Available online at www.ijpab.com

ISSN: 2320 – 7051
Int. J. Pure App. Biosci. SPI: **6 (3):** 640-652 (2018)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Groundwater Level Prediction Using Som-RBFN Multisite Model in Tumaria Canal Command

M. Kumar[1*] and Sachin Kumar Singh[2]
[1]PhD Student, Dept. of Irrigation and Drainage Engineering,
[2]PhD Student, Dept. of Soil & Water Conservation Engg.,
G.B. Pant University of Agriculture and Technology, Pantnagar – 263145, (U.S. Nagar), Uttarakhand, India
*Corresponding Author E-mail: mantu06.iitkgp@gmail.com
Received: 5.05.2018 | Revised: 14.06.2018 | Accepted: 25.06.2018

ABSTRACT

In this paper, a groundwater level forecasting model was proposed by combining the theory of self-organizing map (SOM) and radial basis function network (RBFN). The proposed model was referred to as SOM-RBFN model. Recently, RBFN has been applied in time series forecasting. Traditionally, the number of hidden units and the positioning of the radial basis centers are crucial problems for establishing RBFN. The proposed model could decide the number of RBFN's hidden units with using the two-dimensional feature map which was constructed by SOM, and then it could determine the positioning of the radial basis centers easily. The proposed model was applied to groundwater level data in Tumaria canal command area from 1994 to 2015. It was found that the multisite model can predict the groundwater level more precisely than the single-site model. Moreover, it was also found that the six-site model was more competent in predicting groundwater level as compared to the single-site model and four-site model. For groundwater level prediction, the SOM-RBFN multisite model was recommended as an alternative to the other methods because it has a clear principle and a simple structure.

Key words: *Radial basis function network; Self-organizing map; Neural networks; Groundwater level prediction.*

INTRODUCTION

Groundwater resource is commonly the most important water resource in Tumaria canal command area of Udham Singh Nagar and Moradabad districts that are often subject to water shortage. It plays a fundamental role in supplying clean and safe water to competing uses for domestic, industrial and agricultural sectors, and increasing attentions are also paid to its significance for ecological integrity. However, groundwater aquifer systems always feature complexity, high nonlinearity, being multi-scale and random as a result of the frequent interactions between surface water and groundwater as well as acute human disturbance[8]. Thus, effective modeling techniques would be required for providing efficient ground water management strategies.

Cite this article: Kumar, M. and Singh, S.K., Groundwater Level Prediction Using Som-RBFN Multisite Model in Tumaria Canal Command, *Int. J. Pure App. Biosci.* SPI: **6(3):** 640-652 (2018).

As for dynamic groundwater level (GWL) prediction, physical based or conceptual models represent the hydrological variables and physical processes in real-world systems[6], but they have practical limitations in terms of prediction accuracy as a result of unavoidable discrepancies between the model and the real-world system[1,8,9]. As far as increasingly scarce water resources accompanying with expanding population growth are concerned, improvements and innovations in groundwater predictions become critical.

A number of mathematical models describing groundwater levels have been developed in the past[11,2,10,3]. These models arise from different conceptual models and many different numerical schemes are used to solve the governing equations. However, the calibrations of these models are very difficult because a lot of parameters need to be determined. Furthermore, these models require a large quantity of good quality data and a comprehensive understanding of the underlying physical process in a system.

A groundwater level forecasting model was developed based on combination of radial basis function network (RBFN) and self-organizing map (SOM). This proposed model was referred to as SOM-RBFN model[4]. The proposed model was applied to actual groundwater level data in southern Taiwan from 1997 to 2003. It was found that the multisite model can predict the 1 month ahead groundwater level more precisely than the single-site model. Model based on the combination of the back-propagation network (BPN) and the self-organizing map (SOM), named improved multisite SOM-BPN model for groundwater stations of the alluvial fan of the Zhuoshuixi River in southern Taiwan[5]. In the proposed model, the SOM was used to determine the number of hidden layer neurons, and the autoregressive integrated moving-average (ARIMA) model and semivariogram were used to determine the number of input neurons. Self-organizing map (SOM) was applied to identify spatially homogeneous clusters of GWL piezometers, while GWL time series forecasting was performed through developing a stepwise cluster multisite inference model with various predictors including climate conditions, well extractions, surface runoffs, reservoir operations and GWL measurements at previous steps Han *et. al.*[7].

Computer aided techniques are versatile tools in its own towards the direction of precise and readily solution for various applied science and technology based problems. The groundwater engineering is one of the most significant users of such techniques for analysis, design, simulation, modeling etc. Whatever may be the stream or problem domain for any engineering stream, the most important aspect is to conceptualization the logic which is to be analyzed, designed and accordingly coded in computer software language. The presentation of such conceptualization can be in the form of algorithm, flow charts, UML charts etc. several such computer soft such as ANN and GA etc. are very useful to solve the flow problems in ground water hydrology and other such related field, which can be used in field condition studies.

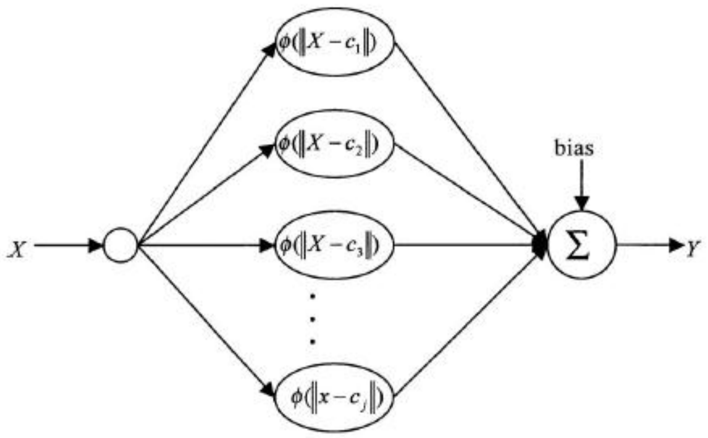
In this paper, a groundwater level forecasting model is developed. First, a detailed review of the RBFN and SOM algorithm is presented, and then the SOM-RBFN model and fitted parameters are used to predict the groundwater level in the Tumaria canal command area. The results and conclusions are then given in the final two sections.

