Deriving Reservoir Operating Rules via Fuzzy Regression and ANFIS

Seyed Jamshid Mousavi[1], Ebrahim Jabbari[1], K. Ponnambalam[2], and F. Karray[2]
1 Structures and Hydro-Structures Research Centre, Civil Engineering Department, University of Science and Technology, Iran, Fax: +9821-745-4053,
e-mail: jmosavi@iust.ac.ir & jabbari@iust.ac.ir
2 Department of Systems Design Engineering, University of Waterloo, Canada,
e-mail: ponnu@uwaterloo.ca & karray@uwaterloo.ca

Abstract
The methods of ordinary least-squares regression (OLSR), fuzzy regression (FR), and adaptive network fuzzy inference system (ANFIS) are compared in inferring operating rules for a reservoir operations problem. Dynamic programming (DP) is used as an optimization tool to provide the input-output data set to be used by OLSR, FR, and ANFIS models. The OLSR, FR, and ANFIS based rules are then simulated and compared. The methods are applied to a long-term planning problem as well as to a medium-term implicit stochastic optimization model. The results indicate that FR is useful to derive operating rules for a long-term planning model, where imperfect and partial information is available. ANFIS is beneficial in medium term optimization as it is able to extract important features of the system from the generated input-output set.
Keywords: Reservoir operation, operating rules, fuzzy regression, fuzzy inference.

1 Introduction

Fuzzy regression (FR) and adaptive-network-based FIS (ANFIS) are used in inferring operating rules for reservoir operations. A summary of different methods used for surface reservoir management can be found in Yeh [1985] and Ponnambalam [2002]. Inferring operating rules is to get general rules by which reservoir operations can be controlled. Young [1967], Bhaskar and Whitlach [1980] and Karamouz and Houck [1982] used multiple linear regression to derive the operating rules. Other methods included using the Artificial Neural Networks (ANN) and fuzzy rule-based technique. In this study fuzzy regression, first introduced by Tanaka et al. [1982], is examined in deriving operating rules for reservoir operations. Two problems are addressed here, a long-term optimization problem and an implicit stochastic optimization model. The OLSR, FR, and ANFIS are used to derive operating rules and they are then simulated and compared.

2 Methods of inferring operating rules

Let us assume that Dynamic Programming (DP) is used to derive optimal releases for a sequence of inflows, as here. Once the optimal releases are available, the problem is to derive an operating policy for any feasible storage and inflow from these releases.

Given the various inflow scenarios and the corresponding optimal releases and storage volumes, a multiple regression analysis can be used in the following manner.

$$r_t = ai_t + bs_t + ci_t^2 + ds_t^2 \quad (1)$$

where r_t is the release and i_t is the inflow during period t and s_t is the beginning storage volume.

2.1 Fuzzy Regression (FR)

Fuzzy regression (FR) was proposed to deal with fuzzy data and we will describe this method first with a tutorial example. In FR, the estimation errors can be viewed as the fuzziness of the model structure. Suppose a simple linear relation as:

$$\hat{y} = a_0 + a_1 x \quad (2)$$

where $\hat{y}$ is the model predicted value and x is a crisp observed value. The FR problem is to find fuzzy coefficients $a_0[c(a_0), w(a_0)]$ and $a_1[c(a_1), w(a_1)]$ so that the widths of the observed fuzzy values $y(s)$ is within the widths of the predicted fuzzy values $\hat{y}(s)$ at a confidence level $h \in [0,1]$ (see Figure 1). This implies that the interval DE in Figure 1 has to be within the interval BC. Note that $c(a_i)$ and $w(a_i)$ are the center point and half width of fuzzy number a_i for a triangular symmetrical fuzzy number.

