Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment

Sasmita Sahoo · Madan K. Jha

Abstract The potential of multiple linear regression (MLR) and artificial neural network (ANN) techniques in predicting transient water levels over a groundwater basin were compared. MLR and ANN modeling was carried out at 17 sites in Japan, considering all significant inputs: rainfall, ambient temperature, river stage, 11 seasonal dummy variables, and influential lags of rainfall, ambient temperature, river stage and groundwater level. Seventeen site-specific ANN models were developed, using multi-layer feed-forward neural networks trained with Levenberg-Marquardt backpropagation algorithms. The performance of the models was evaluated using statistical and graphical indicators. Comparison of the goodness-of-fit statistics of the MLR models with those of the ANN models indicated that there is better agreement between the ANN-predicted groundwater levels and the observed groundwater levels at all the sites, compared to the MLR. This finding was supported by the graphical indicators and the residual analysis. Thus, it is concluded that the ANN technique is superior to the MLR technique in predicting spatio-temporal distribution of groundwater levels in a basin. However, considering the practical advantages of the MLR technique, it is recommended as an alternative and cost-effective groundwater modeling tool.

Keywords Groundwater-level prediction · Multiple linear regression · Artificial neural network · Statistical modeling · Japan

Received: 14 September 2012 / Accepted: 22 July 2013
Published online: 4 October 2013

S. Sahoo (✉) · M. K. Jha
AgFE Department, Indian Institute of Technology Kharagpur, Kharagpur – 721 302, West Bengal, India
e-mail: sasmitaiit@gmail.com
Tel.: +91-9749935147
Fax: +91-3222-282244

M. K. Jha
e–mail: madan@agfe.iitkgp.ernet.in

Introduction

The accurate prediction of groundwater levels is essential for sustainable utilization and management of vital groundwater resources. Since groundwater is hidden and groundwater processes exhibit a high degree of temporal and spatial variability, modeling groundwater fluctuations is a very difficult task. There exist conceptual and process-based modeling techniques for simulating groundwater flow in a variety of hydrogeological settings. However, the data requirements for the process-based models used to simulate groundwater fluctuations are enormous, and generally difficult or expensive to obtain (Coulibaly et al. 2001; Nikolos et al. 2008). Despite large investments in time and resources, the prediction accuracy possible with distributed numerical flow models has not improved satisfactorily for many types of water-management problems (Coppola et al. 2005). Therefore, a dynamically predictive model that can tackle the persistent trend and time-variant behavior of hydrological variables is desirable for the efficient planning and management of water resources. Under such circumstances, empirical models such as regression models and artificial neural network (ANN) models serve as attractive alternatives, because they can provide useful results using relatively fewer data, and are less laborious and therefore cost-effective. Despite the inability of multiple linear regression (MLR) models to cope with the non-linearity existing between model inputs and outputs, they have been used in many hydrological studies, possibly due to the fact that the results are quite easy to use and the interpretation of the relationship between the parameters is easier (Heuvelmans et al. 2006; Adeloye 2009). On the other hand, the ANN technique has been found to be very much suited to the modeling of non-linear and dynamic systems such as water-resources systems (ASCE 2000b; Maier and Dandy 2000). The main advantage of the ANN technique over traditional methods is that it does not require the complex nature of underlying processes to be explicitly described in mathematical form. After proper training, ANN models can yield satisfactory results for many prediction problems in the field of hydrology (ASCE 2000a, b).

In the last decade or so, tools such as ANN and statistical techniques such as MLR have attracted the attention of some

 DOI 10.1007/s10040-013-1029-5

hydrologists and hydrogeologists for the purposes of prediction/forecasting, due to their parsimony in data requirement, simplicity and cost-effectiveness. Although several studies are reported in the literature that use MLR as a modeling technique in the field of surface-water hydrology (e.g., McCuen et al. 1979; Sinnakaudan et al. 2006), the use of the MLR technique in groundwater modeling is very limited. Hodgson (1978) used MLR for the simulation of groundwater-level responses in the Vryburg aquifer of South Africa by considering rainfall and pumping as input parameters.

A comprehensive review of the application of ANN to hydrology can be found in the ASCE Task Committee reports (ASCE 2000a, b) and in Maier and Dandy (2000). In the recent past, several researchers have successfully used ANN for the prediction of groundwater levels in unconfined aquifers (e.g., Coulibaly et al. 2001; Lallahem et al. 2005; Daliakopoulos et al. 2005; Nayak et al. 2006; Affandi et al. 2007; Krishna et al. 2008; Trichakis et al. 2009; Sethi et al. 2010). Coulibaly et al. (2001) developed four types of ANN models, namely: input-delay neural network (IDNN), recurrent neural network (RNN), generalized radial-basis function (RBF) network, and probabilistic neural network (PNN) to simulate groundwater-level fluctuations in four observation wells in the Gondo aquifer, Burkina Faso, Africa. Monthly water-level depth, precipitation, temperature and river-water level were used as inputs to the networks, and it was found that the generalized RBF network is not suitable for deep groundwater-level modeling, whereas IDNN and PNN are effective for predictions up to 2 months ahead. Lallahem et al. (2005) developed a multi-layer feed-forward network, trained with a standard back-propagation algorithm, for estimating groundwater levels in 13 piezometers installed in the unconfined chalky aquifer of northern France, using monthly rainfall, temperature and potential evapotranspiration as inputs. Daliakopoulos et al. (2005) and Affandi et al. (2007) compared different types of back-propagation algorithms for predicting water-level fluctuations, and found that the performance of the ANN model improved from the gradient-descent algorithm to the Levenberg-Marquardt algorithm, but decreased in the case of the RBF algorithm. Nayak et al. (2006) developed multi-layer feed-forward networks, trained with standard back-propagation algorithms, for forecasting groundwater-level fluctuations at two sites in a shallow unconfined aquifer of the Central Godavari Delta System, South India, using groundwater levels of different lags, rainfall and canal releases as inputs to the model. Krishna et al. (2008) used feed-forward neural networks and RBF networks, with Levenberg-Marquardt and Bayesian Regularization training algorithms, to predict water levels 1 month ahead in six wells installed in an unconfined aquifer in Andhra Pradesh, South India. Monthly rainfall, temperature and evapotranspiration were used as inputs, and the feed-forward neural network with Levenberg-Marquardt algorithm was reported to be suitable for predicting water levels in the six wells. Trichakis et al. (2009) developed an ANN model using a multi-layer perception network, trained with an error back-propagation algorithm, to predict water levels in two wells installed in an unconfined karstic aquifer in Mavrosouvala, Greece. The inputs to the ANN model consisted of daily temperature, rainfall, pumping rate and hydraulic head. A differential-evolution algorithm was used to optimally define the time lag in the rainfall measurements, as well as the ANN architecture and training parameters. Sethi et al. (2010) developed multi-layer feed-forward networks, trained with back-propagation algorithms using gradient-descent with momentum, to predict groundwater depths in 64 dug wells located in the Munijhara micro-watershed in Orissa, eastern India. Data on monthly rainfall, potential evapotranspiration, and groundwater depth (from nearby wells) were used as inputs to the model, with the output being groundwater depths, 1 month ahead. Similarly, several researchers have used ANN for predicting groundwater levels in confined aquifers (Hani et al. 2006; Uddameri 2007), in leaky aquifers (Coppola et al. 2003, 2005; Mohanty et al. 2009), and in multi-layer aquifer systems (Coppola et al. 2003; Feng et al. 2008; Nourani et al. 2008).

As far as the combined application of statistical and ANN techniques to the simulation of groundwater levels is concerned, only two studies are reported to date (Uddameri 2007; Izady et al. 2012). Uddameri (2007) applied MLR and ANN techniques to forecast monthly and quarterly time-series of piezometric levels in one deep well in South Texas, by using bias, time and dummy variables as inputs to the models. In this study, no real-world data were used for developing the ANN and MLR models, and the study concluded that the ANN model is more suitable than the MLR model for forecasting piezometric levels at the selected well. In a recent study, Izady et al. (2012) investigated the performance of fixed-effects and random-effects panel-data regression models for predicting groundwater levels in six representative wells in the Neishaboor Plain, Iran, using monthly precipitation and evapotranspiration as inputs to the models. The panel-data model results were compared with those of the ANN models and the former performed better than the latter.

The aforementioned review of literature reveals that a multi-layer feed-forward neural network, trained with the Levenberg-Marquardt algorithm, is more efficient than other types of ANN (e.g., traditional gradient-descent ANN and RBF ANN). The number and types of input, and the number of sites used for predicting groundwater levels, vary considerably from one study to another. Furthermore, the review reveals that in all likelihood, no studies have been conducted to date comparing the predictive capabilities of MLR and ANN techniques in simulating groundwater levels using real-world data. Therefore, the focus of this study was on evaluating the efficacy of two data-driven approaches such as MLR and ANN, for predicting the spatio-temporal distribution of water-levels in a groundwater basin, using relevant real-world data. In the present study, standard protocols of MLR and ANN modeling have been strictly followed and all the significant real-world data have been considered as inputs to the models. Thus, this study demonstrates a scientifically sound methodology for the evaluation of two data-driven approaches (modeling tools) in simulating groundwater levels using real-world data.

 DOI 10.1007/s10040-013-1029-5

Methodology

Overview of study area

The study area selected for the present study is known as the Konan groundwater basin, which is located in Kochi Prefecture of Shikoku Island, Japan (Fig. 1). The Konan groundwater basin is bounded by the Monobe River (perennial) in the west and the Koso River (intermittent) in the east. Mountains demarcate the northern boundary and the southern boundary is demarcated by the Pacific Ocean. There are two further intermittent rivers called the Karasu River and the Yamakita River. Land use mainly comprises agricultural land (paddy fields, greenhouses and fisheries), industry and built-up areas. The study area encompasses a total geographical area of 2,200 ha, of which about 1,502.5 ha is paddy fields, 488.0 ha is upland, and 186.5 ha is under greenhouse cultivation.

Cold dry winters, and warm humid summers characterize the regional climate. The average daily maximum temperature is 37 °C during summer, and the average daily minimum temperature is −4 °C during winter. The mean annual rainfall and evapotranspiration in the study area are about 2,600 mm and 800 mm, respectively. Unconfined aquifers consisting of alluvial sand and gravel and/or diluvial silty sand and gravel are predominant over the Konan groundwater basin (Jha et al. 1999). The unconfined aquifer system encompasses the area between the Koso River in the east and the Monobe River in the west (Fig. 1). The hydraulic conductivity of these aquifers varies from 65 to 804 m/day (Jha et al. 1999).

Study sites and data sources

In the present study, 17 sites (A-2, B-3, C-2, C-7, D-6, E-2, E-4, E-5, F-1, F-6, G-2, GH-4.5, H-2, H-3, H-4, H-5 and I-2) over the Konan groundwater basin of Kochi Prefecture, Japan (Fig. 1) were selected, based on the availability and continuity of groundwater-level time-series data. The selected sites (encircled observation wells in Fig. 1) are located in the unconfined aquifer of the basin, and are more-or-less representative of the unconfined aquifer underlying the Konan basin. 1999). Further details of the Konan groundwater basin can be found in Jha et al. (1999).

Daily groundwater-level data for a 6-year period (1999 to 2004) for the 17 sites were collected from Kochi Prefectural Office, Kochi City, Japan. The daily rainfall, and maximum and minimum ambient temperature data for the 1999–2004 period were obtained from the Gomen meteorological station, situated in Nankoku-shi, Kochi Prefecture, about 7 km west of the Monobe River. Besides these data, daily river-stage data for the Monobe River at the Fukabuchi stream-gauging station (Fig. 1) for the 1999–2004 period were also collected from Kochi Work Office, Ministry of Construction, Japan.

Selection of model inputs

One of the most important steps in the development of MLR and ANN models is the selection of influential (significant) input variables. Generally, all of the potential input variables are not equally informative, because of the fact that some variables may be correlated, noisy, or have no significant relationship with the output to be modeled (Maier and Dandy 1998). In order to study the influence of rainfall, temperature, and river stage on groundwater levels at the 17 selected sites, cross-correlation analyses were performed. The values of the correlation coefficients are summarized in Table 1, which shows that rainfall, temperature and river stage have good correlation with the groundwater levels of all 17 sites. Similarly, the effect of the previous month's groundwater levels on the current month's groundwater levels was analyzed by autocorrelation and partial-autocorrelation techniques (Fig. 2) with the help of STATISTICA 6.0 software (Statistica 2001). Apart from the three hydro-meteorological variables, 11 seasonal dummy variables (D_1, D_2, D_3, …D_{11}) were also considered, in order to capture the seasonal fluctuation of groundwater levels in the monthly datasets. Each of these dummy variables is equivalent to a new explanatory variable, and 11 of them (i.e., 12–1) were used to denote monthly periods, as per the rule in Makridakis et al. (2008). These 11 monthly variables are defined as: $D_1=1$, if the month is January and the value of D_1 is zero for the remaining months; $D_2=1$, if the month is February and zero otherwise and so on, with the value of $D_{11}=1$ for November and zero for the remaining months. In the regression equations, the regression coefficients associated with these variables reflect the average difference in the groundwater-level between those months and the omitted month (i.e., December), which is considered as a base period. Therefore, the coefficient associated with D_1 is a measure of the effect of January–February change in groundwater-level at a site compared to December. Similarly, D_2 refers to the effect of February–March change in groundwater-level at a site, D_3 refers to the effect of March–April change in groundwater-level, and so on, with D_{11} representing the effect of November–December change in groundwater-level. Thus, a total of 22 model inputs were selected for each of the 17 sites, which are: monthly rainfall, monthly mean ambient temperature, monthly mean river stage, 11 seasonal dummy variables, and significant lags (1-month and 2-month lags) of rainfall, ambient temperature, river stage and groundwater-level.

