Predicting Settlement of Shallow Foundations using Neural Networks

Article *in* Journal of Geotechnical and Geoenvironmental Engineering · September 2002

DOI: 10.1061/(ASCE)1090-0241(2002)128:9(785)

Predicting Settlement of Shallow Foundations using Neural Networks

Mohamed A. Shahin[1]; Holger R. Maier[2]; and Mark B. Jaksa[3]

Abstract: Over the years, many methods have been developed to predict the settlement of shallow foundations on cohesionless soils. However, methods for making such predictions with the required degree of accuracy and consistency have not yet been developed. Accurate prediction of settlement is essential since settlement, rather than bearing capacity, generally controls foundation design. In this paper, artificial neural networks (ANNs) are used in an attempt to obtain more accurate settlement prediction. A large database of actual measured settlements is used to develop and verify the ANN model. The predicted settlements found by utilizing ANNs are compared with the values predicted by three of the most commonly used traditional methods. The results indicate that ANNs are a useful technique for predicting the settlement of shallow foundations on cohesionless soils, as they outperform the traditional methods.

DOI: 10.1061/(ASCE)1090-0241(2002)128:9(785)

CE Database keywords: Shallow foundations; Settlement; Neural networks; Cohesionless soils; Prediction.

Introduction

It is generally understood that sand deposits are much more heterogeneous than their clay counterparts. As a result, differential settlements are likely to be higher in sand deposits than in clay profiles (Maugeri et al. 1998). Because cohesionless soils exhibit high degrees of permeability, settlement occurs in a short time; immediately after load application (Coduto 1994). Such immediate settlement causes relatively rapid deformation of superstructures, which results in an inability to remedy damage and to avoid further deformation. Furthermore, excessive settlement occasionally leads to structural failure (Sowers 1970).

The two major criteria that control the design of shallow foundations on cohesionless soils are the bearing capacity of the footing and settlement of the foundation. However, settlement usually controls the design process, rather than bearing capacity, especially when the breadth of footing exceeds 1 m (3–4 ft) (Schmertmann 1970). As a consequence, settlement prediction is a major concern and is an essential criterion in the design of shallow foundations.

The problem of estimating the settlement of shallow foundations on cohesionless soils is very complex and not yet entirely understood. This can be attributed to the uncertainty associated with the factors that affect the magnitude of this settlement. Among these factors are the distribution of applied stress, the stress–strain history of the soil, soil compressibility, and the difficulty in obtaining undisturbed samples of cohesionless soil (Moorhouse 1972; Holzlohner 1984). The geotechnical literature has included many methods, both theoretical and experimental, to predict the settlement of shallow foundations on cohesionless soils. Due to the difficulty of obtaining undisturbed samples of granular soil, many settlement prediction methods have focused on correlations with in situ tests, such as the cone penetration test, standard penetration test (SPT), dilatometer test, plate load test, pressuremeter test, and screw plate load test. However, most of the available methods simplify the problem by incorporating several assumptions associated with the factors that affect the settlement of shallow foundations. Consequently, most of the existing methods fail to achieve consistent success in relation to accurate settlement prediction (Poulos 1999). Comparative studies of the available methods (e.g., Jeyapalan and Boehm 1986; Tan and Duncan 1991; Wahls 1997) indicate inconsistent prediction for the magnitude of settlement calculated by these methods. As a result, alternative methods are needed, which provide more accurate settlement prediction.

In recent times, artificial neural networks (ANNs) have been applied to many geotechnical engineering tasks and have demonstrated some degree of success. Recent state-of-the-art ANN applications in geotechnical engineering have been summarized by the Transportation Research Board (TRB) (1999). ANNs are a form of artificial intelligence, which, by means of their architecture, attempt to simulate the biological structure of the human brain and nervous system. Although the concept of artificial neurons was first introduced in 1943, research into applications of ANNs has blossomed since the introduction of the back-propagation training algorithm for feedforward ANNs in 1986 (Rumelhart et al. 1986). ANNs may thus be considered a relatively new tool in the field of prediction and forecasting.

Many authors have described the structure and operation of ANNs (e.g., Hecht-Nielsen 1990; Fausett 1994). When feedforward ANNs are used for prediction and forecasting, the modeling philosophy employed is similar to that used in the development of more conventional statistical models. In both cases, the purpose of the model is to capture the relationship between a historical set

[1]PhD Research Scholar, Dept. of Civil and Environmental Engineering, Adelaide Univ., South Australia 5005.

[2]Senior Lecturer, Dept. of Civil and Environmental Engineering, Adelaide Univ., South Australia 5005.

[3]Senior Lecturer, Dept. of Civil and Environmental Engineering, Adelaide Univ., South Australia 5005 (corresponding author). E-mail: mjaksa@civeng.adelaide.edu.au

Note. Discussion open until February 1, 2003. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on May 8, 2000; approved on March 22, 2002. This paper is part of the ***Journal of Geotechnical and Geoenvironmental Engineering***, Vol. 128, No. 9, September 1, 2002. ©ASCE, ISSN 1090-0241/2002/9-785–793/$8.00+$.50 per page.

of model inputs and corresponding outputs. This is achieved by repeatedly presenting examples of the input/output relationship to the model and adjusting the model coefficients (i.e., connection weights) in an attempt to minimize an error function between the historical outputs and the outputs predicted by the model.

Although some ANN models are not significantly different from a number of standard statistical models, they are extremely valuable as they provide a flexible way of implementing them. Model complexity can be varied simply by altering the transfer function or network structure. In addition, ANNs belong to the class of data driven approaches, whereas conventional statistical methods are model driven. In the latter, the structure of the model has to be determined first, before the unknown model parameters can be estimated. In the former, the data are used to determine the structure of the model as well as the unknown model parameters. Traditional empirical approaches for predicting settlement of shallow foundations are model driven (e.g., Schmertmann 1970; Schultze and Sherif 1973; Burland and Burbidge 1985). This may potentially compromise model performance, as the form of the equation chosen may be suboptimal. As a result, the use of ANN models may overcome the limitations of the traditional methods.