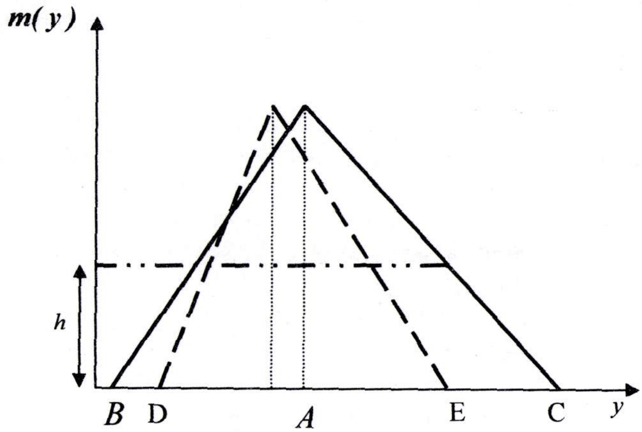

Figure 1: Interpretation of fuzzy regression

2.1.1 Minimum-fuzziness-based FR Models

Therefore, the total fuzziness of FR model relates to the width of fuzzy coefficients used in the model as they are multiplied by the observed values. Hence, the problem of finding the parameters of Equation (2) leads to the solution of an LP with an objective function minimizing the total fuzziness of the model. This LP model is as follows:

$$\textit{Mnimize } z = n.w_0 + w_1.\sum_{j=1}^{n}|x_j|$$

$$s.t: \sum_{i=0}^{\ell} c_i(a_i).x_j + (1-h).\sum_{i=0}^{\ell} w_i(a_i).|x_j| \geq y_j + (1-h).w_j(y_j) \quad \forall j = 1,..n \tag{3}$$

$$\sum_{i=0}^{\ell} c_i(a_i).x_j - (1-h).\sum_{i=0}^{\ell} w_i(a_i).|x_j| \leq y_j - (1-h).w_j(y_j) \quad \forall j = 1,..n$$

$$c_0(a_0), c_1(a_1) \geq 0$$

where ℓ is the number of independent variables and n is the number of data pairs and h is called the credibility level. In the constraints of the above LP model, $c_i(a_i).x_j$ is the center point of the predicted values (point A in Figure 1), a prediction which would be reached to if a non-fuzzy regression equation was used. The second term $\pm(1-h).\sum_{i=0}^{\ell} w_i(a_i).|x_j|$ represents the fuzziness of the predicted values shown by points C and B in Figure 1, when they are added to the center point A. The right hand sides of the constraints are the maximum and minimum values of a fuzzy observed y_i (points E and D in Figure 1), respectively. Bardossy et al. [1995] used another formulation. They considered more general nonlinear, nonsymmetrical LR type fuzzy coefficients (Figure 2) as follows:

$$L(x) = R(x) = (1 - x^p)$$

$$m_{a_i}(x) = L\left[\frac{(c(a_i) - x)}{w_l(a_i)}\right] \qquad \textit{for } x \geq w_l(a_i) \tag{4}$$

$$m_{a_i}(x) = R\left[\frac{(x - c(a_i))}{w_r(a_i)}\right] \qquad \textit{for } x \leq w_r(a_i)$$

where p takes integer values and $m(a_i(x))$ is the membership degree of a_i. Also $w_l(a_i)$ and $w_r(a_i)$ are the left and right widths of a fuzzy number with center of $c(a_i)$.

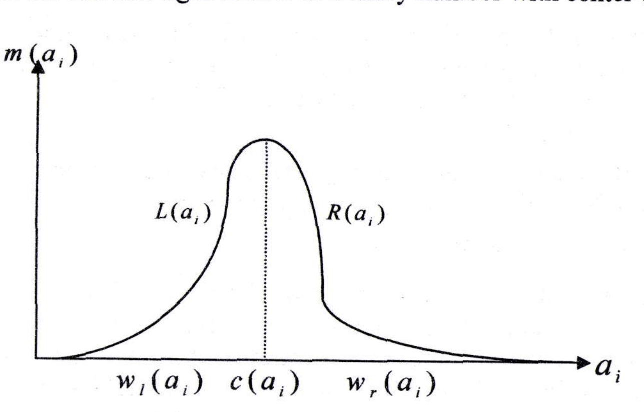

Figure 2: An LR fuzzy number

2.2 Fuzzy inference system (FIS)

Operating rules can be represented by fuzzy "IF-THEN" rules in a Takagi-Sugeno FIS as follows:

$$\text{If } s_t \text{ is } v_{1i} \text{ and } i_t \text{ is } v_{2i} \text{ then } r_t = b_1.i_t + b_2.s_t + b_3$$

where, v_{ki} is the value of the k^{th} explanatory variable and b_i are the parameters of the consequence part of rule "i". Each value of an explanatory variable is represented by a fuzzy set. Therefore, a_{ki} is a fuzzy set. The parameters b_i are estimated using available data or operator experience. There is no systematic way to know what type of membership functions of premise variables is the best in a defined FIS. An efficient way for doing this is using artificial neural nets trained by input-output data like ANFIS [Jang, 1993]. ANFIS uses an initial FIS and tunes it with a back-propagation algorithm.

