Journal of the Operations Research
Society of Japan
Vol. 45, No. 1, March 2002

FUZZY PARTITIONING AND ITS APPLICATION TO RESERVOIR OPERATION PROBLEM (A MULTISTAGE APPROACH USING MARKOV CHAIN)

Adel Azar Mehran Hosseiny Farazmand Ali Asghar Anvary Rostamy
Tarbiat Modarres University

(Received June 19, 2000; Revised July 11, 2001)

Abstract In this paper, we investigated mathematical models on reservoir operation problem and provided a fuzzy model based on Markov chain time series for Karkheh dam in Iran. Based on fuzzy partitioning of monthly streamflows, calculated by time series, a Markovian forecasting model was developed. A deterministic and a fuzzy partitioning stochastic dynamic programming model were formulated for the problem. The goals were formulated with weighting priorities and the optimal reservoir operation was determined. Using historical data, the performance of the dam was simulated. Results of this simulation clearly showed that the fuzzy partitioning proposed stochastic model outperforms the deterministic model.

1. Introduction

After world war II, several mathematical techniques were developed to determine the optimal utilization of reservoirs. Introducing sequential programming algorithm by Little in 50's and developing dynamic programming models facilitated the process of modeling for the operation of reservoirs [4, 8]. The aims of designing these models were to find optimal policies for limited or unlimited time periods. Also, all methods of optimization are classified as either deterministic or stochastic. Deterministic methods use "perfect foresight" of future inflows, while stochastic methods incorporate stochastic models of inflows directly in the optimization process and consider multiple scenarios. Most of these methods were developed based on control theory such as deterministic feedback control and OR such as deterministic/stochastic dynamic/linear programming, mixed integer programming and multiobjective programming [3, 6, 7]. Most of researches use the average historical monthly streamflows and analyze them by ARIMA models or Markov chains.

Since the number of historical data for any given month is limited for practical cases, the classical frequency distribution can not represent the actual behavior of streamflows. The problem is more serious when a large number of new data for updating a classical frequency distribution locate on the borders of classes. In order to overcome the difficulties of the classical models and to improve its effectiveness, fuzzy partitioning is applied. In this paper, fuzzy partitioning model is applied to the Karkheh reservoir in Iran.

2. Fuzzy Partitioning and Markovian Models

Fuzzy sets theory is a powerful tool for more realistic analysis of behavior of human and nature. Fuzzy theory changes the classic logical mathematical structure to a structure with logical continuous values [9]. The idea of fuzzy partitioning stems from fuzzy sets theory. A fuzzy partition includes a given number of fuzzy sets. Each set in turn includes individual elements. These types of fuzzy sets are named an extreme profile. For an element related to each extreme profile, there are a number of grades of membership. So, a fuzzy partitioning is simple to construct since grade of membership scores are constructed relative to the extreme profile. Similar to the bases of vectors in linear spaces, all elements of a fuzzy partition can be expressed as a convex combination of extreme profiles [5]. Based on the definitions of Manton, Woodbury, and Tolley, fuzzy partitioning has the following features:

- For each element in a fuzzy set, there is a grade of membership score (denoted by g_{ik}) that represents the degree to which the element i belongs to the kth sets.
- In order to have a fuzzy partition with k fuzzy sets over the set including all elements of i, the following conditions should be satisfied:

$$g_{ik} \geq 0 \qquad \text{for each element } i \text{ and fuzzy set } k, \tag{1}$$

$$\sum_k g_{ik} = 1 \qquad \text{for each element } i \text{ and over all fuzzy sets.} \tag{2}$$

where

$$\begin{cases} g_{ik} = 0 & i\text{th element is not a member of the } k\text{th fuzzy set,} \\ g_{ik} = 1 & i\text{th element is a complete member of the } k\text{th fuzzy set,} \\ 0 < g_{ik} < 1 & i\text{th element is a relative member of the } k\text{th fuzzy set.} \end{cases}$$

The cases of $g_{ik} = 0$ or 1 are discussed in classical sets theory.