In this paper, ANNs are used to predict the settlement of shallow foundations on cohesionless soils. The objectives of the paper are:

1. To investigate the feasibility of the ANN technique for predicting the settlement of shallow foundations on cohesionless soils and to provide an executable program of the developed ANN model for routine use in practice;
2. To study the effect of ANN geometry and some internal parameters on the performance of ANN models;
3. To explore the relative importance of the factors affecting settlement prediction by carrying out a sensitivity analysis;
4. To compare the performance of the ANN model with some of the most commonly used traditional methods; and
5. To assess the benefits and limitations of the ANN technique over traditional methods.

Development of Neural Network Model

The steps for developing ANN models, as outlined by Maier and Dandy (2000), are used as a guide in this work. These include the determination of model inputs and outputs, division and preprocessing of the available data, the determination of appropriate network architecture, optimization of the connection weights (training), stopping criteria, and model validation. The personal computer-based software *NEUFRAME* Version 4.0 (2000) Neurosciences Corp., Southamption, Hampshire, U.K. is used to simulate ANN operation in this work.

The data used to calibrate and validate the neural network model were obtained from the literature and include field measurements of settlement of shallow foundations as well as the corresponding information regarding the footings and soil. The data cover a wide range of variation in footing dimensions and cohesionless soil types and properties. The database comprises a total of 189 individual cases; 125 cases were reported by Burland and Burbidge (1985), 22 cases by Burbidge (1982), five cases by Bazaraa (1967), and 30 cases by Wahls (1997). Another four cases are given by Briaud and Gibbens (1999), one case by Picornell and Del Monte (1988), and two cases by Maugeri et al. (1998).

Model Inputs and Outputs

A thorough understanding of the factors affecting settlement is needed in order to obtain accurate settlement prediction. Most traditional methods include, as the main factors affecting settlement, footing width (B), footing net applied pressure (q), and soil compressibility within the depth of influence of the foundation. There are some other factors, such as the depth of water table, footing geometry (length to width) (L/B), footing embedment ratio (D_f/B), and the thickness of the soil layer beneath the foundation that contribute to a lesser degree and thus can be considered secondary (Burland and Burbidge 1985). The depth of the water table is not included in this study, as it is believed that its effect is already reflected in the measured SPT blow count (Meyerhof 1965) which, as discussed below, is used as a measure of soil compressibility. There are insufficient data for the thickness of soil layer in the available database and thus it is not considered in this study. Circular footings are also considered to be equivalent to square footings ($L/B=1$), as it was found by Burbidge (1982) that there is no significant difference between the settlement of circular or square footings having the same width (B) on the same soil.

Soil compressibility within the depth of influence of a foundation requires the assignment of soil properties that can accurately reflect this compressibility and the assignment of a depth over which the compressibility of the soil beneath the footing significantly influences the settlement. The SPT is one of the most commonly used tests in practice for measuring the compressibility of cohesionless soils (D'Appolonia and D'Appolonia 1970). While it is not the most accurate in situ method for measuring soil compressibility, it is used extensively worldwide. Consequently, for the purpose of this study, the average SPT blow count/300 mm (N) over the depth of influence of the foundation is used as a measure of soil compressibility.

Burland and Burbidge (1985) recommended no correction to N be taken for overburden pressure or submergence. However, for very fine and silty sand below the water table, Burland and Burbidge (1985) used the submergence correction proposed by Terzaghi and Peck (1948) when $N>15$ as follows:

$$N_{\text{corrected}}=15+0.5(N-15) \tag{1}$$

For gravel or sandy gravel, Burland and Burbidge (1985) proposed a correction for N as follows:

$$N_{\text{corrected}}=1.25N \tag{2}$$

Since most case records in the database used in the present study were obtained from Burland and Burbidge (1985), these corrections were applied to the entire database.

There is no unanimous agreement in the literature for the definition of the depth of influence of a foundation. In this work, the following guidelines, proposed by Burland and Burbidge (1985), are used. When N is decreasing with depth, the depth of influence is taken to be equal to the lesser of $2B$ or the depth from the bottom of the footing to bedrock. On the other hand, when N is constant or increasing with depth, the depth of influence is taken to be equal to $B^{0.75}$.

The aforementioned factors [i.e., footing width (B), footing net applied pressure (q), average SPT blow count (N), footing geometry (L/B), and footing embedment ratio (D_f/B)] are presented to the ANN as model input variables. Settlement is the single output variable. In an attempt to identify which of the input variables has the most significant impact on settlement predictions, a sensitivity analysis is carried out on the trained network. A simple and innovative technique proposed by Garson (1991) is used to interpret the relative importance of the input variables by examining the connection weights of the trained network. For a network with one hidden layer, the technique involves a process of partitioning the hidden output connection weights into components associated with each input node. The method is illustrated in the Appendix for a neural network model with five inputs, two hidden layer nodes, and one output. The results of the sensitivity analysis are discussed later.

The sensitivity analyses are repeated for networks trained with different initial random weights in order to test the robustness of the model in relation to its ability to provide information about the relative importance of the physical factors affecting the settlement of shallow foundations. If the ratio of the number of free parameters (e.g., connection weights) to the number of data points in the training set is too large, the same model prediction error can be achieved with different combinations of weights. While this is acceptable if the trained model is to be used purely for predictive purposes, it makes it difficult to interpret the physical meaning of the relationship found by the ANN. In this research, the ratio of the number of weights to the number of data points in the training set is approximately 1:9 and training is repeated four times with different random starting weights and a sensitivity analysis is carried out.

