Scientific Research and Essays Vol. 7(50), pp. 4242 -4252, 24 December, 2012
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE12.346
ISSN 1992-2248 ©2012 Academic Journals

Full Length Research Paper

A prediction model for the level of well water

Tuğba İnan[1*] and Gökmen Tayfur[2]

[1]Department of Architecture, Faculty of Architecture, Izmir Institute of Technology, Gülbahçe Köyü, Urla, TR-35430, Izmir, Turkey.
[2]Civil Engineering Department, Faculty of Engineering, Izmir Institute of Technology, 35430, İzmir, Turkey.

Accepted 30 November, 2012

There is a nonlinear relationship between rainfall and well water levels. This is one of the most complicated hydrologic phenomena to understand due to the existence of spatial and temporal incoherent geomorphic and climatic factors. The aim of this study is to predict the well water level by the artificial neural networks. A well is located on the campus area of Izmir Institute of Technology, Izmir, Turkey. While precipitation, outside temperature, and evaporation formed the input vector, the water levels were the target outputs. Precipitation and evaporation data were also recorded on the same campus area. The feed forward back propagation neural network is employed using the package program, called NeuroSolutions for Excel, due to its success in learning process, creating graphics for the results and sensitivity analysis. The findings of this study show that the model can successfully predict water levels in the well, with mean absolute error (MAE) of 37 cm and correlation coefficient (R) of 0.91 in the training stage and MAE = 0.40 cm and R = 0.80 in the testing stage. The sensitivity analysis results revealed that the outside temperature is the most effective parameter and the evaporation was least.

Key words: Artificial neural network, well water, prediction, precipitation, evaporation, temperature.

INTRODUCTION

Water availability is a global problem. It is an increasing interest in all areas of the world. Global consumption of water is doubling every 20 years, more than twice the rate of human population growth (Affandi and Watanabe, 2007) and groundwater is no exception. Groundwater levels fluctuate on a seasonal basis, rising in rainy months and falling in dry months. It takes time before you see a drop in water levels and time for the water to be replenished. Water load may depend on climatic factors, such as rainfall, watershed and climatical characteristics (Mencar et al., 2008).

Prediction of groundwater level is difficult, due to the non-linear influence of several factors, sometimes unknown, such as rain in areas close or distant to the wells, ground temperature and humidity, presence of unknown wells, etc. All this factors make the adoption of standard statistical techniques useless (Daliakopoulos et al., 2005). A useful approach for water level prediction involves the use of data-driven models, such as atificial neural networks (ANN). The main advantage of ANN is their ability of predicting water levels without the need of an explicit representation of the geological substrates surrounding the wells. Actually, such a representation is implicitly encoded in the ANN structure and in connection weights (Mencar et al., 2008).

The main disadvantage of ANN is the need of a large amount of historical data that is required to train the network in order to capture the trends between inputs (climatic data and other information) and output (water level of a well). Furthermore, the design of ANN requires several decisions to be taken by trial-and-error. This requires intensive study before arriving at a preferable model. ANN suggests a suitable way for designing predictive models of the water levels in wells (Mencar et al., 2008). There are numerous applications of ANNs in the area of water resources engineering (Chau et al.,

*Corresponding author. E-mail: tugbainan85@gmail.com. Tel: +90-232-750-7053. Fax: +90-232-750-7012.

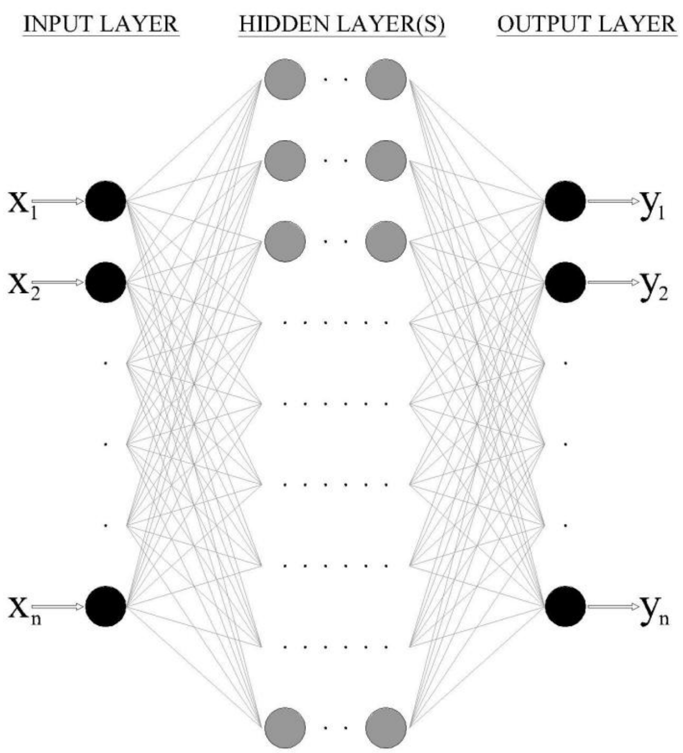


Figure 1. General structure of a neural network.

2005; Muttil and Chau, 2006; Wang et al., 2009; Wu et al., 2010; Tayfur, 2012, among many).

ANN was chosen in this study for its ability to generalize results from unseen data and it is well-suited in modeling dynamic systems on a real-time basis. Hence, they can forecast water levels and missing precipitation even the physical relationships are not well understood (Bustami et al., 2007). ANN can be used as tools for domain experts to understand the main climatic factors that mostly influence water levels. In this study, it is aimed to predict the water levels in wells and find the most sensitive parameters.

