Intelligent control for modelling of real-time reservoir operation

Li-Chiu Chang and Fi-John Chang*

Department of Agricultural Engineering, National Taiwan University, Taipei, Taiwan, 10617

Abstract:

This paper presents a new approach to improving real-time reservoir operation. The approach combines two major procedures: the genetic algorithm (GA) and the adaptive network-based fuzzy inference system (ANFIS). The GA is used to search the optimal reservoir operating histogram based on a given inflow series, which can be recognized as the base of input–output training patterns in the next step. The ANFIS is then built to create the fuzzy inference system, to construct the suitable structure and parameters, and to estimate the optimal water release according to the reservoir depth and inflow situation. The practicability and effectiveness of the approach proposed is tested on the operation of the Shihmen reservoir in Taiwan. The current M-5 operating rule curves of the Shihmen reservoir are also evaluated. The simulation results demonstrate that this new approach, in comparison with the M-5 rule curves, has superior performance with regard to the prediction of total water deficit and generalized shortage index (GSI).

INTRODUCTION

Taiwan is known as ‘Formosa’, which means a beautiful and plentiful island. It is located at the subtropical zone with a land area of 36 000 km$^2$. In the last several decades, the constant population growth and the dramatic economic development have resulted in a tremendous demand for natural resources, especially water. However, owing to non-uniform temporal and spatial distributions of precipitation and the presence of many steep channels all over the island, most of the rainfall flows immediately into the ocean and cannot be used directly as a stable source of water supply. Consequently, reservoirs have become the most important and effective human-made water storage facilities for distributing water among different users in Taiwan. They not only provide water supply, hydroelectric energy and irrigation, but also smooth out extreme inflows to mitigate flood or drought. Therefore, it is very important to determine an optimal reservoir operating schedule for effective water distribution.

Several researchers have adapted simulation approaches to solve the aforementioned problem (Yeh, 1985; Simonovic, 1992; Wurbs, 1993; Chang and Chen, 1998). Currently the most common strategy for reservoir operation in Taiwan involves using rule curves, which can be obtained through an intensive simulation approach (Young, 1967) or an effective searching procedure of the genetic algorithm (Oliveira and Loucks, 1997; Chang and Chen, 1998). Although using these rule curves is straightforward, the ranges between the curves are too large to operate the outflow of a reservoir precisely, rendering water usage inefficient. On the other hand, artificial intelligence, which includes artificial neural networks (ANN), genetic algorithms (GA) and intelligent control theory, recently has been applied extensively to the management of water resources. The concepts of artificial intelligence have been studied by many researchers since the late 1950s, and various techniques have been developed for different purposes. For example, the ANN was shown to be...
highly effective for flow prediction (Hsu et al., 1995; Chang and Hwang, 1999; Liong et al., 2000). The GAs have been used to calibrate rainfall–runoff model parameters (Wang, 1991) and to identify reservoir operating rules (Oliveira and Loucks, 1997; Chang and Chen, 1998). Also, intelligent control has been well recognized for its ability to control aerospace systems (Lin and Su, 2000). However, despite its outstanding ability in controlling complex systems, intelligent control has not been studied for guiding reservoir operation.

Intelligent control is a technology that resembles the human thinking process in decision making and strategy learning. The ANN and fuzzy logic are two potential tools in developing an intelligent control engineering model. Through learning by means of parallel and distributed processes, ANN usually improves the performance. In contrast, fuzzy logic is intended to solve highly non-linear control problems and performance robustness. Therefore, an intelligent control that provides these ANN and fuzzy logic features seems appropriate for reservoir operation.

In this paper, we study reservoir operation by using an intelligent control model that includes a genetic algorithm and adaptive-network-based fuzzy inference system (ANFIS). The GA is first applied to searching the optimal reservoir operating histogram, which is used as the training pattern for ANFIS. The ANFIS is then built to estimate the optimal water release based on the current reservoir depth and inflow conditions. The details of the GA and ANFIS as well as the results of their applications are described in the following sections.

