



Copyright © 2014, Paper 18-06; 68951 words, 7 Figures, 0 Animations, 1 Tables.
<http://EarthInteractions.org>

Estimation of Subsurface Temperatures in the Tattapani Geothermal Field, Central India, from Limited Volume of Magnetotelluric Data and Borehole Thermograms Using a Constructive Back-Propagation Neural Network

Anthony E. Akpan*

Applied Geophysics Programme, Physics Department, University of Calabar, Calabar,
Cross River State, Nigeria

Mahesh Narayanan and T. Harinarayana¹

National Geophysical Research Institute, Hyderabad, India

Received 28 May 2013; accepted 26 October 2013

ABSTRACT: A constructive back-propagation code that was designed to run as a single-hidden-layer, feed-forward neural network (SLFFNN) has been

initialization conditions that can optimally solve nonlinear problems using the random weight initialization approach. Available one-dimensional (1D) MT-derived resistivity data and borehole temperature records from the Tattapani geothermal field, central India, were collated and digitized at 10-m intervals. The two datasets were paired to form a set of input-output pairs. The paired data were randomized, standardized, and partitioned into three mutually exclusive subsets. The various subsets had 52% (later increased to 61%), 30%, and 18% (later reduced to 9%) for training, validation, and testing, respectively, in the first and second training phases. The second training phase was meant to assess the influence of the training data volume on network performance. Standard statistical techniques including adjusted coefficient of determination (R^2_a), relative error (e), absolute average deviation (AAD), root-mean-square error (RMSE), and regression analysis were used to quantitatively rate network performance. A manually designed two-hidden-layer, feed-forward network with 20 and 15 neurons in the first and second layers was also adopted in solving the same problem. Performance ratings were observed to be 0.97, 3.75, 4.09, 1.41, 1.18, and 1.08 for R^2_a , AAD, e , RMSE, slope, and intercept, respectively, compared to an e of 20.33 observed with the manually designed network. The SLFFNN is thus a structurally flexible network that performs better in spite of the small volume of data used in testing the network. The network needs to be tested further.

KEYWORDS: SLFFNN; Magnetotellurics; Resistivity; Borehole thermograms; CBP; Performance; Tattapani, India

1. Introduction

The discovery of the artificial neural network (ANN) technology from a confluence of different scientific and non-scientific disciplines has given birth to a powerful research tool with applications cutting across disciplines. In the geoscience world, the ANN technique has enjoyed enhanced patronage in generating precise information on many geoscientific problems that border on control, prediction, inversion, classification, pattern recognition, and data compression, which were hitherto solved with mathematical and statistical techniques with limited accuracy. Information generated by the ANN procedure is free from the known limitations of the pre-existing, empirically based modeling tools because the ANN technique does not need mathematical equations to guide its operation. Instead, the ANN extracts information directly from a series of input-output data pairs without any prior assumptions about their nature and interrelations (Mandalet al. 2009; Moghaddam et al. 2010; Spichak et al. 2011). To make reliable predictions, the requisite condition that the ANN method needs while solving complicated problems is to have a large volume of

property estimation and characterization (Calderón-Macías et al. 2000; Amínian and Ameri 2005), lithologic boundary discrimination (Maiti and Tiwari 2009, 2010), disaster forecasting (Marzban and Stumpf 1996; Marzban 2000; Böse et al. 2008), weather prediction (Hayati and Mohebi 2007; Hayati and Shirvany 2007; Jin et al. 2008), and many others. The awareness created by the capabilities of the ANNs in solving problems with diverse complexity has attracted immense interest from researchers in both applications and model designs.

Neural networks have the capacity to solve a wide range of problems that was hitherto not satisfactorily solved using the traditional mathematical or statistical tools irrespective of the inherent level of complexity in the problem. The feed-forward, multilayer neural networks (FFMNNs) that utilize function approximation technique in solving problems are very outstanding in solving complex problems because of its structural flexibility, good capabilities, and presence of a large pool of algorithms (Negnevitsky 2005; Stathakis 2009; Sharma and Chandra 2010). Neural network architecture and topology are the major factors that influence the accuracy of results, generalization capacity, and training speed (Stathakis 2009; Sharma and Chandra 2010). Stathakis (Stathakis 2009) has documented the procedures for selecting optimal network topology needed for tackling any problem. They include the traditional trial and error, heuristic search, exhaustive search, and constructive method, as well as using a synergy between neural networks and genetic algorithms, which automatically searches for optimally performing topology based on a novel fitness function.

The conventional trial and error method is the most popular training method commonly used by many ANN users. The procedure involves manual selection of a network topology that is usually implemented using a trial and error approach. Such networks sometimes lack the needed capacity to yield optimal results especially if designed by inexperienced users (Stathakis 2009) and compulsorily requires a large volume of data for use in training the network. Performance optimization is a very important consideration that must be made while planning to design a multilayer perceptron (MLP) network that can optimally solve problems. Networks with minimal complexity and optimal performance capability are the most desirable for the ANN user community. Automated design network architecture has recently been introduced and is becoming very popular considering the level of desirability (Maiti et al. 2011). A major problem that has continued to plague the manual design approach is that all factors that can influence network performance (network topology; the nature of the applicable activation function for each node; type of optimization method; and the values of the training parameters such as epoch, learning rate, momentum, initial weight adjustment procedure, and so on) must be properly set before such networks can perform optimally. Other-

objective is not to develop a new method but to adapt an existing constructive back-propagation (CBP) neural network algorithm to estimate subsurface temperatures from a small volume of MT data and borehole thermograms. A single, hidden multilayer perceptron feed-forward neural network (SLM LPNN), which works automatically using the principle of random weight initialization process, was implemented in the neural network toolbox available in MATLAB (Demuth and Beale 2002). The constructive neural networks (CNNs) belong to a group of algorithms with adaptive structure. Such algorithms normally solve problems from the primitive perceptron model and dynamically increase the node population in steps of one until the maximum number computed from the data structure is attained (Ash 1989; Parekh et al. 2000; Sharma and Chandra 2010). This architectural adaptation can only stop if any of the preset stopping conditions (the upper bound number of nodes calculated from the size of the training dataset, the initialization conditions, and the minimum threshold performance) have been attained. In any case, the best optimized solution will be returned.

2. A review of some methods of estimating subsurface temperature

Researchers have always wanted to develop a dependable relationship that can be used to estimate temperature of Earth's interior from its physical properties. Consequently, many empirical and semi-empirical relationships have been developed and are usually used in estimating subsurface temperatures although some with limited applicability. Archie's law (Archie 1942) is a popular empirical relationship. The modified form of Archie's law connects the bulk electrical conductivity of a formation (s_b) with the conductivity of the saturating fluid (s_w) and the solid rock matrix according to Equation (1) as

$$s_b = a s_w^m S^n s_s, \quad (1)$$

where a , m , and n are formation-dependent constants; S is the saturation index; and s_s is the conductivity of the rock matrix, which was assumed to be nonexistent in the original Archie's law. The conductivity of the solid phase of a rock matrix (s_s) has been found to be dependent on its absolute temperature especially under laboratory conditions according to the semiconductor equation as

$$s_s = s_o \exp\left(\frac{2 E_a}{kT}\right), \quad (2)$$

where s_o is a constant that represents the conductivity of a theoretically infinite temperature. E_a is the activation energy and k is Boltzmann's constant (Meu-

property relationship. Such relationships include a temperature-resistivity relationship of altered basalt that Flvenz et al. (Flvenz et al. 1985) introduced in their work. The relationship is given by

$$r_T = \frac{r_o}{[1 + a(T - T_o)][1 + b(T - T_o)]} \quad (3)$$

where r_T and r_o are the resistivities at temperature T and some reference temperature T_o . The empirical constants a and b are also defined at temperature T_o .