ARTIFICIAL NEURAL NETWORKS

ANNs are defined as massively parallel-distributed information-processing systems that are look-alike biological neural networks of the human brain and can create connections between mathematical processing elements (American Society of Civil Engineers (ASCE), 2000; Tayfur, 2012). The idea of artificial neural networks was firstly introduced over seventy years ago by McCulloch and Pitts (1943). However, the large-scale progress started only in 1982, when Hopfield found the iterative procedures for neural networks (Hopfield, 1982).

ANN consists of information processing units. Information processing occurs at neurons. Signals are transmitted between neurons with connection weights. Each connection link is represented by its connection strength. As Lam et al. (2008) stated, ANNs learn the relationship between the input and output variables. It has the ability to learn from experience and examples and then to adapt to changing situations rapidly in each case. ANNs resembles computer programs which simulate the biological structure of the human brain. It resembles the human brain in two aspects; knowledge is obtained by the learning process and neuron connection strengths are used to store the knowledge (Ayed, 1997; Tayfur, 2012).

The general structure of the neural network is given in Figure 1. It consists of neurons, layers and connection weights. The first layer is called input layer and it may have many input neurons such as x_1, x_2, x_3,…., $x_{n.}$ Each input neuron defines an input data. The second layer is called hidden layer. There may be one or more hidden layer (s) consisting many neurons. The last layer is called the output layer. The neural network may have one or more output depending on the prediction problem. It consists of predicted values. There may be unaccounted parameters affecting the process. In order to avoid an indefinite effect of those parameters, bias neurons are used in the input and hidden layers. They are not used in the output layer. Moreover, it takes generally -1 or + 1 values as input values. The usage of a bias neuron is not compulsory in a neural network (Tayfur, 2012).

Tayfur (2012) defines the process of ANN as follows; a neuron receives inputs over its incoming connections and then combines the inputs. At the end of this process the network outputs the final results. At the training stage of the network both the inputs and outputs are presented to the network for thousands of cycles. Inputs represent the parameters of problem and outputs represent the solution options. The network evaluates the error between the actual and desired output at the end of each cycle. After this work it uses this error to update values of the connection weights according to the chosen training algorithm (Ayed, 1997; Tayfur, 2012).

The ANN models can learn and generalize the problems even when the input data include errors or are incomplete. The prediction capability of the network is tested by the data that are selected from the whole data set (Hoo et al., 2002). Training and testing of the network is continued until no improvement in the output is achieved. This process is performed after a predetermined number of iterations (Lam et al., 2008).

Modeling software (NeuroSolutions)

The software used in this study is called NeuroSolutions which works in Microsoft Excel. The data are firstly arranged in Microsoft Excel. Software adds a tool to Excel tool bar called “NeuroSolutions” and this tool is used for any action in designing the model with the data.

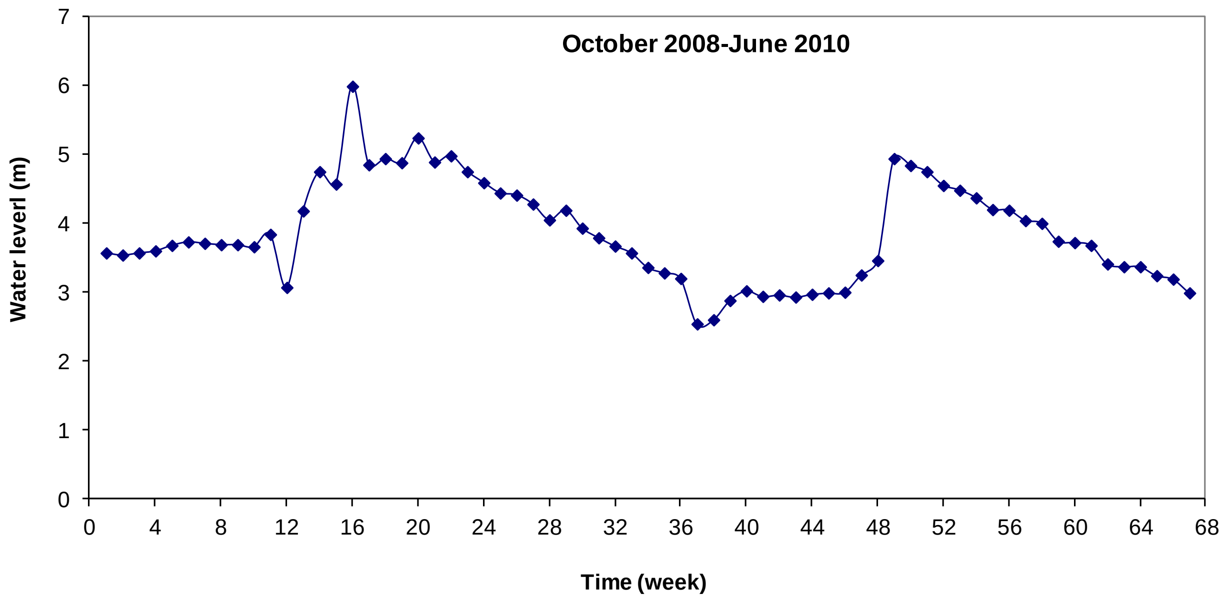


Figure 2. Change in the measured well water levels from 17/10/2008 to 06/08/2010.

NeuroSolutions should be opened together with Excel in order to model the network.

The usage of the NeuroSolutions is so simple. The software has watching demos about modeling the network. One should first watch these videos in order to understand the software. It provides imitating the steps of the demo with your own data. Many practices are needed to be done in order to find better results with different parameters. The success of the network can be seen clearly by generating many different networks.

NeuroSolutions software is based on two columns called input column and output column. The software compares the results obtained from these columns. Before NeuroSolutions for Excel can train a neural network, it first needs to know which columns to use as inputs and which columns to use as the desired outputs. Then, the training, cross validation and testing data sets should be designed. NeuroSolutions does this action by separating the data sets with default percentages; however these percentages can also be organized in the way that the user wants. It presents the results of each function of the data and draws graphics automatically reflecting the values of the functions.

