罾八俳┓ 映ハ③▫ da削吴H 俵 46 ぼ BH 句亍 15 削
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No.46 B, 2003

Multi-objective Storage Reservoir Operation under Uncertainty

Paulo CHAVES*, Toshiharu KOJIRI
Water Resources Research Center, DPRI, Kyoto University
* phchaves@yahoo.com

Synopsis

Water quantity and quality are considered to be the main driving forces the reservoir operation. Barra Bonita reservoir, located in the southeast region of Brazil, is chosen as the case study for the application of the proposed methodology. Herein, optimization and artificial intelligence (AI) techniques are applied in the simulation and operation of the reservoir. A fuzzy stochastic dynamic programming model (FSDP) is developed for calculating the optimal operation procedures. Optimization is applied to achieve multiple fuzzy objectives. Markov chain technique is applied to handle the stochastic characteristics of river flow. Water quality analysis is carried out using an artificial neural network model. Organic matter and nutrient loads are modeled as a function of river discharge through the application of a fuzzy regression model based on fuzzy performance functions. The obtained results show that the proposed methodology provides an effective and useful tool for reservoir operation.

Keywords: fuzzy regression, artificial neural networks, fuzzy stochastic dynamic programming, uncertainty, water quality, Barra Bonita reservoir

1. Introduction

1.1 Background

Water quantity and quality in the environment are affected by many different factors. Such factors include increasing water demand, multiple-use, water pollution due to rapid urbanization, high development of water resources, eutrophication and degradation of water bodies and increasing costs related to water treatment.

Recently, assessment of seasonal, social and ecological changes has become more relevant to reservoir management. Application of stochastic and artificial intelligence (AI) techniques, such as fuzzy set theory or neural networks, in the development of decision support systems can provide suitable ways to analyze all these complex connections and uncertainties.

Historically, reservoir management has aimed to achieve a single and purpose. This approach has been used in many countries worldwide, and has been mostly based on economic cost-benefit analysis. As a consequence, many water-related problems have arisen. This has created the need for a better and more comprehensive evaluation process. The consideration of various purposes (e.g. recreation, environmental conservation and navigation) must be addressed from an integrated water resource management point of view.

Water quality directly affects virtually all water uses. Fish survival, diversity and growth, recreational activities, municipal industrial and domestic water supplies, agricultural uses, such as irrigation and livestock, waste disposal and general aesthetics are all affected by the physical, chemical, biological and microbiological conditions that exist in water bodies.

Many factors influence water quality. The chemistry of bedrock and superficial geology, and the drainage characteristics of soil can determine whether natural waters are acidic or alkaline, high in heavy metals and dissolved salts or relatively free of those constituents. Physical processes like erosion can add large quantities of suspended sediment to surface waters.

The minimum acceptable quality of water depends very much on the water use. Water for irrigation, for example, should be low in dissolved salts, but water intended for livestock should be low in bacteria. Water used in industrial processes should usually be of a much higher quality than water used for industrial cooling.

As for municipal supply, water must not only be safe to drink, but ideally contain low concentrations of materials such as calcium, iron or similar materials, as they may cause costly infrastructure damage, or add unpleasant characteristics even after treatment.

Thus, assessment of water quality is a very complex task for water resources managers. Some reasons why water quality continues to be left out from the decision-making framework process are:

- Multiple stakeholders and objectives are presented in most cases,
- Great number of water quality parameters,
- Variability of quality (daily and seasonal),
- Difficulty to economically evaluate water and environmental quality,
- Water quality is highly influenced by human activities and natural conditions, and
- Lack of spatial and long term observed data.

1.2 Objectives

The aim of this research is to achieve integrated management of water resources with respect to quantity and quality issues. The long-term planning operation to achieve multi-objectives, such as flow stabilization, power generation and quality characteristics, is discussed.

Zalewski et al. (1997) emphasizes the importance of adequate management of quantities to improve water quality. The proposed concept of ecohydrology introduces a creative way to improve the prediction of large scale, long-term processes as a background to sustainable management. Moreover, it presents an example of the relationship between storage levels and eutrophication process for the Sulejow Reservoir.

Quantity and quality of water are often related, as for, when the optimization of two outlets has to be considered. As water in a reservoir may present different characteristics in the vertical water column, the effective operation of gates could improve water quality. Moreover, considering that the manner in which a reservoir is operated has a great influence on its volume and its release quality, it is easy to imagine that in a system of reservoirs, the volumes and releases of each reservoir have important effects on one another.

Both examples cited above indicate the necessity of integrated analysis. However, such systems present with a high degree of complexity for analysis, due to the large number of variables that should be considered. Therefore, this research focuses on a simpler case, shown in Fig. 1.1, which considers the relationship between storage volumes, release from a single outlet, and quality within the reservoir.

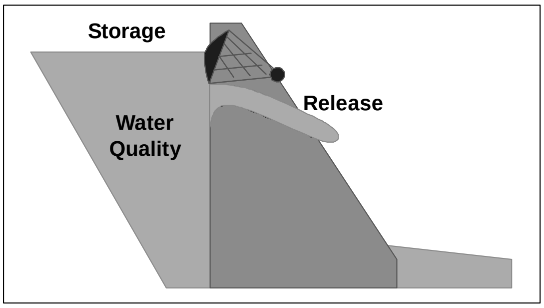

Fig. 1.1 Storage and water quality

The purpose of this research is to generate a broader formulation for the integration of quantity and quality of water within multi-purpose reservoirs, which can serve as a basis for future development in the field. The basic components considered here are shown in Fig. 1.2, where the reservoir water quality is represented by the central part of the water body, just after the junction of the two main tributaries, which present different water quality condition. Storage, inflow and release act as key factors on the system behavior.

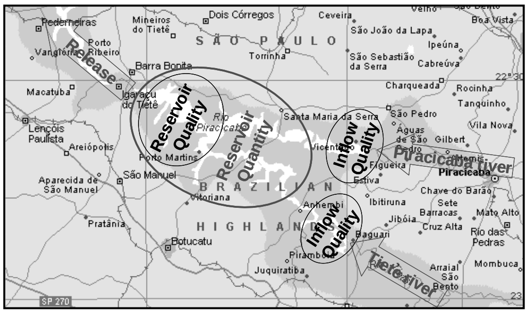

Fig. 1.2 Basic components of the Barra Bonita reservoir system

2. Methodology

Maximizing water quantity and water quality benefits are considered the primary purposes of the reservoir management. Barra Bonita reservoir is chosen as the case study for the application of the proposed methodology.

Techniques such as dynamic programming (DP), time series analysis, Markov chain process, fuzzy regression, genetic algorithm (GA), neural networks and fuzzy sets theory are applied with different degrees of complexity or combined one with the other. The general optimization framework is presented in Fig. 2.1. Details of each of the techniques used in this research are given in the following sections.

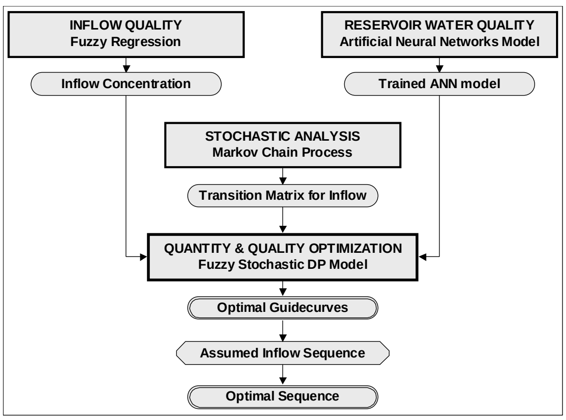

Fig. 2.1 Proposed methodology for the water quantity and quality optimization

Analysis is conducted where quantity and quality are considered at each time-step in the optimization process. For calculation, a time step of one month is used, since the analysis is intended to find the best procedure for long-term reservoir operation. Day-based analysis could be developed for the short-term real time operation through some modifications to the current approach.

A fuzzy stochastic dynamic programming model (FSDP) is developed for calculating optimal operation procedures. Optimization is applied to achieve multiple objectives, which are defined as fuzzy membership functions. This provides a more flexible method for dealing with multi-purpose reservoir management. For example, improvement of water quality conditions can be extremely difficult to define using economic units alone. The stochastic component of the FSDP model is the Markov chain technique, which is applied to handle the probabilistic characteristics related to river discharge into the reservoir inflow.