There are several applications for partitioning in mathematics. In classical statistics, partitioning idea is used to form a classical frequency when a large number of data or observations exists. In practice, frequency of a class of data is obtained by counting the elements of that class in order to estimate population parameters. In statistics, based on the fuzzy partitioning, this estimates can be implemented by using grades of membership. The advantage of this kind of frequency calculation is to better represent the effects of real data that are located on the borders of classes. Comparing to the classical frequency distribution, and specially when sizes of samples are small, fuzzy frequency distribution is smoother.

Suppose X_n denotes an element of time series for monthly streamflows. Then, all observations for the jth month of a year can be defined as

$$A_j = \{X_i \mid i = 12m + j,\ m = 1, 2, \cdots\}. \tag{3}$$

If all observations for the jth month locate between interval $[x_{j\,\min}, x_{j\,\max}]$, then the interval can be divided into a number of cross classes. The membership function of each element relative to each class can be shown as Figure 1:

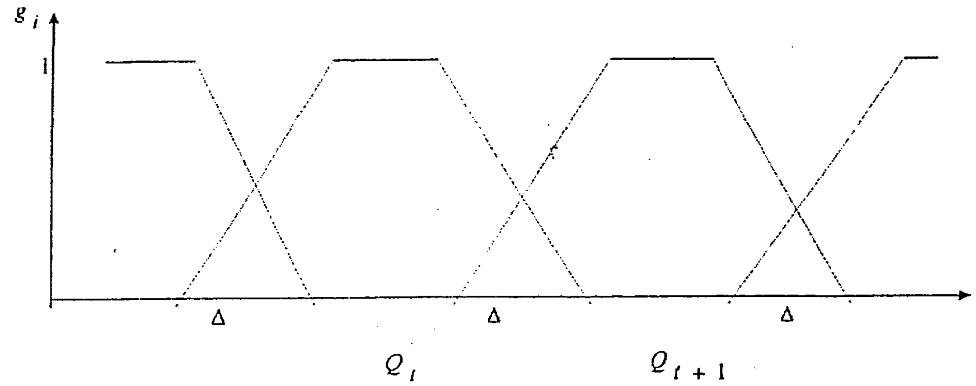


Figure 1: Grade of membership of observations for overlapped classes

The fuzzy statistical model developed by Manton, Woodbury, and Tolley is based on five assumptions [5]. The fourth and the fifth assumptions are:

- **Assumption 4:** The probability of a response l for the jth question by the individual with kth extreme profile is λ_{kjl} (where $\lambda_{kjl} \geq 0$ and $\sum_i \lambda_{kjl} = 1$). It is assumed that at least one, possibly theoretical, individual exists in the kth profile.
- **Assumption 5:** The probability of a response of level l to the jth question by individual i, conditional on the g_{ik} scores, is given by the bilinear form $P_{ijl} = \sum_k \lambda_{kjl} g_{ik}$.

If the question is "Does the element belong to the jth extreme profile?" and if it has two possible answers "yes" or "no", then for "yes", $\lambda_{kjl} \neq 0 = 1$, where $k = j$. Therefore, the probability of "the ith element being a member of the jth extreme profile" will be $P_{ijl} = g_{ij}$. Partitioning all sets of A_j (the historical data for jth month), time series for streamflows can be considered as Markov chain and it will be possible to calculate point estimation of transition probabilities. These are similar to the typical conditional probabilities used in reservoir operation models.