Data Division and Preprocessing

It is common practice to divide the available data into two subsets; a training set, to construct the neural network model, and an independent validation set to estimate model performance in the deployed environment (Twomey and Smith 1997). However, dividing the data into only two subsets may lead to model overfitting. As a result, and as discussed later, crossvalidation (Stone 1974) is used as the stopping criterion in this study and, consequently, the database is randomly divided into three sets: training, testing, and validation. In total, 80% of the data are used for training and 20% are used for validation. The training data are further divided into 70% for the training set and 30% for the testing set.

Recent studies have found that the way the data are divided can have a significant impact on the results obtained (e.g., Tokar and Johnson 1999). Like all empirical models, ANNs are unable to extrapolate beyond the range of their training data. Consequently, in order to develop the best possible model, given the available data, all patterns that are contained in the data need to be included in the training set. Similarly, since the test set is used to determine when to stop training, it needs to be representative of the training set and should therefore also contain all of the patterns that are present in the available data. If all the available patterns are used to calibrate the model, then the most challenging evaluation of the generalization ability of the model is if all of the patterns are also part of the validation set. Consequently, it is essential that the data used for training, testing, and validation represent the same population (Masters 1993). In order to achieve this in the present study, several random combinations of the training, testing, and validation sets are tried until three statistically consistent data sets are obtained. The statistical parameters considered include the mean, standard deviation, minimum, maximum, and range. Despite trying numerous random combinations of training, testing, and validation sets, there are still some slight inconsistencies in the statistical parameters for the training, testing, and validation sets that are most closely matched (Table 1). This can be attributed to the fact that the data contain singular, rare events, that cannot be replicated in all three data sets. However, on the whole, the statistics are in good agreement and all three data sets may be considered to represent the same population. The data ranges used for the ANN model variables are given in Table 2.

Once the available data have been divided into their subsets, it is important to preprocess the data to a suitable form before they are applied to the ANN. Preprocessing the data by scaling them is important to ensure that all variables receive equal attention during training. The output variables have to be scaled to be commensurate with the limits of the transfer functions used in the output layer. Scaling the input variables is not necessary but is always recommended (Masters 1993). In this work, the input and output variables are scaled between 0.0 and 1.0, as the sigmoidal transfer function is used in the output layer.

Model Architecture

Determining the network architecture is one of the most important and difficult tasks in the development of ANN models. It requires the selection of the number of hidden layers and the number of nodes in each of these. It has been shown that a network with one hidden layer can approximate any continuous function, provided that sufficient connection weights are used (Hornik et al. 1989). Consequently, one hidden layer is used in this study.

The number of nodes in the input and output layers are restricted by the number of model inputs and outputs. The input layer of the ANN model developed in this work has five nodes, one for each of the model inputs [i.e., width of footing (B), footing net applied pressure (q), average SPT blow count (N), footing geometry (L/B), and footing embedment (D_f/B)]. The output layer has only one node representing the measured value of settlement (S_m).

In order to obtain the optimum number of hidden layer nodes, it is important to strike a balance between having sufficient free parameters (weights) to enable representation of the function to be approximated, and not having too many so as to avoid overtraining and to ensure that the relationship determined by the ANN can be interpreted in a physical sense. Overtraining is not an issue in this study, as crossvalidation is used as the stopping criterion. However, as just discussed, physical interpretation of the connection weights is important, and hence the smallest network that is able to map the desired relationship should be used. In order to determine the optimum network geometry, ANNs with one, two, three, five, seven, nine, and 11 hidden layer nodes are trained. It should be noted that 11 is the upper limit for the number of hidden layer nodes needed to map any continuous function for a network with five inputs, as discussed by Caudill (1988).

Weight Optimization (Training)

The process of optimizing the connection weights is known as "training" or "learning." As mentioned previously, this is equivalent to the parameter estimation phase in conventional statistical models. The aim is to find a global solution to what is typically a highly nonlinear optimization problem. As the prediction of settlement of shallow foundations on cohesionless soils does not involve any time-related parameter components, feedforward, rather than recurrent, networks are used. The method most commonly used for finding the optimum weight combination for feedforward neural networks is the back-propagation algorithm (Rumelhart et al. 1986), which is based on first-order gradient descent. Feedforward networks trained with the back-propagation algorithm have already been applied successfully to many geotechnical engineering problems (e.g., Goh 1994; Najjar and Basheer 1996), and are thus used in this work. Details of the

Table 1. Artificial Neural Network Input and Output Statistics

Model variables and data sets	Statistical Parameters				
	Mean	Std. dev.[a]	Min.[b]	Max.[c]	Range
Footing width, B (m)					
Training set	8.3	9.8	0.8	60.0	59.2
Testing set	9.3	10.9	0.9	55.0	54.1
Validation set	9.4	10.1	0.9	41.2	40.3
Footing net applied pressure, q (kPa)					
Training set	188.4	129.0	18.3	697.0	678.7
Testing set	183.2	118.7	25.0	584.0	559.0
Validation set	187.9	114.6	33.0	575.0	542.0
Average SPT blow count, N					
Training set	24.6	13.6	4.0	60.0	56.0
Testing set	24.6	12.9	5.0	60.0	55.0
Validation set	24.3	14.1	4.0	55.0	51.0
Footing geometry, L/B					
Training set	2.1	1.7	1.0	10.6	9.6
Testing set	2.1	1.9	1.0	9.9	8.9
Validation set	2.1	1.8	1.0	8.1	7.1
Footing embedment ratio, D_f/B					
Training set	0.52	0.57	0.0	3.4	3.4
Testing set	0.49	0.52	0.0	3.0	3.0
Validation set	0.59	0.64	0.0	3.0	3.0
Measured settlement, S_m (mm)					
Training set	20.0	27.2	0.6	121.0	120.4
Testing set	21.4	26.6	1.0	120.0	119.0
Validation set	20.4	25.2	1.3	120.0	118.7

[a]Std. dev. indicates standard deviation.
[b]Min. indicates minimum.
[c]Max. indicates maximum.

back-propagation algorithm are beyond the scope of this paper and can be found in many publications (e.g., Fausett 1994).