The output from the FSDP model is a set of guidecurves. The guidecurves are constructed for each month (stage). It is straightforward to obtain the optimal end-of-period storage from the guidecurves, when the beginning-of-period storage volume and the previous month inflow are known. Then, using the guidecurves produced by the FSDP model, an optimal storage sequence for the water quantity objectives can be found, as described in Section 6.

Water quality modeling presents is particularly complex due to data inadequacies, difficulties of parameter definition, and the limitations of conceptual assumptions regarding biological processes.

Water quality analysis is carried out using an artificial neural networks model (ANN). The ANN model takes input parameters such as season, storage, retention time and organic matter and nutrient loads in to account, as described in Section 4. Organic matter and nutrient loads are modeled as a function of river discharges through the application of a fuzzy regression model based on fuzzy performance functions, as described in Section 5. Fig. 2.2 presents the basic structure of the application of above techniques.

Fig. 2.2 Modeling the Barra Bonita system

2.1 Multi-objective Analysis

In the operation of multi-purpose reservoirs, quantitative and qualitative objectives must be considered. Hence, the calculation becomes complicated and difficult to execute (Simonovic, 1991; Yakowitz, 1998). The need to consider multiple purposes of a reservoir has become even more important in recent years, due to the increased emphasis on environmental protection, especially in water bodies suffering from pollution from nearby urban areas.

Attempting to evaluate objectives only economically, through cost/benefit analysis, may fail to represent the satisfaction among all stakeholders. Moreover, it is almost impossible to only evaluate some objectives economically, such as environmental quality, recreation and health problems.

Various approaches for solving the multi-objective management of natural resources have been attempted by El-Swaify and Yakowitz (1998). Traditional multi-objective optimization techniques applied to reservoir operation includes: the weighting method, the ε-constraint method, the surrogate worth trade-off method, goal programming, and compromise programming (Changchit and Terrell 1989; Simonovic 1991)

The application of the Fuzzy sets theory can provide a viable way to handle situations when problems with objectives are difficult to define due to vagueness and imprecision. Application of the Fuzzy sets theory to water resources analysis can be found in Hipel (1982), Kojiri (1992), Russel and Campbell (1996), Shrestha (1996) and Fontane et al. (1997).

Fuzzy set theory gives the ability to work with measures of satisfaction by using fuzzy membership functions. In the case of reservoir operation, fuzzy membership functions may be described in terms of water level, release, and water quality parameters.

The main objective of this work is not to detail the formulation process of the membership functions. However, it is important to mention that the better the formulation of the functions, the higher the chance of operation success. The functions have to represent the desire of all groups in the water sector and have to be constantly evaluated and updated to properly represent operation objectives.

2.2 Fuzzy Sets Theory

Commonly, objectives of reservoir operation are represented as a function of quantifiable characteristics, such as release, storage and power generation. However, other objectives, such as environment quality, may not be so simple to evaluate, due to difficulties in the reduction to comparable commensurable units.

One attempt to solve this problem is through the comparison of actual values with expected objective targets. These comparisons can generate functions that represent the degree of satisfaction between operation decisions and desirable values.

Most of these satisfaction functions are based on the relationship between targets and expected values – for example, the goal programming. Nevertheless, some objectives cannot easily be perceived and understood by water stakeholders.

With the intention to make the interpretation of those degrees of satisfaction easier (especially because of their complexity and vagueness) linguistic descriptors such as “good” water quality or “adequate” water supply may be used. The degree of satisfaction will depend on stakeholders’ perceptions and experience, the system’s characteristics considering the constraints encountered, knowledge about the state of the system, risk perception and impacts related to water quantity and quality problems.

The basic method of formulating the membership functions is provided by survey with water manager and users who have the knowledge about the subjectivity of the system. Elaboration and analysis of these surveys (Norwich and Turksen (1984), and Rea and Parker (1992)), as well as the fitting of curves on the acquired data (Dombi (1990) and Klir and Yuan (1995)) can be found in various references.

This method of categorizing reservoir objectives through the use of linguistic values can be handled with fuzzy sets, allowing a gradual transition from a situation that completely fulfills a concept to a situation that does not. A more detailed explanation on the theory of fuzzy sets is presented in Kacprzyk (1997).

2.3 Water Quality Assessment

Porto et al. (2000) attempted to integrate evaluated results on water quality and quantity. The paper presented water quantity optimization, followed by a simulation of scenarios applied for the hydrologic net of the Tiete river basin. However, quality analysis within the reservoirs was not performed.

Some relevant aspects of the water system behavior within the reservoir of Barra Bonita can be found in (Tundisi, 1990). Based on some characteristics of the system, the first assumptions for the modeling may be taken.

i) “Limnology events in Barra Bonita seem to be dominated by climatologic factors and by flushing rates and residence time.” It points out the need to better understand and identify such interrelation. For that, a ANN model is developed to simulate water quality conditions within the reservoir.

ii) “During most of the year, stratification is short and weak.” This can give a basis for the assumption of considering the reservoir a mixed system.

3. Study Area – Barra Bonita Reservoir

3.1 Basic Characteristics

Barra Bonita reservoir is located in the middle Tietê River basin, São Paulo, Brazil, (22^{o}29’ S and 48^{o}34’ W) with maximum surface water at an altitude of 453 m. The reservoir has a water surface area of approximately 340 km^2, total volume of 3.6 km^3 and length of 50 km. Maximum and average depth are around 25 and 10 meters, respectively. Average water fluctuation in the reservoir seems to stay around 5 meters. A schematic profile of the dam with main storage levels and elevations is presented in Fig 3.1.

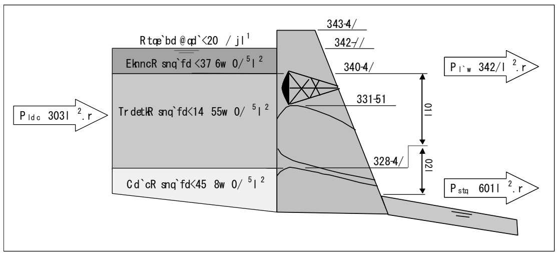

Fig. 3.1 Schematic profile of the Barra Bonita dam site with volumes and elevations

Hydropower energy seems to be the primary water use of the reservoir. Other uses include navigation, recreation, water supply and fishery production. Barra Bonita and other dams located in the Tietê River can be seen in Fig. 3.2.

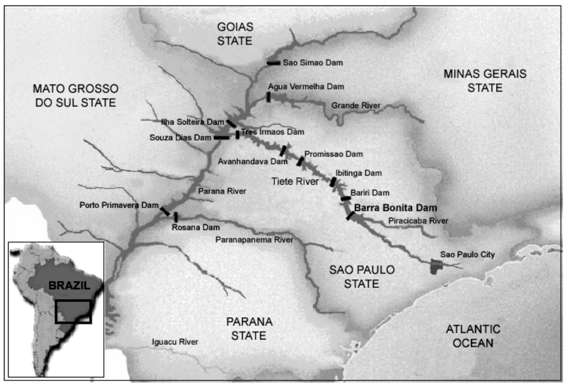

Fig. 3.2: Location of Barra Bonita reservoir

The Reservoir is the first of a series of six reservoirs, around 300 km downstream from Brazil’s biggest city, São Paulo. It can be classified as a subtropical/tropical reservoir with intermediate retention time of around one to two months.

Air temperature normally varies within a range of only 15°C between winter and summer. The wet season occurs between September and March. Annual cumulative precipitation is around 1400 mm, with maximum wind velocity of 5 to 7 m/s during winter.

Water level and volume are related to both climatologic conditions and water use. The average annual flushing rate is 414 m^3/s. Changes in release discharge rates are an important driving force in this system, rapidly modifying ecological conditions within the reservoir, as well as downstream.

Stratification does not seem to be very strong in the reservoir due to the medium residence time and its shallowness. Nevertheless, stratification is significant enough to create undesirable water quality problems.

Two main rivers, Tietê and Piracicaba, flow into the reservoir. Due to the considerable difference of water quantity and quality of the two rivers, water quality in the reservoir is spatial heterogeneous.

The general source of impacts to the Reservoir are nitrogen and phosphorous input from non-point and point sources; input of suspended material from agricultural activities and runoff during precipitation; navigation, tourism and recreation related activities; deforestation in the watershed and, introduction of exotic fish species.

