1	0.4391	0.3874	0.1454
0.45	0.9512	0.3	0.4867	0.3707	0.1213
	0.9912	0.8	0.4793	0.3422	0.1243
0.40	0.9750	0.3	0.5287	0.3158	0.1023
	0.9901	0.5	0.5228	0.3010	0.1038
0.35	0.9897	0.3	0.5741	0.2567	0.0838
0.30	0.9500	<0.3	-	-	-
	0.9900	<0.3	-	-	-

4.5.2.3 Storage-Resilience-Yield (S-Rs-Y) Relationships

It may be observed from the S-Rs-Y relationships (Fig 4.2 and 4.5) that resilience drops significantly with increase in target yield, and this drop with respect to yield is found to be less for flows with higher C_v. Further, the increase in resilience with increasing storage capacity, decreases with increase in target yield and eventually becomes insignificant at high target yields, for both the C_v's considered (for $C_v = 0.5$, $\rho_1 = 0.3$, D > 0.85 MAF; for $C_v = 0.7$, $\rho_1 = 0.3$, D > 0.80 MAF; See Fig 4.2 and 4.5). However, for lower target yields, there is no significant improvement in resilience at low as well as high storage capacities while a significant improvement is noted for the range in between. With increase in target

yield, this transition range widens and moves towards higher storage capacities and eventually ending up in flat storage-resilience relationships for high target yields. This deserves some attention in the planning of water supply systems, wherein the marginal benefits due to increase in resilience are considerable.

4.5.2.4 Storage-Vulnerability-Yield (S-V-Y) Relationships

It is seen from Figures 4.3 and 4.6 that in general, vulnerability increases with C_v. Furthermore, it is observed that at high target yields, vulnerability is very high and this decreases at a mild rate with increase in storage capacity. On the other hand, at lower target yields, appreciable reduction in vulnerability is noted in a certain range of storage capacity and this range widens and moves towards higher storage capacities with increase in target yield and/or C_v. This can be exploited in certain water supply systems, wherein the marginal value of damage reduction due to decrease in vulnerability is quite high.

4.5.2.5 S-P-Y Database and the Query-based Program

The results of the Monte-Carlo simulations are stored in a database created using MS Access. This database contains information regarding the inflow characteristics (ρ_1, C_v), reservoir storage capacity and target yield (expressed in terms of the historical mean annual flow (MAF)) and the corresponding performances including their standard deviations over the 1000 replicates. There are 80,784 records in this database, occupying about 28 MB of space in computer hard disk. To extract information from this database, a front-end query-based program was developed using Visual Basic. Either single values or ranges of values can be pre-specified for the input parameters, autocorrelation, coefficient

of variation of river flow, storage capacity of the reservoir and demand level in the input form (Fig 4.7). The corresponding mean performance indicators (reliability, resilience, vulnerability and average deficit) and their standard deviations over 1000 replicates can be extracted from the S-P-Y database and displayed on the output form (Fig 4.8). This program also allows the user to specify the desirable performance range, so that the output form will not display the results falling beyond this specified range of performance indicators. This helps in avoiding unwanted results and enables in directing the search properly. Furthermore, there is option for sorting the outputs based on the mean performance indicators and their standard deviations. Sorting can be performed at two levels. For example, the first level of sorting may be based on reliability and the second level of sorting may be based on any one of the other indicators (see Fig 4.7). The number of records displayed on the output form will guide deciding further refinement. The outputs can be saved in text files (ASCII format) for further processing or for future reference. It is to be noted that in this study, the performance investigation has been done using the simple Standard Operating Policy (SOP) (Eq. 4.5).

$$R_t = \begin{cases} S_t + Q_t - C & \text{if } (S_t + Q_t - D_t) > C \\ D_t & \text{if } C \geq (S_t + Q_t - D_t) \geq 0 \\ S_t + Q_t & \text{otherwise} \end{cases} \qquad (4.5)$$

where S_t is the initial storage, Q_t is the inflow, D_t is the target demand, R_t is the actual release and C is the active storage capacity of the reservoir. However, such S-P-Y databases can be constructed for other (hedging/optimal) policies also.

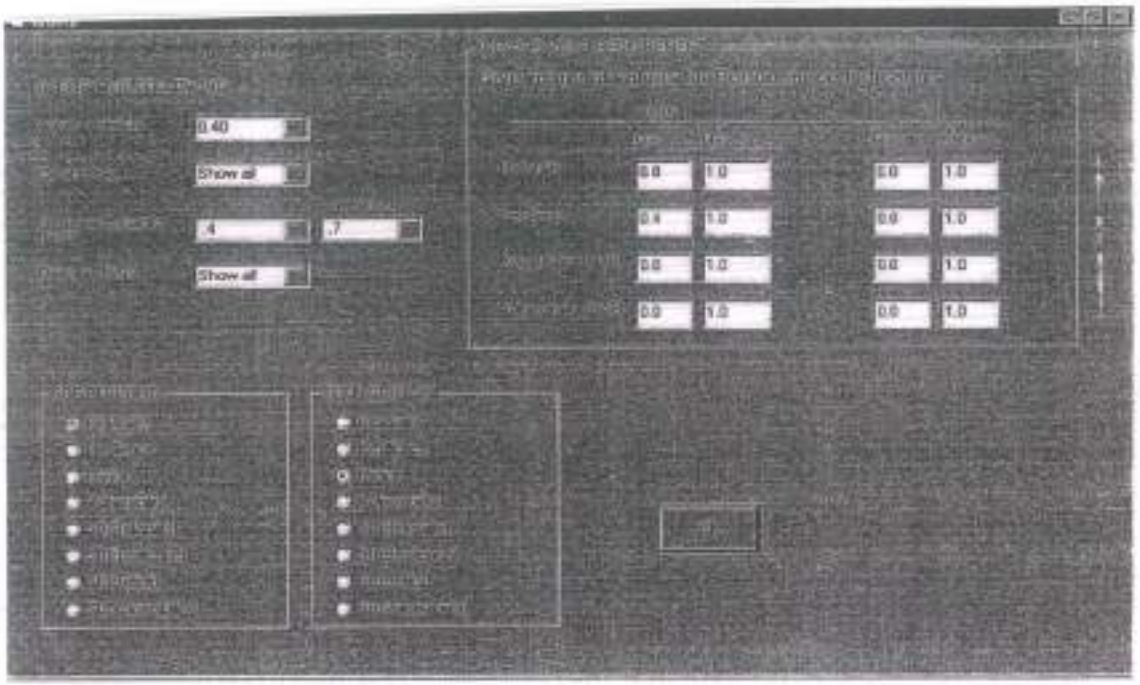


Fig. 4.7 Typical Input Form for S-P-Y Data Base

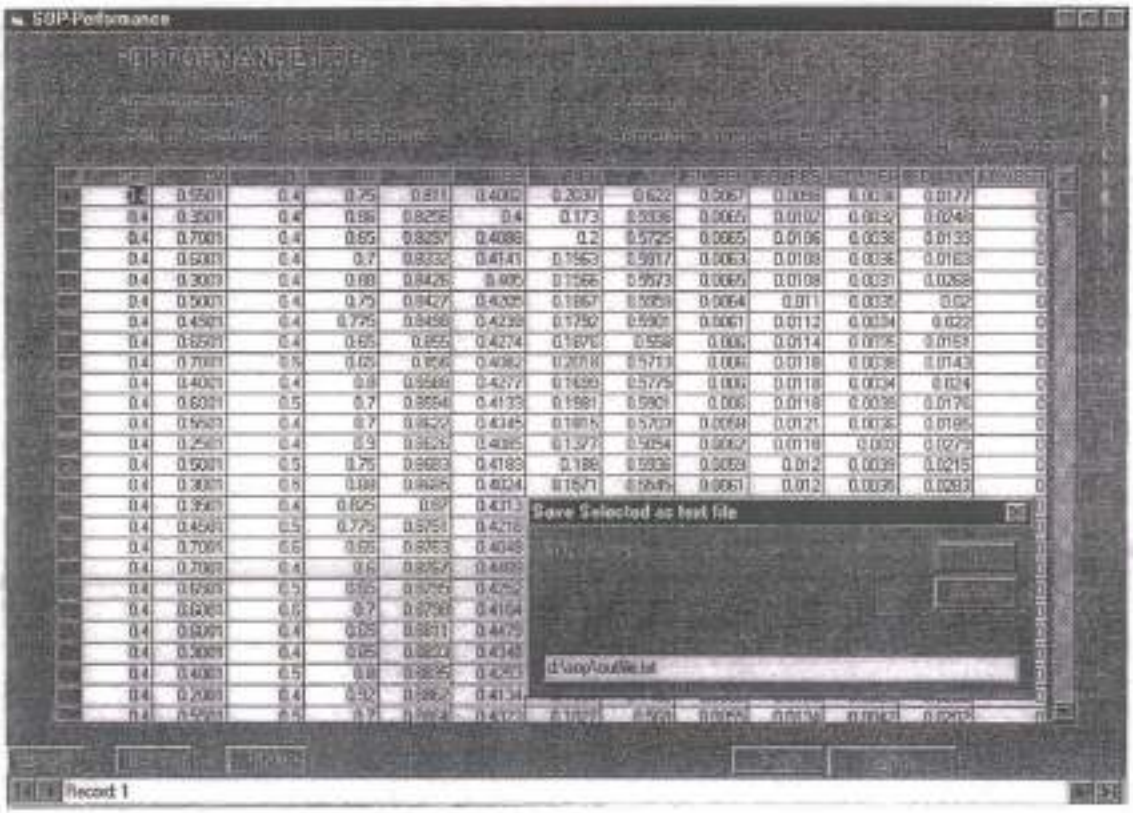


Fig. 4.8 Typical Output Form for S-P-Y Data Base

4.5.2.6 Usefulness of the Performance Database in Decision-making

The over-year reservoir performance database constructed in this study will be useful in: (i) Planning and design of reservoir storage capacity for a pre-specified target yield and desirable performance indicators and (ii) Making decisions regarding capacity expansion or implementation of demand management programs or a combination of the two for an existing reservoir system under operation. However, it is to be mentioned that this database can yield only approximate planning level information concerning the functional performance of an over-year water supply reservoir with regard to meeting a target yield. This is to be used in conjunction with other kinds of inputs such as economical, environmental, social, political and legal in order to arrive at a final decision. The two points mentioned concerning the usefulness of this database are illustrated below.

Consider that the storage capacity of an over-year water supply reservoir is to be fixed for a target yield of 60% MAF and the streamflows into the reservoir have a C_v of 0.7 and ρ_1 of 0.3. A reasonable degree of performance in terms of long-run values of reliability, resilience and vulnerability is expected. Now, the decision-maker can invoke the S-P-Y database constructed in this study and can obtain the kind of storage-performance information presented below:

For K = 1.0 MAF; Reliability = 0.974; Resilience = 0.503; Vulnerability = 0.492 MAF

For K = 1.5 MAF; Reliability = 0.992; Resilience = 0.508; Vulnerability = 0.456 MAF

For K = 2.0 MAF; Reliability = 0.998; Resilience = 0.523; Vulnerability = 0.406 MAF

For K = 2.5 MAF; Reliability = 0.999; Resilience = 0.568; Vulnerability = 0.319 MAF

For K = 3.0 MAF; Reliability = 0.9997; Resilience = 0.626; Vulnerability = 0.204 MAF

A scan of the above information that would be displayed on the output form, suggests that it may be sensible to go in for a storage capacity (K) in the range of 1.5-2.0 MAF. Further investigation of the performance indicator resilience, within the selected range of K shows that K = 1.9 MAF would be an ideal choice of the reservoir capacity.

For K = 1.9 MAF; Reliability = 0.997; Resilience = 0.519; Vulnerability = 0.417 MAF.

Consider that some years after the construction of an over-year water supply reservoir with a storage capacity of 1.0 MAF on a river with inflow characteristics of $C_v = 0.5$ and $\rho_1 = 0.5$, the demand becomes 0.8 MAF, outgrowing the initial target yield of 0.70 MAF. As a result, the long-run reliability is expected to fall from 0.956 to 0.878 (7.8% decrease); the long-run resilience is expected to fall from 0.393 to 0.329 (6.4% decrease); the long-run vulnerability is expected to increase from 0.532 MAF to 0.644 MAF (0.112 MAF increase). In this situation, if the originally intended performance is to be retrieved, then, it may be necessary to decide among capacity expansion or demand management or both. Let the maximum capacity to which the reservoir can be built at the site be restricted to 1.5 MAF from economic, environmental and other considerations. If the existing capacity of 1.0 MAF is increased to 1.5 MAF, then, the long-run mean performance indicators will be 0.922, 0.329 and 0.636 MAF respectively. On the other hand, if the demand management option is exercised (0.05 MAF reduction from the expected demand of 0.80 MAF), then the performance indicators are expected to be 0.922, 0.359 and 0.589 MAF respectively. If both the capacity expansion (from 1.0 MAF to 1.5 MAF) and demand management (0.8 MAF to 0.75 MAF) measures are

implemented, then the performance indicators will be 0.957, 0.360 and 0.574 respectively. Thus, it can be seen that by combining the two measures, reliability is retrieved to the original value, while the other two indicators suffer marginally.

4.6 STORAGE-PERFORMANCE-YIELD RELATIONSHIPS FOR WITHIN-YEAR RESERVOIR SYSTEMS

The development of the storage-performance-yield relationships for within-year reservoir systems cannot be easily generalized since there will be too many parameters involved. Hence, the relationships are to be developed for a particular reservoir under consideration. In this study, to illustrate the construction and the usefulness of such relationships, the case example chosen is the Dharoi reservoir on the river Sabramathi in India.

4.6.1 Performance Indicators for Within-year Reservoir Systems

The definitions of the seven reservoir storage performance indicators for use in within-year reservoir systems are as stated below (equations 4.6 – 4.13):

Occurrence based reliability

$$R_{occ} = 1 - \left(\frac{\sum_{j=1}^{M} d_j}{T} \right) \qquad (4.6)$$

d_j = duration of the j^{th} failure event
M = total number of failure events
T = total number of periods of operations

Volume based reliability

$$R_{vol} = \frac{\sum_{t=1}^{T} R_t}{\sum_{t=1}^{T} D_t} \qquad (4.7)$$

D_t = demand at period t

R_t = release at period t
T = total number of periods of operations

Resilience

$$\text{Res} = \frac{M}{\sum_{j=1}^{M} d_j} \tag{4.8}$$

d_j = duration of the j^{th} failure event
M = total number of failure events

Period vulnerability

$$V_P = \max_t (D_t - R_t) \quad t = 1,2,...,T \tag{4.9}$$

D_t = demand at period t
R_t = release at period t

Event vulnerability

$$V_E = \max\left[E_1, E_2, ... E_j ... E_M\right] \tag{4.10}$$

$$E_j = \sum_{i=1}^{d_j} (D_i - R_i) \quad j=1,2,...,M \tag{4.11}$$

D_i = demand at period i
R_i = release at period i
E_j = total deficit in j^{th} failure event
d_j = duration of the j^{th} failure event

Mean period deficit

$$MD_{period} = \frac{\sum_{t=1}^{T} D_t - R_t}{N} \quad t = 1,2,.....,t \tag{4.12}$$

D_t = demand at period t
R_t = release at period t
N = total number of failure periods
T = total number of periods of operations

Mean event deficit

$$MD_{event} = \frac{\sum_{t=1}^{T} D_t - R_t}{M} \tag{4.13}$$

D_t = demand at period t
R_t = release at period t
M = total number of failure events
T = total number of periods of operations

4.6.2 Case Example: The Dharoi Reservoir on the Sabarmathi River

The Sabarmathi rises in the Aravalli hills and has a length of 300km^2. The drainage area of the river is 21674km^2 of which 19% lies in Rajasthan and the balance in Gujarat (Fig. 4.9). Its main tributaries are the Sei from the right and The Wakul, the Harnav, the Hathmati and the Watrak on the left. At Dharoi the river passes through a gorge and later after 240km of its course, it passes through Ahmedabad and finally falls into the Gulf of Cambay. The important tributaries are the Hathmati (1523km^2), the Sei (946km^2), the Wakul (1625km^2), and the Harnav 972km^2) (Source: NIH, 1987).

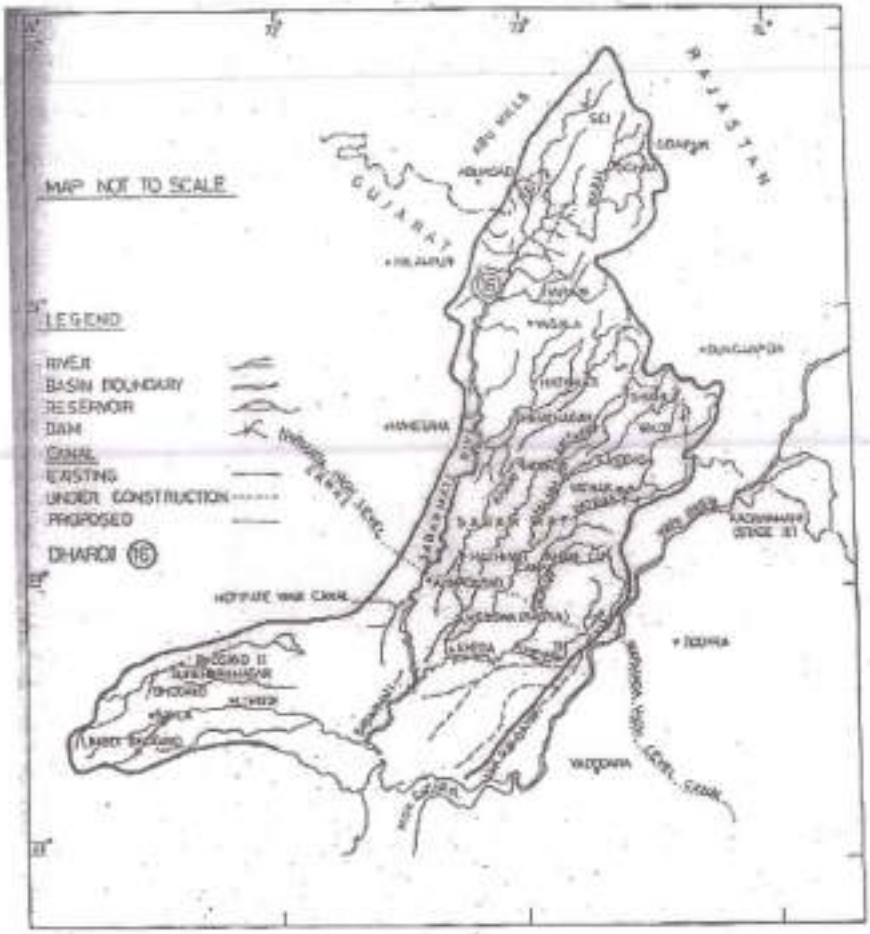


Fig. 4.9: River Sabarmathi – Dharoi Dam Site (Source: NIH. 1989).

The historical monthly flow data at Dharoi dam site are presented in Table 4.8 for 41 water years (June 1935- May 1975).

Table 4.8: Historical Monthly Streamflow Data – River: Sabramathi; Station: Dharoi dam site (Source: NIH, 1987).

