International Journal of Research and Development in Applied Science and Engineering (IJRDASE)

On the use of ANFIS for Predicting Groundwater Levels in Hard Rock Area.

S. Sahay
sulabhsahai@gmail.com

Anurag Banoudha
iisc.anurag@gmail.com

Raghawendra Sharma
robie2007@gmail.com

***Abstract*-Accurate prediction of the ground water level is crucial for optimizing the management of water resources. Since ground water levels are affected by multiple hydrogeological factors, it is difficult to model their complicated, non-linear relationships using conventional approaches. In this paper, two adaptive network based fuzzy inference systems (ANFIS) are presented. These are grid partioned based fuzzy inference system (ANFIS-GRID) and subtractive clustering based fuzzy inference system (ANFIS-SUB). To illustrate the applicability and capability of both these techniques, a hydrograph station in block Naraini, district Banda, India has been chosen as the study area as its ground water resources have been overexploited during the last twenty years and the ground water level has been decreasing steadily. Here different models having different input variables are constructed and the performance of the resultant fuzzy inference system (FIS) has been compared for all the models. FIS have been generated and tested using training and testing data sets. Further model efficiency is evaluated both for training and validating data sets. The results demonstrate that the ANFIS-SUB outperforms ANFIS-GRID due to its fitness in the target problem**

Keywords :-Ground water level, prediction; ANFIS; Grid partitioning; Subtractive Clustering; Model efficiency.

1. Introduction

The importance of groundwater for the existence of human society cannot be overemphasized. Groundwater is the major source of drinking water in both rural and urban India. Besides it is an important source of water both for agricultural and industrial sector. Water resource development projects are vital for countries like India, where high population density demands basic food and fiber requirements. Further more, depletion of ground water supplies and ground water contamination are concerns that will become increasingly important as further aquifer development takes place in any basin. The consequences of aquifer depletion can lead to local water rationing, excessive reduction in yields, wells going dry or producing erratic ground water quality changes, etc.

So constant monitoring of the ground water levels is extremely important. The water levels if properly predicted well in advance can help the administration to plan better ground water utilization. Also, for an overall development of the basin, a continuous forecast of the ground water levels is required to effectively use any simulation model for water management. These models, which may be very simple or highly complex, based on observed data or theoretical principles. provide a framework for decision making for water users and water regulators.

In this direction several studies were carried out for forecasting the groundwater levels using conceptual/physical models that are not only laborious, but also have practical limitations, as many inter-related variables are involved. However, soft computing techniques, which are known for their efficiency in dealing with complicated non linear problems have been applied to may applications such as database management, system design and planning/forecasting of the water resources [Chen et. al., 2006; Chang et. al., 2006; Nayak et. al., 2004b; Firat et. al., 2006]. Here, in the present work two neuro fuzzy systems, ANFIS-GRID and ANFIS-SUB, are employed to model the ground water level data and its influential parameters. To verify the application of this approach Badausa hydrograph station in Banda district, India has been chosen as the case study area. For this quaterly water level records, relative humidity, temperature and rainfall data recorded between the period 1986–2008 years has been used as input variables. All the models are trained and tested and performances of models are compared with observation records. The best fit input structure and method is determined according to performances.

The rest of the paper is organised as follows. Section 2 is a brief introduction to ANFIS, ANFIS-GRID and ANFIS-SUB, section 3 describes study area and data description, section 4 deals with input variables and model structure, section 5 describes performance measurement criteria, section 6 deals with ANFIS model development, section 7 deals with discussion and lastly section 8 deals in conclusion.