DATA AND METHODOLOGY

The well is located near the campus area of Izmir Institute of High Technology, Izmir, Turkey. The level of the water is measured weekly (or bi-weekly) from 17/10/2008 to 06/08/2010 (Figure 2). Sometimes, the measurements were carried out right after the rainfall, resulting in periods shorter than a week. For the same period, the daily collected climatic data have been obtained from the Weather Station in the Department of Mechanical Engineering in Izmir Institute of Technology. The climatic data were measured every 20 min, yet it was possible to obtain hourly, daily, and weekly average values. The values of 67 data sets which are used in this study are presented in Table 1. The rainfall, evaporation, and temperature data are presented in Figures 3, 4 and 5, respectively.

Design

The model comprised three input neurons for the input vector of rainfall, evaporation and temperature and one output neuron for well water level. The number of hidden layers and the neurons in each hidden layer were decided by the trial-and-error procedure. As a result, the model includes two hidden layers with 6 processing elements (neurons). The neural network software, NeuroSolutions, used in this study requires some parameters to start the simulation:

1. The percentage of training data.
2. The percentage of testing data.
3. The type of training algorithm.
4. The number of hidden layers.
5. The learning rule and the momentum constant.
6. The number of iterations (epochs).

The data of the model are categorized in two groups as training data set and testing data set. In this study, we did not use cross validation data sets for two reasons:

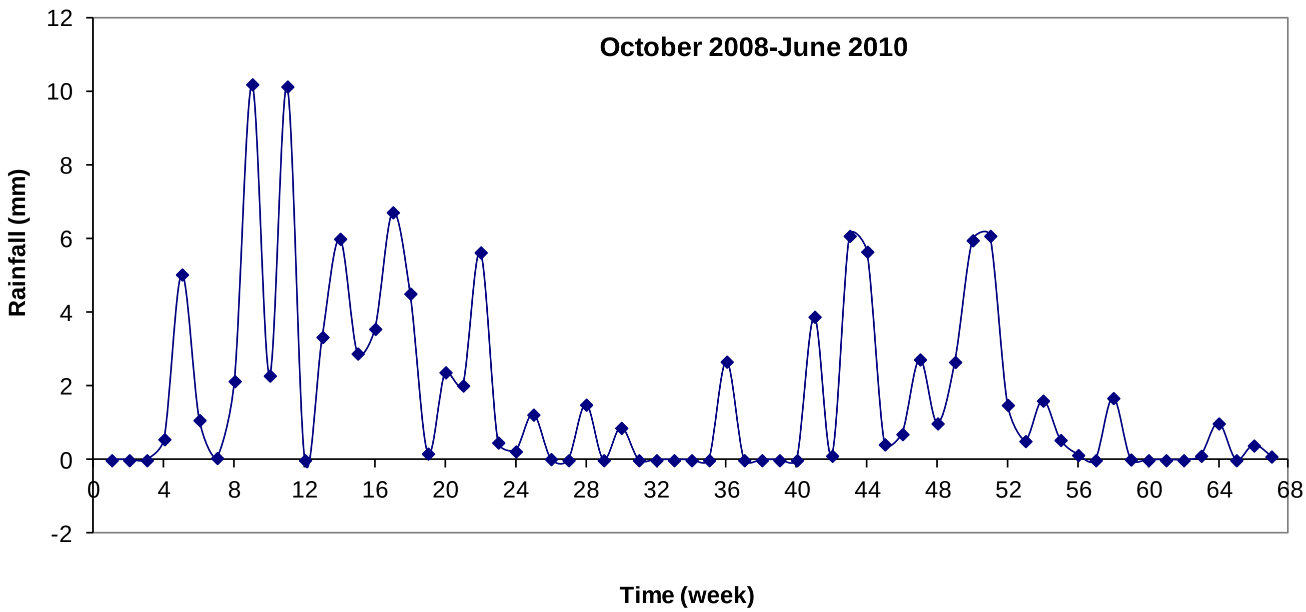


Figure 3. Change in the rainfall amount from 17/10/2008 to 06/08/2010.

Table 1. Measured data.

# of data	Date	Outside temperature (°C)	Rainfall amount (mm)	Evaporation amount (mm)	Well water level (m)
1	17.10.2008	17.37	0.00	3.57	3.58
2	24.10.2008	18.02	0.00	1.74	3.55
3	31.10.2008	16.26	0.00	2.00	3.58
4	07.11.2008	12.61	0.57	1.25	3.61
5	17.11.2008	13.95	5.05	1.01	3.69
6	21.11.2008	14.71	1.09	1.32	3.74
7	28.11.2008	17.61	0.06	1.48	3.72
8	05.12.2008	10.16	2.15	0.91	3.70
9	15.12.2008	11.47	10.22	2.77	3.70
10	18.12.2008	5.26	2.30	3.67	3.67
11	19.12.2008	7.19	10.16	5.20	3.85
12	26.12.2008	6.58	0.00	2.83	3.08
13	02.01.2009	9.35	3.35	1.02	4.19
14	09.01.2009	13.77	6.02	1.27	4.76
15	16.01.2009	11.09	2.90	1.30	4.58
16	23.02.2009	14.20	3.57	1.47	6.00
17	02.03.2009	10.66	6.74	1.36	4.86
18	09.03.2009	6.59	4.53	1.20	4.95
19	16.03.2009	6.38	0.18	7.34	4.89
20	23.03.2009	13.98	2.39	2.32	5.25
21	31.03.2009	10.27	2.03	1.87	4.90
22	06.04.2009	8.22	5.65	1.54	4.99
23	13.04.2009	10.12	0.48	2.08	4.76
24	20.04.2009	15.05	0.24	1.98	4.60
25	28.04.2009	14.46	1.24	0.97	4.45
26	05.05.2009	15.50	0.03	2.26	4.42
27	11.05.2009	13.85	0.00	2.71	4.29
28	16.05.2009	16.47	1.51	3.04	4.06

Table 1. Contd.

