FUZZY LOGIC BASED DECISION MAKING FOR HYDROELECTRIC ENERGY GENERATION IN A CASCADED HYDROPOWERPLANTS

Gabriel OLTEAN*, Mihaela GORDAN*, Emilia ŞIPOŞ*, Eugen STĂNCEL**, Ovidiu DANCEA**
**Bases of Electronics Department, Technical University of Cluj-Napoca*
26-28 G. Baritiu str., Cluj-Napoca, Romania, Tel: +40264401470; Fax: +40264591340;
***IPA SA Cifaat Cluj, Zorilor 15, Cluj-Napoca, Romania Tel: +40264-596155,*
Gabriel.Oltean@bel.utcluj.ro

Abstract: Hydroelectric energy is a major source of renewable electricity in the world. The industry is continuously searching for ways to improve the efficiency and reliability with which it produces energy. A method based on subtractive clustering of historical data, to develop fuzzy rule-based systems for optimal hydroelectric energy generation allocation for a hydropowerplant is proposed in this paper. The applicability of this modeling approach is demonstrated by a study on the cascaded hydropowerplants on Somes River in Romania. The optimal planning of energy generation is determined on total demand of energy production for entire hydropowerplants cascade, water storage stage, and reservoir inflows. Two fuzzy systems, FEA1, encapsulating information only about the hydropowerplant under consideration, and FEA2, encapsulating information about all the cascaded hydropowerplants, are developed. Both fuzzy models was able to model, with good results, the optimal operation of the considered hydropowerplant, with a plus for FEA2 that provides more accurate results in the testing data set.

Keywords: *hydroelectric energy generation, hydropowerplant, fuzzy rule-based system, subtractive clustering.*

I. INTRODUCTION

Hydropowerplant cascades operation constitutes an example of a very complex problem involving many decision variables, multiple objectives as well as considerable risk and uncertainty. In addition, the conflicting objectives lead to significant challenges for operators when making operational decisions [1]. A chain of cascaded hydropowerplants provides huge economy end social benefits in hydroelectric energy generation, water supply, flood prevention, irrigation, navigation, fishing, etc. Hydroelectric energy is a major source of renewable electricity in the world. These are the main reasons explaining the significant interest towards the development of decision support tools, integrated to existing hydropower management and operation systems, able to efficiently support the operators in defining the optimal strategy in respect to various tasks and constraints associated to the hydropowerplants cascade operation.

As in many complex decision making and decision support systems, computational intelligence approaches play an important role in hydropowerplant operation planning support (their use was reported even starting from the '90s [2], and is still employed in recent works nowadays [3 - 6]). It is worth mentioning here that, even with the benefit of the powerful algorithms currently available in the class of computational intelligence, the complexity of the optimal operation planning task for hydropowerplant cascades in respect to all the goals (energy production, societal issues, environment requirements, agriculture, and industry requirements) is so large, that it is practically intractable. Therefore existing approaches aims divide it into sub-problems, with simplified goals and constraints that are easier solvable. A significant class of such approaches refers to the problem of reservoirs operation optimization, considered either individually or in a reservoir cascade. Reservoirs are usually operated to satisfy some constrains of individual reservoir in order to maintain specified flows at downstream control points, and to keep the system in balance. Different approaches to the optimal control of the operation of reservoirs, using computational intelligence methods (fuzzy and neuro-fuzzy techniques, neural networks, genetic algorithms, to mention a few), are reported in the literature. The existing recent works vary in respect to the optimization goal and to the tools implied to achieve it.

Thus, in [3], Pinthong et al. propose the application of a hybrid genetic-neuro-fuzzy algorithm to efficiently control the water management in a multipurpose reservoir system. The optimal releases are determined based on the reservoir inflow, storage stage, sideflow, diversion flow from the adjoining basin, and the water demand; the main tasks considered for the reservoir system are to satisfy the water demand and to maximize the flood prevention (no discussion on the hydroelectric power generation being considered here). The system proposed by Pinthong is based on a neuro-fuzzy model of the optimal release, and the genetic algorithm is used to search for the optimal input configuration of the neuro-fuzzy system; their experiments on real data show a high reliability of the operation, and the