Harinarayana et al. (Harinarayana et al. 2006) used the empirical relationship shown in Equation (4) to estimate subsurface temperature in Puga region of India,

$$T_z = T_s + \frac{Q_s}{k} + \frac{A_s Z^2}{2k} \quad (4)$$

where T_z is the temperature ($^{\circ}\text{C}$) at depth Z (km), T_s is the surface temperature ($^{\circ}\text{C}$), Q_s is the surface heat flux, k is the thermal conductivity of the Earth materials, and A_s is the rate of radioactive heat production. Insufficient knowledge about the nature of the materials inside Earth's interior and the associated prevailing extent of heterogeneity have not permitted reliable estimates to be made. In some cases, absolute lack of correlation between results made under laboratory and field conditions degraded researcher's confidence on the workability of these assumption-laden relationships.

Spichak et al. (Spichak et al. 2007; Spichak et al. 2011) and Spichak and Zakharova (Spichak and Zakharova 2009) reported an entirely new approach of estimating subsurface temperatures from magnetotelluric (MT) data and adjacent borehole (BH) thermograms. Their procedure was based on using an ANN that has been trained on the basis of correspondence between the MT data and BH thermograms. An ANN is a functional imitation of the fascinating and highly successful processing powers of the natural biological neuron. Just like the biological neurons, ANNs have the capability to accept multiple inputs from several sources, integrate, and simultaneously process all the inputs using its parallel processing capabilities (Basand Boyaci 2007; Maiti and Tewari 2010). The ANNs are nonlinear and entail a completely data-driven, modern data modeling procedure that does not require any initial model to stand on. It is now an efficient tool for modeling complex and nonlinear geophysical problems that cannot be successfully handled by conventional empirical equations (Maiti et al. 2007; Maiti et al. 2011). Estimates made by the ANNs are reliable because they are free from problems associated with using empirical relations since the ANNs do not require any mathematical equation to work with but rather extract information directly from sets of the input-output pairs of

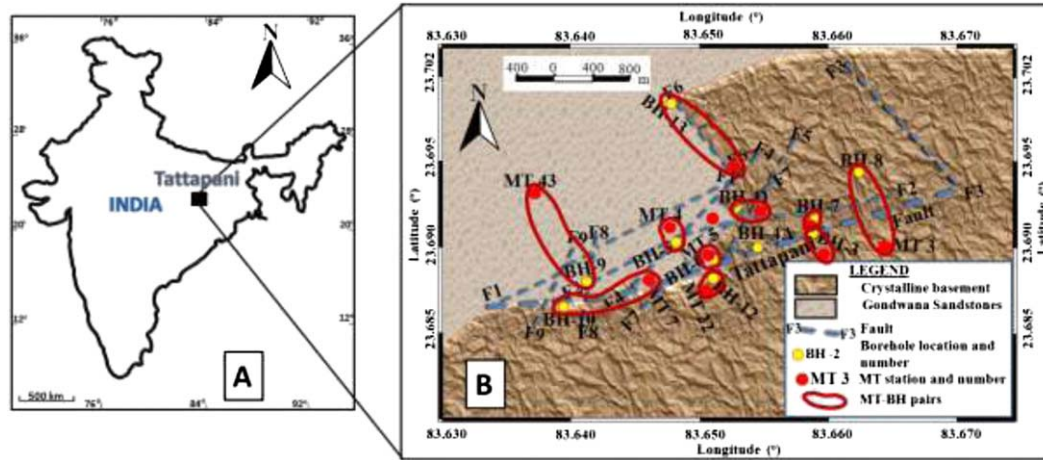


Figure 1. Map of India showing the (a) location and (b) generalized geology of the Tattapani region, borehole locations, and MT stations.

area are in the form of hot-water springs discharging hot water at temperatures of 50–85°C, with siliceous sinter deposited at the base of the springs (Shanker et al. 1987; Jain et al. 1995). All these manifestations occur in marshy grounds and hydrothermally altered clay zones, covering an area of about 0.1 km² (Shanker et al. 1987).

The study area falls within the southern margin of Tattapani-Ramkola coalfield, at the contact of Archaean rocks with the Lower Gondwana rocks. The thermal activity is controlled by the east-northeast-west-southwest Tattapani fault and other northeast-southwest trending cross faults. The Tattapani fault separates the Archaean rocks exposed in the southern side from the lower Gondwana group exposed toward the northern and northwestern parts of the fault (see Figure 1). The Proterozoic basement rocks comprise gneisses, diorites, biotite schists, actinolite-trimolite schists, kyanite-sillimanite schists, granulites, amphibolite bands of phyllites, quartzites, and graphites in some places while the lower Gondwana group consists of sandstones and shales. Effects of shearing are clearly manifested in the rocks near the main Tattapani fault. Most of the numerous hot springs present in the area are located along the Tattapani fault (Sarolkar and Das 2006).

4. MT and borehole temperature data acquisition

can be found in Harinarayana et al. (Harinarayana et al. 2000) and Veeraswamy and Harinarayana (Veeraswamy and Harinarayana 2006). Despite these desirable data treatments, all the datasets were visually checked for problems associated with static shift by directly checking for consistency and correlating adjacent datasets. The lack of static shift problems in the curves that were finally used in this study (see Figure 2) is in accordance with the findings of Harinarayana et al. (Harinarayana et al. 2000). In spite of the seemingly lack of static shift problems as deduced from qualitative assessment and correlation of the curves, the data were still subjected to quantitative dimensionality assessment. From the results of the quantitative dimensional analysis, any data that were observed to be under the influence of 3D structures were summarily dropped. Dimensionality assessment of the data was performed by using the rotationally invariant amplitude skewness ($Skew_S$) parameter of Swift (Swift 1967). According to Swift (Swift 1967), the $Skew_S$ parameter computed from Equation (5) should ideally be 0 for 1D and 2D structures but, in environments where the subsurface is dominated by 3D structures, the $Skew_S$ parameter will ideally be greater than zero ($Skew_S > 0$) (Rybin et al. 2008),

$$Skew_S = \frac{Z_{yx1} Z_{yy}}{Z_{xy2} Z_{yx}} \quad (5)$$

Results observed are shown in Figure 3 where it can be seen that for frequencies above 1 Hz, the $Skew_S$ values are well below 0.1 (< 0.1) (the dashed line in Figure 3), which is suggestive of a subsurface that is dominated by 1D/2D structures. $Skew_S$ values observed for frequencies below 1 Hz were generally greater than 0.1 ($Skew_S > 0.1$), suggesting that 3D structures dominate the subsurface at these depths. Based on the results of earlier MT studies of the area (Harinarayana et al. 2000), regional tectonics, and the observed $Skew_S$ values, the shallow subsurface was assumed to be predominantly 1D. The vertical magnetic field (H_z) records that were observed to be noisy were also dropped.

