Extracting pattern from scattered data — applicability of artificial neural networks to the interpretation of bearing capacity data

by

by John D. McKinley

CUED / D - SOILS / TR.299 (1996)

Extracting pattern from scattered data — applicability of artificial neural networks to the interpretation of bearing capacity data

by John D. McKinley[1]

INTRODUCTION

There has recently been considerable interest in the application of artificial neural networks in civil engineering, frequently to classification problems. Examples include recommending a vertical formwork system given specified information about the building characteristics and plant availability (Kamarthi et *al.*, 1992) and diagnosing damage of prestressed concrete piles given the presence or absence of specified symptoms (Yeh et *al.*, 1993). Other examples can be found in the recent ASCE ***Journal of Computing in Civil Engineering*** special edition on artificial neural networks. Frequently the input parameters to the analysis take discrete values. However, the layered feed-forward back-propagation model commonly employed is based on continuous transfer functions, which indicates that it can be employed where the input and output parameters take continuous values, and therefore to problems where multivariate regression analysis would be applicable. The purpose of this note is to compare the performance of an artificial neural network with that of conventional regression analysis in characterising idealised bearing capacity data.

ARTIFICIAL NEURAL NETWORKS

Detailed descriptions of the philosophy, design and operation of various artificial neural networks can be found in Rumelhart *et al.* (1986) and in Eberthart and Dobbins (1992). There is insufficient space here for a full presentation, but an appreciation of the basic features of the layered feed-forward back-propagation model is necessary, since this is the model used in this note. A brief summary can be found in Appendix I.

Is is possible, though tedious, to write down for a given design of such an artificial neural network (ANN) a single equation giving the value at any one output node as a function of the values at the input nodes and the connection weights. This equation will be large for any non-trivial networks, but the reader should realise that the ANN training algorithm is nothing more than an iterative, trial-and-error method for estimating the parameters in this equation. In essence, ANNs of the type described here fit this overall equation to the training data using a form of regression analysis.

[1] University Assistant Lecturer, Geotechnical Group, Department of Engineering, University of Cambridge.

BEARING CAPACITY OF A STRIP FOUNDATION

Idealised problem

In order to compare the performance of an ANN with that of conventional regression analysis in extracting pattern from scattered data and to evaluate the effect of the level of uncertainty, the author has chosen an idealised problem with a known, exact solution: the drained bearing capacity of a shallow strip foundation on weightless sand. Figure 4 shows the geometry of the problem, a foundation of width B resting on dry sand of apparent cohesion c and friction angle ϕ'.

Various expressions to estimate the ultimate bearing capacity q_u exist; the one used in this study is that suggested by Whitlow (1990):

$$q_u = cN_c + \sigma'_{vo}N_q \quad (1)$$

where $$N_q = \exp(\pi \tan \phi') \tan^2\left(45° + \frac{\phi'}{2}\right) \quad (2)$$

and $$N_c = (N_q - 1)\cot\phi' \quad (3)$$

where σ'_{vo} is the overburden pressure.

Artificially generated test data

The artificial test data was generated from equations (1), (2) and (3) using values for the input variables randomly distributed in the following ranges:

$$0 < c < 50 \text{ kPa} \quad (4)$$

$$0 < \phi' < 40° \quad (5)$$

$$0 < \sigma'_{vo} < 40 \text{ kPa} \quad (6)$$

which gives q_u values in the range O-6333 kPa. The random number generator used returns a random number evenly distributed over the specified range.

Scatter or uncertainty was introduced by applying a further random component ξ to the ultimate bearing capacity such that:

$$4_{u,t} = q_u(1 + \xi) \quad (7)$$

where $q_{u,t}$ is the test ultimate bearing capacity, ξ lies in the following range:

$$-\xi_0 < \xi < \xi_0 \quad (8)$$

and ξ_0 is the range of scatter. No scatter was introduced in the input variables, so the training data simulates the case where the soil parameters and overburden are known precisely, but there is uncertainty in the ultimate bearing capacity determined experimentally.

The maximum value of ξ_0 used in this study was 0·5, equivalent to a 50% 'error' in the overall measurement, giving possible $q_{u,t}$ values in the range O-10000 kPa although there were few $q_{u,t}$ values greater than 5000 kPa. The $q_{u,t}$ values where scaled into the range O-l by dividing by 10000; the input variables c, ϕ' and σ'_{vo} where similarly scaled using the maximum values indicated in equations (4), (5) and (6) respectively.

In order to evaluate the sensitivity of the ANN's performance to the number of training pairs n_{train} and to the amount of scatter, several sets of training data were generated. Table 1 specifies these sets of training data. In test series SF, the training data sets were generated as just described. In test series SE a new input variable E was introduced which takes random values over the range O-l but which has no effect on $q_{u,t}$; in field data E would be a variable which is believed to affect the bearing capacity but in fact does not, so this series evaluates the effect of including irrelevant variables. Test RE2500.50 used the same training data as test SE2500.50 but with the ANN initialised to a new set of random weights, to examine the repeatability of the ANN results.

