Geophys. J. Int. (2000) **142,** 15–26

Artificial neural network inversion of magnetotelluric data in terms of three-dimensional earth macroparameters

Vjacheslav Spichak and Irina Popova

Geoelectromagnetic Research Institute RAS, 142090 *Troitsk, Moscow Region,* PO Box 30, *Russia. E-mail: v.spichak@g23.relcom.ru*

Accepted 1999 November 3. Received 1999 July 27; in original form 1998 December 6

SUMMARY

The possibility of solving the three-dimensional (3-D) inverse problem of geoelectrics using the artificial neural network (ANN) approach is investigated. The properties of a supervised ANN based on the back-propagation scheme with three layers of neurons are studied, and the ANN architecture is adjusted.

A model class consisting of a dipping dyke in the basement of a two-layer earth with the dyke in contact with the overburden is used for numerical experiments. Six macroparameters of the 3-D model, namely the thickness of the top layer, which coincides with the depth of the dyke (D), the conductivity ratio between the first and second layers (C_1/C_2), the conductivity contrast of the dyke (C/C_2), and the width (W), length (L) and dip angle of the dyke (A), are used.

Various groups of magnetotelluric field components and their transformations are studied in order to estimate the effect of the data type used on the ANN recognition ability. It is found that use of only the xy- and yx-components of impedance phases results in reasonable recognition errors for all unknown parameters (D: 0.02 per cent, $C1/C2$: 8.4 per cent, $C/C2$: 26.8 per cent, W: 0.02 per cent, L: 0.02 per cent, A: 0.24 per cent).

The influence of the size and shape of the training data pool (including the 'gaps in education' and 'no target' effects) on the recognition properties is studied. Results from numerous ANN tests demonstrate that the ANN possesses good enough interpolation and extrapolation abilities if the training data pool contains a sufficient number of representative data sets.

The effect of noise is estimated by means of mixing the synthetic data with 30, 50 and 100 per cent Gaussian noise. The unusual behaviour of the recognition errors for some of the model parameters when the data become more noisy (in particular, the fact that an increase in error is followed by a decrease) indicates that the use of standard techniques of noise reduction may give an opposite result, so the development of a special noise treatment methodology is required.

Thus, it is shown that ANN-based recognition can be successfully used for inversion if the data correspond to the model class familiar to the ANN. No initial guess regarding the parameters of the 3-D target or 1-D layering is required. The ability of the ANN to teach itself using real geophysical (not only electromagnetic) data measured at a given location over a sufficiently long period means that there is the potential to use this approach for interpreting monitoring data.

Key words: 3-D inversion, dyke, magnetotellurics, neural network.

1 INTRODUCTION

The successfulness of EM data inversion depends not only on the quality of the data and the tools used to this end, but also, perhaps to an even greater extent, on the suitability of the inversion methods for the purposes of the data interpretation. The methods developed for 3-D inversion of MT data by Mackie & Madden (1993) and Spichak *et al.* (1995) enable the underground conductivity distribution to be estimated based on the measured data and *a priori* information. However, in spite of the different mathematical formalisms used (conjugate gradient relaxation and Bayesian statistics, respectively), both of these methods require (as in all other EM inversion techniques) that the parameters of a layered section are known in advance from some other geological or geophysical method. They are also inefficient for multiple inversions of data within a given class of model (for instance in the monitoring mode) since they do not remember a way of inversion already found. Finally, inversion of very noisy data (if the level of noise is, for example, 30, 50 or even 100 per cent, which is often the case in practice) by these methods may give results that are far from accurate. Hence, it is important to develop fundamentally new approaches to the interpretation that will overcome or at least reduce the difficulties mentioned above.

Methods of data interpretation based on the analogy with the function of the human brain's neural network have proved to be successful in the solution of inversion (recognition) problems in many fields of science. A pattern recognition method, namely the artificial neural network (ANN) technique, has become especially popular during the last decade. The following properties of ANNs make their application successful:

(1) ANNs are very effective for the solution of non-linear problems;
(2) ANNs can reach conclusions from incomplete and noisy data;
(3) ANNs admit the interpolation and extrapolation of the available database;
(4) ANNs provide a means for the synthesis of separate series of observations to obtain an integral response, which allows a joint interpretation of diverse data obtained by different geophysical methods;
(5) ANNs enable simultaneous data processing, thereby essentially reducing the computation time, particularly when special chips are employed;
(6) the time necessary for ANN recognition depends on the dimension of the space of unknown parameters rather than the physical dimension of the medium, which makes ANNs particularly promising for the interpretation of 3-D geoelectric structures.

An excellent review of ANN paradigms and a detailed analysis of their application to various geophysical problems is given in Raiche (1991). ANN methods have been used in geoelectrics for 1-D inversion by Sen *et al.* (1993), Hidalgo *et al.* (1994) and Poulton & Birken (1998), and parameters of 2-D structures have been estimated from synthetic and real time-domain electromagnetic data by Poulton *et al.* (1992a,b). The present paper, developing the ideas formulated in Spichak (1990), is a first attempt to apply the ANN approach to the inversion of electromagnetic data in 3-D geoelectric structures.

2 BACK-PROPAGATION SCHEME OF THE ANN APPLICATION

To solve the inverse problem we use one of the so-called 'methods of learning with a teacher', namely the error back-propagation (BP) technique (Rumelhart *et al.* 1988; Schmidhuber 1989; Silva & Almeida 1990). Such an approach involves two stages in the inversion procedure: the training of the network, and testing, or recognition (the inversion itself). At the learning stage, the 'teacher' specifies the correspondence between chosen input and output data, which is similar to the mechanism for training a human. The analogy with the human brain also includes the similarity of some functional elements of the biological neural system to the non-linear system 'data–parameters of the target' modelled by the ANN (its elements are also called 'neurons'). In both cases, the system could be considered as an *n*-layer network in which every neuron of one layer is somehow connected with the neurons of other layers. A signal arrives at the input layer of neurons from outside the system, but its magnitude at the neurons of the other layers depends on the signal magnitudes and connection weights of all the associated neurons of the previous layer. Moreover, similar to in biological systems, the net response of an artificial neuron is described by a non-linear function.

Even though the BP technique has become a routine procedure (see, in particular, the references above), it is worthwhile to specify the main elements of the scheme. We use a three-layer ANN (Fig. 1) consisting of a layer of input neurons (data), a layer of hidden neurons (their number, generally speaking, is arbitrary and can be adjusted in order to reflect the complexity of the system—see Section 4 below), and a layer of output neurons (unknown parameters of the geoelectric structure).

The propagation of the input signal via a network occurs in the following way. The input signal x_i comes to each ith neuron of the input layer. It is equal to the corresponding element of the input vector, composed of the values of the measured electromagnetic field (or their transformations) at a number of periods. Every kth neuron of a hidden layer receives a summary input signal y_k^{inp} from all neurons of the input layer:

$$y_k^{\mathrm{inp}} = \sum_i w_{ik} x_i \,, \tag{1}$$

where w_{ik} are the connection coefficients (weights) between the input and hidden layers and the summation is carried out over all input neurons. The signals y_k^{inp} are transformed by each kth neuron of the hidden layer into the output signals y_k^{out} by the neuron 'activation functions' G_k^{h}:

$$y_k^{\mathrm{out}} = G_k^{\mathrm{h}}(y_k^{\mathrm{inp}}) \,. \tag{2}$$

The signals then propagate from a hidden layer to the output layer and for each jth neuron of the output layer we obtain

$$u_j = G_j^u \sum_k w_{kj} y_k^{\mathrm{out}} \,, \tag{3}$$

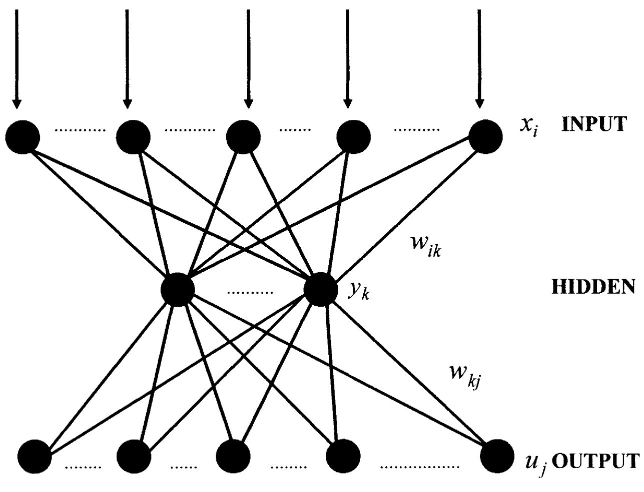

Figure 1. Three-layer artificial neural network.

where u_j are the output signals at the output layer, w_{kj} are the connection weights between the hidden and output layers, and G_j^u are the activation functions for neurons at the output layer. (The activation functions are usually considered to be the same for each neuron of a given layer. We will show in Section 4.1 that the individual choice of activation function type for different output neurons may improve the recognition of some model parameters.)