2. Adaptive Neural Fuzzy Inference System (ANFIS) methodology

2.1 Basic Theory

The fuzzy logic approach is based on the linguistic uncertain expression rather than numerical uncertainty. It is a soft computing technique that has been widely used in hydrological processes. Since Zadeh (1965) proposed the fuzzy logic approach to describe complicated systems, it has become popular and has been successfully used in various engineering problems, especially on control processes [Chang et. al., 2006; Nayak et. al., 2004a; Firat et. al., 2006; Nayak et. al., 2004b]. Nonetheless, the main problem with this approach is that there is no systematic procedure for a design of fuzzy controller. However, a neural network system has the ability to learn its structure from the input-output sets and adapt itself in an interactive manner. ANFIS, consisted of the combination of the ANN and the fuzzy logic, has been used by many researchers to organize network structure itself and to adapt the parameters of fuzzy system for many engineering problems such as time series forecasting. Fuzzy Inference System is a rule based system consists of three conceptual components. These are: (1) a rule-base, contains fuzzy if-then rules, (2) a data-base, defines the membership function and (3) an inference system, combines the fuzzy rules and produces the system results [Şen, 2001]. First phase of fuzzy logic modeling is the determination of membership functions of input –output variables, second is the construction of fuzzy rules and the last is the determination of output characteristics, output membership function and system results [Fırat et. al., 2006]. Two methods, called as back propagation algorithm and hybrid-learning algorithm, provide learning of the ANFIS and construction of the rules, are used to determine the membership function of the input-output variables. A general structure of fuzzy system is given in Figure 1.

2.2 ANFIS Architecture

ANFIS has been shown to be powerful in modeling numerous processes such as wind speed time series and real-time reservoir operation [Chang et. al., 2006; Fırat et. al., 2006]. ANFIS possesses properties such as capability of learning, constructing, expensing and classifying. It has the advantage of allowing the extraction of fuzzy rules from numerical data or expert knowledge and adaptively constructs a rule base. Moreover, it can adapt the complicated conversion of human intelligence to fuzzy systems. The main difficulty of the ANFIS predicting model is the time required for training structure and determining parameters. ANFIS uses the learning ability of the ANN to define the input-output relationship and construct the fuzzy rules by determining the input structure. The system results were obtained by thinking and reasoning capability of the fuzzy logic [17]. The ANFIS architecture consists of five layers (Figure 2). Here the circles denote a fixed node whereas squares denote an adaptive node. For simplicity it is assumed that the examined FIS has two inputs and one output. For a first order Sugeno fuzzy model, a typical rule set with two fuzzy if-then rules can be expressed as

Rule 1: *IF x is* A_1 *and y is* B_1 *THEN* $f_1 = p_1x + q_1y + r_1$

Rule 2: *IF x is* A_2 *and y is* B_2 *THEN* $f_2 = p_2x + q_2y + r_2$

where, x and y are the crisp inputs to the node i , A_i and B_i are the linguistic labels (low, medium, high, etc.) characterized by convenient membership functions and p_i , q_i and r_i are the consequence parameters (i = 1 *or* 2) .

The model is briefly presented step by step in the following way;

Input nodes (Layer 1): Each node in this layer generates membership grades of the crisp inputs which belong to each of convenient fuzzy sets by using the membership functions. Each node's output Oi^1 is calculated by:

$O_i^1 = \mu_{A_i}(x)$ for i= 1,2 ; $O_i^1 = \mu_{B_{i-2}}(y)$ for i= 3,4 (1)

Where μ_{Ai} and μ_{Bi} are the appropriate membership functions for A_i and B_i fuzzy sets, respectively. Many various membership functions such as trapezoidal, triangular, Gaussian function, etc. can be applied to determine the membership grades. In this study, the gauss membership function is used, as;

$$O_i^1 = \mu_{A_i}(x) = e^{\frac{-(x-c)^2}{2\sigma^2}} \quad (2)$$

Where, $\{a_i, b_i, c_i\}$ is the membership functions' parameter set that changes the shape of membership function from 1 to 0. These parameters are referred to as the premise parameters.

Rule nodes (Layer 2): In this layer, the AND/OR operator is applied to get one output that represents the results of the antecedent for a fuzzy rule, that is, firing strength. It means the degrees by which the antecedent part of the rule is satisfied and it indicates the shape of the output function for that rule. The outputs of the second layer [16], called as firing strengths O_i^2 are the products of the corresponding degrees obtaining from layer 1, named as w given below.

$$O_i^2(x) = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i=1,2 \quad (3)$$

Average nodes (Layer 3): Main target is to compute the ratio of firing strength of each ith rule to the sum of all rules' firing strength. Thus the firing strength in this layer is normalized as;

$$O_i^3 = \overline{w_i} = \frac{w_i}{\sum w_i} \quad i=1,2 \quad (4)$$

Consequent nodes (Layer 4): The contribution of ith rule's towards the total output or the model output and/or the function defined is calculated in Equation (5);

$$O_i^4 = \overline{w_i} f_i = \overline{w_i}\left(p_i x + q_i x + r_i\right) \quad i=1,2 \quad (5)$$

Where, w_i is the ith node's output from the previous layer (i.e., demonstrated in the third layer). {p_i ,q_i , r_i } is the parameter set in the consequence function and also the coefficients of linear combination in Sugeno inference system.