MLR modeling

MLR model

The MLR technique models the relationship between two or more explanatory (independent) variables and a response (dependent) variable by fitting a linear equation to the observed data. The generic form of an MLR model is as follows (Makridakis et al. 2008):

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots\cdots\cdots + \beta_k X_{k,i} + \varepsilon_i \quad (1)$$

where Y_i represents the ith observation of the dependent variable Y; $X_{1,\,i}$, $X_{2,\,i}$, ……., $X_{k,\,i}$ represent the ith

DOI 10.1007/s10040-013-1029-5

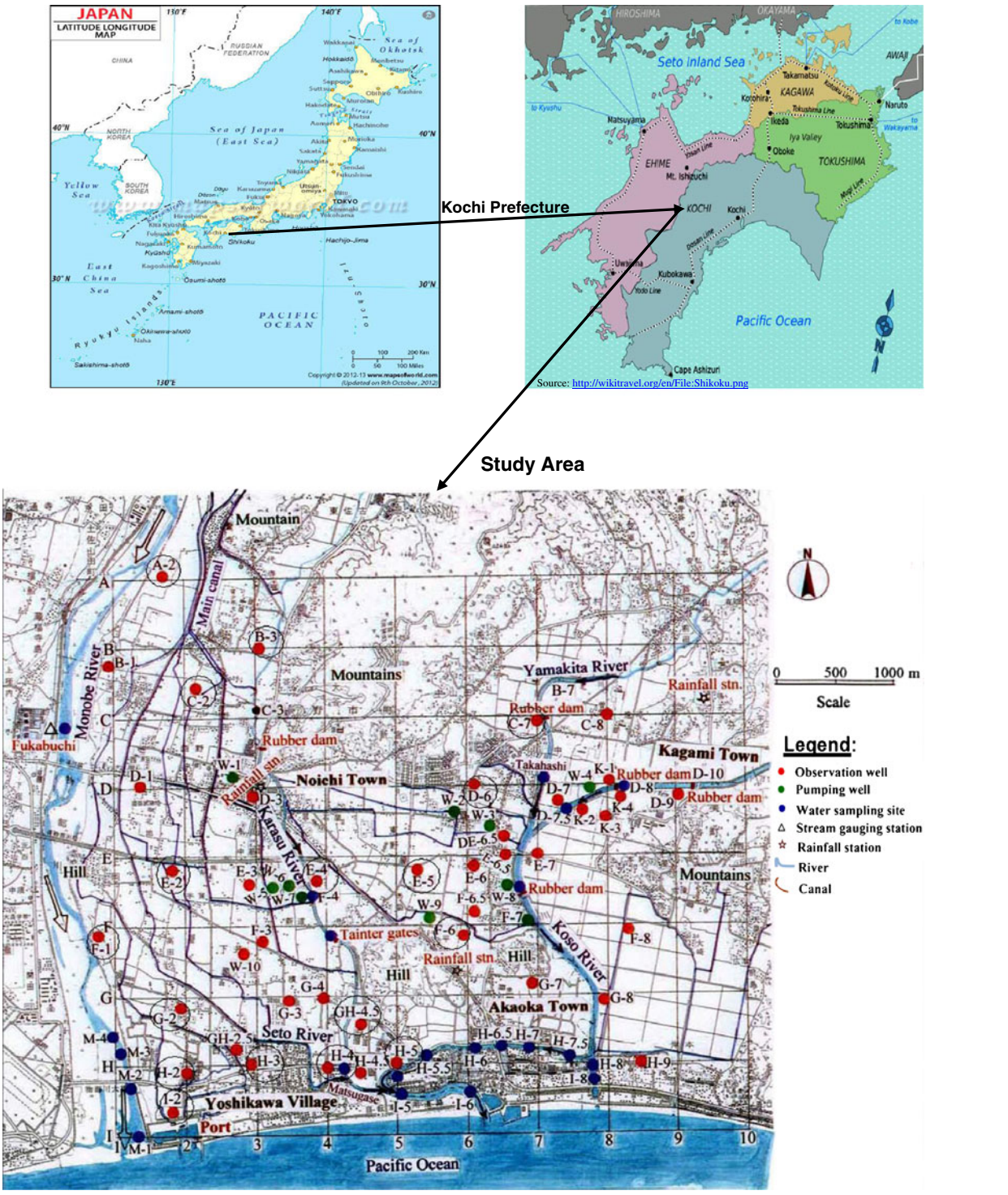

Fig. 1 Location of the study area and observed sites in the Konan groundwater basin

observations of each of the independent variables X_1, X_2,…, X_k respectively; β_0, β_1, $\beta_{2.....}$, β_k are fixed but unknown parameters; and ε_i is a random variable that is normally distributed.

The task of MLR modeling is to estimate the unknown parameters (β_0, β_1, $\beta_{2.....}$, β_k) of an MLR model (Eq. 1). Hence, the pragmatic form of the statistical regression model obtained after applying the least-square technique is given as (Makridakis et al. 2008):

$$Y_i = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + \cdots\cdots\cdots + b_k X_{k,i} + e_i \quad (2)$$

 DOI 10.1007/s10040-013-1029-5

Table 1 Cross-correlation between water levels in selected observation wells and three independent variables (rainfall, temperature and river stage)

Observation well	Values of correlation coefficient (r)		
	Rainfall	Temperature	River stage
A-2	0.818	0.881	0.904
B-3	0.801	0.687[a]	0.744
C-2	0.651[a]	0.881	0.694[a]
C-7	−0.330[b]	−0.108[b]	−0.333[b]
D-6	0.721	0.875	0.797
E-2	0.679[a]	0.888	0.757
E-4	0.676[a]	0.917	0.739
E-5	0.682[a]	0.937	0.765
F-1	0.821	0.866	0.918
F-6	0.700	0.905	0.766
G-2	0.713	0.921	0.801
GH-4.5	0.593[a]	0.771	0.687[a]
H-2	0.666[a]	0.874	0.778
H-3	0.664[a]	0.902	0.770
H-4	0.635[a]	0.854	0.719
H-5	0.558[a]	0.688[a]	0.692[a]
I-2	0.664[a]	0.867	0.793

Strongly correlated ($r>0.7$)

[a] Moderately correlated ($0.5\leq r\leq 0.7$)

[b] Poorly correlated ($r<0.5$)

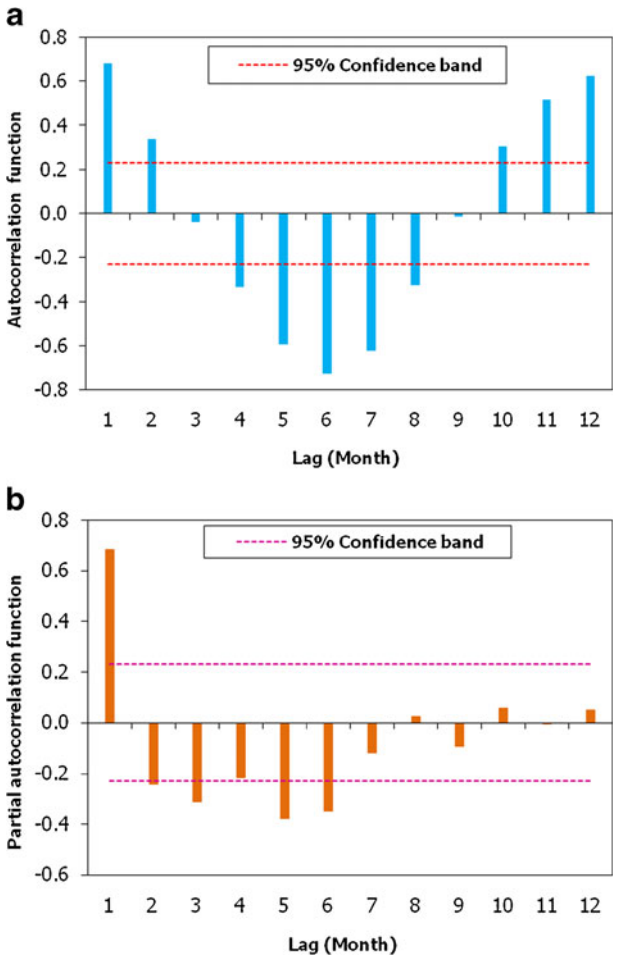

Fig. 2 Correlograms for water levels at site A-2 for various time lags, showing: **a** autocorrelation function; **b** partial-autocorrelation function

where i=1, 2,…., n; b_0, b_1, b_2.....b_k are the estimates or unstandardized regression coefficients of β_0, β_1, β_2,...., β_k respectively; and e_i is the estimated error (residual) for the ith observation.

Therefore, estimate of Y is given by:

$$\hat{Y} = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + \cdots\cdots\cdots + b_k X_{k,i} \quad (3)$$

with the difference between the observed Y and the estimated $\hat{Y}$ being known as the 'residual' (or 'residual error').

Development of MLR models

Site-specific MLR models were developed for all the 17 sites, using influential hydrological variables, 11 seasonal dummy variables and significant multi-period lags of hydrological and hydrogeological variables as independent variables, and the groundwater levels of these sites as dependent variables. In all, 70 % of the data, covering the period 1999–2002 was used for training and 30 % of the data (for the period 2003–2004) was used for testing. The step-wise-regression method was employed using STATISTICA 6.0 software to identify significant inputs to the site-specific MLR models, in order to predict groundwater levels at the 17 sites.

Evaluation of the developed MLR models

The effectiveness of the developed MLR models was measured by a set of standard statistical indicators, namely: standardized regression coefficient (β_j), standard error (SE), t-test, F-test, coefficient of multiple determination (R^2), multiple correlation coefficient (R), adjusted R^2, p-level and standard error of estimate (SEE). The mathematical expressions for these indicators are as follows:

$$\beta_j = b_j \frac{\sigma_j}{\sigma_Y} \quad (4)$$

where b_j is the unstandardized regression coefficient and σ_j and σ_Y are the standard deviations associated with the jth independent variable and the dependent variable Y, respectively.

$$SE = \frac{\sigma}{\sqrt{n-1}} \quad (5)$$

where σ=sample standard deviation, and n=total number of observations. In the t-test, the value of t is given as:

$$t = \frac{b_j}{SE(b_j)} \quad (6)$$

where b_j=unstandardized regression coefficients of unknown parameters. In the F-test, the value of F is given as:

$$F = \frac{\left(\mathrm{SSR}/p\right)}{\left(\mathrm{SSE}/(n-p-1)\right)} = \frac{\mathrm{MSR}}{\mathrm{MSE}} \quad (7)$$

DOI 10.1007/s10040-013-1029-5

where SSR=sum of squares regression, SSE=sum of squared error, SST=total sum of squares, MSR=mean square regression, MSE=mean squared error, p=number of independent variables, and n=total number of observations.

$$R^2 = \frac{\sum\left(\widehat{Y}_i - \overline{Y}\right)^2}{\sum\left(Y_i - \overline{Y}\right)^2} = \frac{\mathrm{SSR}}{\mathrm{SST}} \tag{8}$$

where Y_i=observed value of a dependent variable, $\widehat{Y}_i$ = estimated or predicted value of Y, and $\overline{Y}$ = mean value of a dependent variable.

$$\text{Adjusted } R^2 = 1-\left(\frac{n-1}{n-p-1}\right)\left(1-R^2\right) = 1-\left(\frac{\mathrm{SSE}/(n-p-1)}{\mathrm{SST}/(n-1)}\right) \tag{9}$$

$$\mathrm{SEE} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{n-p-1}} \tag{10}$$

where all the symbols have the same meaning as defined earlier.

The standardized regression coefficient (β_j) measures the impact of a unit change in the standardized value of an independent variable on the standardized value of a dependent variable. Standard error (SE) is a measure of the stability of regression coefficients. A t-test is performed for each regression coefficient, to examine its significance in the presence of all other independent variables. The overall statistical significance of the regression model is evaluated using an F-test in the format of ANOVA (analysis of variance; Makridakis et al. 2008). The coefficient of multiple determination (R^2), is a fraction that represents the proportion of total variation of the dependent variable accounted for (explained) by the explanatory variables. The fit of a multiple regression model can be assessed by R^2. The multiple correlation coefficient (R) is the correlation between the observed value of a dependent variable (Y) and an estimated or predicted value of Y based on multiple independent variables. The 'adjusted R^2' is interpreted in the same manner as the R^2 value, except that the adjusted R^2 takes into consideration the number of degrees of freedom (Statistica 2001). Moreover, the p-level represents the probability of error that is involved in accepting the observed results as valid. In many fields of research, including the field of hydrological sciences, a p-level of 0.05 (i.e., 5 % probability of error or 95 % confidence interval) is customarily treated as a "border-line acceptable" error level (Haan 2002; Statistica 2001). Finally, the standard error of estimate (SEE) is the measure of the amount of error in the prediction of a dependent variable (Y) for each independent variable (X) in the regression equation.

Design of ANN models

A neural network is made up of a number of nodes, which are the processing elements of the network and are usually called 'neurons'. Each neuron is connected to other neurons, receives an input signal, processes it and transforms it into an output signal (Haykin 1994). In order to execute the function of biological neurons artificially, a proper design of artificial neural network (ANN) is necessary. Hence, designing ANN (or ANN models) follows a systematic procedure involving five basics steps: (1) selection of influential inputs; (2) selection of suitable ANN architecture; (3) building the neural network; (4) training and testing of the developed ANN models; and (5) performance evaluation of the ANN models. The first step is described in section "Study sites and data sources" and a brief description of the remaining steps is given in the following.

Selection of ANN architecture

After selecting significant inputs for the ANN, the next step is the selection of suitable network architecture. A multi-layer feed-forward network or multi-layer perceptrons (MLP) with single hidden layer was selected for this study, because this architecture has been widely used in simulation/prediction of groundwater levels for its ease of implementation and also for being considered as a universal approximator (Hornik et al. 1989). In a multi-layer feed-forward neural network, the nodes are generally arranged in layers, comprising an input layer, an output layer and one or more hidden layers. A schematic diagram of a three-layer feed-forward neural network having an input layer (with 22 input neurons), one hidden layer and one output layer, as used in this study, is illustrated in Fig. 3. It has input nodes ($x_1, x_2, x_3, \ldots\ldots\ldots, x_{i,\ldots}, x_n$) including bias ($x_0$), hidden nodes ($z_1, z_2, z_3, \ldots\ldots\ldots, z_{j,\ldots}, z_l$) including bias ($z_0$), and output node ($y_k$), where x, z, and y represent the input, hidden, and output layers respectively, and n, l, and m represent the number of the nodes in each layer (Fig. 3). The weights associated with the connections between the input and the hidden nodes are denoted by w_{ij} and those between the hidden and the output nodes are denoted by w_{jk}. The addition of a bias term to an artificial neuron can be expressed by incorporating a bias element into the input and weight vectors to create extended vectors resulting in an increase of their dimensionality by one. It is useful as it helps the neural network in learning patterns by providing additional weights to the connection between neurons and the bias input is generally assigned a value of positive one (Haykin 1994). The output of a node in the hidden layer (z_j) is determined by applying a nonlinear transformation (activation function) to the sum of the weighted inputs it received from neurons in the input layer. Then the weighted sum of inputs at the output layer is transformed to the network output using a linear activation function between hidden and output