Consider fuzzy partition A_j with related fuzzy sets A_{jk} for time series X_n. Let transition probability from partition A_{jk} to partition $A_{(j+1)l}$ be P^j_{kl}. If we select an element $X_i \in A_j$, randomly and if $X_{(i+1)} \in A_{(j+1)}$ be the next historical data, then

$$P(X_i \in A_{jk}) = g_{ik} \quad \text{and} \quad P(X_{i+1} \in A_{(j+1)l}) = g_{(i+1)l}, \tag{4}$$

$$P(X_i \in A_{jk}, X_{i+1} \in A_{(j+1)l}) = g_{ik} g_{(j+1)l}. \tag{5}$$

Here g_{ik} and $g_{(i+1)l}$ denote the grades of membership of elements X_i and X_{i+1} for extreme profiles A_{jk} and $A_{(j+1)l}$, respectively. The point estimation of transition probability from the kth level of the jth month to the lth level of $(j+1)$, P^j_{kl}, can be estimated as:

$$P^j_{kl} = P(A_{(j+1)l} \mid A_{jk}) = \frac{P(A_{(j+1)l} \mid A_{jk})}{P(A_{jk})} = \frac{E(P(X_i \in A_{jk}, X_{i+1} \in A_{(j+1)l}))}{E(P(X_i \in A_{jk}))},$$

$$\frac{\left(\sum_i g_{(i+1)l} g_{ik}\right)/N}{\left(\sum_i g_{ik}\right)/N} = \frac{\sum_i g_{ik} g_{(i+1)l}}{\sum_i g_{ik}}. \tag{6}$$

Now, it is possible to calculate conditional probabilities and apply them to the model.

3. The Model

The Karkheh river stems from the western mountains of Iran. After going 900 kilometers through five provinces, it ends at Hourolazim area. The Karkheh dam is the biggest dam in Iran. It is in the west of Andimeshk city with capacity of 7.6 billion cubic meters. The goals of building Karkheh reservoir were:

- to prepare about 3.3 billions cubic meters water for about 320000 hectare farm lands,
- to generate 934 gigabits (400 megawatt) of power (energy),
- to control destructive flood water,
- to prepare necessary water for about 40000 hectares of Dasht-e Abbas lands.

Historical streamflow data exist for the last 40 years. The data range from 20 to 1320 cubic meters. The results of a data analysis indicate that the streamflows consist of the following two components:

$$Q_t = S_t + X_t, \tag{7}$$

where Q_t, S_t and X_t denote natural streamflows, the average of streamflows for a given month over the last 40 years, and the deviations from averages, respectively. Time series

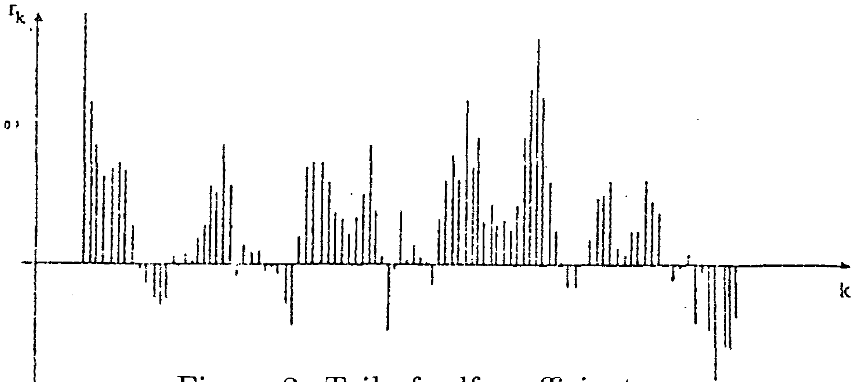


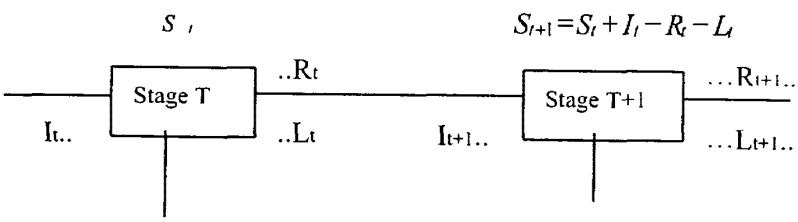
Figure 2: Tail of self-coefficient

analysis shows that X_t has a tail of self-coefficient as shown in Figure 2. Also, based on the Box-Pierce test, an AR(1) Markov process model is suitable.