MATERIAL AND METHODS

In this section, the structures of RBFN and SOM are introduced. Furthermore, a groundwater level forecasting model, which is based on the RBFN and SOM, is developed. The proposed model is referred to as SOM-RBFN model herein.

Radial Basis Function Network

The architecture of RBF networks is a three-layer feed-forward network that consists of one input layer, one hidden layer (also called receptor layer), and one output layer. The hidden layer of an RBFN is nonlinear, whereas the output layer is linear. The argument of the activation function of each hidden unit in an RBFN computes the Euclidean distance between the input vector and the center of hidden unit in the network. The basic architecture of a three-layered neural network is shown in Fig. 1. The learning algorithm can be described as follows.



Input Layer **Hidden Layer** **Output Layer**

Figure 3.2: Structure of RBNF

The input data *Z* is a *P*-dimensional vector, *Z*= $[z_1, z_2, \ldots\ldots, z_p]^T$. In the structure of RBFN, the input layer serves only as input distributor to the hidden layer. The dimensionality of hidden units is the same as that of the input data. The response from the j^{th} hidden unit for the i^{th} input data z_i has the following form:

$$\phi_j(z_j) = \phi\left(\left|\left|z_j - c_j\right|\right|\right) \quad j = 1,2, \ldots\ldots\ldots, Q \tag{1}$$

Where, || denotes the Euclidean norm; c_j=center of the j^{th} unit in the hidden layer; ϕ ( ) =activation function; and *Q*=number of hidden units. In the structure of RBFN, the activation function of hidden units is symmetric in the input space, and the output of each hidden unit depends only on the Euclidean distance between the input vector and the center of the hidden unit. The activation function has different forms, the most commonly used activation function is described by the Gaussian function.

$$\phi(Z) = \exp\left[-\frac{||Z-c||^2}{2\beta^2}\right] \tag{2}$$

where, β=centre width and β can be obtained from (Haykin 1994)

$$\beta = \frac{d_{max}}{\sqrt{2N_h}} \tag{3}$$

Where,

d_{max}=maximum distance between the centers of hidden units and N_h=number of hidden units.

The weighted sum of the inputs at the output layer is transformed to the network output using a linear activation function. The activity of the r^{th} unit in the output layer, $\hat{y}_r$, can be computed using the following equation:

$$\hat{y}_r = w_0 + \sum_{q=1}^{N_h} w_{qr}\phi_q(z) \quad r = 1,2, \ldots\ldots\ldots\ldots\ldots\ldots, N_R \tag{4}$$

Where, $\phi_q(z)$ = response of the q^{th} hidden unit resulting from all input data; w_{qr}= connecting weight between the q^{th} hidden unit and the r^{th} output unit; w_0=bias term; and N_R=number of output units.

3.4.2 Self-Organizing Map

The Self Organizing Map (SOM) algorithm of Kohonen, also called Kohonen feature map, is one of the best known artificial neural network algorithms. SOMs are a unique class of neural networks, since they construct topology preserving mappings of the training data where the location of a unit carries semantic information. Therefore, the main application of this algorithm is clustering of data, obtaining a two dimensional display of the input space that is easy to visualize. The SOM consist of two layers of units: A one dimensional input layer and a two dimensional competitive layer, organized as a 2D grid of units. The SOM is trained using an unsupervised competitive learning algorithm which is a process of self organization. The SOM algorithm can be described as follows.

The input layer is an array of *M* neurons. It can be denoted by

$$X = [x_1, x_2 \dots \dots \dots \dots \dots \dots \dots, x_M]^T \qquad (5)$$

The output layer includes the output neurons u_j, j=1,2,………..,N, which are typically organized in a planar (2D) lattice. The weights from the input layer neuron to the output layer neuron are $w_{ij,}$ i=1,2,……………,M, j=1,2,………….,N. The weight vector of each neuron has the same dimension as the input pattern. The weight vector can be written as

$$W_j = [w_{1\,,}w_2, \dots \dots \dots \dots .., w_M]^T \quad j = 1,2, \dots \dots \dots .., N$$

The training process begins with all weights initialized to small random numbers. The SOM algorithm computes a similarity (distance) measure between the input vector *X* and the weight vector W_j of each neuron u_j. The Euclidean distance d_j between the weight vector W_j and input vector *X* is frequently used as the similarity measure

$$d_j = \left|\left|X - W_j\right|\right| = \sqrt{\sum_{i=1}^{M}(x_i - w_j)^2} \qquad (6)$$

Where, || denotes the Euclidean distance. The output neuron with the weight vector that is the smallest distance from the input vector is the winner. The weights of this winning neuron are adjusted in the direction of the input vector. The winning neuron is the center of the topological neighborhood. A typical choice of topological neighborhood function is Gaussian function

$$h_j = \exp(-\frac{\|u_j - u^*{}_j\|^2}{2\sigma^2}) \qquad (7)$$

Where, $\boldsymbol{h_j}$=topological neighborhood; σ = "effective widt" of the topological neighborhood; and $\boldsymbol{u^*{}_j}$ = winning neuron.

The change to the weight vector W_j can be obtained as

$$\Delta\boldsymbol{W_j} = \boldsymbol{\eta h_j (X - W_j)} \qquad (8)$$

Where, η=learning-rate parameter of the algorithm. Hence, the updating weight vector $W_j(t+1)$ at time t+1 is defined by

$$W_j(t+1) = W_j(t) + \eta(t)h_j(t)[X - W_j(t)] \qquad (9)$$

Where, $\eta(t)$ and $h_j(t)$= learning-rate parameter and the topological neighborhood at time *t*. Eq. (9) is applied to all the neurons in the lattice that lie inside the topological neighborhood of the winning neuron. Upon repeated presentations of the training data, the weight vectors tend to move toward the input pattern due to the neighborhood updating. That is, the adjustment makes the weight vectors to be similar to the input pattern. The winning neuron shows the topological location of the input pattern. The neighborhood of the winning neuron shows the statistical distribution of the input pattern. The output of the SOM is obtained using a dynamic patterns grid, which shows a dynamic representation of the neurons that are winning each pattern. Each cell in the grid represents a neuron in the output layer. Once the clusters are formed in the topology pattern, the data records from each cluster are sampled.