Year	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
1935	24.17	194.79	36.29	204.52	31.64	9.65	4.64	3.34	1.81	0.58	0.29	0.22
1936	35.93	11.25	28.01	24.46	9.00	2.32	0.36	0.00	0.00	0.00	0.00	0.00
1937	7.48	445.33	72.36	121.27	27.29	9.43	6.46	5.66	4.50	3.41	2.47	2.18
1938	130.93	93.19	58.64	14.95	4.28	2.18	2.90	0.00	3.92	3.77	3.56	3.41
1939	1.09	23.30	55.16	82.74	4.14	3.41	3.27	3.41	3.12	3.41	2.03	1.52
1940	93.41	42.24	44.05	9.80	4.72	1.38	0.00	0.00	0.00	0.00	0.00	0.00
1941	49.21	533.62	869.53	51.46	10.96	5.15	3.77	3.85	2.47	1.09	0.44	1.38
1942	2.69	836.66	398.08	498.74	23.44	12.12	2.32	1.74	0.58	0.22	0.15	3.27
1943	57.84	1339.75	493.52	199.73	28.96	3.99	2.32	1.89	1.02	0.44	0.22	5.37
1944	71.85	715.45	305.47	1017.30	31.93	13.14	122.51	7.48	5.15	2.32	1.23	0.58
1945	24.17	1063.46	618.64	129.48	23.22	8.56	5.15	2.32	1.31	0.58	0.29	0.22
1946	2.40	62.71	464.92	180.79	16.18	6.60	1.52	1.31	0.87	0.44	0.36	0.80
1947	0.65	26.85	315.42	245.96	8.93	2.47	3.41	3.77	3.05	1.09	0.44	0.22
1948	0.36	29.32	19.16	11.90	5.37	2.32	0.80	0.22	0.22	0.22	0.15	0.15
1949	1.74	58.93	66.04	19.89	5.01	0.65	0.36	0.36	0.29	0.22	0.15	0.44
1950	0.29	383.20	155.39	1197.72	121.42	35.13	18.36	13.43	7.33	3.19	1.45	0.51
1951	5.81	148.93	139.49	15.89	2.90	1.67	1.74	0.73	0.58	0.29	0.22	0.22
1952	0.36	27.00	128.24	41.80	15.17	5.59	4.14	4.79	1.96	0.87	0.44	0.22
1953	5.59	15.82	9.65	58.79	22.43	7.33	5.01	3.85	2.47	4.14	1.89	1.02
1954	55.74	404.47	135.79	654.71	150.38	20.90	14.01	11.10	8.42	5.95	4.28	3.12
1955	21.19	10.96	331.31	889.49	78.38	31.86	18.00	10.16	5.95	2.40	0.00	0.00
1956	27.80	525.89	588.66	383.06	379.79	34.11	16.33	15.75	12.34	9.00	5.88	4.64
1957	66.62	163.37	109.01	26.85	4.94	2.18	1.45	0.65	0.58	0.36	0.29	0.15
1958	1.96	235.51	23.30	67.42	31.14	9.80	4.64	1.67	1.09	0.73	0.44	0.29
1959	0.29	56.68	187.46	895.22	99.65	50.22	21.99	13.14	11.61	9.80	6.24	4.57
1960	0.00	124.47	239.36	57.26	20.47	9.07	8.85	3.70	2.54	1.45	0.80	0.36
1961	4.94	116.70	71.78	870.48	79.40	45.07	25.18	14.73	27.36	5.59	2.98	1.31
1962	0.51	214.32	187.90	126.86	32.66	10.09	5.59	3.12	3.63	1.31	0.80	0.58
1963	10.52	69.53	231.15	300.32	42.53	15.31	10.16	5.88	4.28	2.69	1.52	1.23
1964	9.73	124.83	391.91	97.76	29.47	12.85	8.78	8.13	5.01	2.25	1.52	0.94
1965	0.73	124.83	86.44	25.11	189.71	2.25	1.81	1.23	0.94	0.80	0.65	0.36
1966	47.97	110.17	24.60	51.17	2.83	0.80	0.58	0.58	0.58	0.51	6.68	1.45
1967	13.50	354.68	182.75	253.80	47.46	14.95	17.93	7.11	4.43	2.25	1.23	0.87
1968	0.58	251.91	99.21	23.95	9.80	3.05	2.10	1.38	1.16	0.94	0.58	0.29
1969	4.72	82.23	64.30	36.43	2.40	0.65	0.36	0.29	3.85	0.36	0.15	0.80
1970	46.09	13.50	352.14	440.46	82.16	31.86	18.43	4.57	2.69	1.16	0.58	0.29
1971	3.56	209.53	99.94	130.20	10.52	3.05	2.18	1.23	1.02	0.51	0.29	0.00
1972	20.25	92.24	60.82	14.37	1.02	0.65	0.58	0.51	0.51	0.36	0.44	0.15
1973	2.47	144.21	970.85	2009.48	205.61	54.00	23.51	13.35	8.78	4.64	2.54	6.75
1974	143.77	44.34	140.29	21.05	18.80	2.54	1.02	23.08	14.73	15.02	11.39	7.26
1975	11.32	258.88	398.37	504.84	133.25	22.72	9.00	5.88	5.52	1.60	0.80	1.23

The summary statistics namely mean, standard deviation and skewness, of the monthly flows are presented in Table 4.9.

Table 4.9: Summary Statistics of Historical Monthly Streamflows – River: Sabramathi; Station: Dharoi dam site.

Month	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Mean (Mm^3)	24.64	238.17	225.74	292.87	49.98	12.47	9.79	5.01	4.09	2.34	1.61	1.42
Standard Deviation (Mm^3)	35.13	295.36	228.16	419.78	73.89	14.32	19.45	5.38	5.12	3.06	2.33	1.87
Skewness	1.90	2.11	1.59	2.23	2.70	1.53	4.92	1.43	2.70	2.39	2.45	1.77

It may be noted that nearly 87% of the flows are received during the period July to September (3 months). The coefficient of variation of the flows are above 1.0, for all the months. The skewness coefficient of the flows is also high in all the months, the highest being in December (about 5). The model parameters for the three periodic stochastic models considered in this study, are tabulated in Table 4.10. While k-NN and PMABB are non-parametric stochastic models, HMBB is a hybrid stochastic model and hence has both parametric and non-parametric components. The periodic AR(1) (PAR(1)) parametric component does the partial pre-whitening that helps in capturing the major part of the linear short-term dependence; and the resampling of the resulting residuals using a moving block of size L = 24 months, helps in capturing the higher-lag linear dependence and a significant part of the non-linearity if present. Moreover, the moving block resampling of the residuals enables reproduction of the features of the marginal distribution, namely, asymmetry and multi-modality. For the k-NN model, the default choice of dependence order d =1 and the number of nearest neighbors $k = n^{0.5}$ (proposed by Lall and Sharma,1996) are adopted, where n is the choice of the sample record. For the PMABB model, a block size of L = 4 months was found to be sufficient, since the flow data are resampled using the matched block bootstrap proposed by Carlstein et al. (1998) which is a method suited for modeling dependent data. The number of elements considered for the matching exercise is 5 and p is the smoothing parameter that aims to

achieve variability in the synthetic simulations. The parameters of the three stochastic models are listed in Table 4.10.

Table 4.10: Parameters of the Stochastic Models

Models	Parameters
k-NN	a) Dependence Order, d = 1; b) No. of neighbours considered for residual resampling, k = 6.
HMBB (a) PAR(1) parameters from June – May	a) **Jun** 0.4612, **Jul** 0.1023, **Aug** 0.4533, **Sep** 0.4620, **Oct** 0.5218, **Nov** 0.6578, **Dec** 0.3611, **Jan** 0.3251, **Feb** 0.8336, **Mar** 0.7450, **Apr** 0.8806, **May** 0.7050 b) Block size for resampling the residuals, L = 24 months
PMABB	a) Length of within-year matched block, L = 4 months; b) No. of data points used for rank matching, w = 5; c) Smoothing parameter, p = 0.90-1.10.

The performance of the three stochastic models in terms of reproducing the summary statistics of the periodic historical flows is presented in Table 4.11. It is seen from the Table 4.11 that:

The k-NN model displays some bias in reproducing the monthly means and the monthly standard deviations of flows, while the other two models perform slightly better.

All the three models are found to reproduce the skewness present in the historical flows in all the 12 months.

Table 4.11: Reproduction of Summary Statistics of Historical Streamflows – River: Sabramathi; Station: Dharoi dam site.

	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Mean (Mm^3)												
Historical	24.64	238.17	225.74	292.87	49.98	12.47	9.79	5.01	4.09	2.34	1.61	1.42
k-NN*	23.19	259.09	233.52	296.60	47.43	12.61	10.22	4.78	3.82	2.11	1.34	1.48
HMBB*	25.98	241.05	223.20	286.41	47.40	12.13	9.35	4.88	3.97	2.29	1.60	1.41
PMABB*	25.18	244.11	231.81	297.49	51.33	12.96	10.31	5.23	4.29	2.44	1.66	1.51
Standard Deviation (Mm^3)												
Historical	35.13	295.36	228.16	419.78	73.89	14.32	19.45	5.38	5.12	3.06	2.33	1.87
k-NN*	32.69	305.66	230.61	392.83	68.34	14.09	20.41	5.03	4.79	2.67	1.82	1.75
HMBB*	33.02	294.71	224.25	410.41	70.47	14.08	18.16	5.23	4.99	2.95	2.27	1.82
PMABB*	35.26	302.90	219.36	405.84	73.39	14.27	19.87	5.44	5.28	3.09	2.33	1.89
Skewness												
Historical	1.90	2.11	1.59	2.23	2.70	1.53	4.92	1.43	2.70	2.39	2.45	1.77
k-NN*	1.88	1.96	1.56	2.03	2.95	1.50	4.70	1.49	3.01	2.48	2.49	1.61
HMBB*	1.65	2.06	1.63	2.24	2.79	1.58	5.13	1.42	2.76	2.39	2.38	1.75
PMABB*	1.84	2.08	1.51	2.12	2.72	1.44	4.70	1.39	2.64	2.29	2.40	1.66

The gross storage capacity of the Dharoi reservoir is 908 Mm^3, and the live storage capacity is 732 Mm^3 and the Full reservoir Level (FRL) of the dam is 189.59 m, as given in the weekly report of 81 important reservoirs of India (CWC, Government of India, 2007). The monthly yield factors as reported in the technical report UM-16 of NIH (NIH,1987) and the same are presented in Table 4.12. These values have been used in this study for the purpose of development of the Storage-Performance-Yield (S-P-Y) relationships.

Table 4.12: Monthly Yield Factors – River: Sabramathi, Dharoi dam site (Source: NIH, 1987).

Month	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Monthly Yield Factor	0.09	0.06	0.06	0.057	0.093	0.0914	0.0914	0.0914	0.0914	0.0914	0.0914	0.0914

The reservoir storage performance measures have been computed from the 41-year long historical flow sequence as well as the 4100-year long synthetic flow sequences

generated from the three stochastic streamflow models considered (k-NN, HMBB and PMABB), for the possible combinations of storage capacity and yield (expressed in % of Mean Annual Flow) given below:

Storage capacities (Mm^3): 500; 600; 700; 732; 800; 900; 1000.

Yield (% MAF) to be supplied by the reservoir: 50, 55, 60,…, 90, 95.

A computer program is developed for the reservoir performance computation. The seven reservoir performance indicators listed in an earlier section, can be computed using this program for the given reservoir storage capacity, yield, monthly yield factor, historical flow sequence, synthetic flow sequence and the reservoir operating policy (either standard operating policy or any hedging policy). A typical output of the simulation program showing the reservoir balance and the computed values of the reservoir performance are presented is given in (Appendix I). This output is obtained by routing the historical flows through the reservoir using standard operating policy.

For brevity, only the salient results are presented and discussed in the following paragraphs.

The reservoir performance-yield relationships obtained for the existing reservoir storage capacity of 732 Mm^3, using the historical flow sequence as well as the three long synthetic sequences generated from the three periodic stochastic models considered are presented in Table 4.13

Table 4.13: Comparison of the Performance-Yield Relations - Historical Vs Synthetic Flows

Yield (%MAF)	OBR	VBR	Resilience	Event Vul (Mm^3)	Period Vul (Mm^3)	Mean Event Deficit (Mm^3)	Mean Period Deficit (Mm^3)
Historical							
50	0.9695	0.9712	0.2000	245.19	39.67	170.64	34.13
55	0.9451	0.9471	0.1852	324.75	43.64	207.29	38.39
60	0.9106	0.9252	0.2273	368.15	47.61	159.78	36.31
65	0.8760	0.8931	0.1803	411.54	51.58	224.73	40.52
70	0.8476	0.8614	0.1600	454.94	55.54	287.73	46.04
75	0.8171	0.8308	0.1556	498.34	59.51	322.51	50.17
80	0.7886	0.8032	0.1346	541.74	63.48	400.18	53.87
85	0.7561	0.7756	0.1417	638.42	67.45	399.32	56.57
90	0.7256	0.7469	0.1556	789.62	71.41	385.95	60.04
95	0.7012	0.7188	0.1565	942.96	75.38	413.32	64.67
k-NN							
50	0.9772	0.9801	0.2422	364.43	39.66	130.21	31.54
55	0.9564	0.9625	0.2240	538.29	43.63	152.88	34.24
60	0.9297	0.9396	0.2175	852.14	47.60	171.55	37.32
65	0.8989	0.9125	0.2083	1018.10	51.57	195.30	40.69
70	0.8669	0.8830	0.1960	1216.30	55.53	226.98	44.50
75	0.8338	0.8532	0.1873	1421.60	59.53	255.77	47.91
80	0.8021	0.8234	0.1804	1626.90	63.57	286.17	51.63
85	0.7737	0.7945	0.1716	1807.90	67.61	325.47	55.84
90	0.7446	0.7667	0.1660	1932.90	71.64	358.23	59.47
95	0.7163	0.7400	0.1594	2057.90	75.68	395.30	63.00
HMBB							
50	0.9557	0.9607	0.2082	367.15	39.67	153.90	32.04
55	0.9307	0.9379	0.1897	691.96	43.64	187.94	35.66
60	0.9023	0.9122	0.1801	817.00	47.61	216.40	38.98
65	0.8693	0.8837	0.1694	918.76	51.58	246.81	41.82
70	0.8394	0.8546	0.1591	1024.00	55.54	288.18	45.85
75	0.8064	0.8246	0.1538	1151.50	59.51	319.48	49.13
80	0.7703	0.7943	0.1538	1321.00	63.48	336.95	51.82
85	0.7386	0.7640	0.1547	1448.60	67.45	358.67	55.49
90	0.7081	0.7346	0.1530	1576.10	71.41	386.95	59.20
95	0.6843	0.7069	0.1482	1703.70	75.38	430.60	63.80
PMABB							
50	0.9609	0.9661	0.2269	574.01	39.66	138.54	31.43
55	0.9342	0.9430	0.2040	710.24	43.63	168.83	34.44
60	0.9045	0.9164	0.1925	819.26	47.60	197.26	37.97
65	0.8728	0.8879	0.1809	908.66	51.57	229.14	41.46
70	0.8406	0.8588	0.1729	998.06	55.58	259.47	44.87
75	0.8098	0.8295	0.1661	1087.50	59.64	292.59	48.61
80	0.7803	0.8012	0.1617	1176.90	63.67	323.86	52.36
85	0.7507	0.7734	0.1579	1266.30	67.71	354.11	55.90
90	0.7237	0.7466	0.1531	1355.70	71.75	389.81	59.69
95	0.6982	0.7214	0.1478	1445.10	75.79	429.04	63.42

Following points may be observed from Table 4.13:

All the three periodic stochastic models preserve the storage performance characteristics well. However, HMBB and PMABB models perform better than k-NN model. The k-NN model overestimates the two reliabilities, resilience and event vulnerability and under estimates the mean event deficit. The PMABB model is seen to be marginally better than the HMBB in predicting both the reliabilities. Hence, the PMABB model is adopted for further analysis and discussion in this research work.

4.6.2.1 Storage - Volume Reliability - Yield Relationships

It is seen from the storage-volume reliability-yield relationships (Fig 4.10) that the volume reliability declines sharply with increasing yield, while the increase in volume reliability with increase in reservoir storage capacity is comparatively less. Moreover, the "increase in volume reliability with increase in storage capacity" decreases for higher values of storage capacity and this decrease is more pronounced at lower yields, since the magnitudes of reliability being dealt with are quite high (above 95%). The actual live storage capacity of 732 Mm^3 would yield a volume reliability of nearly 83% for 75% yield, while it would decrease to a value of 77.3% for 85% yield. If the same volume reliability of 83% is to maintained for an increased yield of 85%, than the storage capacity is to be increased to 1000 Mm^3, which will require a reservoir storage capacity expansion.

4.6.2.2 Storage - Occurrence Reliability - Yield Relationships

It may be noted from Fig 4.11 that the storage-occurrence reliability-yield relationships follow a similar trend as the storage-volume reliability-yield relationships except that the occurrence reliabilities, in general are about 1% to 2% lower than the volume reliabilities.

4.6.2.3 Storage – Resilience - Yield Relationships

It may be observed from Fig.4.12 that the resilience, in general, decreases with increase in the yield (for the given storage capacity). However, the decrease in resilience with increase in yield is observed to be small. This is because, as yield increases, the increase in the number of events (increase in the value of the numerator) is found to be nearly proportional to the increase in the number of failure periods (increase in the value of the denominator). Likewise, the increase in the resilience with increase in the storage capacity (for the given yield), is also negligible, the reason being: the decrease in the number of events (decrease in the value of the numerator) is nearly proportional to the decrease in the number of failure periods (decrease in the value of the denominator).

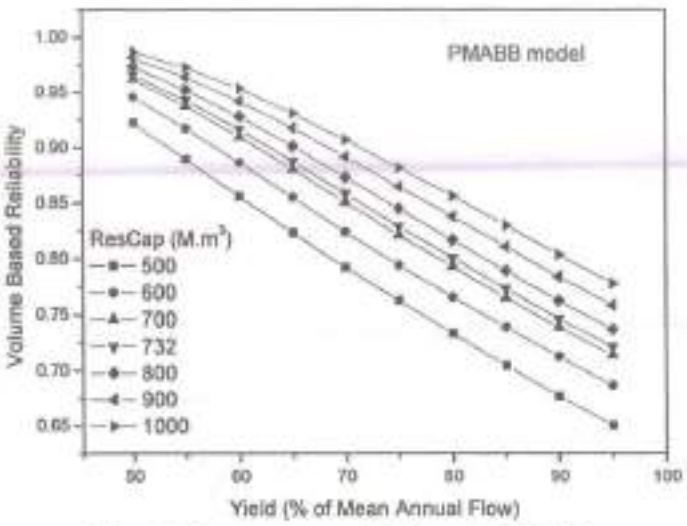


Fig 4.10: Storage - Volume Reliability - Yield Relationships

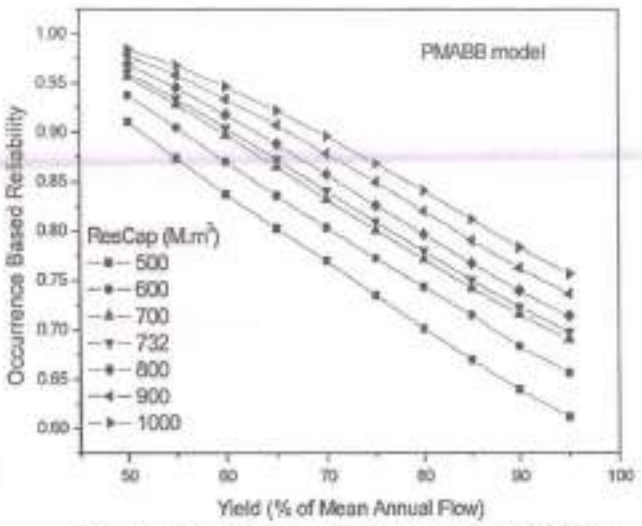


Fig 4.11: Storage - Occurrence Reliability - Yield Relationships

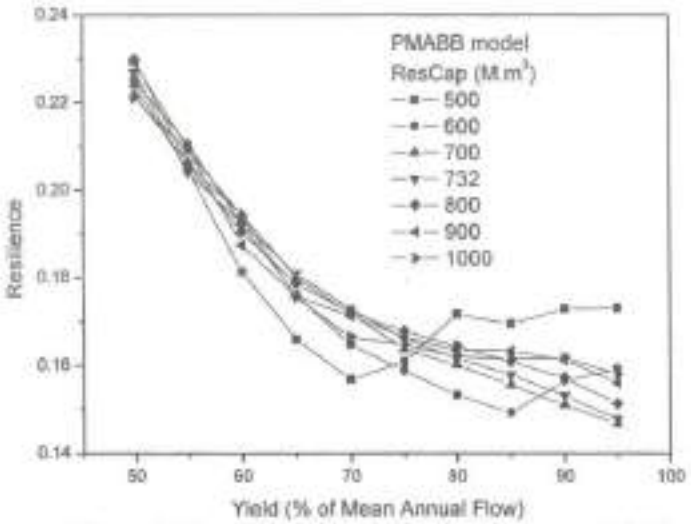


Fig 4.12: Storage- Reslience - Yield Relationships

4.6.2.4 Storage - Period Vulnerability - Yield Relationships

The period vulnerability-yield relationships (Fig. 4.13) do not exhibit any variation with reservoir storage capacity, while the period vulnerability increases significantly with the target yield and this increase is found to be directly proportional to the target yield. This is because at all the storage capacities, the period vulnerability for a given target yield becomes equal to the target yield itself, due to the occurrence of highly critical inflows during a certain period in the 4100-year long synthetic sequence of generated flows; and even if the storage capacity is assumed to be as high as 1000 Mm^3, the situation does not at all result in any improvement despite the additional filling space being available in the reservoir.

4.6.2.5 Storage - Event Vulnerability - Yield Relationships

The storage-event vulnerability-yield relationships (Fig. 4.14) seem to follow nearly the same trend as the storage-period vulnerability-yield relationships (Fig. 4.13), except for a minor local variation at 55% target yield.

4.6.2.6 Storage - Mean Period Deficit - Yield Relationships

For a given storage capacity, with increase in the target yield, the mean period deficit also increases systematically, in line with the decrease in occurrence reliability (Fig. 4.15). This is as per the expected trend. However, for a given target yield, the change in the mean period deficit with the storage capacity is found to be almost insignificant, which is also in tune with the trend of the behaviour of the occurrence reliability with storage capacity.

4.6.2.7 Storage - Mean Event Deficit - Yield Relationships

For a given storage capacity, with increase in the target yield, the mean event deficit also increases systematically (Fig. 4.16). However, for a given target yield, the change in the mean event deficit with the storage capacity is found to be almost insignificant at lower yields, while it is somewhat significant at higher yields. Moreover, it may be noted from Fig. 4.16 that for lower storage capacities and higher target yields, the increase in the mean event deficit with increase in the target yield is insignificant. This is due to the unproportionate increase in the number of events when compared with the number of failure periods.