INTELLIGENT CONTROL OF RESERVOIR OPERATION

The control system of reservoir operation is illustrated in Figure 1. Each time step (i) corresponds to a period of 10 days, which is a traditional reference time frame in the Chinese agricultural society. According to this scale, each month has three time steps, and every year has 36. The system includes the input dynamic variable \( I_i \), the water demand \( D_i \), and output control variables outflow \( O_i \) or storage of reservoir \( S_i \), where \( i \) is the time step. As more than 80% of the rainfall occurs during the typhoon season from May to October, the overconcentrated rainfall often exceeds the capacity of the reservoirs during that period in Taiwan. Therefore, the reservoir operation often faces two conflicting challenges: providing a stable (or sufficient) water supply in a long drought season and maintaining dam safety during typhoons. For long-term reservoir operation, water shortage is considered to be the main index for assessing reservoir performance. A water surplus, on the other hand, usually is ignored, because the water surplus happens mostly during typhoon events, and lasts for less than 72 h. In this study, we use water shortage to assess the long-term performance of reservoir operation; hence, the objective function of reservoir operation is defined as

\[
\min (Obj Func) = \min \left( \sum_{i=1}^{36} \left[ \max \left( 0, \frac{D_i - O_i}{D_i} \right) \right]^2 \times n_i \right)
\]

subject to:

1. the water balance equations

\[
S_i = S_{i-1} + I_i - O_i
\]

Figure 1. The control system of reservoir operation
2. the initial storage

\[ S_0 = 55 \]

3. the limitations of reservoir capacity

\[ 0 \leq S_i \leq 251 \]

4. conditions on final stage of reservoir storage

\[ 0.95 \, S_0 \leq S_{36} \leq 1.05 \, S_0 \]

where \( D_i \) is the water demand of the \( i \)th ten-day, \( O_i \) is the outflow of the \( i \)th ten-day, \( n_i \) is the accumulative deficiency in the \( i \)th ten-day, \( S_i \) is the storage of the \( i \)th ten-day and \( I_i \) is the inflow of the \( i \)th ten-day.

To solve the optimization problem, some conventional techniques can be used. In the present work, we use genetic algorithms (Holland, 1975) to solve the above complex, discontinuous, non-differentiable and multimodal problem.

**GENETIC ALGORITHMS FOR GENERATING THE TRAINING PATTERN**

The GA is similar to Darwinian natural selection, which combines an artificial survival of the fittest and the natural genetic operators. The primary monograph on the topic is Holland’s adaptation (1975). Having become increasingly important, GAs are now widely applied in the engineering fields. In the studies of water resources, GAs also have been applied to solve the problems of rainfall–runoff calibration (Wang, 1991), pipe network optimization (Simpson et al., 1994), operating policies of reservoir systems (Oliveira and Loucks, 1997) and flood reservoir management (Chang and Chen, 1998).

The GAs act as a biological metaphor that tends to mimic some of the processes observed in natural evolution. To evaluate the suitability of the derived solution, an objective function is required. The objective function is chosen in such a way that highly fitted strings (solutions) have high fitness values. The evolution starts from a set of coded solutions (biologically referred to as chromosomes) and proceeds from generation to generation through genetic operations: reproduction, cross-over and mutation. The flowchart of the GA procedure is shown in Figure 2.

**ADAPTIVE NETWORK-BASED FUZZY INFERENCE SYSTEMS (ANFIS)**

The ANFIS model was proposed by Jang (1993) and has been applied successfully to many problems such as motor fault detection and diagnosis (Altug et al., 1999), power systems dynamic load (Djukanovie et al., 1997; Oonsivilai and El-Hawary, 1999). It can serve as a basis for constructing a set of fuzzy if—then rules with appropriate membership functions to generate the preliminary stipulated input–output pairs. This section briefly introduces the basics and architecture of the model.

An adaptive network is a multilayer feed-forward network with a supervised learning scheme. The functions corresponding to nodes of the same layer are similar. For simplicity, the first-order Sugeno fuzzy model (Takagi and Sugeno, 1985) is considered as a fuzzy inference system. Suppose that the model contains two inputs \( x \) and \( y \), one output \( F \) and four fuzzy if—then rules.