3 Implementations

In this study, Dez reservoir located in Iran has been selected to which the methods are applied. For the implicit stochastic model used in this study, a time series model was fitted to the historical inflows. Then it was used to generate a large number (1000) of scenarios of monthly inflows. Using either the historical or the generated inflows in DP, the optimal sets of reservoir storage and release volumes are obtained corresponding to the long-term deterministic model in the first case and the medium term implicit stochastic model. These optimal sets are then used in inferring operating rules using OLSR, FR, and ANFIS. The inferred operating rules are then simulated to see how they perform.

3.1 Long-term planning optimization

3.1.1 Implementation of the FR model

The optimal storages and releases along with historical inflows were used in the FR model to infer the general operating rules. For each month an FR model was built separately. The regression equation used is as follows:

$$r_t = a_0 + a_1 i_t + a_2 s_t \qquad (5)$$

Taking forty years of optimal storage and release sets obtained by DP along with the historical inflows, the FR was applied to the problem.

3.1.2 Implementation of ANFIS

ANFIS is used to extract the relation of storage, inflow, and release variables from the data pairs obtained by DP and represent them as fuzzy if-then rules. The premise part of fuzzy if-then rules is reservoir inflow and storage volumes. The consequent part is the release volume. For each month, an ANFIS model was developed separately.

To make simulation results independent of initial storage volume, reservoir operation was first simulated by OLSR with different initial reservoir storages. These simulated paths and the original DP optimal path were used for training ANFIS. Thus, when comparing different methods, ANFIS results are the results of this retrained ANFIS. Figure 3 shows the mean square error (MSE) of predicted values by the preliminary FIS, ANFIS, OLSR, and FR for different months. In terms of fitness capability, ANFIS is performed superior to other methods. However, the important thing is to analyse their performance in simulation.

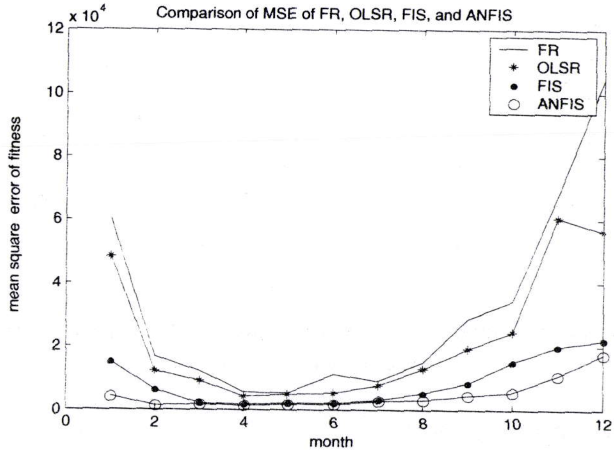

Figure 3: Comparison of mean square error of different models to fit optimal data (long term problem)

3.2 Medium-term implicit stochastic model

In the second problem, we derive operating rules from input-output data obtained from a Monte-Carlo DP model. DP model was solved for 300 synthetic scenarios. Each scenario has one year of horizon with 12 monthly time steps. The final storage at the end of each year was forced to be more than half of the reservoir capacity.

3.2.1 Inferring operating rules

As was done in the long term planning model, OLSR, FR, and ANFIS parameters are determined based on the input-output pairs obtained by Monte-Carlo DP. Different optimal paths of DP, for different initial storage were used in ANFIS training; this is to overcome the difficulty of storage falling out of the training ranges in simulation. Table 1 presents the results for comparing the overall performance of OLSR, FR, FIS, and ANFIS rules in terms of their fitting capability. Also the monthly distribution of MSE of predicted values is illustrated in Figure 4. These results imply that ANFIS performs quite superior to other methods in terms of fitting. However, the simulation will show the real success of the methods.