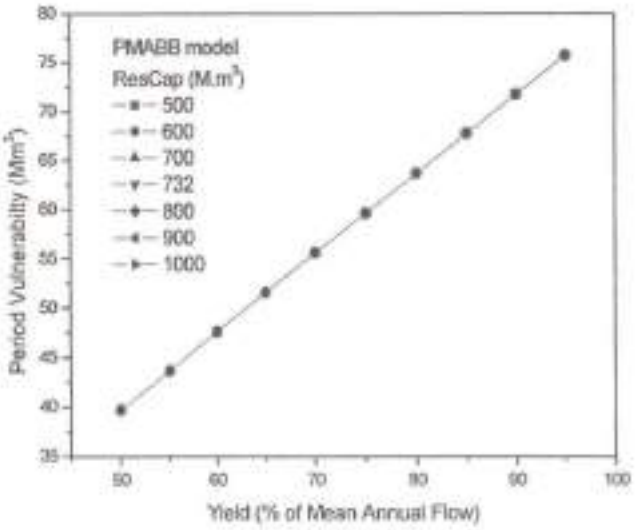


Fig 4.13: Storage – Period Vulnerability - Yield Relationships

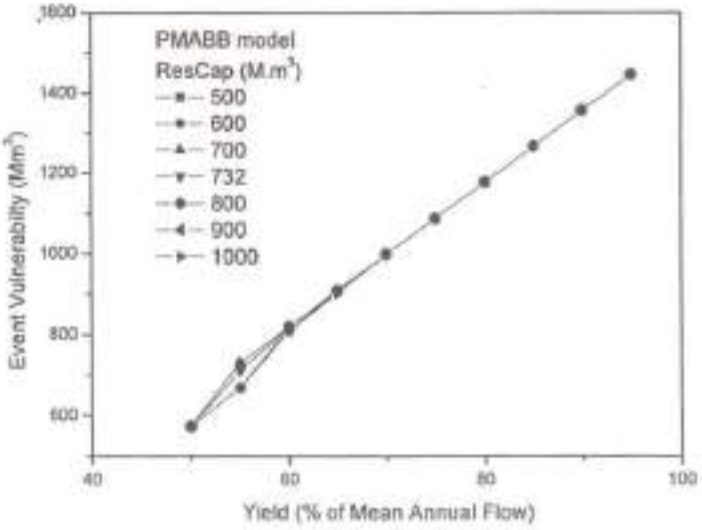


Fig 4.14: Storage – Event Vulnerability - Yield Relationships

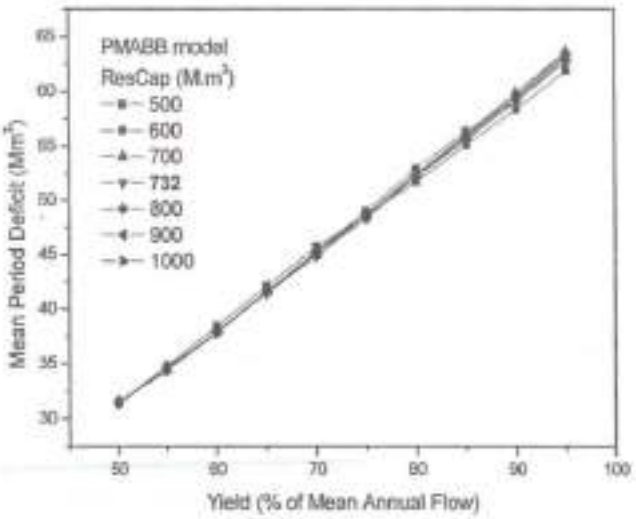


Fig 4.15: Storage – Mean Period Deficit - Yield Relationships

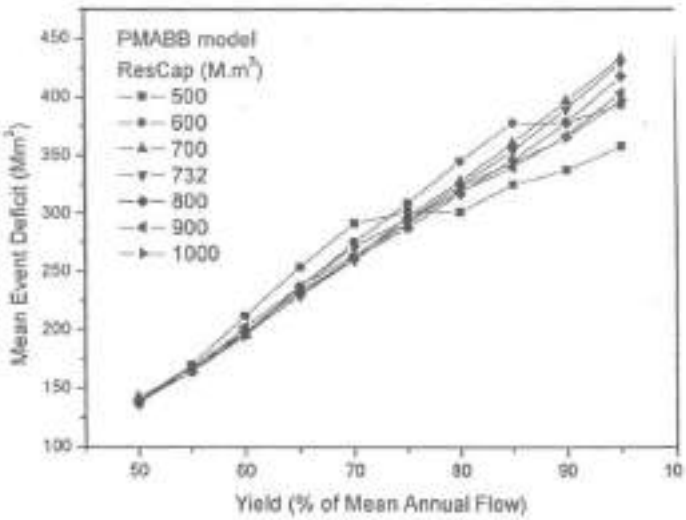


Fig 4.16: Storage - Mean Event Deficit - Yield Relationships

4.7 SUMMARY AND CONCLUSIONS

Over-year Reservoirs

Following behaviour analysis based on stochastic simulation, S-P-Y relationships have been constructed for use in planning and design applications of over-year water supply reservoirs, using standard operating policy. These relationships are useful in: (i) gaining an understanding of the variation of reservoir performance indicators namely, reliability, resilience, and vulnerability on the storage-yield plane; (ii) identifying the sensitive ranges of storage capacity of the over-year reservoirs, with regard to performance characteristics; and (iii) selecting between capacity expansion and demand management options, in case of deficit water supply systems.

For highly over-year reservoirs, resilience and vulnerability do not seem to improve with increase in storage capacity, and hence, the decision regarding storage capacity depends

on reliability. Approximate ranges of over-year storage capacity of water supply reservoirs, required to meet various target yields with a desirable range of reliability (0.95 to 0.99) are presented for a few selected cases of C_y and ρ_1 of annual streamflows. For lower target yields, there is no significant improvement in resilience at low as well as high storage capacities, while a significant improvement is noted for the range in between. With increase in target yield, this transition range widens and moves towards higher storage capacities, eventually ending up in flat storage-resilience relationships for high target yields. In addition, appreciable reduction in vulnerability is noted in a certain range of storage capacity and this range widens and moves towards higher storage capacities with increase in target yield and/or C_y. This can be exploited in certain water supply systems, wherein the marginal value of either increase in resilience or decrease in vulnerability or both, is quite high. A S-P-Y database with a search facility has also been developed that would help in planning and design of reservoir capacity and in decisions regarding capacity expansion or implementation of demand management programs.

Within-year Reservoirs

For the Dharoi reservoir streamflows, the reservoir storage performance measures have been computed from the 41-year long historical flow sequence as well as the 4100-year long synthetic flow sequences generated from the three stochastic streamflow models considered (k-NN, HMBB and PMABB), for seven combinations of storage capacity and ten combinations of yield (expressed in % of Mean Annual Flow) given below:

Storage capacities (Mm^3): 500; 600; 700; 732; 800; 900; 1000.

Yield (% MAF) to be supplied by the reservoir: 50, 55, 60,..., 90, 95.

The results and investigation are presented only for the PMABB model, although the HMBB model yields competitive performance.

The reservoir performance-yield relationships obtained are presented for the existing reservoir storage capacity of 732 Mm^3 of Dharoi reservoir, using the historical flow sequence as well as the three long synthetic sequences generated from the three periodic stochastic models considered. The actual live storage capacity of the Dharoi reservoir (732 Mm^3) would yield a volume reliability of nearly 83% for 75% yield, while it would decrease to a value of 77.3% for 85% yield. If the same volume reliability of 83% is to maintained for an increased yield of 85%, than the storage capacity is to be increased to 1000 Mm^3, which will require a reservoir storage capacity expansion.

The decrease in resilience with increase in yield is observed to be small. Likewise, the increase in the resilience with increase in the storage capacity (for the given yield), is also negligible, the reason being: the decrease in the number of events (decrease in the value of the numerator) is nearly proportional to the decrease in the number of failure periods (decrease in the value of the denominator).

CHAPTER 5

OPTIMAL HEDGING RULES FOR WATER SUPPLY RESERVOIRS

5.1 INTRODUCTION

Water Resource projects involve huge initial outlay and hence it is important to operate them efficiently not only during normal operating conditions but also during extreme situations like droughts. The standard operating policy (SOP) is a simple reservoir operating policy that satisfies the demand when sufficient water is available and if not supplies the available water. This policy minimizes the total shortage over the operating horizon. However during droughts, SOP would result in single period of severe shortage (Vulnerability), which could cause heavy loss of life and property. Thus, effective demand management strategies must be devised to reduce the severity of shortage by distributing the deficits over longer periods. Hedging is one of the simple and common demand management strategies employed by water supply managers to reduce the severity of droughts. Hedging increases water stored in the reservoir by accepting small currents deficits to guard against unacceptable large deficits that may occur in future. Thus, hedging provides insurance for high-valued water uses where reservoirs have low refill potentials or highly uncertain inflows.

5.2 LITERATURE REVIEW

Hedging rule decides the storage allocation of water across time to minimize the impact of the drought. The optimal appropriation of water can be done by analysing the benefits of current release as against the benefits of storing water for future use as carryover

storage (Draper & Lund, 2004). However, it is difficult to derive the actual benefit function since it is time variant and case specific. Hence, the water supply characteristic of reservoirs is used to evaluate their performance. The water shortage characteristics are the primary criteria for evaluating the supply-demand relationship of reservoir during drought (Shiau, 2003). Hashimoto et al. (1982) proposed the indicators reliability, resilience, and vulnerability to measure and monitor the performance of water resource systems. Occurrence based reliability is an indicator of the frequency of occurrence of the deficit, and is computed as the number of times the target demand is satisfied to the total number of operating time periods. Volume based reliability is a measure of the ability to satisfy the volume of demand and is computed as the ratio of total water supplied to the total water demanded. . Shortage ratio is the ratio of total volume of deficits to total volume of water demanded and is an indicator of the total deficit in meeting demand during the operating horizon. The shortage ratio is also the complement of the volume-based reliability. Resilience symbolizes the recovery rate from water shortage (failure) to normalcy (success) and is defined as the number of times the system has moved from failure to success to the total number of periods the system has been in the failure state. Vulnerability is the largest single period shortage encountered during the period of operation of the reservoir and is a measure of the severity of the shortage. The evaluation of the trade-off between these performance indicators would help the decision makers to arrive at the optimal hedging rule (Srinivasan et. al, 1998). Hedging aims to reduce vulnerability by increasing the storage, at the onset of the drought, by proactively accepting small deficits even when sufficient water is available. This would reduce the severity, by distributing the same over more number of periods. However, in some cases,

deficits may be accrued in anticipation of a severe drought in future, but the situation may not become so severe in future. This may increase the total shortage over the operation horizon. Thus, a reduction in vulnerability may result in an increase in shortage ratio and vice-versa. The optimal hedging policy must aim to reduce the vulnerability significantly with a minimal increase in shortage ratio.

The trigger for the initiation and the termination of hedging along with the amount of rationing to be done in each time step typically characterize a hedging rule. The parameters of a hedging rule can be expressed as a function of water available in the reservoir, which is the sum of the current storage and the expected inflows into the reservoir. Bayazit and Unal (1990) defined the two-point hedging rule in terms of starting water availability (volume of water availability above which the reservoir release is hedged, SWA) and ending water availability (the volume of water availability at which hedging is stopped and normal situation is restored, EWA). In case of the two-point hedging rule, when the water availability falls below the SWA, then the entire water present in the reservoir is released towards meeting the demand. So the storage of the reservoir at the end of this time period will become zero. In the next time period, water available becomes the same as the inflow into the reservoir. If the reservoir is still facing drought, then the water availability becomes very less and so the vulnerability will be the same as in case of standard operating policy. So in order to reduce vulnerability, it may be preferable to opt for a hedging policy with a low SWA, and high EWA.

The effectiveness of hedging rules can be enhanced by having control over the amount of water to be released during hedging. Srinivasan and Philipose (1996, 1998) included the *hedging factor* as an independent parameter along with the starting water availability (SWA) and the ending water availability (EWA) to define the modified two-point hedging rule. The hedging factor specified the amount of hedging to be done in each time step. The modified two-point hedging rule essentially provides an offset to the SOP in the period where hedging is done. This hedging rule is not flexible, as the slope of both the phases of rationing, and the amount of offset provided all depend on the single parameter "hedging factor". Hence although this rule can be effective in attaining low vulnerabilities, this may be realized at relatively high values of shortage ratio. They evaluated the trade-off among the reservoir performance indicators for thousands of hedging policies using Monte-Carlo simulation technique. This simulation model does not yield the optimal trade-off surface between the performance indicators considered. Shih and ReVelle (1994) used mixed-integer non-linear programming technique and polytope search procedure to find the optimal linear hedging rule with starting water availability as the only decision vector, by minimizing the maximum shortfall or vulnerability. They found that the mixed-integer non-linear programming technique gives better solution than polytope search procedure but at the expense of more computing time. They also suggested that the optimal hedging rule could be converted into multiple discrete hedging rules for practical implementation. Shih and ReVelle (1995) formulated an explicit two-phase discrete hedging rule and implemented the same using mixed-integer programming model. This

formulation sought to determine the trigger volumes for the different phases of rationing with the objective of maximizing the number of months in which no rationing (or maximizing the reliability) would be required, subject to a constraint on the number of months with second phase of rationing. The amount of the rationing to be done during the different phases of hedging was pre-fixed and by progressively increasing the intensity of rationing, the vulnerability could be decreased. Although the discrete hedging rule can be effective in achieving low vulnerability, it is not flexible as the slopes of both the phase of rationing are fixed at zero. This is likely to reduce the possible number of competent solutions, thus limiting the flexibility to the decision maker. Moreover, since the amount of rationing was not internally optimized within the formulation, the optimal hedging policy could not be arrived at. Also, this formulation was solved for only a single critical drought and cannot be easily solved for long sequences as done in drought analysis, since large number of mixed integer variables would increase the computational burden. Neelakantan and Pundarikanthan (1999) determined the threshold to initiate hedging by minimizing the sum of the squared deficits. Oliveira and Loucks (1997) proposed a piecewise linear hedging rule to derive the optimal hedging operating policy for multi-reservoir systems using genetic algorithm (GA). However, the performance of the hedging rule was evaluated based only on the single objective of minimizing the total deficit. Shiau (2003) developed a reservoir supply index, which was used as an indicator for the onset , the termination and the quantum of rationing to be done at each time period. The reservoir supply index was defined as the probability that reservoir storage plus inflows would

be sufficient to meet the demand. They also used a number of performance indicators to quantify the single period shortage, the event shortage and the long-term effect. Recently, Shiau and Lee (2005) have explored two types of hedging, one based on water availability (defined as storage plus inflow) and the other based on the potential shortage condition within a specific future lead-time period. The length of lead-time is determined by minimizing vulnerability and shortage ratio. Compromise programming was used to solve the multi-objective problem which, in turn, introduced three additional variables, in terms of weights and an exponent. One of the significant drawbacks of the compromise programming is the sensitivity towards weights. In their works, the weights and the exponent are not internally optimized and hence many optimization runs are to be made if a sensitivity analysis is to be done on the weights and the exponent.

5.3 PROPOSED STUDY

After critically reviewing the existing hedging rules for water supply reservoir operation during droughts, a new hedging rule is proposed in this study by effectively combining the two existing hedging rules, namely, modified 2-point hedging rule and discrete hedging rule, with the aim of introducing more flexibility into the hedging formulation and thus obtain more competent trade-off solutions. The new hedging rule proposed herein, being more flexible and more general than the other hedging rules discussed, is expected to provide more efficient, effective and continuous distribution of the trade-off relation between the objective functions. Although a number of hedging rules exist in literature, there has not been

any comparative study that investigates the efficacy of the trade-off solutions obtained using these hedging rules in a multiple objective optimization model framework that seeks to minimize the vulnerability as well as the shortage ratio over the period of operation considered. Hence, there is a need to develop a multi-objective optimization framework that is able to compare across different hedging rules and arrive upon the hedging rule that yields the best set of trade-off solutions for the given problem. This framework should also explicitly incorporate the different probabilistic reservoir performance indicators as constraints within the multi-objective framework. An attempt is made in this study to develop such a framework that would enable the managers of water supply projects to find the most appropriate hedging rule so that the water supply reservoirs can be operated prudently during droughts.

5.4 MULTI-OBJECTIVE OPTIMIZATION FRAMEWORK

The following multi-objective optimization framework is developed in this research work to obtain the optimal trade-off between the two surrogate objective functions mentioned below, for the proposed hedging rule as well as four of the popular hedging rules existing in the literature.

Objective Functions:

1) Minimize Period Vulnerability: $Z_1 = \text{Minimize}\{\text{Max}\,(D_t - R_t)\}$ ...(5.1)

2) Minimize Shortage Ratio: $Z_2 = \text{Minimize}\,\{\sum((D_t - R_t) / \sum D_t\}$ ...(5.2)

where D_t denotes the demand at time t, R_t denotes the release to be made at time t.

Constraints:

Performance Related Constraints: While shortage ratio (complement to volume based reliability) and vulnerability are two of the two primary indicators that need to be minimized during critical drought, it is also desirable to keep some of the other performance related constraints (such as the number of deficit periods during the operation horizon) under check. This will help in increasing the perceived confidence of the stakeholders in the drought management system being practiced. Some of the typical performance related constraints are listed below (eqs. 5.3-5.7).

Occurrence based reliability

$$R_{occ} = 1 - \left(\frac{\sum_{j=1}^{M} d_j}{T} \right) \geq c_1 \qquad (5.3)$$

d_j = duration of the j^{th} failure event
M = total number of failure events
T = total number of periods of operations

Resilience

$$Res = \frac{M}{\sum_{j=1}^{M} d_j} \geq c_2 \qquad (5.4)$$

d_j = duration of the j^{th} failure event

Event vulnerability

$$V_E = \max\left[E_1, E_2, \ldots\ldots E_j \ldots E_M\right] \leq c_3 \qquad (5.5)$$

$$E_j = \sum_{i=1}^{d_j} (D_i - R_i) \quad j=1,2,\ldots,M$$

D_i = demand at period i
R_i = release at period i
E_j = total deficit in j^{th} failure event
d_j = duration of the j^{th} failure event

Mean event deficit

$$MD_{event} = \frac{\sum_{t=1}^{T} D_t - R_t}{M} \leq c_4 \qquad (5.6)$$

Mean period deficit

$$MD_{period} = \frac{\sum_{t=1}^{T} D_t - R_t}{N} \leq c_5 \qquad (5.7)$$

N = total number of failure periods

In equations (5.3)-(5.7), $c_1, c_2, \ldots, c_5$ denote pre-specified limiting performance values.

Constraints specifying the hedging rules:

The constraints that specify the hedging rules are presented and discussed in this section. Four hedging rules are presented including the one proposed in this research work. Constraints corresponding to any one of the four rules (optional) can be activated within the framework and solved.

5.4.1 HEDGING RULES

5.4.1.1 Two-point Hedging Rule

The definition sketch of the two-point hedging rule suggested by Bayazit and Unal (1990) is presented in Fig. 5.1. Equations (5.8) to (5.14) describe this hedging rule.

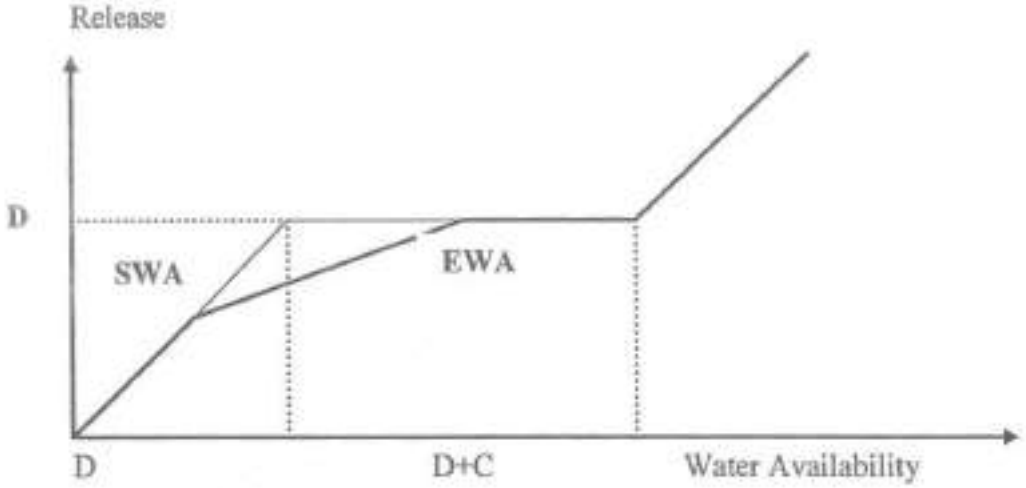


Fig. 5.1 Definition Sketch of two-point hedging rule

$$R_t = S_t + Q_t \quad \text{if } S_t + Q_t \le SWA_t \tag{5.8}$$

$$R_t = SWA_t + \frac{(S_t + Q_t - SWA_t)(D_t - SWA_t)}{(EWA_t - SWA_t)} \quad \text{if } SWA_t \le S_t + Q_t \le EWA_t \tag{5.9}$$

$$R_t = D_t \quad \text{if } EWA_t \le S_t + Q_t \le D_t + C \tag{5.10}$$

$$R_t = St + Q_t - C \quad \text{if } S_t + Q_t > D_t + C \tag{5.11}$$

$$S_{t+1} = S_t + Q_t - R_t \tag{5.12}$$

$$SWA_t = \alpha * D_t \tag{5.13}$$

$$EWA_t = D_t + (C * \beta) \tag{5.14}$$

In eqs. (5.8)-(5.14), S_t denotes the initial storage, S_{t+1} the final storage, R_t the release and Q_t the inflows during time *t* period and C the capacity of the reservoir. Herein, the water availability is defined as the sum of the current storage and the expected inflows. In this hedging policy, a linear hedging is implemented when the water availability falls between the starting water availability (SWA_t) and the ending water availability (EWA_t) (eq. 5.9). Below SWA_t, no hedging is done (eq 5.8) and the all the available water is released to satisfy the demand. If the water availability exceeds EWA_t (eq 5.10), then, no hedging is

implemented and water is released to satisfy the entire demand. If the water availability exceeds the sum of storage capacity and demand (eq 5.11), then after meeting the demand, the surplus amount of water is spilt over from the system. Equation (5.12) is the continuity equation for the single reservoir system that relates the final storage at the end of the period of operation to S_t, Q_t , and R_t . Herein, evaporation is not considered in the formulation.

In the optimization formulation of the two-point hedging rule, the decision vector consists of two parameters namely, starting water availability and ending water availability, expressed in terms of α and β, respectively.

