Forecasting groundwater level using artificial neural networks

P. D. Sreekanth[1,*], N. Geethanjali[2], P. D. Sreedevi[3], Shakeel Ahmed[3], N. Ravi Kumar[4] and P. D. Kamala Jayanthi[5]

[1]National Research Centre for Cashew, Puttur 574 202, India
[2]Sri Krishnadevaraya University, Anantapur 515 003, India
[3]National Geophysical Research Institute, Hyderabad 500 007, India
[4]Central Plantation Crops Research Institute, Kasaragod 671 124, India
[5]Indian Institute of Horticulture Research, Bangalore 560 089, India

The performance of the artificial neural network (ANN) model, i.e. standard feed-forward neural network trained with Levenberg–Marquardt algorithm, was examined for forecasting groundwater level at Maheshwaram watershed, Hyderabad, India. The model efficiency and accuracy were measured based on the root mean square error (RMSE) and regression coefficient (R^2). The model provided the best fit and the predicted trend followed the observed data closely (RMSE = 4.50 and R^2 = 0.93). Thus, for precise and accurate groundwater level forecasting, ANN appears to be a promising tool.

Keywords: Artificial neural networks, back-propagation, feed-forward, forecasting, groundwater level.

WATER is the elixir of life and is crucial for sustainable development. Earlier, it was considered to be a limitless or at least fully renewable natural resource, but during the past 20 years or so, there has been a tremendous pressure on this precious natural resource mainly due to rapid industrialization and human population, as an increase in the human population will simply result in increasing the demand for irrigation purpose to meet food production requirements. Though the advancement in agricultural technology has been impressive, in many regions poor irrigation management has resulted in considerable depletion of the groundwater table, damaged soils and deterioration in the water quality, making the availability of water in the future highly uncertain. Keeping in mind the scarcity of available water resources in the near future and it impending threats, it has become imperative on the part of water scientists as well as planners to quantify the available water resources for its judicial use. Thus, a ready reckoner to monitor the fluctuations in groundwater levels well in advance is the need of the hour to devise sustainable water management protocols.

In this direction several studies were carried out for forecasting the groundwater levels using conceptual/physical models that are not only laborious, but also have practical limitations[1], as many inter-related variables are involved. In the recent past, soft computing tools like artificial neural networks (ANNs) have been used increasingly in various fields of science and technology for prediction purposes[2]. In particular, ANNs have been found useful in the area of groundwater modelling.

The ANN is a general-purpose model with a limited set of variables, and is used as a universal functional approximator[3]. It can forecast many nonlinear time series events[4–7] over conventional simulation methods[8]. Basically, ANNs are intelligent systems that are related in some way to a simplified biological model of the human brain. They are composed of many simple elements called neurons operating in parallel and connected to each other in the forward path by some multipliers called connection weights. Usually, ANNs are trained by adjusting the values of these connection weights between the network elements. These networks have self-learning capability and are fault-tolerant as well as noise-immune, and have applications in various fields like forecasting, system identification, pattern recognition, classification, speech recognition, image processing, etc[9].

In this article, a reliable forecasting model for predicting the groundwater level using weather parameters through ANNs has been developed to have a precision forecasting with added accuracy over the current methods being practised.

Materials and methods

Study area

The input data for the present study were collected from the Maheshwaram watershed, which is situated in the Ranga Reddy District, Andhra Pradesh (AP) at a distance of about 35 km from Hyderabad, AP, India. The watershed has an area of 53 km^2 (Figure 1). The study area is situated between longitude 78°24′30″E and 78°29′00″E, latitude 17°06′20″N and 17°11′00″N and forms a part of the toposheet 56K/8. The topographic elevation is about 600–670 m amsl. Hot/dry summers and cool/dry winters

*For correspondence. (e-mail: pd_sreekanth@yahoo.com)

characterize the area, with a distinct rainy season from June to September. The temperature ranges from 22°C to 44°C, with an average rainfall of about 573 mm. The region receives more than 80% of its rainfall from the southwest monsoon. In general, the topography dips gently from south to north in the watershed.