In this study, the general strategy adopted for finding the optimal parameters that control the training process is as follows. For each trial number of hidden layer nodes, random initial weights and biases are generated. The neural network is then trained with different combinations of momentum terms and learning rates in an attempt to identify the ANN model that performs best on the testing data. The momentum terms used in this study are 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.8, whereas the learning rates used are 0.005, 0.02, 0.1, 0.2, 0.4, and 0.6. Since the back-propagation training algorithm uses a first-order gradient descent technique to adjust the connection weights, it may get trapped in a local minimum if the initial starting point in weight space is unfavorable. Consequently, the model that has the optimum momentum term and learning rate is retrained a number of times with different initial weights and biases until no further improvement occurs.

Table 2. Data Ranges used for Artificial Neural Network Model Variables

Model variables	Minimum value	Maximum value
Footing width, B (m)	0.8	60.0
Footing net applied pressure, q (kPa)	18.3	697.0
Average SPT blow count, N	4.0	60.0
Footing geometry, L/B	1.0	10.5
Footing embedment ratio, D_f/B	0.0	3.4
Measured settlement, S_m (mm)	0.6	121.0

Stopping Criteria

Stopping criteria are those used to decide when to stop the training process. They determine whether the model has been optimally or suboptimally trained. As described earlier, the crossvalidation technique (Stone 1974) is used in this work as the stopping criterion, as it is considered to be the most valuable tool to ensure that overfitting does not occur (Smith 1993) and as sufficient data are available to create training, testing, and validation sets. The training set is used to adjust the connection weights. The testing set measures the ability of the model to generalize, and the performance of the model using this set is checked at many stages of the training process, and training is stopped when the error of the testing set starts to increase. The testing set is also used to determine the optimum number of hidden layer nodes and the optimum internal parameters (learning rate, momentum, and initial weights).

Model Validation

Once the training phase of the model has been successfully accomplished, the performance of the trained model is validated using the validation data, which have not been used as part of the

Table 3. Comparison of Artificial Neural Network (ANN) and Traditional Methods for Settlement Prediction

Factor considered	ANN	Meyerhof (1965)	Schultze and Sherif (1973)	Schmertmann et al. (1978)
Settlement equation	—	$S_c=\frac{8q}{N}$ for $B\leqslant 4$ ft $S_c=\frac{12q}{N}\left[\frac{B}{B+1}\right]^2$ for $B>4$ ft	$S_c=\frac{qF}{N^{0.87}(1+0.4D_f/B)}$	$S_c=C_1C_2q\Sigma\left(\frac{I_z}{E_z}\right)_i\Delta z_i$
Correction for submergence	—	For dense submerged silty sand when $N>15$	N as measured	N as measured
Correction for overburden	—	No	No	No
r^2	0.819	0.160	0.518	0.637
RMSE (mm)	11.04	25.72	23.55	23.67
MAE (mm)	8.78	16.59	11.81	15.69
Remarks	—	S_c (in.), q (t/ft^2), and B (ft)	S_c (cm), q (kg/cm^2), D_f (cm), B (cm), and F (cm^3/kg)	S_c (m), q (kPa), E_z (kPa), and Δz (m)

Note: S_c=calculated settlement; B=footing width; q=footing net applied pressure; N=average SPT blow count; D_f=depth of footing embedment; r^2=coefficient of determination; F=settlement coefficient (obtained from chart); I_z=strain influence factor (obtained from chart); C_1=correction factor for embedment; C_2=correction factor for creep; and E_z=soil Young's modulus at the middle of the ith layer of thickness Δz_i.

model building process. The purpose of the model validation phase is to ensure that the model has the ability to generalize within the limits set by the training data, rather than simply having memorized the input–output relationships that are contained in the training data.

The coefficient of determination (r^2), the root-mean-square error (RMSE), and the mean absolute error (MAE) are the main criteria that are used to evaluate the performance of the ANN models developed in this work. The coefficient of determination is a measure that is used to determine the relative correlation between two sets of variables. The RMSE is the most popular measure of error and has the advantage that large errors receive greater attention than smaller ones (Hecht-Nielsen 1990). In contrast, the MAE eliminates the emphasis given to large errors. Both RMSE and MAE are desirable when the data evaluated are smooth or continuous (Twomey and Smith 1997).

Traditional Methods for Settlement Prediction

Many traditional methods for settlement prediction of shallow foundations on cohesionless soils are presented in literature. Among these, three are chosen for the purpose of assessing the relative performance of the ANN model. These include the methods proposed by Meyerhof (1965), Schultze and Sherif (1973), and Schmertmann et al. (1978). These methods are chosen as they are commonly used, represent the chronological development of settlement prediction, and the database used in this work contains most parameters required to calculate settlement by these methods. The parameters needed for each method are summarized in Table 3, which also includes the performance of the traditional methods and the ANN model for the validation set.

