

**REPLY TO DISCUSSION OF “INTELLIGENT COMPUTING FOR  
MODELING AXIAL CAPACITY OF PILE FOUNDATIONS”**

**Mohamed A. Shahin**

BSc, MSc, PhD, MIEAust, MASCE

Senior Lecturer, Department of Civil Engineering,  
Curtin University of Technology, WA 6845, Australia

Phone: +61-8-9266 1822; Fax: +61-8-9266 2681

E-mail: [m.shahin@curtin.edu.au](mailto:m.shahin@curtin.edu.au)

**Canadian Geotechnical Journal**

**Reply to Discussion of “Intelligent Computing for Modeling Axial Capacity of Pile Foundations” By: Mohamed A. Shahin, Canadian Geotechnical Journal, Vol. 47, 2010, 230–243.**

Mohamed A. Shahin

Department of Civil Engineering, Curtin University of Technology, Perth, WA 6845, Australia

(e-mail: [m.shahin@curtin.edu.au](mailto:m.shahin@curtin.edu.au)).

The author welcomes and appreciates the interest and comments of the discussers. The issues raised by the discussers are responded to as follows. The discussers recommend scaling the inputs and outputs data between  $[-1, 1]$  rather than  $[0, 1]$  because the transfer function used in the hidden layer was *tanh*, which is bounded by  $[-1, 1]$ . It should be noted that the *sigmoid* transfer function was used in the output layer and thus, the outputs must be scaled between  $[0, 1]$  to commensurate with the limits of the sigmoid transfer function. Using *tanh* transfer function in the output layer is not accepted as it does not comply with the underlying physical meaning of the bearing capacity problem. To explain this, consider the minimum value of ultimate bearing capacity of 290 kN and the maximum value of 4500 kN that used in the driven piles model. If these two values are to be scaled between  $[-1, 1]$ , the minimum ultimate bearing capacity will be equal to  $-1$  and the maximum ultimate bearing capacity will be equal to 1. This means that, during ANN model training, the impact of the minimum ultimate bearing capacity will be as if it is exactly opposite to the impact of the maximum ultimate bearing capacity, which is not true from the physical point of view. In the hidden layer, however, *tanh* transfer function was used as it was found to give the best solution. Scaling the input data is not necessary but always recommended to allow the connection weights to have the same order of magnitude (Jang and Sun 2002; Masters

1993), and it is conventional practice to scale the inputs data within the same range used for scaling the outputs data (Teh et al. 1997).

The discussers claim that, based on their experience, it is futile exercise to develop ANN models after proper data division to obtain statistically consistent training and validation sets so that model generalization can be improved. It should be noted that ANNs cannot extrapolate beyond the range of the data used for model calibration. Consequently, in order to develop the best ANN model, given the available data, the calibration data should contain all representative patterns that are present in the available data. For example, if the available data contain cases of extreme values that are excluded from the calibration set, the model cannot be expected to perform well, as the validation data will test the model's extrapolation ability, and not its interpolation ability. In addition, by choosing calibration and validation data sets randomly, without any knowledge of which types of patterns have been included in each set, the quality of the developed model, and hence the model performance on the validation data, will have a large random component associated with it. Consequently, the statistical properties (e.g. mean and standard deviation) of the various data subsets (e.g., training and validation) need to be similar to ensure that each subset represents the same statistical population (Masters 1993). In fact, studies have shown that the performance of ANNs can be "tailored" to be either "good" or "bad" if the statistical properties of the calibration and validation data sets are not considered (Maier and Dandy 1996; Tokar and Johnson 1999). It follows that if all of the patterns that are contained in the available data should be contained in the calibration set, the toughest evaluation of the generalization ability of the model is when all of the patterns (and not just a subset) are contained in the validation data. One could argue that if the statistical properties of the calibration and validation data sets are not taken into account, the quality of the developed model is "pot luck" and it is not possible to assess rigorously the

generalization ability of the model (Maier and Dandy 2000). This issue has been experimentally studied and discussed in detail by Bowden et al. (2002) and Shahin et al. (2004).