The area comprises of granites of Archean age with thin soil cover of sandy loam and clay. These granites are medium-to-coarse grained, pink and grey in colour, and have undergone variable degree of weathering, ranging in depth from 15 to 20 m followed by fracturing which extends up to 20–50 m depending upon the local hydrogeology[10]. A few dolerite dykes and quartz veins also traverse the area. The area in general is undulating and a majority of the area has a slope of 2%.

Maheshwaram is a closed watershed with no major streams in the area. A network of first order and second order streams culminate in a large tank known as Mankal Cheruvu, which forms the discharge area. The weathered zone has become completely dry due to over exploitation. The existing wells tap the fractured bedrock, are in semi-confined situation[11]. In general, the water-striking level is around 25–30 m below ground level (bgl), whereas the water levels are at depths ranging from 15 to 23 m. The water-striking surface is always found to be at greater depth than the static water level, supporting that the aquifers are in semi-confined condition. The tapped groundwater (mostly through bore wells/submersible pumps) is used for irrigation where the discharges are in the range 100–300 m^3/day. The groundwater flow is from southwest to northeast (main drainage) and from south to north in other parts of the watershed.

Monthly water levels have been collected from 22 wells fairly distributed in the area[12] (Figure 1) during the study period 2000–06 (ref. 12). The monitoring wells are all bore wells and water levels were monitored using a graded tape that provides sound and light signals when it touches water in the well, with an accuracy of 2 mm. Care was taken to record the water level in all the wells in the minimum possible time and also when the wells were not being pumped (as the water level could reach its natural condition). Data on weather parameters, viz. evaporation, rainfall, relative humidity and temperature (minimum and maximum) were collected from the local hydro-metereological station established at the centre of the Maheshwaram watershed by the Andhra Pradesh Ground Water Department.

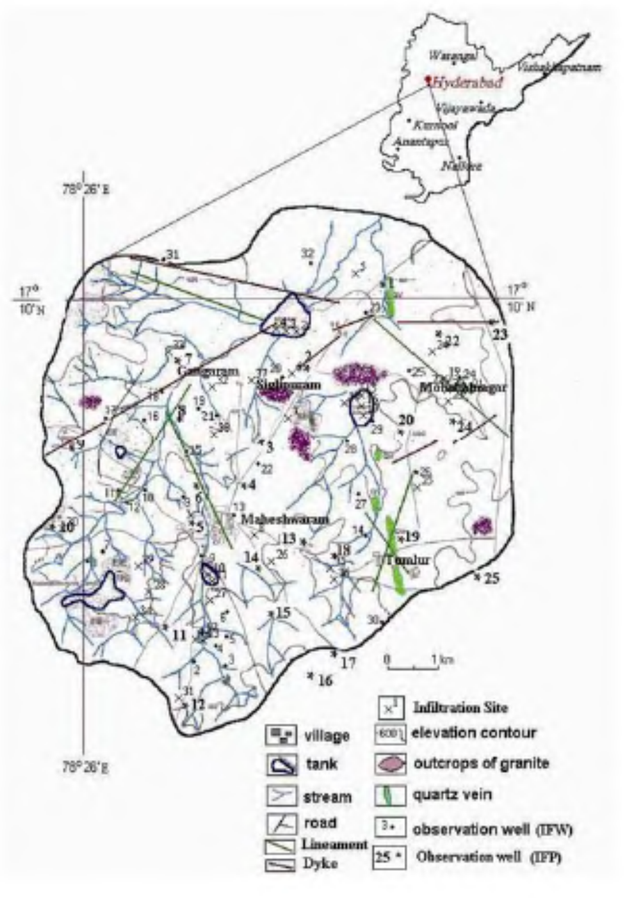

Figure 1. Location of the study area (Maheshwaram watershed).