Output nodes (Layer 5): This layer is called as the output notes in which the single note computes the overall output by summing all the incoming signals and is the last step of the ANFIS. Hence, each rule's fuzzy results are transformed into a crisp output in this layer by defuzzification process, as;

$$f(x,y)=\frac{w_1(x,y)\,f_1(x,y)+w_2(x,y)\,f_2(x,y)}{w_1(x,y)+w_2(x,y)}=\frac{w_1 f_1+w_2 f_2}{w_1+w_2} \quad (6)$$

$$O_i^5=f(x,y)=\sum_i \overline{w}_i f_i=\overline{w}_i f_1+\overline{w}_i f_2=\frac{\sum_i w_i f_i}{\sum_i w_i} \quad (7)$$

The objective is to train adaptive networks to be able to have convenient unknown functions given by training data and be able to find the proper value of the input and output parameters. For this aim, ANFIS applies the hybrid –learning algorithm, consists of the combination of the “gradient descent method” and “the least-squares method”. The gradient descent method is used to assign the nonlinear input parameters (a_i ,b_i ,c_i) , as the least-squares method is employed to identify the linear output parameters (p_i , q_i , r_i) . The antecedent parameter, i.e., membership function given in layer 2 is applied to construct the rules of the ANFIS model. Since the input variables within a range might be clustered into several classes, the structure of input layer needs to be determined accurately. The “subtractive fuzzy clustering” function offering the effective result by less rules, is applied to solve the problem in ANFIS modeling.

2.3 ANFIS-GRID

The ANFIS-GRID fuzzy inference system is the combination of grid partition and ANFIS. Grid partition divides the data space into rectangular sub-spaces using axis-paralleled partition based on pre-defined number of membership functions and their types in each dimension. Premise fuzzy sets and parameters are calculated using the least square estimate method based on the partition and MF types. When con- structing the fuzzy rules, consequent parameters in the linear output MF are set as zeros. Hence it is required to identify and refine parameters using ANFIS.

The wider application of grid partition in FL and FIS is blocked by the curse of dimensions, which means that the number of fuzzy rules increases exponentially when the number of input variables increases. For example, if there are averagely m MF for every input variable and a total of n input variables for the problem, the total number of fuzzy rules is m^n. It is obvious that the wide application of grid partition is threatened by the large number of rules. According to Jang, grid partition is only suitable for cases with small number of input variables (e.g. less than 6).

.

3.2.3. ANFIS-SUB

The ANFIS-SUB fuzzy inference system combines the subtractive clustering method and ANFIS. The subtractive clustering method is proposed by Chiu by extending the mountain clustering method. It clusters data points in an unsupervised way by measuring the potential of data points in the feature space. When there is not a clear idea how many clusters there should be used for a given data set, it can be used for estimating the number of clusters and the cluster centers. Subtractive clustering assumes that each data point is a potential cluster center and calculates the potential for each data point based on the density of surrounding data points. Then data point with highest potential is selected as the first cluster center, and the potential of data points near the first cluster center (within the influential radius) is destroyed. Then data points with the highest remaining potential as the next cluster center and the potential of data points near the new cluster center is destroyed. The influential radius is critical for determining the number of clusters. A smaller radius leads to many smaller clusters in the data space, which results in more rules, and vice versa. Hence it is important to select proper influential radius for clustering the data space.

After clustering the data space, the number of fuzzy rules and premise fuzzy MF are determined. Then the linear squares estimate is used to determine the consequent in the output MF, resulting in a valid FIS. As described above, ANFIS learns and refines the premise fuzzy MF and consequents using the least squares estimate and back propagation. Tuned by ANFIS, the resultant FIS achieves minimum training errors.