29	20.05.2009	20.45	0.00	4.07	4.20
30	02.06.2009	22.02	0.88	3.48	3.94
31	09.06.2009	22.45	0.00	4.75	3.80
32	16.06.2009	22.96	0.00	4.01	3.68
33	23.06.2009	26.03	0.00	4.74	3.58
34	14.07.2009	25.94	0.00	6.36	3.37
35	22.07.2009	24.49	0.00	4.63	3.29
36	28.07.2009	24.06	2.68	2.95	3.21
37	27.08.2009	26.90	0.00	5.42	2.55
38	03.09.2009	26.54	0.00	5.48	2.61
39	10.09.2009	24.97	0.00	4.55	2.89
40	17.09.2009	24.69	0.00	3.83	3.03
41	25.09.2009	21.83	3.90	2.44	2.95
42	01.10.2009	22.82	0.12	3.66	2.97
43	14.10.2009	20.45	6.10	3.20	2.94
44	20.10.2009	20.59	5.67	2.01	2.98
45	27.10.2009	20.62	0.43	1.85	3.00
46	03.11.2009	19.93	0.71	1.39	3.01
47	10.11.2009	19.69	2.74	2.09	3.26
48	17.11.2009	16.81	1.00	3.94	3.47
49	19.02.2010	15.05	2.67	3.79	4.95
50	05.03.2010	14.98	5.98	1.44	4.85
51	12.03.2010	13.99	6.10	1.51	4.76
52	19.03.2010	13.18	1.50	1.19	4.56
53	01.04.2010	13.84	0.52	0.95	4.49
54	08.04.2010	12.08	1.62	1.37	4.38
55	15.04.2010	15.49	0.55	1.09	4.21
56	22.04.2010	12.67	0.14	1.56	4.20
57	30.04.2010	15.14	0.00	2.80	4.05
58	07.05.2010	9.68	1.69	0.75	4.01
59	13.05.2010	13.58	0.02	2.66	3.75
60	27.05.2010	16.17	0.00	2.95	3.73
61	03.06.2010	13.29	0.00	3.80	3.69
62	17.06.2010	16.00	0.00	3.66	3.42
63	24.06.2010	20.23	0.12	3.90	3.38
64	02.07.2010	22.69	1.00	2.36	3.38
65	16.07.2010	17.10	0.00	3.98	3.25
66	23.07.2010	22.69	0.40	3.32	3.20
67	06.08.2010	21.82	0.10	3.98	3.00

1) Since the model performed satisfactorily in the training and testing data sets, there was no over-training problem and thus no need for the cross-validation;
2) The data sets that we have is already a small size. When one is provided with more data, in the order of hundreds, then of course it would be beneficial to carry out such study. 75% of data are used for set and the rest for testing (Figure 6).

Then, NeuroBuilder is used for setting up the network. There were various neural network types. In this study, GFF is chosen for designing a neural network due to its better performance (Figure 7).

The next window is for designing the parameters of hidden layers (Figure 8). The system is designed as to have two hidden layers having same features. As shown in Figure 8 tangent axon is chosen as the activation function. The learning rate parameter and the momentum factor are kept constant throughout the training of network. The values of learning rate parameter and momentum factor are taken as 0.10 and 0.70,

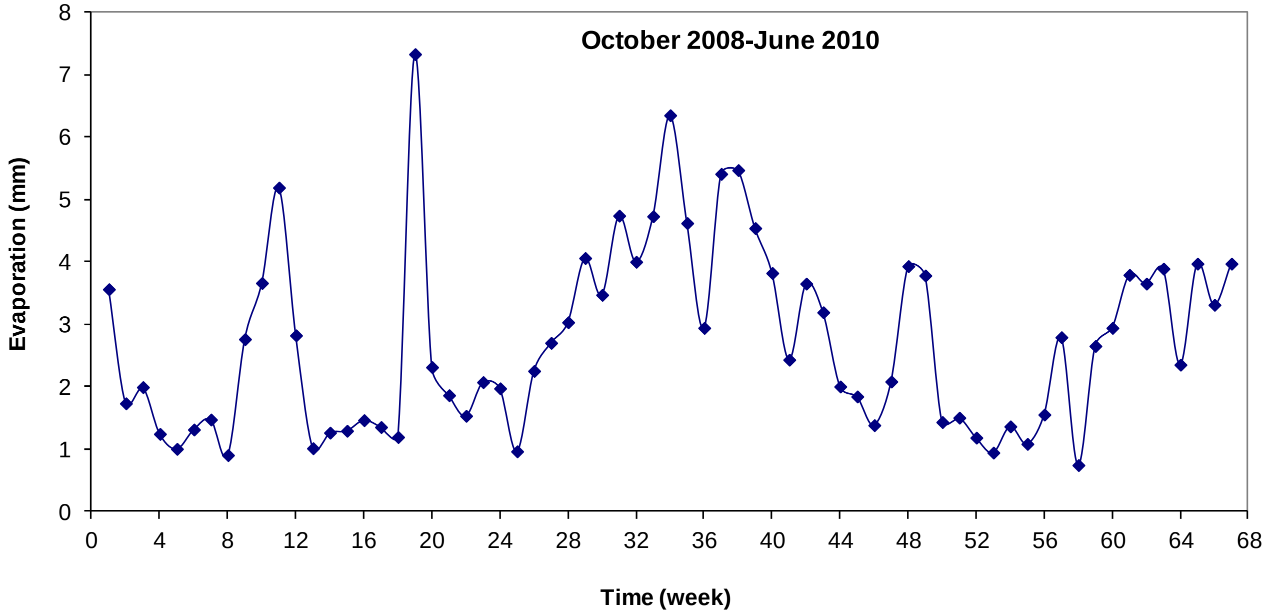


Figure 4. Change in the evaporation amount from 17/10/2008 to 06/08/2010.

