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

Comparative Study of Soft Computing Techniques for Ground Water Level Forecasting in a 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. In this direction use of Soft Computing Techniques is becoming increasingly important in the modeling and forecasting of hydrological and water resource processes. In this study applicability of an adaptive network-based fuzzy inference system (ANFIS) was used to build a prediction model which can simulate the trend of the ground water level and provide an acceptable predictions. To illustrate the applicability and capability of the ANFIS, 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. Different models having different input variables are constructed and the best one is investigated. Further model efficiency is evaluated both for training and validating data sets. Also the best developed model is trained and tested by Artificial Neural Network (ANN) . The results of ANFIS and ANN models are compared and evaluated. The results demonstrate that the ANFIS can be applied successfully and provide high accuracy and reliability for ground water level prediction.**

Keywords-- Soft computing techniques; Ground water level forecasting; ANN; ANFIS; Model Efficiency; Water Resource Management

1. Introduction

Ground water is one of the major sources of supply for domestic, industrial and agricultural purposes. Long term systematic measurements of water levels provide essential data needed to evaluate changes in resources over time, to develop ground water models and to forecast trends and design, implement and monitor the effectiveness of ground water management and protection programs. 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. 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.

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, ANNs have been found as a potentially useful tool for modeling complex non-linear systems and widely used for prediction. The ASCE Task Committee report did a comprehensive review of the application of ANNs to hydrology, as did Maier and Dandy, and also Govindaraju and Rao in a specialized publication.

Recently, Adaptive Neuro-fuzzy inference system (ANFIS), which consists of the ANN and fuzzy logic methods, have been used for many application such as, database management, system design and planning/forecasting of the water resources [Chen et. al., 2006; Chang et. al., 2006; Chang et. al., 2001; Da Silva et. al., 1999; Nayak et. al., 2004b; Firat et. al., 2006].

The main purpose of this study is to investigate the applicability and capability of ANFIS and ANN methods to develop ground water level forecasting models and to compare them in terms of performance criteria.

2. Artificial Neural Networks (ANN) Methodology

An ANN can be defined as a system or mathematical model consisting of many nonlinear artificial neurons running in parallel which can be generated as one or multiple layered. In this study, Feed Forward Neural Network (FFNN) method has been used for forecasting of ground water level. A FFNN consists of at least three layers, input, output and hidden layer. The number of hidden layers and neurons in hidden layer are determined by trial and error method. The strength of connection between the two layers is determined by the weights W_{ij} . The schematic diagram of a FFNN is shown in Fig. 1

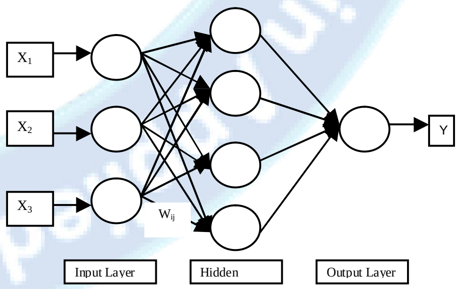

Fig. 1. Schematic diagram of a FFNN

Each neuron in a layer receives weighted inputs from a previous layer and transmits its output to neurons in the next layer. The summation of weighted input signals is calculated by Eq. (1) and is transferred by a nonlinear activation function given in Eq. (2). The responses of network are compared with the observation results and the network error is calculated with equation (3)

$$Y_{net} = \sum_{i=1}^{N} X_i . w_i + w_0 \qquad (1)$$

$$Y_{out} = f(y_{net}) = \frac{1}{1+e^{-Y_{net}}} \qquad (2)$$

$$J_r = \frac{1}{2}.\sum_{i=1}^{k} (Y_{obs} - Y_{out})^2 \qquad (3)$$

Y_{out} is the response of neural network system, f (Y_{net}) is the nonlinear activation function, Y_{net} is the summation of weighted inputs, X_i is the neuron input, w_i is weight coefficient of each neuron input, w_0 is bias, J_r is the error between observed value and network result, Y_{obs} is the observation output value

3. Adaptive Neural Fuzzy Inference System (ANFIS) Methodology

ANFIS has been shown to be powerful in modeling numerous processes such as wind speed time series and real-time reservoir operation [Chen et. al., 2006; 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. ANFIS uses the learning ability of the ANN to define the input-output relationship and construct the fuzzy rules by determining the input structure [19]. The system results were obtained by thinking and reasoning capability of the fuzzy logic. 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* $f_1 = p_1 * x + q_1 * y + r_1$

Rule 2: *IF x is* A_2 *and y is* B_2 *THEN* $f_2 = p_2 * x + q_2 * y + r_2$

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) .