Some MT data were randomly selected and some of the data processing procedures already conducted by the MTs data processing unit of NGR I were repeated. The repetition was done in order to compare the two sets of results and consequently develop confidence in the other datasets. The data processing steps that we repeated included robust single-station processing to down weight or discard outliers of impedance estimates at each frequency band of measurement (Sutarno and Vozoff 1989; Jupp and Vozoff 1997); transforming the remaining data from time domain to frequency domain using the fast Fourier transform algorithm; and computing cross correlations between pairs of mutually orthogonal magnetic

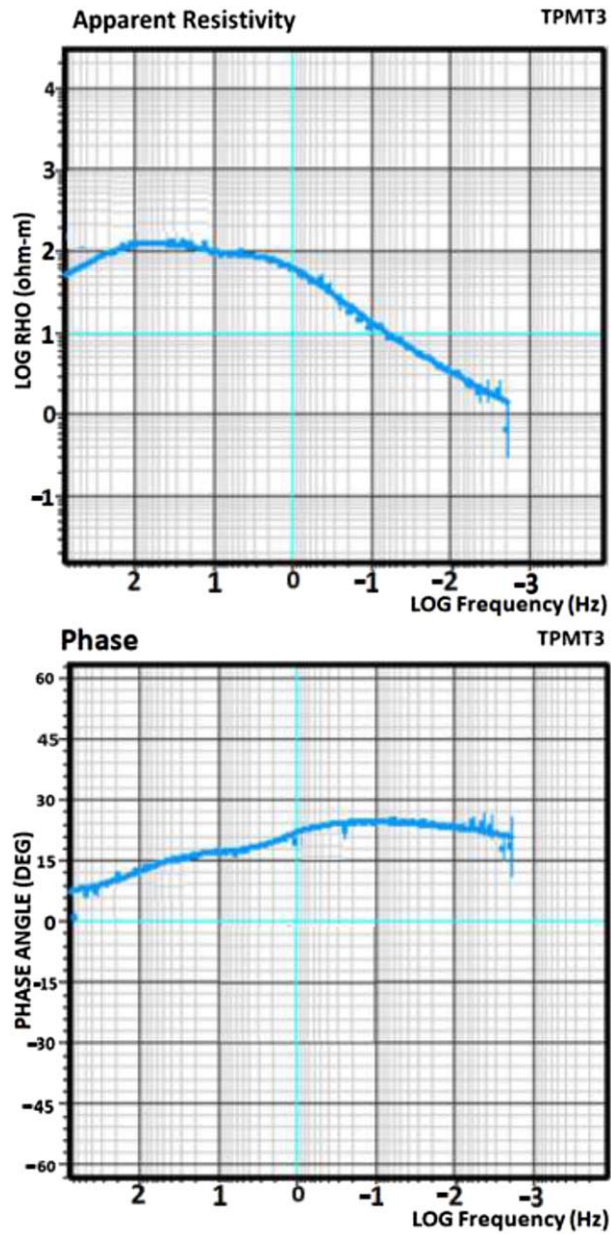


Figure 2. Typical MT resistivity and phase curves.

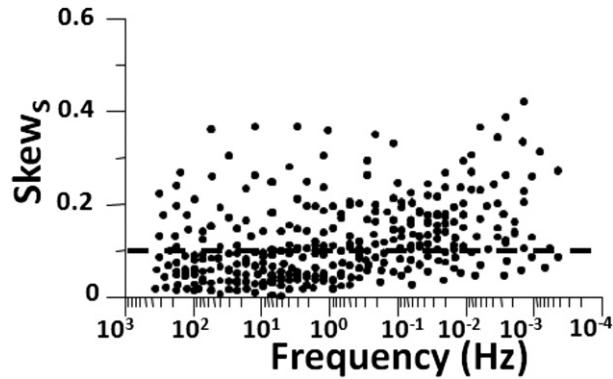
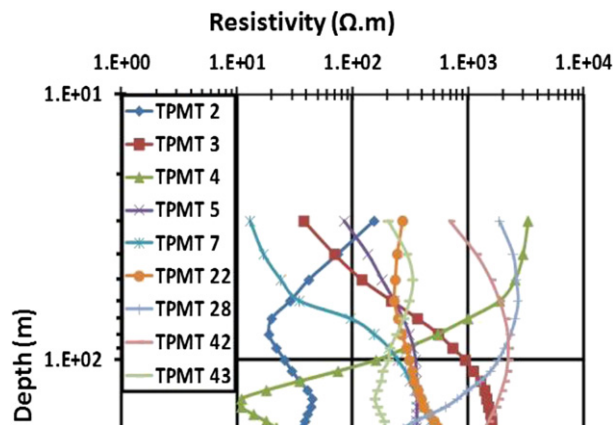


Figure 3. Graph showing variation of Skew_s values with frequency.

Ten borehole thermograms recorded within the shallow depth range of 100–500 m in the study area were acquired from published work of Shanker et al. (Shanker et al. 1987). Most of the temperature recording stations selected were located in close proximity to the MT stations (Figure 1b). The resistivity and temperature variation with depth curves (Figure 5) were digitized in a 10-m interval. Maximum temperature observed from the temperature records was 112°C at 500-m depth, although higher temperature values have been reported in the same area at shallower depths (Sarojkar and Das 2006). Heatflow measurements have been conducted in the area



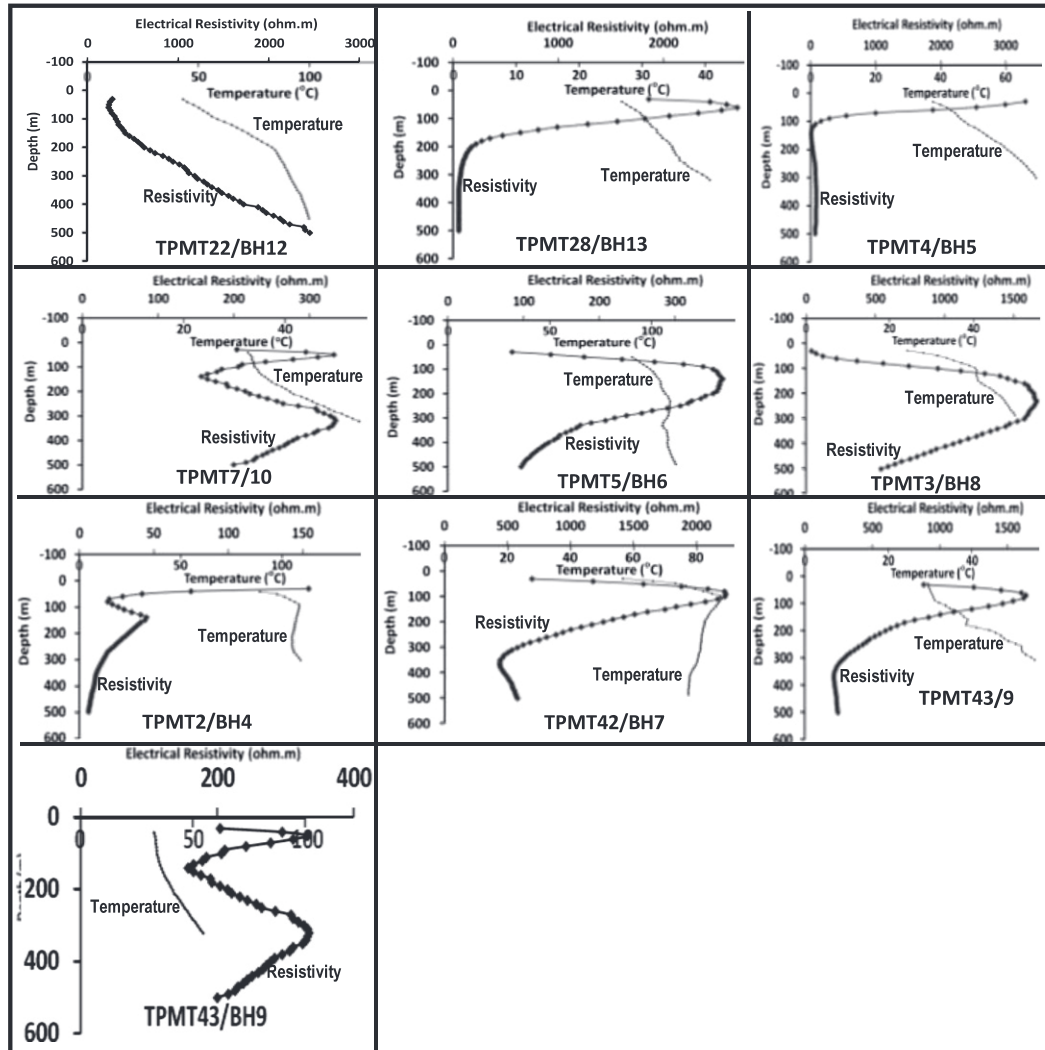


Figure 5. Graphs of M T-derived resistivity and subsurface temperatures.