ANN architecture

Feed-forward neural network (FFNN) along with Levenberg–Marquardt back propagation (LMB) algorithm was used with programing in Matlab 7.0.

FFNN: Here, the source nodes in the input layer of the network, supply the respective elements of the activation pattern, that constitute the input signals to the neurons in the second layer. The output signals of the second layer are used as inputs in the third layer, and so on for the rest of the network. Thus, typically the neurons in each layer of the network have their inputs from the output signals of the preceding layer only. The set of output signals of the neurons in the output layer of the network constitutes the overall response of the network supplied by the source nodes in the input layer[13] as shown in Figure 2, which explains a typical FFNN with one hidden layer.

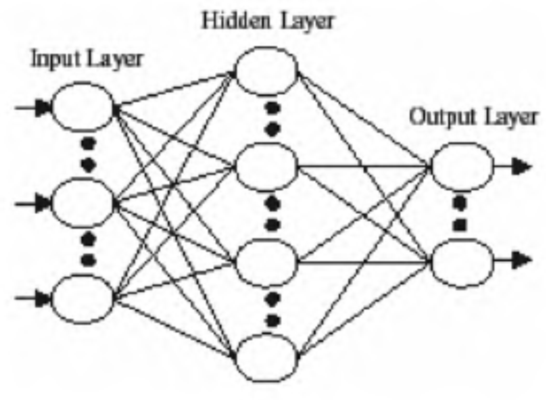

Figure 2. Typical feed-forward neural network.

LMB: The LMB algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (is typical in training FFNNs), then the Hessian matrix can be approximated as

$$H = J^T J,$$

and the gradient can be computed as

$$g = J^T \boldsymbol{e},$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and $\boldsymbol{e}$ is a vector of network errors. The Jacobian matrix can be computed through a standard back-propagation technique that is much less complex than computing the Hessian matrix. Therefore, the Levenberg–Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - (J^T J + \mu I)^{-1} J^T \boldsymbol{e}.$$

When the scalar μ is zero, this is Newton's method using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum. So the aim is to shift towards Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm.

Training and testing the data

The total identified nodes include the monthly observed water levels in 22 sampled well points along with local weather parameters, viz. rainfall, temperature (minimum and maximum), evaporation and relative humidity that have a direct influence on groundwater levels. Here, the input layer contains various layer nodes as given in Table 1. In order to obtain good performance of the ANN, tuning of the ANN architecture and parameters is indispensable. Hence the ANN architecture was tested with various numbers of hidden layers and nodes per hidden layer to find better values and architecture. Thus, the whole dataset was arbitrarily grouped into four different sets, each with the randomly chosen respective well points for training and testing of the selected ANN model (Table 1).

Table 1. Feed-forward neural network–Levenberg–Marquardt back propagation (FFNN–LMB) structures

Data set	ANN structure	Observations (well no.)
Set 1	12–20–5	1, 6, 11, 18, 20
Set 2	15–25–8	2, 5, 7, 9, 13, 15, 16, 22
Set 3	16–30–9	3, 4, 8, 10, 12, 14, 17, 19, 21
Set 4	29–40–22	1 to 22

Table 2. FFNN–LMB performance

Parameter	Set 1	Set 2	Set 3	Set 4
RMSE	4.50	6.97	3.13	4.50
R^2	0.92	0.84	0.96	0.93
Processing time (s)	3.06	13	356	1105
Epochs	12	181	245	56

Table 3. Lower and upper error variation in the estimated groundwater levels

Well no.	Error variation (%) Lower side	Upper side
1	–1.24	1.91
2	–2.67	1.47
3	–1.06	1.11
4	–2.52	1.30
5	–1.86	1.89
6	–1.03	0.64
7	–1.46	1.08
8	–1.06	0.96
9	–1.47	0.51
10	–1.60	0.90
11	–0.91	0.70
12	–0.81	0.71
13	–5.09	4.06
14	–0.62	0.58
15	–1.19	1.77
16	–1.28	2.20
17	–0.90	1.92
18	–1.21	1.00
19	–0.83	1.51
20	–2.08	1.81
21	–0.04	0.91
22	–1.11	2.65

