Investigation into the Robustness of Artificial Neural Networks for a Case Study in Civil Engineering

[1]**M. A. Shahin**, [2]**H. R. Maier and** [3]**M. B. Jaksa**

[1]Research Fellow, School of Civil, Mining & Environmental Engineering, University of Wollongong, NSW, Australia, E-Mail: shahin@uow.edu.au
[2]Associate Professor, School of Civil & Environmental Engineering, University of Adelaide, SA, Australia
[3]Senior Lecturer, School of Civil & Environmental Engineering, University of Adelaide, SA, Australia

Keywords: *Robustness; Artificial neural networks; Settlement; Shallow foundations.*

EXTENDED ABSTRACT

Artificial neural networks (ANNs) have been used as a prediction tool in many areas of engineering. In order to test the robustness and generalisation ability of ANN models, the approach that is generally adopted is to test the performance of trained ANNs on an independent validation set. If such performance is adequate, the model is deemed to be robust and able to generalise. However, this is not necessarily the case. In this paper, the robustness of ANN models is investigated for a case study of predicting the settlement of shallow foundations on granular soils. A procedure that tests the robustness of the predictive ability of ANN models is introduced. The results indicate that good performance of ANN models on the data used for model calibration and validation does not guarantee that the models will perform in a robust fashion over a range of data similar to that used in the model calibration phase. The results also indicate that validating ANN models using the procedure provided in this study is essential in order to investigate their robustness.

1 INTRODUCTION

In recent years, artificial neural network (ANN) models have been used extensively for prediction purposes in civil engineering. Details of the structure and operation of ANNs are beyond the scope of this paper and can be found in many publications (Fausett 1994). ANNs are similar to most traditional statistical models in the sense that model parameters (i.e. connection weights) are adjusted in a model calibration phase called "training" so as to minimise the error between the predicted model outputs and the corresponding actual values for a particular data set (i.e. the training set). Therefore, the purpose of ANNs is to non-linearly interpolate (generalise) in high-dimensional space between the data used for calibration. ANNs have been shown to outperform more traditional statistical methods as they are universal function approximators (Hornik *et al.* 1989). However, one of the difficulties in using ANN models is that the potential number of free model parameters (i.e. connection weights) is generally large compared with that used in traditional statistical models and there is, therefore, a danger of overfitting the training data. In other words, if the number of degrees of freedom of the model is large compared with the number of data points used for training, the model might no longer fit the general trend, as desired, but might learn the idiosyncrasies of the particular data points used for training. In general, one of two methods is used to overcome this problem. The first is to restrict the ratio of the number of connection weights to the number of data points in the training set, and several rules-of-thumb have been developed as a guide. For example, Rogers and Dowla (1994) suggest that the number of weights should not exceed the number of training samples and Masters (1993) suggests that the ratio of the number of training samples to the number of connection weights should be 2 to 1. The second approach to avoiding overfitting is to use the cross-validation method (Stone 1974), in which training is stopped early once the error associated with an independent test set starts to increase. Generally, cross-validation is considered to be the most effective method to ensure overfitting does not occur (Smith 1993).

When cross-validation is used as the stopping criterion, three data sets are needed; a training set, a testing set and a validation set. The training set is used to adjust the connection weights, the testing set is used to decide when to stop training to avoid overfitting and the validation set is used to assess the generalisation ability of the model within the range of the data used for training. If the trained model performs well on the validation set, the model is generally considered robust and ready for use as a predictive tool. However, this is not necessarily the case. The objectives of this paper are: (i) to demonstrate that good performance on separate training, testing and validation sets does not guarantee that a model will perform in a robust fashion over the range of data used for training; and (ii) to introduce a procedure of model validation that tests the robustness of the predictive ability of ANN models. The abovementioned objectives are investigated in this work through a case study of settlement prediction of shallow foundations on granular soils, as will be described below.