by researchers such as Shanker et al. (Shanker et al. 1987) and Jain et al. (Jain et al. 1995), and results show that peak values of heat fluxes are as high as 300 m W m^{-2} . Average heat flow flux is about 190.6 m W m^{-2} which is much higher than the

and Yeung (Kwok and Yeung 1997), Parekh et al. (Parekh et al. 2000), Islam and Murase (Islam and Murase 2001), and Sharma and Chandra (Sharma and Chandra 2010). The CNNs usually start solving problems with minimum network topology (minimum number of layers, nodes in the hidden layer, and connections) and dynamically increase the node, layer, and connection population by unity as may be required before a problem is optimally solved. The cascade-correlation algorithm (CCA), dynamic node creation (DNC), and hybrid algorithms are famous members of the CNN family (Islam and Murase 2001; Sharma and Chandra 2010).

The DNC algorithm (Ash 1989) is an efficient member of the CNN family and was introduced to overcome the problems associated with learning using the back-propagation algorithm including the problem of slow learning rate and rigid topology, the step size problem, the local minima problem, and the moving target problem. The DNC algorithm originally constructs a single, hidden-layer artificial neural network with a zero node in the hidden layer (the naive model); tests the naive network on the problem; and, in the event of failure, automatically adds one node (or batch of nodes) to the existing network and restarts the tests all over. Once an optimum number of nodes in the hidden layer have been found, it is connected to the output side weights. Advantages of using the DNC algorithm include rapid learning rate, establishment of a self-modified network size and topology, and the ability to retain its inbuilt structures even when there is a change in the training dataset. Furthermore, it does not require back propagation of failure signals. The original DNC algorithm has undergone several modifications because it was observed to have difficulty in learning complex problems (Islam and Murase 2001).

The CBP algorithm, for instance, is a modified version of DNC algorithm introduced by Lehtokangas (Lehtokangas 1999). CBP is fast becoming a very popular tool because, apart from enjoying all the benefits attributed to the DNC, it is known to have a simple implementation procedure. The CBP also has the advantage of being suitable for use in fixed size networks as well as its documented ability to utilize stochastic optimization routines (Sharma and Chandra 2010). The CBP is usually designed with only one hidden layer and the traditional method of back propagating the error signal is retained, thereby creating a single-layer, feed-forward neural network (SLFFNN). Such advantages increase the generalization potential of SLFFNN with minimal network architecture and thus influenced our choice in our present study.

6. Data pairing, scaling, and model parameterization

Testing of the network was performed using two different datasets: 1) temperature variations with depth derived from borehole thermograms and 2) resistivity

can be compelled to work in interpolation mode. Moderately wide MT-BH separation (≈ 3 km) was deliberately introduced into the pairing procedure so that geographic factors can be incorporated into the input data and, consequently, forced the network to function in extrapolation mode. The moderately widely spaced set of MT-BH data pairs was included in order to test the function approximating powers of the good network when an optimum architecture has been found (Whiteson and Stone 2006).

The raw data in the database consist of inputs and corresponding correct outputs (targets). All the raw data in the database X_i were converted to standard scores Z_s using the mean m and standard deviation s of the entire dataset (Larsen and Marx 2000; Carroll and Carroll 2002) [see Equation (6)] before they were partitioned into the three mutually exclusive subsets of training, validation, and test sets,

$$Z_s = \frac{X_i - m}{s} \quad (6)$$

The standardization process was performed in order to ensure that all inputs were scaled down to values that vary between -2.1 and 1.1 . This procedure is a standard precautionary measure usually adopted in ensuring that the wide variations in the original values of the input dataset do not force the network to go into saturation while processing the data (Maiti and Tiwari 2010). The "prestd.m" code that is available in the neural network toolbox of MATLAB was used to achieve this. The training subset was further preprocessed using the "prepca.m" code in order to filter out redundant data in the training dataset. Additionally, the validation and test datasets were preprocessed using the "transmit.m" code.

Partitioning the data into three mutually exclusive blocks was necessary in order to guard against the problem of data overfitting associated with back-propagation algorithms. The phenomenon of overfitting can occur if the network "overlearns" and consequently "memorizes" the training dataset such that it cannot make good generalization when confronted with new datasets. Demuth and Beale (Demuth and Beale 2002), Maiti et al. (Maiti et al. 2007; Maiti et al. 2011), Maiti and Tiwari (Maiti and Tiwari 2010), and many other researchers have described how the early stopping technique can improve the generalization capability of a network by eliminating overfitting problems. By utilizing the early stopping method, a threshold learning condition can be set, under which the network can stop the learning process. The early stopping procedure ensures that the network stops learning whenever the best-fitting model beyond which the network cannot make any reasonable increase in the error function is attained. In practice, this is done by randomizing all input data and partitioning it into three mutually exclusive subsets

and 35% for testing. The validation set was used by the network to fine tune its topology with a view to sharpening and clearing any gray areas in its predictive powers (Negnevitsky 2005; Larochelle et al. 2009). It is not directly used for training but is deployed for continuous monitoring of its error during the process of training such that, if the error attains a preset minimum condition, it stops further training of the network. This cross-validation scheme prevents the network from memorizing the training dataset and consequently left in a state that it can make acceptable generalization (Maiti et al. 2007). The testing dataset that consists of data from MT7-BH10 and MT11-BH30 MT-BH data pairs were either fed into the network in full correspondence with a training dataset of 174 (or 52%) or fed into the network one by one. In the latter option, the remaining parts of the testing dataset were usually added to the training dataset, thereby increasing the percentage of the training dataset to 61%.

Threshold upper performance benchmark was fixed at 10% and the initialization conditions that consist of 35 trials for any of the randomly generated numbers (RGNs) were set as variables. The network was set into running mode (training) by using a conditional statement to search for the best fitting initialization conditions [number of trials (N trials) and the preset random number] for a particular number of nodes in the hidden layer. These initialization conditions were used by the network to test the adequacy of a range of nodes (N) in the hidden layer (in our data, N varies from 0 to 9) for optimal performance by the network. For each initialization condition and number of nodes in the hidden layer, the network performance was rated. In the event of the performance index satisfying the preset threshold performance target, the training process terminates and results were displayed in both tabular and graphical forms. Besides the computed performance rating, visual inspection of the graphically presented results was also useful in qualitatively assessing the performance of the network.

7. Network design and adaptation

The neural network toolbox package in MATLAB (Demuth and Beale 2002) was used in designing a structurally flexible SLFFNN. The network was optimally designed to take in four inputs and automatically search for the best performing network topology to solve the problem for each given set of initialization conditions (N trials and random number). A generalized form of the SLFFNN code that was modified to fit into our problem can be found online (at http://www.math-works.com/matlabcentral/newsreader/view_thread/308972). The various steps that the SLFFNN code used in executing the commands have been described in the work of Negnevitsky (Negnevitsky 2005) and Sharma and Chandra (Sharma and

performance of each random number so that, at the end of each round of trial, the best (or near best) performing random number can be selected.