**Rule 1:** if \( x \) is \( A_1 \) and \( y \) is \( B_1 \), then

\[ f_1 = p_{11} x + q_{11} y + r_1 \]

**Rule 2:** if \( x \) is \( A_1 \) and \( y \) is \( B_2 \), then

\[ f_2 = p_{21} x + q_{21} y + r_2 \]

**Rule 3:** if \( x \) is \( A_2 \) and \( y \) is \( B_1 \), then

\[ f_3 = p_{31} x + q_{31} y + r_3 \]

**Rule 4:** if \( x \) is \( A_2 \) and \( y \) is \( B_2 \), then

\[ f_4 = p_{41} x + q_{41} y + r_4 \]
The simplified ANFIS architecture is shown in Figure 3.

Layer 1: input nodes

Every node $i$ in this layer is used to perform a membership function

$$O_{1,i} = \mu_{A_i}(x) \quad \text{for} \quad i = 1, 2$$
$$O_{1,i} = \mu_{B_{i-2}}(y) \quad \text{for} \quad i = 3, 4$$

where $x, y$ are the crisp inputs to node $i$, and $A_i, B_i$ are the linguistic labels characterized by appropriate membership functions $\mu_{A_i}, \mu_{B_i}$, respectively. In other words, the outputs $O_{1,i}$ of this layer are membership functions of $A_i$ and $B_i$. Usually the bell-shaped function is chosen as a membership function.

$$\mu_{A_i} = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2\theta}} \quad \text{and} \quad \mu_{B_{i-2}} = \frac{1}{1 + \left| \frac{y - c_i}{a_i} \right|^{2\theta}}$$

(2)

where $\{a_i, b_i, c_i\}$ is the parameter set of the membership functions in the premise part of fuzzy if—then rules.

Layer 2: rule nodes

After membership functions are generated, the links of this layer correspond to the preconditions in fuzzy logic rules. The $T$-norm operators are used to perform the multiplications of the incoming signals to generate
the outputs of this layer. The $T$-norm operators can be thought of as the extension of fuzzy AND operations. There are several different kinds of $T$-norm operations, e.g., logic product, algebraic product, and bounded product (Tsoukalas and Uhrig, 1997). The algebraic product is used and shown as follows

$$O_{2,k} = w_k = \mu_{A_i}(x) \times \mu_{B_j}(y) \quad k = 1, \ldots, 4; i = 1, 2; j = 1, 2$$

Therefore, the outputs $O_{2,k}$ of this layer are the products of the corresponding degrees from Layer 1.

**Layer 3: average nodes**

The $i$th node’s output in this layer is the ratio of the $i$th node’s output from the previous layer to the total of them

$$O_{3,i} = \frac{w_i}{\sum_{k=1}^{4} w_k} \quad i = 1, \ldots, 4$$

**Layer 4: consequent nodes**

In this layer, the $i$th node corresponds to the following function

$$O_{4,i} = \frac{w_i f_i}{\sum_{i=1}^{4} w_i} = \frac{f_i}{\sum_{i=1}^{4} w_i}$$

where \{\(p_i, q_i, r_i\)\} is the parameter set in the consequent part of the first-order Sugeno fuzzy model.

**Layer 5: output nodes**

In this layer, a defuzzification inference is used to transform the fuzzy results of this model into a crisp output. Hence, the sum of all incoming signals is used to generate the decision crisp output.

$$O_{5,1} = \frac{\sum_{i=1}^{4} w_i f_i}{\sum_{i=1}^{4} w_i}$$

Figure 3. The ANFIS architecture for a two-input Sugeno fuzzy model with four rules
APPLICATION

Case study area: Shihmen reservoir

The above intelligent control is applied to the operation of Shihmen reservoir, which is one of the largest water reservoirs in Taiwan. The Shihmen reservoir, located upstream of the Tahan river (Figure 4), is a multipurpose reservoir for irrigation, hydroelectric energy, public water supply, flood control and tourism. The watershed covers an area of 763.4 km² with an effective storage capacity of $2.51 \times 10^8$ m³. Inflow data for the past 30 years are available, but the water release has not been recorded. Up to the present time, there are only studies based on simulation and/or optimization techniques to search the optimal water release information under specific hydrological conditions.