2 CASE STUDY

In order to meet the objectives set out above, feed-forward ANNs trained with the back-propagation algorithm (Rumelhart *et al.* 1986) are applied to a case study of settlement prediction of shallow foundations on granular soils. Settlement of shallow foundations on granular soils occurs with load application during, or immediately after, the construction period of a structure and is primarily due to the reorientation and distortion of soil grains. Settlement of shallow foundations on granular soils usually causes relatively rapid deformations of superstructures, which results in an inability to remedy damage and to avoid further deformation. As a consequence, settlement is a major concern and is an essential criterion in the design process of shallow foundations on granular soils.

3 DEVELOPMENT OF ANN MODELS

The PC-based software package NeuralWorks *Predict* Release 2.1 (NeuralWare 1997) is used to simulate artificial neural network operation. The first step in the development of ANN models is the determination of appropriate model inputs and outputs. It is generally accepted that five parameters have the most significant impact on the settlement of shallow foundations on granular soils (Burland and Burbidge 1985) and are thus used as the ANN model inputs. These include the footing width, B; footing net applied pressure, q; soil compressibility, which can be represented by the average blow count, N, obtained using the standard penetration test (SPT) over the depth of influence of the foundation, footing geometry, L/B; and footing embedment ratio, D_f/B. The single model output is foundation settlement, S_m.

The next step in the development of ANN models is dividing the available data into their subsets. The data used to calibrate and validate the neural network models are obtained from the literature and comprise a total of 189 individual cases (Shahin *et al.* 2002) that include field measurements of settlement of shallow foundations, as well as the corresponding information regarding footings and soils. As cross-validation (Stone 1974) is used as the stopping criterion in this study, the data are randomly divided into three sets: training, testing and validation. When dividing the data into their subsets, it is essential to check that the data used for training, testing and validation represent the same statistical population (Masters 1993). This is done by trying several random combinations of training, testing and validation sets until three statistically consistent data sets are obtained. The statistical parameters considered are the mean, standard deviation, minimum, maximum and range. 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. Once the available data have been divided into their subsets, the input and output variables are pre-processed by scaling them to eliminate their dimension and to ensure that all variables receive equal attention during training. Scaling has to be commensurate with the limits of the transfer functions used in the hidden and output layers (e.g. –1.0 to 1.0 for the tanh transfer function and 0.0 to 1.0 for the sigmoid transfer function).

One of the most important and difficult tasks in the development of ANN models is determining the model geometry. In NeuralWorks *Predict*, the optimal network geometry (i.e. the number and connectivity of the hidden layer nodes) is found with the aid of *Cascade* learning (Fahlman and Lebiere 1990). Cascade learning is an automatic constructive algorithm in which hidden layer nodes are added as training progresses until there is no further improvement in model performance. This is designed to result in the smallest network that can adequately map the design input-output relationship. Cascade learning can be characterised by the following steps (NeuralWare 1997):

- Initially, the network is trained without hidden nodes and with direct connection between the input and output layers;
- Hidden nodes are added randomly one or a few at a time;
- New hidden nodes have connections from both the input layer and previously established hidden nodes; and
- Construction is stopped when performance on the testing set shows no further improvement.

The process of optimising the connection weights is applied using the default parameters of the software package. Use of the default parameters is considered reasonable since the focus of this study is on the evaluation of the robustness of ANN models rather than studying the impact of varying the networks parameters. Details of the default parameters are discussed in NeuralWare (1997) and are as follows:

- Learning rule: Adaptive gradient learning rule;
- Learning rate: 100 for the hidden layer and 0.01 for the output layer;
- Transfer function for the hidden layer: *Tanh* transfer function; and
- Transfer function for output layer: *Sigmoid* transfer function.