 DOI 10.1007/s10040-013-1029-5

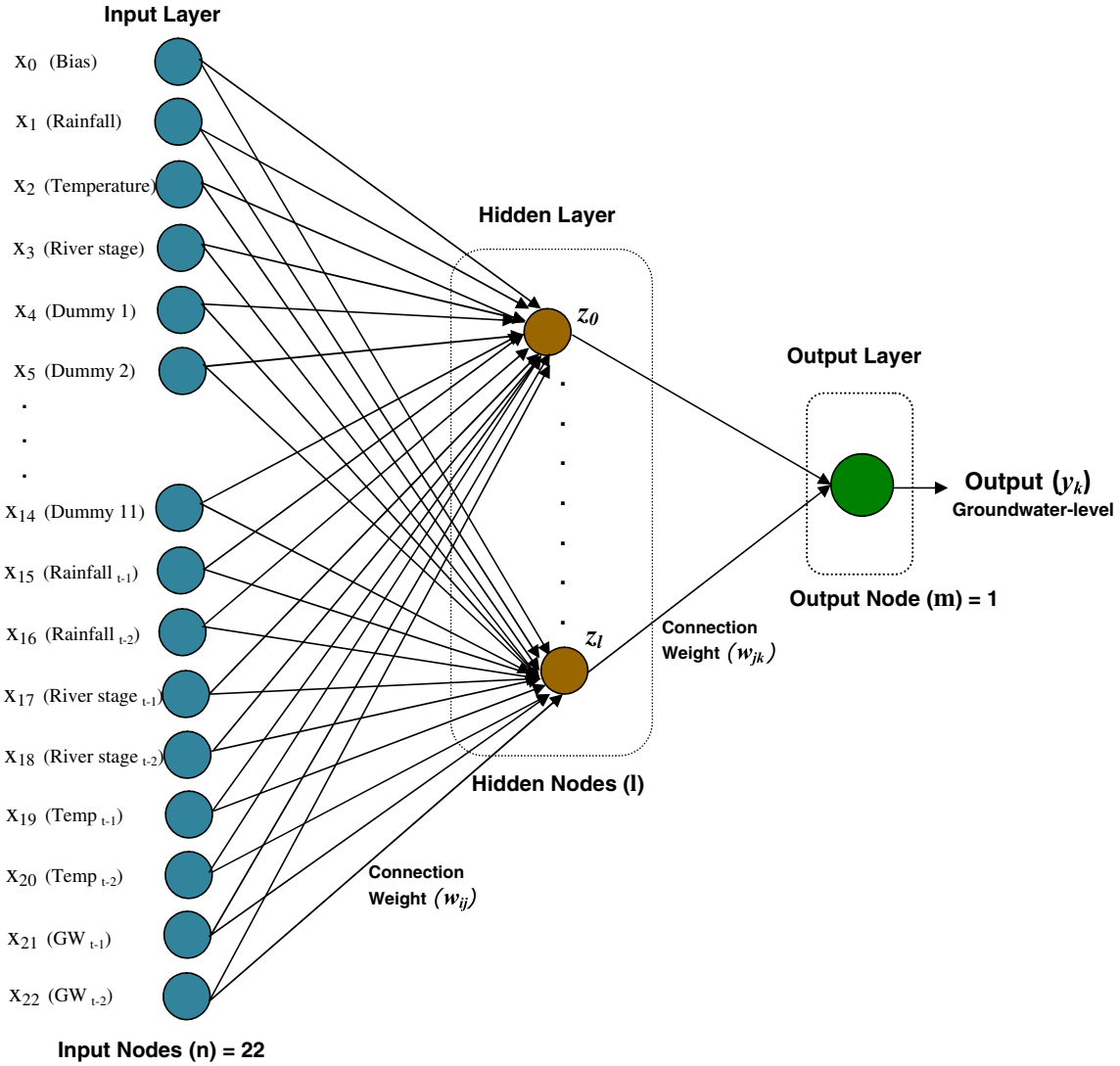

Fig. 3 The architecture for the multi-layer feed-forward network used in the study (parameters defined in the text)

layer. Thus, the prediction of the ANN model (y_k) can be expressed as follows (Haykin 1994):

$$y_k = \sum_{j=1}^{l} w_{jk} z_j + b_j \qquad (11)$$

where w_{jk} is the connection weight between jth node of hidden layer and output node k; l is the number of neurons in the hidden layer; z_j is the output of the jth hidden neuron resulting from all the input data and b_j is the connection weight for bias term.

Training algorithm

After selecting the ANN architecture, the next step is the training of the network. The purpose of training is to ensure that the network extracts the fundamental characteristics or pattern from the datasets used in ANN modeling. The training process consists of determining the ANN weights and biases. The ANN is trained with a set of input and known output data to determine the connection weights, so that the ANN-simulated values are as close as possible to the known output (observed) values. The most common method used for training feed-forward neural networks is back-propagation training (Hagan and Menhaj 1994; Hagan et al. 1996), which is a two-step approach. In the first step, the input signal is propagated forward to estimate the outputs. In the second step, a backward step is performed to adjust the weight vectors between the layers to minimize the network error. Although various weight-optimization techniques are available for the training of ANNs, the Levenberg-Marquardt (LM) algorithm was used in this study, owing to its ability to achieve convergence quickly (Moré 1978; Press et al. 1992). The LM algorithm trains neural networks at a rate 10 to 100 times faster than the usual gradient-descent back-propagation algorithm, and is more effective at finding optimal results than standard back-propagation (Hagan and Menhaj 1994). The LM algorithm combines the Gauss-Newton algorithm and the method of gradient descent, and it uses an approximation of the Hessian matrix to update weights as given below (Bishop 1995):

 DOI 10.1007/s10040-013-1029-5

Table 2 Categorization of sensitivity and associated ranks

Category of sensitivity	Value of sensitivity index	Rank
Very high sensitivity	>0.1	1
High sensitivity	>0.05–0.1	2
Moderate sensitivity	>0.01–0.05	3
Low sensitivity	>0.005–0.01	4
Very low sensitivity	≤0.005	5

$$\Delta w = \left[\mathbf{J}^T(w)\mathbf{J}(w) + \lambda I\right]^{-1}\mathbf{J}^T(w)\mathbf{e}(w) \quad (12)$$

where, w=weight, $\mathbf{J}$ =Jacobian matrix, $\mathbf{J}^T$=transpose matrix of $\mathbf{J}$, $\mathbf{J}^T\mathbf{J}$=Hessian matrix, λ=learning parameter, $\mathbf{I}$=identity matrix and $\mathbf{e}$=vector of network errors. The parameter λ governs the step size and is automatically adjusted based on the direction of the error at each iteration, in order to secure convergence. In this study, an initial λ value of 0.15 was used for optimizing weights with the LM algorithm.

Activation function

The activation (transfer) function determines the response of a node to the total input signal it receives. The most commonly used activation function, named logistic sigmoid-type function (Haykin 1994; Dekker et al. 2001) was used in this study for the hidden layer. However, a linear-type activation function was used for the output layer, as suggested by Maier and Dandy (2000) and Rumelhart et al. (1995). The sigmoid function is a bounded, monotonic, non-decreasing function that provides a graded, non-linear response (Shamseldin 1997), whereas a linear-transfer function calculates a neuron's output by simply returning the value passed to it. The mathematical expressions for these two functions are as follows:

$$\text{Logistic sigmoid function}: \varphi(v) = \frac{1}{1+\exp(-v)} \quad (13)$$

$$\text{Linear function}: \varphi(v) = v \quad (14)$$

where v in Eq. (13) represents the weighted sum for a node in the hidden layer and exp denotes the natural exponential function and in Eq. (14) v is the weighted sum of inputs at the output layer as considered in the present study.

Number of hidden-layer neurons

The function of the hidden-layer neurons is to detect relationships between network inputs and outputs. If there is an insufficient number of hidden neurons, it may be difficult to obtain convergence within the training time. On the other hand, if too many hidden neurons are used, the network may lose its ability to generalize (ASCE 2000a). Although a trial-and-error method is widely used for determining the optimal number of hidden neurons, some researchers have demonstrated the effectiveness of a genetic algorithm (GA) for optimizing hidden neurons (Jones 1993; Castillo et al. 2000; Gerken et al. 2006). In particular, GA has been found to be quite useful and efficient when the exploration space of the network is extensive. Consequently, the heuristic optimization technique GA was used in this study for optimizing the hidden neurons for all the 17 site-specific ANN models.

In the GA optimization, initially, chromosomes (consisting of 'genes' of binary numbers) were characterized (encoded) by a set of unknown parameters (i.e., hidden neurons). The number of hidden neurons was set in a range between a lower bound of 1 and an upper bound of 30. Thereafter, the GA parameters such as selection method, crossover probability, mutation probability, population size, and number of generations, were decided based on trial and error and standard literature (Goldberg 1989; Reed et al. 2000 and Mayer et al. 2001). The GA was initiated with 50 randomly generated chromosomes, with a maximum number of generations of 200, with gene structures as mentioned in the preceding. The best-fit chromosomes were chosen by roulette-wheel selection, based on the ranking-algorithm method, and the crossover and mutation operators were applied with a crossover probability of 0.9–1.0 and a mutation probability of 0.001–0.01, to produce a new population (of chromosomes) for the next generation. The new chromosomes thus reproduced were evaluated based on their fitness values, and this procedure of evaluation and reproduction of the chromosomes was repeated until a pre-defined termination criterion (i.e., 'fitness threshold') of 0.001 m^2 was satisfied. The combination of

Table 3 MLR coefficients and their significance for the MLR model for site F-6 using a step-wise regression technique (parameters defined in the text). *SE* standard error; *Temp* temperature

Parameter	Beta	SE (m)	B	SE (m)	*t* (37)	*p*-level
Intercept			1.625	0.238	6.831	0.000000
Temp	0.259	0.106	0.025	0.010	2.448	0.001921
D_4 (Apr–May)	0.236	0.027	0.618	0.070	8.840	0.000000
GW_{t-1}	0.497	0.080	0.495	0.080	6.175	0.000000
D_7 (Jul–Aug)	0.173	0.027	0.453	0.071	6.356	0.000000
Rainfall	0.141	0.027	0.001	0.000	5.160	0.000009
$Rainfall_{t-2}$	−0.074	0.028	0.0003	0.000	−2.621	0.000637
D_{10} (Oct–Nov)	−0.036	0.024	−0.095	0.064	−1.483	0.006509
D_2 (Feb–Mar)	−0.065	0.025	−0.170	0.066	−2.593	0.001354
$Temp_{t-2}$	−0.266	0.102	−0.026	0.010	−2.620	0.001267
$Temp_{t-1}$	0.329	0.157	0.032	0.015	2.089	0.004366

Regression summary: multiple R=0.941, multiple R^2 =0.885, adjusted R^2 =0.869, *F-statistic*=216.414, *p*-level=0.0000 and *SEE*=0.107 m; D_2, D_4 etc. refer to seasonal dummy variables

 DOI 10.1007/s10040-013-1029-5

Table 4 MLR models for predicting groundwater levels at 17 sites using training datasets

Site	MLR models based on training datasets
A-2	$GWL = 16.1572 + 0.0003 \times R + 0.0129 \times T + 0.8803 \times 0.2878 \times D_4 + 0.0917 \times D_5 - 0.1040 \times D_9 - 0.1020 \times D_{10} + 0.0003 \times R_{t-1} + 0.0133 \times T_{t-1} - 0.2404 \times S_{t-1} - 0.0647 \times S_{t-2}$
B-3	$GWL = 14.2696 + 0.0004 \times R + 0.0375 \times D_3 + 0.0586 \times D_4 + 0.0663 \times D_8 - 0.2739 \times GW_{t-1} + 0.0003 \times R_{t-1} - 0.1529 \times S_{t-1}$
C-2	$GWL = 4.9457 + 0.0011 \times R + 0.1033 \times T + 0.2836 \times D_2 + 1.6220 \times D_4 + 0.6288 \times D_3 - 0.7022 \times D_8 - 0.3603 \times D_9 - 0.4169 \times D_{10} + 0.5431 \times GW_{t-1} - 0.0007 \times R_{t-2} - 0.4311 \times S_{t-2}$
C-7	$GWL = 4.0716 - 0.0002 \times R + 0.0519 \times D_4 + 0.0489 \times D_6 + 0.0666 \times D_9 + 0.4985 \times GW_{t-1} - 0.0003 \times R_{t-1} + 0.1618 \times S_{t-1} + 0.0729 \times S_{t-2}$
D-6	$GWL = 5.4199 + 0.0003 \times R + 0.0432 \times T + 0.6289 \times S - 0.212 \times D_1 + 0.0620 \times D_4 - 0.5236 \times D_2 - 0.3439 \times D_3 - 0.163135 \times D_8 + 0.000358 \times R_{t-1} - 0.0186 \times T_{t-2}$
E-2	$GWL = 2.9648 + 0.0235 \times T + 1.3327 \times S - 0.2348 \times D_1 + 0.2587 \times D_2 + 0.4599 \times D_3 + 1.4086 \times D_4 + 0.3248 \times D_5 + 0.4926 \times D_7 + 0.4037 \times GW_{t-1} + 0.0008 \times R_{t-1} - 1.4295 \times S_{t-1} + 0.0314 \times T_{t-1}$
E-4	$GWL = 2.6132 + 0.0011 \times R + 0.0748 \times T + 1.0500 \times D_4 - 0.3106 \times D_1 + 0.4598 \times D_7 - 0.5201 \times D_2 - 0.1631 \times D_8 + 0.1911 \times D_{11} + 0.3776 \times GW_{t-1} + 0.0951 \times GW_{t-2} - 0.0004 \times R_{t-1} - 0.0009 \times R_{t-2} - 0.0248 \times T_{t-2}$
E-5	$GWL = 2.0439 + 0.0011 \times R + 0.0349 \times T + 1.0076 \times D_4 - 0.2406 \times D_2 + 0.9641 \times D_7 + 0.5285 \times GW_{t-1} + 0.0005 \times R_{t-2} + 0.0421 \times T_{t-1} - 0.0421 \times T_{t-2}$
F-1	$GWL = 2.3141 + 0.0004 \times R - 0.0079 \times T + 1.0911 \times S - 0.2895 \times D_4 + 0.0426 \times D_5 + 0.1080 \times D_6 + 0.1486 \times D_7 + 0.2649 \times GW_{t-1} - 0.2179 \times GW_{t-2} - 0.0003 \times R_{t-t} + 0.0003 \times R_{t-2} - 0.5139 \times S_{t-1} + 0.0357 \times T_{t-1} + 0.0121 \times T_{t-2}$
F-6	$GWL = 1.6247 + 0.0006 \times R + 0.0252 \times T + 0.6183 \times D_4 - 0.1702 \times D_2 + 0.4527 \times D_7 - 0.0948 \times D_{10} + 0.4591 \times GW_{t-1} - 0.0003 \times R_{t-2} + 0.0322 \times T_{t-1} - 0.0262 \times T_{t-2}$
G-2	$GWL = 0.4252 + 0.0008 \times R + 0.0615 \times T + 0.3063 \times S + 0.1820 \times D_1 - 0.2577 \times D_2 - 0.8261 \times D_4 + 0.2460 \times D_7 + 0.2432 \times D_{11} + 0.2498 \times GW_{t-1} + 0.1022 \times GW_{t-2} + 0.0005 \times R_{t-1} - 0.2998 \times S_{t-1} - 0.0006 \times R_{t-2}$
GH-4.5	$GWL = -0.9839 + 0.0010 \times R + 0.2445 \times D_1 + 0.4051 \times D_2 + 0.7339 \times D_3 + 0.9149 \times D_4 + 0.3509 \times D_{11} + 0.6470 \times GW_{t-1} - 0.2079 \times GW_{t-2} + 0.0005 \times R_{t-1} + 0.0902 \times T_{t-1} - 0.6591 \times S_{t-1}$
H-2	$GWL = -0.0774 + 0.0003 \times R + 0.0257 \times T + 0.0861 \times S + 0.1150 \times D_4 - 0.1496 \times D_5 - 0.1080 \times D_6 + 0.0900 \times D_{11} + 0.5011 \times GW_{t-1} + 0.3718 \times GW_{T-2} + 0.0004 \times R_{t-2} + 0.0097 \times T_{t-1} - 0.0199 \times T_{t-2} + 0.2208 \times S_{t-1}$
H-3	$GWL = -0.1759 + 0.0007 \times R + 0.0261 \times T + 0.2050 \times D_3 + 0.7763 \times D_4 + 0.0758 \times D_6 + 0.2191 \times D_7 + 0.2100 \times D_{11} + 0.3995 \times GW_{t-1} + 0.0005 \times R_{t-1} - 0.0002 \times R_{t-2} + 0.0373 \times T_{t-1} - 0.3886 \times S_{t-1}$
H-4	$GWL = -0.0068 + 0.0007 \times R + 0.2399 \times D_3 + 0.6511 \times D_4 + 0.1570 \times D_{11} + 0.1432 \times D_7 + 0.4849 \times GW_{t-1} - 0.0927 \times GW_{t-2} + 0.0005 \times R_{t-1} + 0.0584 \times T_{t-1} - 0.0102 \times T_{t-2} + 0.4554 \times S_{t-1}$
H-5	$GWL = -0.5338 + 0.0003 \times R + 0.2817 \times S + 0.1116 \times D_2 - 0.1053 \times D_7 + 0.2362 \times D_3 + 0.2652 \times D_4 - 0.0859 \times D_5 + 0.5748 \times GW_{t-1} + 0.0356 \times T_{t-1} - 0.4258 \times S_{t-1}$
I-2	$GWL = -0.0626 + 0.0175 \times T + 0.2311 \times S + 0.0004 \times R_{t-2} + 0.0748 \times D_5 + 0.4272 \times GW_{t-1} - 0.0002 \times R_{t-1} - 0.0033 \times T_{t-1} + 0.1847 \times S_{t-2}$

GWL groundwater-level, *R* rainfall, *T* temperature, *S* river stage, D_1, D_2, D_3,..,D_{11} seasonal dummy variables

parameter that produced the lowest fitness value during training runs was selected as the optimal ANN network.