Manuscript received February 9, 2009; revised April 16, 2009

predicted releases give the lowest amount of deficit and spill, indicating the usefulness of such an approach. A similar approach is proposed by Mohan and Prasad in [7], where a fuzzy rule based system is designed for modeling and generating the optimal decision policy of the operation of cascaded reservoirs (and illustrated on two cascaded reservoirs) in respect to the water releases; the decision policy was "learned" based on the historical data (that is, the fuzzy rules were inferred from the historical operation of a set of cascaded reservoirs). However, as input data, only the storage and inflow variables were considered (without mentioning the hydroelectric power generation). The optimization of the reservoirs release considering flood control as a constraint is another possible goal of the operation planning for a cascade of hydropowerplants, as addressed and solved by Wei and Hsu [5]. The challenge in the task they propose to solve is to find a procedure applicable to the real time operation of the reservoirs; this is done by the joint use of two models, namely, a hydrological forecasting model and a reservoir operation model, for which they apply the so-called balanced water level index method.

Another category of systems aims at solving the hydroelectric power generation optimization on a hydropowerplant cascade; the approaches differ in respect to the definition of the optimality and to the algorithms employed to achieve it. Thus, in [4], the authors use a niche genetic algorithm to derive a mathematical model for the operation of the hydropowerplant cascade, aiming to maximize the hydroelectric power production for the current given input conditions, while ensuring the demand formulated as needed (guaranteed) output power in the control period. The resulting system was applied by the authors on the cascaded hydropower stations of the Qing River, showing better accuracy than the standard genetic algorithm optimization procedure alone. Another approach of this type proposes to use a novel technique, called Particle Swarm Optimization (PSO), to optimally operate a hydropower station, where the optimality is viewed as the problem of maximizing the generated power while satisfying a set of water constraints/restrictions, namely: water balance restriction; output power restriction; water outflow restriction; water storage restriction [6]. The output of the PSO algorithm is the presentation of the best operation course for the particular hydropower station.

Finally, more complex approaches to the optimal operation planning and decision support of hydropowerplant cascades, which aim to embed various constraints in the task, are worth to be mentioned, as the solutions proposed in [8] and [9]. Thus, Li and Jiang [8] address the short-term hydropower plant scheduling problem to optimize its daily operating income, based on the forecasted day ahead marginal price, on the reservoir volume constraints, on the power generation constraints and the water discharge constraints. This is achieved by creating a model of short-term hydropowerplant scheduling, using improved evolutionary programming techniques. In [9], a similar task is addressed using multi-time Markov decision processes to achieve an efficient hydropower portfolios management with hydropower producers bidding in the regional electricity market (REM) of China. The approach considers a two-level financial optimization of the operation of the hydropower station: one at an "upper level", of decisions in respect with the contract market, where a Markov decision process is used to establish the hydropower contracts trading; and one at a "lower level", in day-ahead market, where an optimal generation management model under the risks of forecast water inflow and electricity prices uncertainties is employed to "adjust", if necessary, the upper level generation strategies to the current situation imposed by the hydropowerplant current conditions.

As resulting from the literature and customers' requests, one of the most important aspects in the operation of cascaded hydropower stations is the optimal programming of electric power generation, but the set of conditions and constrains to be fulfilled can be defined as a rather complex one. Among these, one should definitely mention the following: energy production planning on the national level – which is an important restriction to be obeyed; provision of the necessary water release (outflow) for the downstream reservoir; other requirements are more or less important as well, depending on the particular situation of the hydropowerplant cascade to be operated. The storage state of each reservoir must be kept in the safe range: the water level has to be greater than the minimum energy production level, and below the maximum level to avoid water waste spill, and to accommodate incoming flood flow.

The purpose of this paper is to develop an intelligent decision making system for optimal hydroelectric energy generation allocation for the first hydropowerplant in the cascaded hydropowerplants of the Somes River in Romania. An appropriate way to obtain such an intelligent decision making system is to translate the expertise and knowledge contained in historical data sets into a form that can be manipulated by a computer. This is realized assuming that a large amount of historical data is available and that this data reflects the optimal decision strategy for the hydropowerplant cascade operation planning at all units level. To perform such a difficult task we will use a computational intelligent technique, namely fuzzy rule-based system. These systems have the capability to reconstruct behavior observed in learning sequences, can form rules of inference and generalize knowledge in situations when they are expected to make prediction or to classify the object to one of previously observed categories [10].

The principle of employing fuzzy systems in this fashion is also encountered in [7], but while that work was only concerned with water operation policies, in our proposed solution, we consider the most important criterion is energy production and planning. Another novel aspect of our work is the concept behind the global model of the hydropowerplant cascade: each hydropowerplant operation is individually modeled by a fuzzy rule-based system, which allows a greater flexibility in modeling different configurations of cascades.

