CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

The development of optimal operating rules for reservoir has been a major research focused for many years. A real world reservoir operation is complex and often offers substantial increase in benefits even for relatively small improvements in operating efficiency. System analysis, which is driven by the development in computer technology, plays a major role in the improvements in operating efficiency. Over the last four decades significant advances have been made in reservoir operation studies and the application of system analysis techniques to water resources problems. As a consequence, the relevant literature has grown enormously. Hence the depth of coverage in this review has been restricted to the literature relevant to the objectives of the study undertaken.

As the main objective of the study is to evolve suitable release rules for an irrigation purpose reservoir using implicit stochastic optimisation, a critical review is carried out from the state-of-the-art of literature, which covers different models to infer operating rules from the optimisation models. As the present study also uses models for extending the limited historic inflow data, a detailed literature review is carried out to identify suitable monthly stream flow generation models.
2.2 RESERVOIR OPERATION MODELS

The origin of system analysis in water resources may be said to be in 1950’s in United States as reported in Maass et al (1962). Since then, the importance of system analysis in water resources has been increasingly recognized and continuous advancement is being made (Buras 1972, Loucks et al 1981). Large numbers of research publications on applying system analysis techniques to water resources problems, especially to reservoir operation models have appeared in the literature during last two decades. The reservoir system analysis models have been applied (a) for sizing reservoir storage, and establishing operation policy during reservoir planning (b) for evaluating the operation policies of existing reservoirs and (c) for supporting release decisions during real time operations. Yeh (1985) provided a comprehensive review of system analysis methods used in reservoir operation. Based on extensive review of the different technologies for optimisation of reservoir operation, he concluded that there is no general algorithm exist that can be adopted for reservoir operation problems. The choice of a method is dictated by the characteristics of the system being modeled, the objectives and the constraints to be satisfied.

2.2.1 Optimisation for Reservoir Operation

Since the early 1960s simulation models have been used to help plan and manage complex water resources system (Maass et al 1962). Typically simulation models are used to evaluate the consequences of a set of decisions (what-if-analysis) over a hydrologic period of interest. In a pure simulation model, reservoir releases are determined by a set of predefined operation rules.

Growing demands for water calls for more efficient water use while increasing environmental awareness had lead to a greater scrutiny of proposed
system expansion. In this context, optimisation models provide a valuable additional tool for planning purposes. One of the most important advances in the field of water resources engineering during the last few decades has been the evolvement and application of optimisation techniques in planning design and operation of complex water resources systems. Many optimisation techniques have been successfully applied in reservoir studies and most of these techniques are based on some type of mathematical programming methods. Traditionally, optimisation techniques have been divided into three distinct categories (i) Linear Programming (LP), (ii) Dynamic Programming (DP), (iii) Nonlinear Programming (NLP). In recent years a new group of optimisation techniques based on probabilistic search concepts has appeared in the literature. Each of these techniques can be applied in a deterministic or stochastic environment.

The role of optimisation model is to supplement rather than replace simulation modeling. Together optimisation and simulation allow investigation to focus on areas of promising benefits. An important goal of the combined optimisation/simulation approach has been the development of reservoir operation rules to aid reservoir management. Operation rules have often developed and evolved over time but are usually based on simulation model results. However before rules can be tested using simulation models, they must first be identified or defined. Optimisation models help in defining this point of departure. Simulation models that have operating rules derived from optimisation models are reported by Young (1967), Bhaskar and Whitlatch (1980), Karamouz et al (1992), Panigrahi and Mujumdar (2000), and Bessler et al (2003).

reservoir operation highlighting computation considerations. LP was covered in elaborate manner by Yeh (1985). Various simulation models and real time operation models used in water resources problems and their applications were also reviewed. Wurbs (1993) extended the work of Yeh by producing a state-of-the-art review together with an annotated bibliography of system analysis techniques applied to reservoir operation. Esogbue (1989) presented an over view of various aspects of DP. Simonovic (1992) presented a short review of mathematical models in reservoir management and operation. The work was aimed at generating ideas for closing the gap between the theory and practice. Wurbs (1993) in his review article discussed the suitability of the optimisation models in various types of decision support situations. Mujumdar (2002) dealt with optimisation and simulation models for planning and management of irrigation system. Nandalal and Simnovic (2002) in their state-of-the-art review explained the role of systems approach for resolution of conflicts over water and provide some basic knowledge of tools and techniques as they apply to water management and conflict resolution. Recently, Labadie (2004) reviewed the optimal operation of the multi-reservoir systems, along with the application of new techniques, such as Genetic Algorithm (GA), Artificial Neural Network (ANN) and fuzzy theory.

Most of the review of models discussed above had the form of linear or dynamic programming. Several of these models are sited in the following sections as a representative sampling of the variety of ways in which optimisation techniques have been applied.

2.2.1.1 Linear programming models

The most important and the widely discussed optimisation techniques used in water resources management is Linear Programming (LP). LP is basically an analytical technique for allocating scarce resources, to maximize or minimize some objective function which is a mathematical
statement of the overall system. The necessary condition for its use is that all relations among the different variables are linear, both in the constraint and the objective function to be optimized. Even though water resources problems are nonlinear in nature and approximations are necessary to formulate the model, LP is applied to various case studies in water resources planning and management. This is because of efficient algorithms and generalized computer software packages that are readily available and also, LP algorithms always converge to global optimum if the solution is feasible. The application of LP to water resources problems started in 1950’s and then on, a variety of water resources problems have been modeled using LP.