The author agrees with the discussers that the use of the coefficient of correlation,  $r$ , as one of the criteria considered for ranking might be misleading and this was clearly illustrated by the discussers. On the other hand, the discussers recommend the use of the coefficient of efficiency,  $E$ , proposed by Nash and Sutcliffe (1970) as a better method for ranking. It should be noted that  $r$  was not the only error measure used for ranking in the original paper and six other measures were also utilized. These include  $Q_{\text{fit}}/Q_u$ ,  $\mu(Q_{\text{fit}}/Q_u)$ ,  $\sigma(Q_{\text{fit}}/Q_u)$ ,  $P_{50}$ , and  $\pm 20\%$  accuracy of the histogram and lognormal distributions of the ratio  $Q_p/Q_u$ , where:  $Q_{\text{fit}}$  = pile capacity of the best-fit line of predicted versus measured pile capacities;  $Q_u$  = actual measured pile capacity;  $Q_p$  = ANN predicted pile capacity;  $P_{50}$  = cumulative probability at 50%;  $\mu$  = mean; and  $\sigma$  = standard deviation. The author has examined the performance of the developed ANN models and CPT-based methods used for comparison by applying the discussers' recommended  $E$  measure as well as  $r$  and the results are given in Table 1. It can be seen that the effects of  $r$  and  $E$  on the ranking performance of the different pile capacity prediction methods are comparable.

The discussers suggest using the maximum absolute error (MAE) or normalized mean biased error (NMBE) rather than the normalized mean squared error (NMSE) to compare the ANN models for model architecture and optimization. The author would like to mention that MSE is usually preferable for ANN model optimization for the following reasons (Hecht-Nielsen 1990; Masters 1993): (i) large errors receive much greater attention than small errors; and (ii) it lies close to the heart of the normal distribution in which, if the errors can be assumed to be normally distributed, minimizing the MSE is optimal.

The author agrees with the discussers that it is important to consider the cumulative probability of  $Q_p/Q_u$  at 90% as well as at 50% for ranking the different pile capacity prediction

methods, and not only the cumulative probability at 50%. The author would like to mention that the cumulative probability at 90% was meant to be included in the original paper but unfortunately was accidentally overlooked. Table 2 shows the performance of ANN models against available CPT-based methods after considering the cumulative probability of  $Q_p/Q_u$  at 90%. It can be seen that the ANN driven piles model still has the lowest rank index, RI, and thus ranks first and performs the best among all driven pile methods used for comparison. It can also be seen that over the four CPT-based methods used, Eslami and Fellenius (1997) performs second, followed by the LCPC (Bustamante and Gianceselli 1982) and the European method (de Ruiter and Beringen 1979). On the other hand, for drilled shafts, the results show that the ANN model and Alsamman's method (Alsamman 1995) perform best with similar values of RI; however, as shown in Fig. 8 of the original paper, the ANN model outperforms Alsamman's method as it gives better predictions at high values of pile capacity. The results also demonstrate that over the four CPT-based methods used for capacity of drilled shafts, Schmertmann (1978) performs third followed by the LCPC (Bustamante and Gianceselli 1982).

The discussers pointed out that the weights and biases presented in Table 3 of the original paper were reported with up to 4 or 5 decimals, although in the final models (i.e. Equations 5 to 13 in the original paper) only up to three decimals were presented. The author would like to mention that the final models were originally derived with numbers that have up to 4 or 5 decimals and were then simplified by mathematical truncation and rounding to up to 3 decimals. The impact of this process on the ANN predicted values, was marginal and can be neglected.

The author agrees with the discussers that in order to investigate the relative importance (significance) of ANN model inputs, the methods that should be used include, for example, the Garson's algorithm (Garson 1991), the connection weight approach proposed by Olden et al. (2004) or the step-by-step method suggested by Liong et al. (2000). However, the sensitivity analysis

carried out in the original paper was used as an additional validation tool to investigate the generalization ability (robustness) of the developed ANN models over the range of the data used for model training. The generalization ability was determined by examining how well the models' predictions are in agreement with the known underlying physical processes of the problem in hand (i.e., the problem of bearing capacity of pile foundations). This is because it was found by Shahin et al. (2005) and Kingston et al. (2005) that good performance of an ANN model in the training and validation sets does not guarantee that the model will perform well over a range of data similar to that used for model training.