Historical daily data are available for a period of almost two year, showing what the operating policy of this hydroenergetic complex was.

II. MODELING APPROACH

Our fuzzy logic based approach for model development of hydroelectric energy generation uses subtractive data clustering.

The objective of data clustering is a partition of data set into clusters of similar data. Operation of automatic data clustering algorithms would result in a fixed structure of data partition, i.e. location and shape of the clusters and membership degrees of each sample to each cluster. Data clustering is a complicated issue as the structures hidden in the data set may have any shapes and sizes. Moreover, the number of clusters is usually unknown.

Generally, data partition should have two important features [10]:

- homogeneity in clusters, i.e. data within a given cluster should be as similar to each other as possible,
- heterogeneity between clusters, i.e. data belonging to different clusters should be as different from each other as possible.

The idea of fuzzy clustering is to divide the data space into fuzzy clusters, each one representing a specific part of the system behavior.

Various methods of data partitioning and algorithms for automatic data clustering exists, subtractive clustering being one of them. Subtractive clustering finds the optimal data point to define a cluster center based on the density of surrounding data points. All data points within a certain distance (radius) of this point are then removed, in order to determine the next data cluster and its center. This process is repeated until all of the data is within the radius of a cluster center. Specifying a smaller cluster radius will usually yield more (smaller) clusters in the data [11].

Subtractive clustering was introduced by [12]. For this method, data points have to be rescaled to [0,1] range in each data dimension. Each data point is assigned a potential P_i, according to its location to all other data points:

$$P_i = \sum_{j=1}^{N} e^{-\alpha \left\| x^i - x^j \right\|^2} \tag{1}$$

where

$$\alpha = \frac{\gamma}{r_a} \tag{2}$$

P_i is the potential value of i-data as a cluster centre

N is the data size

α is the weight between i-data to j-data

x is the data point

γ is the variable (commonly set to 4)

r_a is a positive constant called cluster radius.

The potential of a data point to be a cluster center is higher when more data points are closer. The data point with the highest potential, denoted by P^*_i is considered as the first cluster center $c_1 = (d_1, e_1)$. The potential is then recalculated for all other points excluding the influence of the first cluster center. Again, the data point with the highest potential P^*_k is considered to be the next cluster center $\boldsymbol{c}_k$, and so on. The clustering procedure ends if the following condition is fulfilled:

$$P_k^* < \varepsilon P_i^* \tag{3}$$

where ε is the reject ratio.

Indicative parameters values for $r_a \in [0;1]$ and $\varepsilon = -0.15$ have been suggested in [12].

A sigma value σ_j that specify a range of influence of a cluster center in each of a data dimension is computed for each cluster according with relation:

$$\sigma_j = \frac{r_j \left(x_{j\_\max} - x_{j\_\min} \right)}{\sqrt{8}} \tag{4}$$

where

σ_j is the sigma value for the j^{th} data dimension

r_j is the radius value for the j^{th} data dimension

x_{j_max} is the maximum data value in the j^{th} data dimension

x_{j_min} is the minimum data value in the j^{th} data dimension.

All cluster centers share the same set of sigma value.

To build a fuzzy system that model the data behavior, one will consider one data dimension as output variable and the rest of data dimensions as input variables. Next, the membership of learning data of each cluster should de determined, to obtain the fuzzy sets for all the variables.

For example, if Gaussian membership functions are used, the relation for the membership degree is:

$$\mu_{x_j}\left(A_{kj}\right) = e^{\frac{\left(x_j - C_{kj}\right)^2}{2\sigma_j^2}} \tag{5}$$

where

$k = 1, 2 \ldots K$; K – number of clusters

$j = 1, 2 \ldots M$; M – data dimension

x_j – data on j^{th} data dimension

A_{kj} – fuzzy set fot the k^{th} cluster on the j^{th} data dimension

C_{kj} – cluster centers fot the k^{th} cluster on the j^{th} data dimension

σ_j – sigma value for the j^{th} data dimension.

Each cluster with the associated fuzzy sets is considered as a fuzzy rule describing the system behavior of a specific part of the data set.

An illustration of two fuzzy rules for a three-input TS (Takagi-Sugeno) fuzzy system is further presented with allusion to Figure 3.