SOM-RBFN Model

In an RBFN, once the centers and widths of hidden units are determined, each weight in Eq. (4) can be determined by the least-squares method. In general, there are two steps in the design of RBFN. The first step initializes the centers using a clustering method. The second step determines the parameters and minimizes the error with respect to the connecting weights. Between the existing learning algorithms, the main difference resides in the first step. In other words, the positioning of the radial basis centers is a crucial problem for RBFN. In this paper, the number and centers of hidden units are determined using SOM, which projects high-dimensional data onto a low-dimensional grid and visually reveals the topological order of the original data. The proposed model is referred to as SOM-RBFN model herein.

Study Area and Data

The SOM-RBFN model is applied to forecast groundwater level in the Tumaria canal command area which is covered in Jaspur of Udham Singh Nagar and Thakurdwara of Moradabad districts. It is located between latitude 28° 20' N and 29° 23' N and laterally extends between longitudes 78° 24' E and 80° 08' E as shown in figure 1. There are six groundwater stations (Thakurdwara HP, Jaspur HP, Surjan Nagar HP, Patrampur HP, Angadpur HP, and Kashipur DW) in this area. Pre and post monsoon ground water level data (below ground surface) of last twenty one years has been obtained from Central Ground Water Board, Dehradun, Uttarakhand.

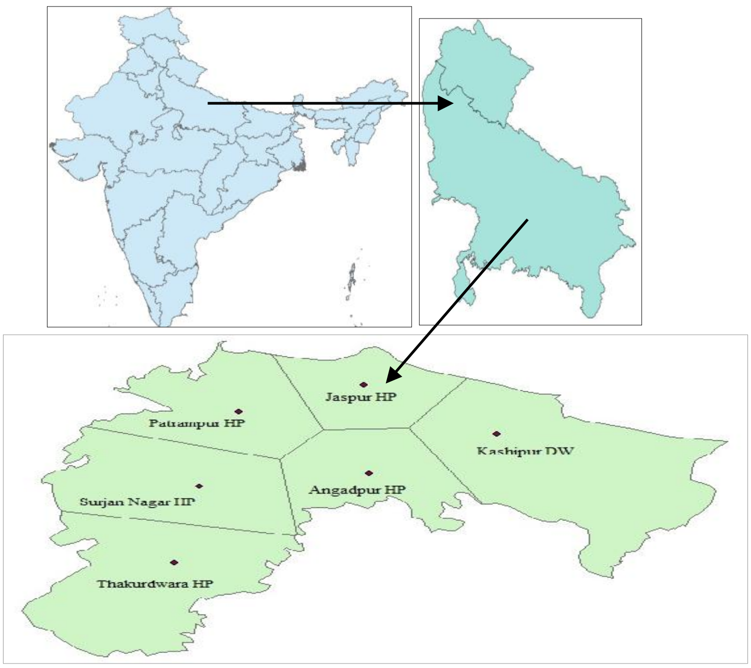


Fig. 1. Well location map of the study area.

Model Parameters

First, we used the application of autocorrelation function (ACF) and partial ACF (PACF) could be employed in software R version 3.2.3 (2015-12-10) (Figure2) for time series of depth to water table. According to the results of the sample ACF and PACF for the original time series, the past 4-period, past 2-period, and past 1-period seasonally average groundwater levels were used as input for the SOM-RBFN model with single station.

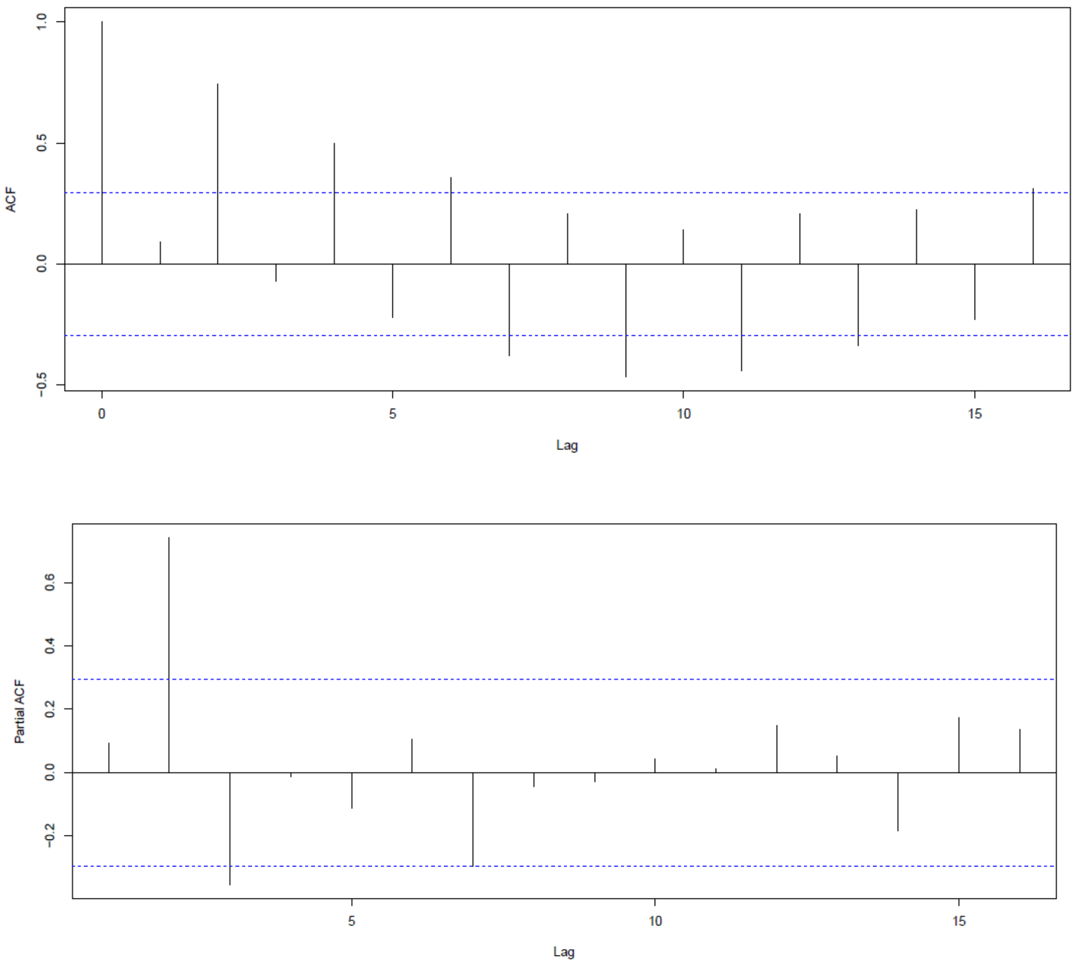


Fig. 2. ACF and PACF of the water table series.