The discussers suggest developing two separate ANN driven pile models, one for steel and another for concrete, rather than the single developed model, which adopts the input numeric values "1" for steel and "2" for concrete, to represent the material type. It should be noted that, as mentioned in the original paper, ANNs usually work well when they deal with binary or numeric variables. Consequently, the pile material was translated from the text format (i.e., steel or concrete) into the arbitrary values of "1" for steel and "2" for concrete. This is a valid process that was used successfully in many applications in geotechnical engineering (e.g., Goh 1994; Shahin and Jaksa 2005). The author, however, agrees with the discussers that developing two separate models, one for steel and another for concrete, might have more physical meaning but this was not possible as the available database did not include sufficient data to apply such process.

In regard to the final comment raised by the discussers that the performance of developed ANN models is less than that of other CPT-methods at values of pile capacity  $< 1000$  kN for driven piles and  $< 2000$  kN for drilled shafts, as can be seen from Figs. 7 and 8 of the original paper. The author would like to emphasize the fact that ANNs are a form of statistical methods and, therefore, are unable to extrapolate beyond the range of the data used for model calibration (Flood and Kartam 1994; Ripley 1993). Consequently, the purpose of ANNs is to non-linearly interpolate in high-

dimensional space between the data used for model calibration. In order to develop the best ANN model, the calibration data should contain all representative patterns that are present in the available database, and the toughest evaluation of the generalization ability of the model is to achieve good overall performance in all patterns that represent the practical range and not just good performance in a certain subset.

## References

- Alsamman, O.M. 1995. The use of CPT for calculating axial capacity of drilled shafts. PhD Thesis, University of Illinois-Champaign, Urbana, Illinois.
- Bowden, G.J., Maier, H.R., and Dandy, G.C. 2002. Optimal division of data for neural network models in water resources applications. *Water Resources Research*, **38**(2): 2.1-2.11.
- Bustamante, M., and Gianceselli, L. 1982. Pile bearing capacity prediction by means of static penetrometer CPT. *In Proceedings of the 2nd European Symposium on Penetration Testing*. Amsterdam, Vol.2, pp. 493-500.
- de Ruiter, J., and Beringen, F.L. 1979. Pile foundation for large North Sea structures. *Marine Geotechnology*, **3**(3): 267-314.
- Eslami, A., and Fellenius, B.H. 1997. Pile capacity by direct CPT and CPTu methods applied to 102 case histories. *Canadian Geotechnical Journal*, **34**(6): 886-904.
- Flood, I., and Kartam, N. 1994. Neural networks in civil engineering I: Principles and understanding. *Journal of Computing in Civil Engineering*, **8**(2): 131-148.
- 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 network. *Journal of Geotechnical & Geoenvironmental Engineering*, **120**(9): 1467-1480.
- Hecht-Nielsen, R. 1990. *Neurocomputing*. Addison-Wesely Publishing Company, Reading, MA.

- Jang, J.-S., and Sun, C.T. 2002. Neurofuzzy and soft computing. Prentice-Hall, New Delhi.
- Kingston, G.B., Maier, H.R., and Lambert, M.F. 2005. Calibration and validation of neural networks to ensure physically plausible hydrological modeling. *Journal of Hydrology*, **314**(2005): 158-176.
- Liong, S.Y., Lim, W.H., and Paudyal, G.N. 2000. River stage forecasting in Bangladesh: neural network approach. *Journal of Computing in Civil Engineering*, **14**(1): 1-8.
- Maier, H.R., and Dandy, G.C. 1996. The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research*, **32**(4): 1013-1022.
- Maier, H.R., and Dandy, G.C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environmental Modeling & Software*, **15**(2000): 101-124.
- Masters, T. 1993. Practical neural network recipes in C++. Academic Press, San Diego, California.
- Nash, J.E., and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models. *Journal of Hydrology*, **10**(3): 282-290.
- Olden, J.D., Joy, M.K., and Death, R.G. 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modeling*, **178**(3-4): 389-397.
- Ripley, B.D. 1993. Statistical aspects of neural networks. *In Networks and Chaos: Statistical and Probabilistic Aspects*. Chapman & Hall, London.
- Schmertmann, J.H. 1978. Guidelines for cone penetration test, performance and design, Report No. FHWA-TS-78-209, U. S. Department of Transportation, Washington, D. C.
- Shahin, M.A., and Jaksa, M.B. 2005. Neural network prediction of pullout capacity of marquee ground anchors. *Computers and Geotechnics*, **32**(3): 153-163.