Results and Discussion

The impact of the number of hidden nodes on ANN performance is shown in Fig. 1. It can be seen that the number of hidden layer nodes has little impact on the predictive ability of the ANN. Even a network with only one hidden layer node is able to adequately map the underlying relationship. For networks with larger numbers of hidden layer nodes, there is no sign of overtraining, as evidenced by fairly consistent prediction errors. This is to be expected, as crossvalidation is used as the stopping criterion. Fig. 1 shows that the network with five hidden layer nodes has the lowest prediction error. However, the network with two hidden layer nodes is considered optimal, as its prediction error is not far from that of the network with five hidden layer nodes (the error difference is only 0.17 mm) coupled with a smaller number of connection weights.

The effect of the internal parameters controlling the back-propagation algorithm (i.e., momentum term and learning rate) on model performance is shown in Figs. 2 and 3, respectively. It can be seen from Fig. 2 that the performance of the ANN model is relatively insensitive to momentum, particularly in the range 0.01–0.6. The best prediction was obtained with a momentum value of 0.8. Fig. 3 shows that the optimum learning rate was found to be 0.2. At smaller learning rates, prediction errors were higher, probably as a result of the inability of the networks to

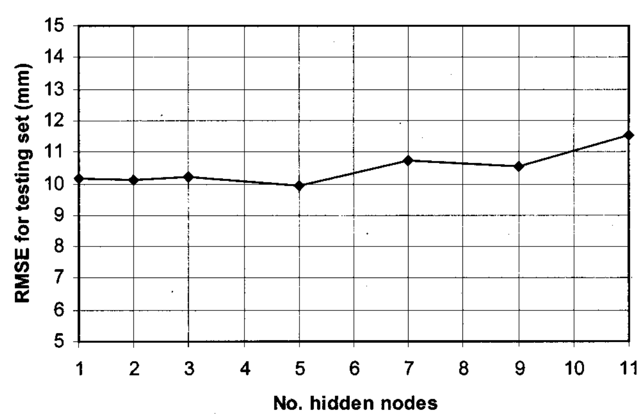

Fig. 1. Performance of artificial neural network models with different hidden layer nodes (learning rate=0.2 and momentum term=0.8)

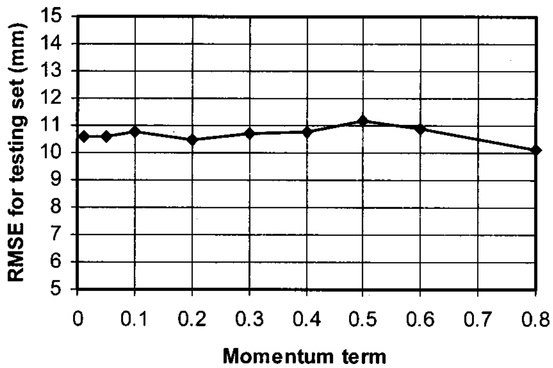

Fig. 2. Effect of various momentum terms on artificial neural network performance (hidden nodes=two and learning rate=0.2)

escape local minima in the error surface due to the small step sizes taken. At larger learning rates, prediction errors increased slightly, possibly as a result of the pseudorandom behavior of the optimization algorithm near the local minima in the error surface due to the large step sizes taken in weight space.

The effect of different random starting positions in weight space on prediction error was negligible for the ANN trained with two hidden layer nodes, a momentum value of 0.8, and a learning rate of 0.2. One possible reason for this is that the error surface in weight space is relatively uncomplicated for the problem under consideration. In addition, as discussed herein, by using a learning rate of 0.2, the model is likely to escape local minima in the error surface during training.

The predictive performance of the optimal neural network model (i.e., two hidden layer nodes, momentum value of 0.8, and learning rate of 0.2) is summarized in Table 4. The results indicate that the ANN model performs well, with an r^2 of 0.819, an RMSE of 11.04 mm, and an MSE of 8.78 mm for the validation set. Table 4 also shows that the results obtained for the model during validation are generally consistent with those obtained during training and testing, indicating that the model is able to generalize within the range of the data used for training.

Comparisons of the results obtained using the ANN and the three traditional methods for the validation set are presented in Fig. 4 and Table 3. Table 3 shows that the ANN method performs better than the traditional methods for all three performance measures considered. As just mentioned, the coefficient of determination, r^2, the RMSE, and MAE obtained using the ANN model are

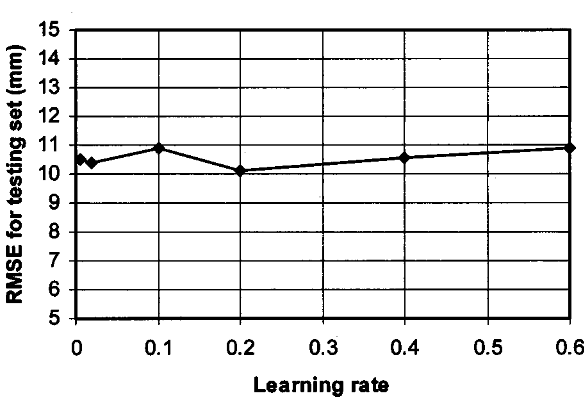

Fig. 3. Effect of various learning rates on artificial neural network performance (hidden nodes=two and momentum term=0.8)

Table 4. Artificial Neural Network Results

Data set	r^2	RMSE (mm)	MAE (mm)
Training set	0.865	10.01	6.87
Testing set	0.863	10.12	6.43
Validation set	0.819	11.04	8.78

0.819, 11.04, and 8.78 mm, respectively. In contrast, these measures range from 0.160 to 0.637, from 23.55 to 25.72 mm, and from 11.81 to 16.59 mm, respectively, when the traditional methods are used. Fig. 4 shows that the ANN model performs reasonably well for the full range of measured settlements considered. In contrast, the traditional methods only appear to work well for small settlements, in the range of 10–20 mm. The method of Schmertmann et al. (1978) tends to overpredict larger settlements, the method of Schultze and Sherif (1973) tends to severely underpredict larger settlements, and the method of Meyerhof (1965) appears to both over- and underpredict larger settlements, although all settlements in excess of 60 mm are generally underpredicted.