Optimization of neuron weights

During the training process, the three factors that are associated with optimization of weights are: (1) initial weight matrix; (2) error function; and (3) termination criteria, such as fixing the number of epochs setting a target error-goal and fixing the minimum performance gradient (derivatives of network error with respect to those weights and bias). An initial weight range of −0.5 to 0.5 was selected in this study, as suggested by earlier researchers (e.g., Sietsma and Dow 1991; Looney 1996). These weights were then optimized using Eq. (11). The following objective

 DOI 10.1007/s10040-013-1029-5

Table 5 Statistical indicators for the developed MLR models based on training datasets

Site	Multiple R	Multiple R^2	Adjusted R^2	F-statistic	p-level	SEE (m)
A-2	0.981	0.963	0.951	84.546	0.0000	0.090
B-3	0.884	0.781	0.742	20.337	0.0000	0.057
C-2	0.973	0.947	0.931	58.217	0.0000	0.351
C-7	0.738	0.544	0.436	5.040	0.0002	0.067
D-6	0.969	0.939	0.923	57.305	0.0000	0.163
E-2	0.944	0.891	0.848	72.630	0.0000	0.227
E-4	0.986	0.973	0.962	93.055	0.0000	0.227
E-5	0.975	0.950	0.936	210.716	0.0000	0.177
F-1	0.958	0.917	0.897	99.452	0.0000	0.089
F-6	0.941	0.885	0.869	216.414	0.0000	0.107
G-2	0.986	0.973	0.962	93.319	0.0000	0.174
GH-4.5	0.979	0.958	0.946	75.270	0.0000	0.221
H-2	0.975	0.951	0.932	50.661	0.0000	0.090
H-3	0.949	0.901	0.885	154.325	0.0000	0.117
H-4	0.974	0.948	0.932	146.570	0.0000	0.104
H-5	0.927	0.860	0.851	49.270	0.0000	0.120
I-2	0.933	0.870	0.864	45.805	0.0000	0.070

function (E) was used for the optimization of weights (Principe et al. 2000):

$$E = \frac{1}{2n}\sum_{i=1}^{n}\left(h_{oi}-h_{pi}\right)^2 \quad (15)$$

where, h_{oi}=observed groundwater-level at the ith time [L], h_{pi}=predicted groundwater-level at the ith time [L], and n=total number of observations.

Monthly groundwater-level data for the 4-year period 1999–2002 were used for training the 17 site-specific ANN models, with the 2-year period 2003–2004 being used for testing. The entire ANN modeling exercise was carried out using the Neural Builder Wizard of NeuroSolutions (version 5.05) software (Principe et al. 2000).

Model evaluation criteria

Performance evaluation of the 17 site-specific MLR and ANN models was carried out using four statistical indicators, namely: mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (r^2) and Nash-Sutcliffe efficiency (NSE), in order to examine their effectiveness in predicting transient groundwater levels at individual sites. MAE indicates the average of the total model errors (both under-prediction and over-prediction) and is used to measure how close forecasts/predictions are to the eventual outcomes (observed values). RMSE indicates an overall (global) discrepancy between the observed values and the calculated (predicted or simulated) values. r^2 describes the proportion of the total variance in the observed data that can be explained by a model. NSE, on the other hand, is an index for assessing the predictive accuracy of a model, and it represents an improvement over r^2 for model evaluation, because it is sensitive to the differences in observed and model-simulated or predicted means and variances (Legates and McCabe 1999). The mathematical expressions for MAE, RMSE, r^2 and NSE are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|h_{oi}-h_{pi}\right| \quad (16)$$

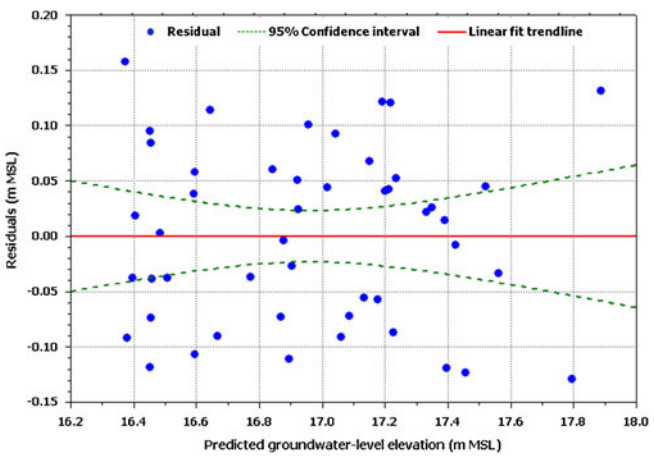

Fig. 4 Plot of residuals versus predicted groundwater levels for MLR model of site GH-4.5

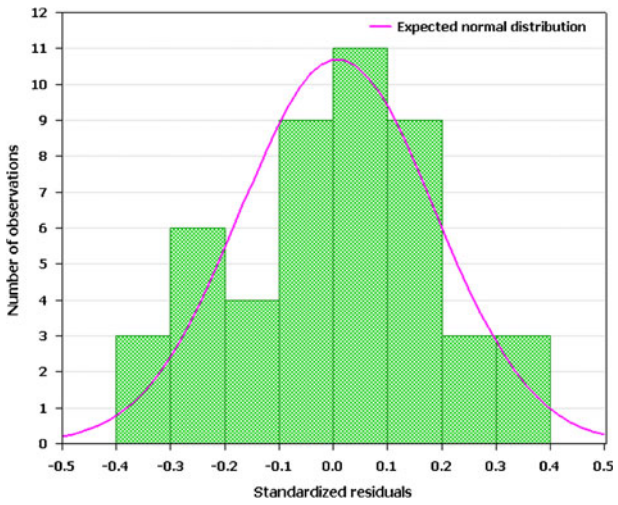

Fig. 5 Histogram of residuals for MLR model of site GH-4.5

DOI 10.1007/s10040-013-1029-5

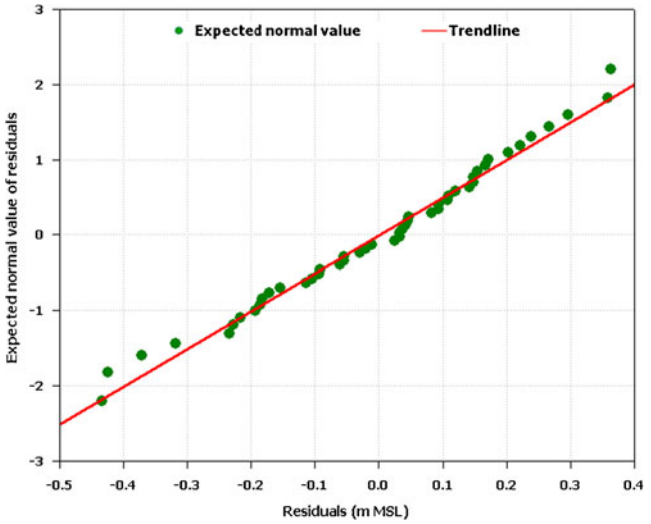

Fig. 6 Normal probability plot of residuals for MLR model of site GH-4.5

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(h_{oi}-h_{pi}\right)^2} \tag{17}$$

$$r^2 = \left[\frac{\sum_{i=1}^{n}\left[\left(h_{pi}-\overline{h}_p\right)\left(h_{oi}-\overline{h}_o\right)\right]}{\sqrt{\left[\left(\sum_{i=1}^{n}\left(h_{pi}-\overline{h}_p\right)^2\right)\left(\sum_{i=1}^{n}\left(h_{oi}-\overline{h}_o\right)^2\right)\right]}}\right]^2 \tag{18}$$

$$\mathrm{NSE} = 1-\frac{\sum_{i=1}^{n}\left(h_{oi}-h_{pi}\right)^2}{\sum_{i=1}^{n}\left(h_{oi}-\overline{h}_o\right)^2} \tag{19}$$

where, h_{oi}=observed groundwater-level at the ith time [L], h_{pi}=predicted groundwater-level at the ith time [L], $\overline{h}_o$ = mean of the observed groundwater levels [L], $\overline{h}_p$ = mean of the predicted groundwater levels [L], and n=total number of observations.

Table 7 Neural network architecture and optimal GA parameters for the 17 ANN models

Site	ANN architecture (n-l-m)	No. of generations	Best chromosome	Fitness (m^2)
A-2	22-14-1	2	29	0.001
B-3	21-8-1	1	34	0.001
C-2	22-8-1	1	32	0.004
C-7	21-11-1	2	30	0.001
D-6	22-10-1	1	29	0.059
E-2	22-6-1	2	35	0.048
E-4	22-7-1	2	40	0.017
E-5	22-6-1	1	26	0.017
F-1	22-7-1	2	20	0.008
F-6	22-11-1	1	19	0.017
G-2	22-6-1	1	39	0.003
GH-4.5	22-10-1	2	34	0.0005
H-2	22-7-1	1	32	0.0002
H-3	22-9-1	2	21	0.021
H-4	22-10-1	2	39	0.0057
H-5	22-9-1	1	42	0.0079
I-2	22-12-1	1	32	0.0016

n number of input neurons, l number of hidden layer neurons, m number of output neurons

In addition to quantitative evaluation using statistical indicators, the efficacy of the MLR and ANN techniques in predicting transient groundwater levels at the 17 sites was assessed by graphical indicators such as visual checking of simulated and observed groundwater-level hydrographs for individual sites, as well as scatter plots of simulated versus observed groundwater levels along with 1:1 lines for individual sites. Prediction errors (as a measure of model uncertainty) of all the MLR and ANN models were also analyzed.

Sensitivity analysis

Sensitivity analyses were conducted for the ANN model (i.e., MLP-LM) to investigate the relative importance of each input variable for accurately predicting groundwater-levels. This analysis was carried out for all the 17 sites by imposing certain changes on individual inputs and observing their effects on the model output. The change in the input was made by adding a random value

Table 6 Results of the multicollinearity test for site GH-4.5

Variable	R^2	Multicollinearity statistic		t (36)	p-Level
		VIF	Tolerance		
Temp_{t-1}	0.728	3.671	0.272	8.697	0.0000
D_4 (Apr–May)	0.571	2.329	0.429	5.198	0.0000
GW_{t-1}	0.712	3.478	0.288	5.601	0.0000
Rainfall	0.443	1.796	0.557	4.243	0.0001
D_3 (Mar–Apr)	0.659	2.930	0.341	3.717	0.0007
D_{11} (Nov–Dec)	0.303	1.436	0.697	2.539	0.0056
$\mathrm{GW}_{(t-2)}$	0.752	4.032	0.248	3.346	0.0046
D_2(Feb–Mar)	0.654	2.887	0.346	2.068	0.0025
River stage_{t-1}	0.763	4.214	0.237	4.841	0.0038
D_1 (Jan–Feb)	0.611	2.568	0.389	3.323	0.0043
$\mathrm{Rainfall}_{t-1}$	0.733	3.749	0.267	3.178	0.0047

VIF variance inflation factor; D_1, D_2, D_3 etc. refer to seasonal dummy variables

 DOI 10.1007/s10040-013-1029-5

Hydrogeology Journal (2013) 21: 1865–1887

DOI 10.1007/s10040-013-1029-5

Table 8 Goodness-of-fit statistics for the ANN and MLR models of the 17 sites during training and testing periods

Site	Dataset used		ANN models				MLR models			
			MAE (m)	RMSE (m)	r^2	NSE	MAE (m)	RMSE (m)	r^2	NSE
A-2	Training		0.072	0.05	0.987	0.984	0.122	0.078	0.963	0.963
	Testing		0.121	0.128	0.935	0.912	0.137	0.127	0.914	0.914
B-3	Training		0.023	0.033	0.925	0.907	0.041	0.052	0.780	0.780
	Testing		0.038	0.04	0.825	0.822	0.056	0.050	0.736	0.736
C-2	Training		0.075	0.085	0.999	0.995	0.213	0.304	0.947	0.947
	Testing		0.311	0.397	0.918	0.901	0.297	0.474	0.858	0.858
C-7	Training		0.021	0.036	0.832	0.828	0.057	0.059	0.544	0.544
	Testing		0.042	0.038	0.781	0.776	0.061	0.080	0.007	0.007
D-6	Training		0.047	0.059	0.993	0.990	0.102	0.144	0.939	0.939
	Testing		0.104	0.157	0.930	0.890	0.127	0.255	0.709	0.709
E-2	Training		0.221	0.304	0.910	0.905	0.246	0.394	0.891	0.891
	Testing		0.241	0.220	0.935	0.995	0.278	0.263	0.930	0.930
E-4	Training		0.157	0.182	0.976	0.975	0.173	0.191	0.973	0.973
	Testing		0.287	0.411	0.89	0.861	0.306	0.471	0.887	0.887
E-5	Training		0.245	0.186	0.973	0.973	0.265	0.257	0.950	0.950
	Testing		0.323	0.264	0.944	0.931	0.317	0.269	0.928	0.928
F-1	Training		0.096	0.123	0.937	0.936	0.102	0.174	0.917	0.917
	Testing		0.098	0.132	0.912	0.911	0.114	0.148	0.888	0.888
F-6	Training		0.212	0.186	0.937	0.934	0.228	0.246	0.885	0.885
	Testing		0.221	0.187	0.922	0.918	0.243	0.071	0.989	0.989
G-2	Training		0.061	0.075	0.995	0.993	0.072	0.146	0.973	0.973
	Testing		0.193	0.277	0.912	0.895	0.197	0.253	0.910	0.910
GH-4.5	Training		0.015	0.03	0.998	0.998	0.045	0.191	0.958	0.958
	Testing		0.101	0.129	0.971	0.967	0.115	0.288	0.841	0.841
H-2	Training		0.017	0.020	0.997	0.997	0.036	0.075	0.951	0.951
	Testing		0.057	0.100	0.937	0.914	0.062	0.152	0.802	0.802
H-3	Training		0.159	0.205	0.925	0.922	0.178	0.300	0.901	0.901
	Testing		0.177	0.246	0.941	0.884	0.192	0.280	0.938	0.938
H-4	Training		0.055	0.107	0.970	0.969	0.076	0.121	0.948	0.948
	Testing		0.112	0.161	0.919	0.917	0.127	0.217	0.849	0.849
H-5	Training		0.107	0.126	0.903	0.899	0.132	0.145	0.860	0.860
	Testing		0.087	0.118	0.906	0.837	0.152	0.115	0.845	0.845
I-2	Training		0.045	0.056	0.926	0.924	0.076	0.073	0.870	0.870
	Testing		0.055	0.068	0.905	0.893	0.089	0.077	0.863	0.863
Range of statistical indicators	Training	Min:	0.015 (site GH-4.5)	0.02 (site H-2)	0.832 (site C-7)	0.828 (site C-7)	0.036 (site H-2)	0.052 (site B-3)	0.544 (site C-7)	0.544 (site C-7)
		Max:	0.245 (site E-5)	0.304 (site E-2)	0.999 (site C-2)	0.998 (site GH-4.5)	0.265 (site E-5)	0.394 (site E-2)	0.973 (site E-4)	0.973 (site E-4)
	Testing	Min:	0.038 (site B-3)	0.038 (site C-7)	0.781 (site C-7)	0.776 (site C-7)	0.056 (site B-3)	0.050 (site B-3)	0.007 (site C-7)	0.007 (site C-7)
		Max:	0.323 (site E-5)	0.411 (site E-4)	0.971 (site GH-4.5)	0.995 (site E-2)	0.317 (site E-5)	0.474 (site C-2)	0.989 (site F-6)	0.989 (site F-6)