At the training stage the actual output signals u_j are compared with known 'correct answers' u_j^t, which correspond to given input signals, and a standard error

$$Er_p = \sum_j (u_{p,j} - u_{p,j}^t)^2 \tag{4}$$

is calculated for each pth learning sample; here the summation is carried out over all neurons of the output layer. In this paper, the term 'learning sample' means a pair of 'calculated synthetic EM data or their transforms at a number of periods and the corresponding set of model parameters'. Such input–output pairs are defined by the 'teacher' and comprise the ANN training sequence. The total error to be minimized is

$$Er = \sum_p Er_p, \tag{5}$$

where the summation is performed over all learning samples.

The connection weights w_{ik} and w_{kj} are the parameters that determine the signal propagation through the network and therefore the final error. BP is actually a gradient descent technique, minimizing the error Er by means of adjusting the connection weights:

$$\Delta w_{ij}^{(n)} = -\alpha \partial Er / \partial w_{ij}, \tag{6}$$

where $\Delta w_{ij}^{(n)}$ is the increment of the weight matrix at the nth step of the iteration process and α is a non-negative convergence parameter called the learning rate. In order to accelerate the process, an inertial term proportional to the weight change at the previous step $(n-1)$ is often added to the right-hand side of (6):

$$\Delta w_{ij}^{(n)} = -\alpha \partial Er / \partial w_{ij} + \beta \Delta w_{ij}^{(n-1)}, \tag{7}$$

where β $(0 \leq \beta \leq 1)$ is the inertial coefficient called the 'learning momentum'. The momentum can speed up training in very flat regions of the error surface and suppresses the weight oscillations in steep valleys or ravines (Schiffman *et al.* 1992).

Learning starts with small random values of the weights. The input signal comes via the network to the output. The output signal of the output layer is then compared with the desired value and the misfit is calculated. If it exceeds a predetermined small number, the signal propagates back through the network to the input, and so on. This procedure is repeated for the whole learning pool and ends when a user-specified threshold value Eps $(Er < Eps)$, known as a 'teaching precision', is reached.

The testing process uses the ANN interpolation and extrapolation properties. Unlike the training procedure, which requires many transits of the signal back and forth through the network, the recognition procedure requires only one passage of the recognizable signal from the input to output layer and uses the connection weights specified at the learning stage. The final set of output values can be treated as a result of the testing data inversion in a given model class.

3 CREATION OF THE SYNTHETIC DATABASE

In order that the ANN learns the correspondence between data and desired geoelectric parameters, it is first necessary to formulate the hypothesis on a class of inversion models, for instance dyke, geothermal reservoir, magma chamber, oil or gas deposit, etc. (Note that we mean only the assumption on the *class of models* for which the solution is sought, rather than the considerably more stringent constraints on the parameters of 1-D layering and/or target geometry used in the applications of other inversion methods.) This may be difficult in the general case, if we have no initial guess about the *type* of the geoelectrical model to be searched for, but is quite possible in some cases of practical importance. The recognition of crust dykes from surface measurements of the electromagnetic field is an example of such a formulation of the problem. It is easily parametrized, and the inversion is reduced to the determination of a few macroparameters of the target itself as well as of the host medium.

To apply the ANN method, it is necessary to create first a fairly representative base of models. A 3-D dipping dyke in the bottom layer of a two-layer earth with the dyke in contact with the overburden (Fig. 2) was considered as a 'class-generating' model and used for numerical experiments. It is characterized by the following parameters of the dyke and the host medium: the thickness of the upper layer (H_1); the conductivity contrast between the two layers (C_1/C_2); the conductivity contrast between the dyke and the host layer (C/C_2); the depth of the upper edge of the dyke (D), its width (W), length (L) and the dip angle in the plane xOz (A). It was assumed for simplicity that the upper boundary of the dyke always lies at the interface between the first and the second layers, so that $D = H_1$, and that the conductivity of the second layer is fixed: $C_2 = 0.01$ S m^{-1}. Thus, six parameters of 3-D geoelectric structure (D (H_1), W, L, A, C_1/C_2 and C/C_2) were to be reconstructed.

The following grading of unknown parameters was used in the forward modelling: D $(H_1) = 50$, 200 m; $C_1 = 0.00333$, 0.01 and 0.03 S m^{-1}; $C_2 = 0.01$ S m^{-1} (fixed); $C = 0.0002$, 0.001, 0.003333, 0.01, 0.02, 0.034, 0.06, 0.1, 0.17, 0.3 and 0.5 S m^{-1}; $W = 16.65$, 25, 50, 66.6, 100 and 200 m; $L = 16.65$, 25, 50, 66.6, 83.25, 100, 125, 200, 250, 330, 500 and 1000 m; $A = 0°$ (180°),

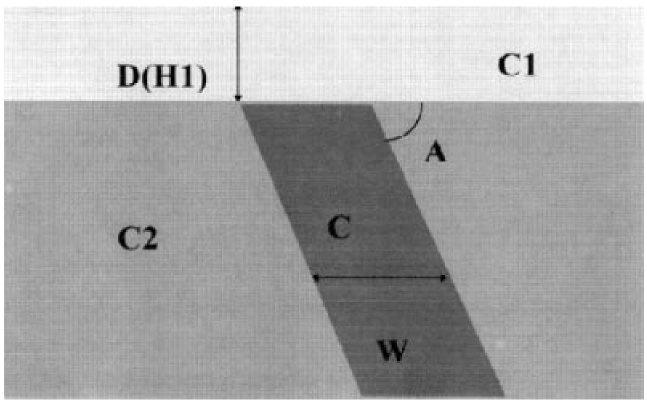

Figure 2. Cross-section of a 3-D model containing a dyke buried in the second layer of a two-layer earth.

45°, 66°, 90°, 114° and 135°. Note that, due to restrictions on computation time, not all possible combinations of the parameter values mentioned above were used for the creation of the synthetic database. In particular, the total number of calculations was decreased by conditions such as $D/W = 1, 2, 3$; $L/W = 1, 5$; 'basic' values of the conductivity contrast $C/C_2 = 2, 10, 50$; and so on.

In order to create a synthetic database the software package FDM3D (Spichak 1983) was used, which has proved to be efficient in the solution of forward and inverse 3-D geoelectric problems (Zhdanov & Spichak 1992; Spichak 1999). All calculations were carried out for two primary field polarizations within the period range typical for audiomagnetotellurics: $T = 0.000333$, 0.001, 0.00333, 0.01, 0.0333, and 0.1 s. Note that, due to the well-known electrodynamic similitude relation (Stratton 1941)

$$w\sigma L^2 = inv\,, \tag{8}$$

where ω is the frequency, σ is the electrical conductivity, and L is the geometrical scale, the same synthetic database could be used for ANN inversion in other period ranges and geometrical parameter scales satisfying (8).

4 ANN ARCHITECTURE

Since ANN architecture is of great importance for the recognition of the model parameters, a comprehensive study was carried out aimed at finding the appropriate values of the following parameters of the ANN: types of activation function for hidden and output layers as well as for neurons at the output layer; number of neurons in a hidden layer; and the effect of a second hidden layer. Finally, a teaching precision was estimated which enabled reasonable inversion results to be obtained.

In order to reduce computation time (without loss of generality), the database used in these experiments was smaller than the total one: only 90 synthetic data sets randomly selected from the total database were used for teaching, while 10 were randomly selected for testing. The ANN architecture in this experiment was as follows: the input layer consisted of 80 neurons, the hidden layer consisted of 20 neurons, while the output layer consisted of six neurons, corresponding to the six model parameters to be recognized. The threshold level (*Eps*) for rms errors in teaching was equal to 0.0075. The learning rate was equal to 0.01 and the momentum to 0.9. In the process of teaching, the rms errors were used to estimate the misfits between the calculated and 'true' responses, so the total error for the test set for all parameters was determined as follows:

$$Err = \left[\frac{1}{N_{\text{test}} N_{\text{par}}} \sum_{n,j} err_{n,j}^2 \right]^{1/2}, \tag{9}$$

where

$err(n, j) = [\text{target}\,(n, j) - \text{neural}\,(n, j)]/[\max(j) - \min(j)]$ $(j = 1, \ldots, N_{\text{par}};\ n = 1, \ldots, N_{\text{test}})$;

j is the number of the neuron in the output layer corresponding to the jth model parameter;

n is the number of the tested sample;

N_{par} is the number of output neurons (=6);

N_{test} is the number of testing data sets;

$\min(j)$, $\max(j)$ are the minimum and maximum values of the jth parameter in the teaching pool, respectively;

neural (n, j) is the recognition result for the jth parameter in the nth testing sample; and

target (n, j) is the target value of the jth parameter in the nth testing sample.