Using the above method, two neural networks, each with 2 hidden layer nodes, are found to perform best. The structure of the models developed is shown in Figure 1 and their performance is summarised in Table 1, which includes three different measures of ANN model performance; the coefficient of correlation, *r*; the root mean square error, RMSE; and the mean absolute error, MAE. The coefficient of correlation, *r*, determines the goodness-of-fit between the predicted and observed data. The RMSE has the advantage that large errors receive much greater attention than small errors. In contrast, MAE eliminates the emphases given to large errors. It can be seen from Table 1 that the two models have similar performance and their predictive ability with regard to the validation set is generally consistent with those of the training and validation sets, indicating that both models are able to generalise within the range of the data used for training. These results indicate that the two models can be used as practical tools for predictive purposes and suggest that Model 2 might slightly outperform Model 1. It should be noted that the two models are developed using the same model parameters, except that they are optimised with different sets of random starting weights.

In order to confirm the robustness of the generalisation ability of the models obtained over the range of the data used for training, an additional validation approach is proposed. The approach consists of carrying out a sensitivity analysis as part of which the response of the ANN model output to changes in its inputs is investigated. Similar approach for carrying out parametric studies to evaluate the effects of various ANN model inputs on the corresponding outputs was used by Goh (1995). All input variables, except one, are fixed to the mean values used for training and a set of synthetic data (between the minimum and maximum values used for model training), are generated for the input that is not set to a fixed value. The synthetic data are generated by increasing their values in increments equal to 5% of the total range between the minimum and maximum values. The response of the model is then examined. This process is repeated using another input variable and so on until the model response is tested for all input variables. The robustness of the model can be determined by examining how well the predicted settlements are in agreement with the known underlying physical processes over the range of inputs examined. The above approach is deemed to complement the approach that is generally used in the literature for model robustness and generalisation (i.e. testing the performance of trained model on an independent validation set).

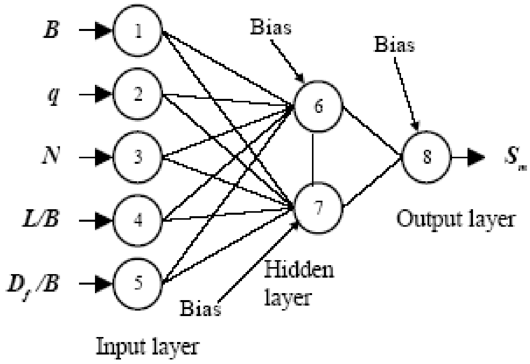

Figure 1. Structure of the ANN models

Table 1. Performance of the ANN models

Model No.	Data set	*r*	RMSE (mm)	MAE (mm)
1	Training	0.92	10.8	7.4
	Testing	0.94	8.4	5.8
	Validation	0.88	12.9	9.8
2	Training	0.94	9.1	6.3
	Testing	0.94	9.1	6.8
	Validation	0.89	11.8	9.6

4 RESULTS AND DISCUSSION

The sensitivity analysis approach proposed above for testing ANN model robustness is applied to the two models developed in this study (i.e. Models 1 and 2) and the results are shown in Figure 2. It can be seen that the results obtained for Model 1 are in agreement with what one would expect based on the known physical behaviour of settlement of shallow foundations on granular soils. For example, in Figures 2(a), (b) and (d), there is an increase