The results of the sensitivity analysis are shown in Table 5. It can be seen that N has the most significant effect on the predicted settlement when the network is retrained with different initial weights. However, the relative importance of the remaining input variables changed depending on which initial weights were used (Table 5). Although N was found to be the most important input in all trials, the rank of B varied between two and three, the rank of q between two and five, the rank of L/B between three and five, and the rank of D_f/B between three and five. The sensitivity analyses carried out indicate that as expected, N, B, and q are the most important factors affecting settlement with average relative importance equal to 33.3, 23.2, and 17.7%, respectively. The results also indicate that D_f/B has a moderate impact on settlement with average relative importance equal to 15.7%, while L/B has the smallest impact on settlement with 9.8% average relative importance.

Settlement analysis of shallow foundations on cohesionless soils, as in many situations in geotechnical engineering, is a complex problem that is not well understood. For most mathematical models that attempt to solve this problem, the lack of physical understanding is usually supplemented by either simplifying the problem or incorporating several assumptions into the models. In contrast, as shown in this study, ANNs use the data alone to determine the structure and parameters of the model. In this case, there is no need to simplify the problem or to incorporate any assumptions. Moreover, ANNs can always be updated to obtain better results by presenting new training examples as new data become available.

Despite the good performance of ANNs, they suffer from a number of shortcomings, notably, the lack of theory to help with their development, the fact that success in finding a good solution is usually obtained by trial and error, and their limited ability to explain the way they use the available information to arrive at a solution. Consequently, there is a need to develop some guidelines, which can help in the design process of ANNs. This paper has provided some of these guidelines. There is also a need for more research into methods that provide a comprehensive explanation of how ANNs arrive at a prediction.

Summary and Conclusions

A back-propagation neural network was used to demonstrate the feasibility of ANNs to predict the settlement of shallow founda-

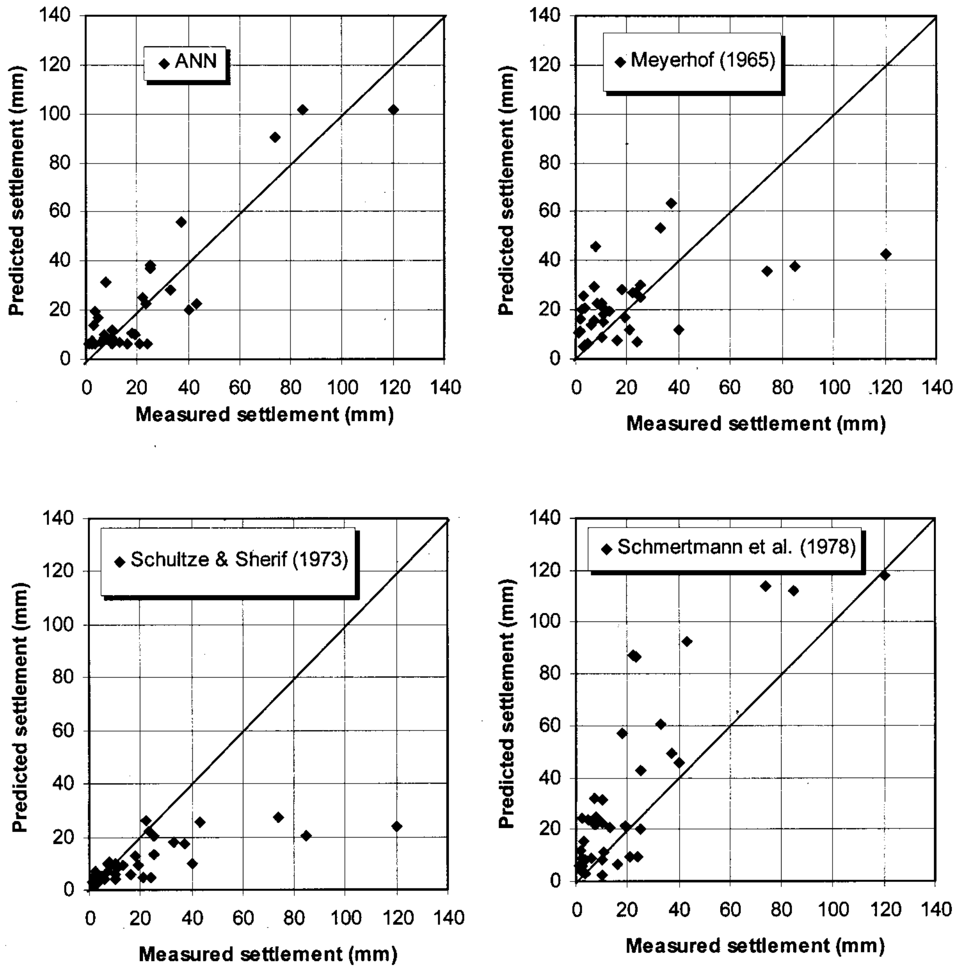

Fig. 4. Measured versus predicted settlement for artificial neural network and traditional methods

tions on cohesionless soils. A database containing 189 case records of actual field measurements for settlement of shallow foundations on cohesionless soils was used for model development and verification. The effect of using various learning rates and momentum terms on the results of the ANN models was investigated. A sensitivity analysis was carried out to study the relative importance of the factors that affect settlement. The results between the predicted and measured settlements obtained by utilizing ANNs were compared with three traditional methods.