ANALYSIS OF ARTIFICIAL BEARING CAPACITY DATA

The training data sets were applied separately to a series of identical ANNs, one per set. For test series SF the ANN had three input nodes (representing the three input variables c, ϕ', and σ'_{vo}), four hidden nodes, and one output node (representing the bearing capacity), while for test series SE the ANN had four input nodes (as before, plus E), five hidden nodes and one output node. There is no generally accepted algorithm for determining how many hidden nodes to use, but values for $n_j \approx n_i + n_l$ are typical. Following some initial experimentation, bias was applied to the input nodes. This unusual approach was found to shift the mean operating point of the input nodes towards the range where the nodal output is most sensitive to changes in the input values, although the effect is slight. Consideration of the ANN equations show that there are twenty-four adjustable parameters for the {3,4,1} design and thirty-five adjustable parameters for the { 4,5,1} design, both of which are smaller than the number of data points in all tests except those in series SF0020 and SE0020. These series are under-specified, and the ANNs trained on these data sets should perform poorly, as would any model generated by regression analysis where there are more adjustable parameters than there are known data points.

The tolerance value for the mean sum-squared error was set at 0·0001, giving a mean error in the scaled bearing capacity of 0.01. This is equivalent to an error of ±100 kPa in any one data point. From the results of the initial experimentation, the learning coefficient and momentum factor were both set to 0·5.

In addition, conventional regression analysis was done of the training data sets. Two fitting schemes to estimate the ultimate bearing capacity were considered, an exact scheme and a simple scheme. For test series SF these were:

$$q_u = a_1 c N_c + a_2 \sigma'_{vo} N_q \tag{9}$$

and $$q_u = a_3 c + a_4 \sigma'_{vo} + a_5 \sigma'_{vo} \tan\phi' \tag{10}$$

respectively, while for test series SE they were:

$$q_u = a_6 c N_c + a_7 \sigma'_{vo} N_q + a_8 E \tag{11}$$

and $$q_u = a_9 c + a_{10} \sigma'_{vo} + a_{11} \sigma'_{vo} \tan\phi' + a_{12} E \tag{12}$$

respectively, where $a_1 \ldots a_{12}$ are dimensionless constants determined by least squares adjustment. The exact scheme has the same form as equation (1). The simple scheme is a linear combination of components similar to that used by Kramer (1977) to determine the carrying capacity of grouted ground anchors by regression analysis of field data. Fitting the exact scheme, equations (9) or (11), could recover the exact solution, while equations (10) or (12) would be a reasonable simple relationship to use if the actual solution were unknown.

In order to evaluate the performance of the ANN and the equations generated by regression analysis, a further data set was produced. This evaluation set consisted of 250 data points with values for c, ϕ', σ'_{vo} and E (for test series SE) generated randomly in the same way as for the training data. The ultimate bearing capacities predicted by the trained ANNs and by the results of the regression analysis were compared to the true value from equation (1), and two error terms calculated: the actual error, χ_{act}, and the relative error, χ_{rel}, calculated from:

$$\chi_{act} = q_{u,p} - q_u \tag{13}$$

and

$$\chi_{rel} = \left| \frac{q_{u,p} - q_u}{q_u} \right| \tag{14}$$

respectively, where $q_{u,p}$ is the predicted ultimate bearing capacity.

RESULTS

In some cases the tolerance error values for ANN training could not be reached. Instead, the ANN's training error gradually reduced to some apparent minimum value but further training caused the ANN to settle in a stable state where the weights did not change significantly but the training error was very high. This generally occurred only for those training sets with high degrees of scatter, for which the tolerance criterion was relaxed and the weights corresponding to this apparent minimum error were accepted.

The means of χ_{act} and χ_{rel}, $\overline{\chi}_{act}$ and $\overline{\chi}_{rel}$ respectively, were calculated for each data set. These are tabulated in Table 1. Figure 5 shows plots of the bearing capacity predicted by each of the three schemes against the theoretical bearing capacity for test SF2500.00, while figures 6, 7, and 8 show the resulting plots for tests SF2500.50, SF0020.50 and SE0020.50 respectively.

The basic pattern is quite clear. The plotted results show that the predicted values lie in a band whose width increases as the bearing capacity increases, indicating that $\bar{\chi}_{rel}$ is a meaningful measure of the scatter in the predictions, and that both the ANN and the exact scheme generate predicted values which are scattered around the true value over the whole range while the simple scheme tends to predict lower than actual bearing capacities towards the upper and lower ends. In some cases, the simple scheme predicts negative bearing capacities.

The tabulated results indicate that the absolute error $\bar{\chi}_{act}$ is generally small for all three schemes except when there are a small number of training pairs, or large amounts of scatter in the training data. Both the actual error and the relative error tend to increase as the sample size decreases and as the degree of scatter increases, which is to be expected. None of the schemes appears to be disastrously affected by the introduction of irrelevant input variables.

Comparison of test RE2500.50 with SE2500.50 shows that there could be a significant variation in the performance of a particular ANN design even for a single set of training data, for different initial weights. This indicates that it will be worthwhile to repeat the training several times and to then select the ANN with the best performance over a set of evaluation data.