The combination of ANFIS and subtractive clustering has been widely applied in automation control, function approximation and resolving engineering problems.

3. Study area and data description

To investigate ANFIS as a robust method for solving non-linear problems such as groundwater level forecasting, Badausa hydrograph station in district Banda, Uttar Pradesh has been chosen as the study area. It is located in south western part of Uttar Pradesh. It lies between latitude $25^0$00” and $25^0$59’00” and longitude $80^0$06’00” and $81^0$00’00”.

3.1 Geology of Banda district

Geologically the district is characteised by Bundelkhand granit/gneiss, Vindhayan formation and younger alluvium near the bank of river Ken and Yamuna.

3.2 Climate

The district is characterised by extreme hot and cold weather as observed in other parts of Bundelkhand. Between November and March the weather is cool and from March onwards the temperature gradually rises, becoming unbearable between May and June.

3.2.1 Rainfall

The Rainfall as southwestern monsoon from Arabian sea take place in the district from June to October. To monitor the amount of precipitation, four raingauge stations are existing in the district at Blocks Banda, Naraini, Baberu and Kamasin. As per IMD report, the normal annual rainfall in the district is 980.0 mm, however it varies over different parts of the district. Deficient rain and high rate of evapotranspiration causes draught conditions.

3.2.2 Temperature

Banda is in general hot and becomes unbearable during summer season. It varies from 9.1^{0}c to 43.1^{0}c Mercury shoots during May and reaches even upto 48-49^{0}c sometimes.

3.2.3 Humidity

It is less during April and May and becomes maximum during August. The average relative humidity in the morning is 61.58%, while it becomes 51.00% during evening.

3.3 Hydrology

The district is endowed with huge surface water resources available from rivers viz. Ken, Yamuna and Baghain. Surface water resources is available from tanks, rivers and canals.

3.4 Data description

The data records consists of ground water level data (recorded four times a year) between the period 1986 - 2008 are procured from Central Ground Water Board and rainfall and surface data (temperature, relative humidity, etc.) are procured from IMD, Pune. The plot of time series record of water level and rainfall data for the period 1986 to 2008 is shown in fig. 3.

4. Input Variables and Model Structure

One of the most important steps in the development of any prediction model is the selection of appropriate input variables. The main reason for this is that ANFIS belongs to the class of data-driven approaches. The ground water level forecasting can be characterized as a function of various variables such as spatial and temporal distribution of rainfall, relative humidity, temperature and previous water level records. Here quaterly water level data has been simulated.

Normally, the data set for ANFIS needs to be divided into three parts. The first part is for the training, the second part for validation and the third part for testing. However, because the length of our sample data was not very big, we considered only two parts: training and testing. The only difference between a testing phase and a validation phase is that if the error rate of the validation phase increases, then the training stops. In this study, those two terms are used synonymously. Different combinations of Water level (WL), Rainfall (RF), Temperature(T) and Relative Humidity(RH) has been used as input variables, keeping the number of input variables to three. Also to make the training data cover all the characteristic of the problem in order to get effective forecasting model, the data has been further subdivided into three sub models based on the principles of Cross Validation. Here the structure of forecasting model is depicted in Table–1, cross validated data sets in Table–2 and data selection types using cross validated data sets in Table-3

5. Performance Measuring Criteria

The performance of ANFIS models using training and testing data are evaluated and compared using Correlation Coefficient (r), Efficiency (E) and Root Mean Square Error (RMSE)

5.1 Correlation Coefficient (r)

It provides information on the strength of linear relationship between the observed and the computed values. The value r close to 1.0 indicates good model performance and can be calculated using the following formula,

$$r = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^{n}(x_i - x)^2 \sum_{i=1}^{n}(y_i - y)^2}}$$

5.2 Root Mean Square Error (RMSE) :-

It is the most easily interpreted statistic, since it has the same units as the parameters estimated. The RMSE is thus the difference, on average, of an observed data and the estimated data. RMSE evaluates the residual between measured and forecasted ground water level. RMSE is a frequently-used measure of the difference between values predicted by a model or an estimator and the values actually observed from the thing being modeled or estimated. These differences are also called residuals. Theoretically, if this criterion equals zero then model represents the perfect fit, which is not possible at all.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n}}$$