The results indicate that back-propagation neural networks have the ability to predict the settlement of shallow foundations on cohesionless soils with an acceptable degree of accuracy ($r^2 = 0.819$, RMSE=11.04 mm, MAE=8.78 mm) for predicted settlements ranging from 0.6 to 121 mm. The predictions obtained using the ANN model were relatively insensitive to the number of hidden layer nodes and the momentum term. The impact of learning rate on model predictions was more pronounced, with step sizes that are too small or too large resulting in reduced model performance. The optimum network geometry was found to be 5-2-1 (i.e., five inputs, two hidden layer nodes, and one output node), the optimum momentum value was found to be 0.8, and the optimum learning rate was found to be 0.2. The results obtained also demonstrate that the ANN method outperforms the traditional methods considered for an independent validation set.

The sensitivity analysis indicates that the SPT blow count, the footing width, and the footing net applied pressure are the most important factors affecting settlement of shallow foundations on cohesionless soils. The sensitivity analysis also shows that the footing embedment ratio and the footing geometry have less impact on settlement.

ANNs have the advantage that once the model is trained, it can be used as an accurate and quick tool for estimating the settlement without a need to perform any manual work such as using tables or charts. The main shortcomings of ANNs are the lack of theory to help with their development and their limited ability to explain the way they use the available information to arrive at a solution. In addition, like all empirical models, the range of applicability of ANNs is constrained by the data used in the model calibration phase and ANNs should thus be recalibrated as new data become available. However, despite the aforementioned limitations, the

Table 5. Sensitivity Analyses of the Relative Importance of Artificial Neural Network Input Variables

	Relative Importance for Input Variables (%)				
Trial no.	B	q	N	L/B	D_f/B
1	22.4	28.6	30.8	3.1	14.8
2	26.3	11.4	33.5	15.6	12.9
3	18.6	16.5	42.7	10.2	11.8
4	25.6	14.1	26.1	10.6	23.3
Average	23.2	17.7	33.3	9.8	15.7

results of this study indicate that ANNs have a number of significant benefits that make them a powerful and practical tool for settlement prediction of shallow foundations on cohesionless soils.

Acknowledgments

The writers would like to thank the anonymous reviewers of this paper for their suggestions and comments. The executable program for the ANN model presented in this paper is provided at www.civeng.adelaide.edu.au/staff/mjaksao1/publications.html.

Appendix. Computational Process of Artificial Neural Network Sensitivity Analysis

The optimum ANN model obtained in this work has five input nodes, two hidden nodes, and one output node with connection weights as follows:

Hidden nodes	(B)	(q)	(N)	(L/B)	(D_f/B)	(S_m)
Hidden 1	0.227 354	0.481 161	0.229 593	−0.017 031	0.067 341	0.725 351
Hidden 2	−2.442 513	−1.114 891	4.239 639	−0.498 853	2.500 301	−2.984 165

The computational process proposed by Garson (1991) is as follows:

1. For each hidden node i, obtain the products P_{ij} (where j represents the column number of the weights just mentioned) by multiplying the absolute value of the hidden-output layer connection weight by the absolute value of the hidden-input layer connection weight of each input variable j. As an example, $P_{11} = 0.227\,354 \times 0.725\,351 = 0.164\,911$.

	(B)	(q)	(N)	(L/B)	(D_f/B)
Hidden 1	0.164 911	0.349 011	0.166 536	0.012 353	0.048 846
Hidden 2	7.288 861	3.327 017	12.65 178	1.488 659	7.461 311

2. For each hidden node, divide P_{ij} by the sum of all input variables to obtain Q_{ij}. As an example, $Q_{11} = 0.164\,911/(0.164\,911 + 0.349\,011 + 0.166\,536 + 0.012\,353 + 0.048\,846) = 0.222\,355$.

	(B)	(q)	(N)	(L/B)	(D_f/B)
Hidden 1	0.222 355	0.470 582	0.224 545	0.016 656	0.065 859
Hidden 2	0.226 238	0.103 267	0.392 698	0.046 206	0.231 591

3. For each input node, sum Q_{ij} to obtain S_j. As an example, $S_1 = 0.222\,355 + 0.226\,238 = 0.448\,593$.

	(B)	(q)	(N)	(L/B)	(D_f/B)
Sum	0.448 593	0.573 849	0.617 243	0.062 863	0.297 451

4. Divide S_j by the sum for all input variables to get the relative importance of all output weights attributed to the given input variable. As an example, the relative importance for input node 1 is equal to $(0.448\,593 \times 100)/(0.448\,593 + 0.573\,849 + 0.617\,243 + 0.062\,863 + 0.297\,451) = 22.4\%$.

	(B)	(q)	(N)	(L/B)	(D_f/B)
Relative importance (%)	22.4	28.6	30.8	3.1	14.8

Notation

The following symbols are used in this paper:

B = footing width;
C_1, C_2 = correction factors for embedment and creep;
D_f = footing embedment depth;
E_z = soil Young's modulus of elasticity;
F = settlement coefficient;
I_z = strain influence factor;
L = footing length;
N = average standard penetration test blow count/300 mm to depth of influence of foundation;
$N_{\text{corrected}}$ = corrected standard penetration test blow count;
q = footing net applied pressure;
r^2 = determination coefficient;
S_c = calculated settlement; and
S_m = measured settlement.

References

Bazaraa, A. R. S. S. (1967). ''Use of the standard penetration test for estimating settlements of shallow foundations on sand,'' PhD thesis, Univ. of Illinois, Urbana, Ill.

Briaud, J. L., and Gibbens, R. (1999). ''Behavior of five large spread footings in sand.'' *J. Geotech. Geoenviron. Eng.*, 125(9), 787–796.

Burbidge, M. C. (1982). ''A case study review of settlements on granular soil.'' MSc. thesis, Imperial College of Science and Technology, Univ. of London, London.