The least square estimation of parameters for the regress process will be $X_t = 0.57715X_{t-1}$, so that it will be possible to define time series by Markov chain. Moreover, based on the fuzzy partitioning for each month, we can classify data as follows:

$$g_{ik} = \begin{cases} \dfrac{X_i - (100k - 120)}{40} & (100k - 120) < X_i < (100k - 80), \\ 1 & (100k - 80) \leq X_i \leq (100k - 20), \\ \dfrac{(100k + 20) - X_i}{40} & (100k - 20) \leq X_i \leq (100k + 20). \end{cases} \tag{8}$$

From (8), the length of each class is 140 units and there are 40 units overlap for each pair of classes. Each data (with a grade of membership value greater than 0) belongs to at least one or at most two classes. Transitional probabilities between different classes and for sequential months can be calculated as $\frac{\sum_i g_{ik} g_{(i+1)l}}{\sum_i g_{ik}}$. The sequential structure for the reservoir operation model can be shown as follow:



The relations and the constraints are as follows:

1. The upper and lower limits for the volume of water in reservoir are

$$S_{\min} \leq S_t \leq S_{\max}. \tag{9}$$

2. The continuity equation that expresses changes in the volume of water in reservoir is

$$S_{t+1} = S_t + I_t - R_t - L_t. \tag{10}$$

3. Empty space that should always exist in reservoir in order to control probable destructive flood water in the future is

$$S_t \leq S_{t\ \max} - S_{\text{con}}. \tag{11}$$

4. The relation between power and energy for the station is

$$K = \frac{9.81 \times QH_\varepsilon}{3600}, \tag{12}$$

where $H > H_{\min}$. In order to formulate an objective function, the following variables and parameters are defined:

Variables:

$R_{t1} =$ monthly released volume of water to lands at the lower section of the dam,
$R_{t2} =$ monthly released volume of water to the Dasht-e Abbas lands,
$I_t =$ monthly input volume of water to the lake of dam,
$L_t =$ monthly vaporized volume of water from the lake of dam,
$H_t =$ height of water at the lake of dam from the turbine tunnel.

Parameters:

$R_{\text{operate}} =$ minimum volume of water required for the turbine operation (150 cubic meters per second),
$Q_{t1} =$ monthly expected output volume of water from the dam toward the lower section,
$Q_{t2} =$ monthly expected output volume of water from the reservoir toward the Dasht-e Abbas lands,
$H_{\min} =$ minimum height of water at the lake of dam from the turbine tunnel,
$R_{\max} =$ maximum volume of water that can be released to the lower section (monthly 1000 cubic meters per second and 2.592 billions cubic meters per month),
$S_{\max} =$ maximum volume of reservoir (7.5 billion cubic meters),
$S_{\text{control}} =$ the volume of water space to control water more than 1000 cubic meters per second (75 million cubic meters).

According to priorities and constraints, a cost function with several components and different weights can be defined [3]. To do this a total cost function is formulated with three main terms. The cost function is defined as:

$$F(t) = F_1(t) + F_2(t) + F_3(t). \tag{13}$$

The cost function elements are as follows:

- In order to avoid flooding water at the lower lands, the constraint $R_{t1} \leq R_{\max}$ should be satisfied in the 11$^{\text{th}}$ and 12$^{\text{th}}$ months of Iran (middle of February–middle of March), specially $S_t \leq S_{\max} - S_{\text{control}}$ should be satisfied. Any deviation from the above considerations is shown in the following cost function

$$F_1(t) = \begin{cases} M^3 \times \min\{(R_{\max} - R_{t1}), 0\} + 0.1 \times M^3 \times \min\{(S_{\max} - S_{\text{control}} - S_t), 0\} & \text{if } t = 11, 12, \\ M^3 \times \min\{(R_{\max} - R_{t1}), 0\} & \text{otherwise.} \end{cases}$$

  In the above relation, the value of 0.1 is based on the probability of having a flood water. If the required space in reservoir is not enough, the cost will be very high.