Generally, the exact scheme has the lowest $\bar{\chi}_{act}$ while the ANN has the highest, and the exact scheme has the lowest $\bar{\chi}_{rel}$ while the simple scheme has the highest. That is, all three schemes predict approximately the correct bearing capacity on average, but in terms of the error in a single prediction the exact scheme significantly outperforms the ANN, and the ANN in turn significantly outperforms the simple scheme.

Curiously, all three schemes predict lower than actual bearing capacities at the upper end in test SE0020.50. This was common to the tests in series SE, and is reflected in the generally negative $\bar{\chi}_{act}$ values calculated. Also curious is the fact that all three schemes tended to give lower relative errors in test series SE than in the corresponding tests in series SF, which may reflect differences in the distribution of the original input variables.

CONCLUSIONS

The results of this study indicate that artificial neural networks can successfully characterise the underlying patterns in scattered and uncertain data such as might be obtained from bearing capacity tests. For the artificial data used, the ANNs performed less well than conventional regression analysis where the basic relationship was known, but much better than a simple regression scheme where the basic relationship was not known. In the latter situation, the civil engineer should consider the ANN technique a useful tool for extracting pattern from large volumes of existing data for the purposes of predicting the results of future, similar setups. Suitable applications would be predictions of pile settlement or grouted ground anchor capacities (McKinley, 1993). Some experimentation in network design and training parameters will be necessary.

APPENDIX I

Design and structure of artificial neural networks

An artificial neural network (ANN) is a signal processing unit, mapping a set of input data on to a set of output data. The basis of the ANN is the node, which receives a number of inputs, implements some transfer function and generates an output. The node represented in Figure 1 has five variable inputs plus a bias input, which is constant, and each input has an associated weighting factor w.

Nodes are organised into a network structure as shown in Figure 2, with three layers-an input layer of n_i nodes, a hidden layer of n_j nodes and an output layer of $n,$ nodes. There is one input node per input variable and one output node per output variable. Conventionally, the set of input data to the ANN and the corresponding set of output data from the ANN are together referred to as a 'pair'. Each node in a given layer receives input from all of the nodes in the previous layer, including the bias input, and each connection between nodes has an associated weight. In this particular ANN bias is also applied to the input nodes.

The networks in this note employed a linear transfer function for the input nodes, where the nodal output is the sum of the nodal input and the bias, and a sigmoidal transfer function for hidden and output nodes, defined by equation (15):

$$output = \frac{1}{1 + \exp\left(-\sum weights \ x \ inputs\right)} \qquad (15)$$

Figure 3 illustrates the relationship between the nodal output and the weighted sum of the nodal inputs, where the bias is treated as a constant input. The nodal output tends towards one as the weighted sum becomes large and positive, towards zero as the weighted sum becomes large and negative, and is bounded between zero and one. Clearly, the nodal output is most sensitive to variations in the nodal input values where the weighted sum is small.

Initialisation and training

The weights associated with the nodal connections encode the mapping between the input data and the output data. These weights are determined by presenting the ANN with numerous examples where both the output and the input are known, and adjusting the weights until the ANN successfully maps the input data on to the output data for these training pairs. This process is called 'training'. The training algorithm is described in Rumelhart *et al.* (1986), and in principle the procedure is as follows: present the ANN with the input data set of a pair and calculate the nodal error as the desired output value minus the computed output value, for each output node; propagate this nodal error back through the network by decreasing those weights which tend to increase the nodal error and increasing those weights which tend to decrease the nodal error, a process called 'gradient descent'; calculate the sum of the squares of the nodal errors for the pair; repeat this process for all training pairs and calculate the mean value of the sum-squared error; the goal of the training process is to reduce this mean sum-squared error over all training pairs to below some acceptable tolerance value set by the ANN designer, so if the error is larger than the tolerance value iterate the process with the new, updated weights.

On a practical level, there are other considerations. Firstly, the weights are not generally updated after each training pair but instead the changes are accumulated separately during each iteration and applied after all of the training pairs have been presented. Secondly, instead of updating the weights by the whole amount calculated, the change made is a proportion of this amount plus a proportion of the change made in the previous iteration. These proportions are the learning coefficient and the momentum factor respectively, will both be in the range 0-1, and their use has been found to lead to faster, less oscillatory training (Rumelhart *et al.,* 1986). Thirdly, in order to feed the first set of input data forward the weights need to have initial values and the training algorithm breaks down if these are zero, so they are initialised to random values in the range -1 to +1. Other, tighter ranges are sometimes used. Fourthly, because the nodal outputs must be between 0 and 1, it is necessary to pre-scale the output values into this range and it is usual to pre-scale the input values similarly.