Table 1: Comparison of inferred operating rules in terms of fitted input-output data

Method	OLSR	FR	FIS	ANFIS
MSE	10142	13081	3781	1703

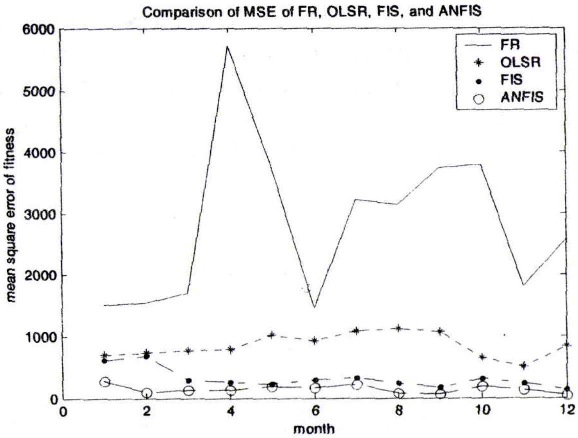

Figure 4: Monthly distribution of MSE of predicted values by different methods (medium term problem)

4 Results and comparisons

The OLSR, FR, and ANFIS are compared based on simulating the reservoir operations using corresponding policies. The simulation performance indices are computed. The indices selected are: mean value of the simulated objective function (LOSS), coefficient of variation of LOSS (CV), and the percentage of the time in which the release is greater than 0.9 of demand (REL).

4.1 Long term operations

For comparison purposes, the results of the so-called SOP policy that simply sets release volume equal to demand if possible and the original DP releases (for long term planning only) are included. Table 2 presents the results. As is clear in Table 2, mean values of the simulated objective function of the OLSR, FR, and ANFIS methods are more or less close while FR has the smallest (best) LOSS value. In terms of reliability of meeting the water demand, the SOP policy is the best. SOP produces severe shortages or spills over the simulated horizon and hence the larger LOSS value. FR has the smallest coefficient of variation (CV) of the objective function. Overall, the FR is the best and its results are the closest to the DP results, although it is the worst in terms of the MSE. Several cases are found when, for example, just 20 years of streamflow record is available and it is hard or inaccurate to extend the record using hydrological methods. On the other hand, planning and design of the projects need to be done for 40 or 50 years of their projected life. To examine whether FR is useful in such a situation, the results were repeated with either half or one-third of the streamflow record for estimating parameters, but carrying out the simulation for the whole forty years. ANFIS was not able to tackle this case because of the lack of data for training. Therefore, only

OLSR and FR are compared in this case. The results are in Table 3. Again, the FR performs quite well especially in terms of the value of simulated objective function.

Table 2: Comparison of simulation performance of OLSR, FR and ANFIS, long term planning model

Method	Simulated objective function		MSE	REL
	LOSS	CV		
OLSR	2.0908e5	3.91	2.7547e5	88.6
FR	1.9821e5	3.33	3.8433e5	84.6
ANFIS	2.0121e5	3.80	8.5320e4	90.6
SOP	3.4946e5	3.52	-	98.1
DP	1.4393e5	2.42	0.0	94.8

Table 3: Comparison of OLSR and FR where partial data are used, long term planning model

Method	Scenario with half parts of data			Scenario with one-third parts of data		
	LOSS	CV	REL	LOSS	CV	REL
OLSR	2.6079e5	4.12	90.8	2.8945	3.86	90.8
FR	2.1398e5	4.06	85.8	2.0543e5	3.53	83.8
SOP	3.4946e5	3.52	98.1	3.4946e5	3.52	98.1
DP	1.4393e5	2.42	94.8	1.4393e5	2.42	94.8

4.2 Medium term operations

For this problem, simulation was carried out using 300 synthetically generated scenarios not used in training stage. The results are presented in Table 4. According to Table 4, unlike the planning model, ANFIS is superior to other methods and it performs well both in fitting and simulation modes. It is obvious that SOP is the best in terms of the reliability, but its simulated loss as the most important index is much higher than other policies. The results of the implicit stochastic model show that the performance of the FR is better than OLSR model but not as good as ANFIS. As in long-term planning, we are interested to the effect of a reduced number of the scenarios used in the optimization model and the parameter estimation of the inferring methods. The results are in Table 5 for 50 and 30 scenarios, respectively. As we see, there is not a significant difference between the cases with 300 and 50 or even 30 scenarios.

