International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:4, No:2, 2010

Using Artificial Neural Network to Forecast Groundwater Depth in Union County Well

Zahra Ghadampour, and Gholamreza Rakhshandehroo

***Abstract*—**A concern that researchers usually face in different applications of Artificial Neural Network (ANN) is determination of the size of effective domain in time series. In this paper, trial and error method was used on groundwater depth time series to determine the size of effective domain in the series in an observation well in Union County, New Jersey, U.S. different domains of 20, 40, 60, 80, 100, and 120 preceding day were examined and the 80 days was considered as effective length of the domain. Data sets in different domains were fed to a Feed Forward Back Propagation ANN with one hidden layer and the groundwater depths were forecasted. Root Mean Square Error (RMSE) and the correlation factor (R^2) of estimated and observed groundwater depths for all domains were determined. In general, groundwater depth forecast improved, as evidenced by lower RMSEs and higher R^2s, when the domain length increased from 20 to 120. However, 80 days was selected as the effective domain because the improvement was less than 1% beyond that. Forecasted ground water depths utilizing measured daily data (set #1) and data averaged over the effective domain (set #2) were compared. It was postulated that more accurate nature of measured daily data was the reason for a better forecast with lower RMSE (*0.1027 m* compared to *0.255 m*) in set #1. However, the size of input data in this set was 80 times the size of input data in set #2; a factor that may increase the computational effort unpredictably. It was concluded that 80 daily data may be successfully utilized to lower the size of input data sets considerably, while maintaining the effective information in the data set.

***Keywords*—**Neural networks, groundwater depth, forecast.

I. Introduction

GROUNDWATER is one of the major sources of supply for different purpose such as industrial and agricultural purposes. In some areas groundwater is the only dependable source of supply, while in some other regions it is chosen because of its availability [1]. Groundwater models provide a scientific and predictive tool for determining appropriate solutions of water problem. It may helps the administrators plan groundwater utilization more effectively [1, 2]. To date, a wide variety of models have been developed and applied for groundwater forecasting [3]. These models can be categorized into empirical time series models and physical descriptive models. The major disadvantage of empirical approach is that they are not sufficient for forecasting when the dynamical behavior of the hydrological system changes with time [3]. Similarly, physics based model requires enormous data to simulate water table fluctuations [4]. In a water table aquifer, relationships between affecting parameters on groundwater level are likely nonlinear rather than linear [3, 4]. In recent years, artificial neural networks (ANNs) have been used for forecasting purposes in many areas of science and engineering especially when the relationship of affecting parameters are in a nonlinear form. This makes ANN an attractive tool for modeling water table fluctuations. A few applications of ANN approach in aquifer system modeling have been recently reported in the literature [5, 6, 7, 8].

A concern that researchers usually face in different applications of ANNs is determination of the size of effective domain in time series. The main approach to address this concern has been trial and error [8].

In this paper, artificial neural network was applied to forecast groundwater depth. Trial and error was used on groundwater depth time series to determine the size of effective domain in the series. The effective domain was then utilized in an ANN to optimize its performance in forecasting groundwater depth in an observation well in Union County, New Jersey, U.S. In order to determine the effectiveness of this combination, different domains of the series were fed as inputs to the ANN and the results were compared.

Z. Ghadampour is with the Islamic Azad University, Estahban Branch, Iran (corresponding author, phone: +(98) 9177005440 e-mail: zahra.ghadampour@gmail.com).

G. Rakhshandehroo is with Civil Engineering Department, Shiraz University, Iran.

II. Artificial Neural Networks

Neural Networks have gone through two major development periods; the early 60's, and the mid 80's. They were a key development in the field of machine learning. Artificial Neural Networks were inspired by biological findings relating to the behavior of the brain as a network of units called neurons [9].

The fundamental building block in an Artificial Neural Network is the mathematical model of a neuron as shown in Fig. 1. The three basic components of an artificial neuron are:

1. The synapses or connecting links that provide weights, w_j, to the input values, x_j for j = 1,..., m;

The weights in neural nets are also often designed to minimize the error in a training data set

2. An adder that sums the weighted input values to compute the input to the activation function:

$$v = w_0 + \sum_{j=1}^{m} x_j w_j ,$$

where w_0 is called the bias is a numerical value associated with the neuron

3. An activation function g that maps v to g(v) the output value of the neuron. This function is a monotone function.

A. Network Architecture

While there are numerous different ANN architectures that have been studied by researchers, the most successful applications have been multilayer feedforward networks. Fig. 2 is a diagram for this architecture [9].