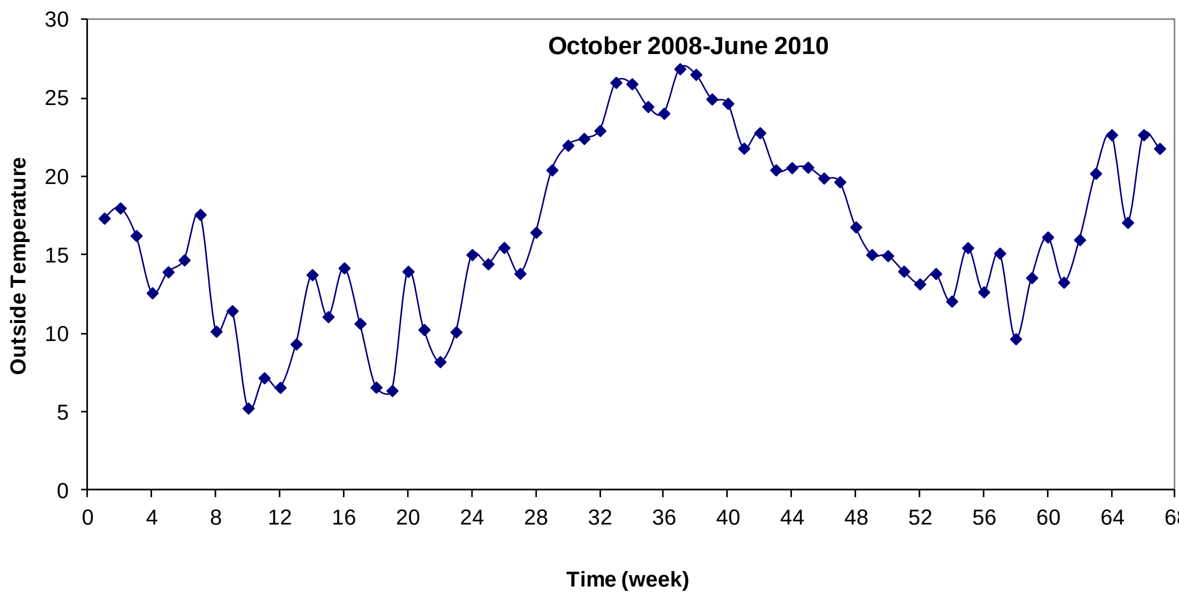


Figure 5. Change in the temperature values from 17/10/2008 to 06/08/2010.

respectively. After designing the hidden layers, output layer is opened in the next step and the same parameter values are used.

This step is for designing the parameters of maximum epochs that is generated during training part. The software chooses training data set for this action to represent the values of mean square error (MSE) and 1000 epochs (Figure 9).

In the last window box of the NeuroBuilder part, configuration of the parameters can be chosen. In this study, default parameters are used for the configuration.

RESULTS AND DISCUSSION

Figure 10 shows the trend of the error function as the number of iterations increase. As seen, the error sharply decreases in 200 iterations and continues to decrease and levels off around (0.0169 m = 1.69 cm) in 1000 iterations. This is a desired trend suggested in the

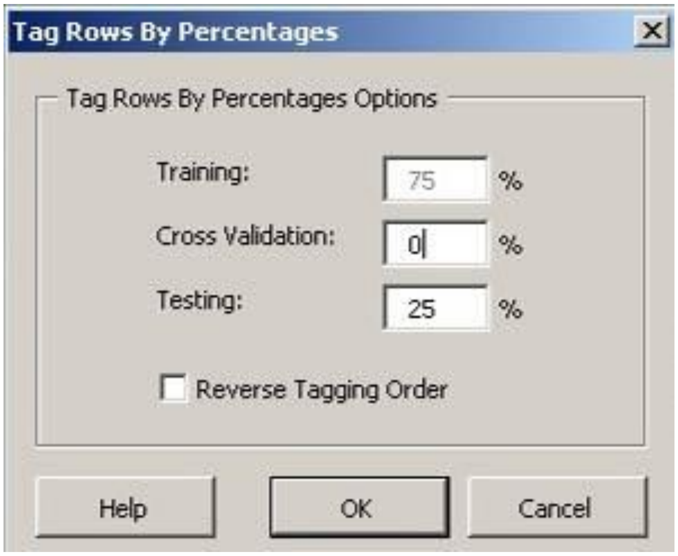


Figure 6. Tagging rows by percentages.

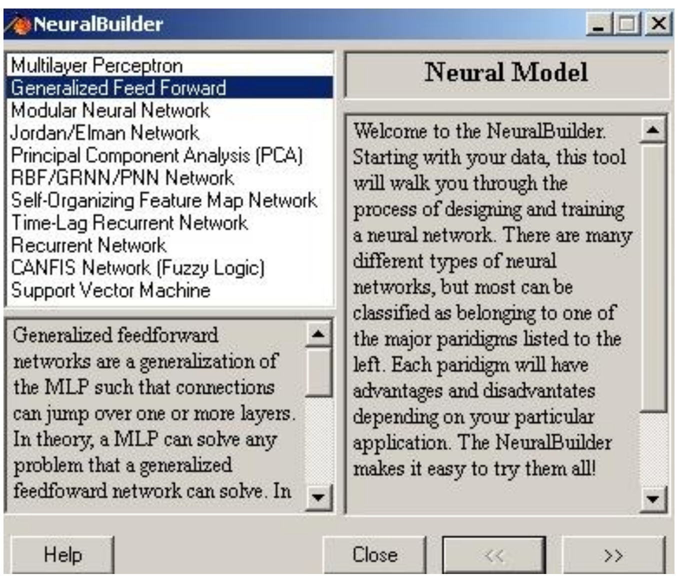


Figure 7. Designing a neural network.

literature (Tayfur, 2012).