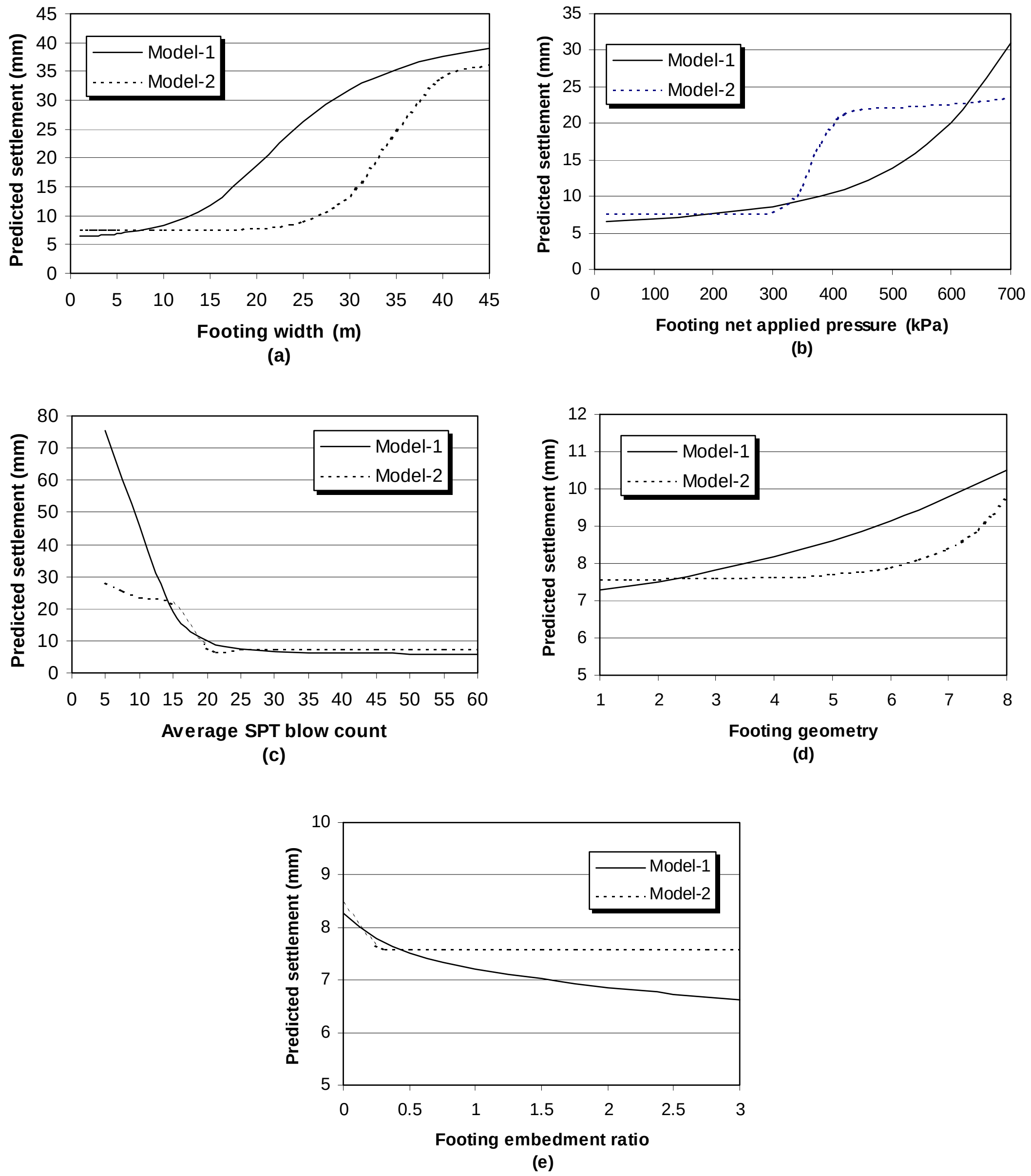

Figure 2. Results of the sensitivity analysis to test the robustness of the ANN models

