Geomechanics and Engineering, Vol. 3, No. 4 (2011) 285-290
DOI: http://dx.doi.org/10.12989/gae.2011.3.4.285

Multivariate adaptive regression spline applied to friction capacity of driven piles in clay

Pijush Samui*

Centre for Disaster Mitigation and Management, VIT University, Vellore-632014, India

(Received September 12, 2011, Revised November 15, 2011, Accepted November 22, 2011)

Abstract: This article employs Multivariate Adaptive Regression Spline (MARS) for determination of friction capacity of driven piles in clay. MARS is non-parametric adaptive regression procedure. Pile length, pile diameter, effective vertical stress, and undrained shear strength are considered as input of MARS and the output of MARS is friction capacity. The developed MARS gives an equation for determination of f_s of driven piles in clay. The results of the developed MARS have been compared with the Artificial Neural Network. This study shows that the developed MARS is a robust model for prediction of f_s of driven piles in clay.

Keywords: multivariate adaptive regression spline; driven pile; clay; friction capacity; artificial neural network.

1. Introduction

The determination of friction capacity (f_s) of driven piles in clay is an imperative task in geotechnical engineering. Geotechnical engineers use various methods such as a method (Tomlinson 1971, McClelland 1972), effective stress method (Chandler 1968, Burland 1973, Meyerhof 1976, Parry and Swain 1977a and 1977b), etc. The main limitation of the effective stress method is the lack of knowledge of the effective stress state around the pile (Randolph *et al*. 1979). Therefore, the available methods are not so reliable. Artificial Neural Network (ANN) model has been successfully used for determination of f_s of driven piles in clay (Goh 1995). ANN has been successfully used to solve many problems in geotechnical engineering (Mayoraz and Vulliet 2002, Akin and Karpuz 2008, Samui and Sitharam 2010, Zaman *et al*. 2010, Zhang *et al*. 2011, Miranda *et al.* 2011). However, ANN has some limitations such as black box approach, arriving at local minima, low generalization capability, overfitting, etc (Park and Rilett 1999, Kecman 2001).

This article adopts an alternative method based on Multivariate Adaptive Regression Spline (MARS) for prediction of f_s of driven piles in clay. MARS is a non-parametric modelling method and can capture the complicated relationship in high dimensional data (Friedman 1991). It can be considered as a generalisation of classification and regression trees (CART) (Hastie *et al*. 2003). It has been used successfully for solving different problems in engineering (Lewis and Stevens 1991,

*Associate Professor, E-mail: pijush.phd@gmail.com

Ekman and Kubin 1999, Yang *et al*. 2004, Crino and Brown 2007, Deconinck *et al*. 2007, Attoh-Okine *et al*. 2009, Andrés *et al*. 2011, Vidoli 2011). This article uses the database collected by Goh (1995). This study has the following aims

- To examine the capability of MARS for determination of f_s of driven piles in clay
- To develop an equation for prediction of f_s of driven piles in clay based on the MARS
- To carry out a comparative study between the MRAS and ANN model developed by Goh (1995)

2. Details of MARS

The main aim of MARS is to predict the values of the output variable (y), from a set of input variables (x). The MARS uses the following equation for prediction of y (Friedman and Roosen 1995)

$$y = f(x) = c_0 + \sum_{m=1}^{M} c_m B_m(x) \tag{1}$$

where c_0 is a constant, $B_m(x)$ is the mth basis function, and c_m is the coefficient of the mth basis function. The four input variables used for the MARS model in this study are pile length (L), pile diameter (D), effective vertical stress (σ'_v), and undrained shear strength (S_u). The output of the MARS model is f_s. So, in this study, $x = \lfloor L, D, \sigma_v', S_u \rfloor$ and $y = f_s$.

MARS uses the following two-sided truncated power functions as spline basis functions (Sekulic and Kowalski 1992)

$$[-(x-t)]_+^q = \begin{cases} (t-x)^q, & \text{if } x<t \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$[+(x-t)]_+^q = \begin{cases} (t-x)^q, & \text{if } x \geq t \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where q is the power and t is knot.

The MARS uses the following two steps:

Building the global MARS-model: Basis functions are added to define Eq. (1). Performance is improved by adding more basis functions. The MARS usually shows overfitting due to many basis functions.

Pruning: To prevent overfitting, basis functions are deleted from Eq. (1). Basis functions are deleted based on Generalised Cross-Validation (GCV). The value of GCV has been calculated from the following expression

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{N} [y_i - f(x_i)]^2}{\left[1 - \frac{C(B)}{N}\right]^2} \tag{4}$$

where N is the number of data and $C(B)$ is a complexity penalty that increases with the number of

BF in the model and which is defined as

$$C(B) = (B+1) + dB \tag{5}$$

where d is a penalty for each basis function (*BF*) included into the model. The details about d are given by Friedman (1991).