Dorfman (1962) used LP to three versions of a model each one with increasing complexities. The storage capacity and the target releases are the decision variables. In the first version, over year storage is not considered and each year is treated like all other years with the given long term average seasonal inflows for dry and wet seasons. The second version incorporates over year storage estimated through critical period analysis. The inflows are treated as deterministic in these two versions, while in the third version they are treated as stochastic.

Thomas and ReVelle (1966) used LP technique to obtain the optimal operating policies for a multipurpose dam. The release in each month for power production and irrigation are the decision variables in the problem and the demand for hydropower was relatively constant for each month. The trade-off between hydropower and irrigation was examined. Young (1968) used LP to determine the optimal operation policies for a low flow augmentation cum flood control reservoir. Discrete inflow probabilities and low flow and flood penalty functions were assumed to be known. Solution for the example indicated the degree to which flood and water quality control
were complementary and concluded that even if the flow penalty was known only within a range, the optimal policy could be determined.

To select capacity variable and operating variables in an irrigation project planning, Rogers and Smith (1970) used a static and deterministic LP models. Windsor and Chow (1972) used LP model for multireservoir operation studies. Piecewise linear approximation was used to replace the nonlinear functions involving a set of decision variables. Windsor (1973) presented a methodology employing recursive linear programming as the optimisation tool for the analysis of multireservoir flood control systems. By means of dividing the flood period in shorter operational periods, it is shown how the system policies may be adjusted to incorporate the latest forecast information and thus ensure maximum flexibility under actual operating conditions.

Clyde and King (1973) studied the optimal allocation of water resources of Utah state in USA using a LP model. The optimal cost minimizing system consists of various combinations of ground water, surface water and interregional transfer activities which minimize the cost. Becker and Yeh (1974) developed a methodology using LP for period by period optimisation, for the selection of an optimal reservoir storage policy path through a specified number of policy periods.

Lakshminarayana and Rajagopalan (1977) applied LP model for determining the extent of allocation of irrigation area to alternative water sources namely canals and tube wells. It was also observed that an increase in the available area for irrigation would give rise to increase the benefits from irrigation activities.

Rydzewski and Rashid (1981) described an approach to find optimal allocation of surface and ground water resources to three agricultural areas in
the Jordan valley under conditions of scarce water supply. They used both LP and NLP models and the LP model was to allocate water from eight sources to three irrigation areas.

For determining firm yields for single and multiple reservoir system in the Potomac River basin, Palmer et al (1982) used LP model. Trade-off analysis was performed to determine the impact of stream flow requirements constraint on the system yield. The application of sequential linear programming technique for optimisation of reservoir operation was demonstrated by Grygier and Stedinger (1985). Martin (1987) derived optimal daily operation of surface water system using a successive LP, forcing to satisfy all the conditions. They concluded that if the objective function was not convex then the global optimum was not guaranteed.

Palmer and Holmes (1988) incorporated a LP model into a decision support system to determine optimal operating policy. The LP model is based on the twin objectives of maximizing yields and minimizing the economic losses associated with deficit from a specified target. Ellis and ReVelle (1988) presented a deterministic separable linear algorithm for maximizing aggregate hydro power production. The method is amenable to solution using standard LP software.

Bessler et al (2003) used a linear network modeling approach for the development of a general operating policy for a water supply system. To solve the LP they used RELAX algorithm. The approach assumes that a water supply system can be modeled as a linear network, given linear constraints and linear cost.

A common difficulty in linear programming formulations for reservoir operations and management problems is that the solution of the optimisation model may report spills, even when reservoir is full. Shin and
ReVelle (1994) attempted to remove this drawback by Mixed Integer Linear Programming (MILP). MILP deals with linear programming in which some of the variables assume integer values. The exhaustive application of MILP in environmental sciences has been reported. Srinivasan et al (1999) developed a MILP model for the reservoir performance optimisation. Tu et al (2003) demonstrated the applicability of MILP model for a multipurpose, multireservoir system.

While linear programming models have been used rather extensively in reservoir problems, the sequential nature of the decisions to be made in reservoir operation problems renders the dynamic programming (DP) models more amenable to solutions.

### 2.2.1.2 Dynamic programming

Dynamic Programming (DP) has been described as the theory of multi-stage (sequential) decision processes (Bellman 1957). DP is not a precise mathematical algorithm but a general approach of solving optimisation problems. A problem formulated as a dynamic programming problem has several defining characteristics (Hillier and Lieberman 2001). The division of the problem into a series of stages that require a decision at each stage; a set of state variables that describe the system condition at the beginning of the stage; and a set of decision variables that transform the current state to the state at the start of the following stage. In this manner a highly complex problem is effectively decomposed into a series of sub-problems that are solved recursively. The validity of this approach rests on the “principle of optimality” (Bellman 1957). The probability of any future outcome is independent of the past event(s) and depends only on the present state. Given the current state, an optimal policy for the remaining stages is independent of the policy decisions adopted in the previous stages. Since DP is based on a decomposition technique, it requires separability and
monotonicity of the objective function (Rao 1996). The problem must be expressed as the sum or product of functions for each stage. There have been numerous studies applying DP to reservoir operation. The popularity of the technique is ascribed to the fact that both the nonlinear and stochastic nature of water resources problems can be readily represented by a DP formulation (Yeh 1985). Dynamic programming can handle non-convex, nonlinear, discontinuous objective and constraint functions. However, where the objective function can be linearised, discrete DP offers few advantages over LP.