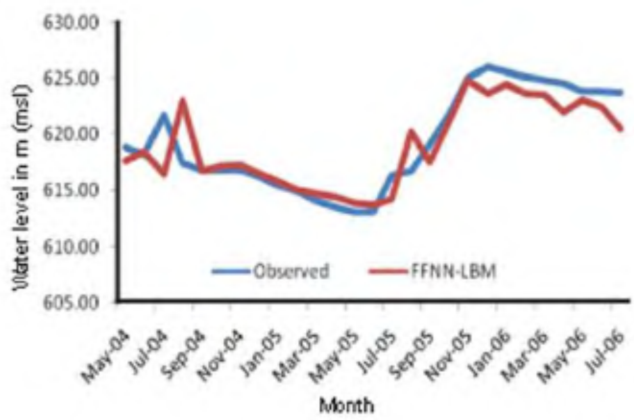

Figure 3. Overall mean trend showing observed/estimated water levels using the FFNN–LMB model.

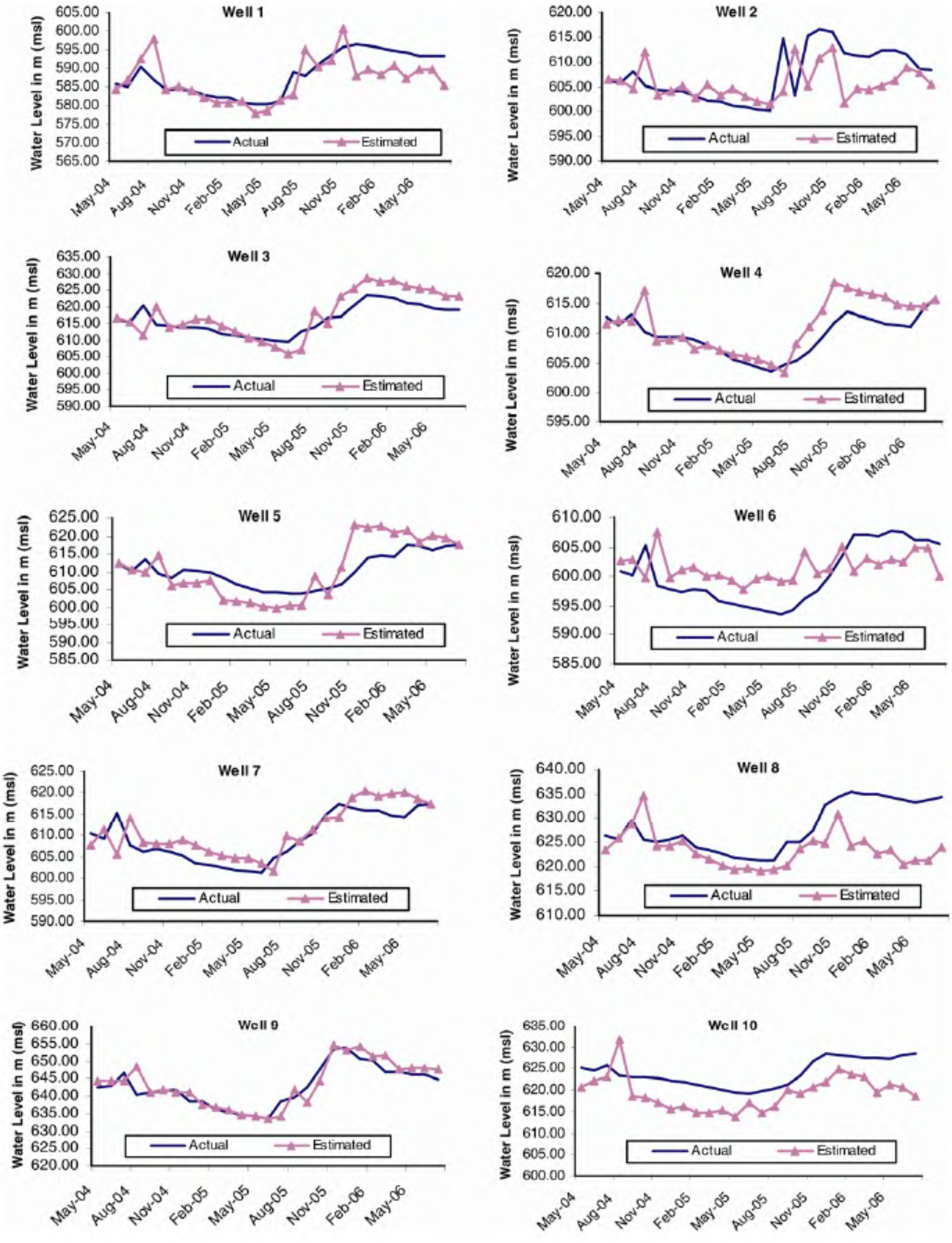

Figure 4.

(*Contd...*)