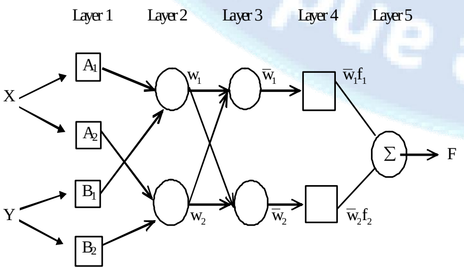

Fig. 2 ANFIS Architecture

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

Layer 1 Each node in this layer generates a membership grades of linguistic label. A_1, A_2, B_1 and B_2 are the linguistic labels which used to define the membership functions. Every node I in this layer is an adaptive node. Parameters in this layer are called premise parameters.

Layer 2. Every node in this layer is a fixed node labeled Π, whose output is the product of all the incoming signals. Each node output represents the firing strength of a rule.

Layer 3. Every node in this layer is a fixed node labeled N. The ith node calculates the ratio of the ith rule's firing strength ($\overline{w_1}$ and $\overline{w_2}$). Thus the outputs of this layer are called normalized firing strengths.

Layer 4. Every node i in this layer is an adaptive node. Parameters in this layer are referred to as consequent parameters {p_i ,q_i , r_i }.

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.

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 modelling.

4. Study area and data description

To investigate the ANN and 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". Geologically, Banda district is characterised by Bundelkhand granit/gneiss, Climate is extreme hot and cold weather as observed in other parts of Bundelkhand. Rainfall as southwestern monsoon from Arabian sea take place in the district from June to October. Temperature is in general hot and becomes unbearable during summer season.

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.

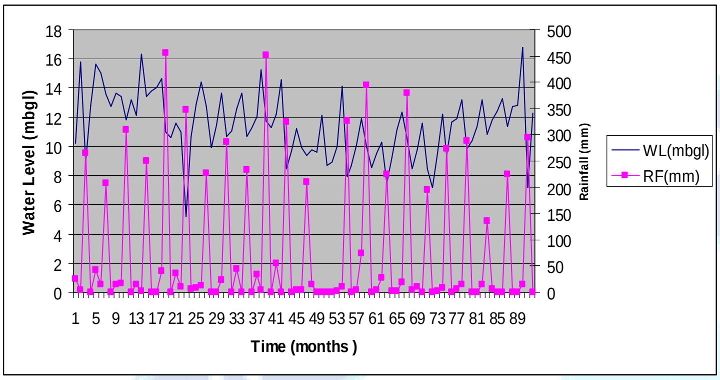

Figure 3 Time series plot of water level and rainfall

5. 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 data set both for ANN and ANFIS has been divided into training and testing. 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. 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 and cross validated data sets in table – 2

Table – 1 Structure of Forecasting model both for ANN and ANFIS

Model	Input Variables	Output Variable
M1	WL(t-3), WL(t-2), WL(t-1)	WL(t)
M2	WL(t-2), WL(t-1), RF(t)	WL(t)
M3	WL(t-1), RF(t), RH(t)	WL(t)
M4	WL(t-1), T(t), RH(t)	WL(t)
M5	WL(t-1), T(t), RF(t)	WL(t)

where, WL(t, t-1, t-2, t-3) is the water level at time periods (t), (t-1), (t-2) and (t-3) RF = Rainfall; RH = Relative Humidity; T = Temperature; WL = Water Level

Table – 2 Cross Validation of data for selection of training / testing datasets.