NOTATION

$a_1 \ldots a_{12}$	dimensionless constants determined by regression analysis
B	width of strip footing
c	apparent cohesion
E	input variable which has no effect on the bearing capacity
N_c, N_q	bearing capacity factors
n_{train}	number of training pairs
n_i	number of nodes in ANN input layer
n_j	number of nodes in ANN hidden layer
n_l	number of nodes in ANN output layer
q_u	true ultimate bearing capacity
$q_{u,p}$	predicted ultimate bearing capacity
$q_{u,t}$	test ultimate bearing capacity
w	weighting factor of a connection between nodes
ξ	scatter factor applied to artificial data
ξ_0	range of scatter in artificial data
σ'_{vo}	overburden pressure
ϕ'	friction angle
χ_{act}	actual error in the prediction of q_u
$\bar{\chi}_{act}$	mean of the actual error in the prediction of q_u
χ_{rel}	relative error in the prediction of q_u

$\overline{\chi}_{rel}$ mean of the relative error in the prediction of q_u

REFERENCES

Eberhart, R.C. and Dobbins, R.W. (1990). ***Neural network PC tools: a practical guide.*** Academic Press Ltd..

Kamarthi, S.V., Sanvido, V.E., and Kumara, S.R.T. (1992). "Neuroform-neural network system for vertical formwork selection". ***ASCE Journal of Computing in Civil Engineering,*** 6(2), 178-199.

Kramer, H. (1977). "Determination of the carrying capacity of ground anchors with the correlation and regression analysis". ***Proceedings of the Ninth International Conference on Soil Mechanics and Foundation Engineering Speciality Session Four,*** pp. 76-8 1.

McKinley, J.D. (1993). ***Grouted ground anchors and the soil mechanics aspects of cement grouting.*** PhD thesis, University of Cambridge.

Rumelhart, D.E., McClelland, J.L., and the PDP Research Group, (1986). ***Parallel distributed processing.*** MIT Press, Vol. 1: Foundations.

Whitlow, R. (1990). ***Basic soil mechanics.*** Longman Scientific and Technical, 2nd.

Yeh, Y.C., Kuo, Y.H., and Hsu, D.S. (1993). "Building KBES for diagnosing PC pile with artificial neural network". ***ASCE Journal of Computing in Civil Engineering,*** 7(1), 71-93.

Test	n_{train}	ξ_0	$\overline{\chi}_{act}$:kPa			$\overline{\chi}_{rel}$: %		
			ANN	Exact	Simple	ANN	Exact	Simple
SF2500.00	2500	0·00	19·2	0·0	5·4	26.5	0·0	111·1
SF2500.05	2500	0·05	31·3	-2·1	3·7	26·7	0·4	110·7
SF2500.10	2500	0·10	-8.9	-1·6	4·3	23.2	0·2	110·7
SF2500.25	2500	0.25	-35.2	1·9	6·4	23·6	0·7	111·3
SF2500.50	2500	0·50	80·9	5·9	11·0	36.1	2·1	113·4
SF0500.00	500	0·00	28.6	0·0	17.7	28.2	0·0	106·4
SF0500.05	500	0·05	40·1	1·4	18·4	28.6	0·2	106.8
SF0500.10	500	0·10	7.2	-4.6	14·2	24.6	o-7	105·4
SF0500.25	500	0.25	-2.6	8.7	22·6	23.1	1·1	107·8
SF0500.50	500	0·50	87.4	48·5	50·1	33·4	7·5	119.0
SF0100.00	100	0·00	61.1	0·0	30·1	31·0	o-o	107·1
SF0100.05	100	0·05	70.7	-0·3	29.7	31·4	0·6	107·1
SF0100.10	100	0·10	41·2	-6.3	25.3	29·8	1·0	106·4
SF0100.25	100	0.25	60·5	6·0	31·6	30·8	0·8	109·1
SF0100.50	100	0·50	277-O	45·4	79.9	92.5	6.4	118·3
SF0020.00	20	0·00	131.8	0·0	72.3	38.6	o-o	149-l
SF0020.05	20	0.05	122.8	12·6	83.9	37·5	1·6	152.6
SF0020.10	20	0·10	125·3	-2.4	70.1	37·3	1·9	149·7
SF0020.25	20	0·25	-18·2	4·1	64.4	23.9	0·6	154·3
SF0020.50	20	0·50	29.7	38·3	80·6	34·6	5·1	165·3
SE2500.50	2500	0·50	-78.4	-15·5	-1.1	13·9	5·9	74·3
SE0500.50	500	0.50	-78.3	-16.1	18.5	17.0	5·4	78.9
SE0100.50	100	0·50	-108·1	-8.2	15·0	26.5	3·7	78.7
SE0020.50	20	0·50	41·5	-110·8	60.5	39.8	12.7	72.3
RE2500.50	2500	0·50	-92.0	–	–	25.4	–	–