In order to estimate the quality of the ANN inversion of the synthetic data (when the true result is known in advance) we calculated for each jth unknown parameter the relative error averaged over all testing samples:

$$Err_j = \frac{1}{N_{\text{test}}} \sum_n \frac{|\text{target}_{n,j} - \text{neural}_{n,j}|}{\text{target}_{n,j}} 100 \text{ per cent}\,. \tag{10}$$

4.1 Types of activation function at the hidden and output layers

Since the type of activation function used is crucially important for a proper simulation of the behaviour of a real system, some experiments are made before using ANN to interpret real data. In spite of the fact that any monotonically increasing and continuously differentiable function can be used as an activation function for BP-type networks, the most commonly used ones are sigmoidal functions: hyperbolic tangent,

$$G(z) = \frac{1}{2}(1 + \tanh z) = \frac{1}{2}\left(1 + \frac{e^z - e^{-z}}{e^z + e^{-z}}\right) = \frac{1}{1 + e^{-2z}}$$

and 'logistic',

$$G(z) = \frac{1}{1 + e^{-z}}\,.$$

Their derivatives $G'(z) = G(z)[1 - G(z)]$ have a Gaussian shape that helps stabilize the network and compensate for over-correction of the weights (Caudill 1988).

ANN *interpolates* the parameters of the model using these activation functions quite satisfactorily, but it completely fails to *extrapolate* their values because the neural output for logistic and hyperbolic tangent functions lies in the interval [0, 1]. No values outwith this interval are achievable, so no real extrapolation can be carried out.

The simplest way to overcome this difficulty probably consists of using the linear activation function at the output layer. Comparative testing of neural networks with linear and hyperbolic tangent activation functions has revealed that networks with non-linear outputs extrapolated low values of the conductivities reasonably well but failed in the extrapolation of high conductivities, and, contrariwise, networks with linear outputs extrapolated high values of the target conductivity reasonably well but failed in the extrapolation of low values. Based on this preliminary experience, we compared the effects of the following two types of activation function for the neurons at the output layer:

$$G^{\text{lin}}(x) = 0.5(1 + x) \quad \text{(linear function)}, \tag{11}$$

$$G^{\text{lin}}(x) = 0.5 \begin{cases} 1 + \tanh x, & x < 0 \\ (1 + x), & x > 0 \end{cases} \quad \text{(mixed function)}. \tag{12}$$

The ANN we used for experiments had the *same* type of activation function at *each* neuron of the hidden or output layer, so, in order to estimate the effects of *different* activation functions, we had to teach *six* ANNs independently, each having only one output neuron, corresponding to the appropriate model parameter. All neurons in the hidden layers had the same mixed activation function, while each (single!) output neuron of the appropriate ANN had an activation function depending on the nature of the corresponding model parameter: *linear* (11) or *mixed* (12) for the dimensional parameters (D, W, L) of the model, and only *mixed* for the conductivity contrasts (C/C_2, C_1/C_2) and dip angle (A).

Table 1 gives the recognition results for two types of activation function and two ways of teaching/testing mentioned above: (1) six ANNs, three of them having 'linear', and three 'mixed' activation functions; (2) six ANNs, all 'mixed'; and (3) one ANN, 'mixed'. (Hereafter, average relative errors and their bars (in per cent) are placed in round brackets and separated by a comma for all model parameters).

It could be concluded from the comparison of the first two rows of Table 1 that, under the condition that the activation function at the output layer has a linear part, the recognition errors for all parameters are reasonable and are practically independent of the type of activation function used. The errors even decrease slightly (though by less than 2.8 per cent except for C/C_2) if all model parameters are reconstructed by the *same* ANN taught to be able to recognize *all* model parameters (third row). It is important to note, however, that, in spite of the fact that the total time of teaching in the latter case is much less than in the former, the best recognition accuracy for the dyke conductivity contrast C/C_2 is achieved if this parameter is recognized *independently* (in a 'partial solution' mode) by an ANN having only one output neuron and taught in an appropriate way. In the latter case, the relative error may decrease by 10 per cent (from 33.2 per cent to 23.6 per cent).

4.2 Number of neurons in a hidden layer

Unfortunately, there is no general theory on the dependence of the recognition errors on the number of neurons in a hidden layer. However, the *approximation* properties of an ANN are improved when the number of hidden neurons increases. In particular, Yoshifusa (1992) has proved that a non-linear perceptron with one hidden layer can approximate any continuous function with a given precision if the number of hidden neurons tends to infinity.

The *recognition* ability of an ANN increases when the number of hidden neurons increases if it is dealing with *familiar* data (in particular, those used for training). On the other hand, it may decrease if the ANN is dealing with *unknown* testing data, because in the general case its recognition ability (called the 'generalization property') depends in a complex way on its architecture (number of hidden layers, number of neurons, type of activation function, etc.), and the volume and structure of the training data pool, etc. Therefore, the *optimal* numbers of hidden layers and hidden neurons are usually found by trial-and-error techniques (Baum *et al.* 1989; Kung *et al.* 1988; Soulie *et al.* 1987).

We studied the effect of the number of neurons in a hidden layer on the accuracy of the recognition of model parameters by means of testing the data sets used in the previous section. The ANN architecture was 80–Nh–6, where Nh is the number of neurons in a hidden layer. The values of Nh were assigned as follows: 10, 20, 30, 40, 50. The teaching precision was equal to 0.0075.

Fig. 3 shows the dependence of the accuracy of the model parameter recognition (in terms of relative errors averaged over all testing data sets and appropriate bars, both in per cent) on the number of neurons in a hidden layer. In Fig. 3 and subsequent figures (unless stated otherwise) (a) shows the depth of the dyke and thickness of the upper layer (D (H_1)), (b) shows the conductivity contrast of the upper layer (C_1/C_2), (c) shows the width (W) of the dyke, (d) shows the conductivity contrast of the dyke (C/C_2), (e) shows the length (L), and (f) shows the dip angle (A).

It is seen from Fig. 3 that the relative errors for four parameters (D, W, L and C_1/C_2) are generally less than 3–4 per cent, while the maximum relative errors for C/C_2 and A are around 14 per cent. The total average error for all six parameters ranges from 3.9 to 5.5 per cent. In spite of the fact that the recognition errors are not very sensitive to the number of neurons in a hidden layer, the minimal total error is achieved at $Nh = 40$. It is worthwhile to note that, for all values of Nh, the standard deviations of the relative errors for W, L and C_1/C_2 were fairly small (<5 per cent), while they reached 15 per cent for C/C_2 and A. This result indicates that the reconstruction of the dyke's depth, conductivity contrast and angle is a less stable procedure than the recognition of other unknown parameters.

4.3 Effect of a second hidden layer

In order to estimate whether two hidden layers are better than one, a second hidden layer was incorporated into the ANN. Based on the results of the previous experiments, the number of neurons in the first hidden layer was fixed at $Nh = 40$, while the number of neurons in the second one varied. The new ANN architecture was 80–40–$Nh(2)$–6, where $Nh(2)$ was successively equal to 10, 20, 30 and 60.

Fig. 4 shows the results of the recognition of model parameters compared with the case when the ANN consists of only one hidden layer with the optimal number of neurons [$Nh(1) = 40$, $Nh(2) = 0$]. The total average error is minimal (4.3 per cent) when $Nh(2) = 0$ and is maximal if $Nh(2) = 60$ (mean value equals 7.7 per cent). The relative error graphs do not change significantly in comparison with the case of only one hidden layer (Fig. 3). Thus, it can be concluded that the architecture of the ANN is quite adequate for the complexity

Table 1. The results (in terms of averaged relative errors and bars, both in per cent) of the model parameter recognition for two types of activation function and two ways of testing.