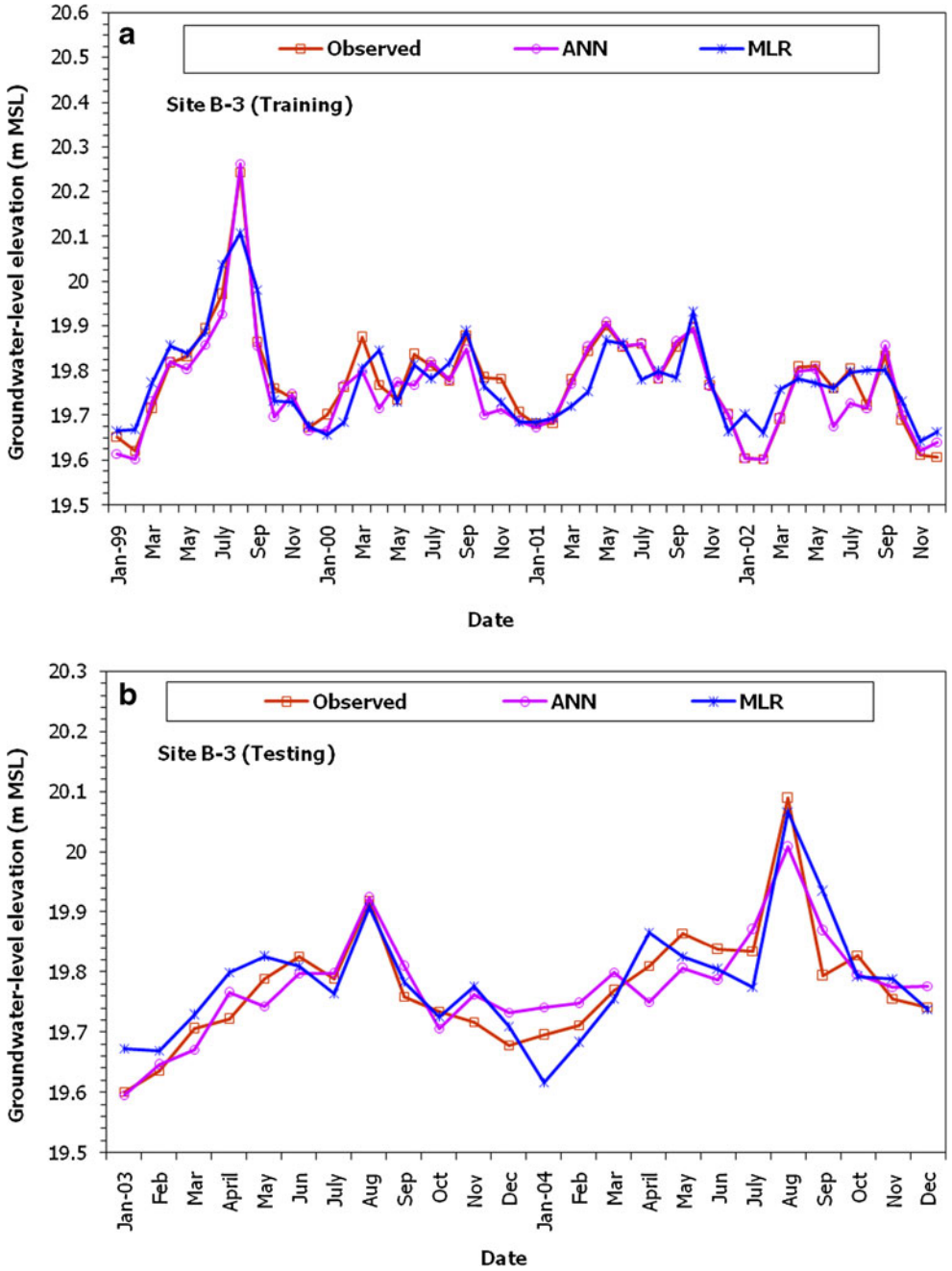

Fig. 7 Observed and predicted groundwater levels using *ANN* and *MLR* models at *site B-3*: **a** for the training period (1999–2002); **b** for the testing period (2003–2004) (*m MSL* m above mean sea level)

of known variance to each sample, while keeping the other inputs at their mean values and then calculating the change in the model output (i.e., simulated water level). The changes in the model results were assessed using a sensitivity index, which is expressed as follows (Principe et al. 2000):

$$S_k = \frac{\sum_{p=1}^{p}\sum_{i=1}^{n}\left(y_{ip}-\overline{y}_{ip}\right)^2}{\sigma_k{}^2} \qquad (20)$$

where, S_k=sensitivity index for input k, $\overline{y}_{ip}$ = ith output obtained with the fixed weights for the pth pattern, n=number of network outputs, p=number of patterns, and $\sigma_k{}^2$ = variance of the input k. The training dataset was used for computing sensitivity indices of each input. After calculating the sensitivity index for all the inputs at each of the 17 sites, they were classified into five categories, depending upon their sensitivity index values and ranked on a scale of 1–5 (Table 2).

 DOI 10.1007/s10040-013-1029-5

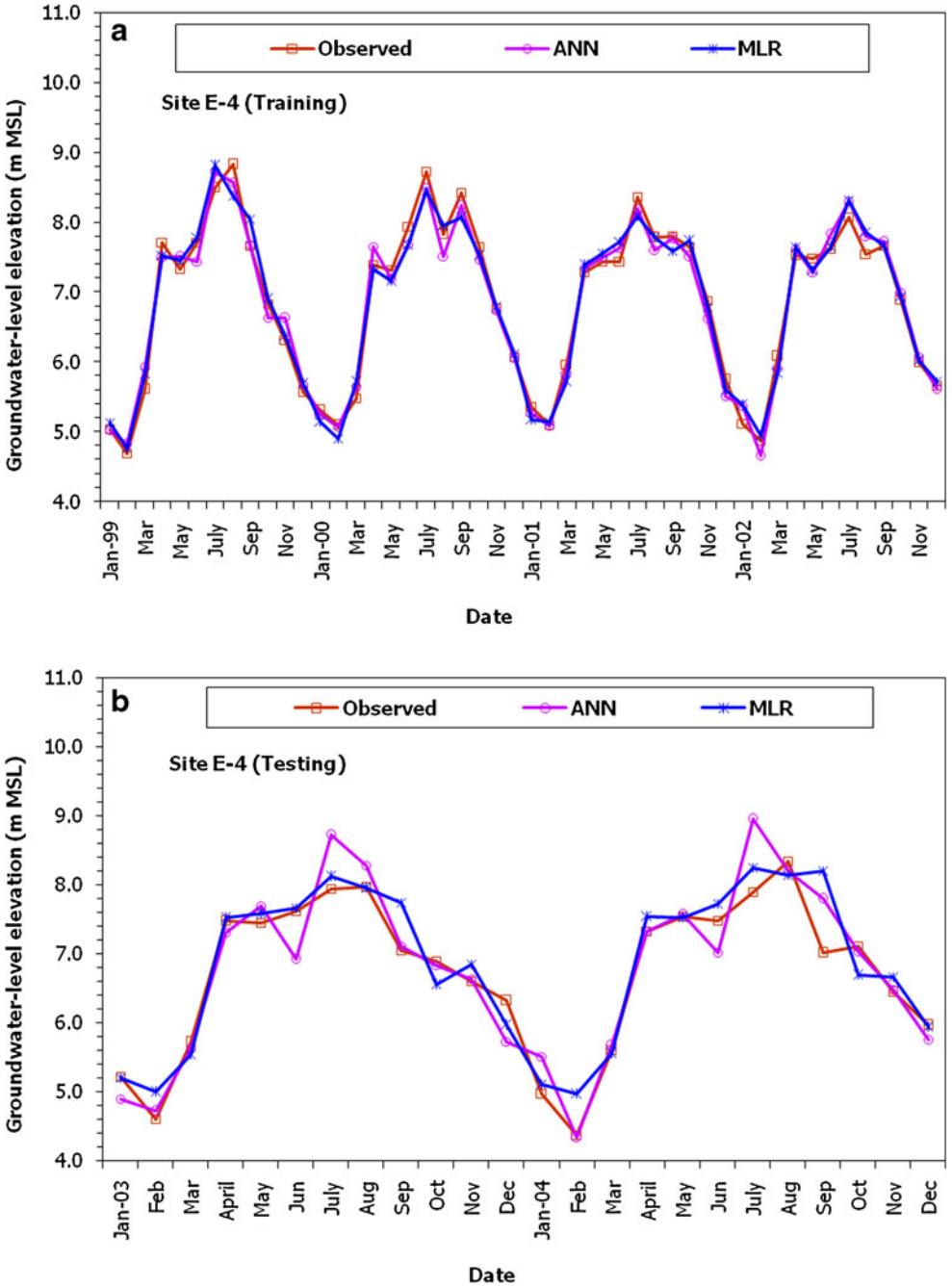

Fig. 8 Observed and predicted groundwater levels using *ANN* and *MLR* models at *site E-4*: **a** for the training period (1999–2002); **b** for the testing period (2003–2004)

Results and discussion

Developed MLR models

The results of MLR analysis for site F-6 using 22 independent variables and a step-wise regression technique are summarized in Table 3 as an example. In Table 3, the 'beta' (standardized regression coefficients) values show the relative contribution of each independent variable in prediction of groundwater levels. The 'B' refers to the raw or unstandardized regression coefficients or weights, which are included in the regression equation. Thus, the MLR model obtained using a step-wise regression technique to predict groundwater levels (GWL) at site F-6 is:

$$\begin{aligned} \mathrm{GWL(F\text{-}6)} = {} & 1.6247 + 0.0006 \times R + 0.0252 \times T \\ & + 0.6183 \times D_4 - 0.1702 \times D_2 + 0.4527 \\ & \times D_7 - 0.0948 \times D_{10} + 0.4951 \times \mathrm{GW}_{t-1} \\ & - 0.003 \times R_{t-2} + 0.0322 \times T_{t-1} - 0.0262 \\ & \times T_{t-2} \end{aligned} \tag{21}$$

It is obvious from Table 3 that D_4 (t=8.84, p=0.0000), D_7 (t=6.36, p=0.0000), GW_{t-1} (t=6.17, p=0.0000), rainfall

 DOI 10.1007/s10040-013-1029-5

(t=5.16, p=0.0000) are highly significant explanatory variables, indicating higher t values and very low p-level (i.e., 0.0000) at the 95 % significance level, followed by rainfall$_{t-2}$ (t=−2.62, p=0.0006)$_{)}$, temp$_{t-2}$ (t=−2.62, p=0.001), D_2(t=−2.59, p=0.001), temp (t=2.45, p=0.002), temp$_{t-1}$ (t=2.08, p=0.004) and D_{10} (t=−1.48, p=0.006) in predicting groundwater levels at site F-6. The remaining variables are excluded from the model during step-wise regression, because their regression coefficients were found to be statistically insignificant in predicting groundwater-levels.

Moreover, an F-test was used to check the overall significance of the developed MLR model for predicting groundwater levels at site F-6. From the F-statistic (F=216.414) and p-value (p=0.0000) of site F-6, it is concluded that it is indeed a significant model, i.e., the independent variables explain a significant amount of variability in the fluctuation of groundwater-levels, and it is also confirmed that the multiple R^2 is highly significant for this model. The results of other statistical indicators for site F-6, namely, multiple correlation coefficient (R), adjusted R^2 and standard error of estimate (SEE) are also summarized in Table 3.