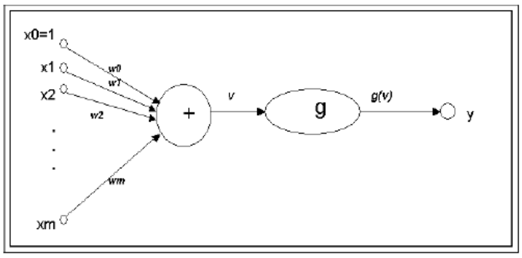

Fig. 1 Schematics for a mathematical model of a neuron

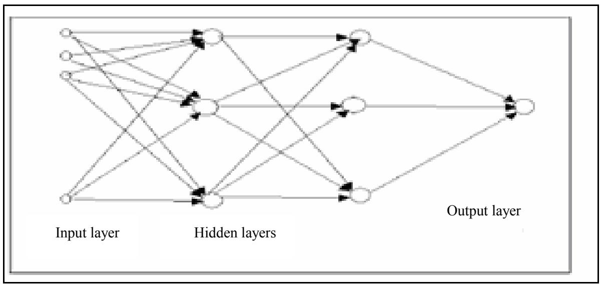

Fig. 2 Diagram of a Feed Forward Back propagation

B. Feed Forward Back Propagation (FFBP) Neural Network

Present study employed a standard back propagation algorithm for training, and the number of hidden neurons is optimized by a trial and error procedure. In these networks, there is an input layer consisting of nodes that simply accept the input values and successive layers of nodes that are neurons as depicted in Fig. 1. The outputs of neurons in a layer are inputs to neurons in the next layer. The last layer is called the output layer. Layers between the input and output layers are known as hidden layers [9].

C. Evaluation Criteria

Root Mean Square Error (RMSE) criterion is used by researchers in order to evaluate the effectiveness of each network in its ability to make precise predictions [9]. It is calculated by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

Where y_i is the observed data, $\hat{y}_i$ is the calculated data, and N is the number of observations. Qualitatively speaking, RMSE reflects the discrepancy between the observed and calculated values. The lower the RMSE, the more accurate the prediction.

III. Study Area

Union County Well is located in New Jersey, US in the hydrologic unit 02030104with 40˚41´06´´north latitude and 74˚ 14´19 ´east longitude. The Well depth is equal to 290 feet and ground surface elevation is 69.00 feet above sea depth. The well is completed in "Early Mesozoic basin aquifers" (N300ERLMZC) national aquifer (Fig. 3). The daily data are recorded for a period of 65 years (from 1943 to 2008) except for an 8 year gap(1975 to 1983) [10]. In this paper ANN (FeedForward Backpropagation) was used to forecast ground water depths in Union County Well.

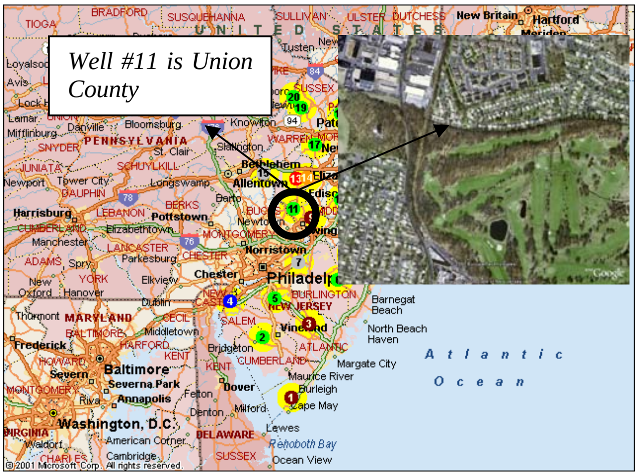

Fig. 3 Location of Union County well in New Jersey, US

IV. Model Structure

A. Selection of Input Vector

As a first step in design of network architecture, the input parameters which affect the outputs should be determined [11]. Generally, experience knowledge is used to specify the initial set of candidate inputs [12, 13].

In this paper, daily depth to ground water was used as the input. Two sets of data, measured daily data (set #1) and data averaged over the effective domain (set #2) were used to forecast ground water depths.