5.4.1.2 Modified two-point hedging

The basic definition sketch of the modified two-point hedging rule suggested by Srinivasan and Philipose (1996, 1998) is presented in Fig 5.2. Equations (5.15) to (5.22) describe this hedging rule. This hedging rule rations on the demand when the water availability in any period falls in the range between the demand (D_t) and the ending water availability (EWA_t) (eq. 5.17); while, hedging is done on the water availability, when the water availability itself falls in the range between the starting water availability (SWA_t) and the demand (eq. 5.16). Hedging factor (HF) specifies the amount of rationing/hedging to be done. For simplicity, a constant HF is used in both stages. Below a water availability of SWA_t, the available water is released (eq. 5.15), whereas above a water availability of EWA_t, the full demand is released (no hedging) (eq. 5.18). When the water availability is greater than (D_t + C), the demand is fully satisfied and the surplus water is spilt from the reservoir (eq. 5.19).

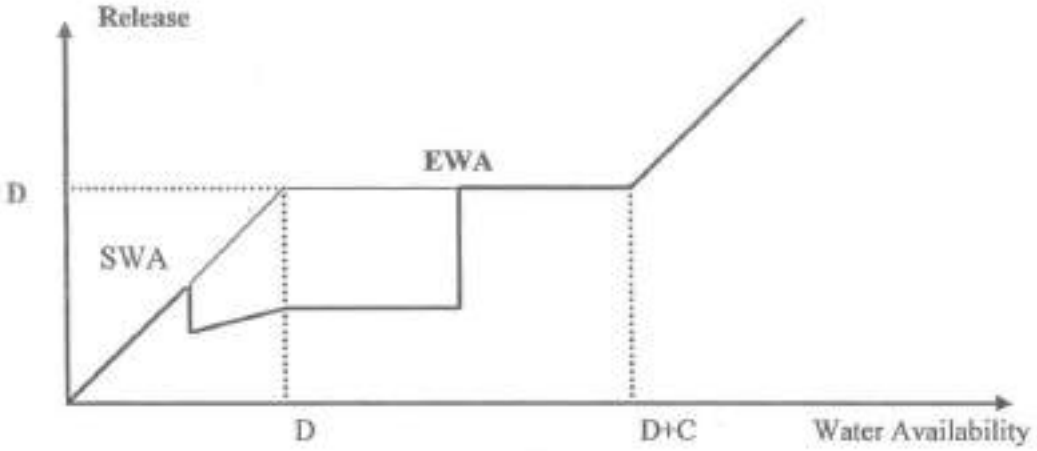


Fig.5.2 Sketch of modified two-point hedging rule

$$R_t = S_t + Q_t \quad \text{if } S_t + Q_t \leq SWA_t \tag{5.15}$$

$$R_t = (1\text{-}HF)^*(S_t + Q_t) \quad \text{if } SWA_t < S_t + Q_t \leq D_t \tag{5.16}$$

$$R_t = (1\text{-}HF)^*(D_t) \quad \text{if } D_t < S_t + Q_t \leq EWA_t \tag{5.17}$$

$$R_t = D_t \quad \text{if } EWA_t < S_t + Q_t \leq D_t + C \tag{5.18}$$

$$R_t = S_t + Q_t - C \quad \text{if } S_t + Q_t > D_t + C \tag{5.19}$$

$$S_{t+1} = S_t + Q_t - R_t \tag{5.20}$$

$$SWA_t = \alpha * D_t \tag{5.21}$$

$$EWA_t = D_t + (C * \beta) \tag{5.22}$$

The decision variables in case of the modified two-point hedging rule are α, β and hedging factor (HF).

5.4.1.3 Discrete hedging

The definition sketch of the discrete hedging rule proposed by Shih and Revelle (1994, 1995) is presented in Fig.5.3. Equations (5.23) to (5.31) describe this hedging rule. In this rule two phases of rationing of the demand are introduced

progressively and the amount of rationing depends on the severity expressed in terms of water availability. When the water availability exceeds V_{1p} (eq. 5.26), no rationing is implemented. When the availability goes below this value, only a part of the demand (α_1 * D_t) would be released. This is the first phase of rationing. If the water availability becomes lesser than V_{2p} (eq. 5.24), then a more severe second phase of water rationing is implemented. If the water availability becomes lesser than V_{3p} (eq. 5.23), no water is released from the system. On the other hand, if the water availability is greater than the reservoir storage capacity even after satisfying the full demand, then the surplus over the capacity will be spilt. This condition is described by eq. (5.27).

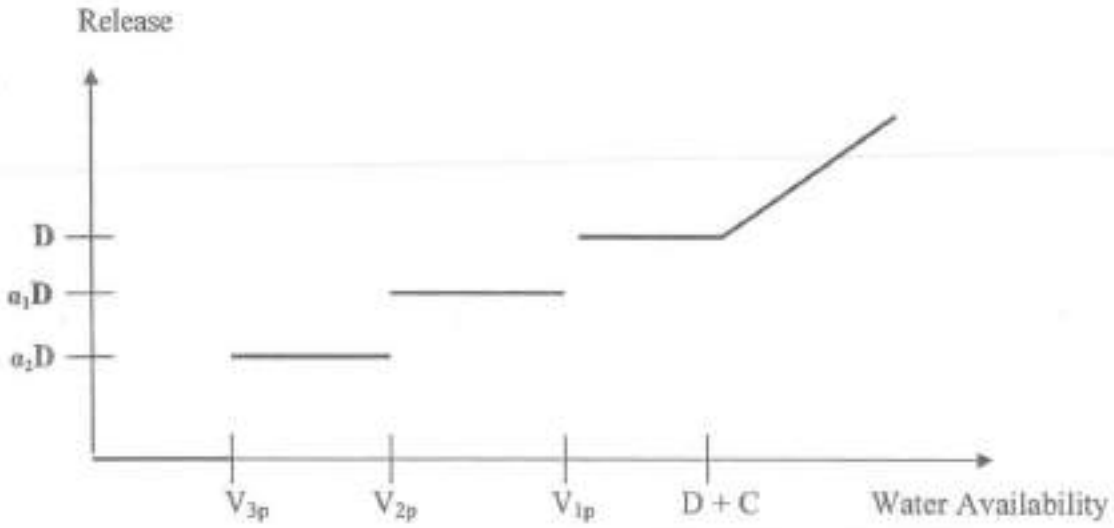


Fig. 5.3: Definition Sketch of discrete hedging rule

$$R_t = 0 \quad \text{if } S_t + Q_t \le V_{3p} \tag{5.23}$$

$$R_t = \alpha_2 D_t \quad \text{if } V_{3p} < S_t + Q_t \le V_{2p} \tag{5.24}$$

$$R_t = \alpha_1 D_t \quad \text{if } V_{2p} < S_t + Q_t \le V_{1p} \tag{5.25}$$

$$R_t = D_t \quad \text{if } V_{1p} < S_t + Q_t \le D_t + C \tag{5.26}$$

$$R_t = S_t + Q_t - C \quad \text{if } S_t + Q_t > D_t + C \tag{5.27}$$

$$S_{t+1} = S_t + Q_t - R_t \quad (5.28)$$

$$V_{1p} = D_t + (k_3 * C) \quad (5.29)$$

$$V_{2p} = D_t - (k_2 * C) \quad (5.30)$$

$$V_{3p} = D_t - (k_1 * C) \quad (5.31)$$

The decision variables of discrete hedging rule are α_1, α_2, k_1, k_2 and k_3.

5.4.1.4 Proposed Hedging Rule

The definition sketch of the new hedging rule proposed in the present study is shown in Fig. 5.4. Equations (5.32) to (5.39) describe this hedging rule. The new hedging rule consists of two phases of rationing and the amount of release in each phase is linear function of the water availability. The slope of each linear phase is a decision vector. In order to provide added flexibility the starting point of each phase of hedging is provided with an offset just like in case of modified two point hedging rule(Fig 5.2). The hedging rule proposed in the present study is a more generic hedging rule as the other twohedging rule are special case of this hedging rule. When the slope of both the phases of rationing is zero then the hedging rule becomes discrete hedging rule. If the slope of each linear phase equals the amount of offset provided then the proposed hedging rule becomes modified two-point hedging rule.

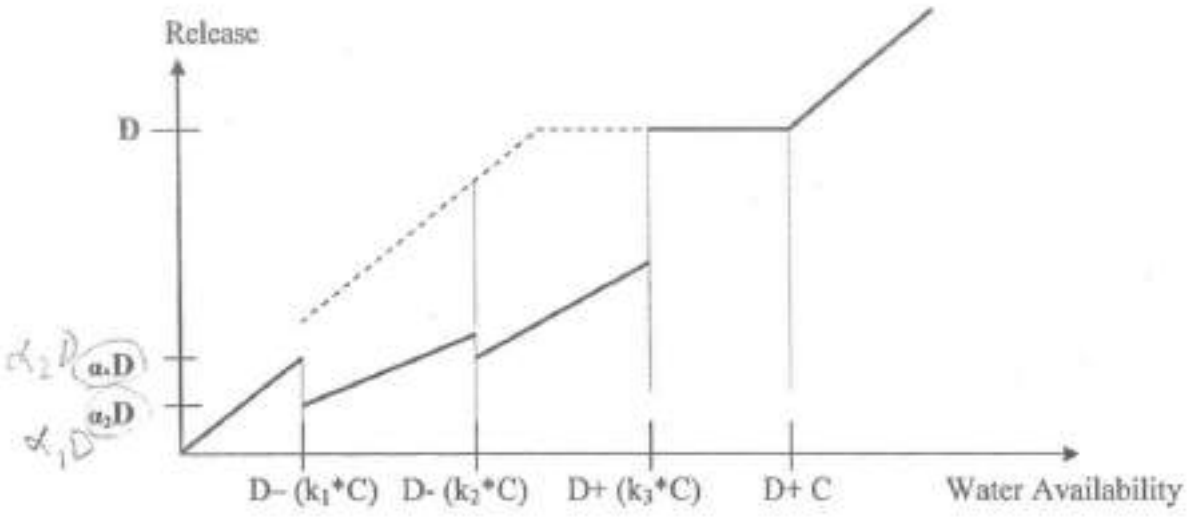


Fig 5.4: Definition Sketch of the Proposed Hedging Rule

According to this rule, when the water availability exceeds V_{1p} (eq. 5.35), no rationing is implemented. If the water available is less than V_{1p} and more than V_{2p} (eq 5.34) then the first linear phase of rationing is implemented. If the water availability goes below V_{2p} but more than V_{3p} (eq. 5.33) then a more severe phase of linear rationing is enforced. If the water available is less V_{3p} (5.32) then the available water is released. If the water available in reservoir after satisfying the demand exceeds the reservoir capacity then the excess water is spilled over (eq. 5.36).

$$R_t = S_t + Q_t \quad \text{if } S_t + Q_t <= V_{3p} \tag{5.32}$$

$$R_t = m_1*(S_t + Q_t - V_{3p}) + \alpha_1 D_t \quad \text{if } V_{3p} < S_t + Q_t <= V_{2p} \tag{5.33}$$

$$R_t = m_2*(S_t + Q_t - V_{2p}) + \alpha_2 D_t \quad \text{if } V_{2p} < S_t + Q_t <= V_{1p} \tag{5.34}$$

$$R_t = D_t \quad \text{if } V_{1p} < S_t + Q_t <= D_t + C \tag{5.35}$$

$$R_t = S_t + Q_t - C \quad \text{if } S_t + Q_t > D_t + C \tag{5.36}$$

$$V_{1p} = D_t + (k_3 * C) \quad (5.37)$$

$$V_{2p} = D_t - (k_2 * C) \quad (5.38)$$

$$V_{3p} = D_t - (k_1 * C) \quad (5.39)$$

The decision vectors of proposed hedging rule are α_1, α_2, m_1, m_2, k_1, k_2 and k_3.

5.5 SOLUTION TECHNIQUE

Genetic Algorithm (GA) is a search technique random search that explores the solution space for promising regions and then search for solutions more intensely in these promising regions. Genetic Algorithm are based on a simple assumption that the best solution is found in regions of solution space having high proportion of good solution (Oliveira and Loucks, 1997).The evolution starts from a population of completely random individuals and happens in generations based on the principle of the "survival of the fittest". Since GA deals with a population of points rather than a single point like in classical methods, it reduces the chances of getting trapped at some local optimum. Moreover multiple objective evolutionary algorithms (MOEAs) are suitable for handling complex problems involving discontinuities, disjoint feasible spaces and noisy function evaluation (Fonseca and Fleming, 1995). A powerful multi-objective genetic algorithm known as Non-Dominant Sorting Genetic Algorithm – II (Deb et al., 2002) is used for solving the multiple objective optimization model.

NSGA-II was proposed by Deb et.al. (2002) to overcome the high computational complexity of non-dominant sorting, lack of elitism and the need for specifying a sharing parameter in the non-dominant sorting based MOEAs. NSGA-II evolves the

trade-off surface from a random parent population of size *N*. The parent population is ranked based on non-domination and each non-dominated solution is assigned a fitness rank. The child population of size *N* is created from the parent population through binary tournament selection, crossover and mutation. Binary tournament selection ensures that the solution having better fitness function has higher probability of selection for cross-over and mutation. Crossover operator combines two chromosomes to produce a new chromosome. The new chromosome may give better fitness function than the parents if it takes best characteristics from the parents. Mutation operator alters one or more gene in a chromosome. As a result a new gene is added to gene pool and genetic algorithm may be able to arrive at better solution through a new search path. Mutation occurs according to user defined probability and it is set at a low value to prevent the search from turning to random search. A mating pool of size 2N is formed by combining the parent population and the child population. The mating pool is fast non dominated sorting procedure to identify the non-dominant fronts. If the number of non-dominant solutions in the fronts exceeds the population size N then the crowded comparison operator is used to reject the solutions from the last front. The crowded distance operator is used to preserve the diversity of the solutions. The above procedure is repeated till the stopping generation is reached. Fig. 5.5 gives the block diagram of the working of the multi-objective optimization framework.

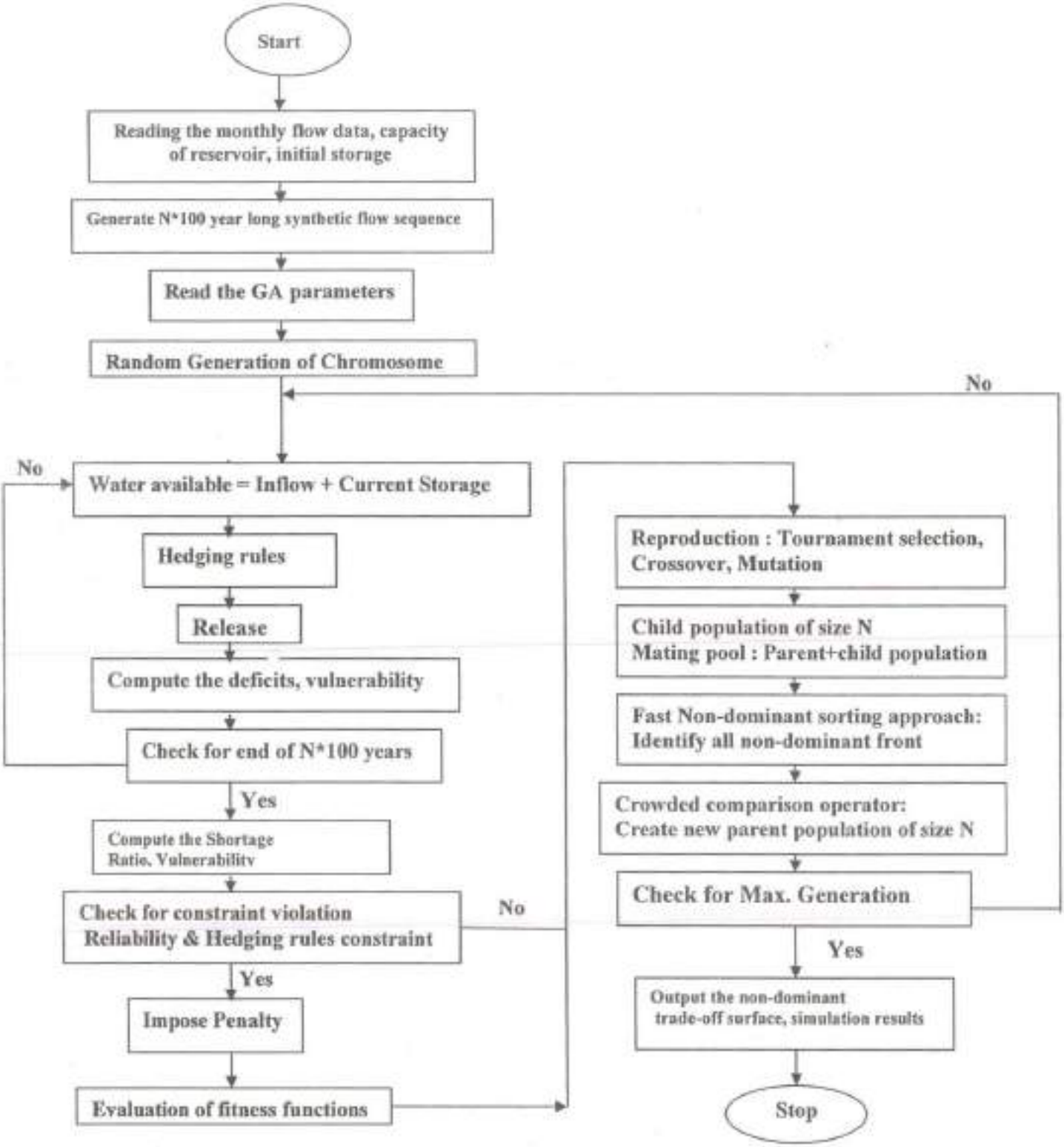


Fig 5.5: Block diagram of the Multi-objective Optimization Framework

5.6 CASE EXAMPLE – DHAROI RESERVOIR

For the purpose of evaluating the storage performance of the alternative hedging policies, one long synthetic sequence of monthly inflows of 4100 years similar to the historical sequence of 41 years of monthly flow data measured near Dharoi reservoir site, India, has been generated using the Perturbed Matched Block Bootstrap (PMABB) proposed in the third chapter of this report. The monthly streamflow data, model structure identification, estimation/selection of parameters, generation of data, verification and validation exercises have already been discussed as part of Chapter 4 of this report. The performance of the four different hedging rules is evaluated by routing the synthetic flows through a single hypothetical water supply reservoir (Dharoi) using the synthetic streamflows from PMABB. The multi-objective genetic algorithm described earlier is linked to this simulation program for evaluating the fitness functions and checking if the constraints are satisfied. A typical simulation based output obtained by routing the historical flows through the Dharoi reservoir is shown in Appendix-2. This work assumes two different yield levels, namely, 75% and 85% MAF. Also, two formulations, one unconstrained and another constrained have been run and the results have been processed and tabulated. These will be discussed in a later section.

Sensitivity Analysis

A brief sensitivity analysis of the various parameters of the multi-objective genetic algorithm was done. Based on these results, one set of parameters is adopted for all the four hedging cases for both unconstrained and constrained formulations. This set of parameters is given in Table 5.1.

Table 5.1 Parameters of NSGA-II

S.No.	Parameters of NSGA-II	Two- point Hedging Rule	Modified Two-point Hedging Rule	Discrete Hedging Rule	Proposed Hedging Rule
1	Crossover Probability	0.70	0.70	0.70	0.70
2	Mutation Probability	0.010	0.010	0.010	0.010
3	Random Seed	0.40	0.40	0.40	0.40
4	Population	100	100	100	100
5	Generation	300	300	300	300

5.7 RESULTS AND DISCUSSION

5.7.1 Discussion of Results - Unconstrained Formulation

The non-dominant fronts that indicate the trade-off between the period vulnerability and the shortage ratio, obtained from the four different cases of hedging corresponding to the unconstrained formulation are presented in Figs. 5.6 and 5.7 respectively for 75% MAF and 85% MAF yields.