Without information of optimal input–output patterns, the training and learning procedure for building an intelligent control system cannot be performed. Consequently, it is impossible to build an intelligent reservoir operation system. Obviously, we must find a way to construct an optimal input–output pattern.

Application of GA

Given the water demand and the inflow information each year, we applied the genetic algorithm to searching the optimal release in the corresponding circumstance. The application of the genetic algorithm requires:

1. an encoding scheme that groups the unknown variables together;
2. a fitness function that assigns an overall performance;
3. the initial multiple sets of feasible values for each of the decision variables;
4. the choice of values for the genetic algorithm parameters.

In this study, real-valued chromosomes are used, and the fitness function is set as a combination of objective function and constraint formula in equation (1). There are 36 ten-day reservoir releases each year to be determined. The parameters in the genetic algorithm are (i) cross-over probability $D_0$, (ii) mutation rate $D_0$, and (iii) population size $= 1000$. The population size is the number of chromosomes (simulation results) in each generation. A large population gives more diversity and better final solutions, but longer computation times are needed. Roulette wheel parent selection is used such that the selected chromosomes for the next generation are directly proportional to their fitness values. To improve the convergence, an elite strategy is implemented in the above process. The terminal conditions are set when the process has been executed more than 8000 generations or when the fitness function has reached a value less than $10^{-6}$.

The large number of variables, the huge domain of searching and the complexity of the problem could make the GA searching procedure impossible to converge. In most cases, it terminates at many local minima. To avoid these situations, we modified the initial multiple sets of feasible values for these variables. According to our experience, the decision variables (water release) at each time step should fall into the neighborhood of water demand, or inflow at the corresponding time step.

Therefore, the one thousand initial sets of decision variables (the chromosomes in the first generation) are divided into four different searching spaces as follows.

1. One hundred initial sets (10%) are distributed randomly in the neighbourhood of water demand.
2. One hundred initial sets (10%) are distributed randomly in the neighbourhood of inflow.
3. One hundred initial sets (10%) are distributed randomly in the neighbourhood around the value of $\sum_{i=1}^{36} I_i \times \frac{D_i}{\sum_{j=1}^{36} D_j}$.
4. The remaining 70% are spread randomly in the original feasible solution domain. This setting can prohibit the solution from quickly falling into a local minimum.

With this modification, the final solutions have dramatically improved, and an approximately convergent solution is achieved. The results of reservoir operating performance by using GA are far superior to those based on the M-5 rule curves (see Table I).

Application of the ANFIS

Given the input–output patterns from the above results, we apply the ANFIS model as described in Figure 5. There are five input variables. Each variable might have several values (in terms of rules), and each rule includes several parameters of membership functions. For example, if each variable has three rules, and

![Figure 5. The structure of ANFIS for reservoir operation](image-url)
Table I. The results of three models in different periodic inflow series

<table>
<thead>
<tr>
<th>Periodic Year</th>
<th>Total inflow</th>
<th>GA</th>
<th>ANFIS</th>
<th>M-5 rule curves</th>
</tr>
</thead>
<tbody>
<tr>
<td>(year)</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(1)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1409-96</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1346-72</td>
<td>37-60</td>
<td>102</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1553-86</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1463-4</td>
<td>87.36</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1297-8</td>
<td>106-37</td>
<td>1.810</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1585-33</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1096-94</td>
<td>314-70</td>
<td>18-501</td>
</tr>
</tbody>
</table>

(1) The number of water shortages within the 36 periods of 1 year.
(2) Total water shortage (10^6 m^3).
(3) GSI.