In the SOM process, a SOM of a small dimension is the first choice. If the clustering result is reasonable and satisfactory, the SOM process is accepted. Otherwise, another SOM of a larger dimension is chosen to analyze input patterns. This step is continued until a satisfactory result is obtained. According to our experiments, a 2D feature map obtained on a network of 8×8 cells is adopted herein. Moreover, the initial value of the weight vector between input layer and hidden layer is a random value between 0.0 and 1.0, and the learning-rate parameter begins at 0.1 and ends at 0.01. After a total of 10000 iterations (200 times the number of output neurons), the SOM has been constructed and the 2D feature map has been obtained. In the RBFN process, the initial value of the weight vector between input layer and hidden layer is also a random value between 0.0 and 1.0.

In this paper, the seasonal average groundwater level data for six stations are six groundwater stations (Thakurdwara HP, Jaspur HP, Surjan Nagar HP, Patrampur HP, Angadpur HP, and Kashipur DW) are available. The primary objective of this paper is to investigate the effects of the multisites for seasonal average groundwater level forecasting. First, the single-site

SOM-RBFN model configuration is evaluated using the seasonal average groundwater level data for single station. Then the multisite SOM-RBFN model includes six-site model and four-site model. The former contains all stations and the latter contains four stations. Fig. 3 shows that the groundwater levels for KP station and JP station are similar to those for AG station and PM station, respectively. Hence, the four-site model is only applied for the stations KP and JP in this paper. The input variables for single-site model, six-site model, and four-site model are summarized in Table 1.

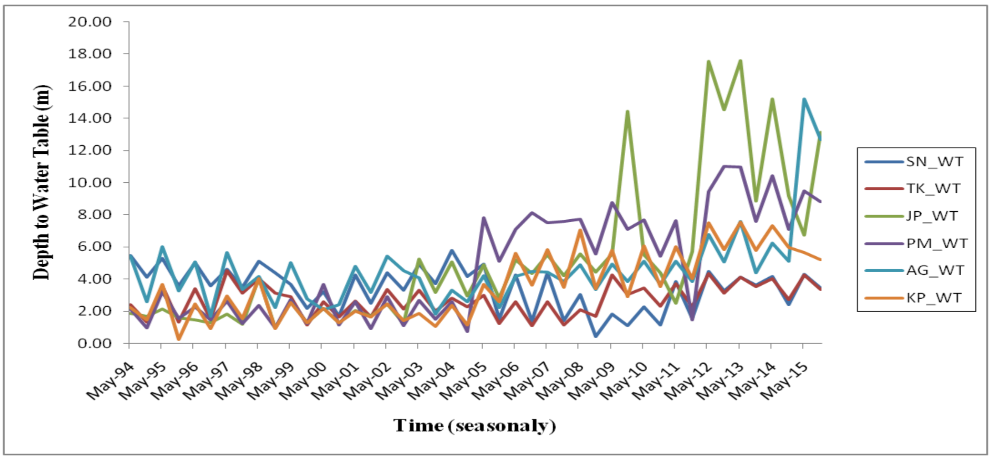


Fig. 3. Time series plots groundwater level data at the six groundwater stations

Table 1. Input Variables of Single-Site Model, Six-Site Model, and Four-Site Model

Model	Input variables
Single-site	The past 4-month, past 2-month, and past 1-month monthly average groundwater levels for individual site
Six-site	The past 4-month, past 2-month, and past 1-month monthly average groundwater levels for all sites
Four-site	The past 4-month, past 2-month, and past 1-month monthly average groundwater levels for the four sites

Criteria for Evaluating Model Performance

In order to evaluate the forecasting accuracy, the following two criteria are used:

1. Root-mean-square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}[L_{est}(t) - L_{obs}(t)]^2}{n}}$$

Where $L_{est}(t)$=estimated value at time t; $L_{obs}(t)$= observed value at time t ; and n= number of estimated values. The RMSE can be used to compare the performance of the three models. The model that yields the smallest value of RMSE is the best.

2. Coefficient of efficiency (CE)

$$CE = 1 - \frac{\sum_{t=1}^{n}[L_{est}(t) - L_{obs}(t)]^2}{\sum_{t=1}^{n}[L_{est}(t) - \bar{L}_{obs}(t)]^2}$$

where $\bar{L}_{obs}(t)$=average observed groundwater level. The CE can take a value between negative infinity and 1. The good model has relatively high CE.

RESULTS AND DISCUSSIONS

Simulated Result of Each Station

Surjan Nagar HP

First, the SOM has been constructed in the proposed model development process. Fig. 4 shows the 2D feature map obtained on a network of 8×8 cells. As shown in Fig. 4(a), the map of a single station can be divided into six regions. Therefore, the 28 training data can be grouped into six clusters. That is to say, the single-site SOM-RBFN model had six neurons in a hidden layer.

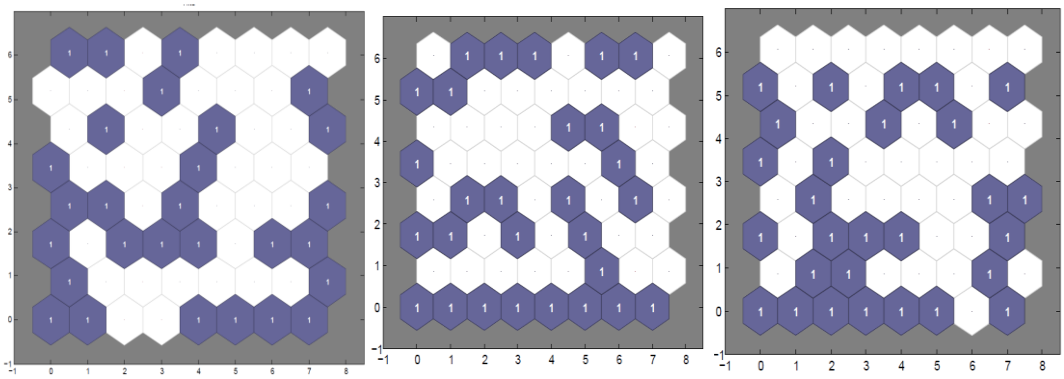


Fig. 4. 2D feature map for (a) single-site model; (b) six-site model; and (c) four-site model of Surjan Nagar HP station.