5.3 Efficiency (E)

The efficiency (E) is one of the widely employed statistics to evaluate model performance. An efficiency of 1 (E=1) corresponds to a perfect match.

$$E = \frac{E_1 - E_2}{E_1} \quad E_1 = \sum_{i=1}^{n}\left(x_i - \overline{x_i}\right)^2 \quad E_2 = \sum_{i=1}^{n}(y_i - x_i)^2$$

where x_i = observed ground water level

$\overline{x_i}$ = mean of x_i

y_i = predicted ground water level

$\overline{y_i}$ = mean of y_i

n = number of data sets used for evaluation

6. ANFIS Model Development

In this study firstly, the five models having various input variables are trained and tested by ANFIS method using MATLAB R2009a software and the performances of the ground water level forecasting models are compared and evaluated based on training and testing performances. The models generate Fuzzy Inference System structure from data using subtractive clustering and grid partitioning.

When generating an FIS using ANFIS, it is important to select proper parameters, including the number of MF for each individual antecedent variable. It is also important to select proper parameters for the training and testing process, including the initial step size S, the step size increase rate R_{Inc}, and the step size decrease rate R_{Dec}. For specific training and testing data sets, we analyze the effect of these parameters on the final ANFIS performance, including the training and testing RMSE.

In case of ANFIS-GRID the number of membership functions used is two and the type of membership function is Generalised bell and the output membership function is linear where as in case of ANIS-SUB, Gaussian membership function has been used as the input membership function and linear membership function for the output function. Here separate sets of input and output data has been used as input arguments. Subtractive clustering has been used as the rule extraction method to determine the number of rules and antecedent membership function and least square estimation to determine each rule's consequent equation. The hybrid learning algorithm, which is the combination of least square estimation and back propagation gradient descent has been applied in both the cases. The best fit model structure is determined according to criteria of performance evaluation. The Cross correlation, Efficiency and RMSE performances of the ANFIS models are given in tables 4, 5, and 6 and their comparative graphs are shown in Fig. 4(a,b,c,d and e), 5(a,b,c,d and e) and 6(a,b,c,d and e) respectively.

7. Discussion

Comparing the results of all the five models using the performance criteria, it is seen that for small size training data, ANFIS_SUB has outperformed ANFIS_GRID. The performance of ANFIS_GRID is hard to predict due to the curse of dimensionality whereas in case of ANFIS_SUB proper selection of influential radius which affects the clusters results directly has resulted in reduction of RMSE. It is seen that the best developed model for ground water level forecasting is the M4 model with input variables as groundwater level, temperature and relative humidity. Using the cross validated data for M4 model, it is seen that M4 MD3 is the best predictive model, with MD3 being the best cross validated data wherein the training data covers all the characteristic of the problem in order to get effective forecasting model. This M4 MD3 model is very close to M4 MD2 model in terms of the performance criteria. It appears that both the M4 MD3 and M4 MD2 models are generally accurate and the values of RMSE are small enough and corr. Coeff. and efficiency values close to unity as compared to other models. The prediction versus observed value graphs for M4 MD3 and M4 MD2 models are shown in fig. 7 and fig. 8 respectively. The Gaussian membership function curves and generalized membership function curves for three input variables for M4 MD3 model are shown in figure 9 and 10 respectively. Closely following the M4 model is the M3 model, having input variables as water level, rainfall and relative humidity. This model is followed by M5, M2 and M1 in decreasing order of best fit model developed. From the above tables 4, 5 and 6 it is seen that Model M4 MD3 has E, Corr. Coeff. and RMSE values both for training and testing datasets as (0.54, 0.73), 0.73, 0.89) and (1.48, 1.11) respectively. The least developed model in terms of model efficiency is the M1 model with all the three input variables as ground water levels only. The results of ANFIS models demonstrate that ANFIS can be applied accurately to establish accurate and reliable ground water level forecasting.

8. Conclusions

As an efficient neuro-fuzzy system, ANFIS can be applied to learn FIS and to identify and refine the antecedent and consequent parameters in MF and fuzzy rules using training data sets. It provides an effective approach for many complicated engineering problems in various fields. In this paper, studies on ANFIS-GRID and ANFIS-SUB indicate that selection of appropriate neuro-fuzzy systems depends on the problem and available data sets.