Table 1 Training data sets: specifications and results

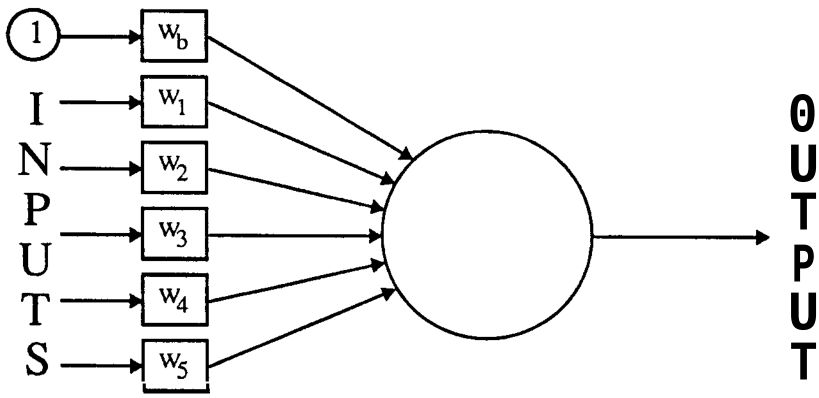


Figure 1 Basic element of an artificial neural network

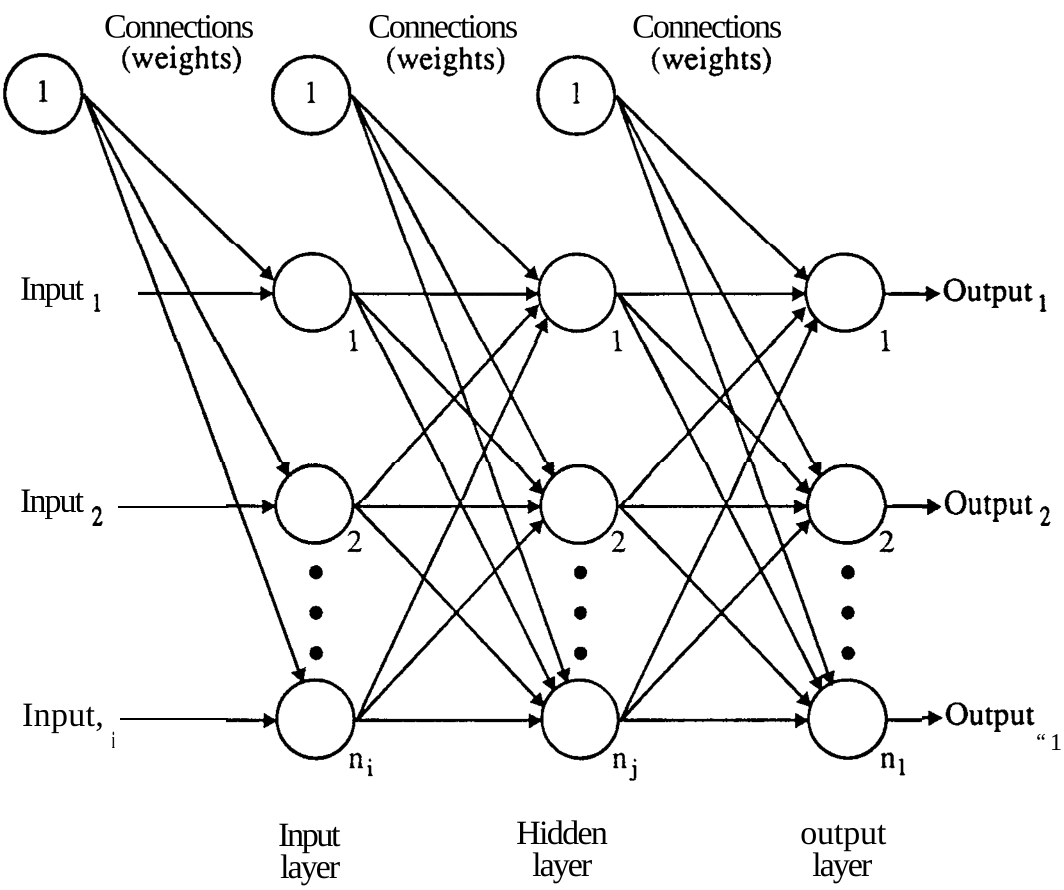


Figure 2 Back-propagation network structure (after Eberthart and Dobbins, 1992)