Analysis of Variance (ANOVA) decomposition of the MARS model is given by the following expression (Friedman 1991)

$$f(x) = \beta_0 + \sum_{B=1} f_i(x_i) + \sum_{B=2} f_{ij}(x_i, x_j) + \sum_{B=3} f_{ijk}(x_i, x_j, x_k) + \dots \tag{6}$$

$\sum_{B=1} f_i x_i$ is over all basis functions that involve only a single variable; $\sum_{B=2} f_{ij}(x_i, x_j)$ is over all basis functions that involve exactly two variables; and $\sum_{B=3} f_{ij}(x_i, x_j, x_k)$ represents (if present) the contributions from three variable interactions and so on.

This article uses the above MARS for prediction of f_s of driven piles in clay. In carrying out the formulation, the data has been divided into two sub-sets; such as:

(a) A training dataset: This is required to construct the model. In this study, 45 out of the 65 cases of pile load test are considered for training the dataset.

(b) A testing dataset: This is required to estimate the model performance. In this study, the remaining 20 data is considered as testing dataset.

The data is normalized between 0 and 1. The program of MARS has been developed by using MATLAB.

3. Results and discussion

The following section presents the results from the MARS analysis for prediction of f_s of driven piles in clay. Ten basis functions have been used to define Eq. (1). In pruning step, six basis functions have been deleted. Therefore, the final MARS model contains four basis functions. The final expression of MARS is given below (by putting $y = f_s$, $M = 4$ and $c_0 = 0.096$ in Eq. (1))

$$f_s = 0.096 + \sum_{m=1}^{4} c_m B_m(x) \tag{7}$$

Table 1 shows the value of c_m and $B_m(x)$. The performance of training and testing dataset has been determined by Eq. (7). The developed MARS uses coefficient of correlation (R) to asses the performance of training and testing dataset. For a good model, the value of R should be close to one. Fig. 1 depicts the performance of training and testing dataset. It is observed from Fig. 1 that the value of R is close to one. So, the developed MARS has shown good predictive abilities for determination of f_s of driven piles in clay. Table 2 shows the ANOVA decomposition of the developed MARS. The value of GCV is maximum for D. So, D has maximum impact on the predicted f_s. Geometry of pile foundation has more effect on f_s than soil properties. The results of MARS have been compared with the ANN model developed by Goh (1995). Comparison has been

Table 1 Basis functions and their corresponding coefficient

Basis functions	Equation	Coefficient (c_m)
$B_1(x)$	$\max(0, S_u - 0.020)$	2.832
$B_2(x)$	$\max(0{,}020 - S_u)$	–2.349
$B_3(x)$	$B_2(x)$* $\max(0{,}020 - S_u)$	53.977
$B_4(x)$	$\max(0{,}0.274 - D)$	–0.183

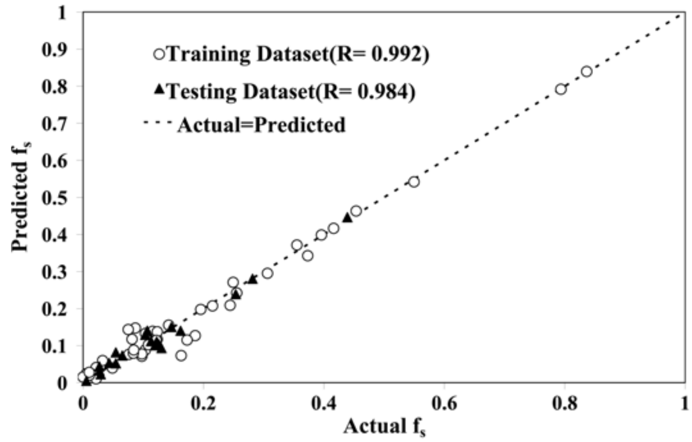

Fig. 1 Performance of the developed MARS

carried out in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The values of RMSE and MAE have been determined by using the following relation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (f_{sai} - f_{spi})^2}{n}} \tag{8}$$

$$MAE = \frac{\sum_{i=1}^{n} |f_{sai} - f_{spi}|}{n} \tag{9}$$

where f_{sa} and f_{sp} are actual and predicted f_s values respectively and n is the number of data. The performance has been compared only for the testing dataset. Fig. 2 illustrates the bar chart of

Table 2 ANOVA decomposition of the developed MARS

GCV	Standard deviation	Variables	Basis function	Functions
0.193	0.017	D	1	1
0.075	0.43	S_u	2	2
0.189	0.02	D, S_u	1	3

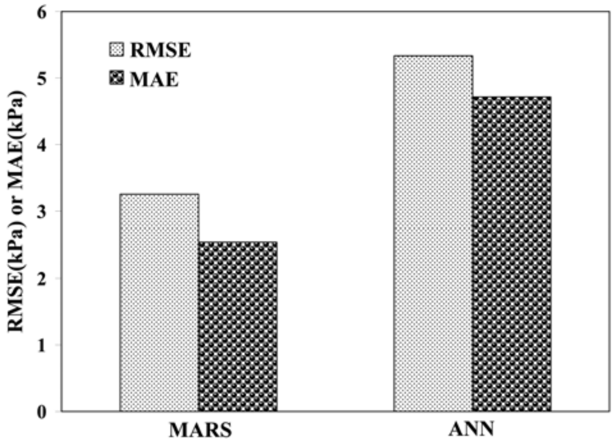

Fig. 2 Bar chart of RMSE and MAE for the MARS and ANN models

RMSE and MAE for the MARS and ANN models. It is observed from Fig. 2 that the value of RMSE and MAE of MARS is smaller than the ANN model. The developed MARS outperforms the ANN model. The developed ANN (Goh 1994) did not give any equation for determination of f_s of driven piles in clay. The MARS gives an equation (Eq. (7)) for determination of f_s.