8. Methodology of subsurface temperature estimation

The problem of designing a structurally flexible network that can be used to estimate subsurface temperatures from a small volume of MT-BH data pairs of data, which the conventionally manually designed MLP networks cannot solve satisfactorily, is what this study is out to address. Consequently, we attempted to solve the problem using both the conventionally manually designed MLP and the SLFFNN. A static structured MLP network was designed and tested with the data. The input layer of the manually designed network had five inputs followed by two hidden layers while the output layer had one fixed neuron. The number of neurons in the first hidden layer was fixed at 20 while the number of neurons in the second hidden layer was 15. The learning rate (α) and momentum (β) were kept at 0.01 and 0.9, respectively. The MLP network was taught repeatedly with the goal of attaining a threshold performance benchmark of 5% using error as performance indicator.

A structurally flexible, single-layer, feed-forward neural network was designed. The actual architecture of the intended feed-forward, back-propagation MLP was not predefined before training began. However, the structure of the input and output layers were uniquely known since they depend on the structure of the input and output data pairs. Thus, the input layer was structurally designed to have five neurons. The neurons were designed to take in three position coordinates of nearby boreholes and MT locations, resistivity, and temperature values at different depths. The position coordinates consist of average values of longitudes and latitudes of the MT and borehole positions and depth and one neuron in the output layer. The single neuron in the output layer was responsible for the transmission of all results (temperature) from the network to the outside world.

The neurons in the hidden layer acts individually as a summing junction that integrates and modifies all the inputs that it has received to produce a net output that can be expressed (Demuth and Beale 2002; Hagan et al. 2002; Negnevitsky 2005; Maiti et al. 2007) as

$$\text{net}_j^{(l)} = \sum_{i=1}^N w_{ji}^{(l)} x_i^{(l-1)} + b_j^{(l)}, \quad (7)$$

where $\text{net}_j^{(l)}$ is the net output from the j th neuron in a hidden layer (l), $x_i^{(l-1)}$ is an input variable for the i th node in hidden layer (l) as received from another node in the $(l-1)$ th layer, $w_{ji}^{(l)}$ is the strength of the connection weight between the i th

transformed from their original plus and minus infinity variations to their binary equivalents. For instance, if the net inputs into node (j) in the hidden layer (l) is $net_j^{(l)}$, then the log sigmoid transfer function ensures that the output from this node j is transformed according to Equation (8) as

$$a_j^{(l)} = f_j(net_j^{(l)}), \quad (8)$$

where f_j is the log-sigmoid transfer function and is mathematically defined as

$$f_j = \frac{1}{1 + e^{-x}}, \quad (9)$$

where x is an input parameter and e denotes the natural logarithm (Benaouda et al. 1999; Van der Baan and Jutten 2000; Negnevitsky 2005; Spichak et al. 2011). The linear transfer function does not transform the outputs from the output neurons in any way but rather ensures that they are transmitted directly with their original variations preserved.

In the output layer, the computed network outputs at node j (O_j) are compared with the known outputs (D_j) and the overall network error E summed over all the output nodes, and the samples were internally computed using

$$E = \frac{1}{2} \sum_{j=1}^N (D_j - O_j)^2. \quad (10)$$

The local error from only the jth node in hidden layer (l) can be calculated according to Benaouda et al. (Benaouda et al. 1999) and Maiti et al. (Maiti et al. 2007) as

$$e_j^{(l)} = \frac{\partial E}{\partial (net_j^{(l)})} = f_j'(net_j^{(l)}) \sum_k e_k^{(l+1)} w_{kj}^{(l+1)}. \quad (11)$$

In the network training process, the network minimizes E by systematically changing the strength of the connection weights Dw_{ji} using the failure history as guide. This operation is implemented by the gradient descent learning rule according to the relation

$$Dw_{ji}^{(l)} = \eta h_{ji}^{(l)} \frac{\partial E}{\partial w_{ji}^{(l)}} = \eta h_{ji}^{(l)} e_j^{(l)} x_i^{(l+1)}, \quad (12)$$

where η is the learning rate

the same (Demuth and Beale 2002; Sin et al. 2006; Maiti et al. 2007; Moghaddam et al. 2010). The R^2 values are computed using the expression

$$R^2 = \frac{\sum_{i=1}^N (T_{known,i} - T_{estimated,i})^2}{\sum_{i=1}^N (T_{known,i} - \bar{T}_{known,i})^2}, \quad (13)$$

where T_{known} is the known temperature at a given depth observation point i , $T_{predicted}$ is the estimated temperature at the same observation point i , \bar{T}_{known} is the mean of the T_{known} , and N is the total number of observation data points in the dataset (Gemitzi et al. 2009; Moghaddam et al. 2010). For a given random number R^2 , other performance tests were calculated for all N trials made by the network to find a suitable structure for the network architecture. The standard deviation, mean, median, and variance of all R^2 values were also computed and plotted against number of nodes in the hidden layer (see Figure 6, top and middle). This was done to check the robustness and sensitivity of the predictions made by the network to the different datasets.

The RMSE is calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (T_{known,i} - T_{estimated,i})^2}{N}}, \quad (14)$$

while e is computed as

$$e = \frac{\sum_{i=1}^N (T_{known,i} - T_{estimated,i})^2}{\sum_{i=1}^N T_{known,i}^2} \times 100\% . \quad (15)$$

Finally, the AAD is computed as

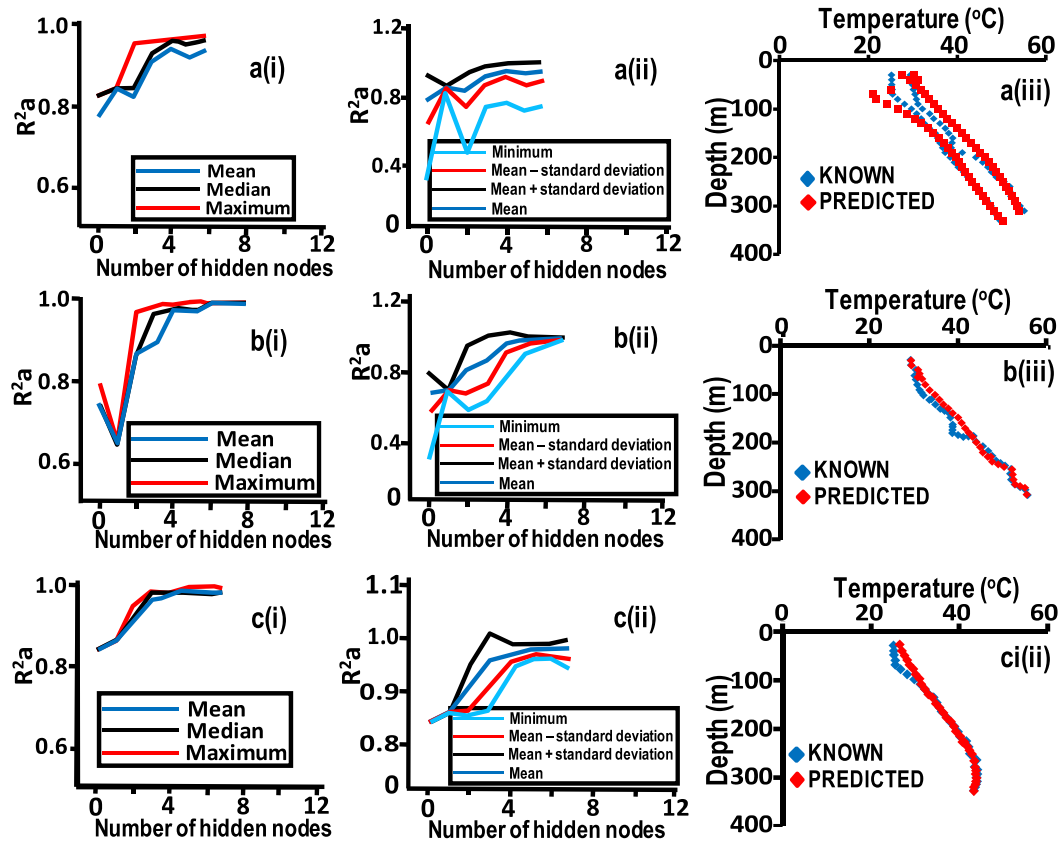


Figure 6. (top) Graphs showing variation in network performance assessed using R^2_a as a function of number of nodes in the hidden layer for (a) TPM T7-BH10 and TPM T30-BH11 pairs combined, (b) TPM T30-BH11, and (c) TPM T7-BH11; (middle) statistical analyses of network performance for (a) TPM T7-BH10 and TPM T30-BH11 pairs combined, (b) TPM T30-BH11, and (c) TPM T7-BH11; and (bottom) graphs of estimated and known temperatures for (a) TPM T7-BH10 and TPM T30-BH11 pairs combined, (b) TPM T30-BH11, and (c) TPM T7-BH11.