*R*1: IF *TotalEnergy* is A_1 and *WaterLevel*1 is B_1 and *Inflow*1 is C_1
THEN *Energy*1 is
$$E_1 = a_1\,TotalEnergy + b_1\,WaterLevel1 + c_1\,Inflow1 + d_1$$

*R*2: IF *TotalEnergy* is A_2 and *WaterLevel*1 is B_2 and *Inflow*1 is C_2
THEN *Energy*1 is

$$E_2 = a_2\,TotalEnergy + b_2\,WaterLevel1 + c_2\,Inflow1 + d_2$$

where A_i, B_i, and C_i are fuzzy sets defined on input variable; E_i are linear functions to determine the value of the output variable with a_i, b_i, and c_i linear coefficient of these functions.

III. MODEL DEVELOPMENT FOR A HYDROPOWERPLANT

The data used for experiments is provided by a SCADA (Supervisory Control and Data Acquisition) system that monitors the hydropowerplants cascade on Somes River basin, and by specific capacity and flow rate – power curves of the reservoirs and hydropowerplants. The artificial lakes created in Somes basin have several purposes, including water supply, electric energy production, and attenuation of floods, pisciculture, and irrigations. The hydropowerplants cascade on Somes River contains eight hydropower plants situated on Somesul Mic and Somesul Cald rivers. The water that supplies these hydropowerplants is taken either directly from the storage reservoirs or by the secondary adductionfeed pipe placed in their area.

The historical data set contains the total demand of energy production for the entire hydropowerplants system (MWh), the current water level (m) and the inflow (m^3/s) for each hydropowerplant, on the daily basis for a period of 678 days for years 2004-2005. The electric energy generated by each power plat is also known. The role of our fuzzy model is to predict the optimal energy to be allocated for generation by the first hydropowerplant in the cascade.

All data was normalized by a linear transformation and mapped in the [0; 1] range. The initial data set was divided into two data sets: a classification set (577 data points) and a testing set (96 data points).

In our first approach we assumed that each hydropowerplant can be operated independently, without considering the influence of the rest of the hydropowerplants. This is why the first fuzzy model (FEA1) is built considering only the inputs specific to this station, namely its current water level, its inflow and the total energy production imposed for the entire complex. The labels used for the linguistic variables of the fuzzy systems are:

- *TotalEnergy* for the total hydroelectric energy to be produced by the entire hydropowerplants complex,
- *WaterLevel*1 for the water level in the reservoir of the first hydropowerplant,
- *Inflow*1 for the water inflow in the reservoir of the first hydropowerplant,
- *Energy*1 for the energy to be produced by the first hydropowerplant.

The structure of this fuzzy model (first order TSK fuzzy logic system) is presented in Figure 1:

Figure 1. Structure of the fuzzy model FEA1

FEA1 fuzzy model was generated in Matlab using the subtractive clustering algorithm to determine the cluster centers and *genfis2* function to determine the fuzzy sets for the inputs and the output, corresponding to each cluster, and also to generate the fuzzy inference system.

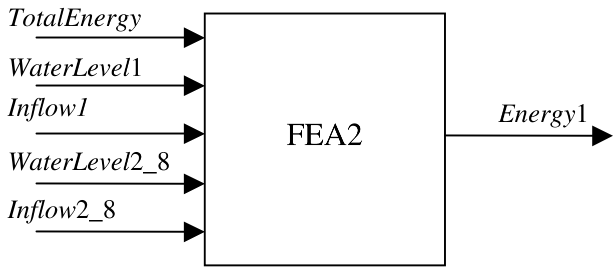

Figure 2. Structure of the fuzzy model FEA2

To study the relevance of the influence of all hydropowerplants in the cascade to one of them, we also developed a more complex model that embeds information about the remaining hydropowerplants in the complex. Two new inputs are added: the normalized sum of the current water level for the other 7 reservoirs, (*WaterLevel*2_8) and the normalized sum of inflow for the other 7 reservoir (*Inflow*2_8). FEA2 models the allocated energy to be generated by the first hydropowerplant as a function of current water level and inflow for the first reservoir, the sum of the current water level and the sum of inflow for the other 7 reservoirs and the total demand of energy. The block diagram of FEA2 is presented in Figure 2.