- In order to supply enough water for lower lands, R_{t2} should be equal to Q_{t2}. If $H_t > H_{\min}$, then $R_{t2} = Q_{t2}$. Any deviation from these conditions is reflected in the following cost function:

$$F_2(t) = \begin{cases} M^2 \times \min\{(R_{t1} - Q_{t1}), 0\} - M^2 \times R_{t2} & \text{if } H_t < H_{\min}, \\ M^2 \times \min\{(R_{t1} - Q_{t1}), 0\} & \text{otherwise.} \end{cases}$$

- To prevent power disgeneration in the power station and to utilize its full capacity, we must have $H_t > H_{\min}$ and $H_t R_{t1} \geq (H_t \times R_{t1})_{\min}$. Any deviation from these conditions is reflected in the next cost function

$$F_3(t) = \begin{cases} -M \times R_{\text{operate}} & \text{if } H_t < H_{\min}, \\ M \times R_{\text{operate}} \times \min\left\{\left(\frac{H_t \times R_{t1}}{(H_t \times R_{t1})_{\min}} - 1\right), 0\right\} & \text{otherwise.} \end{cases}$$

In addition, the continuity equation of reservoir is

$$S_{t+1} = S_t + I_t - R_{t1} - R_{t2} - L_t(S_t). \tag{14}$$

In a dynamic model for reservoir operation, the state of the system at any stage is determined by (S_t), which is the volume of water in the reservoir in the beginning of a month, and (I_{t-1}) is the average streamflow of the previous month. The volume of water is divided into 50 stages and the average streamflow for the previous month is partitioned as in Figure 1.

Using the forecasting relation and having streamflows for the previous month in deterministic case, the streamflows for a given period in the future can be estimated. The optimal policy for current stage was determined by solving the model. The regress relation for deterministic dynamic model is:

$$f_n^*(S_n) = Opt_i\{c(S_n, X_{ni}) + f_{n+1}^*(S_{n+1})\} \tag{15}$$

where

S_n: the position of the stage n,

X_{ni}: various decisions at stage n (released water),

$c(S_n, X_{ni})$: the additional costs of stage n as a consequence of making decision X_{ni},

f^*: minimum cost from the current stage to the next stages.

Assume $S_{12} = S_0$, as the annual equilibrium performance equation, we analyzed the deterministic dynamic model to obtain the optimal policies $TG(j, S_t, I_{t-1})$ for the first state of (S_t, I_{t-1}) and the jth month. For stochastic model, the backward method with infmite stages is used so that the optimal policy for each month over years is only depends on the first state (S_t, I_{t-1}). The recursive relation for stochastic dynamic model can be explained as:

$$f_t^*(S_t, I_{t-1}) = Opt_{i\ st}\left\{\sum_k Pr\left(I_t = i_k \mid I_{t-1} = i_{t-1}\right)\left[c(S_t, I_t) + f_{t+1}^*(S_{t+1}, I_t)\right]\right\}. \tag{16}$$

In the model we used conditional probability for a given streamflow. It is also assumed that the streamflow for current month is unknown. In our case the last equation is applied to determine the optimal policy in terms of the first state, (S_t, I_{t-1}) and for each month in the form of $TG(j, S_t, I_{t-1})$ [1].

Comparisons of results between deterministic and fuzzy partitioning stochastic models implemented for four possible situations and for different months are shown in Table 1. The measurement unit is 10 cubic meter per second. An S in Table 1 indicates that the result of the fuzzy partitioning stochastic model is meaningfully less than that of the deterministic model. A D, on the other hand, indicates that the result of the deterministic model is meaningfully less than that of the fuzzy partitioning stochastic model. From Table 1, it can be concluded that the fuzzy partitioning stochastic model proposes less released water for lower lands, except for the 12th month. Only for the situation with high volume of water in the 12th month, the stochastic model suggests more released water for lower lands.