in the predicted settlement, in a relatively consistent and smooth fashion, as footing width, footing net applied pressure and footing geometry, respectively, increase. On the other hand, in Figures 2(c) and (e), the predicted settlement decreases, also in a consistent and smooth fashion, as the average SPT blow count and footing embedment ratio, respectively, increase. In contrast, it can be seen from Figure 2 that the curves obtained for Model 2 have an unexpected shape that is difficult to justify from a physical understanding of footing settlement. For example, there are abrupt changes in predicted settlement in some instances and no change in predicted settlement for a range of inputs in others. However, as both models gave similar performance on the training, testing and validation sets (Table 1), one would expect the models to behave in a similar manner when presented with a range of inputs, as was done during the sensitivity analysis. The above results indicate that there appears to be some kind of overfitting for Model 2, however, as mentioned previously, cross-validation was adopted in this work, which is considered to be the most valuable tool to ensure overfitting does not occur (Smith 1993). In addition, an independent validation set was used to test the predictive ability of Model 2 and the model was found to perform well (see Table 1). It appears that the actual values of the connection weights are the only possible reason for the different behaviour exhibited by Models 1 and 2 over the range of the data used during the sensitivity analysis, as this is the only difference in model parameters between the two models, as mentioned previously. Consequently, the actual connection weights (including biases) of the two models are examined as part of this study and are shown in Figure 3. Examining connection weights for interpreting ANN model behaviour was proposed by Garson (1991). It can be seen that the values of the weights obtained for Model 1 are more consistent than those of Model 2. Some values of the weights obtained for Model 2 are significantly larger than the others, which can often indicate a problem with the model (Bailey and Thompson 1990), and result in erratic behavior. The large values of the weights of Model 2 generally cause the node activations to be large, and as a result, the nodal outputs can become trapped in the flat spots at the extreme values of the transfer functions used in the hidden and output layers of Model 2.

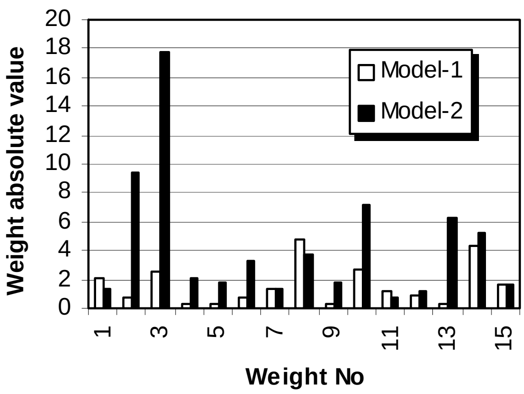

Figure 3. Bar charts of the weights obtained for the ANN models

5 CONCLUSIONS

Based on the results of this study, it is concluded that good performance of ANN models on training, testing and validation sets does not guarantee the robustness of the predictive ability of ANN models over a range of data similar to that used for model training. It is recommended that the following procedure for testing ANN model robustness be used routinely as part of ANN model development.

1. The performance of an ANN model should be tested on an independent validation set, the statistical properties of which should be consistent with those of the training and validation sets (this is the method that is commonly used in the literature for testing ANN model robustness);
2. The connection weights obtained should be checked for any inconsistency (e.g. whether some weights are large compared with other weights in the network); and
3. A sensitivity analysis, similar to the one proposed in this study, should be carried out to ensure that the model can be used for predictive purposes with confidence.

6 REFERENCES

Bailey, D. L., and Thompson, D. M. (1990), Developing neural-network applications, *AI Expert*, 5(9), 34-41.

Burland, J. B., and Burbidge, M. C. (1985), Settlement of foundations on sand and gravel, *Proc. Institution of Civil Engineers*, London, 1325-1381.

Fahlman, S. E., and Lebiere, C. (1990), The cascade-correlation learning architecture, Advances in Neural Information Processing Systems 2, D. S. Touretzky, ed., Morgan Kaufmann, San Mateo, California, 524-532.

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

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

Goh, A. T. C. (1995), Back-propagation neural networks for modelling complex systems, *Artificial Intelligence in Engineering*, 9, 143-151.

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

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

NeuralWare. (1997), *NeuralWorks Predict Release 2.1*, NeuralWare Inc., Pittsburgh.

Rogers, L. L., and Dowla, F. U. (1994), Optimization of groundwater remediation using artificial neural networks with parallel solute transport modelling, *Water Resources Research*, 30(2), 457-481.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986), Learning internal representation by error propagation, Parallel Distributed Processing, D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge.

Shahin, M. A., Maier, H. R., and Jaksa, M. B. (2002), Predicting settlement of shallow foundations using neural networks, *J. Geotech. & Geoenv. Eng.*, ASCE, 128(9), 785-793.

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

Stone, M. (1974), Cross-validatory choice and assessment of statistical predictions, *J. Royal Statistical Society*, B 36, 111-147.