IV. RESULTS AND DISCUSSION

As was stated above, the cluster radius has an important influence of the number on clusters and consequently on the structure of the generated fuzzy inference systems. In the model development process we have tried different values for cluster radius. For the FEA1 model, best results from the point of view of accuracy was obtained considering for cluster radius a value of 0.5 in all data dimensions. For this value, data was divided into 4 clusters, so the resulted fuzzy system has 4 rules, as one can see in Figure 3. The fuzzy rules are automatically built by the *genfis*2 function in Matlab, that implements the procedure described above in Section II. The fuzzy systems uses 'prod' operator for the "and" method that connects the partial premises of each rule, 'probor' (probabilistic or) for the "or" method, and 'wtaver' (weighted average) as the defuzzification method. The operation of the FEA1 fuzzy system for a particular input can also be seen in Figure 3.

FEA2 model has a similar structure, but with 5 inputs instead of 3. Best results for FEA2 were obtained for a value of 0.2 for the cluster radius in all data dimensions, a fuzzy system with 12 rules being generated.

The energy allocation obtained from our models, compared with the data in the testing data set (reference data) is displayed in Figure 4. Our models provide accurate results, in the majority of data points the simulated values being almost superimposed on the reference data. There are some point with a slightly higher difference between simulated data and reference data. Some aspects regarding the accuracy of the simulation are further presented.

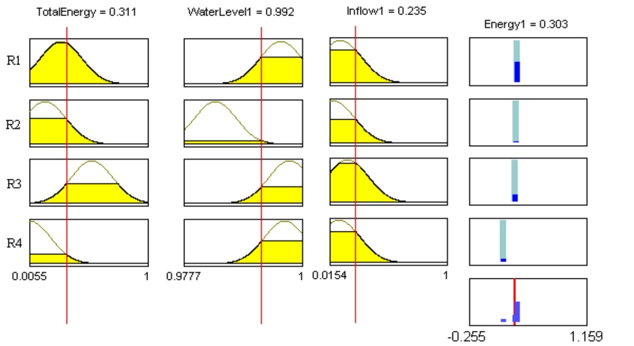

Figure 3. Fuzzy sets and fuzzy rules for FEA1 fuzzy model

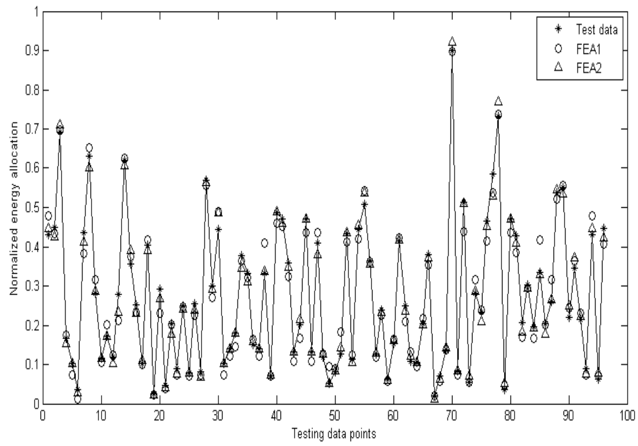

Figure 4. Energy production allocation: computed values by FEA1 and FEA2 compared with reference values from testing data set

Table 1 presents the mean absolute percent error (*MAPE*) and maximum value of absolute percent error (APE_{max}) in the testing data sets for both fuzzy models.

The formula used to compute *MAPE* is:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{Energy1\_FEA - Energy1\_reference}{Energy1\_reference} \quad (6)$$

where
N is the number of testing data
*Energy*1_*FEA* is the energy computed by fuzzy model
*Energy*1_*reference* is the reference energy form testing data set

Table 1. MAPE and APE_{max} for fuzzy models

Errors \ Fuzzy models	FEA1	FEA2
MAPE [%]	11.38	6.47
APE_{max} [%]	76.99	43.28

Hence it appears that the second model (FEA2) is more accurate than the first model (FEA1). This means that the decision on energy generation distribution for one hydropowerplant must consider the characteristics of all (hydropowerplants cascade on Somes River basin).

Percent errors for FEA1 and FEA2 are presented in Figure 5 and respectively Figure 6. FEA1 presents two quite large errors, -76.99% at the 49th data point and +65.61% at the 6th data point. These points can be interpreted as being non-specific data in normal operation of the hydropower plant. If these data point are neglected, the MAPE error drops to a lower value (10.1%), while the maximum value of absolute percent error became 45.17%. FEA2 also gives some larger errors for two data points: -43.28% at 79th data point, respectively +42.77% at 67th data point. If these point are neglected the mean absolute percent error drops to 5.69%, while the maximum value of absolute percent error became 26.19%.