Similarly, for the remaining sites, the MLR models developed using training datasets (1999–2002) are presented in Table 4. It is apparent from Table 4 that there are some big differences in the values of regression coefficient between the sites, which indicates the heterogeneity of the aquifer and/or the variation in local conditions. This finding emphasizes the need to develop site-specific MLR models. Furthermore, the statistical indicators of step-wise regression for the 17 sites, using 22 inputs, are summarized in Table 5. It is apparent from Table 5 that the MLR model for sites E-4 and G-2 have

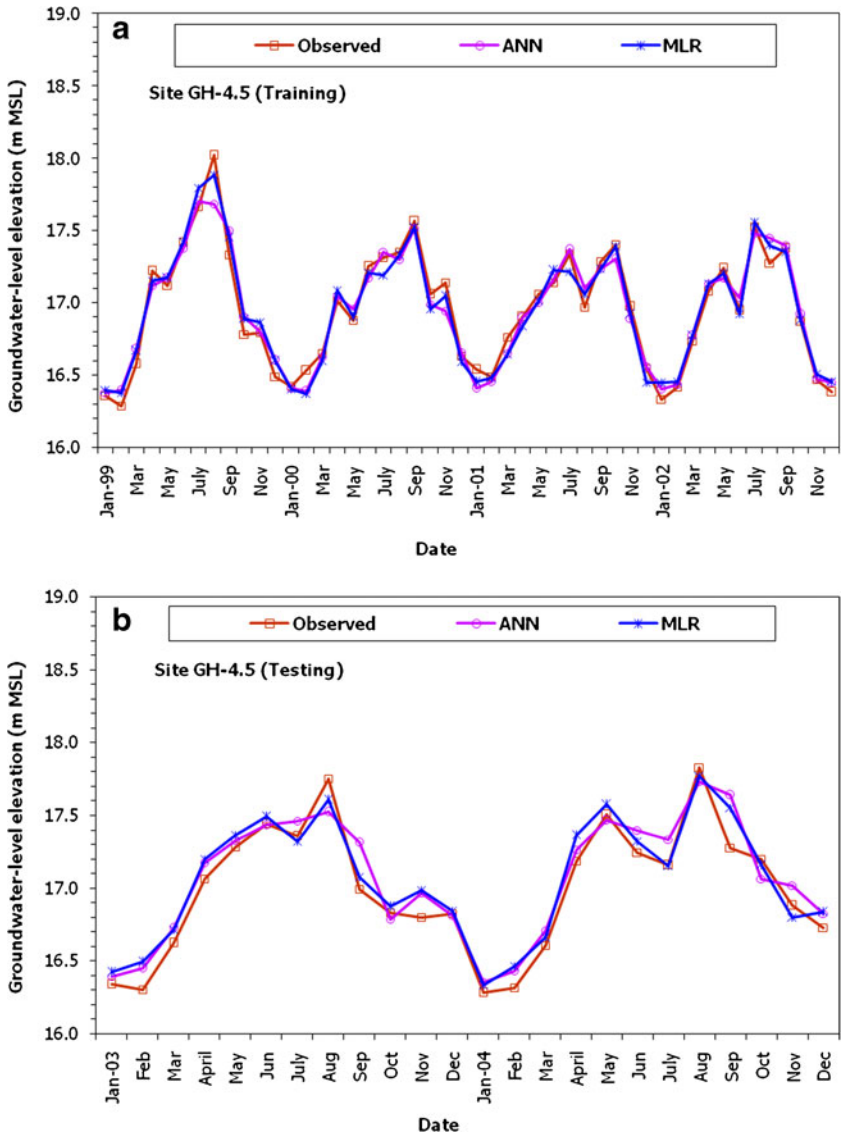

Fig. 9 Observed and predicted groundwater levels using *ANN* and *MLR* models at site *GH-4.5*: **a** for the training period (1999–2002); **b** for the testing period (2003–2004)

 DOI 10.1007/s10040-013-1029-5

the highest multiple R (0.986) and adjusted R^2 (0.962) values, and their significance is supported by the corresponding F-statistic and p-level. Thus, the values of multiple R, adjusted R^2, F-statistic and p-level for the MLR models for almost all the sites are statistically significant, which suggests that the developed MLR models can simulate/predict groundwater-level fluctuations reasonably. However, the MLR model of site C-7 has the lowest multiple R (0.738) and adjusted R^2 (0.436) among the 17 sites. This indicates that the groundwater-level at site C-7 is not significantly influenced by the environmental factors under consideration, and that the groundwater-level fluctuation at this site is affected by some anthropogenic factors such as the existence of a rubber dam very close to site C-7.

Results of diagnostic checks on the developed MLR models

After developing the MLR models for the 17 sites, the models were examined for basic assumptions to justify the use of MLR models for prediction. The four assumptions of MLR models are (Makridakis et al. 2008): (1) linearity

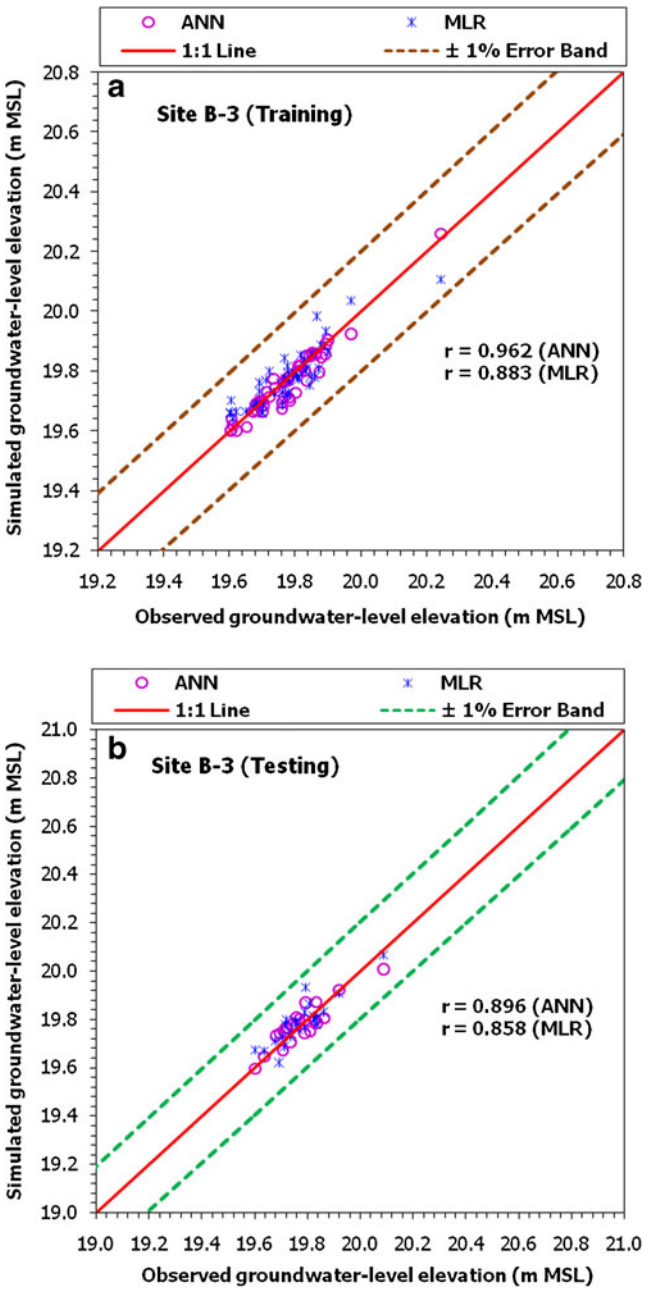

Fig. 10 Scatter plots of observed and predicted groundwater levels by *ANN* and *MLR* models with ± *1 % error band* at *site B-3*: **a** for the training period (1999–2002); **b** for the testing period (2003–2004) (r correlation coefficient between observed and simulated groundwater-level elevation)

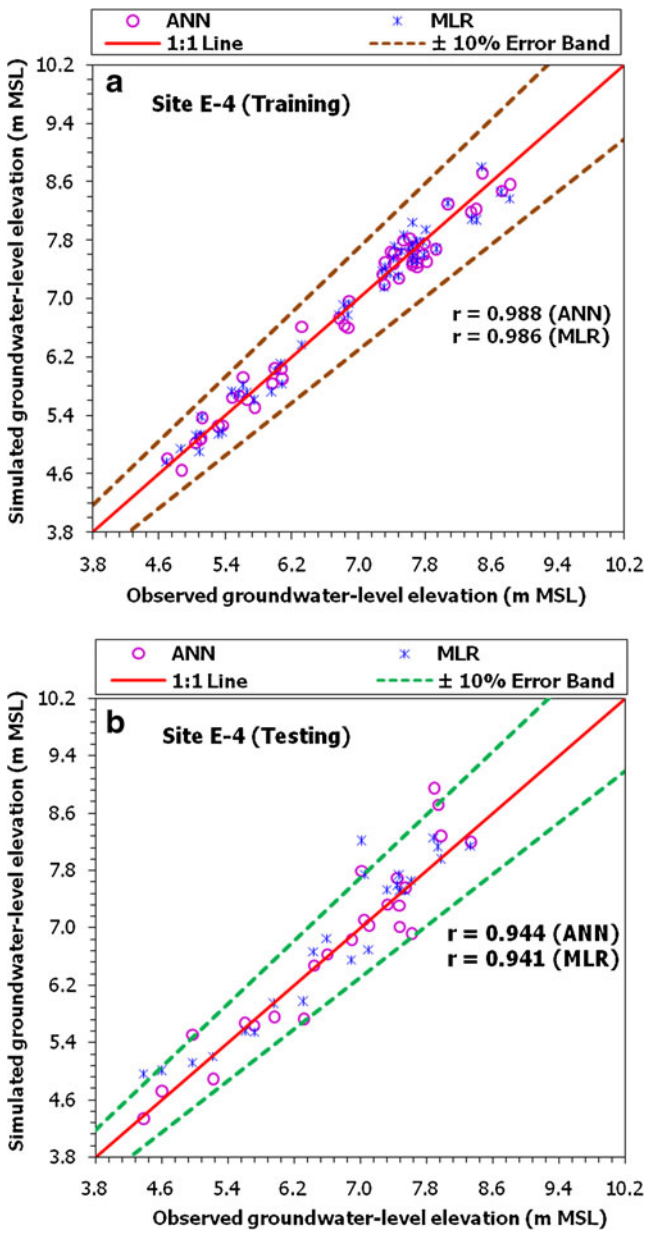

Fig. 11 Scatter plots of observed and predicted groundwater levels by *ANN* and *MLR* models with ± *10 % error band* at *site E-4*: **a** for the training period (1999–2002); **b** for the testing period (2003–2004) (r correlation coefficient between observed and simulated groundwater-level elevation)

DOI 10.1007/s10040-013-1029-5

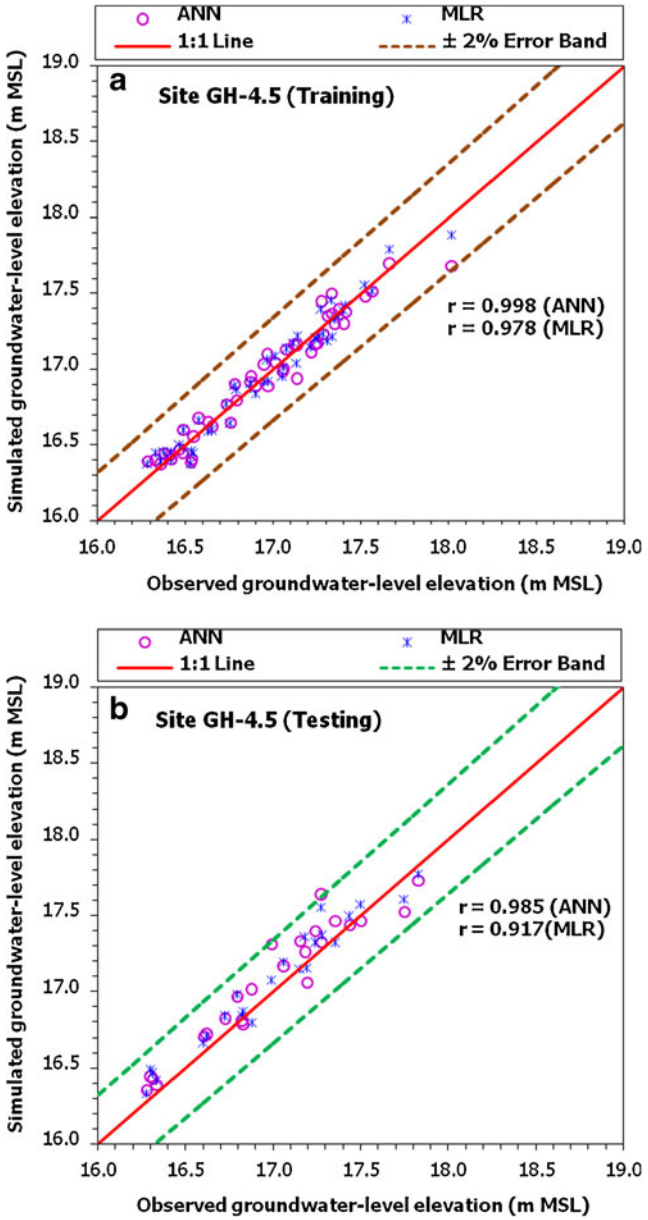

Fig. 12 Scatter plots of observed and predicted groundwater levels by *ANN* and *MLR* models with ± *2 % error band* at *site GH-4.5*: **a** for the training period (1999–2002); **b** for the testing period (2003–2004) (*r* correlation coefficient between observed and simulated groundwater-level elevation)

of the relationship between dependent and independent variables; (2) independence of the errors (no autocorrelation); (3) homoscedasticity (constant variance) of the errors with respect to the predicted values; and (4) normality of the error distribution. The plots of the standardized residuals (model errors) and predicted water tables at site GH-4.5 are shown in Fig. 4, in order to check the assumption of homoscedasticity. This figure reveals that there is no increasing spread of residuals from left to right, and hence the assumption of constant variance of errors is satisfied. The assumption of normality was examined with the help of histogram of residuals (Fig. 5) and normal probability plots (Fig. 6). In Fig. 5, the residuals are distributed in such a manner that they resemble the normal distribution curve, thereby satisfying the assumption of normality. Also, Fig. 6 reveals that a majority of the residuals fall on the straight line; it can therefore be inferred that the residuals obtained are normally distributed. The autocorrelation of residuals was checked by using the Durbin-Watson statistic (Makridakis et al. 2008). For example, the value of Durbin-Watson statistic for the residuals obtained at site GH-4.5 is 1.78, which is significant at a 95 % confidence interval, thereby satisfying the condition of no autocorrelation at lag 1 in the groundwater-level time-series of this site. These diagnostic checks were employed for all of the 17 sites, and the results indicated that all 17 MLR models satisfy the basic assumptions of MLR technique.

In addition to the basic assumptions, the degree of multicollinearity was checked using the variance inflation factor (VIF) and tolerance. As a rule of thumb, a variable with a VIF value of greater than 5, or tolerance <0.2, corresponds to high multicollinearity (Statistica 2001). Table 6 shows the check for multicollinearity at site GH-4.5 as an example, which reveals that the magnitude of VIF is less than 5 and also the tolerance is greater than 0.2 for all the independent variables used in developing the MLR models, and hence no multicollinearity exists in the independent variables used. Similarly, the multicollinearity check was employed for all 17 sites, and it was found that the multicollinearity is absent in the independent variables used for developing the MLR models. Thus, the developed MLR models are technically acceptable for predicting groundwater levels at almost all the sites, except for site C-7.

Optimal neural network parameters of the ANN models

The optimal number of hidden neurons, together with the corresponding number of generations, population size and fitness values for the 17 ANN models, is summarized in Table 7. Clearly, the optimal number of hidden neurons varies from 6 (sites E-2, E-5 and G-2) to 14 (site A-2) for the ANN architecture selected in this study. It is also evident from this table that the number of generations required to obtain optimal neural network parameters for the 17 ANN models varies from 1 (sites B-3, C-2, D-6, E-5, F-6, G-2, H-2, H-5 and I-2) to 2 (sites A-2, C-7, E-2, E-4, F-1, GH-4.5, H-3 and H-4), with the fitness values ranging between 0.0016 m^2 (site I-2) and 0.059 m^2 (site D-6).