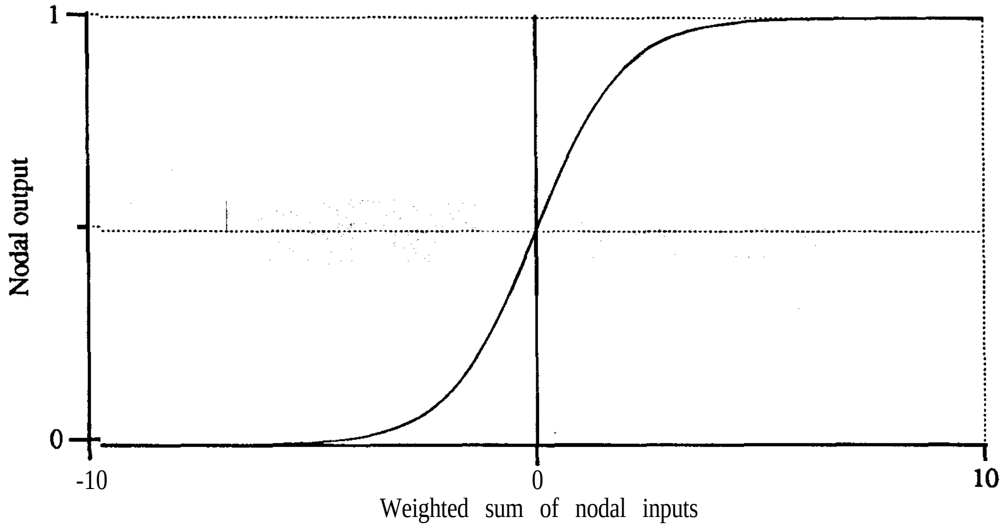


Figure 3 Sigmoidal transfer function

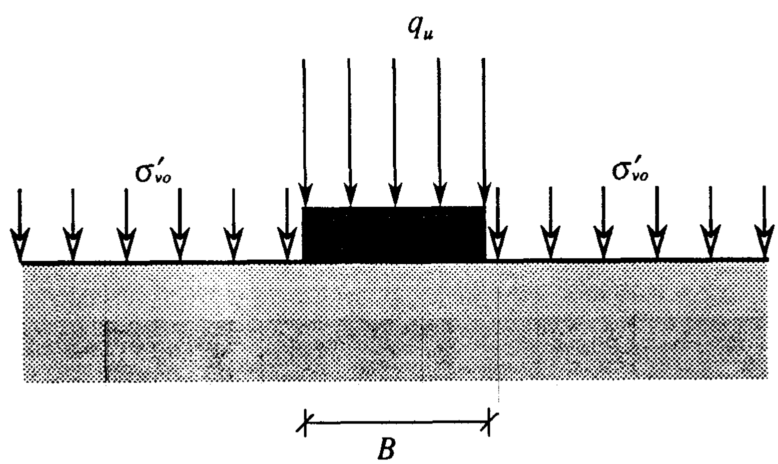
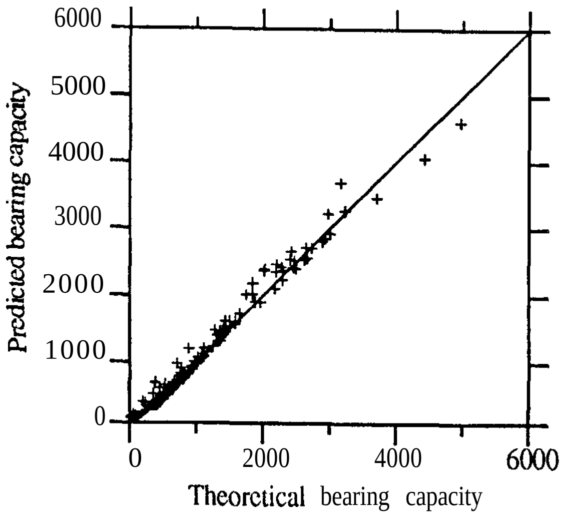
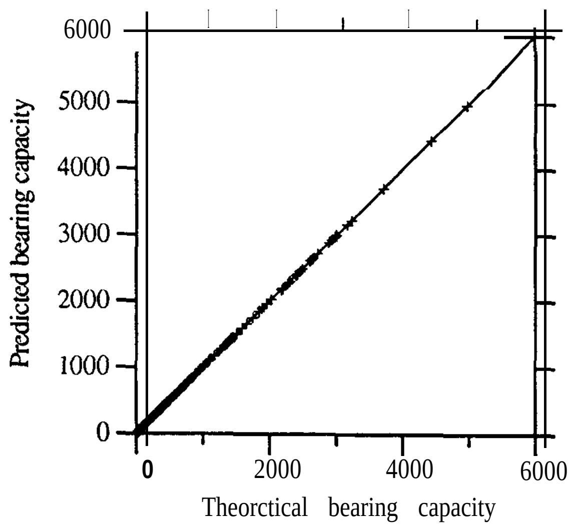


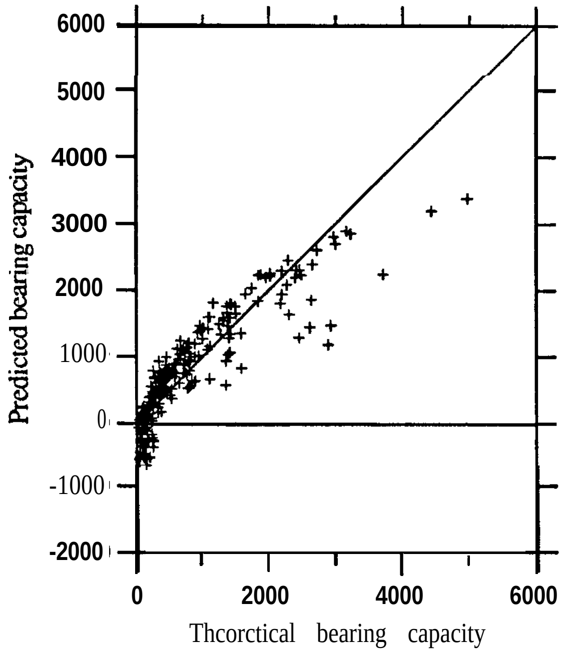
Figure 4 Geometry of the shallow strip foundation