After the RBFN process is constructed, it can be applied for groundwater level time series analysis. As the same reason, the six-site model had six neurons and the four-site model had seven neurons. For training data, the comparisons of observed groundwater heads with values forecasted using the single-site SOM-RBFN model, the six-site model, and the four-site model are given in Fig. 5. Fig.6 presents the groundwater level forecasts using the three different models for testing data.

Thakurdwara HP

The 2D feature map obtained on a network of 8×8 cells is shown in Fig. 7. As shown in Fig. 7(a), the 28 training data can be grouped into five clusters. Therefore, the single-site SOM-RBFN model had five neurons in the hidden layer. As the same reason the six-site model and the four-site model had seven and seven neurons in the hidden layer, respectively. Fig. 8 presents the comparisons of observed groundwater heads with values forecasted using the single-site SOM-RBFN model, the six-site model, and the four site model for training data. Fig. 9 presents the groundwater level forecasts using the three different models for testing data.

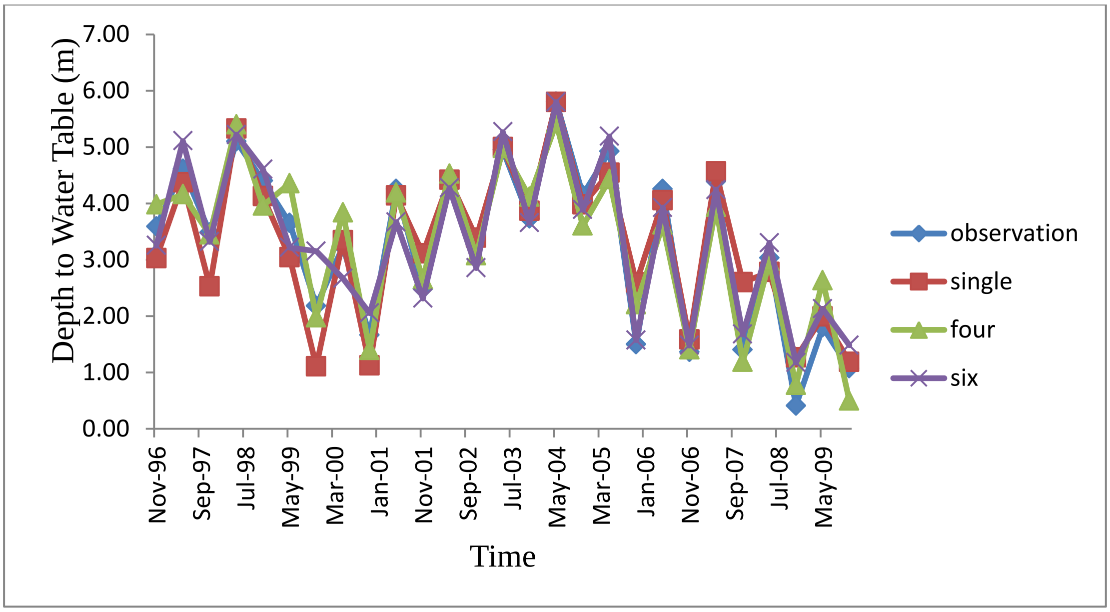


Fig. 5. Comparison of observed groundwater heads with values forecasted using the three models for training data of Surjan Nagar HP

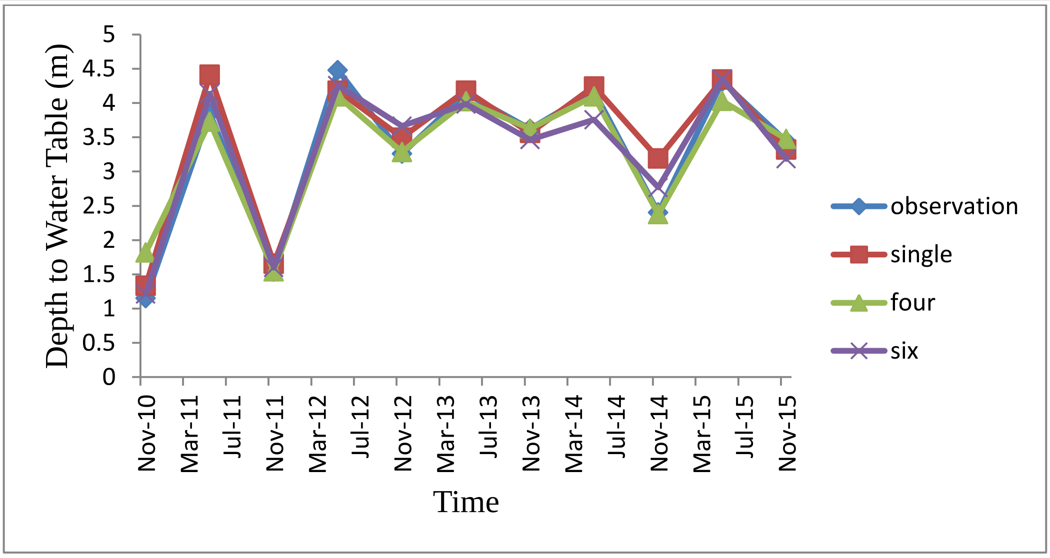


Fig. 6. Comparison of observed groundwater heads with values forecasted using the three models for testing data of Surjan Nagar HP

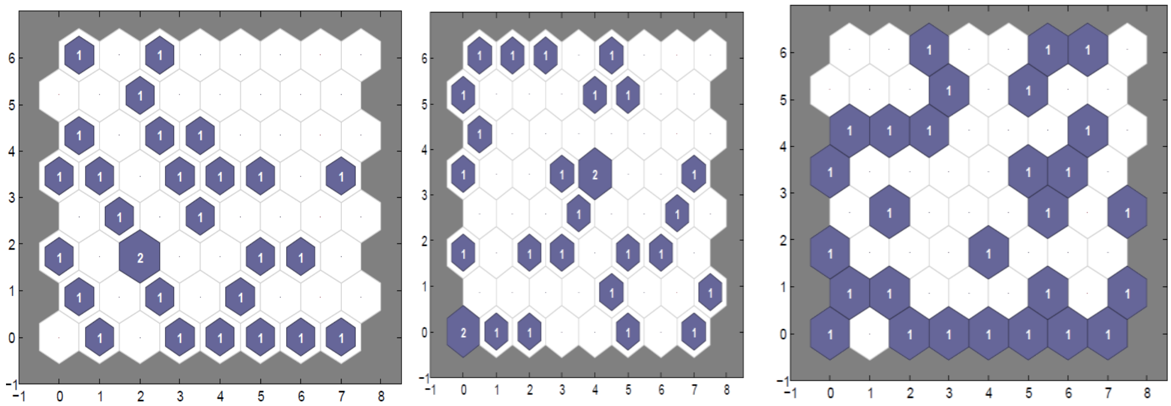


Fig. 7. 2D feature map for (a) single-site model; (b) six-site model; and (c) four-site model of Thakurdwara HP station.