Table 1: $D1, D2, D3, D4$ denote the deviations for four situations. $(1/3S_{\max} < S_t < 2/3S_{\max})$ is for the general situation of the reservoir, $(S_t < 1/3S_{\max})$ is for the situation that reservoir is relatively empty, and $(S_t > 2/3S_{\max})$ when reservoir is relatively full. $C11, C12, C13, C14$ are confidence intervals.

Month	Less	$D1$	$C11$	Less	$D2$	$C12$	Less	$D3$	$C13$	Less	$D4$	$C14$
1	S	-3.69	(-.8053, .8053)	S	-1	—	S	-1.32	(-.17, .17)	S	-9.07	(-2.77, 2.77)
2	S	-2.42	(-.5088, .5088)	S	-1.9	(-.1736, .1736)	S	-.91	(-.375, .375)	S	-4.53	(-1.4714, .4714)
3	S	-1.9	(-.2982, .2982)	S	-1.15	(-.1287, .1287)	S	-.82	(-.3224, .3224)	S	-3.85	(-.722, .722)
4	S	-1.715	(-.21678, .2168)	S	-1.41	(-.3341, .3341)	S	-1.67	(-.549, .549)	S	-2.09	(-.07345, .07345)
5	S	-1.59	(-.2649, .2649)	S	-1.35	(-.3729, .3729)	S	-1.4	(-.71, .71)	S	-2	—
6	S	-1.7	(-.2103, .2103)	S	-2	—	S	-1.75	(-.635, .635)	S	-1.312	(-.1698, .1698)
7	S	-1.47	(-.2064, .2064)	S	-1.77	(-.1351, .1351)	S	-1.563	(-.212, .212)	S	-1	—
8	S	-2.26	(-.3719, .3719)	S	-2.6	(-.5735, .5735)	S	-2.9	(-.155, .155)	S	-1.156	(-.207, .207)
9	S	-3.43	(-.4186, .4186)	S	-3.2	(-.5621, .5621)	S	-4.52	(-.944, .944)	S	-2.75	(-.6348, .6348)
10	S	-3.49	(-.471, .471)	S	-3.79	(-.5376, .5376)	S	-5.39	(-.85, .85)	S	-1.22	(-.7912, .7912)
11	S	-10.44	(-.838, .838)	S	-8.68	(-.6773, .6773)	S	-9.34	(-.855, .855)	S	-13.62	(-2.26, 2.26)
12	D	4.01	(-1.46, 1.46)	S	-1.8	(-.2349, .2349)	S	-2.259	(-.416, .416)	D	-16.71	(-3.59, 3.59)

4. Results of Simulation

Having the historical data and suggested policies by the two models, it is possible to simulate the performance of the reservoir with respect to each model. Assuming equilibrium situation at the first state, the results are as follows:

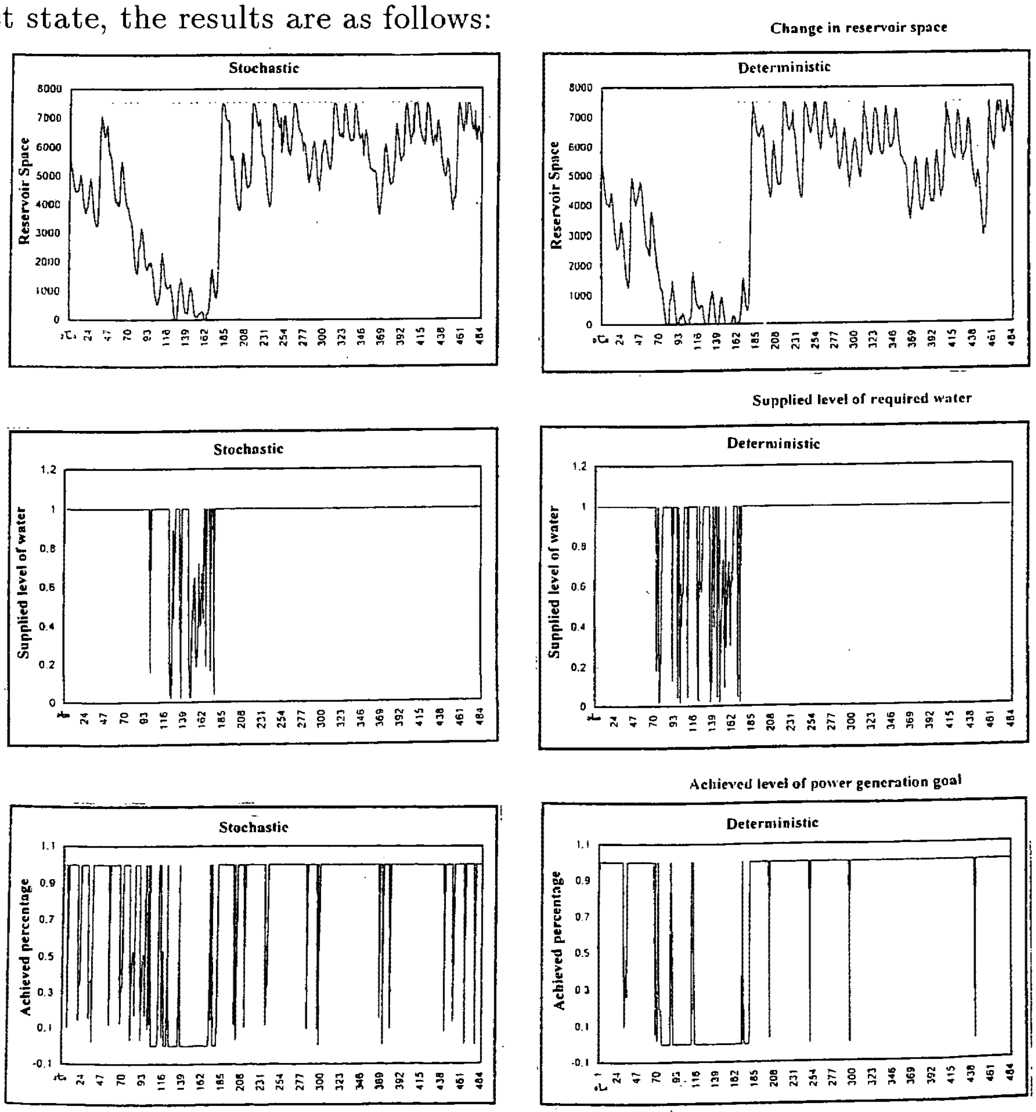


Figure 3: Simulation of system performance over 40 years

1. Figure 3 shows changes in volume of water in reservoir. In Figure 3, the scale of volume in reservoir is in million cubic meters per month. It is clear that for the years that the average streamflows are less than 178 cubic meters per second, there are low water situation. Comparing to the deterministic model, the stochastic model provides fewer number of months in which the water of the reservoir is less than the minimum level (1100 millions cubic meter). Inadequacy in water reservoir causes the stochastic model for 0.39% of months and the deterministic model for 2.09% of months do not supply required policies, successfully.

Table 2: Comparison the results of the deterministic and fuzzy partitioning stochastic models

Camparing Criteria	Fuzzy Partitioning Stochastic Model	Deterministic Model
number of months in which the water of the reservoir are less than the minimum level	fewer number ofmonths in which the water of the reservoir are less than the minimum level (0.39% of months)	number of months in which the water of the reservoir are less than the minimum level is higher than fuzzy partitioning stochastic model (2.09% of months)
supplying required water to lower lands.	lack ofwater for Dasht-e Abbas is about 44 months. (degree of success =95.67%)	lack of water for Dasht-e Abbas is 76 months. (degree of success =94.4%)
goal achievement degree for the power generation objective	the rate of goal achievement is 78.3%	the rate of goal achievement is 77.1 %
volume of water that is required to control flood water	for 2 months can not satisfy the volume of water that is required to control flood water	for 7 months could not satisfy the volume of water that is required to control flood water
the damages creates by implementing each of the models over 40 years	total costs is lower (lower damages)	higher costs or higher damages