Cross Validated Data	Data Period	Number of data
CV1	1986 - 1993	30
CV2	1994 – 2001	30
CV3	2002 - 2008	29

Table – 3 Data Selection for optimum water level prediction

Main Models	Sub - models	Cross Validation data		
		CV1	CV2	CV3
M1 to M5	MD1	Testing	Training	Training
	MD2	Training	Testing	Training
	MD3	Training	Training	Testing

where Training = TRG, Testing = TST

6. Performance Measuring Criteria

The performance of both the ANN and ANFIS models using training and testing data are evaluated and compared using Correlation Coefficient (r) and Root Mean Square Error (RMSE)

Correlation Coefficient (r) 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,

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

Root Mean Square Error (RMSE) is the difference, on average, of an observed data and the estimated data. RMSE evaluates the residual between measured and forecasted ground water level.

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

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

7. ANN Model Development

In this study five models M1, M2, M3 M4 and M5 using Feed Forward Neural Network (FFNN) are constructed with Back Propagation learning algorithm using EasyNN Plus Software. Here the summation of the input signals is transferred to next layer using sigmoid activation function. The faster learning of the network has been achieved by selecting optimum learning rate and momentum, both found by trial and error. For better performance of the network, the number of hidden layers has been kept to minimum, i.e one and the number of neurons in this layer has been varied. In the training and testing of the ANN models, the same data set is used and performances of the models is evaluated using the above mentioned criteria. Figure 4 shows the performances of all the models with respect to RMSE and Corr. Coeff.

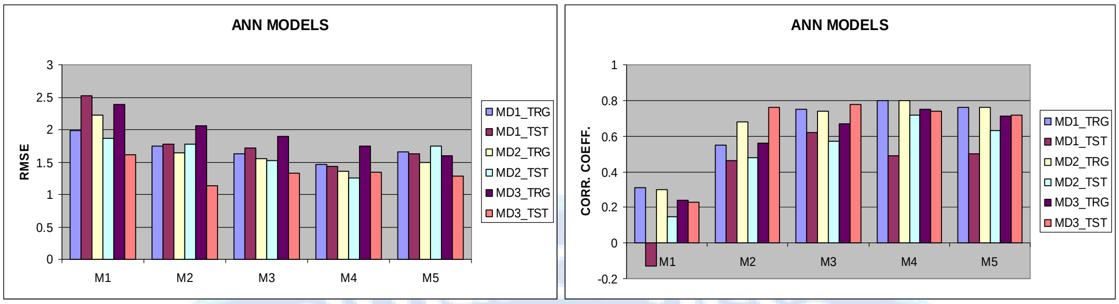

Figure 4 Plot of ANN model for RMSE and Corre. Coeff. values

The details of the Corr. Coeff. and RMSE values for of all the models are given below in table 4.

Table 4. ANN model performance values

	Correlation Coefficient					RMSE				
MODELS	**M1**	**M2**	**M3**	**M4**	**M5**	**M1**	**M2**	**M3**	**M4**	**M5**
MD1_TRG	0.31	0.55	0.75	0.8	0.76	1.99	1.74	1.63	1.47	1.66
MD1_TST	-0.13	0.46	0.62	0.49	0.5	2.52	1.77	1.72	1.44	1.62
MD2_TRG	0.3	0.68	0.74	0.8	0.76	2.23	1.64	1.55	1.36	1.5
MD2_TST	0.15	0.48	0.57	0.72	0.63	1.86	1.78	1.52	1.25	1.75
MD3_TRG	0.24	0.56	0.67	0.75	0.71	2.39	2.06	1.9	1.74	1.6
MD3_TST	0.23	0.76	0.78	0.74	0.72	1.61	1.14	1.33	1.34	1.28

8. ANFIS Model Development

Here also same methodology has been used for model development using MATLAB R2009a software. The models generate Fuzzy Inference System structure from data using subtractive clustering. 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. The performances of the ANFIS models are given in table 5 and their comparative graphs are shown in Fig. 5.