N	D	C_1/C_2	W	L	A	C/C_2
1	(5.5, 16.3)	(11.8, 16.8)	(5.5, 16.3)	(4.8, 14.5)	(7.3, 12.1)	(23.6, 24.7)
2	(6.7, 14.6)	(11.8, 16.8)	(6.7, 14.6)	(5.4, 13.0)	(7.3, 12.1)	(23.6, 24.7)
3	(4.9, 16.1)	(9.0, 15.4)	(4.9, 16.1)	(4.3, 14.4)	(7.1, 7.9)	(33.2, 35.3)

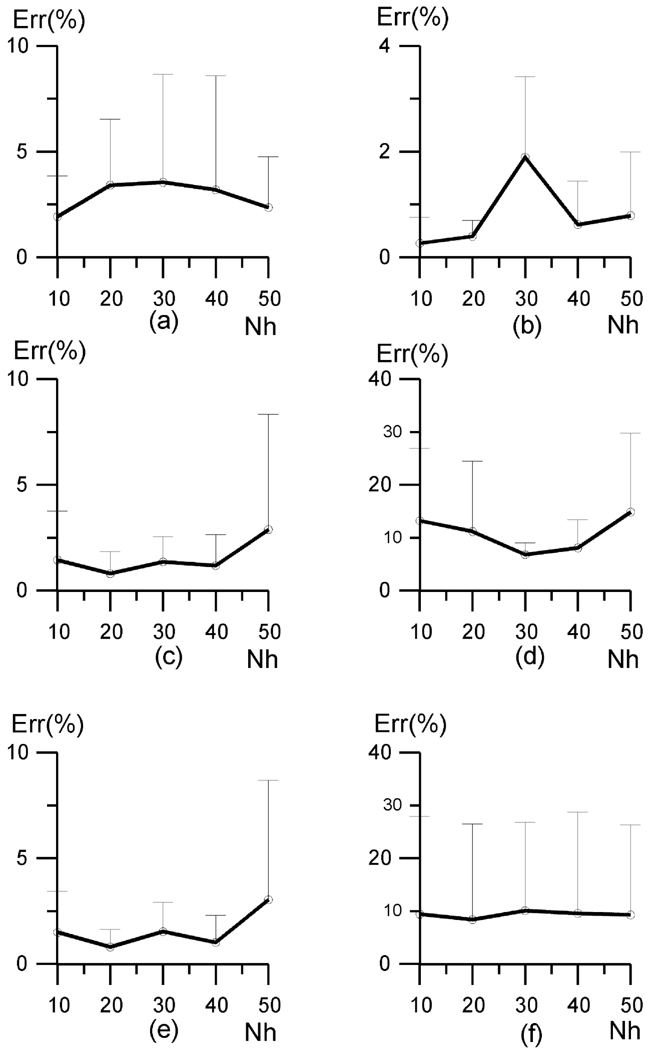

Figure 3. The dependence of the recognition errors on the number of neurons in a hidden layer (*Nh*). Here and in the following figures (except Fig. 7), (a) is the depth of the dyke (*D*), (b) is the conductivity ratio between the upper and second layers (C_1/C_2), (c) is the width of the dyke (*W*), (d) is the conductivity contrast of the dyke (C/C_2), (e) is the length of the dyke (*L*), and (f) is the dip angle of the dyke (*A*).

of the problem considered; so, addition of the second hidden layer has a very similar effect to increasing the total number of neurons in intermediate layers, the total number $Nh=40$ in all hidden layers being the optimal value.

4.4 Threshold level

To select the optimal value of the teaching precision (*Eps*) the parameters of the model were reconstructed using ANNs having the same architecture, but taught using different values of the stopping criterion *Eps* (0.005, 0.0075, 0.01, 0.02, 0.05). Fig. 5 shows the dependence of the relative recognition errors and their bars (in per cent) on the threshold value *Eps* (the horizontal axis has $\log_{10} Eps$ in inverse order). It is seen that the average errors and bars for all model parameters (besides the dip angle of the dyke) sharply decrease when *Eps* tends to zero and stabilize when it becomes less than 0.01.

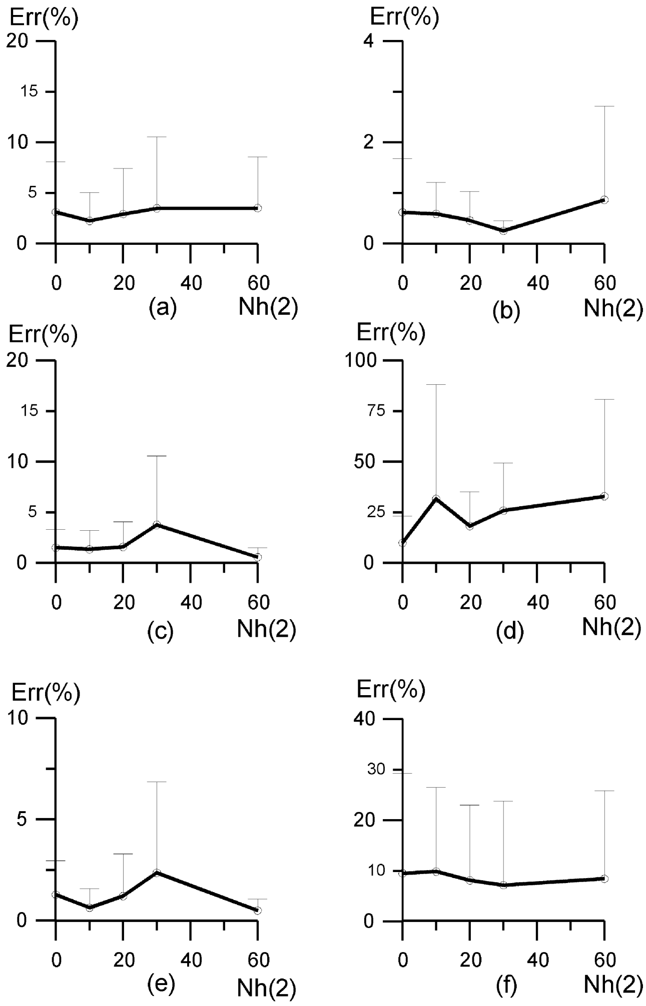

Figure 4. The dependence of the recognition errors on the number of neurons in a second hidden layer ($Nh(2)$).

5 EFFECT OF THE INPUT DATA TYPE

The results of ANN-based parameter recognition depend on the data type used (both for teaching and inversion itself), the volume and structure of the teaching data pool, etc. In order to estimate the effect of data type on the results of inversion, the following five types of input data were studied:

(1) normalized components of electrical and magnetic fields: $(|E_{y,x}| - E^{n}_{y,x}|)/|E^{n}_{y,x}|$, $(|H_{x,y}| - |H^{n}_{x,y}|)/|H^{n}_{x,y}|$;

(2) components of not normalized electrical field parallel to the polarization of the primary field ($\mathcal{R}e E_{x,y}$, $\mathcal{I}m E_{x,y}$);

(3) only apparent resistivities ρ^{a}_{xy}, ρ^{a}_{yx};

(4) apparent resistivities ρ^{a}_{xy}, ρ^{a}_{yx} and phases φ_{xy}, φ_{yx} of impedance;

(5) only phases φ_{xy}, φ_{yx} of impedance.

Table 2 shows the recognition results for the five groups enumerated above.

It is seen from Table 2 that the recognition of the dimensional parameters (*D*, *W* and *L*) is carried out with errors of less than 0.5 per cent irrespective of the type of input data used. The dip angle (*A*) is determined fairly well (the maximum error of 4.14 per cent corresponds to using only apparent resistivities,

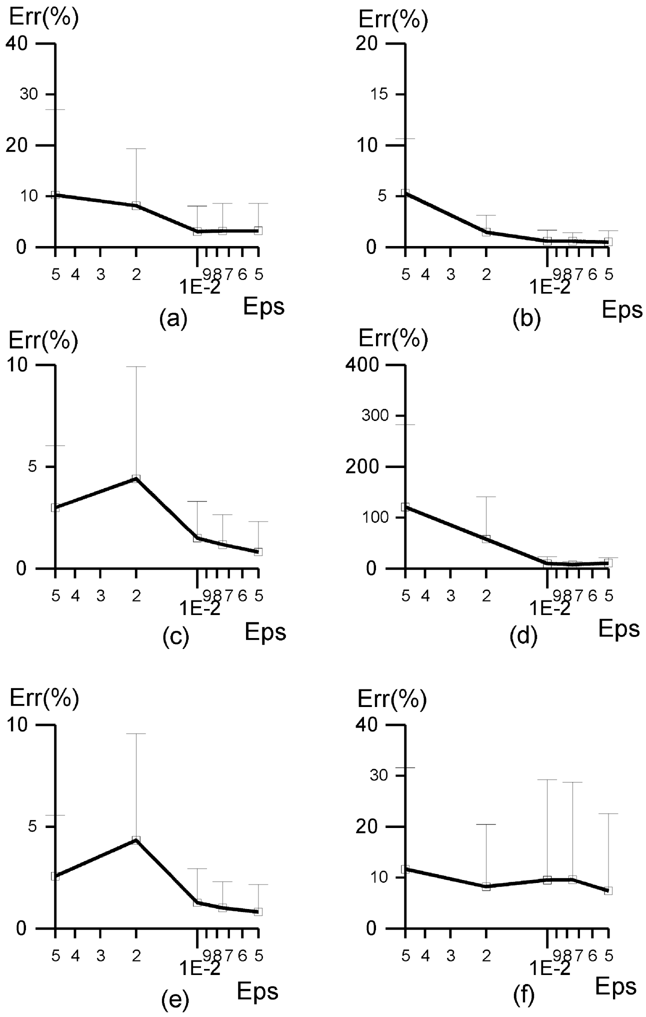

Figure 5. The dependence of the recognition errors on the rms error achieved in the teaching process (*Eps*).