Posttraining regression analysis was also performed by testing the dataset to assess how the known (actual) output and the predicted network outputs are quantitatively related, formally establishing their reliability limit. Demuth and

10. Discussion of results

The performance of an SLFFNN MLP has been optimized and adapted for solving a nonlinear problem of subsurface temperature estimation from a limited volume of MT-BH data pairs using the CBP algorithm. The CBP algorithm bequeaths to the network architecture an adaptable structure, which enables it to automatically adjust its topology to meet with the challenges of the problem to be solved. The static-structured manually designed network performs well during training but fails to make good generalization when confronted with the test data (see Figure 7). This observation suggests that either the volume of the training dataset might not have been large enough for proper learning or the number of hidden neurons that was manually fixed in the network for the purpose of acquiring the needed intelligence for use in solving the problem might have caused the network to overlearn and consequently fail to make good generalization when confronted with the testing dataset. The possibility of the latter option being responsible for the poor performance of the manually designed network was ruled out because the performance of the network was satisfactory during the validation phase. Although it has been pointed out by many researchers including Lawrence (Lawrence 1994), Helle et al. (Helle et al. 2001), Stathakis (Stathakis 2009), and others that there is no actual rule to guide users in selecting an optimal number of hidden layers and neurons, it thus appears that allowing the network to adjust its topology to cope with the challenges of the data to be solved is a better practice in network design.

Solving problems from any dataset containing limited volume of data has continued to pose serious challenges to the neural network users' community since neural networks require a large volume of data pairs from which to learn (see Moghaddam et al. 2010). Therefore, a structurally flexible network can be deployed to solve a wide range of problems including those involving limited volume of data, although more tests need to be performed to ascertain the level of reliability. The automated search for the optimal number of neurons in the hidden layer from zero (the naive model) with corresponding increase in R^2 values toward unity suggest that the network is structurally flexible.

Figure 6 (bottom) shows graphical plots of estimated and known subsurface temperatures from the SLFFNN. The SLFFNN network seems to have performed well even with the limited volume of data used in training the network. The gradual build up in the number of neurons in the hidden layer and their corresponding performance can be seen in the computed values of R^2 plotted in Figure 6 (top). The R^2 values approach unity when the maximum number of neurons in the hidden layer computed from the size of the training dataset has been loaded in each

tuatbns in network performance for different volumes of input data .

Data volumes (%)		Initialization conditions		Regression results			Statistical results		
Validation	Testing	Random number	N trials	Slope	Intercept	R ²	AAD	e	RMSE
30	18	90151941	33	0.96	0.91	0.97	2.29	4.09	1.14
30	9	50151942	12	0.94	1.03	0.99	2.45	3.05	1.27
30	9	80151940	14	1.08	2.118	0.99	3.75	3.19	1.47

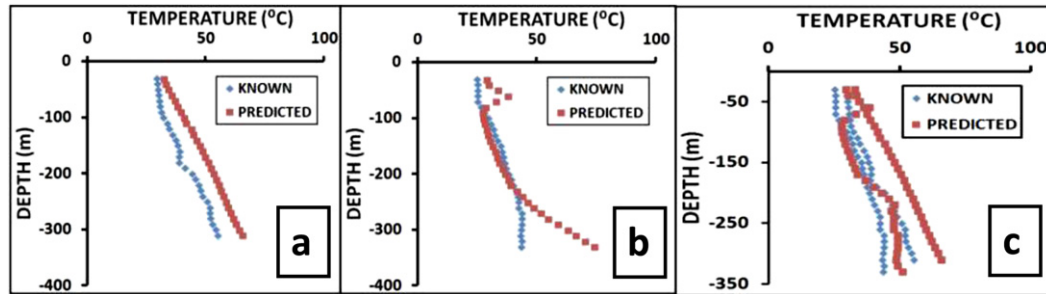


Figure 7. Graphs of known temperatures against annually estimated temperatures at (a) TPM T30-BH11, (b) TPM T7-BH10, and (c) combination of TPM T30-BH11 and TPM T7-BH11 pairs.

readings, even with the low volume of data used for training the network, is suggestive of a high approximating capacity network that is typical of reinforcement learning in neural networks (Sutton and Barto 1998; Ferrari and Stengel 2005; Zainuddin and Pauline 2008). Therefore, a better performance level that is associated with the random weight initialization and degrees of freedom procedure that this procedure used in tackling the problem can be achieved when the volume of data is large enough to properly train the network.

Changes in subsurface structural geology (e.g., faults, fractures, etc.) and hydrogeological conditions are known to adversely affect the normal subsurface conductivity distribution pattern (see Spichak et al. 2011). Consequently, estimates of subsurface conditions made in such heterogeneous environments are likely to be fraught with errors and therefore unreliable (Spichak et al. 2011). Function approximating procedures that are usually implemented by the SLFFNN could serve as a reliable tool in predicting subsurface conditions in such heterogeneous environments. Neural networks with adaptive structures have been reported to be characterized by high function approximating capability and consequently high-performance ratings (Sharma and Chandra 2010; Qing-Lai et al. 2010; Nicoletti et al. 2009). The ability of the SLFFNNs to optimally estimate subsurface temperatures in spite of the constraints imposed by performance limiting factors such as the small volume of the training dataset, the geological complexity prevalent at the site and the sometimes wide separation of the MT-BH data pairs is indicative of a high-performance network. Such level of nonlinear problem solving capability, even with a small size of data, is characteristic of networks with good

quantity of the training dataset to 62%. The results of the repeated trainings for each half and full dataset are shown in Table 1. In all the cases, better performance is obtained when the MT–BH data pairs used for training the network was more than 60%. In the case of our limited volume of dataset, the tabulated results (see Table 1) try to capture the little improvement in network performance with increase in the volume of the training dataset. This indicates that better predictions can always be obtained from the network if a greater number of MT–BH pairs is used in the network training phase. Consequently, the anticipated influence of changes in subsurface structural, hydrogeological, and geological conditions can be drastically reduced since the network is expected to generate enough neurons that can tackle the problem head on. Thus, the training methodology has a significant influence on the accuracy of the predictions made by the network.