(a) Artificial neural network

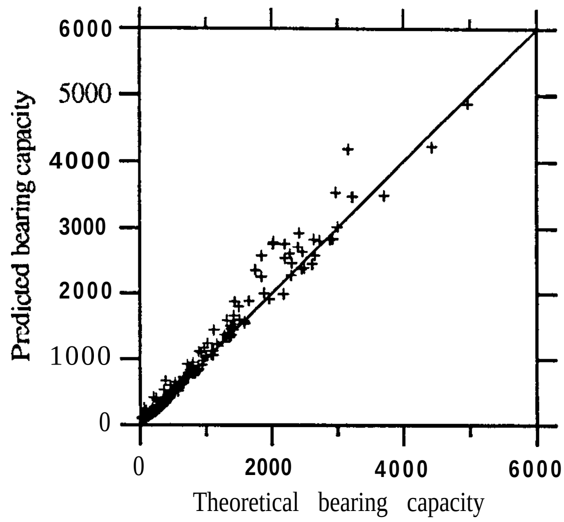


(b) Exact regression equation

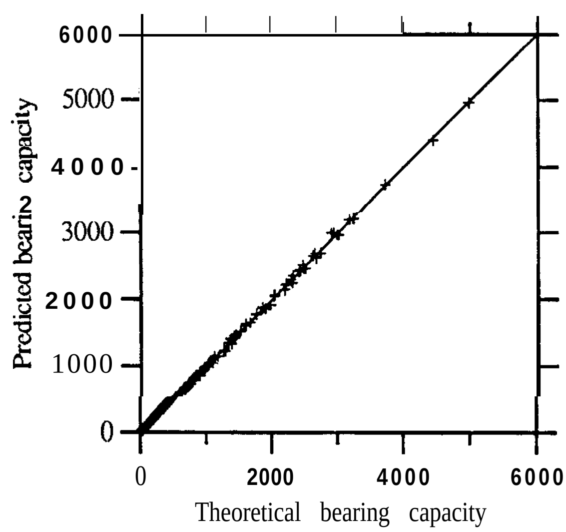


(c) Simple regression equation

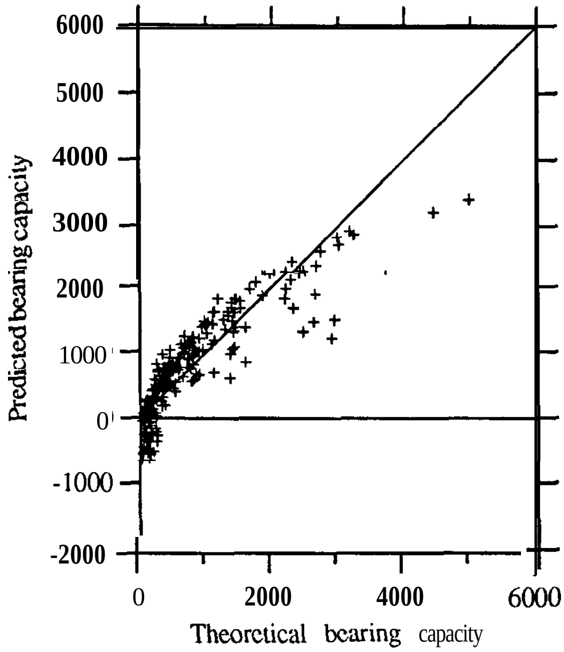
Figure 5 Predicted bearing capacities for data set SF2500.00