Each rule includes three parameters, then there are 45 (5 (variables) × 3 (rules) × 3 (parameters)) parameters needing to be determined in layer 1. In layer 2, these rules will generate 3^2 nodes, and there are 1458 (3^2 × 6 (coefficients in the first-order Sugeno model)) parameters undetermined within the defuzzification process in layer 4.

Apparently, there are too many parameters to be determined for an efficient model. To solve this problem, we apply subtractive fuzzy clustering (Chiu, 1994) to establish the rule base relationship between the input and output variables. (This subtractive fuzzy cluster is summarized in the Appendix) After the number of the rules is reduced, the ANFIS model’s parameters can be adjusted much more efficiently.

In using the 30-year data of input–output patterns obtained from GA to construct the ANFIS, we could not obtain a convergent solution even though different approaches were tried. This probably is the result of the following reasons: (i) the inflow is very diversified, and (ii) the input–output patterns of 30-year data are not consistent. Namely, the same input information might lead to very different output results. Therefore, instead of using the 30-year historical inflow, we opted to use three short inflow series to construct the ANFIS model.

Based on the 1080 records from the 10-day inflow data in the last 30 years, we use the 10-day moving average process to generate 1-year, 2-year and 5-year periodic inflow time-series as shown in Figure 6. The inflow time-series are then used as the input. The 1-2 times of the water demand in the year of 2001 is adopted to represent the requirement for the simulation period.

Results

The overall results of M-5 curves, GA and ANFIS are presented in Table I. For comparison, the simulation results of the M-5 curves are also shown. The figures for reservoir storages for different time periods obtained by the GA and ANFIS methods are given in Figure 7. The results show that the ANFIS method can obtain very similar results to the GA method, which represents the outstanding learning ability of the ANFIS. The data of water demand and reservoir water release for different time periods are also shown in Figure 8. Apparently, water releases given by the ANFIS match or greater than water demands most of the time, although during the final year of the 5-year periodic inflow series, which is a drought year, water releases, as expected, are less than water demands.

The simulation results demonstrate that this new approach, in comparison with the M-5 rule curves, has superior performance with regard to the generalized shortage index (GSI). The GSI has been proposed by
Figure 6. Three different periodic inflow time-series: (a) 1 year, (b) 2 year and (c) 5 year
Figure 7. Reservoir Storages of the GA and ANFIS methods: (a) 1 year periodic series; (b) 2 year periodic series; (c) 5 year periodic series
Figure 8. Reservoir outflows from the ANFIS versus water demands: (a) 1 year periodic series; (b) 2 year periodic series; (c) 5 year periodic series
Hsu (1995) and is defined as follows

\[
DPD = \sum (DDR(\%) \times NDC)
\]

\[
GSI = \frac{100}{N} \sum_{i=1}^{N} \left( \frac{DPD_i}{100 \times DY_i} \right)^k
\]

where DPD is the deficit per cent 10-day index; DDR is the 10-day deficit rate, i.e. total deficit in a period/water demand in the period ×100%; NDC is the number of 10-day periods in a continuous deficit; DY is the number of 10-day periods in the \(i\)th year (36); \(N\) is the number of sample years; and \(k\) is a coefficient, usually taken as 2.

The total water demand is set as \(1329 \times 10^6\) m\(^3\). It is easy to find from Table I that as the total inflow in a year is greater than the water demand, there will not be much of a water shortage by all methods. Table I shows that the GA has smallest GSI values in all cases, whereas the ANFIS model has better performance in terms of smaller water shortages than the M-5 rule curves. Our results demonstrate that the GA is superior to the M-5 curves in providing ideal input–output training patterns. The ANFIS model has the ability to learn from these patterns, and therefore it can be applied in real-time reservoir operation.

CONCLUSIONS

Based on artificial intelligent algorithms, we presented a framework for the intelligent control of reservoir operation. The combination of genetic algorithm and ANFIS provides a framework that can be used to pre-classify the input–output pattern and determine the optimal real-time reservoir release operation. The genetic algorithm is used to search the optimal reservoir operating schedule in a given inflow time-series. Given reasonable initial sets, GA has produced superior results. Learning from the input–output patterns obtained from GA, the ANFIS model can estimate the reservoir release in predefined conditions. The ANFIS model, once given sufficient information to construct the fuzzy rules, has more efficient operation of the reservoir than the rule-curve-based model. Our study of the Shihmen reservoir operation demonstrates both the intelligent capability and superior performance of this framework in contrast to the traditional rule-curve operating strategy.