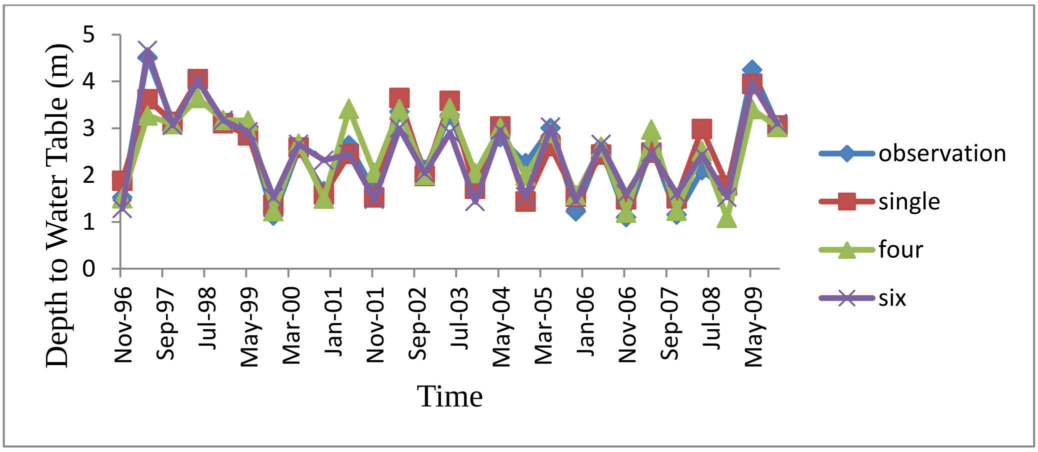


Fig. 8. Comparison of observed groundwater heads with values forecasted using the three models for training data of Thakurdwara HP station

Jaspur HP

Fig. 10 shows the 2D feature map obtained on a network of 8×8 cells for the training data of the three models. The single site SOM-RBFN model had eight neurons in the hidden layer according to Fig. 10(a). As the same reason the six-site model and the four-site model had five and five neurons in the hidden layer, respectively. For training data, the comparisons of observed groundwater heads with values forecasted using the single-site SOM-RBFN model, the six-site model, and the four-site model are given in Fig. 11. Fig. 12 presents the groundwater level forecasts using the three different models for testing data.

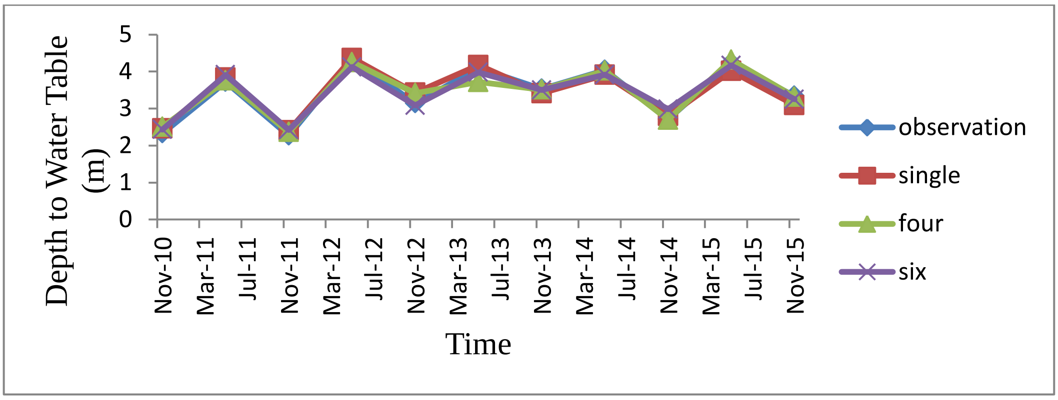


Fig. 9. Comparison of observed groundwater heads with values forecasted using the three models for testing data of Thakurdwara HP station

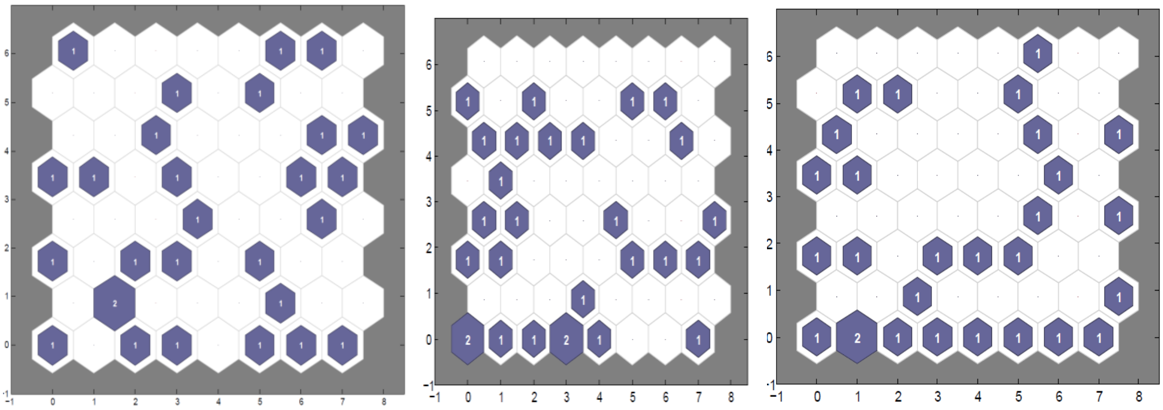


Fig. 10. 2D feature map for (a) single-site model; (b) six-site model; and (c) four-site model of Jaspur HP station.