Some consequences of these impacts include: eutrophication; siltation; blooms of Cyanophyta (Microcystis), especially in summer and Anabaema, during the tropical winter – this impact may be minimized by controlled operation of the spill water (Tundisi, 1990).

4. Artificial Neural Networks

Artificial neural network (ANN) models have been widely used in different fields such as aerospace, finance, robotics, environmental assessment and hydrology, for different purpose, such as classification, pattern identification, simulation and prediction. The concept of neural networks was first introduced in 1943, but it was not until the middle 80's that applications of ANN became widespread, as noted by Maier and Dandy (2000).

The ANN model developed here is used for simulation and prediction of water quality within the reservoir of Barra Bonita. Here, an ANN model is used in preference to a physical model, as the latter tends to demand a lot of computational effort, particularly when combined with optimization techniques, such as dynamic programming, which is used here for storage optimization.

Additionally, physical models usually require much diversified data and estimation of various parameters, which can considerably increase the uncertainty in model output, particularly when data and information are limited.

Also as demonstrated by Chaves (2002), in using a stochastic DP model, forward calculation is not applicable. As most physical models use forward calculation algorithms with time dependence, a basic conflict combining both models arises.

4.1 ANN model for Reservoir Water Quality

Modeling water quality presents a great degree of complexity. This is due to two basic reasons; the first is related to the number of parameters and the complexity of the processes itself (see Section 1). The second comes about as a result of the few available data for water quality. For example, only 24-observations data set is available for this research. In addition, to make the set useful for training, some gaps within the observation data set had to be interpolated, as observation is made only every one or two months, which is a too long period between observations, as water quality variability is known to occur even on a daily scale in some cases.

Therefore, time-dependence is not considered in the developed model due to lack of appropriate data. However, as stated above, the lack of time-dependence makes the combination of the ANN model for water quality and the stochastic DP model more straightforward, as the ANN model can be used with the backward calculation recommended for SDP.

Five quality parameters are simulated and predicted with the ANN model: DO, BOD, TP, TN and chlorophyll (CHA). The concentrations of these parameters are represented in the ANN model as the neurons of the output layer. Here, the ANN model is developed to handle all five parameters in the same model. A different ANN model could also be developed for each parameter independently. However, as these parameters have some interdependence amongst themselves and the various input variables, it is assumed that a single model would best reflect these interrelations.

As for the input variables or the input layer, eight nodes are applied, representing the current month, the average storage value for the month, average retention time considering average storage and inflow for the current month and the inflow quality loads for the same parameters as for the output layer: DO, BOD, TP, TN and CHA loads.

To summarize the characteristics of the input layer neurons, it can be said that they may represent:

- Month: influences due to seasonal characteristics, such as temperature and solar intensity.
- Storage: influences due to size, area and depth of reservoir.
- Retention Time: characteristics such as the ratio of inflow to storage, and phosphorous and nitrogen retention.
- Nutrients and organic matters loads: inflow quality into the reservoir system.

The hidden layer is defined through a trial-and-error process. It is important to note that with limited data available for, it is recommended to avoid using a large number of nodes in the hidden layer, so as to avoid the problem of over-training (Hagan et al.; 1995).

The ANN model structure used here is shown in Fig. 4.1, where the log-sigmoid and linear transfer functions are used for the hidden and output layers, respectively. The model with two layers and the above-mentioned transfer functions can be trained to approximate most functions (Hagan et al.; 1995).

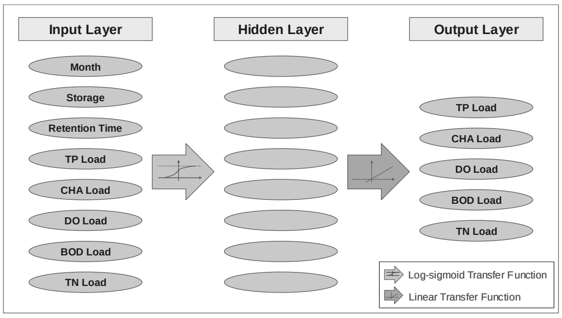

Fig. 4.1 Basic structure of the ANN water quality model

4. 2 Training Process

The proposed reservoir water quality ANN model is trained using 18 of the 24 observations. Six observations are used for the validation of the model. The training data sets are obtained as being equal to the average of two observation points located within the main body of the reservoir. This is intended to make the ANN model more representative at the situation of spatial variability of the quality parameters within the reservoir.

The model is trained through the use of a genetic algorithm (GA) model, further information on the GA model can be found in Galvao (1999). Backpropagation training was also carried out. However, here, the GA training presented better results; therefore, it is used as the final training process. A basic flowchart for the development of the ANN model with GA training is shown in Fig. 4.2.

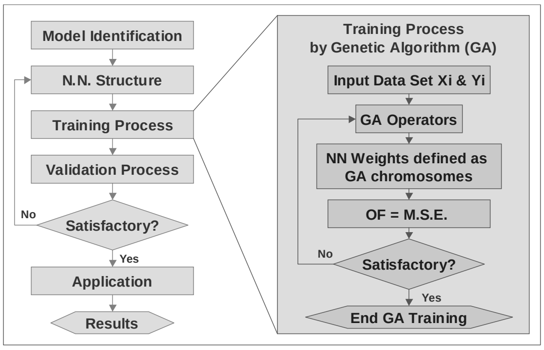

Fig. 4.2 Flowchart for the overall model construction and GA training process

Here, the mean square error (MSE) function is used as the evaluation function in the training process. The MSE function can be mathematically represented as below:

$$MSE = \frac{1}{N}\left[\sum_{i=1}^{N}\left(Y_{obs} - Y_{calc}\right)^2\right] \qquad (1)$$

where, N is the total number of observations, and Y_{obs} and Y_{calc} are the observation and calculated values, respectively.

Only results for TP, CHA and DO from the training and validation processes are presented in Figs. 4.3. The calculated outputs are plotted together with the values from the two observation stations used in the training process.

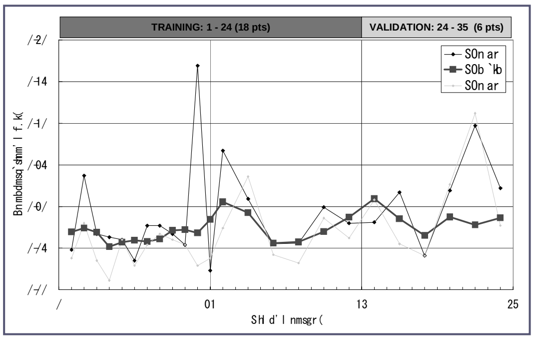

Fig. 4.3a Results for phosphorous after training

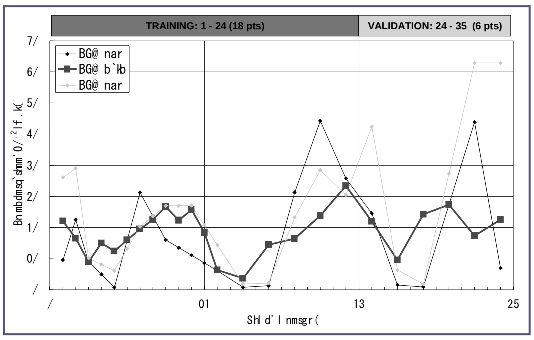

Fig. 4.3b Results for chlorophyll after training

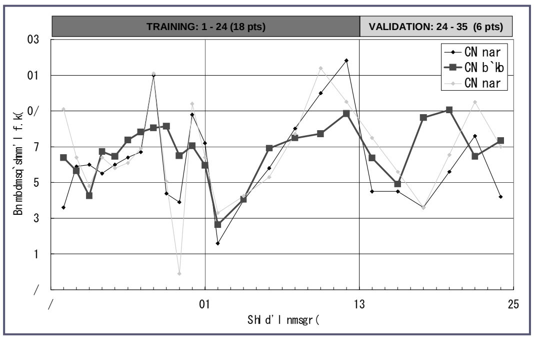

Fig. 4.3c Results for DO after training

The outputs of the developed ANN model are considered to be representative, as they can represent the same temporal variability and same magnitude of the observed data. Therefore, it can be concluded that the developed ANN model may be used to the evaluation of water quality parameters within the reservoir during the optimization process.