Table 9 Percentage error bands for the ANN and MLR models of 14 sites

Error band with respect to the 1:1 line	Sites
± 2 %	A-2, C-7, D-6 and E-5
± 5 %	C-2, E-2 and H-3
± 10 %	F-1, F-6, G-2 and H-4
± 20 %	H-2, H-5 and I-2

DOI 10.1007/s10040-013-1029-5

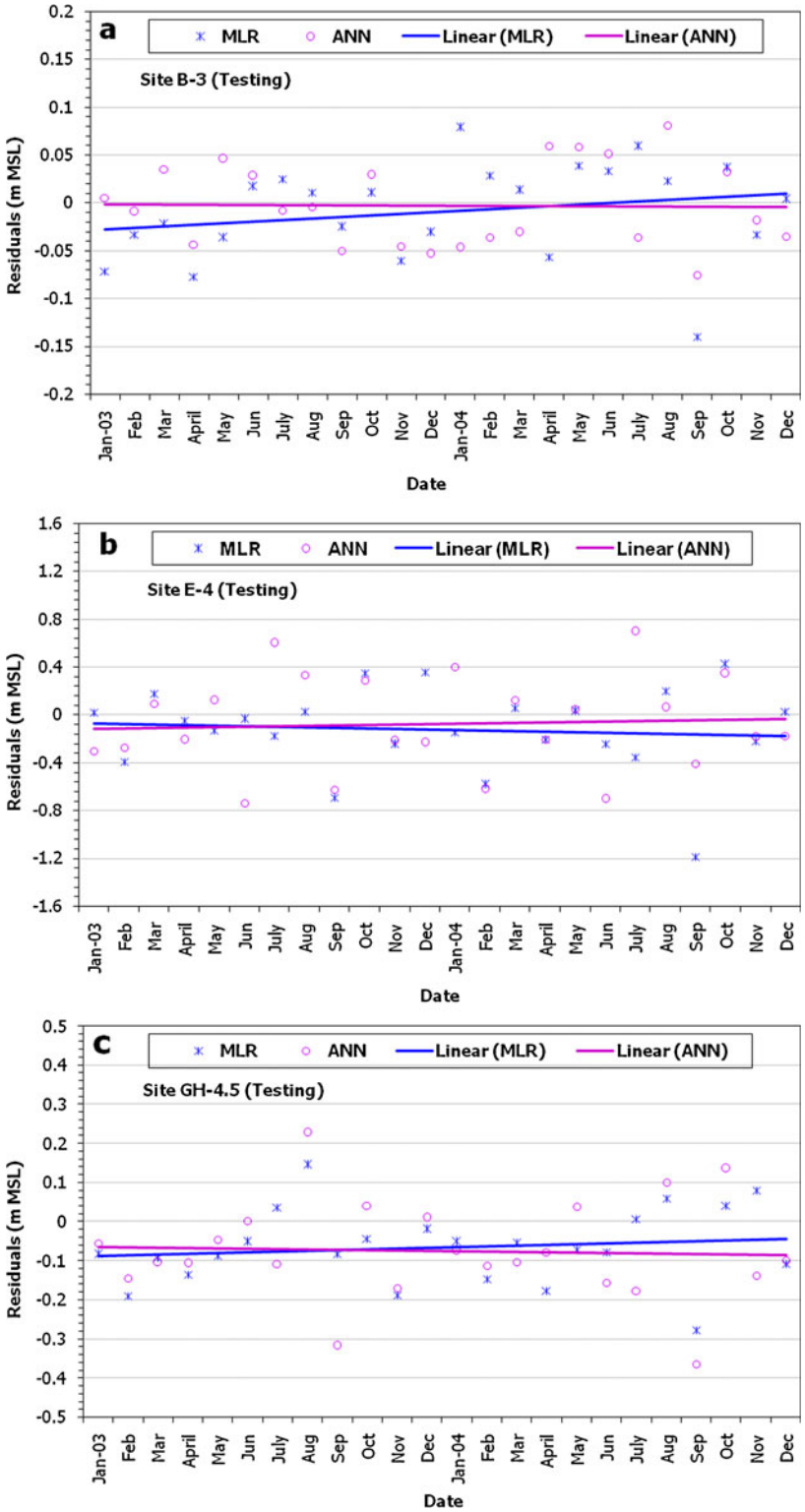

Fig. 13 Comparison of prediction errors of groundwater levels for *ANN* and *MLR* models at: **a** site B-3; **b** site E-4; **c** site GH-4.5; all for the testing period (2003–2004) (*Linear line* represents best linear fit trendline for ANN and MLR models)

Comparative performance of the ANN and MLR models

Quantitative evaluation

The performances of the 17 site-specific ANN models in predicting groundwater levels were compared with those of the corresponding MLR models during both training and testing periods by using MAE, RMSE, r^2 and NSE goodness-of-fit statistics. The results of this comparison are presented in Table 8.

Like the training results, the testing results also indicated appreciably high values of r^2 and NSE and considerably low

 DOI 10.1007/s10040-013-1029-5

Table 10 Summary of sensitivity index values and level of sensitivity (rank) for different inputs at the 17 sites

Input	Values of sensitivity index (S_k) and rank (R) at different sites																															
	A-2		B-3		C-2		C-7		D-6		E-2		E-4		E-5		F-1		F-6		G-2		GH-4.5		H-2		H-3		H-4		H-5		I-2	
	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R	S_k	R
GW_{t-1}	0.052	2	0.035	3	0.304	1	0.045	3	0.049	3	0.053	2	0.333	1	0.034	3	0.303	1	0.155	1	0.161	1	0.166	1	0.121	1	0.048	3	0.045	3	0.132	1	0.028	3
GW_{t-2}	0.061	2	–	–	0.263	1	–	–	0.062	2	0.070	2	0.070	2	0.051	2	0.066	2	0.073	2	0.033	3	0.010	4	0.022	3	0.008	4	0.007	4	0.023	3	0.052	2
Rainfall	0.073	2	0.036	3	0.124	1	*0.004*	5	0.025	3	0.024	3	0.174	1	0.037	3	0.179	1	0.073	2	0.148	1	0.075	2	0.020	3	0.022	3	0.040	3	0.047	3	0.007	4
$Rainfal_{t-1}$	0.037	3	0.014	3	0.101	1	0.026	3	*0.002*	5	*0.003*	5	0.034	3	0.099	2	0.032	3	*0.005*	5	0.124	1	0.220	1	*0.000*	5	0.025	3	0.026	3	*0.005*	5	0.023	3
$Rainfall_{t-2}$	0.012	3	*0.004*	5	0.069	2	0.005	4	*0.004*	5	*0.004*	5	0.076	2	0.022	3	0.077	2	0.019	3	0.046	3	0.142	1	0.044	3	0.007	4	0.040	3	0.008	4	0.069	2
Temperature	0.044	3	0.007	4	0.771	1	0.131	1	0.102	1	0.105	1	0.259	1	0.313	1	0.271	1	0.184	1	0.164	1	0.267	1	0.064	2	0.147	1	0.021	3	0.070	2	0.149	1
$Temp_{t-1}$	0.262	1	0.038	3	0.324	1	0.029	3	0.076	2	0.078	2	0.341	1	0.042	3	0.341	1	0.172	1	0.173	1	0.820	1	0.021	3	0.079	2	0.087	2	0.080	2	0.041	3
$Temp_{t-2}$	0.169	1	0.013	3	0.444	1	0.030	3	0.345	1	0.349	1	0.480	1	0.014	3	0.482	1	0.117	1	0.153	1	0.193	1	0.009	4	0.043	3	0.010	4	0.116	1	0.091	2
River stage	0.106	1	0.008	4	0.117	1	0.008	4	0.049	3	0.047	3	0.092	2	0.093	2	0.084	2	0.019	3	0.085	2	0.006	4	0.017	3	0.050	3	0.020	3	0.011	3	0.011	3
River $stage_{t-1}$	0.084	2	0.014	3	0.132	1	0.034	3	0.057	2	0.057	2	0.037	3	0.032	3	0.043	3	0.031	3	0.040	3	0.234	1	0.009	4	0.023	3	0.017	3	0.047	3	0.046	3
River $stage_{t-2}$	0.027	3	0.016	3	0.227	1	*0.004*	5	0.078	2	0.079	2	0.071	2	0.021	3	0.068	2	0.079	2	0.038	3	0.209	1	0.016	3	*0.004*	5	0.058	3	0.018	3	0.021	3
D_1 (Jan–Feb)	0.145	1	0.031	3	0.090	2	0.020	3	0.070	2	0.070	2	0.230	1	0.045	3	0.228	1	0.011	3	0.024	3	0.178	1	0.009	4	0.036	3	0.023	3	0.015	3	*0.004*	5
D_2 (Feb–Mar)	0.040	3	0.017	3	0.093	2	0.021	3	0.226	1	0.229	1	0.099	2	0.041	3	0.104	1	0.054	2	0.027	3	0.326	1	0.009	4	0.057	2	0.058	2	0.054	2	0.054	2
D_3 (Mar–Apr)	0.011	3	0.022	3	0.080	2	0.013	3	0.124	1	0.125	1	0.074	2	0.049	3	0.074	2	0.044	3	0.054	2	0.303	1	0.007	4	0.038	3	0.037	3	0.047	3	0.036	3
D_4 (Apr–May)	0.067	2	*0.000*	5	0.033	3	0.025	3	0.157	1	0.159	1	0.104	1	0.103	1	0.111	1	0.053	2	0.141	1	0.094	2	0.013	3	0.051	2	0.056	2	0.025	3	0.067	2
D_5 (May–Jun)	0.051	2	0.008	4	0.165	1	0.047	3	0.114	1	0.108	1	0.052	2	0.094	2	0.037	3	0.061	2	0.008	4	0.184	1	0.019	3	*0.004*	5	0.083	2	0.010	4	0.022	3
D_6 (Jun–Jul)	*0.002*	5	*0.003*	5	0.161	1	0.054	2	0.173	1	0.1741	1	0.086	2	0.029	3	0.091	2	0.042	3	0.057	2	0.224	1	*0.001*	5	0.009	4	0.040	3	*0.004*	5	0.010	4
D_7 (Jul–Aug)	0.042	3	0.008	4	0.251	1	0.070	2	0.056	2	0.055	2	0.045	3	0.035	3	0.040	3	0.094	2	0.082	2	0.331	1	0.015	3	0.028	3	0.059	2	*0.002*	5	0.009	4
D_8 (Aug–Sep)	0.034	3	0.016	3	0.227	1	0.073	2	0.070	2	0.072	2	0.161	1	0.089	2	0.168	1	0.070	2	0.055	2	0.259	1	*0.001*	5	0.031	3	0.078	2	0.033	3	0.006	4
D_9 (Sep–Oct)	0.009	4	0.013	3	0.012	3	0.037	3	0.079	2	0.079	2	0.095	2	0.024	3	0.091	2	0.097	2	0.029	3	0.069	2	0.012	3	0.028	3	0.047	3	0.064	2	0.075	2
D_{10} (Oct–Nov)	*0.003*	5	*0.004*	5	0.065	2	0.015	3	0.031	3	0.030	3	0.014	3	0.080	2	0.010	4	0.027	3	0.049	3	0.133	1	0.027	3	0.009	4	0.051	2	0.095	2	0.031	3
D_{11} (Nov–Dec)	*0.003*	5	0.008	4	0.038	3	0.007	4	0.052	2	0.052	2	0.110	1	0.027	3	0.118	1	0.009	4	0.046	3	0.007	4	0.007	4	0.007	4	0.056	2	0.064	2	0.029	3

Rank 1: very high sensitivity (S_k>0.1); Rank 2: high sensitivity (S_k=>0.05−0.1); Rank 3: moderate sensitivity (S_k=>0.01–0.05); Rank 4: low sensitivity (S_k=>0.005–0.01); Rank 5: very low sensitivity (S_k≤0.005). The rank 5 values are *italicized*

D_1–D_{11} indicate seasonal dummy variables

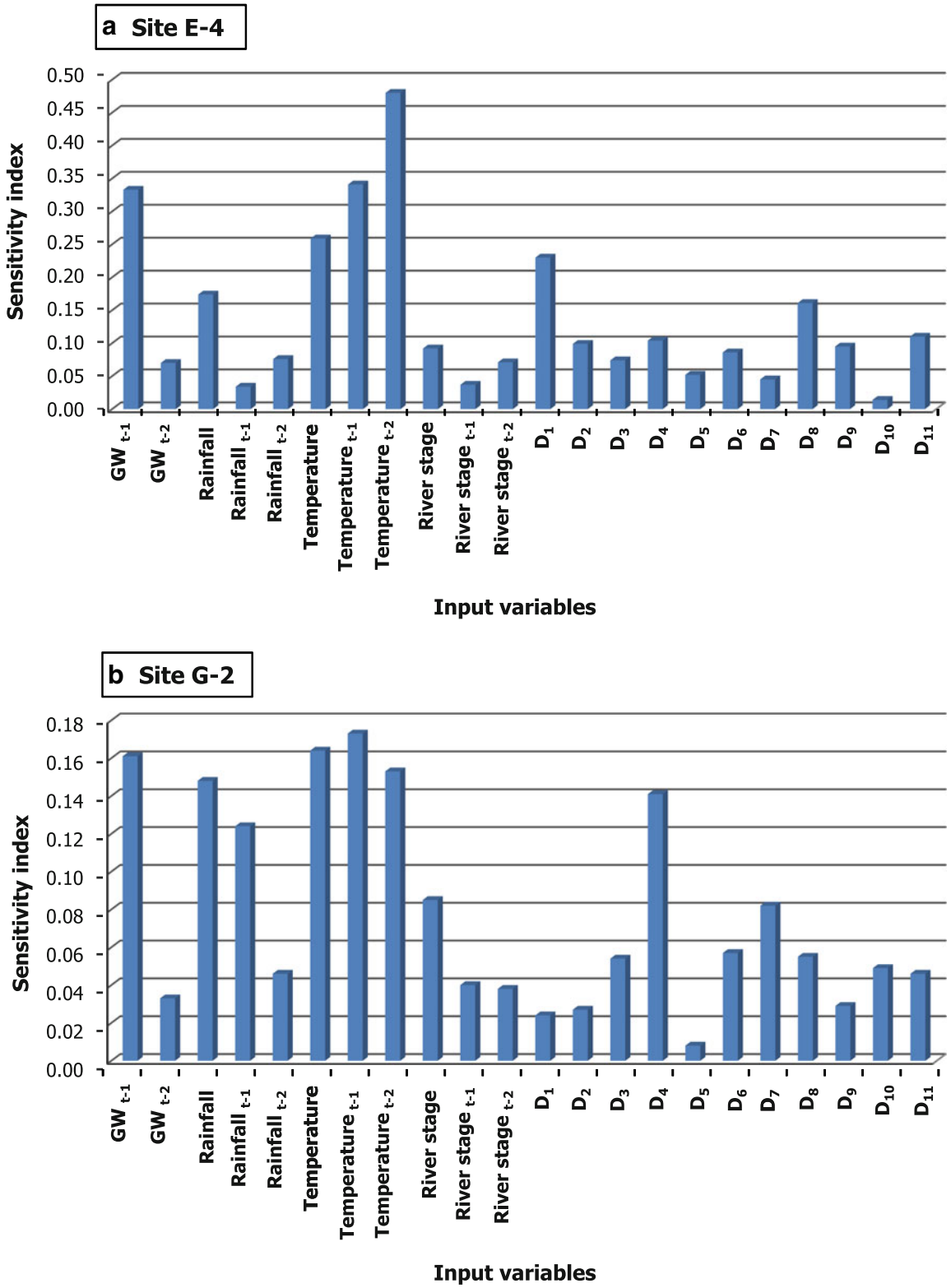

Fig. 14 Sensitivity of the MLP-LM model to the input variables for: **a** site E-4; **b** site G-2. In these graphs, D_1–D_{11} indicate seasonal dummy variables

values of MAE and RMSE for the ANN models compared to those for the MLR models. The ANN (MLP-LM) models have the highest overall prediction accuracy at site GH-4.5 (MAE=0.015 m, RMSE=0.03 m, r^2=0.998 and NSE=0.998) and site H-2 (MAE=0.017 m, RMSE=0.02 m, r^2=0.997, NSE=0.997), whereas the prediction accuracy at site C-7 is the lowest (MAE=0.021 m, RMSE=0.036 m, r^2=0.832 and NSE=0.828), but it is still acceptable,

 DOI 10.1007/s10040-013-1029-5

because both r^2 and NSE are >0.8. Here, it is interesting to note that the MLR model fails to predict reliable groundwater levels at site C-7, while the ANN model is capable of providing reasonable groundwater-level predictions at this site.