Table 1. Performance of Single-Site Model, Six-Site Model, and Four-Site Model for Training Data

Criterion						
	RMSE			CE		
Site	Single-site	Six-site	Four-site	Single-site	Six-site	Four-site
Surjan Nagar HP	0.525	0.386	0.424	0.840	0.913	0.902
Thakurdwara HP	0.359	0.311	0.397	0.817	0.875	0.765
Jaspur HP	1.178	1.156	1.267	0.710	0.720	0.674
Kashipur DW	0.542	0.470	0.483	0.871	0.889	0.902

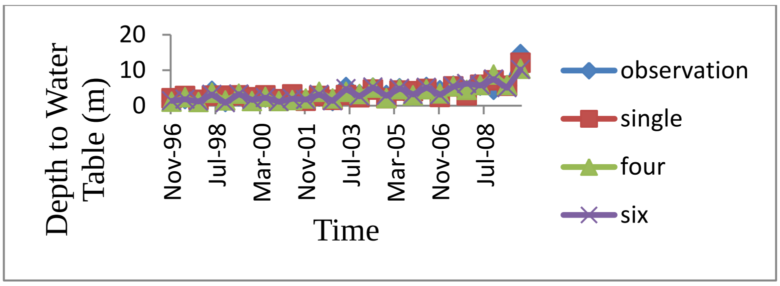


Fig. 11. Comparison of observed groundwater heads with values forecasted using the three models for training data of Jaspur HP station

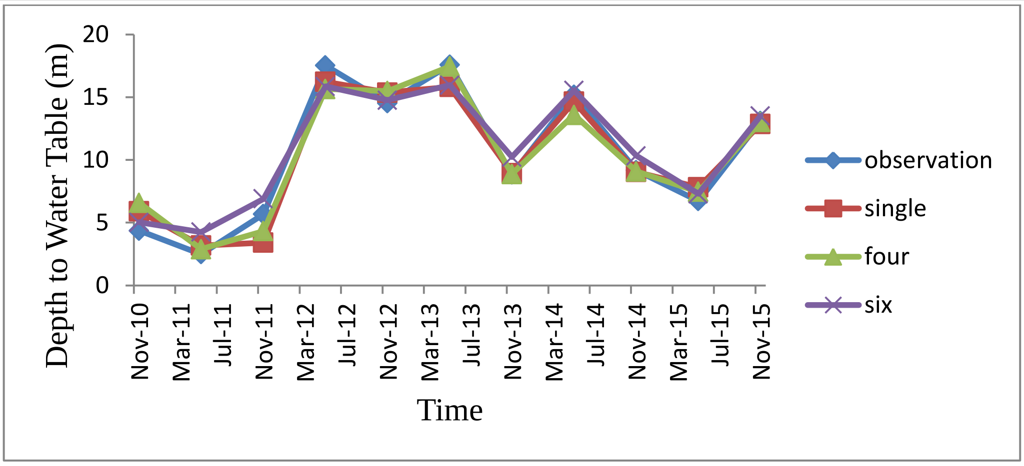


Fig. 12. Comparison of observed groundwater heads with values forecasted using the three models for testing data of Jaspur HP station

Kashipur DW

For the training data of the three models, the 2D feature mapobtained on a network of 8×8 cells is shown in Fig. 13. As shown in Fig. 13(a), the single-site SOM-RBFN model had nine neurons in the hidden layer. As the same reason the six-site model and the four-site model had five and seven neurons in the hidden layer, respectively. Fig. 14 shows the comparisons of observed groundwater heads with values forecasted using the single-site SOM-RBFN model, the six-site model, and the four-site model for training data. Fig. 15 shows the groundwater level forecasts using the three different models for testing data.

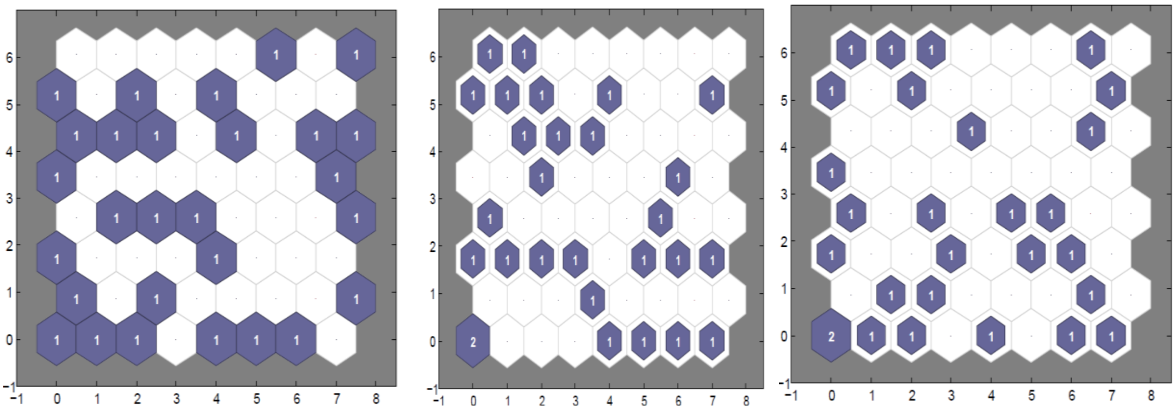


Fig. 13. 2D feature map for (a) single-site model; (b) six-site model; and (c) four-site model of Kashipur DW station.