5. Fuzzy Regression

Statistical regression analysis provides a means to analyze the relation between two, or more in the case of multi-regression, sets of values such as *x* and *y*. Statistical regression techniques are well developed and widely used in many fields of studies. In the case of large data sets, use of statistical techniques can be used to describe relationships and dependence between set of values.

However, in many cases, lack of data remains a great constraint when statistical techniques are to be applied. This is a particular problem when working in fields related to the natural sciences, such as hydrology, where limited data is common.

Fuzzy regression is a documented technique, which can give us a way to deal with the problem of scarce data inadequacy and errors. Besides, data may also be inconsistent or not representative, which will lead to wrong and surprising conclusions. In the next section, an explanation of the basis of fuzzy regression and its mathematical formulation is presented and some relevant references on fuzzy regression are cited.

The purpose of the analysis is to determine the nutrient and organic matter load flowing into the Barra Bonita reservoir, State of Sao Paulo, Brazil.

Other applications for fuzzy regression in the field of hydrology include: uncertainty in discharge curves, sediment transportation relation between flow quantities and suspended sediment, groundwater hydrology in the estimation of aquifer parameters and health risk analyses when decisions have to be made on the basis of only few observed samples (Bardossy et al. 1990).

5. 1 Elements of Uncertainty

Uncertainty reflects our lack of perfect understanding of the phenomena and processes involved in addition to the random nature of the events. Some of the sources of uncertainties related to reservoir operation can be summarized as:

- Data inadequacy and errors: temporal and spatial observations may be inconsistent or not representative
- Modeling inaccuracy: parameters, assumptions,
- Randomness of natural phenomena: climate change, extreme events, alga bloom,
- Operational variability: future socio-economic objectives, maintenance,
- Each component and the system as a whole

Several methods have been developed to deal with these elements of uncertainty. These methods are applied in different levels of complexity and their application will depend on the purpose of analysis and source of uncertainty. Some of these methods are probabilistic based techniques, such as Monte Carlo simulation and first order variance estimation, artificial neural networks, Bayesian theory and fuzzy set theory.

In the field of water quality analysis, hydrologists and limnologists might face all of the uncertainties mentioned above. Therefore, uncertainty analysis is found to play an important role in the water quality assessment. To assess the uncertainties related to nutrient and organic matters loads into the reservoir; fuzzy multi-objective fuzzy regression is formulated.

5.2 Fuzzy Regression

Fuzzy regression was first proposed by Tanaka et al. in 1982 in the article named *Linear regression with fuzzy model*. This first proposal has faced some criticisms and other methods have then been reformulated and proposed (Redden and Woodall, 1996; Savic and Pedrycz, 1991).

Generally, regarding the development of the fuzzy regression models, the main discussion focuses on the objective functions of the model. Basically, there are two main methods use to derive the fuzzy regression coefficients (Chang, 2001). The first method is based on fuzzy linear regression (FLR), including the original FLR model proposed by Tanaka et al. and its variations. The second is the fuzzy least-square regression (FLSR), which was first proposed by Diamond (1988).

Tran and Duckstein (2002) introduced the multi-objective fuzzy regression (MOFR). The proposed method is intended to overcome some of the shortcomings of the fuzzy and statistical approaches. It is claimed that fuzzy regression does not take into account all data points and sensitivity to outlying data points. For statistical regression, it is argued that it is difficult to verify distribution assumptions and to deal with insufficient and inaccurate input or output data, and the vagueness of the relation between input and output in statistical approaches (Tran and Duckstein 2002). Moreover, it is said that other methods of fuzzy regression can be considered as specific cases of MOFR.

Here, the idea of fuzzy multi-objective fuzzy regression is considered. However, different performance criteria are formulated and tested in an actual application. The objective is to find the fuzzy regression parameters considering its relation with the observation data and the level of credibility. This is carried out considering three performance criteria: fitness, constraints and vagueness. The performance criteria are modeled as fuzzy membership functions, resulting in a fuzzy multi-objective fuzzy regression (FMOFR). Detailed explanation on the performance criteria is given later.

5.3 Mathematical Formulation

Fuzzy sets theory (Bellman 1959) is used to describe the regression parameters. Each coefficient is represented by a fuzzy number with its fuzzy membership function. The description of a fuzzy number by the L-R representation introduced by Dubois and Prade (1980) is used (Fig. 5.1). Here, L-R is defined as a non-strictly decreasing linear function defined on [0, 1] interval, such as:

$$L(z) = R(z) = 1 \qquad z \leq 0$$
$$L(z) = R(z) = 0 \qquad z > 1 \tag{2}$$

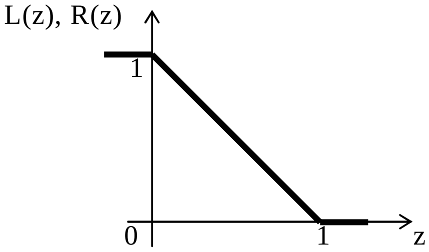

Fig. 5.1 Linear representation of L-R membership function for fuzzy numbers

Curvilinear functions could have been also used. However, as the system already contains some uncertainties, the process of choosing a curve to represent L-R may become another source of uncertainty in the system. For that reason, it is here used and recommended a simpler approach of having linear functions for L and R.

Schematic representation for the fuzzy number is shown in Fig. 5.2. Coefficient *h* is the level of credibility, which represents the confidence in the system.

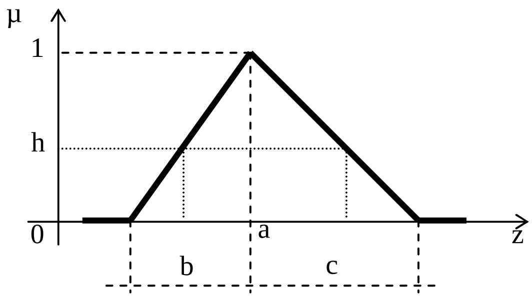

Fig. 5.2 Membership for the fuzzy number *A*(*a, b, c*)

The interval regression, as defined by Chang 2001, is calculated based on the vague width and the level of credibility. The basic equation for the fuzzy regression can be written for the linear case as:

$$y = A1 + A2 \,.\, x \tag{3a}$$

Or, for the general case:

$$y = A1 + \Sigma\,(Ak \,.\, x) \tag{3b}$$

where, *Ai* is a fuzzy number described as *Ai*(*ai*, *bi*, *ci*), where *ai* is the center value and *bi* and *ci* are the fuzzy width, as shown Fig. 5.2.

Based on the previous equations, it is possible to determine the fuzzy regression intervals. The interval is the basis for some of the performance criteria of the FMOFR shown later. The mathematical formulation for the interval is quite simple as shown below:

$$y_{k,low} = (a1 - b1 \,.\, L) + (a2 - b2 \,.\, L) \,.\, x_k \tag{4a}$$

for $k = 1$ to K

$$y_{k,up} = (a1 + c1 \,.\, R) + (a2 + c2 \,.\, R) \,.\, x_k \tag{4b}$$

for $k = 1$ to K

Subject to

$$L(h) = R(h) = 1 - h \text{ and } y_{k,up} \geq 0 \tag{5}$$

Where, $y_{k,low}$ and $y_{k,up}$ are the lower and upper limits for the fuzzy regression, respectively, according to the specified credibility level *h* and observed data *k*. L and R are the left and right fuzzy representation, which is a function of *h*. Moreover, in the application of fuzzy regression for load estimation, values of *y* must be greater than zero, as water quality parameters cannot assume negative values ($y_{k,up} \geq 0$).

Performance Criteria and Objective Function

Three performance criteria – fitness, constraint and vagueness – are formulated as fuzzy membership functions, where the worst value is zero and the best value is one. Fuzzification is considered an easier way to deal with different objectives and its vagueness. Moreover, it is easier for visualization and understanding of the results. In this sense, other criteria might also be formulated and combined for the improvement of fuzzy regression analysis. Besides, it is important to mention that other multi-objective techniques, such as e-constraint and compromise programming, may also be used, helping the decision maker to choose the most appropriate solution.

Fitness (OF1)

The fitness criterion is based on the same criterion used for the traditional regression. It is considered to be equal to the R-square coefficient. It is intended to give the best fit for the fuzzy regression model, including the characteristics of statistic regression.

$$OF1 = \text{R-square} \tag{6a}$$

Constraints (OF2)

During data observation and processing, human, mechanical and methodological mistakes may occur. In a data set some points are clearly outliers and it may be possible to just not consider them. However, it may be a very difficult decision when points are not so clearly apart from the others. Particularly, in water quality analysis, there is a great variability of instruments and methods. Normally, observation is still done manually and laboratory methods may not be so efficient, which may result in useless data.