In the process of searching for optimal initialization conditions, the network tests all available nodes computed from the size of the training data structure using the random weight initialization and degree of freedom scheme. The network usually starts the testing process with a zero node (the naive model) in the hidden layer and runs the test until all the nodes are exhausted. This process may last for some time, although it is considerably less than the time usually spent in training manually designed networks by the trial and error method (see Table 1 in Moghaddam et al. 2010). For instance, a continuously running set of 35 trials for each random number and node population in the hidden layer usually returns satisfactory predictions within 45 min. Once such conditions are found, the process of training the network is usually fast with repeatable results. Thus, the approach, unlike what has been reported in the literature (see Moghaddam et al. 2010), is very economical with time, although the overall time tends to be dependent on the desired level of accuracy. The search for optimal initialization conditions to solve the problem normally starts with the network attempting to solve the problem without any node in the hidden layer corresponding to Rosenblatt's primitive perceptron (constant or naive) model (Rosenblatt 1962; Negnevitsky 2005). If the residual prediction error is greater than the preset tolerance limit after 35 attempts (N trials) for a particular number of nodes in the hidden layer and random number, the network automatically increases the node population by one and restarts the search all over again. The network will keep on trying until the upper bound number of nodes in the hidden layer computed from the size of the training data structure has been exhausted. The upper bound number of nodes was 6 and 9 when the quantity of training data was 52% and 61%, respectively. For each N trials made with a specified number of nodes in the hidden layer, the network returns a local minimum that corresponds to a convergence point with the best value of adjusted coefficient of determination (R^2_a). Thus, by prefixing

11. Conclusions

We have demonstrated how a SLFFNN can be used to estimate subsurface temperatures using a small volume of resistivity data and borehole thermograms from the Tattapani geothermal field in central India. The network uses the random weight initialization and adjusted degrees of freedom approach in searching for optimal performing initialization condition for the network. The network has the capability of dynamically changing the network topology during training in order to cope with the challenges of the problem. The network performed satisfactorily well in spite of the performance limiting constraints imposed by the limited volume of MT-BH data pairs used in training the network, the nonlinearity of the problems to be solved, the wide separation of the MT-BH pairs, and the prevailing geological and structural complexities at the Tattapani geothermal field in central India.

The structural adaptability of the SLFFNN is evident in the network's continuous and desperate search for optimum network architecture to solve the problem. The network attempted solving the problem with no node in the hidden layer corresponding to Rosenblatt's primitive perceptron (the naive or constant) model; if the predicted result does not satisfy some preset performance conditions, it automatically increases the node population by one and repeats the trials all over. The trial can only stop if either an acceptable solution is found or the upper bound number of nodes generated from the dimension of the training dataset is exhausted, in which case it will return the near-best solution.

Acknowledgments. The first author is grateful to the Council for Scientific and Industrial Research (CSIR) and Third World Academy of Sciences (TWAS) for cosponsoring and funding his research at the National Geophysical Research Institute (NGRI), Hyderabad, India, under the CSIR-TWAS postdoctoral fellowship. Thanks are also due to the director of NGRI for his encouragements, understanding, and interest and for kindly granting us permission to publish this work. We benefited immensely from Prof. Greg Heath of MATLAB Central, who generously shared his single-hidden-layer multilayer perceptron neural network code with us. We are also grateful to our colleagues in the Magnetotelluric Division of NGRI for their support, concern, and encouragement. Contributions, suggestions, and critical review made by the reviewers that shaped the original manuscript to the present form are also appreciated.

References

Ali Akcayol, M., and C. Cinar, 2005: Artificial neural network based modeling of heated catalytic converter performance. *Appl Therm Eng* 25, 2341–2350. doi:10.1016/

- Ash, T., 1989: Dynamic node creation in back-propagation. *Connect. Sci.*, 1, 365–375, doi:10.1080/09540098908915647.
- Bas, D., and I. H. Boyaci, 2007: Modeling and optimization II: Comparison of estimation capabilities of response surface methodology with artificial neural networks in a biochemical reaction. *J. Food Eng.*, 78, 846–854, doi:10.1016/j.jfoodeng.2005.11.025.
- Benaouda, D., G. Wadge, R. B. Whitmarsh, R. G. Rothwell, and C. MacLeod, 1999: Inferring the lithology of borehole rocks by applying neural network classifiers to downhole logs: An example from the Ocean Drilling Program. *Geophys. J. Int.*, 136, 477–491, doi:10.1046/j.1365-246X.1999.00746.x.
- Böse, M., F. Wenzel, and M. Erdik, 2008: PreSEIS: A neural network-based approach to earthquake early warning for finite faults. *Bull. Seismol. Soc. Amer.*, 98, 366–382, doi:10.1785/0120070002.
- Calderón-Macias, C., M. K. Sen, and P. L. Stoffa, 2000: Artificial neural networks for parameter estimation in geophysics. *Geophys. Prospect*, 48, 21–47, doi:10.1046/j.1365-2478.2000.00171.x.
- Carroll, S. R., and D. J. Carroll, 2002: *Statistics Made Simple for School Leaders: Data-Driven Decision Making*. Rowman & Littlefield, 146 pp.
- Chen, D., and J. Yang, 2005: Robust adaptive neural control applied to a class of nonlinear systems. *Proc. 17th IMACS World Congress, Paris, France, IMACS, T5-I-01-0911*.
- Constable, S. C., R. L. Parker, and C. G. Constable, 1987: Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 52, 289–300, doi:10.1190/1.1442303.
- Dai, H., and C. MacBeth, 1994: Split shear-wave analysis using an artificial neural network. *First Break*, 12, 605–613.
- Demuth, H., and M. Beale, 2002: *Neural network toolbox for use with MATLAB handbook. The MATLAB user's guide*, 154 pp.
- Fenari, S., and R. F. Stengel, 2005: Smooth function approximation using neural networks. *IEEE Trans. Neural Netw.*, 16, 24–38, doi:10.1109/TNN.2004.836233.
- Flóvenz, Ó. G., L. S. Geirgsson, and K. Á. Mason, 1985: Resistivity structure of the upper crust in Iceland. *J. Geophys. Res.*, 90, 10136–10150, doi:10.1029/JB090iB12p10136.
- Gemitzi, A., C. Petalas, V. Pinaras, and V. A. Tsahrintzis, 2009: Spatial prediction of nitrate pollution in groundwaters using neural networks and GIS: An application to South Rhodope aquifer (Thrace, Greece). *Hydrol. Processes*, 23, 372–383, doi:10.1002/hyp.7143.
- Gentili, S., and A. Michelini, 2006: Automatic picking of P and S phases using a neural tree. *J. Seismol.*, 10, 39–63, doi:10.1007/s10950-006-2296-6.
- Geotools, 1997: *Users guide*. Geotools Rep., 448 pp.
- Hagan, T. H., H. B. Demuth, and M. Beale, 2002: *Neural Network Design*. PWS Publishing, 734 pp.
- Harinarayana, T., and Coauthors, 2000: Magnetotelluric investigations in Tatapani geothermal region, Surguja district, Madhya Pradesh, India. *MNES Project Rep.*, 127 pp.
- , K. K. Abdulazeez, K. Naganjaneyulu, C. Manoj, K. Veerasamy, D. N. Murthy, and S. P. E. Rao, 2004: Magnetotelluric studies in Puga valley geothermal field, NW Himachal Pradesh, India. *J. Volcanol. Geotherm. Res.*, 138, 405–424, doi:10.1016/j.jvolgeores.2004.07.011.