B. Hidden Neurons Optimization

A number of empirical relationships between the number of training samples and the number of connection weights have been suggested in the literature [9]. Nevertheless, the network geometry is highly dependent on the problem and properties of available data. The optimum ANN architecture which can effectively capture the relationship between the input and output data is usually determined by trial and error. The trial and error procedure started with one hidden layer initially, and the number of hidden layers was increased up to 3 with a step size of 1 in each trial. The training was stopped when there was no significant improvement in the efficiency, and the model was then tested for its general properties.

C. Internal Parameters of the Model

A sigmoid function was used as the activation function in both hidden and output layers. As the sigmoid transfer function has been used in the model, the input-output data have been scaled appropriately to fall within the function limits. A standard back propagation algorithm has been employed to estimate the network parameters [9]. The learning rate was held constant throughout training in the standard steepest descent (back propagation) process.

V. Results

Monitored daily ground water depths in Union County Well from Mar 1985 to Mar 2007 are shown on Fig. 4 as a time series. The fluctuating nature of the data reflects a fractal character which has an average, minimum, maximum, and standard deviation of 6.79 *m*, 3.45 *m*, 10.75 *m*, and 0.956 *m*, respectively. Trial and error method was applied to the data set considering different domains of 20, 40, 60, 80, 100, and 120 preceding.

Data sets in different domains were fed to ANN and groundwater depth was forecasted. Figs. 5, 6, and 7 depict estimated versus observed groundwater depths for a 22 year period considering the domains of 20, 80, and 120 preceding days, respectively. Root Mean Square Error (RMSE) and the correlation factor (R^2) of estimated and observed groundwater depths for all domains are also presented in Table I. In general, when the domain length increased from 20 to 120, groundwater depth forecast was improved as evidenced by lower RMSEs (*0.563509* compared to *0.098972 m*) and higher R^2s (*0.69023* compared to *0.99273*). How ever, the improvement was less than 1% beyond the domain length of 80 days.

Considering Table I, the data in the 80-day domain was selected as the effective data which required a reasonable computational effort and yielded an acceptable R^2 (*0.99218*) and RMSE (*0.102719 m.* The calculated effective domain (80 days) confirms well with the effective domain of a few months reported by other researchers on groundwater depth time series.

Statistical parameters of training and testing the network with the two data sets considering different numbers of hidden layers are shown in Table II. As shown, errors are bound to acceptable values (less than *0.037*) and typically smaller in lower numbers of hidden layers for both data sets. No over flowing was observed in either sets of data and it was concluded that data with one hidden layer was sufficient for accuracy of the network.

Comparison of the estimated data (based on data averaged over the effective domain; set #2) and observed data is shown in Fig. 6. Similar comparison for the estimated data (based on daily data in the effective domain; set #1) and the observed data was shown in Fig. 8. Comparing Figs. 6 and 8, it was postulated that more accurate nature of data set #1 (shown in Fig. 6) was the reason for a better prediction with lower RMSE (*0.1027 m* compared to *0.255 m*). However, the size of input data in set #1 was 80 times the size of input data in set #2; a factor that may increase the computational effort unpredictably. Hence, it was concluded that data set #2 may be successfully utilized to lower the size of input data sets considerably, while yielding acceptable RMSE (*0.255 m* in our case) as well.

VI. Conclusion

ANN was utilized to forecast ground water depths in two different sets of data. Trial and error method was applied to the sets considering different domains of 20, 40, 60, 80, 100, and 120 preceding days and 80 days was selected as an effective domain considering lower RMSE and higher R^2 of predicted time series in ANN. This method decreased the computational effort and, at the same time, yielded acceptable R^2 and RMSE. The number of hidden layers was optimized by trial and error to one hidden layer for both data sets. It was postulated that more accurate nature of measured daily data was the reason for a better forecast with lower RMSE (*0.1027 m* compared to *0.255 m*) in set #1. However, the size of input data in this set was 80 times the size of input data in set #2; a factor that may increase the computational effort unpredictably.