2.2.1.3 Nonlinear programming

In contrast to LP and DP relatively little is published on the use of nonlinear programming (NLP) techniques for reservoir system operation. The main reason is that NLP techniques are slow, iterative, and take up large amount of computer storage and time. In his review of system analysis, Yeh (1985) concluded that NLP had proved unpopular due to its demanding computational requirements, resulting in long solution times. However recent rapid advances in computer processing speed and the availability of commercial nonlinear solvers have made NLP a more attractive technique to practitioners. NLP offers a more general mathematical formulation of reservoir problems. NLP include search techniques, quadratic programming, geometric programming, and separable programming. They can be used in conjunction with simulation as well as other programming techniques. Unlike stochastic dynamic programming (SDP), NLP cannot easily handle stochastic inflows, although NLP does not require discretisation of decision and state variables. NLP techniques applied to reservoir operation include the gradient projection method (Lee and Waziruddin 1970), the conjugate gradient method

2.2.1.4 Linear decision rules

Operating rules are commonly deduced indirectly from model results. Deliveries, storage, instream flows and other objectives can be expressed in terms of the parameters of a set of linear decision rules (LDRs) and solved for directly using mathematical programming techniques. Since the introduction of LDRs by ReVelle et al (1969), they have remained a popular research topic and when used in conjunction with chance-constrained reliability models offer the promise of a screening model for both reservoir sizing and optimal operation. Various LDRs have been proposed: S-type where release is a function of storage (ReVelle et al 1969, ReVelle 1999) and SQ-type where release is a function of storage and inflow during the current time step (Loucks 1970). Two typical LDR forms as described by ReVelle et al (1969) and Loucks (1970) are

\[ R_t = S_t - b_t \]  \hspace{1cm} (2.1)

\[ R_t = S_t + I_t - b_t \]  \hspace{1cm} (2.2)

where \( R_t, I_t, \) and \( S_t \) are release, inflow, and beginning storage values, respectively, for period \( t \) and \( b_t \) operation parameters for time \( t \). The optimal value of \( b_t \) is to be found through the application of an appropriate optimisation method.

Early work by Young (1967) suggested that LDRs might be as effective as more complex (non-linear) rules. Many papers published in the 1970s and early 1980s extended the single reservoir model of

Non-linear decision rules have been studied by Colorni and Fronza (1976) and Simonovic and Marino (1980). Loucks and Dorfman (1975) by evaluating various LDRs found that they are useful as preliminary screening models to demonstrate the trade-off between storage capacity and reliability. But found that a detailed simulation modeling is required to determine final reservoir capacities or operating policies.

The interest in LDRs seems questionable. According to Stedinger (1984) LDRs have little to offer except their simplicity. For a single reservoir operated for water supply, recreation and flood control, Stedinger reports that the SOP out-performs LDRs (using total water shortage as a metric). For reservoir sizing Stedinger found that S and SQ type LDRs over-estimate the required capacity for a given reliability; S-type rules performing particularly badly. LeClerc and Marks (1973) concluded that chance-constrained LP has serious drawbacks for the design and analysis of large reservoir networks. The method does not define the magnitude by which the system fails. Highly non-linear economic objectives are unlikely to be well served by LDRs.

2.2.2 Evolutionary Algorithms

Within the last decade, many researchers have shifted the focus of reservoir optimisation from traditional optimisation techniques based on
linear and non-linear programming to the implementation of Evolutionary Algorithms (EA) namely; Genetic Algorithms (GA) simulated annealing and Ant Colony Optimisation (ACO).

Noted advantages that exist with the use of EA for application to reservoir optimisation are; (i) they treat the optimisation problem as discrete, (ii) they deal only with objective function information and avoid complications associated with determining derivatives or other auxiliary information, (iii) they are global optimisation procedures, and (iv) as they deal with a population of solutions numerous optimal or near-optimal solutions can be determined.

Due to the iterative nature of the solution generation of EA, they can be intuitively seen as algorithms that incrementally search through the solution-space using knowledge about solutions that have already been found to further guide the search. The searching behavior of EA can be characterised by two main features (Colorni et al. 1996), (i) exploration, which is the ability of the algorithm to search broadly through the solution space and (ii) exploitation, which is the ability of the algorithm to search more thoroughly in the local neighborhood where good solutions have previously been found. By definition, these attributes are in conflict with one another.

2.2.2.1 Genetic Algorithms

Genetic Algorithms (GA), introduced by Holland (1975) and developed by Goldberg (1989), offers a powerful optimisation approach that has a potential in water resources system analysis; its applications are quite recent. GA modeling is gaining importance, because of its robust random search capability and near global optimal values. Very few works has been found in application of GA for reservoir operating problems

East and Hall (1994) applied a GA to the four-reservoir problem. The objective was to maximize the benefits from power generation and irrigation water supply, subject to constraints on storages and releases from the reservoirs. Fahmy et al. (1994) applied GA to the reservoir system operation and compared the performance of the GA approach with that of dynamic programming. It was concluded that GA performs better than the DP model. Oliveira and Loucks (1997) used a GA model to evaluate operating rules for multi-reservoir systems, demonstrating that GA can be used to identify effective operating policies. A brief review of GA applications to the water resources problems can be found in the work of Wardlaw and Sharif (1999). Sharif and Wardlaw (2000) applied genetic algorithm for the optimisation of multi-reservoir system in Indonesia by considering the existing development situation in the basin and two future water resource development scenarios and proved that the GA was able to produce solutions very close to those produced by dynamic programming.