Table 2. Relative errors (in per cent) of parameter recognition for five types of input data.

N	D	C_1/C_2	W	L	A	C/C_2
1	0.10	3.84	0.10	0.13	0.41	40.77
2	0.13	7.20	0.15	0.17	1.77	67.71
3	0.04	3.28	0.05	0.04	4.14	39.05
4	0.27	4.14	0.40	0.43	1.63	24.45
5	0.02	8.43	0.02	0.02	0.24	26.81

and the minimum corresponds to the case when only impedance phases are used). The errors of the conductivity contrast (C_1/C_2) estimation are larger, ranging from 3.28 to 8.43 per cent. They are maximal (in contrast to the previous case) when only impedance phases are used, and minimal when only apparent resistivities are used. Finally, the dyke conductivity contrast (C/C_2) recognition errors are the largest and vary from 24.45 per cent (apparent resistivities and phases) up to 67.71 per cent (electrical fields only). On average, groups 4 (apparent resistivities and phases) and 5 (phases only) give the best recognition results for all dyke parameters searched for.

Thus, we conclude that: first, the conductivity contrasts are determined worse than other model parameters; second, the recognition errors of the dimensional parameters (D, L and W) are virtually independent of the data type, while the conductivity contrasts (especially of the dyke) are essentially dependent; and third, the phases of impedance could be used to determine not only the geometrical parameters, but also the conductivity distribution. The latter conclusion enables us to reduce (at least by a factor of two in comparison with the routinely used two components of apparent resistivity and two components of the impedance phase) the volume of data required for EM data inversion in 3-D media.

6 EFFECT OF THE VOLUME AND STRUCTURE OF THE TRAINING DATA POOL

6.1 Effect of size

The dependence of the accuracy of the model parameter recognition on the number of data sets in the teaching database was studied. In total, 120 data sets were used in these experiments: 12 of them were used for testing, while the remaining 108 samples were used for teaching. Synthetic electrical and magnetic fields were, in accordance with the results of the previous section, first transformed to the apparent resistivities and phases. In order to reduce the huge volume of data being processed during the teaching, the data were compressed by 2-D FFT (using the first five pairs of sine and cosine coefficients). Thus, the total number of input neurons in the network was $N = 10N_{\mathrm{tr}}N_{\mathrm{T}}$, where N_{tr} is the number of field or transformation components used for the inversion, and N_{T} is the number of periods. 2-D FFT of the data was carried out on a 32×32 grid for each period. The testing data sets were not changed during this experiment, while the data sets used for teaching were selected from the remaining data sets in a random way to make their number successively equal to 54 (50 per cent of the total database), 65 (60 per cent), 76 (70 per cent), 86 (80 per cent), 92 (85 per cent), 97 (90 per cent), 103 (95 per cent). The results of the parameter recognition are shown in Fig. 6. It is seen that the average errors for all parameters decrease with increasing number of data sets used for teaching, and stabilize (the bars for almost all parameter errors decrease) when the size of the training database becomes about 90–100. Thus, an increase of the teaching database size up to 90–100 data sets may improve both the accuracy and robustness of the recognition of all model parameters. The latter point is very important from the point of view of interpretation of a single data set, which is usually the case in practice.

6.2 Effect of structure

It could have been expected that the larger the volume of the training database, the better the results of unknown parameter recognition; however, the effect of the database structure is much less clear. It has been studied in two different ways: first, the effect of the random selection of the training data sets from the database was estimated; second, the influence of the 'gaps' in the database used for teaching was studied. In both cases, the testing data set was not changed and consisted of the same 12 data sets as in the previous section.

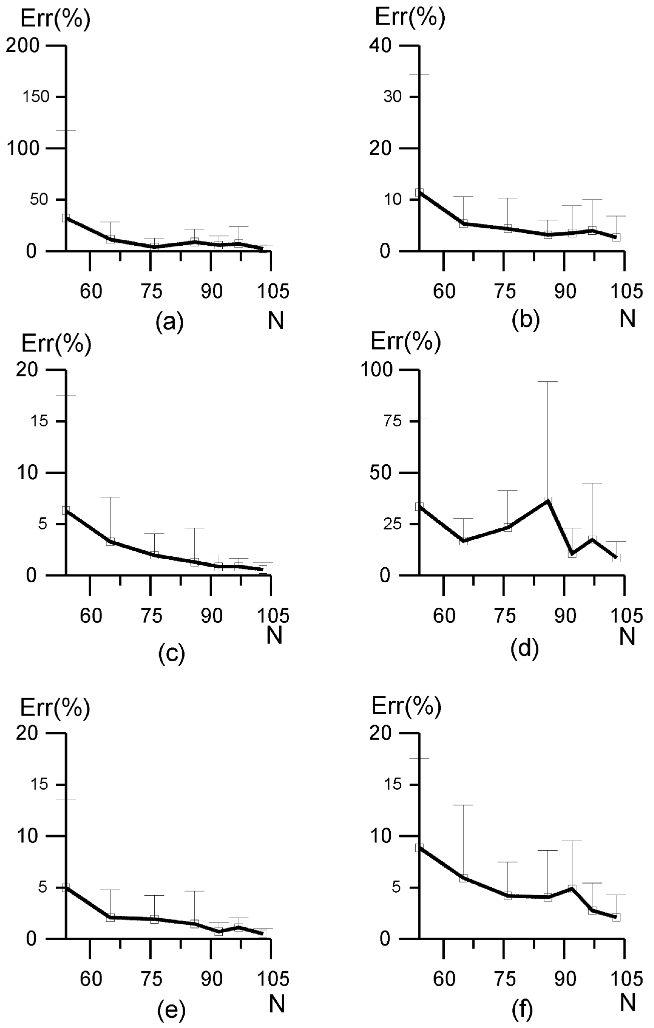

Figure 6. The dependence of the recognition errors on the number of synthetic data sets (N) used for teaching the ANN.

6.2.1 *Random selection of synthetic data samples*

In order to estimate the effect of a random selection of the training data sets, the teaching was carried out using 97 data sets randomly selected from the database. The architecture of the ANN was the same as above. The results of the model parameter recognition for five different random formations of the training data sets are given in Table 3.

Table 3 indicates that for each model parameter both the errors and their bars are quite reasonable [the errors of the dyke width (W) and length (L) recognition are the smallest, while the error of the dyke conductivity contrast (C/C_2) recognition is the largest], and do not depend on the random selection of the data sets in the training database. Thus, a random selection of the data sets used for training hardly affects the results of the recognition. This implies that even a database consisting of only 97 data sets is of good enough 'quality' for teaching the ANN to recognize the model parameters.

Table 3. The model parameter recognition errors and bars (in per cent) for five randomly selected teaching data pools.

N	D	C_1/C_2	W	L	A	C/C_2
1	(2.2, 1.3)	(2.6, 2.7)	(1.0, 0.7)	(1.3, 0.8)	(3.3, 2.5)	(7.1, 7.3)
2	(4.6, 4.6)	(3.4, 2.7)	(1.7, 2.1)	(1.3, 1.5)	(4.9, 3.3)	(8.8, 9.8)
3	(4.5, 6.8)	(5.4, 6.2)	(2.5, 2.3)	(2.3, 1.9)	(5.5, 4.6)	(10.3, 11.6)
4	(4.9, 5.6)	(2.8, 2.3)	(1.4, 1.6)	(1.1, 1.2)	(5.7, 8.3)	(11.3, 10.7)
5	(4.9, 5.6)	(2.6, 3.1)	(2.0, 3.9)	(1.4, 2.0)	(3.7, 2.3)	(12.0, 8.7)

6.2.2 *Gaps in the training database*

It is important to estimate the effect of gaps in the database on the recognition errors. In order to study this, the training database was artificially 'damaged' (decrease in the number of teaching values) separately for each model parameter.

