A Study on Groundwater Level Prediction using Artificial Neural Network and Hybrid Model

1Ankit Panchal, 2Sunil Ajmera
PG Student, Professor
Department of Civil Engineering and Applied Mechanics, S.G.S.I.T.S, Indore, M.P., India

ABSTRACT

Analysis of groundwater level is needed for maintaining groundwater availability. There is always a concern about the lifelong sustainability of groundwater resources and reduction of groundwater level in particular areas in India. It is therefore necessary to implement mechanisms and systems that can be employed to predict the groundwater level. There is an immense importance of groundwater level simulation for the groundwater resource optimization. In this study, Artificial Neural Network (ANN) model and Hybrid Model (HM) have been developed to predict the groundwater level and their performances were compared. A feed forward multilayer perceptron type ANN along with Levenberg-Marquardt (LM) algorithm is adopted for the development of ANN, while HM is developed by the techniques of ANN and regression modelling in which the output layer of ANN was changed to a Non-Linear Regression Model (NLRM). The rainfall and groundwater level data on monthly basis were used to develop the models i.e. taking rainfall and past groundwater level as input and present groundwater level as output. For the present case, the study area is Kasturbagram, Indore, (M.P.). Performances of both ANN model and Hybrid model has been examined using statistical measures viz. Correlation Coefficient (R), Coefficient of Determination (R²), Root Mean Square Error (RMSE), Average Arithmetic Relative Error (AARE), Nash- Sutcliff Efficiency (E), and Threshold Statistics (TS). The result revealed that both the models can be successfully used for prediction of groundwater level, while comparing both the models, the performance of HM was found to be better than ANN model.

Keywords: Artificial Neural Network, Feed Forward Multilayer Perceptron, Hybrid Model, Levenberg-Marquardt, Non-linear regression model

1. INTRODUCTION

Groundwater is one of the most important and dependable resource for agricultural, domestic and industrial prospects. The demand of groundwater is successively rising because of its delightful features such as; slow moving, large storage volume and long retention time, can be drawn on demand, less risk free than surface water sources, etc. In India, the contribution of groundwater to the rural population, urban population and agricultural sectors are about 80%, 50%, and 60% respectively. It is the only reliable source of supply in some areas, also in other areas it is chosen due to its easily available nature. In recent times, the supply of groundwater has been used in excess, especially in the agricultural and industrial sector which results in the groundwater level declination. For the sustainable yield of groundwater resources the necessary step is to precisely predict the groundwater level and for this suitable decision should be made by the water resource managers, policy makers and engineers that overcome the adverse effects in groundwater. One has to manage the groundwater resources in effective manner by knowing the behaviour of groundwater level fluctuations first. Since it is a complex task to model groundwater fluctuations, there exist conceptual modeling techniques for obtaining groundwater level. However, this conceptual based model requires huge amount of data for the simulation of groundwater fluctuations, and are commonly very rigorous or costly to obtain. Earlier studies developed various models for predicting groundwater level by employing soft computing, analytical, stochastic techniques. In modern times, the utilization of Artificial Neural Network (ANN) and Hybrid Model (HM) are seen to be quite suitable in the field of groundwater sector. Artificial Neural Network model is considered as ‘black box’ models and it had also given many promising and satisfactory outcomes in the water resources and hydrology sectors. For modelling a dynamic nonlinear system, ANN properties are greatly suited. The application of ANN in hydrology sector started in 1980s, came to regular practice during 1990s, however during 2000s there has been massive growth of ANN in the hydrology and water resources. Many researchers adopted different type of ANN structures in their studies. Lallahem et al. (2005) showed the flexibility of using ANN technique for the groundwater level estimation in piezometers executed in an unconfined chalky aquifer. Daliakopoulosa et al. (2005) simulated the prediction of groundwater level up to 18 months ahead. Coppola et al. (2005) used ANN for water level prediction of two monitoring wells. Nayak et al. (2005) investigated the ANN technique for forecasting the fluctuations in the groundwater level in an unconfined coastal aquifer. Krishna et al. (2008) studied ANN method for the prediction groundwater levels in individual wells with the lead of one month. Nourani et al. (2008) used ANN method for the groundwater level estimation in some piezometers placed in an aquifer. Sreekanth et al. (2009) studied the performance of the ANN model for forecasting groundwater level by using feed-forward neural network which is trained with Levenberg–Marquardt algorithm. Mohanty et al. (2009) studied ANN model for forecasting of groundwater level in a river island of tropical humid region, eastern India. Moosavi et al. (2012) studied numerous ANN model structures for forecasting groundwater level with 1, 2, 3 and 4 months ahead for two sub-basins under two case studies. Mohanty et al. (2015) studied an application of ANN technique for weekly groundwater level forecasting for multiple wells situated over a river basin. The use of Hybrid technique in prediction of groundwater level is somewhat limited. Although, several researchers have developed Hybrid Model for groundwater level prediction, Jalalkamali et al. (2010), Dash et al. (2010) made an effort to generate a hybrid ANN model with the help of ANN and genetic algorithm technique. Jain and Kumar (2009) adopted a Hybrid ANN regression model structure from a trained neural network for information extraction. Kumar et al. (2013) also developed an ANN model and Hybrid model (HM) to predict the groundwater level fluctuations and compared their performances.
2. MODELS