- Shahin, M.A., Maier, H.R., and Jaksa, M.B. 2004. Data division for developing neural networks applied to geotechnical engineering. *Journal of Computing in Civil Engineering*, **18**(2): 105-114.
- Shahin, M.A., Maier, H.R., and Jaksa, M.B. 2005. Investigation into the robustness of artificial neural network models for a case study in civil engineering. *In Proceedings of the International Congress on Modeling and Simulation, MODSIM 2005*. Melbourne (Australia), pp. 79-83.
- Teh, C.I., Wong, K.S., Goh, A.T.C., and Jaritngam, S. 1997. Prediction of pile capacity using neural networks. *Journal of Computing in Civil Engineering*, **11**(2): 129-138.
- Tokar, S.A., and Johnson, P.A. 1999. Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering*, **4**(3): 232-239.

**Table 1.** Performance of ANN models against available CPT-based methods using the coefficient of correlation,  $r$ , and coefficient of efficiency,  $E$ .

Method	Ranking measure	
	<i>Coefficient of Correlation, <math>r</math></i>	<i>Coefficient of Efficiency, <math>E</math></i>
<b>Driven piles</b>		
ANN	0.97	0.91
European method (deRuiter and Beringen, 1979)	0.85	0.67
LCPC (Bustamante and Gianceselli, 1982)	0.83	0.65
Eslami & Fellenius (1997)	0.95	0.83
<b>Drilled shafts</b>		
ANN	0.97	0.94
Schmertmann (1978)	0.83	0.61
LCPC (Bustamante and Gianceselli, 1982)	0.93	0.73
Alsamman (1995)	0.95	0.90

**Table 2.** Performance of ANN models against available CPT-based methods after considering the cumulative probability at 90%.

Method	Best fit calculations			Arithmetic calculations			Cumulative probability			Accuracy $\pm 20\%$			Overall rank
	$Q_{fit}/Q_u$	$r$	$R_1$	$\mu(Q_p/Q_u)$	$\sigma(Q_p/Q_u)$	$R_2$	$P_{50}$	$P_{90}$	$R_3$	Histogram	Lognormal	$R_4$	RI
<b>Driven piles</b>													
ANN (this study)	0.98	0.97	1	1.05	0.20	1	1.02	1.34	1	75	70	1	4
European method*	0.90	0.85	3	0.93	0.37	3	0.86	1.51	4	38	42	4	14
LCPC (1982) <sup>†</sup>	0.89	0.83	4	0.96	0.35	2	0.88	1.54	3	40	43	3	12
Eslami & Fellenius (1997)	1.10	0.95	2	1.13	0.23	4	1.09	1.47	2	59	61	2	10
<b>Drilled shafts</b>													
ANN (this study)	0.97	0.97	1	1.06	0.40	2	0.98	1.56	1	51	45	2	6
Schmertmann (1978)	0.91	0.83	3	0.93	0.40	3	0.86	1.36	3	49	41	3	12
LCPC (1982) <sup>†</sup>	1.16	0.93	4	1.19	0.51	4	1.14	1.63	4	47	39	4	16
Alsamman (1995)	0.92	0.95	2	1.04	0.38	1	0.96	1.38	2	60	56	1	6

**Note:**  $Q_{fit}$ , pile capacity of the best-fit line of predicted versus measured pile capacity;  $Q_p$ , ANN predicted pile capacity;  $P_{50}$ , cumulative probability at 50%;  $P_{90}$ , cumulative probability at 90%;  $r$ , correlation coefficient;  $\mu$ , mean;  $\sigma$ , standard deviation.

\*de Ruiter and Beringen (1979).

<sup>†</sup>Bustamante and Gianceselli (1982).