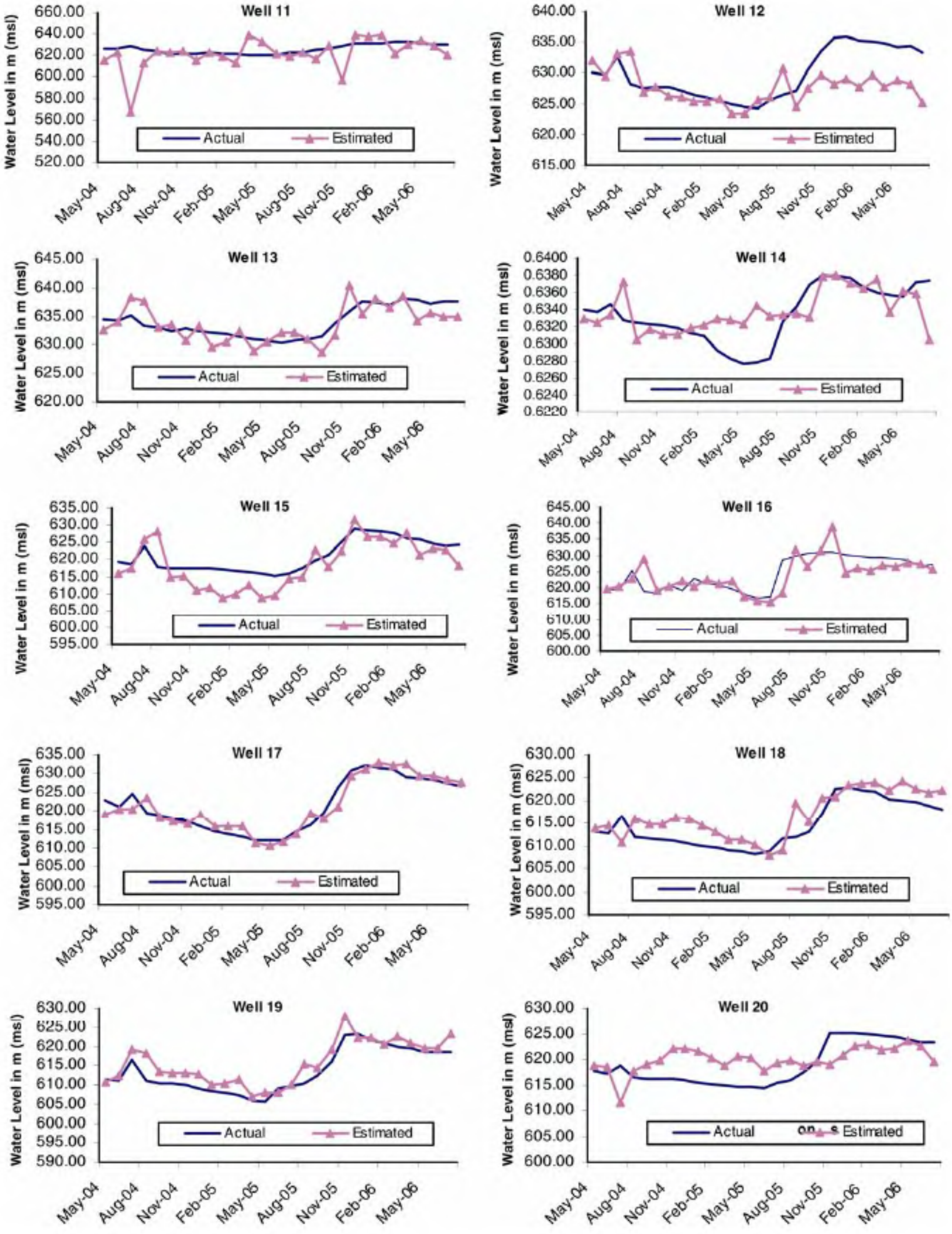

Figure 4.

(*Contd...*)

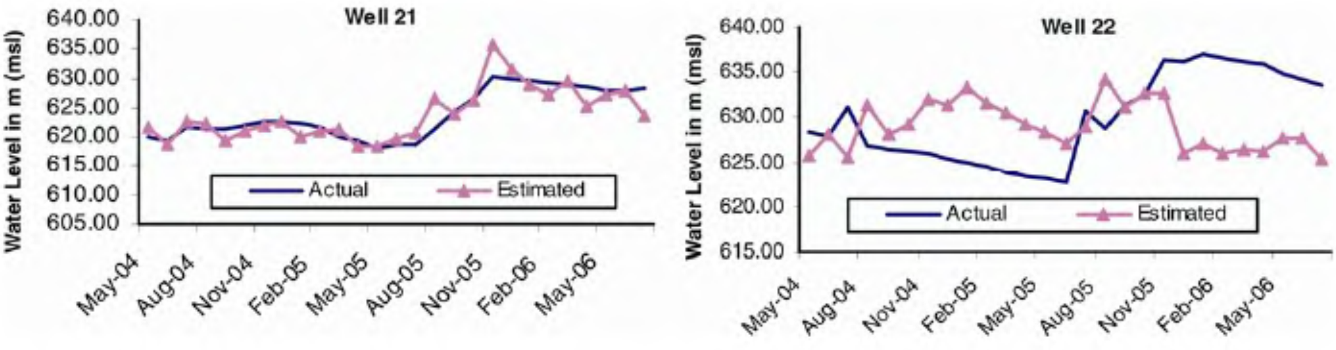

Figure 4. Performance trend of the model.

In the case of set 1, the selected ANN structure was 12–20–5 (i.e. 12 input layer nodes, 20 hidden layer nodes and 5 output layer nodes). Similarly, for sets 2–4, the selected ANN structures were 15–25–8, 16–30–9 and 29–40–22 respectively (Table 1). Mapping for the hidden layer/respective nodes was carried out based on trial and error method, as there is no standard methodology for selecting the same[14]. The applied transfer functions are linear and log sigmoid for all the above structures.

The performance of a trained network can be measured to some extent by the errors on the training, validation and test sets, but it is often useful to investigate the network response in detail using statistical parameters. Therefore, the efficiency/response of the selected network (in different sets) for accurate output was measured using statistical indices, viz. error variation (EV), root mean square error (RMSE) and regression coefficient (R^2) and calculated based on the corresponding measured data, according to eqs (1)–(3).

$$\mathrm{EV} = ((y - \hat{y})/y) * 100, \tag{1}$$

where y is the observed data and $\hat{y}$ the calculated data.