Table 5 ANFIS model performance values

	Correlation Coefficeint					RMSE				
Models	**M1**	**M2**	**M3**	**M4**	**M5**	**M1**	**M2**	**M3**	**M4**	**M5**
MD1_TRG	0.64	0.66	0.87	0.8	0.74	1.57	1.41	1.1	1.01	1.15
MD1_TST	0.18	0.06	0.51	0.6	0.63	2.8	2.98	2.06	2.04	2.1
MD2_TRG	0.61	0.77	0.81	0.8	0.74	1.8	1.41	1.37	1.34	1.41
MD2_TST	0.18	0.62	0.69	0.65	0.66	2.15	1.56	1.54	1.48	1.64
MD3_TRG	0.45	0.77	0.81	0.73	0.68	2	1.38	1.49	1.48	1.47
MD3_TST	0.2	0.45	0.65	0.89	0.41	2.04	2.13	1.23	1.11	1.53

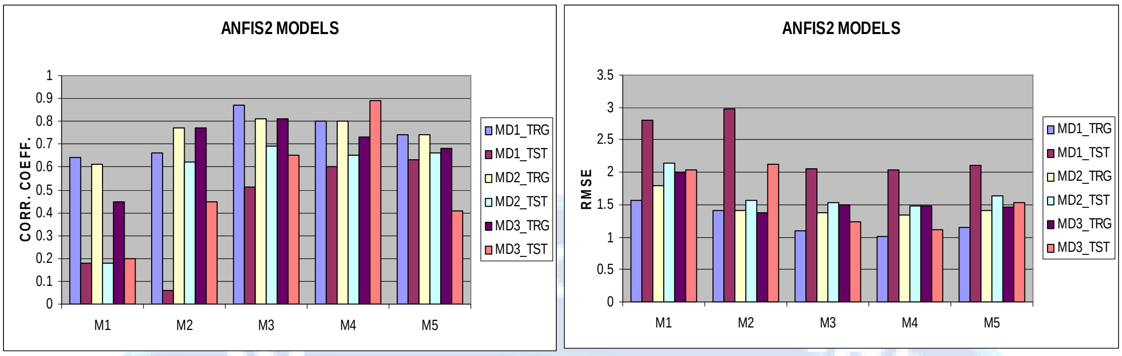

Figure 5 Plot of ANFIS model for RMSE and Corre. Coeff. values

9. Discussion and Conclusions

From the study and evaluation of the models developed using ANN and ANFIS techniques as given in section 7 and 8 above it is seen that both the techniques have been able to develop ground water level prediction models with fairly good accuracy. M4 was found to be the best prediction model having water level, temperature and relative humidity as input variables, closely followed by M3 model with water level, rainfall and relative humidity as the three input variables. The least developed model is the M1 model with all the three input variables as water levels only which shows that taking water level as the only input variable will result in the development of poor prediction model. Hence one has to consider other influencing factors of water level. One thing worth noticing is that 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 estimation in an area. However the second best model has rainfall as one of the input variables. Thus for all practical purposes the second best model M3 can be taken as the best model for ground water level forecasting which closely follows the M4 model in terms of prediction accuracy. The evaluation of M4 model based on the performance criteria are shown in Table 6 and Figures 6 with respect to Correlation coefficient and RMSE respectively. From the comparative study of both the ANN and ANFIS techniques it was seen that ANFIS technique outperformed ANN technique in ground water level forecasting with input variables as (t-1) time lag for water level and (t) time lag both for relative humidity and temperature.

Table 6 Comparative values for ANN and ANFIS M4 model

M4	ANN		ANFIS	
Model	RMSE	Corr. Coeff.	RMSE	Corr. Coeff.
MD1_TRG	1.47	0.8	1.01	0.8
MD1_TST	1.44	0.49	1.04	0.6
MD2_TRG	1.36	0.8	1.34	0.8
MD2_TST	1.25	0.72	1.48	0.65
MD3_TRG	1.74	0.75	1.48	0.73
MD3_TST	1.34	0.74	1.11	0.89