Kim and Heo (2004) applied multi-objective GA to optimize multi-reservoir system of the Han River basin in South Korea. A curve identifying the population points that define optimal solutions were derived. It was reported that GA multi-objective has limited application in multi-reservoir system optimisation. Ahmed and Sarma (2005) developed a GA model for deriving the optimal operating policy and compared its performance with that of Stochastic Dynamic Programming (SDP) for a multi-purpose reservoir. The objective function of both GA and SDP are set to minimise the squared deviation of irrigation release only. The irrigation release rules were assumed as piecewise linear functions, and based upon this number of linear rule
functions four policies were developed using GA model. Reis et al (2006) applied a hybrid method using Genetic Algorithm (GA) and Linear programming (LP) to determine operational decisions for a reservoir system over the optimisation period. This method identifies part of the decision variables called Cost Reduction Factors (CRF) by GA and operational variables by LP. Jothiprakash and Ganesan Shanthi (2006) applied a GA to a single reservoir to derive the optimum operational strategies and found to be successful in real world reservoir operation.

2.2.2.2 Ant colony optimisation

Ant colony optimisation (ACO), called ant system (Colorni et al 1991), was inspired by studies of the behavior of ants (Deneubourg et al 1983). Ant algorithms were first proposed by Dorigo et al (1996) as a multi-agent approach to different combinatorial optimisation problems like the traveling salesman problem and the quadratic assignment problem. The ant-colony metaheuristic framework was introduced by Dorigo and Di Caro (1999), which enabled ACO to be applied to a range of combinatorial optimisation problems. Dorigo et al (2000) also reported the successful application of ACO algorithms to a number of bench-mark combinatorial optimisation problems.

So far, very few applications of ACO algorithms to water resources problems have been reported (Abbaspour et al 2001, Maier et al 2003, Jalali et al 2003, Nageshkumar and Janga Reddy 2006). In which for reservoir optimisation, Jalali et al (2003) proposed ACO algorithms for monthly operation of single purpose reservoir system and Nageshkumar and Janga Reddy (2006) for monthly operation of a multi purpose reservoir system. According to them the ACO algorithm gave much better performance than GA.
Despite the development and growing use of optimisation models (Labadie 1997), the vast majority of reservoir planning and operation studies are based predominantly or exclusively on simulation modeling, and thus require intelligent specification of operating rules. Practical real-time operations also usually require the specification of reservoir operating rules. These rules determine the release and storage decisions for each reservoir at each time-step during the simulation and helps to guide the reservoir operators (Hufschmidt and Fiering 1966).

2.3 MODELING APPROACHES TO DEAL WITH STOCHASTIC NATURE OF INFLOWS

Reservoir operation is a multistage dynamic stochastic control problem (Marino and Loaiciga 1985). Reservoir operators often must make release decisions with incomplete information. Seasonal demand may be relatively fixed. In contrast, variation in natural streamflow between seasons may be highly variable. Long-range reservoir inflow forecasts are unreliable. Release policies for reservoirs with low refill probabilities and variable annual inflows may be oriented towards minimizing risk of losses in subsequent periods as well as maximizing short-term benefits. There are two modeling approaches to treat the stochastic nature of streamflows in the reservoir system namely deterministic approach and stochastic approach. In a deterministic approach, which is generally known as Implicit Stochastic Optimisation (ISO) model, inflows are based on the historic flow record or a sequence of synthetic data. A stochastic approach involves the assignment of probabilities to discrete flow ranges and it is known as Explicit Stochastic Optimisation (ESO) model.
2.3.1 Explicit Stochastic Optimisation (ESO) Model

Reservoir operation planning is inherently stochastic given the uncertain nature of reservoir inflows. Modeling risk is discussed by many authors (e.g. Hashimoto et al 1982, Fiering 1982). However it is often inadequately represented in optimisation models (Watkins and McKinney 1997). Explicit stochastic optimisation model is an approach; where the probabilistic descriptions of streamflow are directly incorporated into optimisation formulas. The explicit stochastic approach is designed to operate directly on probabilistic descriptions of random streamflow processes rather than deterministic hydrologic sequences (Labadie 2004). Stochastic analysis presents computational difficulties for both LP and DP. Three different stochastic techniques are commonly used in conjunction with first-order Markov chains i.e. LP; DP; and policy iteration. Howard (1960) gives an excellent introduction to policy iteration. These techniques are compared by Loucks and Falkson (1970). Gablinger and Loucks (1970) found DP to be less computationally burdensome; however the number of state variables using traditional stochastic DP (SDP) methods is limited to two or three. Unfortunately, many of the iterative schemes that have been developed to reduce the dimensionality of deterministic DP problems are not applicable to their stochastic counterpart (Yakowitz 1982).

The most popular explicit optimisation technique is stochastic dynamic programming (SDP), which has a number of applications for reservoir operations (Stedinger 1984, Kim and Palmer 1997, Mousavi and Karamouz 2003). Kelman et al (1990) proposed sampling stochastic DP (SSDP), which directly incorporates inflow scenarios in the DP recursive equation to reflect the various characteristics of streamflows at all sites within a basin. SSDP does not require classifying stochastic variables and their probability functions into discrete states. Faber and Stedinger (2001)
successfully combined SSDP with updated forecast information from ensemble streamflow prediction (ESP) of the National Weather Service. However the computational effort required makes this approach infeasible for complex systems. Moreover, SDP models are restricted to separable objective functions and constraints.