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2 / n}, \tag{2}$$

$$R^2 = 1 - \frac{\sum (y_i - \overline{y}_i)^2}{\sum y_i^2 - \frac{\sum \overline{y}_i^2}{n}}, \tag{3}$$

where y_i is the observed data, $\overline{y}_i$ the calculated data and n the number of observations.

RMSE indicates the discrepancy between the observed and calculated values. The lower the RMSE, the more accurate is the prediction. The best fit between observed and calculated values[1], which is unlikely to occur, would have R^2 as 1 and RMSE as 0.

Results and discussion

The training performance of the FFNN–LMB for different datasets is given in Table 2. Perusal of the data showed that the minimum RMSE using FFNN–LMB was observed with set 3 (3.13), which has 16–30–9 network structures. However, for sets 1, 2 and 4, the respective RMSE was found to be 4.50, 6.97 and 4.50. The processing time was more for set 4 (1105 s) followed by sets 3, 2 and 1 (356, 13 and 3 s respectively). However, epoch number was found to be more in the case of set 3 (245) followed by sets 2, 4 and 1 (181, 56 and 12 respectively). R^2 was found to be 0.92, 0.84 and 0.93 for sets 1, 2 and 4 respectively. The maximum variability explained through R^2 was found to be 0.96 (in case of set 3). The EV between the calculated and observed data for different sets is given in Table 3. Analysis of data in randomized sets clearly showed that FFNN–LMB is the best-fit ANN model for predicting the groundwater level in terms of statistical significance (EV, RMSE and R^2) as well as processing flexibility (processing time and epochs) as FFNN–LMB recorded lower EV, RMSE, processing time and epochs.

Further, the data were analysed separately for each independent well point to have a clear comparison of the mean observed and estimated water levels. Here also, the results exhibited similar trend as in the case of set-wise data analysis, i.e. FFNN–LMB was found to be the best fit for predicting groundwater levels at the Maheshwaram watershed. This is clearly shown in Figure 3, where the overall mean actual water level has been compared with the predicted water level. Here, the predicted water level trend followed the observed trend closely, showing the accuracy of the network.

A number of studies have indicated that ANN can produce generalized models of environmental systems with greater accuracy than conventional statistical techniques[15–17]. Recent literature reviews reveal that neural networks, specifically the FFNNs, have been successfully used for water resource modelling and prediction[18,19]. The performance trend of the model, individually for each of the 22 wells along with actual and predicted values is indicated in Figure 4, showing that the difference between actual and predicted values is close to zero.

In this study, a better forecasting model using ANNs has been developed for predicting monthly groundwater level fluctuations in the Maheshwaram watershed. The most suitable configuration for this proved to be the FFNN–LMB method, as it showed the most accurate prediction, and the overall accuracy of this model is around 93% (Figure 3). Further, a significant advantage of this model is that it can provide satisfactory predictions with limited groundwater level records also. Earlier also, many researchers have proved that ANN models are the best tools for predicting groundwater levels[20–24]. Future research efforts should be envisaged towards exploring the use of soft computing tools for predicting groundwater levels with more accuracy and stability over conventional methods.