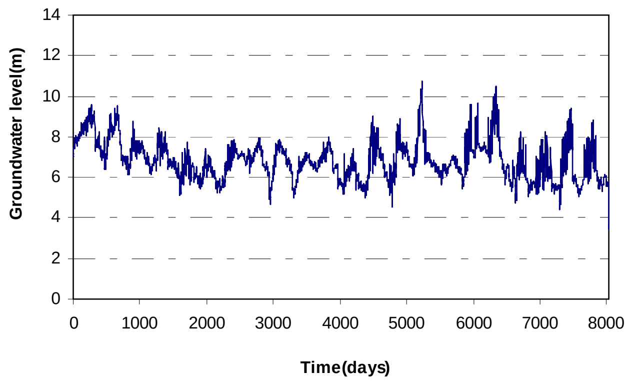

Fig. 4 Observed ground water depth in Union County Well

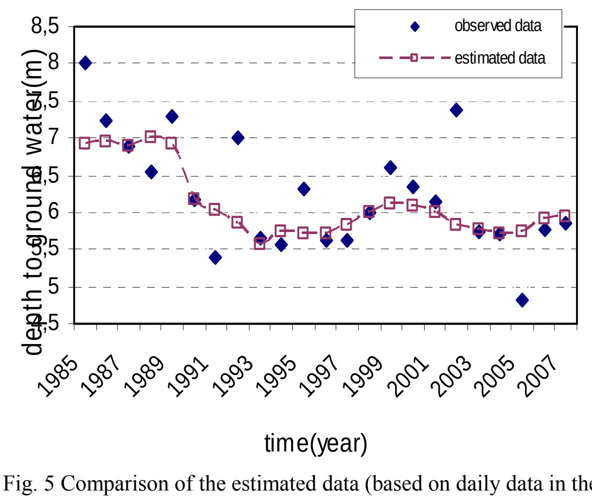

Fig. 5 Comparison of the estimated data (based on daily data in the 20-day domain) and observed data

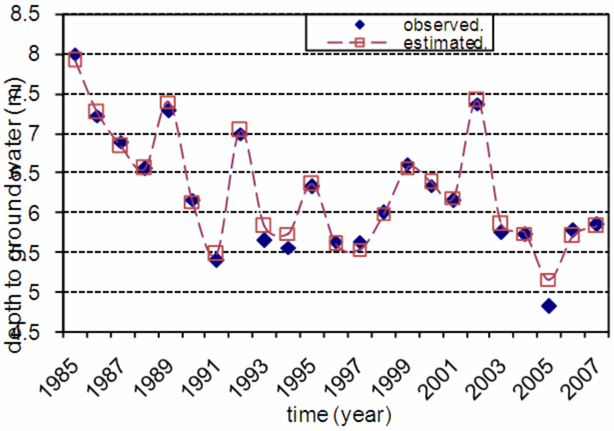

Fig. 6 Comparison of the estimated data (based on daily data in the effective domain; set #1) and observed data

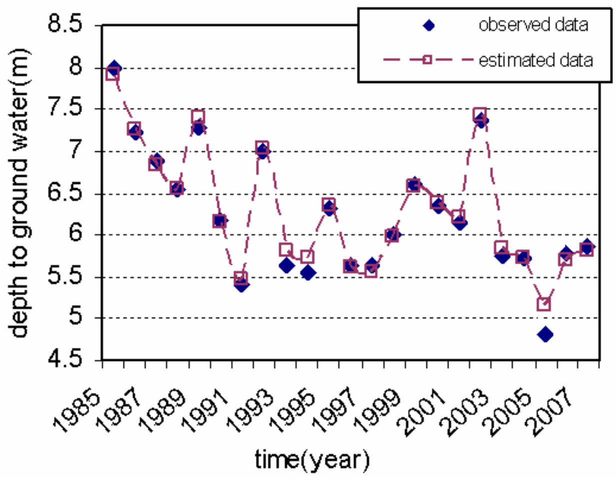

Fig. 7 Comparison of the estimated data (based on daily data in the 120-day domain) and observed data