2. Table 2 shows the degree of success of supplying water to lower lands. This degree is equal to 1 when the volume of released water is greater than the required volume. Otherwise, the degree of success is equal to the ratio of released volume to the required volume. The result shows that the stochastic model outperforms the deterministic model in terms of supplying required water to lower lands. The degree of success is 95.67% for the stochastic model and 94.40% for the deterministic model. Given the stochastic model, the lack of water for the Dasht-e Abbas is about 44 months. This number is 76 months for the deterministic model.
3. Goal achievement degree for the power generation objective is shown in Table 2. The rates are 78.3% and 77.1% for the stochastic and the deterministic models, respectively.
4. Both models are successful in controlling flood water. However, the stochastic model for 2 months and the deterministic model for 7 months could not satisfy the volume of water that is required to control flood water.
5. The damages created by implementing each of the models over 40 years. The damages are estimated by the first component of equation (13). The stochastic model shows better results.

5. Summary and Final Remarks

In this paper, a fuzzy model based on the Markov chain time series was suggested for Karkheh dam in Iran. Using fuzzy partitioning of monthly streamflows, a Markovian forecasting model was developed. A deterministic and a fuzzy partitioning stochastic dynamic programming models were formulated for the problem. The goals were formulated by weighting priorities and the optimal utilization policy for reservoir operation was determined. Using historical data, the performance of the dam is simulated. The results of simulation clearly showed that the proposed fuzzy partitioning stochastic model outperforms the deterministic model.

References

[1] I.C. Goulter and T.F. Tai: Practical implication in the use of stochastic dynamic programmmg for reservoir operation. *Water Resource Bulletin*, **21**-1 (1985) 65–74.

[2] T. C. Hellendar: Defuzzification in fuzzy controllers. *Journal of Intelligent & Fuzzy Systems*, **1** (1993) 109–123.

[3] Q. Liang, J.E. Johnson and Y.S. Yu: A comparison of two methods for multiobjective optimization for reservoir operation. *Water Resources Research*, **32**-2 (1996) 333-340.

[4] D.P. Loucks, J.R. Stedinger and D.A. Haith: *Water Resource Systems Planning and Analysis* (Prentice Hall, 1981).

[5] K.G. Manton, M.A. Woodbury and H.D. Tolley: *Statistical Applications Using Fuzzy Sets* (John Wiley and Sons Inc, 1994).

[6] C. Russ, J.R. Philbrick and P.K. Kitanidis: Limitation of deterministic optimization applied to reservoir operations. *Journal of Water Resource Planning and Management*, **125**-3 (1999) 135–142.

[7] K. Srinivasan, T.R. Neelakantan, P. Shyam and C. Nagarajukumar: Mixed integer programming model for reservoir performance optimization. *Journal of Water Resource Planning and Management*, **125**-5(1999) 298–301.

[8] S. Yakowitz: Dynamic programming applications in water resources. *Water Resources Research*, **18**-4 (1982) 673–696.

[9] H.J. Zimmermann: *Fuzzy Set Theory and Its Applications* (Kluwer-Nijhoff Publishing, U.S.A., 1984)

Adel Azar
Department of Management
Tarbiat Modarres University
Tehran, Iran
E-mail: Adel_Azar2000@yahoo.com

Ali Asghar Anvary Rostamy
Iran Management & Productivity
Study Center (IMPSC)
Tarbiat Modarres University
Tehran, Iran
E-mail: amin_armin@yahoo.com

Mehran Hosseiny Farazmand
Iran Management & Productivity
Study Center (IMPSC)
Tehran, Iran