A scenario tree is an example of the explicit approach, where each branch of the tree represents a possible value of future streamflow (Pereira and Pinto 1985, Jacobs et al 1995, Watkins et al 2000). An important issue of the scenario tree approach is determining enough stages and branches so that the optimisation model can properly capture uncertainties in streamflow (Watkins and Wei 2004). Typically, the more stages and scenarios the better the model, but too many may lead to a dimensionality problem. Dempster and Thompson (1999), Dupacova et al (2000), and Heitsch and Romisch (2003) overcame this issue in theory, but additional work is needed to develop scenario trees appropriate for a multistage model (Watkins and Wei 2004).

2.3.2 Implicit Stochastic Optimisation (ISO) Model

Difficulties in explicitly stochastic formulations have led water resource engineers of large integrated systems to rely on Implicitly Stochastic Optimisation (ISO) techniques (Labadie 1997). In the implicit stochastic optimisation, a number of synthetic flow sequences are first generated using time series models or Monte Carlo techniques. Then the system is optimized through mathematical programming model for each of these sequences and the optimal operating policies are derived using multiple regression or multivariate analysis on the optimized data sequences.

The first implicit stochastic optimisation model reported in the literature is Monte Carlo dynamic programming proposed by Young (1967). Optimal operating policies derived from forward dynamic programming
algorithm are used to determine optimal operating rule using regression analysis. Storage level and previous period inflow were taken as independent variables and release as dependent variable. The general operating rule derived by the procedure suggested by Young (1967) could capture some of the optimal solution resulting from the large deterministic optimisation results.

Croley (1974) gave a comprehensive critique of ISO as well as comparison with explicit stochastic optimisation model. McKerchar (1975) used dynamic programming coupled with multivariate stream flow simulation model for considering the stochastic nature of inflows in to a system of two interconnected reservoirs in the southern island of Newsland. The objective of the study was to minimize the expected value of the thermal power generation cost while operating the hydro power generating plants in conjunction with thermal power generating plants.

Gal (1979) applied ISO coupled with policy iteration methods which approximate a quadratic cost function over a substantial portion of a state space and chooses the set of controls that minimise the expected costs. Bhaskar and Whitlatch (1980) generated synthetic stream flow records using Thomas-Fiering model and optimal operating policies were derived using backward DP. They made use of the procedure of deriving general operating policies from deterministic optimisation as initiated by Young (1967). They also considered a quadratic loss function and derived monthly policies by regression analysis of optimal set of releases on the input and state variables.

Karmouz and Houck (1982) used a deterministic dynamic program, regression analysis and simulation model known as a Dynamic Programming – Regression (DPR) model to generate the operating rules. Willis et al (1984) approached optimal reservoir operating problem with a somewhat different perspective. The probability distribution function for the optimal reservoir
release was conditioned on the storage and inflows of the reservoir system. Optimal releases for each time period in operational horizon was then determined using LP. After repeating this procedure for a large number of synthetic stream flow sequences, an operational release policy was determined using probability mass function of the optimal release. Implicit stochastic optimisation is also taken and tested using simplified simulation model by Lund and Ferreira (1996).

Application of regression techniques however finds a limitation in the necessity to define explicitly the mathematical form of the link between independent and dependent variables (Cancelliere et al 2002). For highly nonlinear problems and for problems with high uncertainty, this method may not yield satisfactory results. This places the focus on new-generation techniques and tools emerging to intelligently assist humans in analyzing data, finding useful knowledge and in some cases performing analysis automatically. During the last decade such models became quite popular which are based on tools of Machine Learning (ML) such as Artificial Neural Networks (ANN), Fuzzy Inference System (FIS), Combination of ANN and FIS techniques known as ANFIS, Data mining techniques such as decision tree, rough set approach etc and heuristic method such as Genetic algorithm (GA). ML is an interdisciplinary subject, which is enriched with concepts drawn from diverse fields such as statistics, artificial intelligence, information technology, biology, cognitive science, philosophy, control theory and others (Mitchell 1997).