1. Ioannis, N., Daliakopoulos, Coulibaly, P., Ioannis, K. and Tsanis, Groundwater level forecasting using artificial neural networks. *J. Hydrol.*, 2005, **309**, 229–240.
2. Gail, M., Brion, T. R., Neelakantan and Lingireddy, S., A neural-network-based classification scheme for sorting sources and ages of fecal contamination in water. *Water Res.*, 2002, **36**, 3765–3774.
3. Hornik, K., Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators. *Neural Networks*, 1989, **2**, 359–366.
4. Guan, P., Huang, D. and Zhou, B., Forecasting model for the incidence of hepatitis A based on artificial neural network. *World J. Gastroenterol.*, 2004, **10**, 3579–3582.
5. Hill, T., O'Connor, M. and Remus, W., Neural network models for time series forecasts. *Manage. Sci.*, 1996, **42**, 1082–1092.
6. Tang, Z. and Fishwick, P. A., Feedforward neural nets as models for time series forecasting. *ORSA J. Comput.*, 1993, **5**, 374–385.
7. Zhang, G., Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 2003, **50**, 159–175.
8. French, M. N., Krajewski, W. F. and Cuykendall, R. R., Rainfall forecasting in space and time using a neural network. *J. Hydrol.*, 1992, **137**, 1–31.
9. Ramalingam, V., Palaniappan, B., Panchanatham, N. and Palanivel, S., Measuring advertisement effectiveness – a neural network approach. *Expert Syst. Appl.*, 2006, **31**, 159–163.
10. Raoof Hashimi, S. A. and Engerrand, C., Status report for Maheshwaram watershed, Andhra Pradesh, India. NGRI Technical Report No. APGWD, 1999.
11. Subrahmanyam, K., Ahmed, S. and Dhar, R. L., Geological and hydro geological investigations in the Maheswaram watershed, AP, India. Technical Report No. NGRI-2000-GW-292, 2000, p. 16.
12. Ahmed, S., Engerrand, C., Sreedevi, P. D., Kumar, D., Subrahmanyam, K., Ledoux, E. and Marsily, G. de, Geostatistics, aquifer modeling and artificial recharge. Scientific report: vol. 3, Indo-French collaborative project (2013-1), Technical Report No. NGRI-2003-GW-411, 2003.
13. Haykin, S., *Neural Networks, A Comprehensive Foundation*, Pearson Education, 2006, 2nd edn.
14. Altunkaynak, A., Forecasting surface water level fluctuations of lake van by artificial neural networks. *Water Resour. Manage*, 2007, **21**, 399–408.
15. Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S., Application of neural networks to modelling nonlinear relationships in ecology. *Ecol. Model*, 1996, **90**, 39–52.
16. Comrie, A. C., Comparing neural networks and regression models for ozone forecasting. *J. Air Waste Manage. Assoc.*, 1997, **47**, 653–663.
17. Paruelo, J. M. and Tomasel, F., Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecol. Model*, 1997, **98**, 173–186.
18. Coulibaly, P., Anctil, F. and Bobée, B., Hydrological forecasting using artificial neural networks: the state of the art (in French). *Can. J. Civ. Eng.*, 1999, **26**, 293–304.
19. Maier, H. R. and Dandy, G. C., Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Model. Software*, 2000, **15**, 101–124.
20. Holger, R., Maier, Graeme, C. and Dandy, Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Modell. Software*, 2000, **15**, 101–124.
21. Coulibaly, P., Anctil, F., Aravena, R. and Bobee, B., Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.*, 2001, **37**, 885–896.
22. Mohammadi, K., Eslami, H. R. and Qaderi, K., Water table estimation using Modflow and artificial neural network. *Geophys. Res. Abstr.*, 2005, **7**, 751.
23. Paul, A., Conrads, Edwin A. and Roehl, Estimating water depths using artificial neural networks. In Seventh International Conference on Hydro Informatics, HIC 2006, Nice, France.
24. Emery Coppola Jr, Szidarovszky, F., Poulton, M. and Charles, E., Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions. *J. Hydrol. Eng., ASCE*, 2003, 348–360.

ACKNOWLEDGEMENTS. We are grateful to the Director, NGRI, Hyderabad for the encouragement and permission to publish this work. Senior author thanks the Indian Council of Agricultural Research, New Delhi, for sanctioning study leave. The third author acknowledges the Council of Scientific and Industrial Research, New Delhi for financial assistance. We thank the anonymous reviewers for valuable comments to improve this manuscript.

Received 7 July 2008; revised accepted 2 March 2009