By using GA in many searching processes, we found the following important features of the reservoir operation. In wet years, more than one release histogram can satisfy water demand, which implies that there is more than one optimal solution. In drought years, the optimal release histograms are very close to each other. Overall, there are many optimal solution sets that have the same minimum value in the objective function of reservoir operation. Consequently, the objective function does not represent the reservoir operation characteristics in this circumstance. To solve this problem, one may increase the water demand or add other objectives such as power generation rate. In applying the ANFIS for reservoir operation, we recommend using the subtractive fuzzy clustering to reduce the number of input rules. In this way, the ANFIS model’s parameters can be adjusted much more efficiently.

ACKNOWLEDGEMENTS

This paper is based on partial work supported by the National Science Council, R.O.C. (Grant No. NSC89-2313-B-002-042). Suggestions from Professor Chen Yung-Yaw in Department of Electrical Engineering of the National Taiwan University are greatly appreciated.
APPENDIX

Clustering via fuzzy subtractive clustering (Chiu, 1994) and initial model generation

Cluster analysis is the process of classifying a collection of data into subsets and producing an effective representation of a system’s behaviour. When we apply the clustering estimation method to the input–output data observations, cluster centres can be used as rules that describe characteristics of the system. The objective of the fuzzy subtractive clustering in our model is to generate the fuzzy if–then rules of the Sugeno fuzzy model (Takagi and Sugeno, 1985).

Cluster estimation. In an $M$-dimensional space, there are $n$ data points $\{x_1, \ldots, x_n\}$ to be grouped. As each data point is considered as a potential cluster centre, the density measure of a point $x_i$ is defined as

$$ D_i = \sum_{j=1}^{n} \exp \left[ -\frac{\|x_i - x_j\|^2}{(r_a/2)^2} \right] $$

where the positive constant $r_a$ is the radius defining a neighbourhood of a cluster centre. After the density measure of each point $x_i$, we select the point with the highest density $D_{c1}$ as the first cluster centre $x_{c1}$. Then, to prevent neighbourhood points of the first cluster centre from being selected as the second centre, the density measure of each point $x_i$ is revised by Equation (A2) as follows

$$ D_i = D_i - D_{c1} \exp \left[ -\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2} \right] $$

A recommended value of setting $r_b$ is $1.5 r_a$. Again, we select the remaining data with the highest modified density as the second cluster centre. The modification of the density measure in the remaining point set is applied in each step. In general, after the $k$th cluster centre has been determined, the density of the remaining data point is revised by Equation (A3)

$$ D_i = D_i - D_{c_k} \exp \left[ -\frac{\|x_i - x_{c_k}\|^2}{(r_b/2)^2} \right] $$

where $x_{c_k}$ is the $k$th cluster centre and $D_{c_k}$ is its density. The process of determining a cluster centre and its corresponding density repeats until stop conditions for specific parameters have been reached.

Initial model identification. The cluster centres can be used as the centres for the fuzzy rules’ premise in a Sugeno fuzzy model. In the previous process, there are $k$ cluster centres $\{x_{c1}, \ldots, x_{ck}\}$ in an $M$-dimensional space. The $i$th cluster centre $x_{ci}$ can be decomposed into two component vectors $p_i$ and $q_i$, where $p_i$ is the input part and it contains the first $N$ element of $x_{ci}$; $q_i$ is the output part and it contains the last $M-N$ elements of $x_{ci}$. Therefore, given an input vector $p$, the degree of the membership function to $i$th rule is defined as Equation (A4)

$$ \mu_i = \exp \left[ -\frac{\|x - p_i\|^2}{(r_a/2)^2} \right] $$

REFERENCES