Salient methods pertaining to the generation of reservoir operation rules by Machine Learning (ML) are given in Table 2.1. From the table it is clear that the field of ML complement the field of traditional techniques used to derive reservoir operation rules from the inflow-release patterns. Also it is clear that one of the most frequently and successfully used technique in this
<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Release pattern</th>
<th>Method of rule generation</th>
<th>Type of streamflow</th>
<th>Inferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shresta et al (1996)</td>
<td>Historical releases</td>
<td>Fuzzy rule based modelling</td>
<td>Historical flows</td>
<td>According to the authors, the model is very simple and can be easily implement in real world situation.</td>
</tr>
<tr>
<td>Russell and Campbell (1996)</td>
<td>Optimal release pattern using dynamic programming</td>
<td>Fuzzy programming</td>
<td>Inflows generated from probability distribution</td>
<td>Easy to explain and understandable operation rules are developed and concluded that the approach is useful to supplement other types of approaches</td>
</tr>
<tr>
<td>Raman and Chandramouli (1996)</td>
<td>Optimal release pattern using dynamic programming</td>
<td>Neural Network (NN)</td>
<td>Historical flows</td>
<td>By comparing the DPN model with (explicit) stochastic dynamical programming, standard operating policy, and dynamic programming regression, it was found that DPN model out performs the other three operation models</td>
</tr>
<tr>
<td>Jain et al (1999)</td>
<td>Optimal release pattern using dynamic programming</td>
<td>Artificial Neural Network (ANN)</td>
<td>Inflow forecasted using ANN</td>
<td>Compared the result with regression model and non linear regression model and concluded that ANN model are an effective and viable tool for reservoir operation rule generation.</td>
</tr>
<tr>
<td>Panigrahi and Mujumdar (2000)</td>
<td>Optimal release pattern using stochastic dynamic programming</td>
<td>Fuzzy rule based model</td>
<td>Transitive probability matrices</td>
<td>Concluded that the rules obtained from fuzzy rule based modeling with linguistic statement are appealing to the reservoir practitioners and they avoid complex optimisation procedures</td>
</tr>
<tr>
<td>Jolma et al (2001)</td>
<td>Historical releases</td>
<td>Fuzzy rule based model</td>
<td>Historical flows</td>
<td>Concluded that the fuzzy approach in general is suitable for simulating reservoir operation</td>
</tr>
<tr>
<td>Authors and Year</td>
<td>Release pattern</td>
<td>Method of rule generation</td>
<td>Type of streamflow</td>
<td>Inferences</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------</td>
<td>---------------------------</td>
<td>-------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Naresh and Sharma (2001)</td>
<td>Optimal releases from non linear optimisation</td>
<td>Fuzzy neural system</td>
<td>Inflows randomly generated using normal distribution.</td>
<td>The rules from the fuzzy-rule based model appear to be robust and do not change significantly with small changes in the input values. Concluded that the method is simple, easy and reliable for the reservoir operation model</td>
</tr>
<tr>
<td>Chang and Chang (2001)</td>
<td>Optimal releases from Genetic Algorithm(GA)</td>
<td>Fuzzy Rule Based system</td>
<td>Historical flows</td>
<td>The proposed models built on different types of knowledge produced much better performance than the traditional methods of rule curves in real-time reservoir operation</td>
</tr>
<tr>
<td>Hasebe and Nagayama (2002)</td>
<td>Historical releases</td>
<td>Fuzzy system and neural network fuzzy system</td>
<td>Inflows forecasted using filter separation AR method</td>
<td>Fuzzy system is an effective operation system when water use is the main objective and neural network fuzzy system is effective for flood control systems</td>
</tr>
<tr>
<td>Cancelliere et al (2002)</td>
<td>Optimum release pattern using dynamic programming</td>
<td>Artificial Neural network approach</td>
<td>Historical data</td>
<td>Compared the result with standard operating policy and the results shows operating rules obtained from ANN closely resembles the real system operation criteria</td>
</tr>
<tr>
<td>Ponnambalam et al (2003)</td>
<td>Optimum release pattern using nonlinear stochastic optimisation method</td>
<td>Artificial neural network fuzzy inference system(ANFIS) model</td>
<td>Simulation</td>
<td>The authors compared the result with fuzzy rules and multiple regression based rules. The results shows that ANFIS rules perform much better than the regression rules</td>
</tr>
<tr>
<td>Authors and Year</td>
<td>Release pattern</td>
<td>Method of rule generation</td>
<td>Type of streamflow</td>
<td>Inferences</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------</td>
<td>---------------------------------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Bessler et al (2003)</td>
<td>Optimum release pattern using linear network modeling</td>
<td>Data mining approach- decision tree</td>
<td>Historical inflows</td>
<td>Generated rules provided good results for the operation of single reservoir system and provides better simulation results than linear regression method</td>
</tr>
<tr>
<td>Chang et al (2004)</td>
<td>Optimal releases from Genetic Algorithm(GA)</td>
<td>Fuzzy Rule Based system</td>
<td>Historical flows</td>
<td>According to them (1) the GA is an efficient way to search the optimal input-output patterns, (2) the FRB can extract the knowledge from the operating rule curves, and (3) the ANFIS models built on different types of knowledge can produce much better performance than the traditional methods in real-time reservoir operation</td>
</tr>
<tr>
<td>Chandramouli and Deka Paresh (2005)</td>
<td>Optimum release pattern using dynamic programming</td>
<td>Artificial Neural network approach</td>
<td>Historical flows</td>
<td>Compared the results with regression based approach and found ANN method outperforms the regression method.</td>
</tr>
<tr>
<td>Mousavi et al (2005)</td>
<td>Optimum release pattern using dynamic programming</td>
<td>Fuzzy rule based modeling (FRB)</td>
<td>Historical inflows</td>
<td>The operation rules derived from FRB approach performs well in terms of satisfying system target performance and computational requirement</td>
</tr>
<tr>
<td>Authors and Year</td>
<td>Release pattern</td>
<td>Method of rule generation</td>
<td>Type of streamflow</td>
<td>Inferences</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------</td>
<td>--------------------------------------------</td>
<td>----------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Pan Liu et al (2006)</td>
<td>Optimum release pattern using dynamic programming</td>
<td>Artificial Neural Network (ANN) and GA</td>
<td>Historical inflows</td>
<td>The dynamic programming neural-network simplex (DPNS) model, which is trained by GA and back-propagation (BP) algorithm was superior to the DPN model and found that it is an effective and efficient method to derive operating rules.</td>
</tr>
</tbody>
</table>
respect is artificial neural networks (ANN). But it has been demonstrated that there is a whole set of other methods that can at least as accurate, and have additional advantages (Solomatine and Dulal 2003). One of such ML method is classification based on association rules (Liu et al 1998) which come under the category of data mining.