Qualitative evaluation

As far as the results of graphical indicators are concerned, a comparison of observed and simulated/predicted groundwater levels by ANN and MLR models is shown in Figs. 7a,b, 8, 9a,b at sites B-3, E-4 and GH-4.5 for training and testing periods, as examples. It is apparent from these figures that the simulated groundwater levels obtained by ANN models match reasonably well with the observed groundwater-levels, as compared to those obtained by MLR models. Besides the simultaneous plots, the match between observed groundwater levels and the groundwater levels predicted by both the models was also investigated by scatter plots with 1:1 lines and error bands. For instance, the scatter plots with 1:1 lines and error bands at sites B-3, E-4, and GH-4.5 are illustrated in Figs. 10a,b, 11, 12a,b. In these figures, the parallel lines indicate percentage upper and lower error bands with respect to the 1:1 line. Obviously, the simulated groundwater levels yielded by the ANN and MLR models fall within ±1 % error at site B-3, ±2 % error at site GH-4.5 and ±10 % error at site E-4. The error bands for the remaining 14 sites are summarized in Table 9. It is apparent from Table 9 that the groundwater levels simulated by both ANN and MLR techniques at sites A-2, C-7, D-6 and E-5 are within a much lower tolerance limit (i.e., ±2 % error band). The simulated groundwater levels by both the models at sites C-2, E-2 and H-3 lie within an error band of ±5 %, whereas they are within an error band of ±10 % at sites F-1, F-6, G-2 and H-4. However, both the models yield groundwater levels at sites H-2, H-5 and I-2 with larger errors (i.e., within 20 %), which suggests somewhat greater uncertainty in the predicted groundwater levels at these three sites. Nevertheless, a close examination of Figs. 10a,b, 11, 12a,b reveals that the spreading of the groundwater levels simulated by the MLR models with respect to the 1:1 line is larger than the groundwater levels simulated by the ANN models for both training and testing periods. Similar trends were found for the remaining sites. This suggests that the prediction accuracy of the ANN technique is better than the MLR technique.

Furthermore, the residual analysis was performed by plotting prediction errors against time, in order to detect any trend/spread present in the process, and also to assess the accuracy of the groundwater levels predicted by both techniques. As an example, the residual plots for the ANN and MLR models of sites B-3, E-4 and GH-4.5 during testing period are shown in Fig. 13a–c. It is obvious from Fig. 13a that during testing period, the residuals of the ANN model for site B-3 are evenly distributed with respect to the zero reference line, which implies that they exhibit almost no trend with respect to time, i.e., the residuals are independent in nature. In contrast, the residuals of the MLR model for site B-3 follow an increasing trend with time. At site E-4, the residuals of the ANN model show an increasing trend with time, while the residuals of the MLR model have a decreasing trend. Moreover, the residuals of ANN models for site GH-4.5 exhibit a slightly decreasing trend with zero residual line, whereas the residuals of MLR models have a distinct increasing trend with time. Similarly, the residual analyses for the remaining 14 sites were performed, and it was found that the residuals of the ANN models do not have a trend with time for almost all the sites, thereby indicating greater prediction accuracy for the ANN models, except for sites A-2, C-2, D-6, F-1, F-6, G-2 and H-5. On the contrary, the residuals of the MLR models show a distinct trend at a majority of the sites, except for sites C-7, E-5, and F-6, thereby suggesting a lower prediction accuracy for the MLR models at most sites.

Considering the quantitative and qualitative performance indicators, it was found that the ANN technique performs much better than the MLR technique. The inferior performance of the MLR models could be attributed to the fact that the MLR modeling technique is based on the simple least-square method, whereas the ANN modeling technique is based on artificial intelligence, which mimics the functioning of the human brain by using sophisticated neural networks and non-linear activation functions for establishing input–output relationships.

Results of sensitivity analysis

Table 10 presents the results of the sensitivity analyses for the ANN model, and the variations of sensitivity index values with the inputs at sites E-4 and G-2 are shown in Fig. 14a,b, as an example. It can be seen from Fig. 14a, that at site E-4, almost all the inputs have reasonably high values of sensitivity index (the rank ranging between 1 and 3), and hence they significantly affect the groundwater-level at this site. However, at site G-2, input D_5 has a low value of sensitivity index (rank 4) as compared to those for the remaining inputs (rank=1–3; Fig. 14b). Therefore, input D_5 may be discarded at site G-2. Thus, it is evident from Table 10 that at seven sites (C-2, E-4, E-5, F-1, G-2, GH-4.5, and H-4), all the inputs have a strong influence on groundwater-levels, showing greater sensitivity levels (rank <5). However, for the remaining ten sites, only a small number of inputs have very low sensitivity (rank=5), which are highlighted in Table 10. Therefore, the inputs having 'very high' to 'moderate' sensitivity should be considered with greater accuracy so as to ensure reliable prediction of groundwater levels by the ANN model.

Conclusions

This study was carried out to examine the potential of two data-driven approaches, MLR and ANN, for simulating/predicting transient groundwater levels over a groundwater basin using relevant real-world data. It employed the standard protocols of ANN and MLR modeling as well as all the pertinent and influential input variables to achieve this goal. Altogether, 17 site-specific MLR and ANN

 DOI 10.1007/s10040-013-1029-5

models were developed, using all relevant and influential input data—namely, rainfall, ambient temperature, river stage, 11 seasonal dummy variables, and influential lags of rainfall, ambient temperature, river stage and water level. The genetic algorithm (GA) technique was applied to optimize the number of hidden neurons of the ANN models. The performance of the site-specific MLR and ANN models developed for the 17 sites was assessed both quantitatively and qualitatively by using appropriate statistical and graphical indicators.

Analysis of the results of MLR and ANN modeling indicated that the developed ANN models provide more accurate prediction of groundwater levels than the MLR models at almost all the 17 sites, with considerably high values of r^2 and NSE and much lower values of MAE and RMSE. This is due to the fact that the MLR models have inherently limited prediction ability, particularly in the presence of nonlinear relationships and/or noisy data. Therefore, the ANN models developed in this study can be regarded as a robust, powerful and reliable tool for simulating the dynamic and complex behavior of groundwater in a basin. Nevertheless, the performance of the MLR models is reasonably good at most sites, except for site C-7, though their residuals indicated a trend at a majority of the sites under study. The MLR technique has important practical advantages such as the fact that implementation is much easier and less time-consuming and labor-intensive, and it requires minimal skill compared to the ANN technique. Consequently, the MLR technique can serve as an alternative and cost-effective tool for groundwater-level simulation/prediction in the situation where professional skill and time are the constraints, and the field data are favorable (i.e., of good quality and free from noise). The methodology presented in this study can be easily applied to other parts of the world as well, irrespective of the hydrogeological settings. Thus, the methodology and findings demonstrated in this study are very useful to the research community of both developing and developed nations involved in groundwater management or protection, as well as to practising hydrogeologists.

Acknowledgements The authors are grateful to the staff of the Kochi Prefecture Office and of other concerned local offices for providing necessary data and kind assistance during field investigations. The financial support to the second author from the Japan Society for the Promotion of Science (JSPS) is also gratefully acknowledged. The authors are also very much obliged to the three reviewers and the editor for their helpful comments.

References

Adeloye AJ (2009) Multiple linear regression and artificial neural network models for generalized reservoir storage–yield–reliability function for reservoir planning. J Hydrol Eng ASCE 14(7):731–738

Affandi AK, Watanabe K, Tirtomihardjo H (2007) Application of an artificial neural network to estimate groundwater level fluctuation. J Spat Hydrol 7(2):23–46

ASCE (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology) (2000a) Artificial neural networks in hydrology, part I: preliminary concepts. J Hydrol Eng ASCE 5(2):115–123

ASCE (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology) (2000b) Artificial neural networks in hydrology, part II: hydrologic application. J Hydrol Eng ASCE 5(2):124–137

Bishop C (1995) Neural networks for pattern recognition. Clarendon, Oxford, 504 pp

Castillo PA, Merelo JJ, Prieto A, Rivas V, Romero G (2000) GProp: global optimization of multilayer perceptrons using GAs. Neurocomputing 35:149–163

Coppola EA, Szidarovszky F, Poulton MM, Charles E (2003) 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 8(6):348–360

Coppola EA, Rana AJ, Poulton MM, Szidarovszky F, Uhl VW (2005) A neural network model for predicting aquifer water level elevations. Ground Water 43(2):231–241

Coulibaly P, Anctil F, Aravena R, Bobee B (2001) Artificial neural network modeling of water table depth fluctuations. Water Resour Res 37(4):885–896

Daliakopoulos IN, Coulibaly P, Tsanis IK (2005) Groundwater level forecasting using artificial neural network. J Hydrol 309:229–240

Dekker SC, Bouten W, Schaap MG (2001) Analyzing forest transpiration model errors with artificial neural networks. J Hydrol 246:197–208

Feng S, Kang S, Huo Z, Chen S, Mao X (2008) Neural networks to simulate regional ground water levels affected by human activities. Ground Water 46:80–90

Gerken WC, Purvis LK, Butera RJ (2006) Genetic algorithm for optimization and specification of a neuron model. Neurocomputing 69:1039–1042

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, New York, pp 1–145

Haan CT (2002) Statistical methods in hydrology, 2nd edn. The Iowa State University Press, Ames, IA, 496 pp

Hagan MT, Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Network 5:989–993

Hagan MT, Demuth HB, Beale MH (1996) Neural network design. PWS, Boston, MA

Hani A, Lallahem S, Mania J, Djabri L (2006) On the use of finite difference and neural-network models to evaluate the impact of underground water overexploitation. Hydrol Processes 20:4381–4390

Haykin S (1994) Neural Networks. Macmillan College Publishing, New York

Heuvelmans G, Muys B, Feyen J (2006) Regionalisation of the parameters of a hydrological model: comparison of linear regression models with artificial neural nets. J Hydrol 319:245–265

Hodgson FDI (1978) The use of multiple linear regression in simulating ground-water level responses. Ground Water 16(4):249–253

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Networks 2:359–366

Izady A, Davary K, Alizadeh A, Ghahraman B, Sadeghi M, Moghaddamnia A (2012) Application of "panel-data" modeling to predict groundwater levels in the Neishaboor Plain, Iran. Hydrogeol J 20:435–447

Jha MK, Chikamori K, Kamii Y, Yamasaki Y (1999) Field investigations for sustainable groundwater utilization in the Konan basin. Water Resour Manag 13:443–470

Jones AJ (1993) Genetic algorithms and their applications to the design of neural networks. Neural Comput Appl 1:32–45

Krishna B, Rao YRS, Vijaya T (2008) Modeling groundwater levels in an urban coastal aquifer using artificial neural networks. Hydrol Processes 22:1180–1188

Lallahem S, Mania J, Hani A, Najjar Y (2005) On the use of neural networks to evaluate groundwater levels in fractured media. J Hydrol 307:92–111

Legates DR, McCabe GJ (1999) Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resour Res 35(1):233–241

Looney CG (1996) Advances in feed-forward neural networks: demystifying knowledge acquiring black boxes. IEEE Trans Knowledge Data Eng 8(2):211–226

Maier HR, Dandy GC (1998) Understanding the behavior and optimizing the performance of back-propagation neural networks: an empirical study. Environ Modeling Softw 13(2):179–191

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environ Modeling Softw 15:101–124

Makridakis S, Wheelwright SC, Hyndman RJ (2008) Forecasting methods and applications, 3rd edn. Wiley, Singapore, 656 pp

Mayer DG, Belward JA, Burrage K (2001) Robust parameter settings of evolutionary algorithms for the optimization of agricultural systems models. Agric Syst 69:199–213

McCuen RH, Rawls WJ, Whaley BL (1979) Comparative evaluation of statistical methods for water supply forecasting. Water Resour Bull 15(4):935–947

Mohanty S, Jha MK, Kumar A, Sudheer KP (2009) Artificial neural network modeling for groundwater level forecasting in a river island of eastern India. Water Resour Manag 24(9):1845–1865

Moré JJ (1978) The Levenberg-Marquardt algorithm: implementation and theory. In: Watson GA (ed) Lecture notes in mathematics 630. Springer, Berlin, pp 5–116

Nayak PC, Rao YRS, Sudheer KP (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. Water Resour Manag 20:77–90

Nikolos IK, Stergiadi M, Papadopoulou MP, Karatzas GP (2008) Artificial neural networks as an alternative approach to groundwater numerical modeling and environmental design. Hydrol Processes 22(17):3337–3348

Nourani V, Mogaddam AA, Nadiri AO (2008) An ANN-based model for spatiotemporal groundwater level forecasting. Hydrol Processes 22:5054–5066

Press WH, Teukolski SA, Vetterling WT, Flannery BP (1992) Numerical recipes in Fortran, 2nd edn. Cambridge University Press, New York, pp 675–683

Principe JC, Euliano NR, Lefebvre WC (2000) Neural and adaptive systems: fundamentals through simulations. Wiley, New York, 656 pp

Reed P, Minsker B, Goldberg DE (2000) Designing a competent simple genetic algorithm for search and optimization. Water Resour Res 36(12):3757–3761

Rumelhart DE, Durbin R, Golden R, Chauvin Y (1995) Backpropagation: the basic theory. In: Chauvin Y, Rumelhart DE (eds) Backpropagation: theory, architectures, and applications. Erlbaum, Hillsdale, NJ, pp 1–34

Sethi RR, Kumar A, Sharma SP, Verma HC (2010) Prediction of water table depth in a hard rock basin by using artificial neural network. Int J Water Resour Environ Eng 2(4):95–102

Shamseldin AY (1997) Application of a neural technique to rainfall-runoff modeling. J Hydrol 199:272–294

Sietsma J, Dow RJF (1991) Creating artificial neural networks that generalize. Neural Networks 4:67–69

Sinnakaudan SK, Ghani AA, Ahmad MSS, Zakaria NA (2006) Multiple linear regression model for total bed material load prediction. J Hydraul Eng ASCE 132(5):521–528

Statistica (2001) STATISTICA: system reference. StatSoft, Tulsa, OK, 1098 pp

Trichakis IC, Nikolos IK, Karatzas GP (2009) Optimal selection of artificial neural network parameters for the prediction of a karstic aquifer's response. Hydrol Processes 23:2956–2969

Uddameri V (2007) Using statistical and artificial neural network models to forecast potentiometric levels at a deep well in South Texas. Environ Geol 51:885–895

 DOI 10.1007/s10040-013-1029-5