Burland, J. B., and Burbidge, M. C. (1985). ''Settlement of foundations on sand and gravel.'' *Proc., Inst. Civ. Eng.*, Part I, 78(6), 1325–1381.

Caudill, M. (1988). ''Neural networks primer, Part III.'' *AI Expert*, 3(6), 53–59.

Coduto, D. P. (1994). *Foundation design principles and practices*, Prentice-Hall, Englewood Cliffs, N.J.

D'Appolonia, D. J., and D'Appolonia, E. (1970). ''Use of SPT to estimate settlement of footings on sand.'' *Proc., Symposium of Foundation Interbedded Sands*, Division of Applied Geomechanics, Commonwealth Scientific and Industrial Research Organization, Australia and Western Australia of the Australian Geomechanics Society, Perth, 16–22.

Fausett, L. V. (1994). *Fundamentals neural networks: Architecture, algorithms, and applications*, Prentice-Hall, Englewood Cliffs, N.J.

Garson, G. D. (1991). ''Interpreting neural-network connection weights.'' *AI Expert*, 6(7), 47–51.

Goh, A. T. C. (1994). ''Seismic liquefaction potential assessed by neural networks.'' *J. Geotech. Geoenviron. Eng.*, 120(9), 1467–1480.

Hecht-Nielsen, R. (1990). *Neurocomputing*, Addison-Wesly, Reading, Mass.

Holzlohner, U. (1984). ''Settlement of shallow foundations on sand.'' *Jpn. Soc. Soil Mech. Found. Eng.*, 24(4), 58–70.

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

Jeyapalan, J. K., and Boehm, R. (1986). ''Procedures for predicting settlements in sands.'' *Proc., Settlement of Shallow Foundations on Cohesionless Soils: Design and Performance*, Geotechnical Special Publication No. 5, ASCE, Seattle, Wash., 1–22.

Maier, H. R., and Dandy, G. C. (2000). ''Applications of artificial neural networks to forecasting of surface water quality variables: Issues, applications and challenges.'' *Artificial neural networks in hydrology*, R. S. Govindaraju and A. R. Rao, eds., Kluwer, Dordrecht, The Netherlands, 287–309.

Masters, T. (1993). *Practical neural network recipes in C++*, Academic, San Diego.

Maugeri, M., Castelli, F., Massimino, M. R., and Verona, G. (1998). ''Observed and computed settlements of two shallow foundations on sand.'' *J. Geotech. Geoenviron. Eng.,* 124(7), 595–605.

Meyerhof, G. G. (1965). ''Shallow foundations.'' *J. Soil Mech. Found. Div., Am. Soc. Civ. Eng.,* 91(SM2), 21–31.

Moorhouse, D. C. (1972). ''Shallow foundations.'' *Proc., Conf. Performance of Earth and Earth-Supported Structures*, Purdue Univ., Lafayette, Ind., 71–109.

Najjar, Y. M., and Basheer, I. A. (1996). ''A neural network approach for site characterization and uncertainty prediction.'' *Geotechnical Special Publication*, **58**(1), 134–148.

Picornell, M., and Del Monte, E. (1988). ''Prediction of settlement of cohesive granular soils.'' *Proc., Measured Performance of Shallow Found.*, Nashville, Tenn., 55–72.

Poulos, H. G. (1999). ''Common procedures for foundation settlement analysis-Are they adequate?'' *Proc., 8th Australia New Zealand Conf. on Geomechanics*, Hobart, 3–25.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). ''Learning internal representation by error propagation.'' *Parallel distributed processing*, Vol. 1, Chap. 8, D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge, Mass.

Schmertmann, J. H. (1970). ''Static cone to compute static settlement over sand.'' *J. Soil Mech. Found. Div., Am. Soc. Civ. Eng.,* 96(SM3), 7302–1043.

Schmertmann, J. H., Hartman, J. P., and Brown, P. B. (1978). ''Improved strain influence factor diagrams.'' *J. Geotech. Geoenviron. Eng.,* 104(8), 1131–1135.

Schultze, E., and Sherif, G. (1973). ''Prediction of settlements from evaluated settlement observations for sand.'' *Proc., 8th Int. Conf. Soil Mechanics and Foundation Engineering*, 1(3), 225–230.

Smith, M. (1993). *Neural networks for statistical modelling*, Van Nostrand–Reinhold, New York.

Sowers, G. F. (1970). *Introductory soil mechanics and foundations: Geotechnical engineering*, Macmillan, New York.

Stone, M. (1974). ''Cross-validatory choice and assessment of statistical predictions.'' *J. R. Stat. Soc.,* B 36, 111–147.

Tan, C. K., and Duncan, J. M. (1991). ''Settlement of footings on sands-Accuracy and reliability.'' *Proc., Geotechnical Engineering Congress*, Boulde, Colo., 446–455.

Terzaghi, K., and Peck, R. B. (1948). *Soil mechanics in engineering practice*, Wiley, New York.

Tokar, S. A., and Johnson, P. A. (1999). ''Rainfall-runoff modeling using artificial neural networks.'' *J. Hydrologic Eng.,* 4(3), 232–239.

Transportation Research Board (1999). ''Use of artificial neural networks in geomechanical and pavement systems.'' *Transportation Research Circular*, December (E-C012), 1–18.

Twomey, J. M., and Smith, A. E. (1997). ''Validation and verification.'' *Artificial neural networks for civil engineers: Fundamentals and applications*, N. Kartam, I. Flood, and J. H. Garrett, eds., ASCE, New York, 44–64.

Wahls, H. E. (1997). ''Settlement analysis for shallow foundations on sand.'' *Proc., 3rd Int. Geotechnical Engineering Conf.*, Cairo, Egypt, 7–28.