Figure 11 shows predicted water levels by the trained network, using the data set used in the training as a testing set in order to see whether the network is trained successfully or not. As seen, the model perfectly captures the trend, including minimum and maximum values. The mean absolute error (MAE) is 0.37 m, the mean square error (MSE) is 0.22 m and the correlation coefficient (R) is 0.91 (Figure 11). All these results imply the successful training of the network.

Figure 12 shows the model testing. As seen, the overall model predicts water levels, with MAE of 0.40 m, MSE of

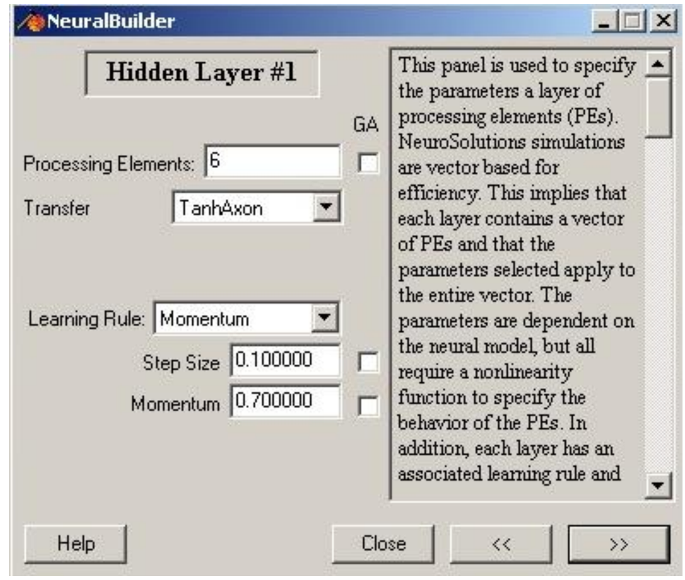


Figure 8. Parameters of hidden layer.

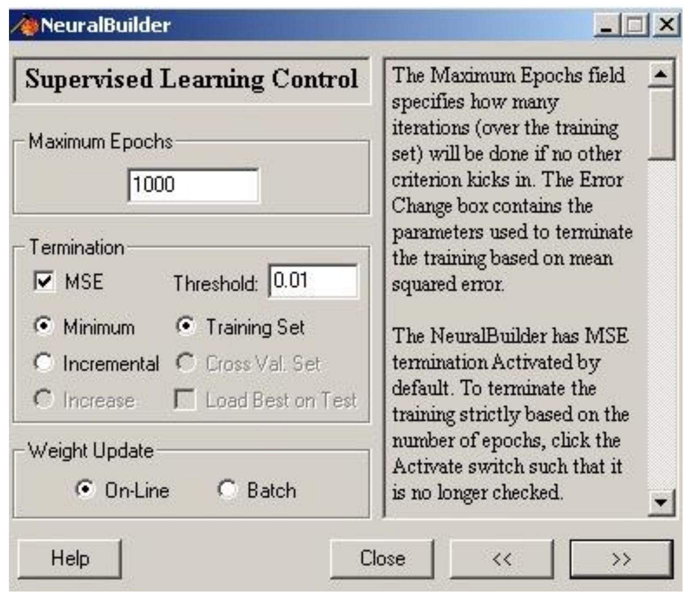
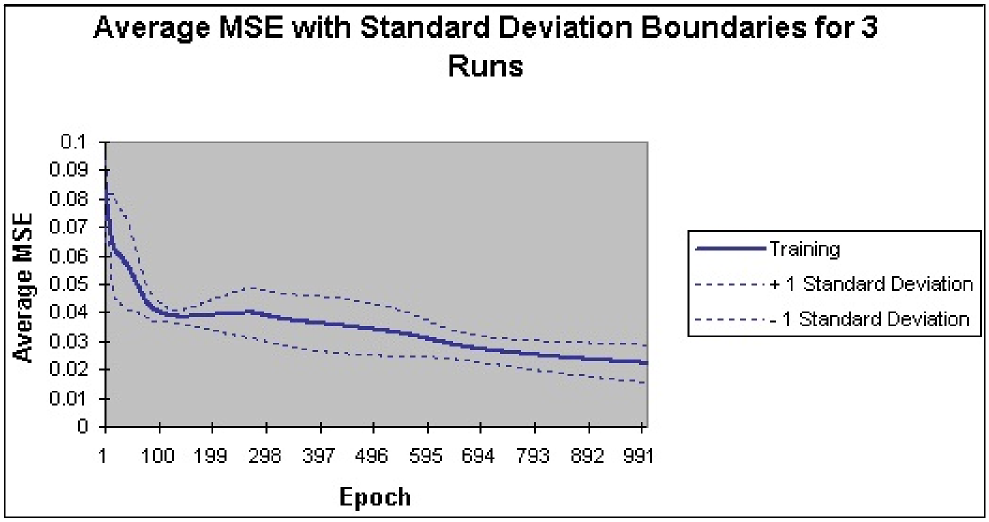


Figure 9. Parameters of supervised learning.

0.37 m and R of 0.80. These can be assumed as satisfactory results. The scatter diagram showing measured versus predicted values is presented in Figure 13. As seen, the values are evenly distributed around the regression line, implying that there was neither overwhelming over prediction not under prediction.