The constraint criterion is intended to deal with observed data that lies outside the calculated regression interval. It is mathematically defined as below:

$$OF2 = 1 - (\text{No. points out of } [y_{i,low}, y_{i,up}]) / (\text{total No. observation points}) \tag{6b}$$

Vagueness (OF3)

The vagueness criterion aims to give the least vagueness in the model. It is also defined between zero and one. It is based on the characteristics of the fuzzy coefficients of the fuzzy regression.

$$OF3 = \text{average } [ai/(ai+bi) \text{ and } ai/(ai+ci)] \tag{6c}$$

Final Objective Function

The objective function is here considered to be the simple average of the three criteria. Nevertheless, any other technique for combination of criteria may be use, such as maximization and weighted-sum.

$$OF = (OF1 + OF2 + OF3) / 3 \tag{7}$$

5.4 Model-fitting Formulation

Many optimization techniques can be used in the model-fitting process for fuzzy regression. Generally, linear programming (LP) is found to be the most applied method. However, as cited by (Chang and Ayub 2001), application of linear programming may involve some difficulties in certain cases.

As each data set results in two constraints in the fuzzy regression formulation, with increase of observed points, the number of constraints will increase proportionally. This can result in computational limitations using linear programming techniques. Moreover, every time data is added to or removed from the independent variables the whole set of constraints should be reformulated. In some cases, this inconvenience may restrict the experiment process to find the optimal number of independent variables. Other problems related to the negative sign of LP variables can also be another problem when using LP. So, the linear programming formulation can limit the applications of fuzzy regression (Chang and Ayub 2001). In this study, genetic algorithm (GA) model is developed for the fuzzy regression model-fitting process. The basic concept used for the GA model is the same used in the GA model for the training process of the ANN model. Further information about the GA model can be found in Galvao et al. (1999)

5.5 Application and Results

It is well known that loads of nutrients and organic matters transported by rivers depend on a variety of factors, such as discharge, land use, point and non-point pollution sources and weather conditions. However, most of these factors influence loads generally in the long term. Nevertheless, plotting loads versus discharge values of the main affluent of the Barra Bonita reservoir, it is easy to note the strong relation between these two variables.

Owing to the cost of data collection, particularly due to the number of water quality parameters to be considered, observation data in the reservoir area are rather limited. Most of the data is no longer than 7 years, with observation done only once every one or two months.

In many analyses, such as the application of physical model to assess reservoir water quality, it is found necessary to estimate inflow quality values. It is not uncommon to have long observation or synthetic discharge data, but no quality observations. This is a great drawback when water quality models have to be used.

Loads of treatment plant affluent should also be considered when variations of inflows occur. Sometimes, such variation may be too sudden and time for water quality analysis may be insufficient. This is another example where load estimation is found to be extremely important.

The proposed methodology was applied for the load estimation of the two main affluent of Barra Bonita reservoir. The five considered quality parameters are: chlorophyll (CHA), biological oxygen demand (BOD), dissolved oxygen (DO), total nitrogen (TN) and total phosphorous (TP). The fuzzy coefficients of the fuzzy regression may assume different shapes for their membership function. However, here, they are assumed to be symmetric triangles. This guarantees that the main tendency, the regression curve when credibility level equal to 1.0, lay in the center of the vagueness interval.

Plotted fuzzy regression curves are presented in Fig. 5.3a,b and Fig 5.5a,b, for Piracicaba and Tiete rivers, respectively. From the curves for loads, it is easy to find the concentration curve, dividing loads by discharges, Fig. 5.4a,b and Fig. 5.6a,b. Results of fuzzy regression are plotted together with statistic regression curve (single line).

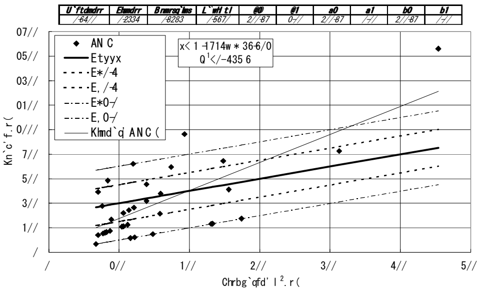

Fig. 5.3a BOD load into Piracicaba river

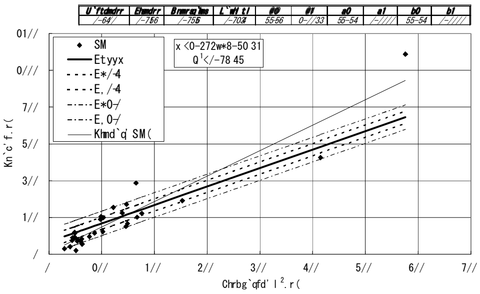

Fig. 5.3b TN loads into Piracicaba river

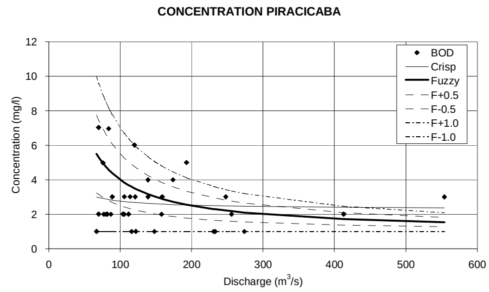

Fig. 5.4a BOD concentration of Piracicaba River

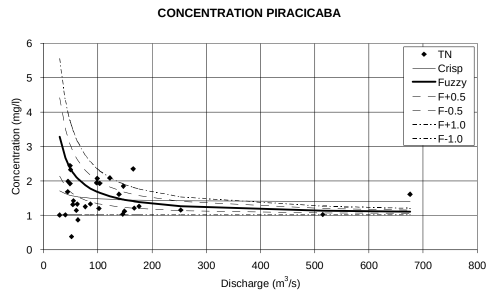

Fig. 5.4b TN concentration of Piracicaba River

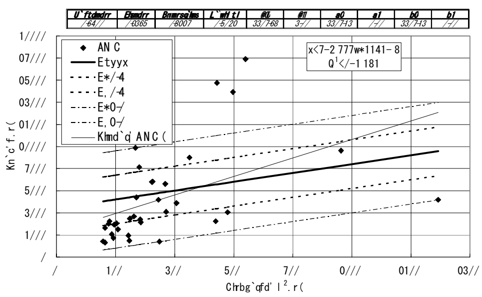

Fig. 5.5a BOD load into Tiete River

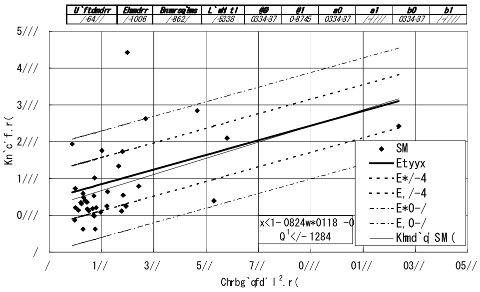

Fig. 5.5b TN load into Tiete River

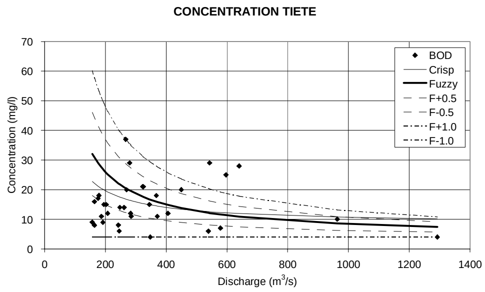

Fig. 5.6a BOD concentrations of Tiete River

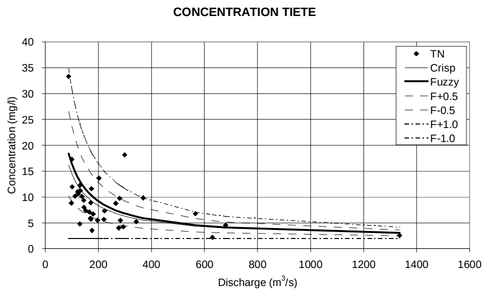

Fig. 5.6b TN concentrations of Tiete River

The fuzzy multi-objective fuzzy regression model has been introduced and illustrated. The fuzzy formulation for the fuzzy regression objectives gives users a much more flexible method to define the fuzzy regression performance criteria. The fuzzy formulation for the MOFR has still the same advantages stated by Tram and Duckstein (2002), combining central tendency and statistical properties and fuzzy regression. It is possible to overcome the shortcomings of these two techniques, fuzzy and statistic regression, when they are applied alone.