6.2.2.1 Lack of data in each of the model parameters. In order to estimate the effects of gaps in the teaching data pool on the recognition abilities of the ANN, the following experiments for four characteristic model parameters (A, D, C_1/C_2, C/C_2) were carried out. The teaching data pool was successively 'damaged' by deleting all possible data sets for one of the parameter values ($A=66$, $C_1/C_2=0.333$, $D=50$ m, $C/C_2=3.4$) from the teaching data. In each experiment two testing data pools were used: one (unchanged in all experiments) consisting of 12 reference data sets randomly selected from the total database, and the second consisting of those data sets that had been deleted from the teaching data.

The architecture of the ANN was the same as in the previous sections, except for the type of activation function at the output layer. Tables 4 and 5 give the results of the model parameter recognition for two types of output activation function—linear and mixed, respectively—when the testing data pool consists of the 12 reference data sets mentioned above. The last rows in both Tables contain, for comparison, the recognition errors for the case when no artificial gaps are created in the teaching data pool.

Similarly, Tables 6 and 7 give the results of the model parameter recognition for two types of output activation function—linear and mixed, respectively—when the testing data pool consists of data sets removed from the teaching data pool.

Analysis of Tables 4, 5, 6 and 7 results in the following conclusions:

(1) When using a 'damaged' training data pool, recognition errors depend on the type of output activation function [in contrast to the case of 'homogeneous' training data (Table 1)]. The mixed function is preferable for determining the conductivities and dip angle, while the linear function gives better results for dimensional parameters. It enables us to decrease the recognition errors of C/C_2, even if the grading of C_1/C_2 and A in the training database is insufficient.

(2) A lack of training data sets in C_1/C_2 is more crucial for the recognition of C/C_2 than the lack of values of C/C_2 itself.

(3) A lack of training data sets in dip angle A affects the recognition error of D even more than the recognition error of A itself.

(4) The largest recognition errors of all model parameters occur if the number of data sets is insufficient in D. On the

Table 4. Effect of the gaps in the training data sets on the results of testing 12 reference data sets (in the case of a linear output activation function).

Errors in:	D	C_1/C_2	W	L	A	C/C_2
A	(70, 125)	(10, 17)	(44, 106)	(30, 71)	(17, 25)	(56, 87)
C_1/C_2	(23, 61)	(36, 52)	(2.4, 3.2)	(1.6, 1.4)	(10, 19)	(179, 388)
D	(136, 161)	(33, 65)	(88, 131)	(58, 86)	(16, 23)	(33, 39)
C/C_2	(5.1, 6.1)	(3.6, 3.4)	(4.9, 10)	(4.6, 9.5)	(7.8, 9.1)	(23, 23)
No gaps	(3.2, 8.2)	(4.3, 5.1)	(1.9, 2.1)	(1.9, 2.3)	(4.1, 3.2)	(23, 17)

Table 5. Effect of the gaps in the training data sets on the results of testing 12 reference data sets (in the case of a mixed output activation function).

Errors in:	D	C_1/C_2	W	L	A	C/C_2
A	(67, 149)	(9, 13)	(48, 127)	(34, 87)	(14, 21)	(29, 32)
C_1/C_2	(19, 26)	(67, 98)	(8, 12)	(8, 13)	(12, 20)	(37, 35)
D	(149, 197)	(16, 24)	(84, 149)	(58, 100)	(18, 29)	(32, 32)
C/C_2	(8, 16)	(4, 5)	(6, 15)	(6, 16)	(8, 16)	(17, 14)
No gaps	(8.7, 11.4)	(5.9, 6.0)	(2.2, 3.2)	(2.1, 2.9)	(4.7, 5.3)	(23, 31)

Table 6. Effect of the gaps in the training data sets on the results of testing deleted data sets (in the case of a linear output activation function).

Errors in:	D	C_1/C_2	W	L	A	C/C_2
A	(127, 153)	(6, 10)	(117, 152)	(80, 103)	(41, 32)	(58, 60)
C_1/C_2	(22, 40)	(104, 16)	(6, 8)	(7, 7)	(32, 31)	(297, 492)
D	(261, 145)	(69, 80)	(219, 102)	(152, 63)	(45, 43)	(80, 66)
C/C_2	(35, 74)	(12, 34)	(32, 37)	(30, 34)	(28, 46)	(59, 53)

Table 7. Effect of the gaps in the training data sets on the results of testing deleted data sets (in the case of a mixed output activation function).

Errors in:	D	C_1/C_2	W	L	A	C/C_2
A	(178, 246)	(9, 13)	(142, 181)	(100, 122)	(34, 24)	(29, 25)
C_1/C_2	(56, 31)	(200, 0.3)	(19, 20)	(15, 18)	(21, 22)	(89, 72)
D	(320, 202)	(28, 19)	(221, 129)	(155, 82)	(46, 35)	(66, 19)
C/C_2	(31, 31)	(6.0, 7.8)	(36, 28)	(34, 29)	(32, 24)	(27, 25)

other hand, the smallest errors occur when the number of data sets corresponding to C/C_2 is close to other parameters' grading.

Thus, it could be concluded that, in order to obtain good results of ANN recognition, the numbers of the intervals (grading) in each parameter in the training data pool should be close to each other. Since it is often difficult to follow this recommendation in practice, it is generally better to use different types of activation function for two groups of parameters: mixed for essentially 'non-linear' parameters (C/C_2, C_1/C_2 and A), and linear for dimensional ones (D, W and L).

6.2.2.2 'No dyke' case. It is also important to estimate the ability of the ANN to recognize situations where there is *no* anomalous body embedded in the layered earth. To this end, the dependence of the 'no dyke' recognition on the number of corresponding synthetic data sets in the training data pool was studied.

12 data sets corresponding to the 'no dyke' case were randomly selected from the total database and then successively shared by teaching and testing data pools as follows: 4/8, 6/6, 8/4 and 10/2. Thus, the teaching data pool included 4, 6, 8 and finally 10 data sets corresponding to the absence of the dyke, while the testing data consisted only of the 'no dyke' data sets (8, 6, 4 and 2, in turn). The ANN architecture was the same as in the previous sections.

The model parameter recognition errors and their bars versus the number N of the 'no dyke' data sets in the training data pool are given in Fig. 7 [(a) D (H_1), (b) C_1/C_2, (c) C/C_2, (d) total average relative error].

Fig. 7 shows that the average errors and bars for the model parameters sharply decrease if the number of the 'no dyke' data sets in the teaching data pool increases. The recognition errors of the upper layer parameters (C_1/C_2 and H_1) manifest similar behaviour: they become quite reasonable if the number of 'no dyke' data sets included in the teaching data pool is at least 8 (6.7 per cent of the total number of the teaching data sets used).

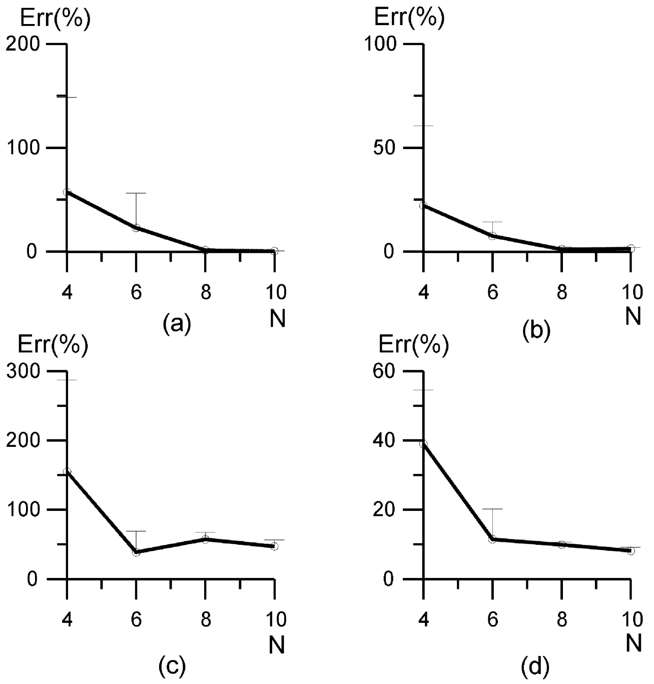

Figure 7. The dependence of the recognition errors on the number of synthetic data sets, corresponding to 'no fault' models (N), in the teaching data pool: (a) D (H_1), (b) C_1/C_2, (c) C/C_2, (d) total average relative error.