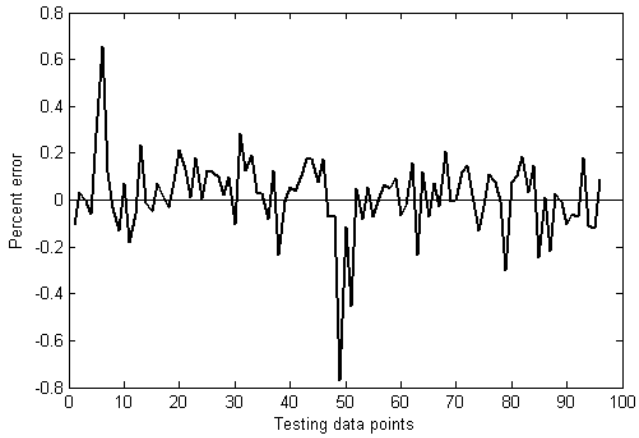

Figure 5. Percent errors in the testing sets for FEA1

The accuracy of our fuzzy models can be also appreciated by a statistical interpretation of error distribution. The statistical distribution of the errors among the testing data set can be derived from the histograms of the errors. Figure 7 present the histogram of absolute percent error (APE) for FEA1. Computations on these statistics show that 53% of the data give APE of at most 10%, and that 90% of the data in testing data sets produces APE less than 23.13%. This maximum statistical bound of the APE for 90% of the testing data is given by the 90% percentile of the APEs in the testing data set, which provides a reliable way of estimating the APE value under which most of the new test data absolute percent errors will fall.

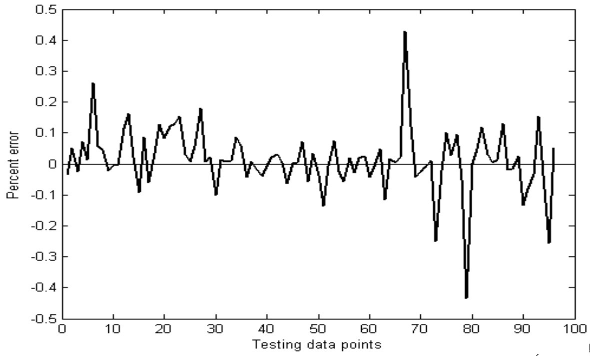

Figure 6. Percent errors in the testing sets for FEA2

whose histogram of APE is shown in Figure 8: 76% of the data points give absolute percent error of at most 10%, and 90% of the data in testing data sets produces APE smaller than 13.59%. This means that, statistically speaking, the model behaves quite well for the large majority of the testing data points (relevant data points), especially FEA2.

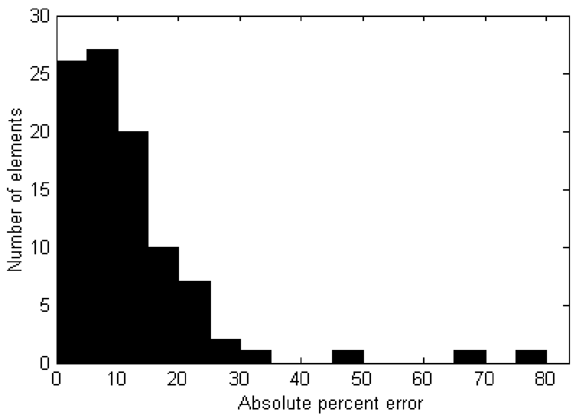

Figure 7. Histogram of APE in the testing sets for FEA1

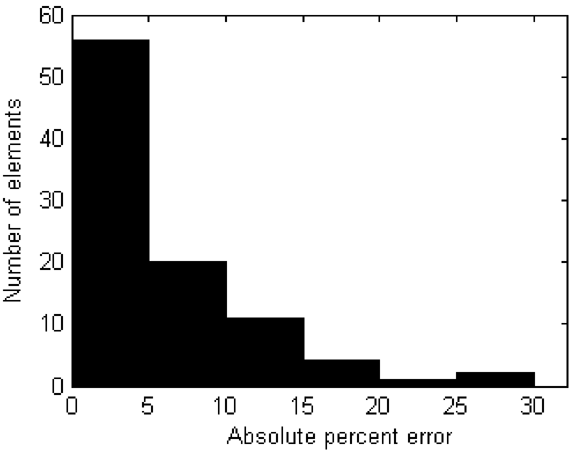

Figure 8. Histogram of APE in the testing sets for FEA2

V. CONCLUSIONS

A new method to build fuzzy models, by subtractive clustering, for the energy generation allocation for a hydropowerplant in a cascaded hydropowerplants system on Somes River basin was presented in this paper.