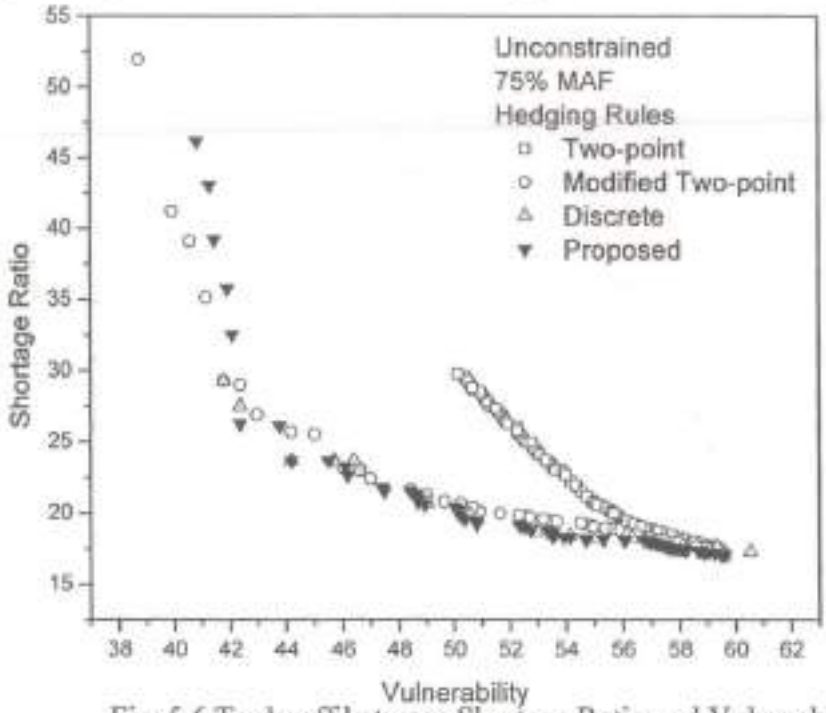


Fig. 5.6 Trade-off between Shortage Ratio and Vulnerability - Unconstrained case; Yield = 75% MAF

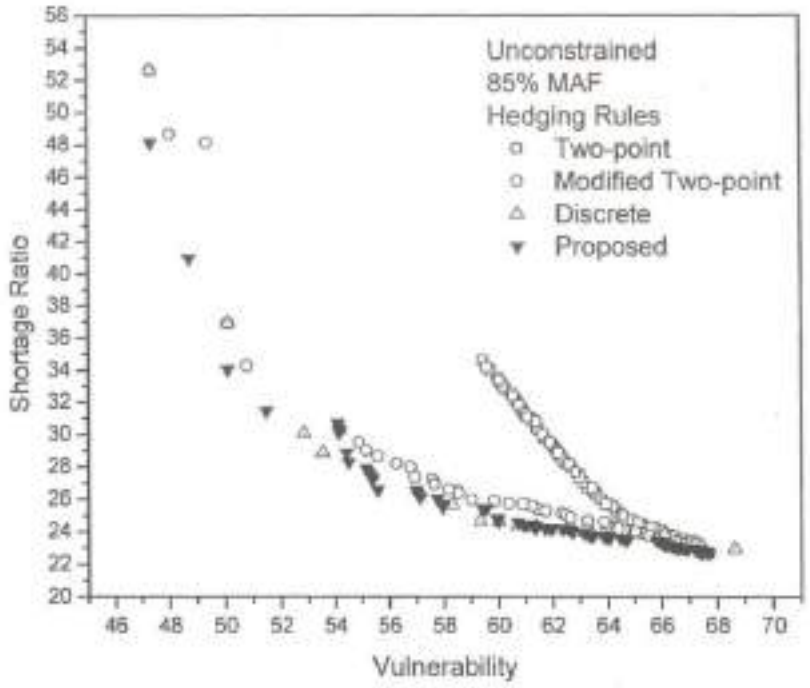


Fig 5.7: Trade-off between Shortage Ratio and Vulnerability - Unconstrained case; Yield = 85% MAF

It is observed from Fig. 5.6 that the trade-off curve obtained using the two-point hedging rule does not yield solutions in the domain of low vulnerability, although on the other extreme, it is able to yield a solution with the least SR. The non-dominant front obtained in this case is apparently far from pareto-optimality. This is because the flexibility offered by this hedging rule is limited due to the constant ratio of the release of the water availability throughout the hedging period Fig.5.1 the same behavior is also portrayed by Fig. 5.7., which also refers to the unconstrained case but with a higher yield of 85% MAF.

A further observation from Fig. 5.6 is that the discrete hedging rule yield very less number of non-dominant solutions in the low vulnerability domain, thus offering less flexibility for decision-making. However, this rule also performs well in the low

shortage ratio region. On the other hand, the modified two-point hedging rule and the proposed hedging policy offer a good spread of non-dominant solutions in the low vulnerability as well as the low shortage ratio domains. On the low vulnerability side the modified two-point hedging rule is seen to perform slightly better, while the proposed hedging rule performs marginally better in the low shortage ratio domain. However, the coverage and the density of non-dominant solutions obtained are somewhat better in case of the proposed hedging rule. Similar behavior of these two hedging rules is also observed in case of 85% yield (Fig. 5.7.), except that the proposed hedging rule performs better in the low vulnerability region as well.

More detailed results obtained from the simulation of the three selected hedging policies (one from either extreme of the non-dominant front and one compromising policy based on nearness to the origin on the multi-objective space) for all the four cases of hedging, are shown in Table 5.4 and 5.5. Also, the corresponding decision vectors and the evaluated objective function values of these three selected hedging policies for all the four cases of hedging are presented in Tables 5.2 and 5.3 respectively.

Table 5.2 Hedging Parameters for three selected Non-dominated solutions – Yield = 75%; Case: Unconstrained; Reservoir: Dharoi

Hedging Policy and Objectives	Decision Vector (Hedging Parameters)							Period Vul. (Mm^3)	SR (%)
2 Point Hedging	**α**	**β**							
Obj.1: Min. (Vul)	0.15	1.00						50.156	29.789
Compromise Policy	0.12	0.82						51.937	26.285
Obj.2: Min. (SR)	0.23	0.00						59.638	17.049
Modified 2-point Hedging	**α**	**β**	**HF**						
Obj.1: Min. (Vul)	0.31	0.96	0.64					38.753	51.945
Compromise Policy	0.20	0.46	0.71					42.992	26.876
Obj.2: Min. (SR)	0.99	0.58	0.00					59.638	17.049
Discrete Hedging	**α_1**	**α_2**	**k_1**	**k_2**	**k_3**				
Obj.1: Min. (Vul)	0.31	0.01	0.12	0.57	0.52			41.781	29.253
Compromise Policy	0.27	0.22	0.25	0.3	0.36			44.203	23.664
Obj.2: Min. (SR)	0.31	0.12	0.13	0.32	0.00			60.552	17.300
Proposed Hedging	**α_1**	**α_2**	**m_1**	**m_2**	**k_1**	**k_2**	**k_3**		
Obj.1: Min. (Vul)	0.01	0.31	0.19	0.01	0.21	0.32	0.91	40.838	46.130
Compromise Policy	0.01	0.30	0.22	0.00	0.05	0.32	0.45	42.386	26.272
Obj.2: Min. (SR)	0.67	0.67	0.95	0.58	0.68	1.00	0.00	59.617	17.054

Vul: Period Vulnerability (Mm^3); SR: Shortage Ratio (%)

Table 5.3 Hedging Parameters for three selected Non-dominated solutions – Yield = 85%; Case: Unconstrained; Reservoir: Dharoi

Hedging Policy and Objectives	Decision Vector (Hedging Parameters)							Period Vul. (Mm^3)	SR (%)
2 Point Hedging	α	β							
Obj.1: Min. (Vul)	0.11	1.00						59.444	34.584
Compromise Policy	0.08	0.72						61.691	29.794
Obj.2: Min. (SR)	0.23	0.00						67.710	22.660
Modified 2-point Hedging	α	β	**HF**						
Obj.1: Min. (Vul)	0.70	0.88	0.69					47.351	52.619
Compromise Policy	0.93	0.48	0.74					50.782	34.270
Obj.2: Min. (SR)	0.99	0.58	0.00					67.710	22.660
Discrete Hedging	α_1	α_2	k_1	k_2	k_3				
Obj.1: Min. (Vul)	0.31	0.09	0.15	0.93	0.88			47.351	52.619
Compromise Policy	0.23	0.08	0.13	0.24	0.36			52.841	30.023
Obj.2: Min. (SR)	0.79	0.27	0.27	0.80	0.00			68.625	22.919
Proposed Hedging	α_1	α_2	m_1	m_2	k_1	k_2	k_3		
Obj.1: Min. (Vul)	0.08	0.31	0.75	0.00	0.16	0.40	0.80	47.351	48.155
Compromise Policy	0.13	0.25	0.95	0.00	0.18	0.51	0.41	51.469	31.436
Obj.2: Min. (SR)	0.71	0.71	0.95	0.62	0.71	0.90	0.00	67.710	22.670

Vul: Period Vulnerability (Mm^3); SR: Shortage Ratio (%)

Table 5.4 Comparison of Dharoi Reservoir Performance for SOP and Four Hedging Policies for three selected Non-dominant solutions; Demand – 75%; Case: Unconstrained

Policy	Non-dominant Solutions	Reservoir Performance Indicators*						
		OBR	VBR	Res.	Period Vul. (Mm^3)	Event Vul. (Mm^3)	Avg. Period Def	Avg. Event Def
SOP		0.8098	0.8295	0.1661	1087.50	59.64	292.59	48.61
Two point Hedging	Obj.1: Min. (Vul)	0.1088	0.7022	0.0505	4688.30	50.16	358.73	18.13
	Compromise Policy	0.2503	0.7372	0.0688	4636.90	51.94	276.37	19.02
	Obj.2: Min. (SR)	0.8098	0.8295	0.1661	1087.50	59.64	292.59	48.61
Modified Two-point Hedging	Obj.1: Min. (Vul)	0.2299	0.4806	0.0910	3662.10	38.75	402.13	36.58
	Compromise Policy	0.6409	0.7313	0.1737	3277.60	42.99	233.70	40.59
	Obj.2: Min. (SR)	0.8098	0.8295	0.1661	1087.50	59.64	292.59	48.61
Discrete Hedging	Obj.1: Min. (Vul)	0.5974	0.7075	0.1707	3266.70	41.78	230.82	39.41
	Compromise Policy	0.6932	0.7634	0.1654	2232.80	44.20	252.92	41.84
	Obj.2: Min. (SR)	0.8170	0.8270	0.1808	1036.90	60.55	283.58	51.28
Proposed Hedging	Obj.1: Min. (Vul)	0.2515	0.5388	0.0896	3874.30	40.84	373.03	33.43
	Compromise Policy	0.6434	0.7374	0.1711	3231.40	42.39	233.51	39.96
	Obj.2: Min. (SR)	0.8098	0.8295	0.1662	1087.50	59.62	292.67	48.63

* The reservoir performance indicators have been computed using 4100-year long monthly synthetic streamflow sequence generated from PMABB model

Table 5.5 Comparison of Dharoi Reservoir Performance for SOP and Four Hedging Policies for three selected Non-dominant solutions; Demand – 85%; Case: Unconstrained

Policy	Non-dominant Solutions	Reservoir Performance Indicators*						
		OBR	VBR	Res.	Period Vul. (Mm^3)	Event Vul. (Mm^3)	Avg. Period Def	Avg. Event Def
SOP		0.7507	0.7734	0.1579	1266.30	67.71	354.11	55.90
Two point Hedging	Obj.1: Min. (Vul)	0.1012	0.6541	0.0468	5857.70	59.44	505.95	23.66
	Compromise Policy	0.2805	0.7020	0.0721	5777.20	61.69	353.21	25.46
	Obj.2: Min. (SR)	0.7507	0.7734	0.1579	1266.30	67.71	354.11	55.90
Modified Two-point Hedging	Obj.1: Min. (Vul)	0.2770	0.4737	0.1008	4474.60	47.35	444.21	44.76
	Compromise Policy	0.5603	0.6573	0.1670	3971.30	50.78	286.92	47.91
	Obj.2: Min. (SR)	0.7507	0.7734	0.1579	1266.30	67.71	354.11	55.90
Discrete Hedging	Obj.1: Min. (Vul)	0.2770	0.4737	0.1008	4474.60	47.35	444.21	44.76
	Compromise Policy	0.6311	0.6997	0.1697	2669.10	52.84	294.91	50.05
	Obj.2: Min. (SR)	0.7574	0.7708	0.1656	1266.00	68.63	350.92	58.10
Proposed Hedging	Obj.1: Min. (Vul)	0.3377	0.5184	0.1089	4428.00	47.35	410.63	44.71
	Compromise Policy	0.6022	0.6856	0.1654	3587.90	51.47	293.84	48.60
	Obj.2: Min. (SR)	0.7509	0.7733	0.1581	1266.30	67.71	354.07	55.96

* The reservoir performance indicators have been computed using 4100-year long monthly synthetic streamflow sequence generated from PMABB model

5.7.2 Discussion of Results - Constrained Formulation

After carefully analyzing the simulated performance results of the unconstrained formulation, it was decided to make some more runs with a view to converge to a narrow domain in the decision space to enable selection of a compromising policy for implementation. The constraints of the formulation were activated in such a way that there should be sufficient flexibility while searching for the non-dominant solutions, and at the same time convergence to a narrower domain should be achieved. For the 75% MAF yield scenario, after a careful inspection of Table 5.2 (that gives the range of performance evaluation of the non-dominant solutions for the unconstrained formulation), the following lower/upper limits were introduced for three of the performance indicators (other than shortage ratio and period vulnerability):

Occurrence Reliability	≥ 0.60
Mean Period Deficit	≤ 45.0 Mm^3
Mean Event Deficit	≤ 300.0 Mm^3

The decisions regarding the constraints to be activated and the limits to be plugged in for the same were done after sufficient trials. The results of this exercise are presented in Fig. 5.8, Tables 5.6 and 5.7.

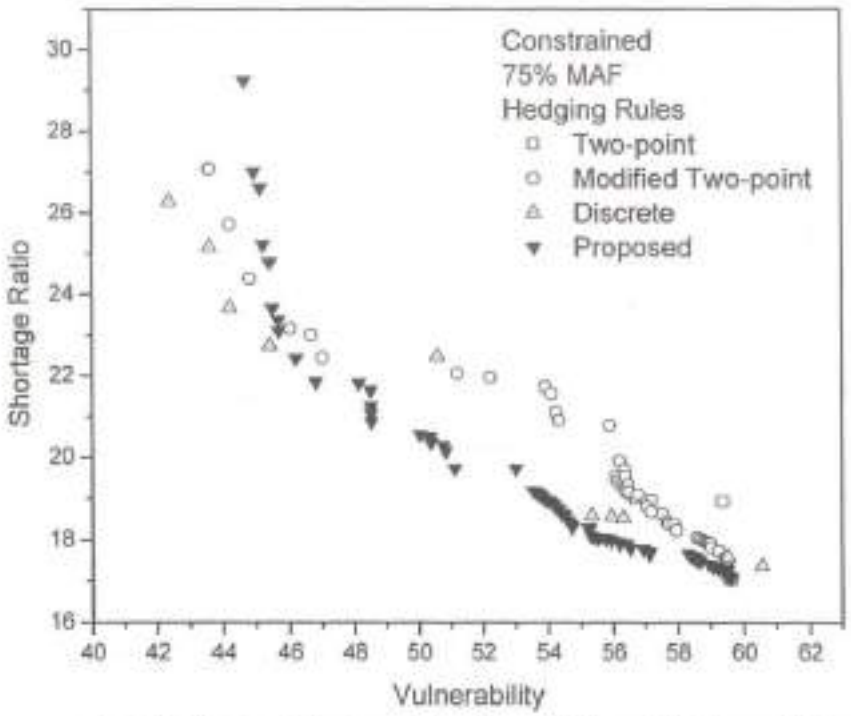


Fig 5.8 Trade-off between Shortage Ratio and Vulnerability - Constrained case; Yield = 75% MAF

Table 5.6 Hedging Parameters for three selected Non-dominated solutions – Yield = 75%; Case: Constrained; Reservoir: Dharoi

Hedging Policy and Objectives	Decision Vector (Hedging Parameters)							Period Vul. (Mm^3)	SR (%)
2 Point Hedging	**α**	**β**							
Obj.1: Min. (Vul)	0.05	0.38						56.075	19.527
Compromise Policy	0.05	0.38						56.075	19.527
Obj.2: Min. (SR)	0.95	0.03						59.638	17.054
Modified 2-point Hedging	**α**	**β**	**HF**						
Obj.1: Min. (Vul)	0.94	0.46	0.72					43.597	27.077
Compromise Policy	0.90	0.38	0.74					44.809	24.387
Obj.2: Min. (SR)	0.95	0.04	0.01					59.638	17.052
Discrete Hedging	**α_1**	**α_2**	**k_1**	**k_2**	**k_3**				
Obj.1: Min. (Vul)	0.3	0.06	0.08	0.37	0.45			42.386	26.272
Compromise Policy	0.3	0.06	0.08	0.37	0.45			42.386	26.272
Obj.2: Min. (SR)	0.91	0.34	0.34	0.92	0.05			60.552	17.348
Proposed Hedging	**α_1**	**α_2**	**m_1**	**m_2**	**k_1**	**k_2**	**k_3**		
Obj.1: Min. (Vul)	0.08	0.23	0.80	0.01	0.16	0.24	0.51	44.679	29.238
Compromise Policy	0.00	0.23	0.80	0.01	0.16	0.24	0.34	45.684	23.089
Obj.2: Min. (SR)	0.27	0.79	1.00	0.31	0.27	0.8	0.51	59.617	17.075

Vul: Period Vulnerability (Mm^3); SR: Shortage Ratio (%)

Table 5.7 Comparison of Dharoi Reservoir Performance for SOP and Four Hedging Policies for three selected Non-dominant solutions; Demand – 75%; Case: Constrained

Policy	Non-dominant Solutions	Reservoir Performance Indicators*						
		OBR	VBR	Res.	Period Vul. (Mm^3)	Event Vul. (Mm^3)	Avg. Period Def	Avg. Event Def
SOP		0.8098	0.8295	0.1661	1087.50	59.64	292.59	48.61
Two point Hedging	Obj.1: Min. (Vul)	0.6042	0.8048	0.1027	4211.70	56.08	260.54	26.76
	Compromise Policy	0.6042	0.8048	0.1027	4211.70	56.08	260.54	26.76
	Obj.2: Min. (SR)	0.7925	0.8295	0.1551	1602.50	59.64	287.30	44.57
Modified Two-point Hedging	Obj.1: Min. (Vul)	0.6437	0.7293	0.1763	2925.30	43.60	233.84	41.22
	Compromise Policy	0.6883	0.7562	0.1705	1752.50	44.81	248.94	42.44
	Obj.2: Min. (SR)	0.7857	0.8295	0.1483	1602.90	59.64	290.95	43.15
Discrete Hedging	Obj.1: Min. (Vul)	0.6434	0.7374	0.1711	3231.40	42.39	233.51	39.96
	Compromise Policy	0.6434	0.7374	0.1711	3231.40	42.39	233.51	39.96
	Obj.2: Min. (SR)	0.7864	0.8265	0.1511	1600.90	60.55	291.53	44.06
Proposed Hedging	Obj.1: Min. (Vul)	0.6097	0.7077	0.1753	3116.90	44.68	231.73	40.63
	Compromise Policy	0.7014	0.7691	0.1612	1750.50	45.68	260.20	41.94
	Obj.2: Min. (SR)	0.7911	0.8293	0.1570	1447.70	59.62	282.49	44.34

* The reservoir performance indicators have been computed using 4100-year long monthly synthetic streamflow sequence generated from PMABB model

The comparison of the non-dominant fronts presented for the four cases of hedging (Fig 5.8) shows that:

i) The two-point hedging rule does not yield any solution in the intermediate and lower ranges of vulnerability, and in general, performs poorly.

ii) The modified two-point hedging rule yields some good non-dominant solutions in the lower vulnerability region, but results in higher shortage ratios in most parts of the intermediate and higher vulnerability ranges.

iii) The discrete hedging rule yields limited number of non-dominant solutions compared with the modified two-point hedging rule and the proposed hedging rule. Although this is yielding some very good solutions near the lower vulnerability region, only a single solution is obtained in the entire intermediate ranges of shortage ratio and vulnerability (which form the major part of the compromising domain). Thus, this rule, when implemented into the multi-objective optimization formulation, does not seem to yield sufficient number of trade-off solutions to facilitate the decision-making process from a practical stand-point.

iv) The proposed hedging policy, is able to yield sufficient number of non-dominant solutions when compared with all the other three rules and provides very good non-dominant solutions near the minimum shortage ratio domain. However, the non-dominant solutions obtained for the proposed hedging rule, although well spread and are sufficient in number that offer the flexibility for decision-making its pareto-optimality near the low vulnerability region, is somewhat poorer than the other two competing rules

5.8 SUMMARY & CONCLUSIONS

A multi-objective optimization framework has been developed to evaluate optimal hedging rules. This employs NSGA-II, an efficient multi-objective genetic algorithm technique that can handle constrained formulations. The evaluation of the non-dominant solutions on the trade-off surface between the conflicting objectives of minimization of vulnerability and minimization of shortage ratio

helps us to compare the performance of the different hedging rules under water shortage conditions.

A new hedging rule has been proposed in the present study which is more generic than the discrete hedging rule and the modified two-point hedging rule. From the results of the optimal hedging studies done using the Dharoi reservoir data, the proposed hedging rule is shown to produce efficient non-dominant fronts containing well-distributed non-dominant solutions. The compromising hedging policies obtained using the proposed rule is shown to yield a number of trade-off solutions that exhibit good performance with regard to the different reservoir storage performance indicators.

CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 Hybrid Non-linear Data-driven Model for Annual Streamflows

A hybrid model that blends the two non-linear data-driven models, ANN (deterministic) and MBB (stochastic) is proposed for modeling annual streamflows of rivers that exhibit complex dependence. First, a nonlinear deterministic model, ANN (radial basis function network) is fitted to the historical annual streamflows, which captures the nonlinear trend in the data effectively. Then, the resulting residuals from the ANN model are resampled using a non-parametric resampling technique, moving block bootstrap with a view to capture the weak linear as well as the nonlinear dependence and any distributional information retained in the residuals. The proposed model has been applied to three annual streamflow data sets that exhibit complex dependence, drawn from different geographic regions with varying record lengths. The effective blending of the two data-driven models is shown to result in efficient simulations of the long-term storage and drought-related characteristics.

The ANN based Hybrid Model (ANNHM), being a completely data-driven model, reproduces the features of the marginal distribution more closely compared to Linear Parametric based Hybrid Model (LPHM), but offers less smoothing and little extrapolation value. However, the linear dependence structure is better reproduced by LPHM than ANNHM.

Despite a better preservation of the linear dependence structure, LPHM does not seem to effectively predict the variation of critical drought duration with respect to

truncation level. On the contrary, ANNHM is able to model the variation of critical drought duration better, even though the preservation of linear dependence structure is inferior to LPHM. This is plausibly due to the effective blending of the two nonlinear models. Also, the mean drought characteristics are more efficiently modeled by ANNHM.

The relative bias in predicting the reservoir storage statistics at lower demand levels is found to be high in case of LPHM. Moreover, a large spread of the same is observed at all demand levels, thus increasing the relative RMSE significantly compared with ANNHM.

Future research should address the extension of the proposed ANN-based hybrid model to single-site and multi-site modeling of periodic stream flows. Also, different hybrids could be tried and some smoothing can be introduced to get better variety and variability of the generated flows and the predicted water use characteristics thereof.