2.4 CLASSIFICATION BASED ON ASSOCIATION RULES

Classification has been studied extensively in literatures and techniques developed for classification include a variety of learning algorithms, such as $k$-nearest neighbors, decision tree induction, Bayesian classification, neural networks, hidden Markov models, support vector machines, etc. These, majority of traditional classification techniques use heuristic-based strategies for building the classifier (Witten and Frank 2000). In constructing a classification system, they look for rules with high accuracy. Once a rule is framed, they delete all positive training objects associated with it. Thus, these methods often produce a small subset of rules, and may miss detailed rules that might play an important role in some cases. The heuristic methods that are employed by traditional classification techniques often use domain independent biases to derive a small set of rules, and therefore, rules generated by them are different in nature and more complex than those that users might expect or be able to interpret (Pazzani et al 1993). In recent years, a promising new approach that mainly uses association rule mining in classification has been proposed.

Association rule discovery and classification are analogous tasks in data mining, with the exception that classification main aim is the prediction of class labels, while association rule mining discovers associations between attribute values in a data set. In the last few years, association rule mining has been successfully used to build accurate classifiers, which resulted in a new approach coming to life, known as classification based on association rule
mining (CAR) (Ali et al 1997, Liu et al 1998). The aim of association rule discovery is to find rules with strong associations between items from the training data. It focuses on detecting relationships between the items.

The richness and profoundness of association rules lay concrete foundations for building classification models. In addition, the easy interpretability of the rules by humans and the competitive performance exhibited in many application domains made such rule-based models especially popular. Empirical studies (Liu et al 1998, Antonie et al 2003, Yin and Han 2003) showed that CAR often builds more accurate classifiers than traditional classification techniques and many of the rules found by these methods can not be discovered by traditional classification algorithms.


In water resources area, Dingsheng et al (2007) used association rule mining technique in hydrological time series for helping to leverage the hydrological data more effectively and extract insightful information from the data. An improved method of mining quantitative association rules is proposed to implement the hydrological analysis, as well as the optimisation of the rules. Also an improved clustering method is exploited for discretisation of attributes in the process of mining quantitative association rules. The experiment results indicate the feasibility and practicability of association rule mining to analyse the time series in hydrology.
For predicting the tropical cyclone intensity using the satellite based precipitation observation, Tang et al (2005) used association rule mining technique. Chatzidimitriou and Andrew Sutton (2005) used the approach of mining classification rules for the tropical cyclone intensity prediction based on association rule discovery and compared the results with another rule discovery approach known as Swarm intelligent rule discovery. According to them the two approaches differ very much. Swarm intelligent rule discovery have more degrees of freedom in the search, while classification rules based on association rule (CAR) discovery performs a more exhaustive search based on certain criteria. So it is often the case that CAR would found better quality rules, which was reflected in their results reported.

Harms et al (2001a and 2001b) described the application of association rule mining in a Geo-spatial decision support system which focuses on drought risk management. Tadesse et al (2004), used association rule mining techniques to identify drought episodes and associate these episodes with climatic and oceanic indices. According to them the generated rules can be used for proactive management of drought and improving the reliability of drought predictions.

According to Ali et al (1997) association rules can be used in domains where conventional classifiers would be ineffective due to one or more of the following conditions:

i) There are a very large number of attributes.

ii) Most of the values of each attributes are missing.

iii) The class distribution is much skewed, and the user is interested in understanding some low frequency classes.

iv) Each of the attributes must be modeled based on the other attributes.

v) The number of training examples is very large.
Generally in implicit stochastic optimisation (ISO) models, for deriving operation rules from data set, the main idea behind is that, any function that provides a better fit to the state-decision vectors can better operate the system. However, this assumption is actually not valid for various reasons. For example, all of the state decision points are considered to be worth the same in the fitting process. However, operations during some exceptional conditions, such as droughts and floods, can be more important than operations during normal conditions. But the frequent occurrence of normal conditions may dominate the fitting process and reduce the weight of exceptional conditions to influence the optimal values of operating policy parameters. Thus they are ineffective due to the condition (iii) as stated above. Hence in the present study, classification using association rules is proposed for the first time for derivation of the reservoir operation rule from the inflow-release pattern.

2.5 STOCHASTIC SIMULATION OF STREAMFLOW

Stochastic hydrology, the goal of which is to generate synthetic stream flow sequences that are statistically similar to observed streamflow sequences, plays a significant role in the field of water resources planning and management. Synthetic river flow series are useful for determining the dimensions of hydraulic works, for flood and drought studies, for optimal operation of reservoir systems, for determining the risk of failure of dependable capacities of hydroelectric systems, for planning capacity expansion of water supply systems, and for many other purposes (Salas 1993).

Streamflow data used in deriving planning alternatives and operating policy of water resource systems are restricted by available historical records because information content in a single historical record has its own limitation as that sequence is not going to repeat itself in future.
Streamflows can be modeled and forecasted by a number of different methods, including physically-based (deterministic) methods and statistically-based (empirical) methods. Stochastic models capture streamflow variability by generating ensembles- multiple scenarios of plausible streamflow values which include extreme events such as floods and droughts and which preserve the statistics of the observed data. The ensembles can be used to quantify the uncertainty of the forecast and to calculate exceedence probabilities. Both statistically-based and physically-based models can generate ensemble forecasts. Empirically-based stochastic models often selectively sample from the range of past streamflow data to generate ensembles. In both frameworks, the models operate on the premise that the statistics (mean, standard deviation, lag-1 correlation, and skew) of the historical flow (or precipitation) are likely to occur in the future, i.e. the stationary assumption. Statistical models require less initial data and parameters and do not need to be calibrated like deterministic models do. Deterministic models, however, typically have several parameters to be calibrated, thus requiring large amounts of data.