As already mentioned before, load is a variable, which does depend on different factors. However, in some cases when data is scarce or due to any other constraints, the correlation between discharge and load can be of much help to load estimation.

6. Fuzzy Stochastic Dynamic Programming

6.1 Dynamic Programming

Dynamic Programming (DP) was developed in the 1950's by RAND Corporation sponsored by the US Air Force. It was named and described in a series of papers by Richard Bellman (1959). DP is used in the resolution of the optimization problem posed here. Applications of dynamic programming to water resources systems can be found in many works, such as Yakowitz (1982) and Esogbue (1989).

DP presents various advantages over other methods to approach water resources management aspects, and can be associated with other programming methods, named such as stochastic DP and fuzzy DP. The basic characteristics of water resources and reservoir operation that lead to the use of DP are: stage-wise structure and non-linearity of the system. DP can also be divided into two different approaches depending on the problem, continuous or discrete – with the latter the most commonly used. One important characteristic of the DP is the possibility to develop the calculation in different directions (known as forward-looking and backward-looking).

The advantages and disadvantages of DP can be summarized from Labadie (1993):

Advantages:

i) DP is suitable for solving sequential decision problems. Reservoir operation can be easily modeled in a stage-sequence problem, facilitating the use of DP, with storage as the state variable and release as the decision variable.

ii) DP allows simple handling of nonlinear modeling. This characteristic can be very important when modeling variables such as water quality and hydropower functions.

iii) Efficiency increases with increasing number of constraints, because possible calculation iterations decrease.

Disadvantages:

The biggest disadvantage is definitely the curse of dimensionality, where required computer time increases linearly with the number of stages and exponentially with the number of state and decision variables considered. The maximum possible number of variables able to be handled with DP is usually six or seven.

Dynamic Programming converts a large, complicated optimization problem into a series of smaller interconnected ones, each containing only a few variables. The result is a series of partial optimizations requiring a reduced effort to find the optimum. The DP algorithm can be applied to find the optimum of the entire process by using the connected partial optimizations of smaller problems.

In each process, the functional equation governing the process was obtained by an application of the following intuitive principle stated by Bellman that says:

"Principle of Optimality. An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

This principle was stated mathematically as the dynamic programming algorithm to maximize a serial process with *n* stages. A unit of time can be represented as a stage, like weeks or months. Every stage has a variety of variables and functions. The evaluation function (E) gives the measure of profit or cost for the stage. Decision variables, for example release and end-of-period storage, are those that can be manipulated independently. State variables, like storage and water quality parameters, are inputs to the stage from an adjacent stage. Consequently, such variables cannot be manipulated independently.

The transition function for the storage state variable is based on the continuity equation. The water quality state variable is based on the results given by the water quality model. The return function depends on the decision and state variables. In order to determine the optimal value for the return function for each stage, it is necessary to exhaustively list individual values of the state variables, searching for the correspondent decision variables.

The objective function is based on the maximum or minimum from the sum of the return functions for each stage. The accumulation of the objective function between each stage will guarantee the continuity of optimization.

(1) Stochastic DP

The variables of stochastic models will have time-dependent probabilities (conditional probability). Some applications of stochastic DP for water resources management is found in Torabi (1973) and Fontane (1997).

(2) Fuzzy DP

Dynamic programming, as stated before, may be combined with other computational techniques. When probabilities of occurrence are unknown (uncertainty), fuzzy logic based models may be applied. In fuzzy logic based models, variables are imprecise or vague, and the source of uncertainty is not merely due to randomness of the natural event.

Another advantage is that the fuzzy optimization approach can address the problem of subjective and noncommensurable objectives in an easily interpretable way. It indicates relatively how each objective has been satisfied. Fuzzyfication allows decision makers to specify the goals and/or constraints in subjective and linguistic terms. In a Fuzzy DP, decision and state variables as well as constraints, can be set as fuzzy membership functions.

6.2 Fuzzy Objective Functions

The objective function that was used in the optimization model is based on fuzzy sets. They represent the degree of satisfaction related to the objectives of the operation. In this work, Fuzzy membership functions refer to four basic water quantity and three water quality objectives assumed for the reservoir operation.

The mathematical formulation for the fuzzy optimization is demonstrated as follows. At each time step (month) n of the DP optimization, the integration of the fuzzy objectives is defined as:

$$f_n(S_n, Q_n) = \sum_{w=1}^{W} \alpha_w \mu_w (S_n, Q_n) \quad (9)$$

where μ_w represents each fuzzy membership function, and α_w stands for the relative weights associated with the *w*th fuzzy element where

$$\sum_{w=1}^{W} \alpha_w = 1 \quad (10)$$

Recursive objective function for the overall maximum for the formulated DP can be written as:

Backward-looking stochastic DP:

$$F_n(S_n, I_{n-1}) = \max_{Sn+1} \left[\sum_{k=1}^{K} [f_n(S_n, R_{n,k}) + F_{n+1}(S_{n+1}, I_{n,k})].P(I_n / I_{n-1}) \right] \quad (11)$$

subjected to:

$$R_{n,k} = S_n - S_{n+1} + I_{n,k} \quad (12a)$$

$$S_{\min n} \leq S_n \leq S_{\max n} \quad (12b)$$

$$R_{\min n} \leq R_n \leq R_{\max n} \quad (12c)$$

where $f_n(S_n, R_n)$: stage (month) return function combining the membership objectives; *I* is the inflow associated with each month, and k refers to the discrete probabilistic inflow; $V_{\min n}; V_{\max n}$: minimum and maximum storage for each stage n; and $R_{\min n}, R_{\max n}$ = minimum and maximum releases for each stage n.

For the quantity optimization the objective function is based on the average of the two quantity evaluation functions.

$$f_n^1(S_n) = \frac{1}{N} \sum_{i=1}^{N} \mu_i \quad (13a)$$

Where *i* represents each quantity objective

For the quality optimization the above membership functions are calculated together with a combination of the quality evaluation function for each parameter, represented as:

$$f_n^2(Q_{i,n}) = \frac{1}{N} \sum_{i=1}^{N} \mu_i \quad (13b)$$

Where *i* represents each quality parameter

So that, the final evaluation function can be written as:

$$f_n(f_n^1, f_n^2) = (f_n^1 + f_n^2)/2 \quad (14)$$

The first two membership functions are related to release from the reservoir. The values 400 and 450 m^3/s of release are assumed to represent maximum satisfaction in the called flow stabilization membership function. They are based on the average historical inflow into the reservoir. The function is applied to guarantee minimum release downstream attending basic needs, such as domestic and industrial supplies and irrigation.

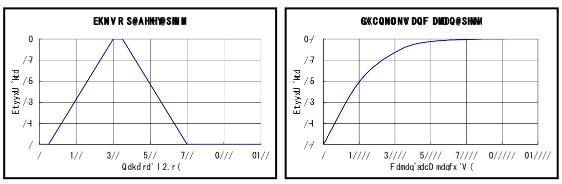

Fig. 6.1 Fuzzy membership functions for flow stabilization and hydropower generation as a function of release and produced energy, respectively

It is very important, particularly in the case of Brazil, to consider power generation in the optimization. The membership function is based on the produced and demanded energy. Due to lack of data regarding the energy production, an average for all months was used here to represent energy demand. It may be improved with realistic data and generation targets for the specific reservoir. Another way to improve the power generation analysis is by using statistical analysis to predict demand. However, for the purpose of this study, a value of 100MW is assumed as the monthly average demand for all months.

The membership functions explained until now are related to quantity objectives. However, for quality analysis it is also necessary to elaborate functions that represent the objectives related to the water quality analysis.

In this research five quality parameters are used total phosphorous, total phosphorous (TP), chlorophyll (CHA), total nitrogen (TN), dissolved oxygen (DO) and BOD. As for the quality optimization, the five fuzzy objective functions are integrated assuming the same relative weight for each. The functions are constructed based on the Brazilian water quality indices and international standards. More on the elaboration of water quality parameters can be found in Bollman and Marques (2000).