6.2 Periodic Stochastic Models for Monthly Streamflows

A new nonparametric method of conditional bootstrap is presented for simulating multi-season hydrologic time series. It resamples non-overlapping within-year blocks of hydrologic data (formed from the observed time series) using the rank matching rule of *Carlstein et al.* [1998]. This algorithm searches the historical record to find neighbouring blocks whose ends closely match the end element of the current block and subsequently resamples their successor blocks. The resampled blocks are perturbed using a weighted smoothing strategy with a window size of 12 months to achieve smoothing and extrapolation in simulations.

The proposed method, namely, perturbed matched-block bootstrap (PMABB), is shown to be efficient in reproducing a wide variety of statistical attributes for both hypothetical and real data sets. The verification and validation results presented here support PMABB as a plausibly better alternative to the non-parametric method "k-nearest neighbour bootstrap of Lall and Sharma (1996)" and the hybrid periodic model HMBB *of Srinivas and Srinivasan* [2001a,b] in simulating periodic streamflows. It is believed that PMABB can provide a rather flexible and adaptive method for simulating time series at finer time scales (e.g., weekly, daily and hourly), where there is progressively more structure to exploit.

The method provides simulations that are efficient in reproducing summary statistics, dependence structure and the salient features of the marginal distribution, without compromising on smoothing, extrapolation and variety in simulations. As a result, better prediction of storage capacity and critical run characteristics for water resources design is achieved.

6.3 Storage-Performance-Yield (S-P-Y) Relationships for Reservoirs

Over-year Reservoirs

Following behaviour analysis based on stochastic simulation, S-P-Y relationships have been constructed for use in planning and design applications of over-year water supply reservoirs, using standard operating policy. These relationships are useful in: (i) gaining an understanding of the variation of reservoir performance indicators namely, reliability, resilience, and vulnerability on the storage-yield plane; (ii) identifying the sensitive ranges of storage capacity of the over-year reservoirs, with regard to

performance characteristics; and (iii) selecting between capacity expansion and demand management options, in case of deficit water supply systems.

For highly over-year reservoirs, resilience and vulnerability do not seem to improve with increase in storage capacity, and hence, the decision regarding storage capacity depends primarily on reliability improvement. Approximate ranges of over-year storage capacity of water supply reservoirs, required to meet various target yields with a desirable range of reliability (0.95 to 0.99) are presented for a few selected cases of C_v and ρ_1 of annual streamflows. For lower target yields, there is no significant improvement in resilience at low as well as high storage capacities, while a significant improvement is noted for the range in between. With increase in target yield, this transition range widens and moves towards higher storage capacities, eventually ending up in flat storage-resilience relationships for high target yields. In addition, appreciable reduction in vulnerability is noted in a certain range of storage capacity and this range widens and moves towards higher storage capacities with increase in target yield and/or C_v. This can be exploited in certain water supply systems, wherein the marginal value of either increase in resilience or decrease in vulnerability or both, is quite high. A S-P-Y database with a search facility has also been developed that would help in planning and design of reservoir capacity and in decisions regarding capacity expansion or implementation of demand management programs.

Within-year Reservoirs

For the Dharoi reservoir streamflows, the reservoir storage performance measures have been computed from the 41-year long historical flow sequence as well as the 4100-year long synthetic flow sequences generated from the three stochastic streamflow models

considered (k-NN, HMBB and PMABB), for seven combinations of storage capacity (including the existing capacity) and ten combinations of yield (expressed in % of Mean Annual Flow). The results and investigation are presented only for the PMABB model, although the HMBB model yields competitive performance.

The reservoir performance-yield relationships obtained are presented for the existing reservoir storage capacity of 732 Mm^3 of Dharoi reservoir, using the historical flow sequence as well as the three long synthetic sequences generated from the three periodic stochastic models considered. The actual live storage capacity of the Dharoi reservoir (732 Mm^3) would yield a volume reliability of nearly 83% for 75% yield, while it would decrease to a value of 77.3% for 85% yield. If the same volume reliability of 83% is to maintained for an increased yield of 85%, than the storage capacity is to be increased to 1000 Mm^3, which will require a reservoir storage capacity expansion.

The decrease in resilience with increase in yield is observed to be small. Likewise, the increase in the resilience with increase in the storage capacity (for the given yield), is also negligible, the reason being: the decrease in the number of events (decrease in the value of the numerator) is nearly proportional to the decrease in the number of failure periods (decrease in the value of the denominator).

6.4 Optimal Hedging Rules for Water Supply Reservoirs

A multi-objective optimization framework has been developed to evaluate optimal hedging rules. This employs NSGA-II, an efficient multi-objective genetic algorithm technique that can handle constrained formulations. The

evaluation of the non-dominant solutions on the trade-off surface between the conflicting objectives of minimization of vulnerability and minimization of shortage ratio helps us to compare the performance of the different hedging rules under water shortage conditions.

A new hedging rule has been proposed in the present study which is more generic than the discrete hedging rule and the modified two-point hedging rule. From the results of the optimal hedging studies done using the Dharoi reservoir data, the proposed hedging rule is shown to produce efficient non-dominant fronts containing well-distributed non-dominant solutions. The results are found to be consistent for moderate as well as highly critical water supply conditions. The compromising hedging policies obtained using the proposed rule is shown to yield a good number of trade-off solutions that exhibit good compromising performance with regard to the different reservoir storage performance indicators.

REFERENCES

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a), "Artificial neural networks in hydrology I: Preliminary concepts", *J. Hydrol. Engg.*, ASCE, 5 (2), 115-123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b), "Artificial neural networks in hydrology II: Hydrologic applications" *J. Hydrol. Engg.*, ASCE, 5 (2), 124-137.

Bayazit, M.., (1982), "Ideal Reservoir Capacity as a Function of Yield and Risk", *Journal of Hydrology*, **58**,1-9.

Bendat, J. S. and A. G. Piersol (1986), *Random Data: Analysis and Measurement Procedures*, John Wiley, New York.

Bose, A. (1988) "Edgeworth corrections by bootstrap in autoregressions". *Annals of Statistics*, 16, 1709-1722.

Bowman, A. W. and A. Azzalini (1997), Applied Smoothing Techniques for Data Analysis, Clarendon press, Oxford.

Bras, R. L. and I. Rodriguez-Iturbe (1985), *Random Functions and Hydrology*, Addison-Wesley, Reading, Mass.

Buchberger, S.G., and Maidment. D.R.(1989), "Diffusion approximation for equilibrium distribution of reservoir storage", Water Resources Research, **25**(7), 1643-1652.

Burges, S. J., and Linsley, R. K. (1971), "Some Factors Influencing Required Reservoir Storage", Journal of Hydraulic Engineering Division, ASCE, **97**(HY7), 977-991.

Burn, D.H., Venema, H.D., and Simonovic, S.P.(1991), "Risk-based performance criteria for real-time reservoir operation", Canadian Journal of Civil Engineering, **18,** 36-42.

Buscema M., and Sacco P. L. (2000). "Feedforward networks in financial predictions: The future that modifies the present", *Expert Systems*, 17(3): 149-169.

Campolo, M., Andreussi, P., and Soldati, A. (1999) "River flood forecasting with neural network model", *Water Resources Research,* 35(4), 1191-1197.

Carlson, R. F., A. J. MacCormick, and D. G. Watts (1970) "Application of linear models to four annual streamflows series". *Water Resources Research*, 6, 1070-1078.

Carlstein, E. (1986) "The use of subseries values for estimating the variance of a general statistic from a stationary sequence". *Annals of Statistics*, 14, 1171-1179.

Carlstein, E., K-A. Do, P. Hall, T. Hesterberg, and H. R., Matched-block bootstrap for dependent data, Bernoulli, 4(3), 305-328, 1998.

Cayan, D. and L. Riddle (1992) Atmospheric circulation and precipitation in the Sierra Nevada, Managing water resources during global change, paper presented at *Conference of American Water Resources Association*, Tucson, Ariz., 1992.

Chen, S., Cowan, C. F. N., and Grant, P. M. (1991). "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks." *IEEE Transactions on Neural Networks*, 2(2), 302-309.

CWC (2007). Weekly report of 81 important reservoirs of India, Central Water Commission, Government of India.

Daniel, T. M. (1991). Neural networks—applications in hydrology and water resources engineering. *Proc., Int. Hydrol. and Water Resour. Symp.*, Institution of Engineers, Perth, Australia.

Datta, B and S.J. Burges. (1984), Short – term Single Multipurpose Reservoir Operation; Importance of Loss Functions and Forecast Errors, Water Resource Research, pp 1167- 1176.

Davison, A. C. and D. V. Hinkley *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge, 1997.

Dawson, D.W. and Wilby, R. (1998), An artificial neural network approach to rainfall-runoff modeling, *Hydrol. Sci. J.*, 43 (1), 47-65.

Dawson, C.W. and Wilby, R.L. (2001). Hydrological modeling using artificial neural networks. Progress in Physical Geography, 25, 80–108.

Deb, k., Agarwal, S., Pratap, A., and Meyarivan, T., (2002), "A fast and elitist multi-objective genetic algorithm: NSGA-II:, IEEE Transactions on Evolutionary Computations 6(2), 182-197.

DeRoach., J. N. (1989). Neural networks: An artificial intelligence approach to the analysis of clinical data, *Australas Phys. Eng. Sci. Med.*, 12: 100-106.

Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.

Efron, B. and R. J. Tibshirani (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statistical Sciences*, 1, 54-96.

Efron, B. and R. J. Tibshirani *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.

Fernando, A.K. and Jayawardena, A.W. (1998). Runoff forecasting using RBF networks with OLS algorithm. *Journal of Hydrologic Engineering*, ASCE, 3(3), 203-209.

Fiering, M. B., 1963, Statistical Analysis of the Reservoir Storage-Yield Relations, In: H. A.Thomas Jr. and R. P. Burden, (Ed), Operations Research in Water Quality Management. Harv. Water Resources Group, Cambridge, Mass.

Fiering, J. D. (1964) Multivariate technique for synthetic hydrology. *Journal of Hydraulic Division*, ASCE, 90, 43-60.

Fiering, M.B., 1967, Streamflow Synthesis, Harvard University Press, Cambridge, Mass.

Freedman, D. A. (1984) On bootstrapping two-stage least-squares estimates in stationary linear models. *Annals of Statistics*, 12, 827-842.

Freedman, D. A. and S. C. Peters (1984) Bootstrapping a regression equation: some empirical results. *Journal of the American Statistical Association*, 79, 97-106.

Fonseca C.M., Fleming P.J. (1995), An overview of evolutionary algorithms in multi-objective optimization, *Evolutionary Computation*, 3, 1, 1-16.

Galeati, G. (1990) A comparison of parametric and non-parametric methods for runoff forecasting. *Hydrological Sciences Journal*, 35(1), 79-94.

Gernoth, K. A., Clark, J. W., Prater, J. S., and Bohr, H. (1993). "Neural network models of nuclear systematics", *Physics Letters B*, 300: 1-7.

Gomide, F. L. S.: 1975, Range and Deficit Analysis using Markov Chains, Hydrology Paper, 79, Colorado State University, Fort Collins, Colorado.

Grygier, J. C. and J. R. Stedinger (1988) Condensed disaggregation procedures and conservation corrections for stochastic hydrology, *Water Resources Research*, 24(10), 1574-1584.

Grygier, J. C. and J. R. Stedinger (1990) SPIGOT, A synthetic streamflow generation software package. Technical description, version 2.5, Cornell University, Ithaca, New York.

Hall, P. (1985) Resampling a coverage pattern. *Stochastic Processes and their Applications*, 20, 231-246.

Haltiner, J. P. and J. D. Salas (1988) Development and testing of a multivariate seasonal ARMA(1,1) model, *J. Hydro.*, Amsterdam, The Netherlands, 104, 247-272.

Hardle, W. and A. W. Bowman (1988) Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83(401), 102-110.

Harms, A. A. and T. H. Campbell (1967) An extension to the Thomas-Fiering model for the sequential generation of streamflow, *Water Resources Research*, 3(3), 653-661.

Hashimoto, T., Stedinger, J.R. and Loucks, D.P. (1982) Reliability, resiliency and vulnerability crireria for water resource system performance evaluation, *Water Resources Research* **18** (1982) (1), pp. 14–20.

Hazen, A., 1914, Storage to be provided in impounding reservoirs for municipal water supply, *Trans. Am. Soc.Civ. Eng.*, 77, 1539-1640.

Helsel, D. R. and R. M. Hirsch *Statistical Methods in Water Resources*. Elsevier, New York, 1992.

Hesterberg, T., Matched-block bootstrap for long memory processes, Research Rep. No.66, Mathsoft Inc., Seattle, 1997.

Hipel, K. W., A. I. McLeod, and W. C. Lennox (1977) Advances in Box-Jenkins modeling- 1. Modeling Construction. *Water Resources Research*, 13(3), 567-575.

Hirsch, R. M. (1979) Synthetic hydrology and water supply reliability. *Water Resources Research*, 15(6), 1603-1615.

Hirsch, R. M., R. B. Alexander, and R. A. Smith (1991) Selection of methods for the detection and estimation of trends in water quality. *Water Resources Research*, 27(5), 803-813.

Hirsch, R. M., D. R. Helsel, T. A. Cohn, and E. J. Gilroy *Statistical Analysis of Hydrologic Data.* pp. 17.1-17.55. *In* D. R. Maidment *Handbook of Hydrology.* McGraw Hill Inc., New York, 1993.

Hjorth, J. S. U. *Computer Intensive Statistical Methods-Validation Model Selection and Bootstrap*, Chapman and Hall, New York, 1994.

Hsu K., Gupta V.H., Sorooshian S. (1995) "Artificial neural network modeling of the rainfall-runoff process", *Wat. Resour. Res.*, 31(10), 2517– 2530.

Hurst, H. E. (1951) Long-term storage capacity of reservoirs, *Trans. American Society of Civil Engineers*, 116, 776-808.

IMSL (1984) IMSL User's Manual. IMSL, Inc. Houston, Texas.

Karlsson, M. and S. Yakowitz (1987a) Nearest-Neighbor methods for nonparametric rainfall-runoff forecasting. *Water Resources Research*, 23(7), 1300-1308.

Karlsson, M. and S. Yakowitz (1987b) Rainfall-runoff forecasting methods, old and new. *Stochastic Hydrology and Hydraulics*, 1, 303-318.

Klemes, V., 1987, One hundred years of applied storage theory, *Water Resources Management*, **1**, 159-175.

Koutsoyiannis, D. (1992) A nonlinear disaggregation model with a reduced parameter set for simulation of hydrologic series. *Water Resources Research*, 28(12), 3175-3191.

Koutsoyiannis, D. and A. Manetas (1996) Simple disaggregation by accurate adjusting procedures. *Water Resources Research*, 32(7), 2105-2117.

Koutsoyiannis, D. (2000) A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water Resources Research*, 36(6), 1519-1533.

Kumar, D. N., U. Lall, and M. R. Peterson (2000) Multi-site disaggregation of monthly to daily streamflow. *Water Resources Research*, 36(7), 1823-1833.

Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17(3), 1217-1241.

Lahiri, S. N. (1993) On the moving block bootstrap under long range dependence. *Statistics and Probability Letters*, 18, 405-413.

Lahiri, S. N. (1995) On the asymptotic behaviour of the moving block bootstrap for normalized sums of heavy-tail random variables. *Annals of Statistics*, 23(4), 1331-1349.

Lall, U. (1995) Recent advances in nonparametric function estimation: Hydrologic applications, United States national report to international union of Geodesy and Geophysics 1991-1994, Reviews of Geophysics, 33, 1093-1102.

Lall, U. and K. Bosworth Multivariate Kernel Estimation of Functions of Space and Time Hydrologic data. In K. Hipel Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Kluwer, Waterloo, 1993.

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

Lall, U., B. Rajagopalan, and D. G. Tarboton (1993) A Nonparametric wet/dry spell model for daily precipitation, Utah Water Research Laboratory, Utah State University, USA.

Lall, U., T. Sangoyomi, and H. D. I. Abarbanel (1994) Nonlinear dynamics of the Great Salt Lake: Nonparametric Short Term Forecasting, Utah Water Research Laboratory, Utah State University, USA.

Lall, U., B. Rajagopalan, and D. G. Tarboton (1996) A nonparametric wet/dry spell model for resampling daily precipitation. *Water Resources Research*, 32(9), 2803-2823.

Lane, W. L. *Applied stochastic techniques* (LAST Computer Package): User Manual. Division of Planning Technical Services, United States Bureau of Reclamation, Denver, Colorado, 1979.

Lane, W. L., Corrected parameter estimates for disaggregation schemes. In V. P. Singh Statistical Analysis of Rainfall and Runoff, Water Resources Publications, Littleton, Colorado, 1982.

Lane, W. L. and D. K. Frevert *Applied stochastic techniques, user's manual, personal computer version 5.2*. Earth sciences division, U. S. Bureau of Reclamation, Denver, Colorado, 1990.

Leger, C., D. N. Politis, and J. P. Romano (1992) Bootstrap technology and applications. *Technometrics*, 34(4), 378-398.

LePage, R. and L. Billard *Exploring the limits of bootstrap*, John Wiley, New York, 1992.

Liu, R. Y. and K. Singh *Moving blocks jackknife and bootstrap capture weak dependence*. pp. 225-248. *In* R. Lepage and L. Billard (eds.) *Exploring the Limits of Bootstrap*. John Wiley, New York, 1992.

Loucks, D. P., J. R. Stedinger, and D. A. Haith *Water Resource Systems Planning and Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

Maass, A., M. M. Hufschmidt, R. Dorfman, H. A. Thomas, S. A. Marglin, and M. F. Gordon *Design of water-resource systems*. Macmillan, London, 1967.

Maier, H. R., and Dandy, G. C. (2000) "Neural networks for the prediction and forecasting of water resources variables: A review of modeling issues and application" *Environmental Modeling and Software*, 15: 101-124.

Mandelbrot, B. B. (1971) A fast fractional Gaussian noise generator, *Water Resources Research*, 7(3), 543-553.

Mandelbrot, B. B. and J. R. Wallis (1969a) Computer experiments with fractional Gaussian noises, 1, Averages and variances, *Water Resources Research*, 5(1), 228-241.

Mandelbrot, B. B. and J. R. Wallis (1969b) Computer experiments with fractional Gaussian noises, 2, Rescaled ranges and Spectra, *Water Resources Research*, 5(1), 242-259.

Mandelbrot, B. B. and J. R. Wallis (1969c) Computer experiments with fractional Gaussian noises, 3, Mathematical appendix, *Water Resources Research*, 5(1), 260-267.

Matalas, N. C. (1967) Mathematical assessment of synthetic hydrology. *Water Resources Research*, 3 (4), 937-945.

McLeod, T. A. and K. W. Hipel (1978) Simulation procedures for Box-Jenkins models. *Water Resources Research*, 14(5), 969-975.

McMahon, T. A. and R. G. Mein *River and Reservoir Yield*, Water Resources Publications, Littleton, Colorado, 1986.

Mejia, J. M. and J. Rousselle (1976) Disaggregation models in hydrology revisited, *Water Resources Research*, 13(2), 185-186.

Mejia, J. M., I. Rodriguez-Iturbe, and D. R. Dawdy (1972) Streamflow simulation, 2, The broken line process as a potential method for hydrologic simulation. *Water Resources Research*, 8(4), 931-941.

Montaseri, M., and Adeloye, A.J., 1999, Critical period of reservoir systems for planning purposes, *Journal of Hydrology*, **224**(3-4), 115-136

Moody, J. and C. Darken (1989). "Fast learning in networks of locally tuned processing units". Neural computation, 1: 281-294.

Moss, M. E. and G. D. Tasker (1991) An intercomparison of hydrological network-design technologies. *Hydrological Sciences Journal*, Oxford, U. K., 36(3), 209-221.

Moy, W. S., Cohon. J. L., and Revelle, C. S., 1986, A Programming Model for Analysis of Reliability, Resilience, and Vulnerability of a Water Supply Reservoir, *Water Resources Research*, **22**(4), 489-498.

Neelakantan, T. R. and Pundarikanthan, N. V.: 1999, 'Hedging rule optimization for water supply reservoir system', Water Resources Management, 13(6), 409–426.

NIH (1986), Hydrological year book Hemavathi subbasin year 1985-86, Report No. TR-53, National Institute of Hydrology, Roorkee.

NIH (1987), Storage Yield Analysis, Report No. UM-16, National Institute of Hydrology, Roorkee.

NIH (1989), Forecasting of Monsoon Runoff Using Data from Specific Basins, Report No. TR-43, National Institute of Hydrology, Roorkee.

NIH (1991), Application of HEC-1 to Hemavathi (up to Sakleshpur) Basin, Report No. CS-55, National Institute of Hydrology, Roorkee.

NIH (1996), Multi-objective Optimization of Operation of a Dam, Report No. TR(BR)-143, National Institute of Hydrology, Roorkee.

Pegram, G.G.S., Salas, J.D., Boes, D.C., and Yevjevich, V., 1980, Stochastic properties of water storage, Hydrology Paper 100, Colorado State University, Fort Collins, Colorado, USA.

Perrens, S. J., and Howell, D. T., 1972, Effects of Serially Correlated Inflows on Reservoir Behaviour, *Water Resources Research*, **8**(3), 642-647.

Phatarford, R.M., 1989, Riverflow and reservoir storage models, Math. Comput. Modelling, **12**(9), 1057-1077.

Philipose, M.C. and Srinivasan, K. (1995), "Construction of Storage-Performance-Yield Relationships for a Reservoir Using Stochastic Simulation", Water Resources Development, Vol. 11, No-3, 1995.