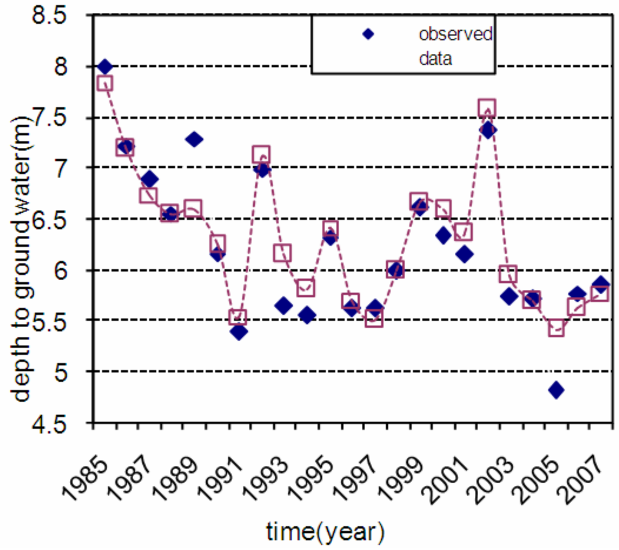

Fig. 8 Comparison of the estimated data (based on data averaged over the effective domain; set #2) and observed data

TABLE I
ROOT MEAN SQUARE ERROR (RMSE) AND CORRELATION FACTOR (R^2) FOR DIFFERENT DOMAIN LENGTHS

Domain Length (days)	R^2	*RMSE (m)*
20	0.69023	0.563509
40	0.85301	0.284730
60	0.89763	0.163450
80	0.99218	0.102719
100	0.99225	0.099437
120	0.99273	0.098972

TABLE II
STATISTICAL PARAMETERS FOR TRAINING AND TESTING DATA SETS

Data Sets	No. of Hidden Layers	Error in Training Data			Error in Testing Data		
		Ave (m)	Max (m)	Pd (%)	Ave (m)	Max (m)	Pd (%)
Set #1	1	0.026	0.13	91.4	0.035	0.16	88.46
	2	0.03	0.14	89.14	0.037	0.15	87.5
	3	0.09	0.33	43.67	0.11	0.32	35.58
Set #2	1	0.037	0.16	85.52	0.046	0.16	79.81
	2	0.037	0.15	85.7	0.046	0.15	77.88
	3	0.095	0.33	42.08	0.11	0.34	35.58

Pd: percent of data with less than *0.05 m* error

REFERENCES

[1] P. D. Sreekanth, N. Geethanjali, P. D. Sreedevi, Shakeel Ahmed, N. Ravi Kumar and P. D. Kamala Jayanthi, (2009), “Forecasting groundwater level using artificial neural networks”, CURRENT SCIENCE 96, 933-939

[2] Van Geer, F. C. and Zuur, A. F., (1997), ‘An extension of Box-Jenkins transfer noise models for spatial interpolation of groundwater head series’, Journal of Hydrology 192, 65–80.

[3] Bierkens, M. F. P., (1998), “Modeling water table fluctuations by means of a stochastic differential equation”, Water Resources Research 34, 2485–2499.

[4] Knotters, M. and Bierkens, M. F. P., (2000), “Physical basis of time series models forwater table depths”, Water Resources Research 36(1), 181–188.

[5] Gail, M., Brion, T. R., Neelakantan and Lingireddy, S., (2002), “A neuralnetwork- based classification scheme for sorting sources and ages of fecal contamination in water”, Water Res., 36, 3765–3774.

[6] Guan, P., Huang, D. and Zhou, B., (2004), “Forecasting model for the incidence of hepatitis A based on artificial neural network”, World J. Gastroenterol., 10, 3579–3582.

[7] Morshed, J. and Kaluarachchi, J. J., (1998), “Parameter estimation using artificial neural network and genetic algorithm for free-product migration and recovery”, Water Resources Research 34(5), 1101–1113

[8] Coulibaly, P., Anctil, F., Aravena, R., and Bobee, B., (2001), “Artificial neural network modeling of water table depth fluctuations”, Water Resources Research 37(4), 885–896.

[9] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., (1986), Learning representations by back propagating errors. Nature 323, 533–536.

[10] http://www.usgs.org/

[11] Maier, H. R. and Dandy, G. C., (2000), “Neural networks for the prediction and forecasting of water resources variables: A review of modeling issues and application”, Environmental Modeling and Software, 15: 101–124.

[12] Campolo, M., Andreussi, P., and Soldati, A., (1999), “River flood forecasting with neural network model”, Water Resources Research 35(4), 1191–1197.

[13] Thirumalaiah, K., and Deo, M. C., (2000), “Hydrological forecasting using neural networks”, Journal of Hydrologic Engineering 5(2), 180–189.