ANFIS_SUB has out performed ANFIS_GRID which is known for the "curse of dimensionality". In problem modeling using ANFIS-SUB, it is important to investigate the size of the training set and the architecture of FIS, with the size of the training set being larger than the total number of parameters (e.g. for premise fuzzy sets and linear output) in the FIS.

In the present study, applicability and capability of ANFIS techniques for ground water level forecasting has been investigated. For this ground water level data of Badausa hydrograph station along with temperature, relative humidity and rainfall data of Banda district were chosen as the study area. Five models (M1 to M5) having different input variables were trained and tested using ANFIS techniques. M4 model was found to be the best forecasting model with water level, temperature and relative humidity as the input variables, closely followed by M3 model with water level, relative humidity and rainfall as the three input variables. and the least developed forecasting model was M1 model with all the three input variables as water level. Thus the best developed model does not have rainfall as one of the input variables, although rainfall is the most important influencing factor for ground water level in an area. However the second best model has rainfall as one of the input variables may be given more preference, being a more practical model.

References:

[l] Abonyi J., Andersen H. J., Nagy L., Szeifert F., 1999, Inverse fuzzy-process-model based direct adaptive control, Mathematics and Computers in Simulation l19–132.

[2] Chang F.J., Chang Y.T., 2006 : Adaptive neuro-fuzzy inference system for prediction of water level in reservoir, Advances in Water Resources, 29 1-10.

[3] Chiu S., 1994, Fuzzy model identification based on cluster estimation, Journal of Intelligent and Fuzzy Systems 2 (3) 267–278.

[4] Chiu S., 1997, Extracting fuzzy rules from data for function approximation and pattern classification, in: D. Dubois, H. Prade, R. Yager (Eds.), Chapter 9

in the Fuzzy Information Engineering: A Guided tour of Applications, Springer, Berlin, pp. l49–l62.

[5] Demirli K., Cheng S.X., Muthukumaran P., Subtractive clustering based modeling of job sequencing with parametric search, Fuzzy Sets and Systems l37 (2) 235–270.

[6] Demirli K., Muthukumaran P., 2000, Higher order fuzzy system identification using subtractive clustering, Journal of Intelligent and Fuzzy Systems 9 (3–4) l29–l58.

[7] Firat M., 2008, Comparison of Artificial Intelligence Techniques for river flow forecasting, Hydrol. Earth Syst. Sci., 12, 123-139.

[8] Fritzke B., 1997, Incremental neuro-fuzzy systems, in: Proc. Application of Soft Computing, SPIE International Symposium on Optical Science, Engineering and Instrumentation, San Diego, pp. 86–97.

[9] Jang J.R., 1993, ANFIS: A daptive-network-based fuzzy inference system, IEEE transaction on Systems, Man and Cybernetics 23 (3) 665–685.

[10] Jang J.R., Frequently asked questions – ANFIS in the fuzzy logic toolbox, <http://www.cs.nthu.edu.tw/~jang/anfisfaq.htm>.

[l1] Jang J.R., Sun C., Mizutani E., 1997 Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice Hall Inc., Englewood Cliffs.

[l2] Kasabov N., Song Q., 2002, DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction, IEEE Transactions on Fuzzy Systems l0 (2) l44–l54.

[13] Nayak P.C., Sudheer K.P., et al., 2004a: Fuzzy computing based rainfall-runoff model for real time flood forecasting, Hydrol. Process., 17, 3749-3762.

[14] Sen Z.,: 2001, Fuzzy Logic and Foundation, ISBN 9758509233, Bilge Kultur Sanat Publisher, Istanbul, 172 pp.

[l5] Zadeh L.A., 1965, Fuzzy sets, Information and Control, 338–353.

[16] S. Sahay, et. al., " On the use of ANFIS for Ground Water Level Forecasting in an Alluvium Area" International Journal of Research and Development in Applied Science and Engineering, Volume 2, Issue 1, November 2012.

[17] S. Sahay, et. al.., "Comparative Study of Soft Computing Techniques for Ground Water Level Forecasting in a Hard Rock Area" International Journal of Research and Development in Applied Science and Engineering, Volume 4, Issue 1, November 2013.