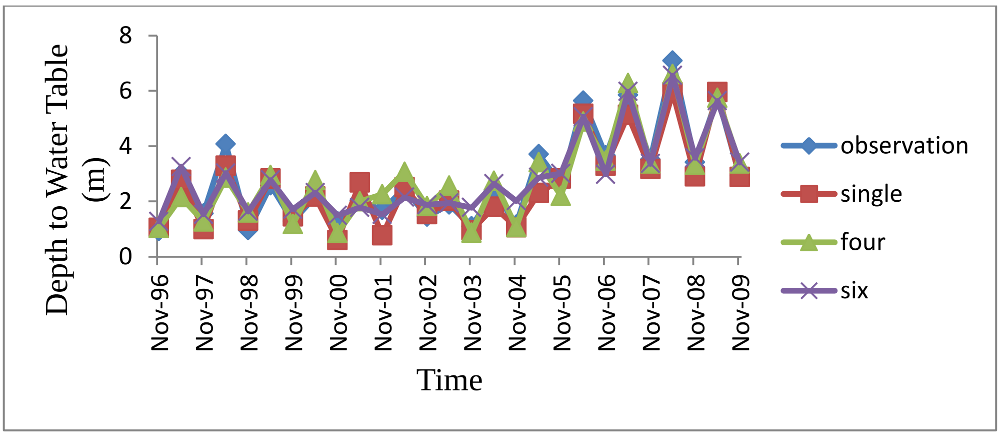


Fig. 14. Comparison of observed groundwater heads with values forecasted using the three models for training data of Kashipur DW station

Table 2. Performance of Single-Site Model, Six-Site Model, and Four-Site Model for Testing Data

Criterion						
	RMSE			CE		
Site	Single-site	Six-site	Four-site	Single-site	Six-site	Four-site
Surjan Nagar HP	0.324	0.257	0.254	0.899	0.935	0.919
Thakurdwara HP	0.161	0.148	0.160	0.939	0.942	0.938
Jaspur HP	1.168	1.143	1.144	0.938	0.929	0.940
Kashipur DW	0.372	0.299	0.306	0.903	0.920	0.914

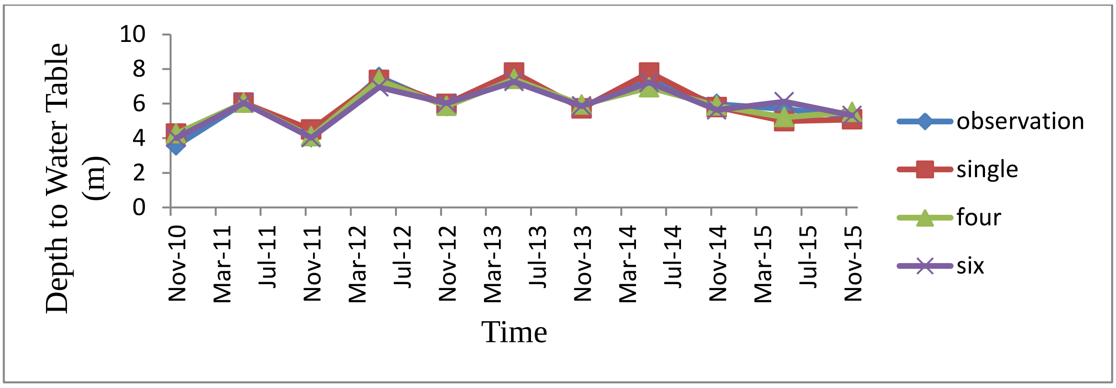


Fig. 15. Comparison of observed groundwater heads with values forecasted using the three models for testing data of Kashipur DW station

Evaluation of Prediction Accuracy

The performance of the single-site model, six-site model, and four-site model during training and testing is summarized in Tables 1 and 2 in terms of RMSE and CE. As shown in Table 1, one can see that the values of RMSE and CE for the six-site model both are better than those of the single-site model and four-site model, respectively. These results indicated that the six site model could increase the efficiency of the forecasting model. In same manner, the RMSE and CE from single-site model, six-site model and four-site model for testing data are summarized in Table 2. Again, the six-site model produces better performance than the single-site model and six-site model.

Summary and Conclusions

In this paper, a groundwater level forecasting model was developed. The model was based on the combination of RBFN and SOM. The SOM was used to construct the 2D feature map from which the number of clusters (i.e., the number of hidden units in the RBFN) could be figured out directly by eyes, and then the radial basis centers could be determined easily. In such a manner, the crucial problem for RBFN, i.e., the positioning of the radial basis centers could be solved. The proposed methodology was finally applied to Tumaria canal command area to find the forecasts of groundwater. The result showed that the six-site model could forecast groundwater level more accurately as compared to the single-site model and four-site model. That is, the SOM-RBFN multisite model could forecast more precisely than the single-site model.

REFERENCES

1. Adamowski, J., Chan, H.F., A wavelet neural network conjunction model for groundwater level forecasting". *J. Hydrol.* 407, 28e40 (2011).
2. Bracq, P., and Delay, F., "Transmissivity and morphological features in a chalk aquifer: A geostatistical approach of their relation." *J.Hydrol.*, **191:** 139–160 (1997).
3. Chen, Z. H., Stephen, E. G., and Kirk, G. O., "Predicting average annual groundwater levels from climatic variables: An empirical model." *J. Hydrol.*, **260:** 102–117 (2002).
4. Chen, L.H., Chen, C.T., and Pan, Y.G., "Groundwater Level Prediction Using SOM-RBFN Multisite Model" *J. Hydrol. Eng.*, **15(8):** 624-631 (2010).

5. Chen, L.H., Chen, C.T., and Lin, D.W., "Application of Integrated Back-Propagation Network and Self-Organizing Map for Groundwater Level Forecasting" *J. Water Resour. Plann. Manage.*,**137(4):** 352-365 (2011).
6. Han, J.C., Huang, G., Huang, Y., Zhang, H., Li, Z., Chen, Q., "Chance-constrained overland flow modeling for improving conceptual distributed hydrologic simulations based on scaling representation of sub-daily rainfall variability". *Sci.Total Environ*. 524e525, 8-22 (2015).
7. Han, J.C., Huang, Y., Li, Z., Zhao, C., Cheng, G., Huang, P., "Groundwater level prediction using a SOM-aided stepwise cluster inference model" *Journal of Environmental Management* **182(2016):** 308-321 (2016),
8. Nourani, V., Alami, M.T., Vousoughi, F.D., "Wavelet-entropy data preprocessing approach for ANN-based groundwater level modeling". J. Hydrol.524, 255e269 (2015).
9. Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., " Applied Modeling of HydrologicalTime Series". Water Resources Publications, Denver. (1990).
10. Savic, D. A., Walters, G. A., and Davidson, J. W., "A genetic programming approach to rainfall-runoff modeling." *Water Resour. Manage.*, **13(3):** 219–231 (1999).
11. Vandycke, S., Bergerat, F., and Depuis, C. H., "Meso-cenozoic faulting and inferred palaeostresses in the Mons Basin, Belgium." *Tectonophysics*, **192:** 261–271 (1991).