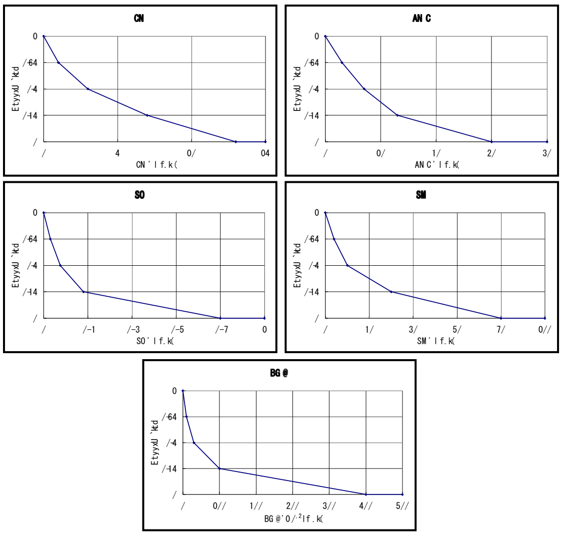

Fig. 6.2 Fuzzy membership functions for water quality evaluation within the reservoir

6.3 Stochastic Analysis

(1) Markov chain

The Markov chain is applicable for discrete-valued processes or, for computational convenience, discrete continuous processes. The inflow into reservoir can be considered as a Markov process, where the system state, within a certain stage, is considered to be dependent on the state of the previous stage and the known probabilities.

For the Markov chain process, a transition matrix of frequencies is constructed. Where inflows in a certain month depend on the previous month inflow. To represent this dependence, a transition matrix containing the conditional probabilities can be obtained by dividing the frequencies by the total number of occurrences for each previous inflow values. This can be seen in the figures below.

TO J(t)

		J_1	J_k	
FROM	I_1	A_{11}	A_{1k}	ΣA_i
I(t – 1)	I_k	A_{k1}	A_{kk}	ΣA_i

Fig. 6.3a Frequency Transition matrix

TO J(t)

		J_1	J_k	
FROM	I_1	$A_{11}/\Sigma A_i$	$A_{1k}/\Sigma A_i$	$\Sigma = 1$
I(t – 1)	I_k	$A_{k1}/\Sigma A_i$	$A_{kk}/\Sigma A_i$	$\Sigma = 1$

Fig. 6.3b Probability Transition matrix

Where, *Ik* and *Jk* are the previous and actual monthly inflow values or intervals, with conditional frequency, having a probability of *Akk/ΣAi* for *Jk* to happen if *Ik* has occurred.

Values related to 10%, 30%, 50%, 70% and 90% of probability are used as the limits for the discrete inflow grid. The actual inflow used in the calculation of storage and release refer to the probability values of 5%, 20%, 40%, 60%, 80% and 95, in other words, the mid points of each interval.

(2) Stationary Policies

Since the transition probabilities repeat every 12 months, the undiscounted stochastic DP calculations are repeated via successive approximations to confirm and guarantee that the optimum guidecurves of end-of-period storage for each month are converging to stationary values. As a result, optimal guidecurves may be applied to each year over the entire operational horizon for any sequence of inflow. As referred to by Labadie (1993), if this procedure converges, then the solution must be optimum.

The result from a stochastic DP model is the set of guidecurves, which give the optimal end-of-period storage as a function of previous inflow and beginning-of-period storage value. The guidecurve for the month of November is showed in Fig. 6.4.

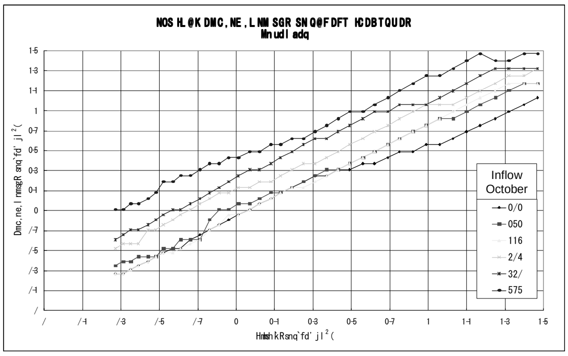

Fig. 6.4 Guidecurves for the month of November

7. Optimization Results

After having concluded the development of all components, as fuzzy regression, ANN and Markov chain, it is possible to combine all of them into the optimization scheme through the fuzzy stochastic DP model, considering water quantity and quality, under uncertainty of inflow quality and stochastic characteristics of inflow quantity.

The results of storage, TP and quantity fuzzy evaluation functions are presented for two situations of inflow load: low and high loads represented by the credibility level of the fuzzy regression being equal to 0.5. Each load scheme is optimized for “only quantity” and “quantity combined with quality” objectives, represented by the QT and QT+QL, respectively, in the legend of the figures below.

The last nine years of the observed historical data is used as the inflow sequence in the optimization schemes.

From Figs. 6.5, it can be clearly seen that quality optimization presents a higher values of storage, which indicates a greater dissolubility of input pollutants. For storage, however, results for low and high loads did not presented visually great difference; however, small variations of storage may result in large variation of release values.

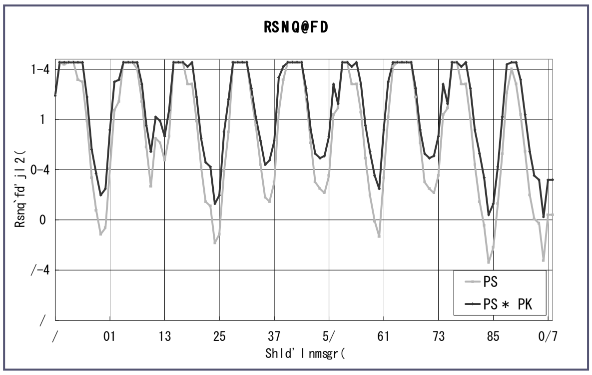

Fig. 7.5a Optimized storage for low load situation

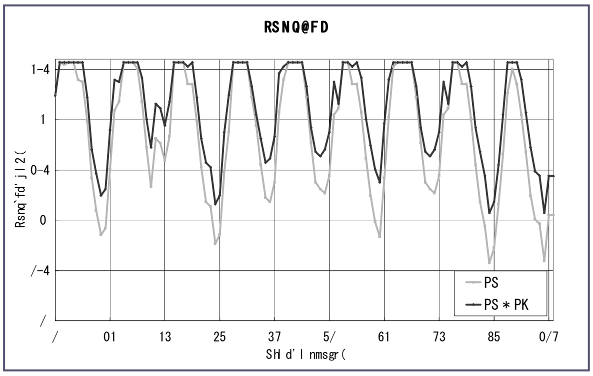

Fig. 7.5b Optimized storage for high load situation

From Figs. 6.6, it can be seen that consideration of quality objectives in the optimization scheme yielded lower concentration values within the reservoir, to what was naturally expected. Moreover, results also showed that a quality optimization may decrease as well the increasing trend of TP concentrations. Other quality parameters results showed the same trends but for the sake of conciseness they are not presented here.

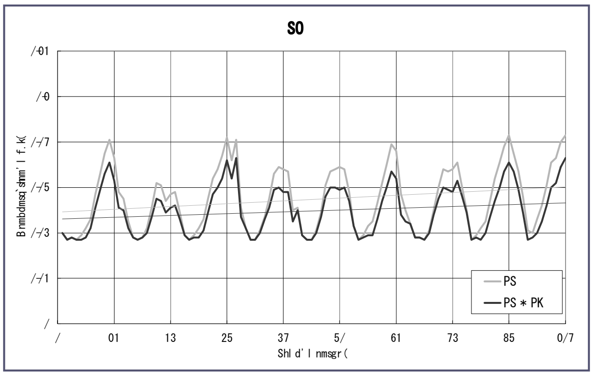

Fig. 7.6a Results of TP for high load situation

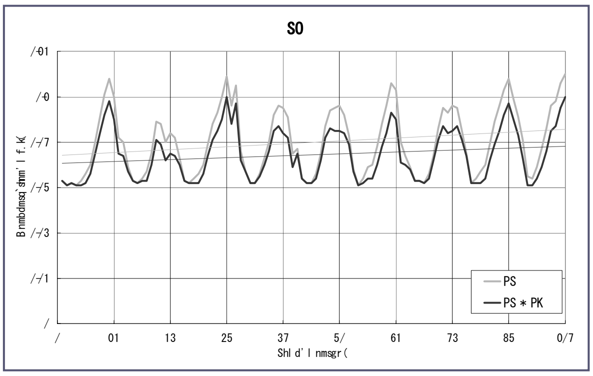

Fig. 7.6b Results of TP for high load situation

Increase of water quality benefits is achieved, but on the other hand, quantity objectives have to pay some of the cost for it. Figs. 6.7 show the lost of quantity objectives for both flow stabilization and hydropower generation, respectively.