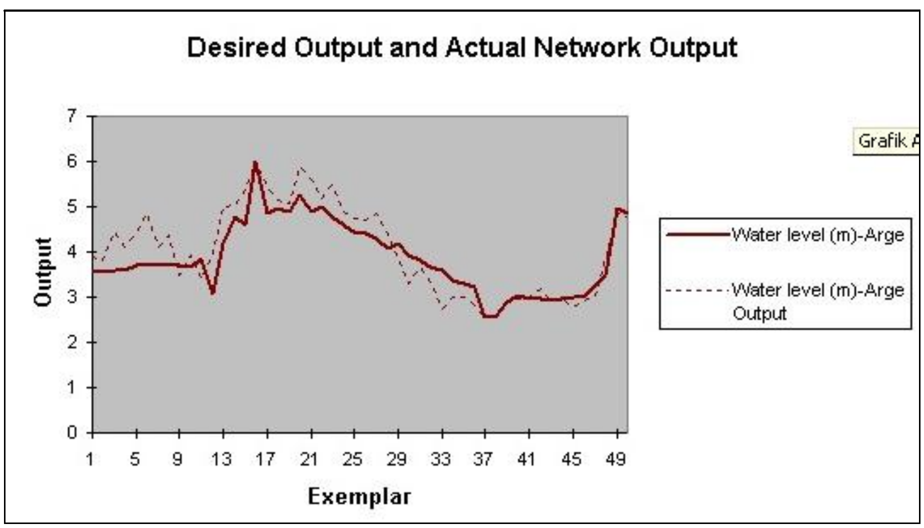
Figure 14 presents the sensitivity analysis result. As seen, the outside temperature turns out to be the most sensitive parameter. Evaporation is the least sensitive. However, we have to comment here that rainfall is the main source of the water in groundwater and therefore,



All Runs	*Training Minimum*	*Training Standard Deviation*
Average of Minimum MSEs	0.022253089	0.006587322
Average of Final MSEs	0.022253089	0.006587322

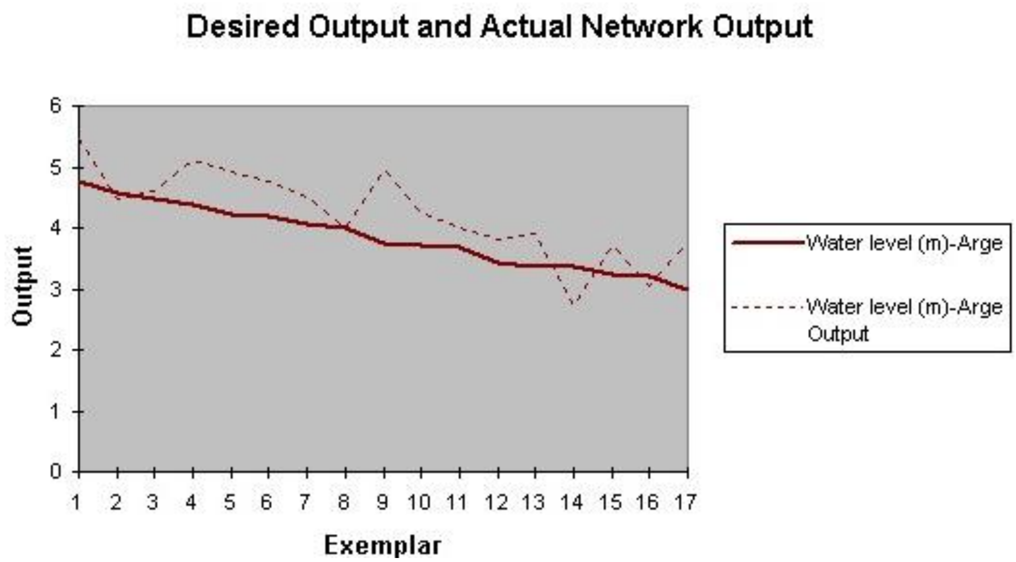
Best Network	*Training*
Run #	2
Epoch #	1000
Minimum MSE	0.016859077
Final MSE	0.016859077

Figure 10. Mean square errors versus epoch for training data set.



Performance	*Water level (m)-Arge*
MSE	0.215670954
NMSE	0.355043723
MAE	0.370865094
Min Abs Error	0.000772542
Max Abs Error	1.124808012
r	0.905641349

Figure 11. Measured versus predicted data (training stage) and error measures.



Performance	Water level (m)-Arge
MSE	0.327713365
NMSE	1.231979489
MAE	0.490568167
Min Abs Error	0.004859394
Max Abs Error	1.205223045
r	0.801181684

Figure 12. Predicted versus measured data (testing stage) and error measures.