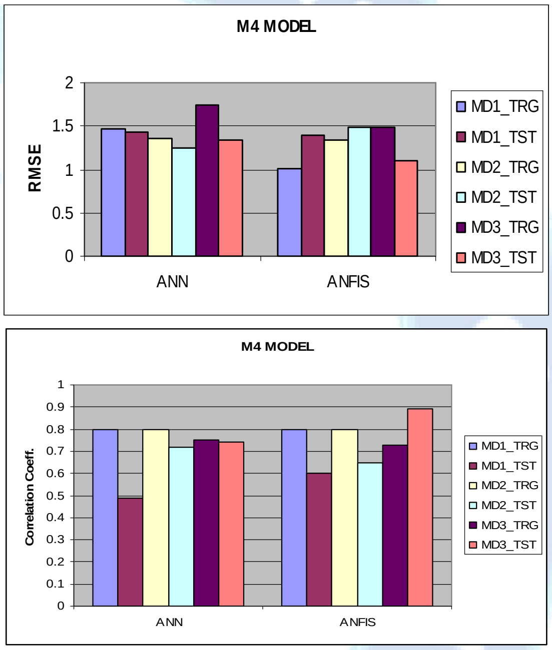

Figure 6 Comparative plot for ANN and ANFIS M4 model using RMSE and Corr. Coeff. values

References:

[1]. Artificial Neural Networks – A Neural Network Tutorial

[2]. Ahmad S., Simonic S.P., "An artificial neural network model for generating hydrograph from hydro-meteorological parameters", Journal of hydrology315(2005) 236 – 251.

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

[4]. Coulibaly, P., Anctil F., et. al., "Artificial Neural Network modelling of water table depth fluctuations"., Water Resources Research, vol.37, No. 4, pages 885-896, April, 2001.

[5]. Coppola E., Poulton M., et.al. "Application of Artificial Neural Networks to complex groundwater Management Problems". Natural Resources Research, vol.12, No. 4, 303 – 320, Dec., 2003. SPRINGER

[6]. Coppola E., Poulton M., et.al. "Artificial Neural network approach for predicting transient water levels in a multilayered groundwater system under variable state pumping and climatic conditions", Journal Hydrologic engineering, Vol.8, Issue 6, pp.348-360, Nov.-Dec. 2003.

[7]. Govindraju, R.S.," Artificial Neural Networks in Hydrology II – Hydrologic applications", J. Hydrologic Engineering, Vol.5, issue 2, pp.

124-137 (April 2000). The ASCE Task Committee report on application of ANN in Hydrology.
[8]. Govindraju, R.S., "Book on Artificial Neural Networks in Hydrology", Water Science and Technology Library, vol. 36
[9]. Jang, J.S. R., Sun, C.T., et al,: Neuro-Fuzzy and Soft Computing, PrinticeHall., ISBN 0-13-261066-3, 607s., United States of America, 1997.
[10]. Jang, J.S. R., Sun, C.T.,: Neuro-Fuzzy Modeling and control, Proceedings of IEEE, 1995.
[11]. Jang, J.S.R., ANFIS:Adaptive- network based fuzzy inference systems, IEEE Trans. On systems, Man and Cybernatics, 23(03):665-685, May, 1993.
[12]. Nayak, P.C., Sudheer, K.P., el al.,: Fuzzy computing based rainfall-runoff model for real time flood forecasting, Hydrol. Process., 17, 3749-3762, 2004a.
[13]. Sen Z: Fuzzy Logic and Foundation, ISBN 9758509233, Bilge Kultur Sanat Publisher, Istanbul, 172 pp., 2001
[14]. "Use of ANN and Fuzzy logic for integrated water management : Review of applications", Project report, Delft 2000, Hydroinformatics.
[15]. Watanabe K., Weesakul U., "Hydrological Monitoring System Based on the ANN: Application to the ground water management".
[16]. Zadeh, L.A.,: Fuzzy Sets, Information and Control, 8(3), 338-353, 1965.
[17]. Zhang, G.P.: Time Series Forecasting using Hybrid ARIMA and neural network model, Neurocomputing, 50, 159-175, 2003.
[18]. Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. Int. J. Forecasting 14, 35–62.
[19]. 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.