Powell, M. J. D. (1987). "Radial basis function approximations to polynomials", In proceedings of the 12th biennial numerical analysis conference (Dundee): 223-241.

Prairie, J.R. (2002). Long-Term Salinity Prediction With Uncertainty Analysis: Application For Colorado River Above Glenwood Springs, CO, M.S thesis, University of Colorado, Boulder, Colorado

Pretto, P.B., Chiew, F.H.S., McMahon, T.A., Vogel, R.M., and Stedinger, J.R., 1997, The (mis)behavior of behavior analysis storage estimates, *Water Resour. Res.*, **33**(4), 703-709.

Rajagopalan, B., U. Lall, and D. G. Tarboton (1993) Simulation of daily precipitation from a nonparametric renewal model, Utah State University.

Rajagopalan, B., U. Lall, and D. G. Tarboton (1994) A Nonparametric Renewal Model for Modeling Daily Precipitation. *In* K. Hipel (ed.) *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, Kluwer, Waterloo, Ontario, CA.

Rajagopalan, B., U. Lall, and D. G. Tarboton (1996) Nonhomogeneous Markov model for daily precipitation, *Journal of Hydrologic Engineering*, 1(1), 33-40.

Rajagopalan, B., Lall, U., Tarboton, D. G., and Bowles, D. S. (1997). Multivariate nonparametric resampling scheme for simulation of daily weather variables. Journal of Stochastic Hydrology and Hydraulics 11(1), 65-93.

Rajagopalan, B., and Lall, U. (1999). A k-Nearest neighbor simulator for daily precipitation and other weather variables. Water Resources Research 35(10), 3089-3101.

Rasmussen, P. F., J. D. Salas, L. Fagherazzi, J-C Rassam, and B. Bobee (1996) Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resources Research*, 32(10), 3151-3160.

Rivera, R. J. C. O., Bartual, G. and Andreu, J. (2002). "Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks". Hydrology and Earth System Sceinces, 6(4): 641-654.

Rippl, W., 1883, The Capacity of Storage Reservoirs for Water Supply, *Proceedings of the Institute of Civil Engineers*, London, **71,** 270-278.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (1980). Applied modeling of hydrologic time series. Water Resources Publications, Lttleton, Colarado.

Salas, J. D. and G. G. S. Pegram (1977) A seasonal multivariate multilag auto-regressive model in hydrology. Proceedings of the *Third International Symposium on Theoretical and Applied Hydrology*, Colorado State University, Fort Collins, Colorado.

Salas, J. D., D. C. Boes, and R. A. Smith (1982) Estimation of ARMA models with seasonal parameters, *Water Resources Research*, 18(4), 1006-1010.

Salas, J. D., Smith, R. A., and M. Markus (1992) Modeling and generation of univariate seasonal hydrologic data (Programs CSU001 and CSU002). Technical report 2, Colorado State Univ., Fort Collins, Colorado.

Sangoyomi, T. and U. Lall (1993) Nonlinear dynamics of the great salt lake: Insights from the 1848-1992 time series, Utah Water Research Laboratory, Utah State University.

Santos, E. G. and J. D. Salas (1992) Stepwise disaggregation scheme for synthetic hydrology, *Journal of Hydraulic Engineering*, 118(5), 765-784.

SAS/ETS (1988) SAS/ETS User's Guide, version 6, 1st ed., SAS Inc.,Cary, N. C.

Scott, D. W. *Multivariate density estimation: Theory, practice, and visualization*. John Wiley, New York, 1992.

Sharma, A. and U. Lall (1999) A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation*, 48, 361-371.

Sharma, A., D. G. Tarboton and U. Lall (1997) Streamflow simulation: A nonparametric approach. *Water Resources Research*, 33(2), 291-308.

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

Shih, J.S, Re Velle, C. (1994), "Water supply operations during drought: a discrete hedging rule" , European Journal of Operational Research, 82(1995), 163-175

Silverman, B. W. *Density estimation for statistics and data analysis*. Chapman and Hall, New York, 1986.

Simonovic, S. P., 1992, Reservoir Systems Analysis: Closing Gap Between Theory and Practice, *Journal of Water Resources Planning and Management*, **118**(3), 262-280.

Smith, J., G. N. day, and M. D. Kane (1992) Nonparametric framework for long range streamflow forecasting, *ASCE Journal of Water Resources Planning and Management*, 118(1), 82-92.

Smith, J. A. (1991) Long-range streamflow forecasting using nonparametric regression, *Water Resources Bulletin*, 27(1), 39-46.

Srinivasan, K. and Philipose M.C (1998), ' Effect of hedging on over year reservoir performance' , Water Resource Management, 12(2), 95-120.

Srinivasan, K., B. S. Thandaveswara, and P. C. Naik (1992) Monthly streamflow data generation – A case study. *Journal of Indian Water Resources Society*, 12(1&2), 29-38.

Srinivas, V. V., and Srinivasan, K. (2000). Post-blackening approach for modeling dependent annual streamflows. Journal of Hydrology 230 (1-2), 86-126.

Srinivas, V. V., and K. Srinivasan (2001a) Post-blackening approach for modeling periodic streamflows, J. Hydrol., 241 (3-4), 221-269.

Srinivas, V. V., and K. Srinivasan (2001b) A Hybrid stochastic model for multiseason streamflow simulation, Water Resour. Res., 37 (10), 2537-2549.

STATAGRAPHICS (1984) Statgraphics User Manual. Statistical Graphics Corp., Rockville, Md.

Stedinger, J. R. and M. R. Taylor (1982) Synthetic streamflow generation - 1. Model verification and validation. *Water Resources Research*, 18(4), 909-918.

Stedinger, J. R., and D. Pei, *An annual-monthly streamflow model for incorporating parameter uncertainty into reservoir simulation*. pp. 520-529. In A. H. El-

shaarawi and S. R. Esterby (eds.) *Time Series Methods in Hydroscience, Dev. Water Sci.*, vol. 17, Elsevier, New York, 1982.

Stedinger, J. R. and R. M. Vogel (1984) Disaggregation procedures for generating serially correlated flow vectors, *Water Resources Research*, 20(11), 47-56.

Stedinger, J. R., D. P. Lattenmaier, and R. M. Vogel (1985) Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. *Water Resources Research*, 21(4), 497-509.

Sudheer, K.P., , Srinivasan, K., Neelakantan, T. R and Srinivas, V.V. (2007). A nonlinear data-driven model for synthetic generation of annual streamflows. *Hydrological Processes* (In Press).

Sudheer, K. P., and Jain, S. K. (2003). "Radial basis function neural networks for modeling stage discharge relationship", *J. Hydrol. Engg., ASCE*, 8(3): 161-164.

Sudheer, K. P., Gosain, A. K., and Ramasatri, K. S. (2002). "Comparisons between Back Propagation and Radial Basis Function based Neural Networks in Rainfall-Runoff Modeling", Proceedings of the international conference on advances in civil engineering, Kharagpur, India, Vol 1: 449-456.

Sudheer, K. P., Nayak, P. C., and Ramasastri, K. S. (2003) Improving peak flow estimates in artificial neural network river flow models, *Hydrological Processes*, 17(3): 677-686.

Tarboton, D. G. (1994) "The source hydrology of severe sustained drought in the southwestern U.S.", *Journal of Hydrology*, 161, 31-69.

Tarboton, D. G., A. Sharma, and U. Lall (1993) "The use of non-parametric probability distributions in streamflow modeling", *Proceedings of the Sixth South African*

National Hydrological Symposium, ed. S. A. Lorentz et al., University of Natal, Pietermaritzburg, South Africa, September 8-10, 315-327.

Tarboton, D. G., A. Sharma, and U. Lall (1998)" Disaggregation procedures for stochastic hydrology based on nonparametric density estimation". *Water Resources Research*, 34(1), 107-119.

Tasker, G. D. (1987) "Comparison of methods for estimating low flow characteristics of streams". *Water Resources Bulletin*, 23(6), 1077-1083.

Tasker, G. D. and P. Dunne (1997) "Bootstrap position analysis for forecasting low flow frequency". *ASCE Journal of Water Resources Planning and Management*, 123(6), 359-367.

Thomas, H.A., and Burden, R.P., 1963, Operation research in water quality management, Harvard University Press, Cambridge, Mass.

Todini, E. (1980), "The preservation of skewness in linear disaggregation schemes". *Journal of Hydrology*, 47, 199-214.

Tong, H. *Nonlinear Time Series Analysis: A Dynamical Systems Perspective*. Academic, San Diego, California, 1990.

Troutman, B. M., 1978, Reservoir Storage with Dependent, Periodic Net Inputs, *Water Resources Research*, **14**(3), 395-401.

Valencia, D. R. and J. L. Schaake (1973) Disaggregation processes in stochastic hydrology. *Water Resources Research*, 9(3), 580-585.

Vecchia, A. V., J. T. B. Obeysekera, J. D. Salas, and D. C. Boes (1983) Aggregation and Estimation for low-order periodic ARMA models, *Water Resources Research*, 19(5), 1297-1306.

Velleman, P. F., and D. C. Hoaglin, Applications, Basics, and Computing of Exploratory Data Analysis, Duxbury Press, Boston, MA, 1981.

Vogel, R. M. and A. L. Shallcross (1996) The moving blocks bootstrap versus parametric time series models. *Water Resources Research*, 32(6), 1875-1882.

Vogel, R. M., and Stedinger, J. R., 1987, Generalized Storage - Reliability - Yield Relationships, *Journal of Hydrology*, **89,** 303-327.

Vogel, R.M., 1985, The variability of reservoir storage estimates, PhD thesis, Cornell University, Ithaca, N.Y.

Vogel, R.M., Fennessey, N.M., and Bolognese, R.A., 1995, Storage-reliability-resilience-yield relations for northeastern United States, *Jl. of watre resources planning and management*, **121**(5), 365-374.

Vogel, R.M., and McMahon, T.A., 1996, Approximate reliability and resilience indices of over-year reservoirs fed by AR(1) gamma and normal flows, *Hydrological Sciences Jl.*, **41**(1), 75-96.

Vogel, R.M., and Stedinger, J.R., 1988, The value of stochastic streamflow models in overyear reservoir design applications, *Water Resour. Res.*, **245**(9), 1483-1490.

Weeraratne, J.R., Logan, L., and Unny, T.E., 1986, Performance evaluation of alternate policies on reservoir system operation, *Canadian Jl. of Civil Engineering*, **13**, 203-212.

Woo, M. K. (1989) Confidence intervals of optimal risk-based hydraulic design parameters. *Canadian Water Resources Journal*, 14(2), 10-16.

WRDO *Master plan for equitable use of water resources in the Cauvery river basin – An outline*. Water Resources Development Organisation, Government of Karnataka, 13 Vols., 1976.

Yakowitz, S. (1987) Nearest-Neighbor methods for time series analysis. *Journal of Time series Analysis*, 8(2), 235-247.

Yakowitz, S. (1993) Nearest-neighbor regression estimation for null-recurrent Markov time series, *Stochastic Processes Their Applications*, 48, 311-318.

Yakowitz, S. and M. Karlsson Nearest-neighbor methods with application to rainfall/runoff prediction, in *Stochastic Hydrology*, edited by J. B. Macneil and G. J. Humphries, pp.149-160, D. Reidel, Norwell, Mass., 1987.

Yevjevich, V. (1967) An objective approach to definitions and investigations of continental hydrologic droughts. Hydrology paper 23, Colorado State University, Fort Collins, Colorado.

Zucchini, W. and P. T. Adamson (1988) On the application of the bootstrap to assess the risk of deficient annual inflows to a reservoir. *Water Resources Management*, 2, 245-254.

Zucchini, W. and P. T. Adamson (1989) Bootstrap confidence intervals for design storms from exceedence series. *Hydrological Sciences Journal*, Oxford, U.K., 34(1), 41-48.

APPENDIX – 1

Information Regarding Source Files, Input Files and Output Files for Running the PMABB Model and Generating the Multi-season Synthetic Streamflow Replicates and Computing the Streamflow Statistics and the Water Use Statistics, given the observed/historical streamflow data measured at a site:

The executables of the files have to be run in the following order:

Step-1: First Synthetic replicates by Matched-block bootstrap have to be generated by executing the file "mabb.exe" (created from mabb.cpp). Results get stored in the file "synth.dat".

Program File: MABB.CPP
Executable File: MABB.exe
Input Files: apple.dat, mabb.dat (for runs=1), mabb-par.dat, replno.dat
Output Files: stat.det, synth.dat, check.out, matchbb.dat

apple.dat: This file contains historical record of streamflows. The values are arranged such that each row has the appropriate number of multi-season flow records corresponding to one typical water year (12 records for monthly streamflows).

mabb-par.dat: This file has five entries.

Row 1: It contains a value for the number of innovation series that have to be dispensed at the start of simulation.

Row 2: It contains the value of burn-in period, in years. Size of each synthetic replicate (innovation) is equal to the sum of length of the historical record and burn-in period.

Row 3: It contains the value of neighbourhood parameter "w", which is used for computing bandwidth (Bandwidth=2w+1)

Row 4: It contains either 1 or 0 for the option "runs"
If runs=1, the program would read block sizes from mabb.dat file
If runs=0, the program would make user specify the block sizes

Row 5: It contains replicate size, in years

replno.dat: This file has two numbers.

The first number denotes the number of synthetic replicates to be generated by the program.

The second number denotes the number of lags to be considered for computation of correlation structure. The second number is not useful for synthetic sequence generation. It is useful at a later stage while computing statistics for replicates.

mabb.dat: The number of within year blocks and the sizes of each of the blocks have to be entered in this file.

For example, if there are 12 within-year blocks each of size 1 year, the following values have to be entered in the file.

12 1 1 1 1 1 1 1 1 1 1 1 1

If there are 5 within-year blocks of size 2, 4, 3, 1, 2 months, then, the entries will look like:

5 2 4 3 1 2

stat.det: The output file shows the summary statistics (mean and standard deviation) computed for the synthetic streamflows computed for each period/season. This helps in making a quick check if the streamflow data generated are in order and the basic summary statistics of the historical streamflows are statistically reproduced.

synth.dat: Output file containing synthetic replicates obtained using the non-parametric technique Matched-block bootstrap.

Output files that can be discarded: check.out, matchbb.dat
check.out: is used for checking
matchbb.dat: is useful in

Step-2: Next, executable file created from "perturber.cpp" program "pertur–1.exe" has to be executed to perturb synthetic replicates that got stored in "synth.dat" file created in the previous step. The perturbed synthetic replicates are overwritten in the file "synth.dat". The current "synth.dat" file contains the required synthetic streamflow replicates.

For example, consider that 500 synthetic replicates are generated. In such a case, the "synth.dat" file created will contain 501 sets. The first set is historical record (echo-printed for checking), with number of years written in front of it. The following 500 sets contain synthetic streamflow replicates, with the replicate number written in front of them. Note that the replicates are numbered as: 0, 1, 2,…488, 499.

Program File:	Perturber.cpp
Executable File:	Perturber.exe
Input Files:	apple.dat (contains historical data) synth.dat (contains synthetic replicates from MABB program) perturb.dat (contains user preference for pertubation)
Output File:	synth.dat (contains historical streamflows and synthetic streamflow replicates generated using PMABB).

Perturb.dat: The file has four entries.

Row 1: Lower limit of perturbation (1-δ). For example, if δ = 0.10, 1-δ = 0.90
Row 2: Upper limit of perturbation (1+δ). For example, if δ = 0.10, 1+δ = 1.10
Row 3: Interval size (for Incrementing)
Row 4: Number of intervals between lower and upper limits of perturbation. Example, if interval size is 0.01, the number of intervals between lower and upper limits is equal to (1.10-0.90)/0.01 = 20.

Step-3:

The synthetic streamflows generated in Step-2, need to be evaluated for the model performance in terms of:

i) ability to reproduce the basic summary statistics (at multi-season and aggregated levels), marginal distributions (at multi-season and aggregated levels);

ii) ability to preserve the dependence structure in respect of a number of serial correlations specified (multi-season and aggregated annual levels), state-dependent correlations (indicative of non-linear dependence);

iii) ability to accurately predict the reservoir storage capacity corresponding to various demand levels (expressed as percent of mean annual flow); and

iv) ability to accurately predict the critical and the mean drought (run) characteristics corresponding to various truncation levels (expressed as percent of mean annual flow).

For evaluating the performance of the model, the following files from the PMABB (Main) folder are to be pasted to the STATISTICS folder (sub-folder):

MABB.DAT; ii) PERTURB.DAT; iii) APPLE.DAT; iv) SYNTH.DAT

The file RUN.DAT in the subfolder STATISTICS has to be modified by the user, corresponding to the flow data set to be modeled.

Now, the program SEAS_PRO can be executed.

Program File: SEAS_PRO.cpp
Executable File: SEAS_PRO.exe

Input Files: apple.dat, mabb.dat, perturb.dat, REPLNO.dat, RUN.dat, SYNTH.dat

Output File: STATE.gen, seascorr.his, seasmean.his, seasstd.his, seasskew.his, anncorr.his, cap.his, abovbwd.his, abovfwd.his, belobwd.his, belowfwd.his, event_no.his, marl.his, mars.his, merl.his, mers.his, unitvol.dat, replcorr.dat, anncorr.dat, REPLDEV.dat, REPLMEAN.DAT, REPLSKEW.DAT, quantile.dat, EVENTdet.dat, MARLdet.dat, MASdet.dat, MERLdet.day, MERSdet.dat, check.det, histstat.det, capdetal.out, correl.out, filter.out, scfilter.out, summary.out, detail.out, qrtjunk1.out, qrtjunk2.out, statecor.out, STATEDET.out, abovbwd.qrt, abovfwd.qrt, belobwd.qrt, belofwd.qrt, cap.qrt, events.qrt, marl.qrt, mars.qrt, merl.qrt, mers.qrt, replcorr.qrt, repldev.qrt, replmean.qrt, replskew.qrt

Summary of Useful Output Files:

1. FILTER.out
   - Mean, standard deviation, skewness and serial correlations (up to 7 lags) – historical statistics, mean synthetic statistics are printed at periodic level
   - Mean, standard deviation, skewness and serial correlations (up to 7 lags) – historical statistics, mean synthetic statistics are printed at

annual level

- Reservoir storage capacity – for each demand level, historical value, mean synthetic value, standard deviation of synthetic value; R-Bias and R-RMSE of storage capacity
- Drought Statistics (number of runs, maximum run length , average run length, maximum run sum, average run sum) for each truncation level – historical value, mean synthetic value, standard deviation of synthetic values, percent exceedance of historical value.

2. SCFILTER.out

- Lag-one state-dependent correlations (above and forward, above and backward, below and forward, and below and backward) periodic level – historical value, mean synthetic value, standard deviation of synthetic values

3. Quantile Files

- Mean synthetic flow statistics and quantiles (minimum, 5 percentile, 25 percentile, median, 75 percentile, 95 percentile and maximum) of each of the statistics are printed in these files.

These files are identified by the extension "qrt".

- Mean – REPLMEAN.QRT, Standard deviation – REPLDEV.QRT, Skewness – REPLSKEW.QRT, Correlations –REPLMEAN.QRT, State dependent correlations – ABOVBWD.QRT, ABOVFWD.QRT, BELOBWD.QRT, BELOWFWD.QRT
- Reservoir Storage – CAP.QRT
- Drought Characteristics – EVENTS.QRT, MARL.QRT, MARS.QRT, MERL.QRT, MERS.QRT

COMPUTATION OF STATISTICS

The listing of the various input files required to run and the various output files that will be generated from the different source programs that are used to compute a variety of streamflow statistics, including the validation (water-use) statistics (reservoir storage capacity and critical and mean drought characteristics) for the historical as well as the synthetic streamflows is given below.