- Islam, M. M., and K. Murase, 2001: A new algorithm to design compact two-hidden-layer artificial neural networks. *Neural Networks*, 14, 1265–1278, doi:10.1016/S0893-6080(01)00075-2.
- Jain, S. C., K. K. Nair, and D. B. Yedekar, 1995: Geology of the Son-Narmada-Tapti lineament zone in central India. *Geoscientific Studies of the Son-Narmada-Tapti Lineament Zone*, Geological Survey of India, 1–154.
- Jin, L., C. Yao, and X.-Y. Huang, 2008: A nonlinear artificial intelligence ensemble prediction model for typhoon intensity. *Mon. Wea. Rev.*, 136, 4541–4554, doi:10.1175/2008MWR2269.1.
- Jupp, D., and K. Vozoff, 1997: Reply by the authors to F. J. Esparza and E. Gómez-Treviño. *Geophysics*, 62, 692, doi:10.1190/1.1487031.
- Kwok, T.-Y., and D.-Y. Yeung, 1997: Objective functions for training new hidden units in constructive neural networks. *IEEE Trans. Neural Networks*, 8, 1131–1148, doi:10.1109/72.623214.
- Lagoudakis, M. G., and R. Parr, 2003: Least-squares policy iteration. *J. Mach. Learn. Res.*, 4, 1107–1149.
- Larochelle, H., Y. Bengio, J. Louradour, and P. Lamblin, 2009: Exploring strategies for training deep neural networks. *J. Mach. Learn. Res.*, 10, 1–40.
- Larsen, R. J., and M. L. Marx, 2001: *An Introduction to Mathematical Statistics and Its Applications*. 3rd ed. Prentice Hall, 790 pp.
- Lawrence, J., 1994: *Introduction to Neural Networks: Design, Theory and Applications*. California Scientific Software Press, 348 pp.
- Lehtokangas, M., 1999: Modeling with constructive backpropagation. *Neural Networks*, 12, 707–716, doi:10.1016/S0893-6080(99)00018-0.
- Maiti, S., and R. K. Tiwari, 2009: A hybrid Monte Carlo method based artificial neural networks approach for rock boundaries identification: A case study from the KTB bore hole. *Pure Appl. Geophys.*, 166, 2059–2090, doi:10.1007/s00024-009-0533-y.
- , —, and —, 2010: Automatic discriminations among geophysical signals via Bayesian neural network approach. *Geophysics*, 75, E67–E78, doi:10.1190/1.3298501.
- , —, and H. J. Kijmpel, 2007: Neural network modeling and classification of lithofacies using well log data: A case study from KTB borehole site. *Geophys. J. Int.*, 169, 733–746, doi:10.1111/j.1365-246X.2007.03342.x.
- , G. Gupta, V. C. Enam, and R. K. Tiwari, 2011: Inversion of Schlumberger resistivity sounding data from the critically dynamic Koyna region using the hybrid Monte Carlo-based neural network approach. *Nonlinear Processes Geophys.*, 18, 179–192, doi:10.5194/npg-18-179-2011.
- Manal, S., P. V. Sivaprasad, S. Venugopal, and K. P. N. Murthy, 2009: Artificial neural network modeling to evaluate and predict the deformation behaviour of stainless steel type AISI 304L during hot torsion. *Appl. Soft Comput.*, 9, 237–244, doi:10.1016/j.asoc.2008.03.016.
- Manoj, C., and N. Nagarajan, 2003: The application of artificial neural networks to magnetotelluric time-series analysis. *Geophys. J. Int.*, 153, 409–423, doi:10.1046/j.1365-246X.2003.01902.x.
- Marquardt, D. W., 1963: An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, 11, 431–441, doi:10.1137/0111030.

- Negnevitsky, M., 2005: Artificial Intelligence: A Guide to Intelligent Systems. 2nd ed. Pearson, 415 pp.
- Nicoletti, M. C., J.R. Bertini Jr., D. Elizondo, L. Franco, and J.M. Jerez, 2009: Constructive neural network algorithms for feedforward architectures suitable for classification tasks. Constructive Neural Networks, D. Elizondo et al., Eds., Studies in Computational Intelligence, Vol. 258, Springer, 1-23.
- Parekh, R., J. Yang, and V. Honavar, 2000: Constructive neural-network learning algorithms for pattern classification. IEEE Trans. Neural Netw., 11, 436-451, doi:10.1109/72.839013.
- Qing-Lai, W., Z. Hua-Guang, L. De-Rong, and Z. Zhao Yan, 2010: An optimal control scheme for a class of discrete-time nonlinear systems with time delays using adaptive dynamic programming. Acta Autom., 46, 121-129.
- Rosenblatt, F., 1962: Principles of Neurodynamics: Perceptions and the Theory of Brain Mechanisms. Spartan, 616 pp.
- Rybin, A. K., V. V. Spichak, V. Y. Batalev, E. A. Bataleva, and V. E. Mamyukov, 2008: A magnetotelluric soundings in the active seismic area of northern Tien Shan. Russ. Geol. Geophys., 49, 337-349, doi:10.1016/j.rgg.2007.09.014.
- Sarolkar, P. B., and A. K. Das, 2006: Reservoir studies at Tatapani geothermal field, Surguja district, India. Proc. 31st Workshop on Geothermal Reservoir Engineering, Stanford, CA, Stanford University, SGP-TR-179.
- Shanker, R., J. L. Thussu, and J. M. Prasad, 1987: Geothermal studies at Tattapani hot spring area, Sarguja district, central India. Geothermics, 16, 61-76, doi:10.1016/0375-6505(87)90079-4.
- Sharma, S. K., and P. Chandra, 2010: Constructive neural networks: A review. Int. J. Eng. Sci. Technol., 2, 7847-7855.
- Simpson, F., and K. Bahr, 2005: Practical Magnetotellurics. Cambridge University Press, 254 pp.
- Sin, H. N., S. Yusof, N. A. H. Shilkh, and R. R. Abdu, 2006: Optimization of enzymatic clarification of sapodilla juice using response surface methodology. J. Food Eng., 73, 313-319, doi:10.1016/j.jfoodeng.2005.01.031.
- Singh, U. K., R. K. Tiwari, and S. B. Singh, 2005: One-dimensional inversion of geo-electrical resistivity sounding data using artificial neural networks— A case study. Comput. Geosci., 31, 99-108, doi:10.1016/j.cageo.2004.09.014.
- Spichak, V. V., 2006: Estimating temperature distributions in geothermal areas using a neuronet approach. Geothermics, 35, 181-197, doi:10.1016/j.geothermics.2006.01.002.
- , —, and O. K. Zakharova, 2009: The application of an indirect electromagnetic geothermometer to temperature extrapolation in depth. Geophys. Prospect., 57, 653-664, doi:10.1111/j.1365-2478.2008.00778.x.
- , —, and A. K. Rybin, 2007: Possibility of realization of contact-free electromagnetic geothermometer. Dokl. Earth Sci., 417, 1370-1374, doi:10.1134/S1028334X07090176.
- , —, and —, 2011: Methodology of the indirect temperature estimation basing on magnetotelluric data: Northern Tien Shan case study. J. Appl. Geophys., 73, 164-173, doi:10.1016/j.jappgeo.2010.12.007.
- Stathakis, D., 2009: How many hidden layers and nodes? Int. J. Remote Sens., 30, 2133-2147, doi:10.1080/01431160802549278.

- Whiteson, W., and P. Stone, 2006: Evolutionary function approximation for reinforcement learning. *J. Mach. Learn. Res.*, 7, 877–917.
- Zainuddin, Z., and O.N.G. Pauline, 2008: Function approximation using artificial neural networks. *WSEAS Trans. Math.*, 6, 333–338.

Earth Interactions is published jointly by the American Meteorological Society, the American Geophysical Union, and the Association of American Geographers. Permission to use figures, tables, and brief excerpts from this journal in scientific and educational works is hereby granted provided that the source is acknowledged. Any use of material in this journal that is determined to be "fair use" under Section 107 or that satisfies the conditions specified in Section 108 of the U.S. Copyright Law (17 USC, as revised by P.L. 94-553) does not require the publishers' permission. For permission for any other form of copying, contact one of the copublishing societies.