Table 4: Comparison of simulation performance of OLSR, FR and ANFIS, implicit stochastic model

Method	Simulated objective function		REL
	LOSS	CV	
OLSR	59152	3.24	84.4
FR	52704	3.55	82.4
ANFIS	47136	2.45	84.8
SOP	103280	2.89	93.6

Table 5: Comparison of OLSR and FR where partial data are used in parameter estimation implicit stochastic model

Method	With 50 generated scenarios			With 30 generated scenarios		
	Loss	CV	REL	Loss	CV	REL
OLSR	57024	3.53	85.7	59650	3.54	85.9
FR	53446	3.48	87.2	53660	3.47	87.8
ANFIS	47783	3.38	83.5	47252	3.42	84.5
SOP	103280	2.89	93.6	103280	2.89	93.6

4.3 Sensitivity Analysis

The most important factor in an FR is the degree of fuzziness considered for observed values of dependent variable. Take the FR defined in Equation (3); if we simply set $h=0$, the fuzziness of y_i will be defined by $\Delta y_i * y_i = w_i(y_i)$. The variation of simulated objective function (LOSS), with the variation of Δy showed that for $\Delta y = 10\%$, the LOSS is the lowest but the other LOSS values are not too far off. As the degree of fuzziness increases, the degree of fuzziness of the dependent variable y considered in the FR model should be increased too. Earlier we considered the case in which the uncertainty of the model is increased due to lack of data. In this case, the combined effect of randomness of inflows and imprecision of discretization and initial storage effect will provides the possibility for simulated releases being far from the releases to be suggested by regression models. Therefore, it is rational to use an FR model assuming a higher degree of fuzziness of dependent variable y reflected in Δy. Examinations showed that where half parts of data are used in the parameter estimation mode, the best degree of fuzziness for y is around 40 percent and where only one third of data is used; it is around 70 percent using LOSS as the criterion.

We close this section by raising this question: when can a complicated method like ANFIS be useful? We realized that the capability of ANFIS in terms of fitness to data is much better than OLSR and FR. However, ANFIS did not perform well in the long-term planning problem while it performed quite well in the implicit stochastic problem. ANFIS is powerful in terms of fitting and extracting nonlinear and ill-defined relations that may exist in the data. Therefore, if the knowledge and information required are reflected in the input-output data, ANFIS will be able to extract and convert them to its knowledge-based core–fuzzy rule base. Thus, if the input-output data include the information on uncertainty and imprecision of the real world system to be controlled by ANFIS, ANFIS is expected to be successful. However, if those aspects of the system behaviour did not exist in the input-output data or even they existed but are purely white noise, then linear rules are enough to represent them.

5 Conclusion

The problem of inferring operating rules using an ordinary regression, a fuzzy regression, and an adaptive neuro-fuzzy model explored. These methods tested in two types of models, a long-term planning and a medium-term implicit stochastic optimization. For the long-term planning model, the fuzzy regression showed promising results, especially in the situations where the length of streamflow record is limited. For the implicit stochastic optimization model since the main information by which the reservoir should be operated is within the data set, ANFIS performed superior to other methods. Results indicate that the fitness error of the models used in deriving operating rules does not necessarily show how well they will perform while simulating their policies.

References

Bardossy A., Bogardi, I., and Duckstein, L. (1995) Fuzzy Regression in Hydrology. Water Resources Research, 26 (7), pp. 1497-1508.

Bhaskar N. R., and Whitlach Jr. E. E. (1980) Deriving of Monthly Reservoir Release Policies. Water Resources Research, 16 (6), pp. 987-993.

Jang J. S. R. (1993) ANFIS: Adaptive -Network-Based Fuzzy Inference Systems. IEEE Transactions on Systems, Man, and Cybernetics, 23 (3), pp. 665-685

Karamouz M., and Houck M. H. (1982) Annual and Monthly Reservoir Operating Rules. Water Resources Research, 18 (5), pp. 1337-1344.

Ponnambalam K. (2002) Optimization in Water Reservoir Systems. Handbook of Applied Optimization, Edited by Padalos P. M., and Mauricio G. C. R., pp. 933-943.

Tanaka H., Uejiima S. and Asai K. (1982) Linear Regression Analysis with Fuzzy Models. IEEE Transactions on Systems, Man., Cybernetics, 12, pp. 903-907.

Yeh W.-G. (1985) Reservoir Management and Operation Models: A state of the Art Review. Water Resources Research, 21(12), pp.1797-1818.

Young G. K. (1967) Finding Reservoir Operating Rules. Journal of Hydrology, 93 (6), pp. 297-321.