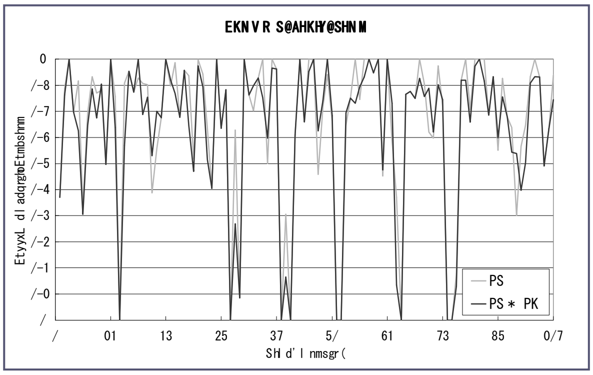

Fig. 7.7a Results of flow stabilization fuzzy evaluation function for low load situation

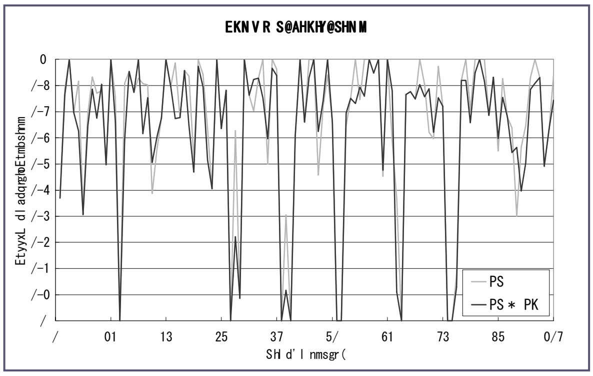

Fig. 7.7b Results for the hydropower fuzzy evaluation function for high load situation

8. Conclusion

Here, the objective functions, especially ones for quantity objectives, were roughly defined. Therefore, it is not easy to say if loss of quantity objectives are preferable than improvements in water quality. To enhance this analysis, a more realistic definition of the objectives and its membership functions is needed. Moreover, economic variables could also be introduced.

Nevertheless, despite of the limitations of each technique, they were found to be effective in dealing with the problems which they were applied to.

The storage reservoir system is successfully optimized accounting for the uncertainties related to input loads. Moreover, stochastic characteristics of inflow are properly handled through the use of Markov chain process.

Further development of the water quality analysis component may be proposed. For example, by increase the number of parameters being used for evaluation or even increasing the accuracy of the water quality simulation model, which can be achieved through the development of more sophisticated models together with a more adequate data collection process.

References

Bardossy, A, et al. (1990): Fuzzy Regression in Hydrology. Water Resource Research, 26 (7), pp. 1497-1508.

Bellman, Richard (1959): Dynamic Programming, Princeton University Press. New Jersey.

Bollmann and Marques (2000): Bases para a Estruturação de Indicadores de Qualidade de Águas. RBRH Vol 5 No.1 Jan/Mar, pp. 37-60. ABRH. Brazil.

Chang, Y.H.O., B.M. Ayyub (2001): Fuzzy regression methods – a comparative assessment. Fuzzy Sets and Systems, 119, pp. 187-203.

Changchit and Terrell (1989): CCGP model for multi-objective reservoir systems. Journal Water Resource Planning and Management, ASCE 115(5), pp. 658-670.

Chaves, P. (2002): Planning Operation of Storage Reservoir for Water Quantity and Quality, Master Thesis, Graduate School of Engineering, Kyoto University, Japan.

Diamond, P. (1988): Fuzzy least square. Information Science, 46, pp. 141-157.

Dombi, J. (1990): Membership function as an evaluation. Fuzzy Sets and Systems, 35, pp. 1-21.

Dubois, D. and Prade, H. (1980): Fuzzy Sets and Systems: Theory and Applications, pp. 393, Academic, California.

El-Swaify and Yakowitz, editors (1998): Multiple Objective Decision Making for Land, Water, and Environment Management, Proceedings of the First International Conference on Multiple Objective Decision Support Systems (MODSS) for Land, Water and Environment Management. CRC Press LLC. Hawaii.

Esogbue, Augustine O., et al. (1989): Dynamic Programming for Optimal Water Resources Systems Analysis. Prentice Hall Inc. USA.

Fontane, D., et al. (1997): Planning Reservoir Operation with Imprecise Objectives, Journal of Water Resources Planning and Management, pp. 154-162, May/June.

Galvao, C. O. and Valenca, M.S. editors (1999): Sistemas Inteligentes: Aplicacoes a Recursos Hidricos e Ciencias Ambientais, ABRH. Brazil.

Hagan et al. (1995): Neural Network Design, PWS Publishing Company, Boston.

Kacprzyk, Janusz (1997): Multistage Fuzzy Control, Wiley. England.

Klir, G. and Yuan, B. (1995): Fuzzy sets and fuzzy logic: theory and applications, Prentice-Hall, Inc., Englewood Cliffs. USA.

Kojiri, Toshiharu (1992): Planning of reservoir Construction through Fuzzy Optimization, DPRI, Japan.

Labadie, J.W. (1993): Combining Simulation and Optimization in River Basin Management, Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization, pp. 345-371. NATO ASI Series. Kluwer Academic Publishers. The Netherlands.

Maier, H. and Dandy, G. (2000): Neural networks for the prediction and forecasting of water resources: a review of modeling issues and applications, Environment Modeling & Software, Vol. 15, pp. 101-124.

Norwick, A. and Turksen, I. (1984): A model for the measurement of membership and consequences of its empirical implementation, Fuzzy Sets and Systems, 12(1), pp.1-25.

Porto, Rubem, et al. (2000): Integration of Water Quantity and Quality in Strategic River Basin Planning, Journal of Water Resources Planning and Management, pp. 85-97, March/April. USA.

Rea, L. and Parker, R. (1993): Designing and conducting survey research: a comprehensive guide, Jossey-Bass Pub., San Francisco, CA.

Redden, D.T., Woodall, W.H. (1996): Further examination of fuzzy linear regression. Fuzzy Sets and Systems, 79, pp. 203-211.

Russel, Samuel and Campbell, Paul (1996): Reservoir Operation Rules With Fuzzy Programming, Journal of Water Resources Planning and Management, pp. 165-170.

Savic, D.A., Pedrycz, W. (1991): Evaluation of fuzzy linear regression models, Fuzzy Sets and Systems, 39, pp. 51-63.

Shrestha, Bijaya (1996): Fuzzy Rule-Based Modeling of Reservoir Operation, Journal of Water Resources Planning and Management, pp. 262-269.

Simonovic, S (1991): Coping with changing objectives in managing an existing multipurpose reservoir, Hydro. Of Natural and Manmade Lakes, Publ. No. 206, G. Schiller et al., eds., Int. Assn. of Hydro. Sci., Wallingford, UK, pp. 181-189.

Torabi, M. and Mobasheri, F. (1973): A Stopchastic Dynamic Programming Model for the Optimum Operation of a Multi-purpose Reservoir, Water Resources Bulletin, American Water Resources Association, Vol. 9, No. 6.

Tran, L., L. Duckstein (2002): Multi-objective fuzzy regression with central tendency and possibilistic properties, Fuzzy Sets and Systems, 130, pp. 21-31.

Tundisi, J.G. and Tundisi, M. (1990): Limnology and eutrophication of Barra Bonita Reservoir, S. Paulo State, Southern Brazil. Arch. Hydrobiol. Beith. Ergebn. Limnology, pp. 661-676. Stuttgart.

Yakowitz, Sidney (1982): Dynamic Programming Applications in Water Resource, Water Resources Research, Vol. 18, No. 4, pp. 673-696.

Zalewski, M. et al. (1997): Ecohydrology: A New Paradigma for Sustainable Use of Aquatic Resources, UNESCO. Paris, France.