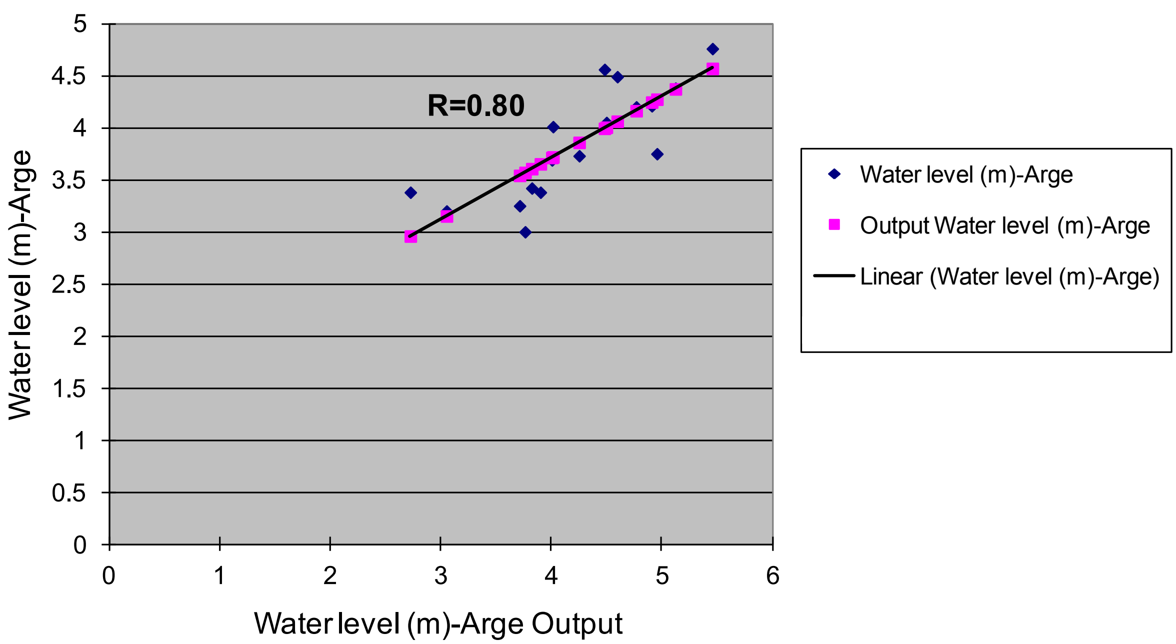
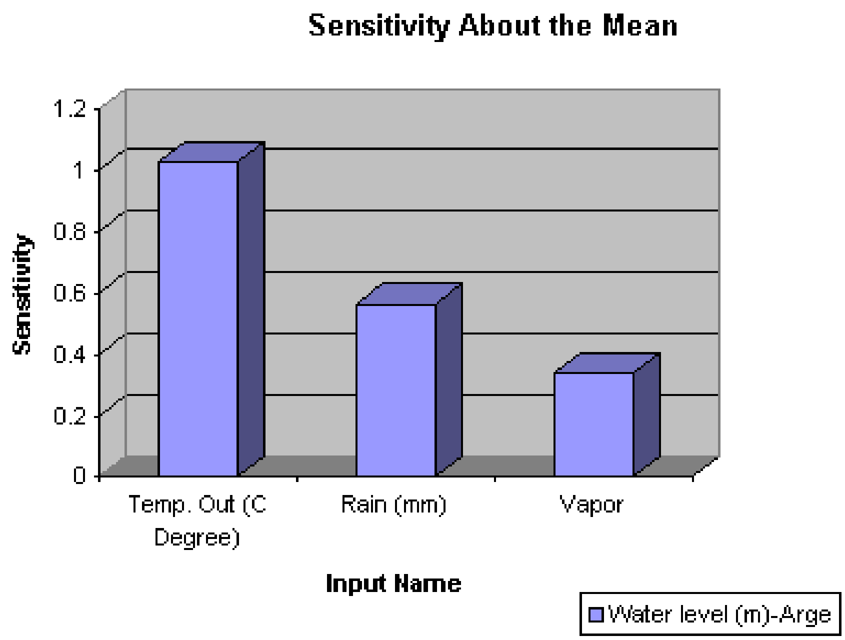


Figure 13. Predicted versus measured data (Testing stage).



Sensitivity	Water level (m)-Arge
Temp. Out (C Degree)	1.025137182
Rain (mm)	0.565314722
Vapor	0.338642631

Figure 14. Sensitivity analysis of the inputs.

one cannot ignore rainfall in the analysis of groundwater levels.

Conclusions

This study revealed that:

1) ANN can be an effective tool for prediction of water levels in a well.
2) NeuroSolutions software is a user-friendly package program.
3) Temperature and precipitation are sensitive parameters in the prediction of water levels.
4) Evaporation amount is the least sensitive parameter in the prediction of water levels.
5) Water levels in a well can be predicted by ANNs using average weekly data.
6) The model performance may be improved by employing more meteorological variables (humidity, solar radiation, etc) and hydrological variables (soil temperature, soil cover, etc.).
7) The model can be tried for hourly, daily, and monthly average and/or seasonal average predictions of well water levels.

REFERENCES

Affandi AK, Watanabe K (2007). Daily groundwater level fluctuation forecasting using soft computing technique. Nat. Sci. 5(2):110-119.

ASCE (2000). Task Committee on Application of Artificial neural networks in Hydrology: “Artificial neural networks in hydrology (I & II)”. J. Hydrol. Eng. 5(2):115-136.

Ayed AS (1997). Parametric Coat Estimating of Highway Projects using Neural Networks, A thesis submitted to the School of Graduate Studies in the partial fulfillment of the requirements for the degree of Master of Engineering, Faculty of Engineering and Applied Sciences. Canada: National Library of Canada.

Bustami R, Bessaih N, Bong C, Suhaili S (2007). Artificial Neural Network for Precipitation and Water Level Predictions of Bedup River. Int. J. Comp. Sci. 34(2):228-239.

Chau KW, Wu CL, Li YS (2005). Comparison of several flood forecasting models in Yangtze River. J. Hydrol. Eng. ASCE. 10(6):485-491.

Daliakopoulos IN, Coulibaly P, Tsanis IK (2005). Groundwater level forecasting using artificial neural networks. J. Hydrol. 309(1-4):229-240.

Hoo KA, Sinzinger ED, Piovoso MJ (2002). Improvements in the predictive capability of neural networks. J. Process Control 4(12):193-202.

Lam JC, Wan KKW, Yang L (2008). Solar radiation modelling using ANNs for different climates in China. Energy Convers. Manage. 49(5):1080-1090.

Muttil N, Chau KW (2006). Neural network and genetic programming for modelling coastal algal blooms. Int. J. Environ. Pollut. 28(3-4):223-238.

Tayfur G (2012). Soft computing in water resources engineering. WIT Press, Southampton, England, UK, 267 p.

Wang WC, Chau KW, Cheng CT, Qiu L (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J. Hydrol. 374(3-4):294-306.

Wu CL, Chau KW, Fan C (2010). Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J. Hydrol. 389(1-2):146-167.