2.5.1 Characteristics of Hydrological Time Series

The structure of hydrological time series consists of mainly one or more of these four basic structural properties and components (Salas 1993):

- Over year trends and other deterministic changes (such as shifts in the parameters). In general, natural and human induced factors may produce gradual and instantaneous trends and shift in hydrological time series. A detailed discussion of trends and shifts in hydrological data and their removal are given in Salas (1993).
• Intermittency in the processes, mainly consisting of the hydrology of intermittent sequences of zero and non-zero values.

• Seasonal or periodic changes of days, weeks or months within the annual cycle. Periodicity means that the statistical characteristic changes periodically within the year. For example, in hydrologic data concerning river flows, we expect high runoff periods in the spring and low flow periods in the summer. Thus the river flow correlations between spring months may be different from the correlations between summer months.

• Stochasticity or random variations.

2.5.2 Modeling and Simulation of Hydrological Series

Traditional statistically-based model fit a regression, often linear, between the response variable and the independent variables. They are of the form:

\[ Y_t = a_1 x_{t-1} + a_2 x_{t-2} + \ldots + a_p x_{t-p} + e_t \]  

(2.3)

where the coefficients \( a_1, a_2, \ldots, a_p \) are estimated from the data. The error, \( e_t \), is assumed to be normally distributed with mean 0 and standard deviation 1.

In the above model, the independent variables can be past values of the response variable itself:

\[ Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + e_t \]  

(2.4)

These models are termed autoregressive moving average (ARMA) and periodic autoregressive (PAR) models. These traditional modeling techniques
are termed “parametric” because they are based on estimating parameters (e.g., determining the coefficients) to fit the model. Parametric models inherently assume that the time series is normally (Gaussian) distributed (Salas 1993). Typically, streamflow data do not fit a Gaussian distribution, thereby violating this assumption. To address this, the data are transformed to a normal distribution using a log or power transformation before fitting a parametric model to the transformed data (Sharma et al 1997). The forecasted values are then back-transformed into the original space. This process of fitting the model on the transformed data and then back transforming it often does not guarantee the preservation of statistics (Benjamin and Cornell 1970, Bras and Iturbe 1985, Sharma et al 1997). There are a number of literature regarding the fitting and testing of such models which are discussed below.

2.5.2.1 Stream flow generation model

Generally, the available historic stream flow record is not long enough to contain extreme conditions of high flow and low flows, hence it is not possible to assess completely the performances or reliability of a reservoir system situated across such a river. Stochastic methods provide a powerful tool for the water resources planners to effectively and efficiently formulate the development proposals by testing different operating policies with generated sequences of hydrologic inputs. The most effective stochastic flow generation model should preserve the skewness and extended tail behavior seen in the natural flow data. Preserving the tails is of particular interest because the tails exhibit the probability of extreme low or high flow.

The parametric models generally preserve the mean, variance, and auto correlations (depending on the order of the model). As the time series is transformed to a Gaussian distribution (or near Gaussian) by appropriate transformation, the skewness is preserved to the extent the transformations are good. Parametric models require estimating multiple model parameters,
depending on the type of model. Considerable uncertainty can exist in the estimation, depending on the length of the historical data, which adds to variability in the simulations. Furthermore, because they are restricted to a Gaussian framework (Gaussian distribution assumption), parametric models cannot reproduce non-Gaussian features such as heavily skewed distributions or bimodal distributions that may be present in the historical data (Lall and Sharma 1996, Sharma et al 1997), which is one of the major drawbacks of the parametric model.

Nonparametric models have been developed to address these drawbacks of the parametric models. The simplest nonparametric model is the Index Sequential Method (ISM), which involves selecting chunks of historic data. For example, if we have 100 years of historical data, without wraparound we can generate 80 sequences of 20-year lengths, 70 sequences of 30-year lengths, etc. The advantages are that it is simple and easy to implement, assumption free, and can reproduce the entire distributional properties of the historic data: the mean, variance, auto-correlation, etc. The main disadvantage is that only historically observed sequences can be generated.

Recently developed nonparametric models, (Lall 1995, Lall and Sharma 1996, Tarboton et al 1998) have tried to address the problems of ISM and parametric models. Several types of nonparametric models exist for streamflow forecasting. These include the kernel based (Sharma et al 1997), nearest neighbor based (Lall and Sharma 1996, Rajagopalan and Lall 1999), modified nearest neighbor based (Prairie et al 2006) and hybrid parametric/nonparametric models (Srinivas and Srinivasan 2001, 2005).

In effect, the nonparametric models estimate the marginal and conditional probability density functions locally, and simulate sequences from them. Nonparametric models, unlike parametric models, are assumption free
and are driven by the data alone and can model any shape of the density function. Nonparametric models do not assume any underlying distribution in the data. No parameter estimations or data transformations are necessary.

The choice of a stream flow generation model for purpose of planning and operation has very important practical implications in terms of system reliability and investment decisions (Pereira et al 1984). Efficient model selection procedures are essential for the successful application of stochastic models in planning studies. In this work for generating stream flow scenarios, non parametric model recently developed by Srinivas and Srinivasan (2001) is used and compared with the parametric approach.

2.6 SUMMARY

A review of literature related to water resources system analysis models particularly with reference to reservoir operation has been presented. It is seen from the literature that, despite development of new techniques, implicitly stochastic optimisation (ISO) models solved using linear programming (LP) remain one of the most readily applicable method for the analysis of complex systems. However, the inability to capture the rules during some exceptional conditions, such as droughts and floods, limits its application. The classification using association rules are found to be effective in such exceptional conditions and hence proposed for derivation of the reservoir operation rule from the inflow- release pattern.