On the other hand, even if the number of 'no dyke' data sets in the training data pool further increases the relative error of the conductivity contrast C/C_2, recognition remains at the level of 50 per cent (Fig. 7c), which is evidently not sufficient for the classification of such inversion results as 'absence of dyke'. This situation could probably be improved by further increasing the proportion of appropriate data sets in the total data pool used for teaching the ANN. Fortunately, in all experiments (see Tables 8 and 9 for test examples) the resulting values of the dyke width (W) and length (L) were close to zero, so, even if other parameters (in particular the dyke conductivity contrast C/C_2) do not indicate the 'absence of the dyke', these two conditions could be used as reliable indicators of such a situation.

Table 8. The recognition results for the 'no dyke' case when the host medium is homogeneous ($C_1/C_2=1$).

	D (m)	C_1/C_2	W (m)	L (m)	A (°)	C/C_2
Target	0.0	1.00	0.0	0.0	0.00	1.00
Result	0.0	1.98	0.0	0.0	3.21	0.59
Err (per cent)	0.0	1.7	0.0	0.0	0.0	40.2

Table 9. The recognition results for the 'no dyke' case when the host consists of two layers ($C_1/C_2=3$).

	D (m)	C_1/C_2	W (m)	L (m)	A (°)	C/C_2
Target	50	3.00	0.0	0.0	0.00	1.00
Result	40	3.00	0.0	0.0	0.53	1.53
Err (per cent)	4.2	0.1	0.0	0.0	0.0	53.9

7 EXTRAPOLATION ABILITY OF ANNS

As was mentioned above, the most important abilities of ANNs consist of interpolation and extrapolation of the parameter values. The testing carried out in the previous sections has demonstrated the good interpolation ability of ANNs. Special tests were made in order to estimate the extrapolation properties with respect to the dyke's depth (D), dip angle (A), length (L) and the conductivity contrast (C/C_2). Appropriate model parameter values to be recognized from extrapolation were first of all removed, if they existed, from the teaching data pool. If necessary, all data sets corresponding to smaller or larger values of the parameter considered were also removed. Based on the results obtained in Section 5, the recognition (inversion) was carried out based on xy and yx components of the apparent resistivities and the impedance phases.

Tables 10, 11 and 12 show the extrapolation abilities of the ANN with respect to the dyke depth (D) (Table 10), dip angle (A) (Table 11) and conductivity contrast (C/C_2) (Table 12).

Tables 10 and 11 demonstrate adequate results for extrapolation in D and A. The recognition errors for these parameters equal 22.8 and 19.7 per cent, respectively, each of them being underestimated. However, it is worthwhile to mention that extrapolation in D results in a rather large error in C/C_2 (51 per cent), which could be explained by relatively sparse grading in D and, vice versa, dense grading in C/C_2 in the teaching data pool (see Section 6.2.2.1).

Table 12 shows the results of extrapolation in conductivity contrast of the dyke (C/C_2). The recognition of the conductivity contrast as well as of other parameters in this test is quite reasonable. A similar test of the extrapolation abilities of the ANN in C/C_2 ($C/C_2=50$, while values of other parameters vary) shows that the results are very little affected by the location of the point selected for extrapolation regarding the parameter space used for teaching. Thus, we can conclude that an ANN with good interpolation and extrapolation properties is created.

Table 10. The recognition results for extrapolation in D.

	D (m)	C_1/C_2	W (m)	L (m)	A (°)	C/C_2
Target	300	1.00	200	200	66.00	10
Result	230	1.00	220	170	67.34	4.9
Err (per cent)	22.8	0.0	10.9	12.4	2.0	51

Table 11. The recognition results for extrapolation in A.

	D (m)	C_1/C_2	W (m)	L (m)	A (°)	C/C_2
Target	200	0.333	200	150	135.0	50
Result	190	0.334	190	140	108.4	46.7
Err (per cent)	2.3	0.35	2.3	2.1	19.7	6.6

Table 12. The recognition results for extrapolation in C/C_2.

	D (m)	C_1/C_2	W (m)	L (m)	A (°)	C/C_2
Target	50	3.0	50	50	66.0	50.0
Result	50.1	2.82	50.1	50.1	59.9	37.1
Err (per cent)	0.2	6.0	0.2	0.2	9.2	25.9

However, the situation changes dramatically if some artificial noise is added to the synthetic testing data. Fig. 8 shows the extrapolation recognition results when synthetic data are mixed with 30, 50, and 100 per cent Gaussian noise. In spite of the fact that the recognition errors for all parameters averaged over all testing samples generally increase when the noise level in the data increases, in some cases their behaviour is quite unusual. In particular, the recognition errors for dip angle and the dyke conductivity contrast often diminish when the noise level increases, which results in 'plateaus' in averaged errors graphs, clearly seen in Fig. 8. This, in turn, means that the routine methods of data noise reduction may not necessarily lead to an improvement of the ANN-based inversion results. A special study of this problem is to be carried out

8 RESULTS

The following results have been obtained.

(1) A 3-D geoelectrical model of a typical dyke zone was used to create a synthetic MT database for the frequency range typical for audiomagnetotelluric studies. It was used further for forming both the training and testing data pools for ANN recognition of the model parameters.

(2) The properties of the supervised ANN (based on using the back-propagation scheme) were investigated and the most appropriate values of its parameters were found. The results of studies aimed at improving the performance characteristics of the ANN as well as at increasing the recognition reliability and accuracy are as follows: one hidden layer is quite sufficient, the preferred number of hidden neurons being about 40; the threshold value used for teaching should be less than 0.01; the best activation function for the hidden layer is mixed, while for the output layer it is linear for dimensional parameters of the model and mixed for conductivity contrasts and the dip angle.

(3) The resolving power of different MT data transformations was estimated. The best recognition results (among the transforms considered) were achieved in the case of the simultaneous use of the apparent resistivity (xy and yx components) and phases of impedance, and also in the case of using only phases. The latter circumstance enables, in turn, the volume of data required for reliable inversion to be decreased. Moreover, it is worthwhile to note that use of phases only may result in even lower recognition errors for the dimensional parameters searched for.

(4) It was found that the volume of the teaching data pool sufficient for good enough recognition of six model parameters is about 100 data sets. Although the method of data set selection hardly affects the recognition errors, the 'damaged' structure of the teaching data pool may significantly affect the results of the recognition (even if the number of training data sets used is sufficient). In particular, gaps in the training data pool increase the errors of the parameter recognition (in a complicated way). It turns that, in order to obtain reliable results of ANN recognition, the numbers of intervals in each parameter grading during the creation of the training data pool should be as close to each other as possible. In other words, the teaching data pool should not be 'skewed' either due to insufficient grading of some parameter in comparison to others (which might be expected) or due to its extra dense division. Since it is often difficult to follow this recommendation in practice, it is generally better to use different types of activation function for two groups of parameters: mixed for essentially 'non-linear' parameters (C/C_2, C_1/C_2 and A), and linear for dimensional ones (D, W and L).

(5) Reliable recognition of macroparameters implies the ability to classify situations when there is no target embedded in the layered earth. Since the recognition errors (especially for conductivity contrast and dip angle) are not usually equal to zero even in numerical experiments, it is unlikely that the theoretical 'necessary' and 'sufficient' conditions of such classification will be useful in practice. It was found that inclusion of 'no target' data sets in the teaching data pool (about 10 per cent of the total volume) enables the ANN to reliably recognize the absence of the target: equality to zero of the target dimensional parameters (except the depth, which is by definition equal to the first layer thickness) could be considered as good indicators for such a classification.

(6) The recognition quality becomes worse when artificial (Gaussian) noise is added to the synthetic testing data. The non-monotonic increase of the recognition errors for the dyke conductivity contrast (C/C_2) and dip angle (A), as the level of noise increases up to 100 per cent, indicates that standard methods of noise reduction may not work in ANN-based inversion, so development of a special noise treatment methodology is required (Spichak 1999; Spichak *et al.* 1999).