2.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is the strong technique in the field of computer modeling. These computational tools are the nonlinear information (signal) processing devices, which are built from interconnection of elementary processing devices called ‘Neurons’ for studying the structural-functional relationship of the human brain. In a basic ANN structure mainly three layer are there as shown in Figure 1. The first layer is the input layer which connects with the input variables. The final layer in which the output variables are present is known as output layer. The intermediate layer that lies in between the input and output layer is called hidden layer. There may be one or more than one hidden layer.

![Figure 1: Structure of Artificial Neural Network](image)

2.2 Hybrid Model (HM)

A Hybrid Model (HM) is a model which is developed by comprising the techniques of Artificial Neural Network and Regression Modeling (kumar et al.; 2013). After desirable and satisfactory outcomes of training and testing, best network of ANN is procured and that structure of ANN is used for the development of HM. As a basic ANN network has three layers (i.e. input, hidden and output), a Hybrid Model is formed when the output layer of the network is replaced by regression model (i.e. Linear Regression Model (LRM) or Non-Linear Regression Model (NLRM)). In present study the output layer of ANN network is replaced with NLRM to form a Hybrid Model as shown in Figure 2.

![Figure 2: Structure of a Hybrid Model](image)

3. STUDY AREA AND DATA ACQUISITION

For the present work the study area is located in Indore (M.P.). Indore city is situated in the western region of Madhya Pradesh, on the southern edge of the Malwa plateau. The geographical location of Indore is 22° 43’ N latitude and 76° 42’ E longitude. The city covers with an area of 530 km², at an altitude of 550 meters above mean sea level. Topography of Indore is full of rocky terrains and undulated plateaus. Also the soil found here is deep medium black soil. The well is located at the Kasturbagram, Indore, (M.P.).The geographical location of the well is 22°38’00”N latitude and 75°54’25”E longitude. The well is a type bore well with the depth 60.06 m. Figure 3 represents the location of the study area.

The data acquired for the present study is rainfall and groundwater level. The monthly rainfall data from 2005-2017 has been collected from the website www.indiawaterportal.org and hydro.imd.gov.in. The monthly groundwater level data from 2005-2017 has been collected from the “Groundwater Department, Indore”.

![Figure 3: Location of the study area](image)
4. METHODOLOGY

4.1 Artificial Neural Network Development

As ANN is described in the article 2.1, for artificially executing the functioning of the biological neurons, it is necessary to develop a proper network. For this various parameters should be adopted to develop ANN model. At first the overall data (monthly rainfall and groundwater level from 2005-2017) is divided into two sets i.e. training and testing data. 70% of the data is taken for training data and the remaining 30% is taken for testing data. The scaling of the data is done in between the range 0.1-0.9 by

\[ X_n = \frac{(X_b - X_a)(X_0 - X_{\text{min}})}{(X_{\text{max}} - X_{\text{min}})} + X_a \]  

where, \( X_n = \) Scaled value, \( X_a = 0.1 = \) Lower limit of the scaled value, \( X_b = 0.9 = \) Upper limit of the scaled value, \( X_0 = \) Value to be scaled, \( X_{\text{max}} = \) Maximum value in the data, \( X_{\text{min}} = \) Minimum value in the data

For the simulation of ANN, MATLAB software is used in which Multi-Layer Perceptron (MLP) programming is written for the suitable data sets. Before this a regression analysis is also done with the help of XL-STAT software for generating the regression equation which is introduced in the programming along with the input parameters. For present case, a non-linear regression equation is generated for training data set. The training data is run until the perfect result of the MLP network is obtained. After this, the validity of developed equation is checked by applying it on testing data set. For the present study ANN structure is found to be 7-5-1 i.e. input layer having 7 inputs, 1 hidden layer having 5 hidden nodes and 1 output layer.

4.1.1 ANN Architecture

4.1.1.1 Feed Forward Neural Network (FFNN)

In present study Feed Forward Neural Network (FFNN) is used, which is the most popularly applied Neural Network in ANN modeling. The architecture of this network is the simplest amongst all neural networks. The main advantage of Feed Forward Neural Network is that they are easy to deal and performs with good accuracy. The arrangement of a basic Feed Forward Neural Network is in layers, in which number of nodes are presented in all layers, the nodes which are presented in a layer are connected to the nodes of another layer next to it, but are not connected to the nodes of same layer. The information processes from input to output side, i.e. the information is processed from the nodes of input layer to the nodes of hidden layer and then to the final output node in the forward direction. Also in this type of network, there are no cycles or loops. There can be one or more hidden layers having number of nodes presented in that particular layer. A Feed Forward Neural Network is shown in Figure 4.