Two fuzzy models, FEA1, that implies only information regarding the hydropowerplant under consideration and FEA2 that implies extra information about the rest of hydropowerplants in cascade, beside the information regarding the hydropowerplant under consideration was developed. Both fuzzy models was able to capture the information comprised in the data set, with a plus for the second model, FEA2, that present a value of 6.47% for the mean absolute percent error in the testing data set. For FEA2, 90% of the data in testing data sets produces absolute percent error smaller than 13.59%, the model behaving quite well for the majority of the testing data point (relevant data points).

It appears that the operating policy of the cascaded hydropowerplants is a global one, the decision about the energy to be produced by a certain hydropowerplant taking into consideration the information about all the hydropowerplants in the system.

Future work will be conducted on two main directions. First – to obtain more accurate results, a genetic algorithm will be used to optimize the range of influence of each cluster center (radius) in every data dimension. Second – fuzzy models will be built for all hydropowerplants in the Somes River basin, and a complex decision making system will be developed connecting all these models together, to emulate the operation of the entire cascaded hydropowerplants.

REFERENCES

[1] S. Mohan and S. Sivakumar, "Development of Multi Reservoir Systems Operation using DP-Based Neuro-Fuzzy Model: A Case Study in PAP Systems, India", *Proc. 4th INWEPF Steering Meeting and Symposium,* Bangkok, Thailand, 5-7, July 2007, paper 2-04, pp. 1-11.

[2] I.Dabbaghch, . D. Christie, G. W. Rosenwald, and C. Liu, "AI Application Areas in Power Systems", *IEEE Expert: Intelligent Systems and Their Applications*, Volume 12, Issue 1, pp: 58 – 66, ISSN:0885-9000, 1997.

[3] P. Pinthong, A. Das Gupta, M.S. Babel and S. Weesakul, "Improved reservoir operation using hybrid genetic algorithm and neurofuzzy computing", *Water Resources Management*, Vol. 23(4), Page 697-720, March 2009.

[4] N. Li, Y.-D. Mei, "An Improved Genetic Algorithm for Optimal Operation of Cascaded Reservoirs", *Proc. of 2007 International Conference on Computational Intelligence and Security*, 15-19 Dec. 2007 Page(s):110 – 113.

[5] C.-C. Wei, N.-S. Hsu, "Multireservoir real-time operations for flood control using balanced water level index method", *Journal of environmental management,* 2008, Sep, Vol. 88(4), pp. 1624-39.

[6] N. Wang, S. Wang, J. Xie, "Modified Particle Swarm Optimization Algorithm and Its Application on Optimal Operation of Hydropower Station", *Proceedings of the Third International Conference on Natural Computation*, Vol. 4, pp. 549-553, 2007.

[7] S. Mohan and M. Anjaneya Prasad, "Fuzzy Logic Model for Multi Reservoir Operation", *Proc. 2nd Conference of Asia Oceanic sciences Society*, Singapore, 20-24, June 2005, pp. 145-148.

[8] S. Li, C. Jiang, "Short-term Hydropower Station Scheduling

Under Deregulated Environment Based on Improved Evolutionary Programming", *Proc. of the 2006 International Conference on Computational Intelligence and Security*, Guangzhou, China, Nov. 2006, Vol. 1, pp. 242-247.
[9] C. Zhu, J. Zhou, W. Wu, L. Mo, "Hydropower Portfolios Management via Markov Decision Process:, *Proc. of the 32nd IEEE Annual Conference on Industrial Electronics, IECON 2006*, Paris, France, 6-10 Nov. 2006, pp. 2883-2888.
[10] L. Rutkowski, *Computational Intelligence, Methods and Techniques*, Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-76287-4, 2008.
[11] *** *Fuzzy Logic Toolbox 2,* User' s Guide, on line version, 2007, The Math Works, Inc.
[12] S. L. Chiu, "A Cluster Estimation Method with Extension to Fuzzy Model Identification", *Proc. IEEE Internat. Conf. on Fuzzy Systems*, pp. 1240-1245, 1994.