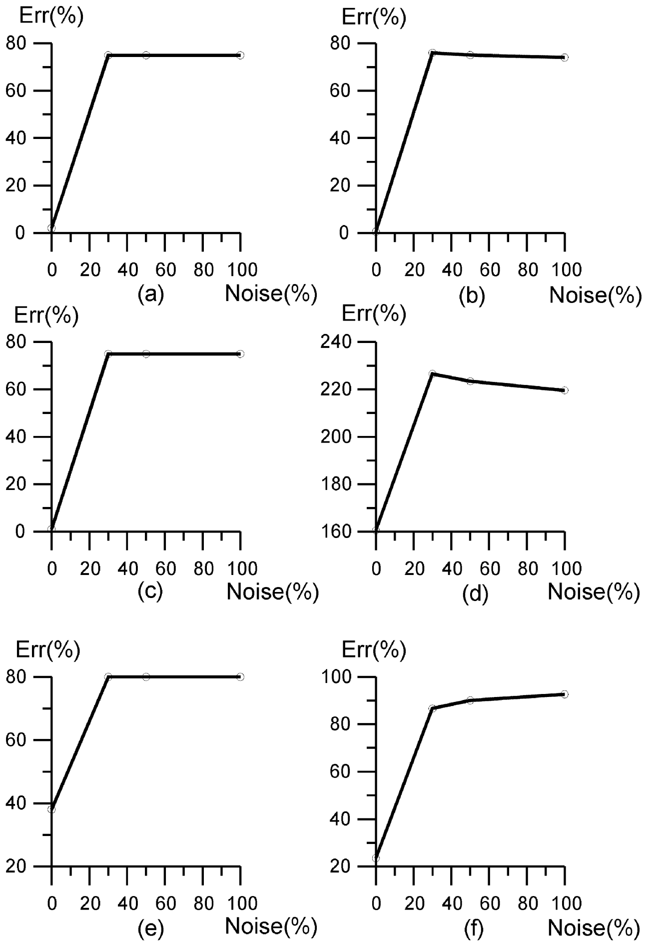

Figure 8. The effect of the noise level (Noise) in the testing data on the recognition errors for Test 3.

9 CONCLUSIONS

A new approach to 3-D magnetotelluric data inversion based on a neural-network paradigm is developed. It is tested using synthetic AMT data created by forward modelling for a 3-D geoelectrical model of a dyke buried in the second layer of a two-layer earth. It is shown that the ANN recognition can be successfully used for inversion if the data correspond to the model class familiar to the ANN. No initial guess regarding the parameters of 1-D layering or the geometry of the domain of search is required.

It is worth noting that, in spite of the fact that the teaching process based on forward modelling is a common feature of practically all known inversion techniques, the ANN approach to inversion differs from others in that here the teaching and recognition/inversion procedures are separated in time, which, in turn, enables practically instantaneous inversion of real data even using microcomputers. Moreover, the ability of the ANN to be taught once to recognize parameters in a given model class enables the effective 'cost' of one inversion to be decreased by a factor equal to the number of data inversions carried out.

Unfortunately, the recognition abilities of an ANN are restricted by its 'education' level: the more models (classes) of different geological situations are familiar to it, the better the inversion results. Meanwhile, the same database could be used for teaching in the case when another frequency range and geometrical parameters' scale satisfy the well-known electrodynamic similitude relation (8).

The ability of an ANN to teach itself by real geophysical (not only electromagnetic) data measured at the same place over a sufficiently long period gives an impetus to using this approach for the interpretation of monitoring data.

ACKNOWLEDGMENTS

The authors wish to thank Dr E. Fomenko and Mrs K. Nalivaiko who carried out intensive forward modelling by means of the program package FDM3D, and Dr A. Ezhov who adjusted the ANN parameters. Special acknowledgments are due to Dr T. Mogi (Hokkaido University, Japan), Dr H. Shima, Mr K. Fukuoka and Mr T. Kobayashi (OYO Corporation, Japan) and Dr M. Poulton (Arizona University, USA) for valuable suggestions and fruitful discussions during our meetings. The authors are also grateful to the referees for their helpful remarks, suggestions and comments.

We gratefully acknowledge financial support for this study from the OYO Corporation (grant EM-95) and Russian Basic Research Foundation (grant 99–05–64552).

REFERENCES

Baum, E. & Haussler, D., 1989. What size net gives valid generalization?, in *Advances in Neural Information Processing Systems,* Vol. 1, pp. 81–90, ed. Touretzky, D., Morgan Kaufmann Publishers, San Mateo, CA.

Caudill, M., 1988. Neural networks primer, Part 4, *AI Expert,* **8,** 61–67.

Hidalgo, H., Gomez-Trevino, E. & Swinarski, R., 1994. Neural network approximation of an inverse functional, in *Proc. IEEE World Congr. on Comput. Intelligence,* pp. 3387–3392, Orlando.

Kung, S. & Hwang, J., 1988. An algebraic projection analysis for optimal hidden units size and learning rates in backpropagation learning, in *1st Int. Conf. on Neural Networks, IEEE Proc.,* pp. I-363–370, eds Caudill, M. & Batler, C., SoS Printing, San Diego.

Mackie, R.L. & Madden, T.R., 1993. Three-dimensional magnetotelluric inversion using conjugate gradients, *J. Geophys.,* **115,** 215–229.

Poulton, M. & Birken, R.A., 1998. Estimating one-dimensional models from frequency-domain electromagnetic data using modular neural networks, *IEEE Trans. GeoScience and Remote Sensing,* **36,** 547–559.

Poulton, M., Sternberg, B. & Glass, C., 1992a. Neural network pattern recognition of subsurface EM images, *J. appl. Geophys.,* **29,** 21–36.

Poulton, M., Stemberg, B. & Glass, C., 1992b. Location of subsurface targets in geophysical data using neural networks, *Geophysics,* **57,** 1534–1544.

Raiche, A., 1991. A pattern recognition approach to geophysical inversion using neural networks, *Geophys. J. Int.,* **105,** 629–648.

Rumelhart, D., Mcclelland, J. & the PDP Research Group, 1988. *Parallel Distributed Processing,* Vol. 1, MIT Press, Cambridge.

Schiffman, W., Joost, M. & Werner, R., 1992. Optimization of the backpropagation algorithm for training multilayer perceptrons, *Technical report,* Institute of Physics, University of Koblenz.

Schmidhuber, J., 1989. Accelerated learning in back-propagation nets, in *Connectionism in Perspective,* pp. 439–445, Elsevier, Amsterdam.

Sen, M.K., Bhattacharya, B.B. & Stoffa, P.L., 1993. Nonlinear inversion of resistivity sounding data, *Geophysics,* **58,** 496–507.

Silva, F.M. & Almeida, L.B., 1990. Speeding up backpropagation, in *Advanced Neural Computers,* pp. 151–158, ed. Eckmiller, R.

Soulie, F., Gallinari, P., Le, C. & Thiria, S., 1987. Evaluation of neural network architectures on test learning tasks, in *1st Int. Conf. on Neural Networks, IEEE Proc.,* II-653–660, eds Caudill, M. & Batler, C., SoS Printing, San Diego.

Spichak, V.V., 1983. The FDM3D software package for the numerical modeling of 3D electromagnetic fields, in *Algoritmy i Programmy Resheniya Pryamykh i Obratnykh Zadachakh Elektromagnitnoi Induktsii v Zemle (Algorithms and Programs for the Calculation and Inversion of the Electromagnetic Fields in the Earth),* pp. 58–68, ed. Zhdanov, M., Izmiran, Moscow (in Russian).

Spichak, V.V., 1990. A general approach to the EM data interpretation using an expert system, in *Proc. X Workshop on EM Induction in the Earth,* Ensenada, Mexico.

Spichak, V.V., 1999. *Magnitotelluricheskie Polja v Trekhmernikh Geoelektricheskikh Modeljakh (Magnetotelluric Fields in 3D Geoelectrical Models),* Scientific World, Moscow (in Russian).

Spichak, V., Menville, M. & Roussignol, M., 1995. Three-dimensional inversion of the magnetotelluric fields using Bayesian statistics, in *Proc. Schlumberger Doll Research Symp. on 3D Electromagnetics,* pp. 347–358, eds Spies, B. & Oristaglio, M., Ridgefield, USA.

Spichak, V., Fukuoka, K., Kobayashi, T., Mogi, T., Popova, I. & Shima, H., 1999. Neural networked based interpretation of insufficient and noisy MT data in terms of the target macroparameters, in *Proc. 2nd Int. Symp. on 3D Electromagnetics,* pp. 297–300, eds Wannamaker, P. & Zhdanov, M., Salt Lake City.

Stratton, J.A., 1941. *Electromagnetic Theory,* McGraw-Hill, New York.

Yoshifusa, I., 1992. Approximation of continious fuctions on R^d by linear combination of shifted rotation of a sigmoid function with and without scaling, *Neural Networks,* **4,** 817–826.

Zhdanov, M.S. & Spichak, V.V., 1992. *Matematicheskoe Modelirovanie Elektromagnitnykh Polei v Trekhmerno Neodnorodnykh Sredakh (Mathematical Modeling of Electromagnetic Fields in 3D Inhomogeneous Media),* Nauka, Moscow (in Russian).