(a) Artificial neural network

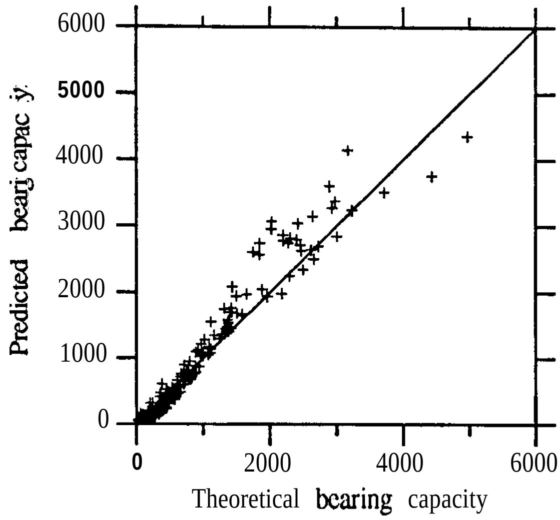


(b) Exact regression equation

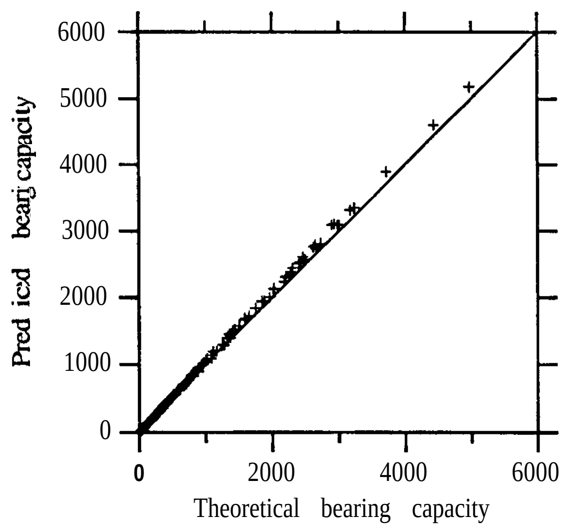


(c) Simple regression equation

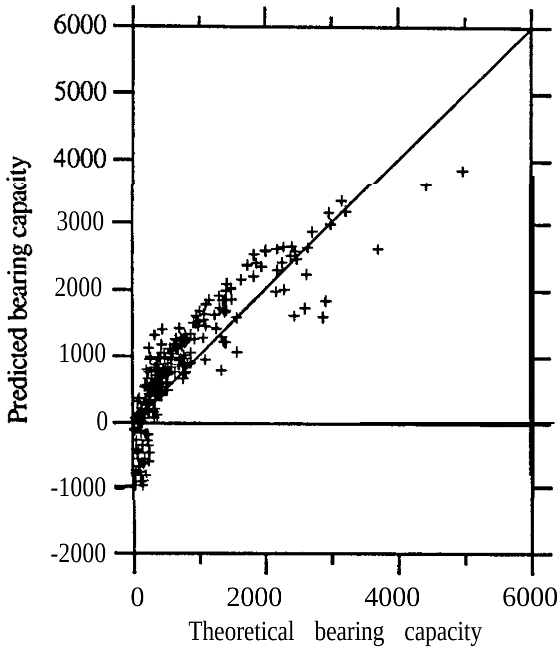
Figure 6 Predicted bearing capacities for data set SF2500.50



(a) Artificial neural network

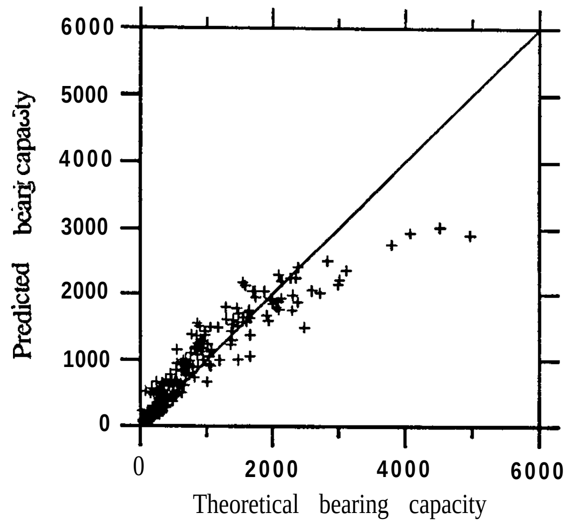


(b) Exact regression equation

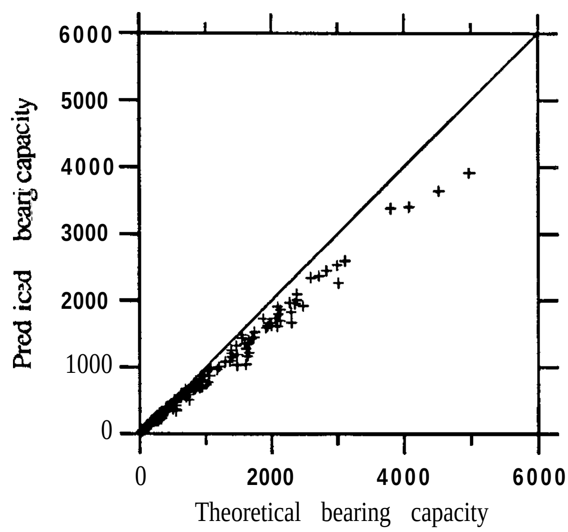


(c) Simple regression equation

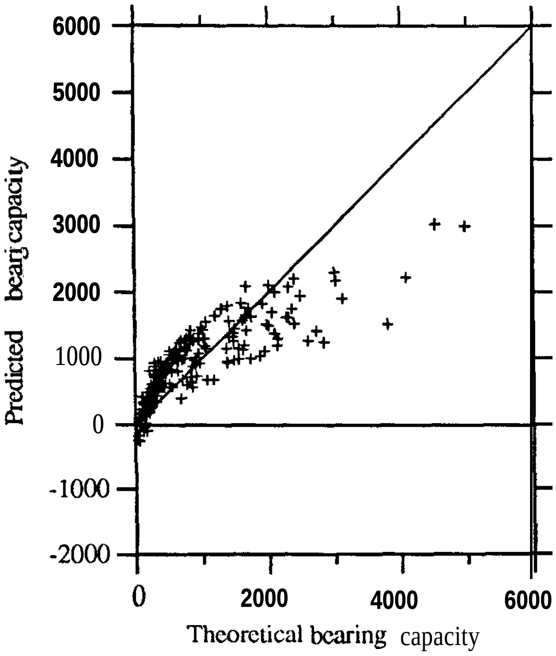
Figure 7 Predicted bearing capacities for data set SF0020.50



(a) Artificial neural network



(b) Exact regression equation



(c) Simple regression equation

Figure 8 Predicted bearing capacities for data set SE0020.50