Figure 3: Location of the Study Area

Figure 4: Feed Forward Neural Network (FFNN)
4.1.1.2 Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) is the first and basic tool of Artificial Neural Network modelling. Mainly a MLP structure consists of three layers viz. input layer, one or more than one hidden layers and output layer. The first layer is the input layer that where the data is given to the network, the final layer where the output results are obtained is known as output layer. The intermediate layer present between the input and output layer where the data is processed is known as hidden layer. In these layers number of nodes are present and each node of a layer is connected with the nodes of another layer, but not of the same layer and consists of particular weight (W) and bias values (b) as shown in Figure 5.

**Figure 5: Multi-Layer Perceptron (MLP) Neural Network**

4.1.1.3 Training Algorithm

As ANN architecture is decided, the next step is to select an algorithm used for training the network. Various types of algorithm are there to train a network like Levenberg-Marquardt algorithm (LM), Back-propagation (GDX), Bayesian regularization (BR). Levenberg-Marquardt algorithm is adopted for the present case. In ANN, Levenberg-Marquardt Algorithm (LM) gives an optimum solution also it is considered as more efficient and stable. It was observed in the past studies that the LM is an effective algorithm to produce an Artificial Neural Network model for hydrological forecasting (Daliakopoulos et al., 2005; Sreekanth et al., 2009). The LM algorithm is an updation of classic Newton algorithm to find out the best solution for a minimization problem (Daliakopoulos et al., 2005). For small networks structures LM Algorithm is widely used because of its huge computation and memory requirements (Mair and Dandy; 1998).

4.1.1.4 Activation functions

The activation functions (transfer function) are the functions which are applied on the hidden layer and output layer of the structure for the calculation the outputs of the neurons. For the present case, the activation function applied at the hidden layer was logistic sigmoid or logsig, also for output layer it is set to purlin. The activation functions are used to weighted sum of the inputs of a neuron to produce the output. Mathematical expression of these function are given

For linear or purlin -

\[ f(x) = x \]  (2)

For logsig -

\[ f(x) = \frac{1}{1 + e^{-x}} \]  (3)

4.2 Hybrid Model Development

As HM is described in the article 2.2, a Hybrid Model is formed by the dissection of the trained network (Jain and Kumar 2008) i.e. dissection of the best ANN network obtained after training/testing. Also according to the study of Kumar et al. (2013), for the present case NLRM is used in place of output layer of the best ANN network. Inputs were given to the input layer and the outputs of the hidden layer i.e. hidden nodes output were calculated by the use of ANN at the hidden layer. Now the outputs obtained from the hidden neurons will work as NLRM’s input of Hybrid model. The hidden neurons output is determined by

\[ H_1 = F_0(X_1W_{11} + X_2W_{12} + X_3W_{13} + \ldots \ldots + X_nW_{1n} + B_1) \]  (4)

where, \( H_1 \) = First hidden neuron of the hidden layer, \( F_0 \) = Activation function used during ANN training/testing, \( X_1, X_2, X_3, X_n \) = First, second, third, \( n \)th input of the input layer, \( W_{11}, W_{12}, W_{13}, W_{1n} \) = Weight of the first, second, third, \( n \)th input layer, \( B_1 \) = Bias value obtained at the hidden layer. The structure of a Hybrid Model with network parameters is shown in Figure 6. After the
hidden neurons output are calculated, generate the regression equation taking that hidden neurons as inputs and observed groundwater level as target in XL-STAT to obtain regression coefficients and predicted groundwater level (i.e. Gp).

\[ G_p = \alpha_1H_1 + \alpha_2H_2 + \ldots + \alpha_nH_n + \beta_1(H_1)^2 + \beta_2(H_2)^2 + \ldots + \beta_n(H_n)^2 \]

where, \(G_p\) = Predicted Groundwater level, \(\alpha_i, \beta_i\) = Regression coefficients, \(H_1, H_2, \ldots, H_n\) = 1st, 2nd, nth Hidden neurons of the hidden layer