Estimation of Summary Statistics:

STAT.CPP

COMPUTATION OF SUMMARY STATISTICS AND SERIAL CORRELATIONS OF MULTI-SEASON (PERIODIC) FLOWS
INPUT FILES : SYNTH.DAT, MABB.DAT, PERTURB.DAT
OUTPUT FILES : FILTER.OUT (APPENDED), SUMMARY.OUT, CORREL.OUT(NOT USEFUL), SEASMEAN.HIS, SEASSTDV.HIS, SEASSKEW.HIS, SEASCORR.HIS, REPLCORR.DAT, REPLNO.DAT

Estimation of Quantiles of Summary Statistics:

QUARTDAT.CPP

PREPARATION OF DATA FILES FOR ESTIMATION OF QUANTILES OF SUMMARY STATISTICS
INPUT FILE : SUMMARY.OUT (OUTPUT OF STAT.CPP)
OUTPUT FILES: REPLMEAN.DAT, REPLDEV.DAT, REPLSKEW.DAT

MEANQRT.CPP

COMPUTATION OF QUANTILE VALUES FOR MEAN FLOW OF PERIODIC DATA
INPUT FILES : REPLMEAN.DAT, REPLNO.DAT, SIZE.DAT
OUTPUT FILE : REPLMEAN.QRT (APPENDED)

SDDEVQRT.CPP

COMPUTATION OF QUANTILE VALUES OF STANDARD DEVIATION OF PERIODIC FLOWS
INPUT FILES : REPLDEV.DAT, REPLNO.DAT, SIZE.DAT
OUTPUT FILE : REPLDEV.QRT (APPENDED)

SKEWQRT.CPP

COMPUTATION OF QUANTILE VALUES FOR SKEWNESS OF PERIODIC FLOWS
INPUT FILES : REPLSKEW.DAT, REPLNO.DAT, SIZE.DAT
OUTPUT FILE : REPLSKEW.QRT (APPENDED)

ANNCHECK.CPP

COMPUTATION OF SUMMARY STATISTICS OF ANNUAL FLOWS FROM PERIODIC FLOWS

INPUT FILES : SYNTH.DAT, SIZE.DAT, REPLNO.DAT, UNITVOL.DAT
OUTPUT FILES : CORREL.OUT, ANNCORR.HIS, ANNCORR.DAT;
SEASMEAN.HIS, SEASSKEW.HIS, SEASCORR.HIS, SEASSTDV.HIS;
REPLMEAN.DAT, REPLDEV.DAT, REPLSKEW.DAT, FILTER.OUT

Quantile Estimation for Correlation Structure:

CORRQRT.CPP

COMPUTATION OF QUANTILE VALUES FOR CORRELATIONS OF PERIODIC DATA
INPUT FILES: REPLCORR.DAT, SIZE.DAT, REPLNO.DAT, ANNCORR.DAT
OUTPUT FILE : REPLCORR.QRT (APPENDED)

Computation of State-dependent Correlations:

Pre-processing of Replicates:

STATECOR.CPP

PROGRAM FOR COMPUTATION OF STATE-DEPENDENT CORRELATIONS
INPUT FILES : SYNTH.DAT, REPLNO.DAT
OUTPUT FILES : STATE.GEN, STATEDET.OUT

Grouping the data into various class intervals:

STATCOR2.CPP

PROGRAM FOR COMPUTATION OF STATE-DEPENDENT CORRELATIONS
INPUT FILES : SIZE.DAT, REPLNO.DAT, SYNTH.DAT, STATE.GEN
OUTPUT FILES : STATECOR.OUT, SCFILTER.OUT,
ABOVFWD.QRT, BELOFWD.QRT, ABOVBWD.QRT,
BELOBWD.QRT,
ABOVFWD.HIS, BELOFWD.HIS, ABOVBWD.HIS,
BELOBWD.HIS

Computation of Cross-correlations:

Processor for the Periodic Cross-correlations:

CROSSPRO.CPP

Computation of Cross-correlations

CROSSCOR.CPP

PROGRAM FOR THE COMPUTATION OF CROSS CORRELATIONS OF PERIODIC STREAMFLOWS
INPUT FILE : SYNTH.DAT
OUTPUT FILE : CROSS.GEN
SUBSEQUENT FILES REQUIRED : CROSPLUS.CPP, CROSSQRT.CPP

CROSPLUS.CPP

PROGRAM FOR THE COMPUTATION OF CROSS CORRELATIONS
MAIN JOB : FURTHER PROCESSING OF THE OUTPUT FROM CROSSCOR.CPP
INPUT FILES : CROSS.GEN (Replicate-wise cross-corr. details),
REPLNO.DAT
OUTPUT FILES :

Main Output Files:
CROSSCOR.OUT (Preservation of Hist. values)
CROSCORR.QNT (Quantiles for box-plot)

Intermediate Files:
CRSJUNK1.QNT (Mean Values of various cross-correlations)
CRSJUNK2.QNT (Quantiles of various cross-correlations)

CROSSCOR.HIS (Hist. values of Month-to-year cross-corr.)

Additional File for Post-Processing:
CROSSQRT.DAT (Data file for computing Month-to-Year cross-corrs.)

Computation of Quantiles of Cross-correlations:

CROSSQRT.CPP

COMPUTATION OF QUANTILE VALUES FOR CROSS CORRELATIONS OF MULTI-SEASON (PERIODIC) DATA
INPUT FILES : CROSSQRT.DAT, REPLNO.DAT, SIZE.DAT
OUTPUT FILE : CROSSCOR.QRT (APPENDED)

TOYUPMA.CPP

PROGRAM FOR SEPERATION OF MONTH-TO-MONTH CROSS-CORRELATION QUANTILES BETWEEN MONTHLY FLOWS IN SUCCESSIVE YEARS
INPUT FILES : REPLCORR.QRT, TOYUPMA.DAT

OUTPUT FILES: TOYUPMA.GEN, TOYUPMA.QRT

Prediction of Reservoir Storage Capacity:

SEQPEAK.CPP

SEQUENT PEAK ALGORITHM FOR DETERMINATION OF NON-FAILURE CAPACITY FOR PERIODIC FLOWS
INPUT FILES : SYNTH.DAT, REPLNO.DAT, SIZE.DAT
OUTPUT FILES : CHECK.DET, QUANTILE.DAT, CAPDETAL.OUT (APPENDED)

Quantile Estimation for Reservoir Storage Capacity:

CAPQRT.CPP

COMPUTATION OF QUANTILE VALUES FOR STORAGE CAPACITY OF PERIODIC DATA
INPUT FILES : QUANTILE.DAT, SIZE.DAT, REPLNO.DAT
OUTPUT FILE : CAP.QRT (APPENDED)

Annual and Multi-season Drought Analysis

NEWRUN.CPP

PROGRAM FOR ANALYSIS OF ANNUAL & PERIODIC DROUGHT (RUN) CHARACTERISTICS

INPUT FILES : RUN.DAT, SIZE.DAT, APPLE.DAT, SYNTH.DAT (REPLICA.DAT)

OUTPUT FILES: FILTER.OUT, EVENTS.QRT, MARL.QRT, MERL.QRT, MARS.QRT, MERS.QRT, QRTJUNK1.DAT, QRTJUNK2.DAT, HISTSTAT.DET, DETAIL.OUT, MARLDET.DAT, MERLDET.DAT, MARSDET.DAT, MERSDET.DAT, EVENTDET.DAT, EVENT_NO.HIS, MARL.HIS, MERL.HIS, MARS.HIS, MERS.HIS

NOTE : NP=1 DENOTES ANNUAL DROUGHT ANALYSIS
SYNTH.DAT IS THE DATA FILE FOR SINGLE-SITE MODELING

NP=12 INDICATES PERIODIC DROUGHT ANALYSIS
REPLICA.DAT IS THE DATA FILE FOR MULTI-SITE MODELING

APPENDIX - 2

Table: Typical Simulation Output from Water Supply Reservoir Operation

year	month	Initial storage	Inflow	Demand	Evaporation	Release	Final Storage	deficit	max_def_ month	max_deficit_s um_event	sum_deficit	No_events	deficit_no	max_deficit_le ngth	Condition	Check Sum	Limits	Demand	D+K	Release for Check
1	1	731.99	24.17	54.69	0	54.69	701.47	0.00	0.00	0.00	0.00	1	0.00	0	2	756.16	0	54.69	786.68	54.69
1	2	701.47	194.79	36.46	0	36.46	731.99	0.00	0.00	0.00	0.00	1	0.00	0	3	896.26	0	36.46	768.45	36.46
1	3	731.99	36.29	36.46	0	36.46	731.82	0.00	0.00	0.00	0.00	1	0.00	0	2	768.28	0	36.46	768.45	36.46
1	4	731.82	204.52	34.64	0	34.64	731.99	0.00	0.00	0.00	0.00	1	0.00	0	3	936.34	0	34.64	766.63	34.64
1	5	731.99	31.64	56.52	0	56.52	707.12	0.00	0.00	0.00	0.00	1	0.00	0	2	763.63	0	56.52	788.51	56.52
1	6	707.12	9.65	55.54	0	55.54	661.23	0.00	0.00	0.00	0.00	1	0.00	0	2	716.77	0	55.54	787.53	55.54
1	7	661.23	4.64	55.54	0	55.54	610.33	0.00	0.00	0.00	0.00	1	0.00	0	2	665.87	0	55.54	787.53	55.54
1	8	610.33	3.34	55.54	0	55.54	558.13	0.00	0.00	0.00	0.00	1	0.00	0	2	613.67	0	55.54	787.53	55.54
1	9	558.13	1.81	55.54	0	55.54	504.40	0.00	0.00	0.00	0.00	1	0.00	0	2	559.94	0	55.54	787.53	55.54
1	10	504.40	0.58	55.54	0	55.54	449.44	0.00	0.00	0.00	0.00	1	0.00	0	2	504.98	0	55.54	787.53	55.54
1	11	449.44	0.29	55.54	0	55.54	394.18	0.00	0.00	0.00	0.00	1	0.00	0	2	449.73	0	55.54	787.53	55.54
1	12	394.18	0.22	55.54	0	55.54	338.86	0.00	0.00	0.00	0.00	1	0.00	0	2	394.40	0	55.54	787.53	55.54
2	1	338.86	35.93	54.69	0	54.69	320.09	0.00	0.00	0.00	0.00	1	0.00	0	2	374.78	0	54.69	786.68	54.69
2	2	320.09	11.25	36.46	0	36.46	294.88	0.00	0.00	0.00	0.00	1	0.00	0	2	331.34	0	36.46	768.45	36.46
2	3	294.88	28.01	36.46	0	36.46	286.43	0.00	0.00	0.00	0.00	1	0.00	0	2	322.89	0	36.46	768.45	36.46
2	4	286.43	24.46	34.64	0	34.64	276.25	0.00	0.00	0.00	0.00	1	0.00	0	2	310.89	0	34.64	766.63	34.64
2	5	276.25	9.00	56.52	0	56.52	228.74	0.00	0.00	0.00	0.00	1	0.00	0	2	285.25	0	56.52	788.51	56.52
2	6	228.74	2.32	55.54	0	55.54	175.52	0.00	0.00	0.00	0.00	1	0.00	0	2	231.06	0	55.54	787.53	55.54
2	7	175.52	0.36	55.54	0	55.54	120.34	0.00	0.00	0.00	0.00	1	0.00	0	2	175.88	0	55.54	787.53	55.54
2	8	120.34	0.00	55.54	0	55.54	64.79	0.00	0.00	0.00	0.00	1	0.00	0	2	120.34	0	55.54	787.53	55.54
2	9	64.79	0.00	55.54	0	55.54	9.25	0.00	0.00	0.00	0.00	1	0.00	0	2	64.79	0	55.54	787.53	55.54
2	10	9.25	0.00	55.54	0	9.25	0.00	46.29	46.29	46.29	46.29	2	1.00	1	1	9.25	0	55.54	787.53	9.25
2	**11**	**0.00**	**0.00**	**55.54**	**0**	**0.00**	**0.00**	**55.54**	**55.54**	**101.84**	**101.84**	**2**	**2.00**	**2**	**1**	**0.00**	**0**	**55.54**	**787.53**	**0.00**
2	12	0.00	0.00	55.54	0	0.00	0.00	55.54	55.54	157.38	157.38	2	3.00	3	1	0.00	0	55.54	787.53	0.00
3	1	0.00	7.48	54.69	0	7.48	0.00	47.22	47.22	204.60	204.60	2	4.00	4	1	7.48	0	54.69	786.68	7.48
3	2	0.00	445.33	36.46	0	36.46	408.86	0.00	0.00	204.60	204.60	2	0.00	4	2	445.33	0	36.46	768.45	36.46
3	3	408.86	72.36	36.46	0	36.46	444.76	0.00	0.00	204.60	204.60	2	0.00	4	2	481.22	0	36.46	768.45	36.46
3	4	444.76	121.27	34.64	0	34.64	531.40	0.00	0.00	204.60	204.60	2	0.00	4	2	566.04	0	34.64	766.63	34.64
3	5	531.40	27.29	56.52	0	56.52	502.17	0.00	0.00	204.60	204.60	2	0.00	4	2	558.69	0	56.52	788.51	56.52
3	6	502.17	9.43	55.54	0	55.54	456.06	0.00	0.00	204.60	204.60	2	0.00	4	2	511.61	0	55.54	787.53	55.54
3	7	456.06	6.46	55.54	0	55.54	406.98	0.00	0.00	204.60	204.60	2	0.00	4	2	462.52	0	55.54	787.53	55.54
3	8	406.98	5.66	55.54	0	55.54	357.10	0.00	0.00	204.60	204.60	2	0.00	4	2	412.64	0	55.54	787.53	55.54
3	9	357.10	4.50	55.54	0	55.54	306.06	0.00	0.00	204.60	204.60	2	0.00	4	2	361.60	0	55.54	787.53	55.54
3	10	306.06	3.41	55.54	0	55.54	253.92	0.00	46.29	204.60	204.60	2	0.00	4	2	309.47	0	55.54	787.53	55.54
3	11	253.92	2.47	55.54	0	55.54	200.85	0.00	55.54	204.60	204.60	2	0.00	4	2	256.39	0	55.54	787.53	55.54
3	12	200.85	2.18	55.54	0	55.54	147.48	0.00	55.54	204.60	204.60	2	0.00	4	2	203.03	0	55.54	787.53	55.54

Break

15	1	0.00	1.74	54.69	0	1.74	0.00	52.95	53.60	417.46	1461.68	6	31.00	9	1	1.74	0	54.69	786.68	1.74
15	2	0.00	58.93	36.46	0	36.46	22.47	0.00	13.16	417.46	1461.68	6	0.00	9	2	58.93	0	36.46	768.45	36.46
15	3	22.47	66.04	36.46	0	36.46	52.05	0.00	0.00	417.46	1461.68	6	0.00	9	2	88.51	0	36.46	768.45	36.46
15	4	52.05	19.89	34.64	0	34.64	37.30	0.00	0.00	417.46	1461.68	6	0.00	9	2	71.94	0	34.64	766.63	34.64
15	5	37.30	5.01	56.52	0	42.31	0.00	14.21	24.55	417.46	1475.89	7	32.00	9	1	42.31	0	56.52	788.51	42.31
15	6	0.00	0.65	55.54	0	0.65	0.00	54.89	54.89	417.46	1530.78	7	33.00	9	1	0.65	0	55.54	787.53	0.65
15	7	0.00	0.36	55.54	0	0.36	0.00	55.18	55.54	417.46	1585.96	7	34.00	9	1	0.36	0	55.54	787.53	0.36
15	8	0.00	0.36	55.54	0	0.36	0.00	55.18	55.54	417.46	1641.14	7	35.00	9	1	0.36	0	55.54	787.53	0.36
15	9	0.00	0.29	55.54	0	0.29	0.00	55.25	55.54	417.46	1696.39	7	36.00	9	1	0.29	0	55.54	787.53	0.29
15	10	0.00	0.22	55.54	0	0.22	0.00	55.33	55.54	417.46	1751.71	7	37.00	9	1	0.22	0	55.54	787.53	0.22
15	11	0.00	0.15	55.54	0	0.15	0.00	55.40	55.54	417.46	1807.11	7	38.00	9	1	0.15	0	55.54	787.53	0.15
15	12	0.00	0.44	55.54	0	0.44	0.00	55.11	55.54	417.46	1862.22	7	39.00	9	1	0.44	0	55.54	787.53	0.44
16	1	0.00	0.29	54.69	0	0.29	0.00	54.40	54.40	454.94	1916.62	7	40.00	9	1	0.29	0	54.69	786.68	0.29
16	2	0.00	383.20	36.46	0	36.46	346.74	0.00	13.16	454.94	1916.62	7	0.00	9	2	383.20	0	36.46	768.45	36.46
16	3	346.74	155.39	36.46	0	36.46	465.66	0.00	0.00	454.94	1916.62	7	0.00	9	2	502.13	0	36.46	768.45	36.46
16	4	465.66	1197.72	34.64	0	34.64	731.99	0.00	0.00	454.94	1916.62	7	0.00	9	3	1663.39	0	34.64	766.63	34.64
16	5	731.99	121.42	56.52	0	56.52	731.99	0.00	24.55	454.94	1916.62	7	0.00	9	3	853.41	0	56.52	788.51	56.52
16	6	731.99	35.13	55.54	0	55.54	711.57	0.00	54.89	454.94	1916.62	7	0.00	9	2	767.12	0	55.54	787.53	55.54
16	7	711.57	18.36	55.54	0	55.54	674.39	0.00	55.54	454.94	1916.62	7	0.00	9	2	729.94	0	55.54	787.53	55.54

contd...

year	month	Initial storage	Inflow	Demand	Evaportation	Release	Final Storage	deficit	max_def_ month	max_deficit_s um_event	sum_deficit	No_events	deficit_no	max_deficit_le ngth	Condition	Check Sum	Limits	Demand	D+K	Release for Check
16	8	674.39	13.43	55.54	0	55.54	632.28	0.00	55.54	454.94	1916.62	7	0.00	9	2	687.82	0	55.54	787.53	55.54
16	9	632.28	7.33	55.54	0	55.54	584.06	0.00	55.54	454.94	1916.62	7	0.00	9	2	639.61	0	55.54	787.53	55.54
16	10	584.06	3.19	55.54	0	55.54	531.71	0.00	55.54	454.94	1916.62	7	0.00	9	2	587.26	0	55.54	787.53	55.54
16	11	531.71	1.45	55.54	0	55.54	477.62	0.00	55.54	454.94	1916.62	7	0.00	9	2	533.17	0	55.54	787.53	55.54
16	12	477.62	0.51	55.54	0	55.54	422.59	0.00	55.54	454.94	1916.62	7	0.00	9	2	478.13	0	55.54	787.53	55.54
17	1	422.59	5.81	54.69	0	54.69	373.70	0.00	54.40	454.94	1916.62	7	0.00	9	2	428.39	0	54.69	786.68	54.69
17	2	373.70	148.93	36.46	0	36.46	486.17	0.00	13.16	454.94	1916.62	7	0.00	9	2	522.63	0	36.46	768.45	36.46
17	3	486.17	139.49	36.46	0	36.46	589.20	0.00	0.00	454.94	1916.62	7	0.00	9	2	625.66	0	36.46	768.45	36.46
17	4	589.20	15.89	34.64	0	34.64	570.45	0.00	0.00	454.94	1916.62	7	0.00	9	2	605.09	0	34.64	766.63	34.64

Break

40	1	456.77	143.77	54.69	0	54.69	545.85	0.00	54.40	454.94	3452.81	13	0.00	9.00	2	600.54	0	54.69	786.68	54.69
40	2	545.85	44.34	36.46	0	36.46	553.74	0.00	22.96	454.94	3452.81	13	0.00	9.00	2	590.20	0	36.46	768.45	36.46
40	3	553.74	140.29	36.46	0	36.46	657.56	0.00	26.81	454.94	3452.81	13	0.00	9.00	2	694.02	0	36.46	768.45	36.46
40	4	657.56	21.05	34.64	0	34.64	643.97	0.00	0.00	454.94	3452.81	13	0.00	9.00	2	678.61	0	34.64	766.63	34.64
40	5	643.97	18.80	56.52	0	56.52	606.25	0.00	24.55	454.94	3452.81	13	0.00	9.00	2	662.77	0	56.52	788.51	56.52
40	6	606.25	2.54	55.54	0	55.54	553.25	0.00	54.89	454.94	3452.81	13	0.00	9.00	2	608.79	0	55.54	787.53	55.54
40	7	553.25	1.02	55.54	0	55.54	498.72	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	554.27	0	55.54	787.53	55.54
40	8	498.72	23.08	55.54	0	55.54	466.26	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	521.80	0	55.54	787.53	55.54
40	9	466.26	14.73	55.54	0	55.54	425.45	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	480.99	0	55.54	787.53	55.54
40	10	425.45	15.02	55.54	0	55.54	384.93	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	440.47	0	55.54	787.53	55.54
40	11	384.93	11.39	55.54	0	55.54	340.78	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	396.33	0	55.54	787.53	55.54
40	12	340.78	7.26	55.54	0	55.54	292.50	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	348.04	0	55.54	787.53	55.54
41	1	292.50	11.32	54.69	0	54.69	249.13	0.00	54.40	454.94	3452.81	13	0.00	9.00	2	303.82	0	54.69	786.68	54.69
41	2	249.13	258.88	36.46	0	36.46	471.55	0.00	22.96	454.94	3452.81	13	0.00	9.00	2	508.01	0	36.46	768.45	36.46
41	3	471.55	398.37	36.46	0	36.46	731.99	0.00	26.81	454.94	3452.81	13	0.00	9.00	3	869.92	0	36.46	768.45	36.46
41	4	731.99	504.84	34.64	0	34.64	731.99	0.00	0.00	454.94	3452.81	13	0.00	9.00	3	1236.83	0	34.64	766.63	34.64
41	5	731.99	133.25	56.52	0	56.52	731.99	0.00	24.55	454.94	3452.81	13	0.00	9.00	3	865.24	0	56.52	788.51	56.52
41	6	731.99	22.72	55.54	0	55.54	699.16	0.00	54.89	454.94	3452.81	13	0.00	9.00	2	754.71	0	55.54	787.53	55.54
41	7	699.16	9.00	55.54	0	55.54	652.62	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	708.16	0	55.54	787.53	55.54
41	8	652.62	5.88	55.54	0	55.54	602.96	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	658.50	0	55.54	787.53	55.54
41	9	602.96	5.52	55.54	0	55.54	552.93	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	608.47	0	55.54	787.53	55.54
41	10	552.93	1.60	55.54	0	55.54	498.98	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	554.53	0	55.54	787.53	55.54
41	11	498.98	0.80	55.54	0	55.54	444.24	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	499.78	0	55.54	787.53	55.54
41	12	444.24	1.23	55.54	0	55.54	389.93	0.00	55.54	454.94	3452.81	13	0.00	9.00	2	445.47	0	55.54	787.53	55.54

Max Deficit length of sequence = 9
Max. Deficit Sum of Sequence = 454.94
Number of Failure periods = 75
Number of Events = 13
Mean Event Deficit Length for the Sequence(Sum_deflen/No_EVNTS) = 6.25
Mean Event Deficit Sum for the Sequence(Sum_defsum/No_EVNTS) = 287.73
Volume Based Reliability (=1-Sum_defsum/(Annual_demand*NY)) = 0.86
Occurrence Based Reliability (=1-Sum_deflen/(NY*NP)) = 0.85
Resilience (No. of Events/Number of failure periods) = 0.16
Vulnerability = 55.54 Occur in Year 2 and Month 11
Total Deficit = 3452.81
Total Demand = 24910.28
Shortage Ratio = 0.14