5. MODEL PERFORMANCE CRITERIA

The performance of developed model is evaluated by applying various statistical measures viz. Correlation Coefficient (R), Coefficient of Determination (R²), Root Mean Square Error (RMSE), Nash-Sutcliff Efficiency (E), Average Absolute Relative Error (AARE), and Threshold Statistics (TS). Mathematical form of these statistical measures are given by

\[ R = \frac{\sum_{i=1}^{N} (G_o_i - \overline{G_o}) \cdot (G_p_i - \overline{G_p})}{\sqrt{\sum_{i=1}^{N} (G_o_i - \overline{G_o})^2 \cdot \sum_{i=1}^{N} (G_p_i - \overline{G_p})^2}} \]

\[ R^2 = \left( \frac{\sum_{i=1}^{N} (G_o_i - \overline{G_o}) \cdot (G_p_i - \overline{G_p})}{\sqrt{\sum_{i=1}^{N} (G_o_i - \overline{G_o})^2 \cdot \sum_{i=1}^{N} (G_p_i - \overline{G_p})^2}} \right)^2 \]

\[ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (G_o_i - G_p_i)^2} \]

\[ AARE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{G_p_i - G_o_i}{G_o_i} \right| \times 100 \]

\[ E = 1 - \frac{\sum_{i=1}^{N} (G_o_i - G_p_i)^2}{\sum_{i=1}^{N} (G_o_i - \overline{G_o})^2} \]

\[ TS_x = \frac{n_x}{N} \times 100 \]

where, \(G_o_i\) = Observed groundwater level at \(i^{th}\) interval, \(\overline{G_o}\) = Average observed groundwater level, \(G_p_i\) = Predicted groundwater level at \(i^{th}\) interval, \(\overline{G_p}\) = Average predicted groundwater level, \(N\) = Total no. of data points, \(n_x\) = No. of data points whose AARE values are less than \(x\%\), \(x\) = percentage at which TS is to be calculated.

6. RESULTS AND DISCUSSIONS

Various statistical measures were performed in the present study for the groundwater level prediction by ANN and HM, the results of statistical measures shown in Table 1, all of them shows good result. For both ANN and HM, values of R and R² are close to 1. The errors RMSE and AARE are less, the values of E is also fair and the results of TS (at 10% and 15%) are also good. All these statistical measures fits in their criteria, which indicates that ANN and HM gives good performance during both training and testing phase. Also while comparing the results of ANN and HM, HM gives better results than ANN.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>R²</th>
<th>RMSE</th>
<th>R</th>
<th>E</th>
<th>AARE %</th>
<th>TS 10%</th>
<th>TS 15%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>During Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>0.9137</td>
<td>0.0238</td>
<td>0.9559</td>
<td>0.8980</td>
<td>6.0773 %</td>
<td>80.73 %</td>
<td>91.74 %</td>
</tr>
<tr>
<td>HM</td>
<td>0.9451</td>
<td>0.0210</td>
<td>0.9721</td>
<td>0.9204</td>
<td>5.9037 %</td>
<td>83.49 %</td>
<td>93.58 %</td>
</tr>
<tr>
<td><strong>During Testing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>0.9051</td>
<td>0.0870</td>
<td>0.9360</td>
<td>0.8751</td>
<td>7.3245 %</td>
<td>76.60 %</td>
<td>80.85 %</td>
</tr>
<tr>
<td>HM</td>
<td>0.9278</td>
<td>0.0663</td>
<td>0.9632</td>
<td>0.9119</td>
<td>7.7530 %</td>
<td>80.85 %</td>
<td>89.36 %</td>
</tr>
</tbody>
</table>
The scatter plots (Figure 7 and Figure 9) and graphical performance (Figure 8 and Figure 10) shows the relation between Go and Gp of ANN and HM during training and testing phase. Both ANN and HM shows good result, also while comparing the scatter plots and graphs of ANN and HM, HM graphs are better than ANN graphs.

Figure 7: (a) Scatter plot of ANN during training (b) Scatter plot of HM during training

Figure 8: (a) Graphical performance of ANN during training (b) Graphical performance of HM during training

Figure 9: (a) Scatter plot of ANN during testing (b) Scatter plot of HM during testing
7. CONCLUSION

In this study, prediction of groundwater level is done by ANN and HM, also performances of these both models were compared. A three layer ANN structure is developed for the training and testing phase. For Hybrid Model, non-linear regression model is used in place of output layer of the ANN structure. Based on the performance and results obtained in the present study, the proposed methodology for predicting groundwater level by Artificial Neural Network (ANN) and Hybrid Model (HM) are suitable for the study area, i.e. Kasturbagram, Indore, M.P. Multilayer perceptron tool of ANN gave best output response for groundwater level prediction. All the statistical measurements performed in the study viz. Correlation Coefficient (R), Coefficient of Determination (R²), Root Mean Square Error (RMSE), Nash-Sutcliff Efficiency (E), Average Absolute Relative Error (AARE), and Threshold Statistics (TS) for the ANN and HM gives good response. Also on comparison between ANN and HM, HM gave better results than ANN for groundwater level prediction.

REFERENCES