4. Conclusions

This article describes MARS model for prediction of f_s of driven piles in clay. The developed MARS has shown good predictive abilities. The comparative study proves that the performance of the developed MARS is better than the ANN. Geotechnical engineers can use the developed equation for determination of f_s of driven piles in clay. The developed MARS models relationship that is nearly additive or involve interactions with fewer variables. In summary, it can be concluded that MARS can be used as an effective tool for solving different problems in geotechnical engineering.

References

Akin, S. and Karpuz, C. (2008), "Estimating drilling parameters for diamond bit drilling operations using articial neural networks", *Int. J. Geomech.*, **8**(1), 68-73.

Andrés, J.D., Lorca, P., Juez, F.J.D.C. and Lasheras, F.S. (2011), "Bankruptcy forecasting: a hybrid approach using fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS)", *Expert Syst. Appl.*, **38**,1866-1875.

Attoh-Okine, N.O., Cooger, K. and Mensah, S. (2009), "Multivariate adaptive regression (MARS) and hinged hyperplanes (HHP) for doweled pavement performance modeling", *Constr. Build. Mater.*, **23**(9), 3020-3023.

Burland, J.B. (1973), "Shaft friction of piles in clay-a simple fundamental approach", *Ground Eng.*, **6**(3), 30-42.

Chandler, R.J. (1968), "The shaft friction of piles in cohesive soil in terms of effective stress", *Civil Eng. Public Works Rev.*, **63**, 48-51.

Crino, S. and Brown, D.E. (2007), "Global optimization with multivariate adaptive regression splines", *IEEE T. Syst. Man Cy. B*, **37**(2), 333-340.

Deconinck, E., Coomans, D. and Heyden, V.Y. (2007), “Exploration of linear modelling techniques and their combination with multivariate adaptive regression splines to predict gastro-intestinal absorption of drugs”, *J. Pharmaceut. Biomed*., **43**,119-130.
Ekman, T. and Kubin, G. (1999), “Nonlinear prediction of mobile radio channels: measurements and mars model designs”, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, **5**, 2667-2670.
Friedman, J.H. (1991), “Multivariate adaptive regression splines”, *Ann Stat.*, **19**, 1-141.
Friedman, J.H. and Roosen, C.B. (1995), “An introduction to multivariate adaptive regression splines”, *Stat. Methods Med. Res*., **4**, 197-217.
Goh, A.T.C. (1995), “Empirical design in geotechnics using neural networks”, *Geotechnique*, **45**(4), 709-714.
Hastie, T., Tibshirani, R. and Friedman, J.H. (2003), *The elements of statistical learning*, Springer-Verlag, New York.
Kecman, V. (2001), *Leaming and soft computing: support vector machines, neural networks, and fuzzy logic models*, The MIT press, Cambridge, Massachusetts, London, England.
Lewis, P.A.W. and Stevens, J.G. (1991), “Nonlinear modeling of time series using multivariate adaptive regression splines (mars)”, *J. Am. Stat. Assoc.*, **86**(416), 864-877.
Mayoraz, F. and Vulliet, L. (2002), “Neural networks for slope movement prediction”, *Int. J. Geomech.*, **2**, 153-173.
McClelland, B. (1972), “Design and performance of deep foundations in clay”, General Report. Am. Sot. Civ. Engrs Specialty Conf Performance of Earth and Earth-supported Structures, **2**, 111-114.
Meyerhof, G.G. (1976), “Bearing capacity and settlement of pile foundations”, *J. Geotech. Eng. Div. -ASCE*, **102**(GT3), 195-228.
Miranda, T., Correia, A.G., Santos, M., Sousa, L.R. and Cortez, P. (2011), “New models for strength and deformability parameter calculation in rock masses using data-mining techniques”, *Int. J. Geomech.*, **11**(1), 44-58.
Park, D. and Rilett, L.R. (1999), “Forecasting freeway link ravel times with a multi-layer feed forward neural network”, *Comput. Aided Civil Infrastruct. Eng.*, **14**, 358-367.
Parry, R.H.G. and Swain, C.W. (1977a), “Effective stress methods of calculating skin friction on driven piles in soft clay”, *Ground Eng.*, **10**(3), 24-26.
Parry, R.H.G. and Swain, C.W. (1977b), “A study of skin friction on piles in stiff clay”, *Ground Eng.*, **10**(8), 33-37.
Randolph, M.F., Carter, J.P. and Wroth, C.P. (1979), “Driven piles in clay-the effects of installation and subsequent consolidation”, *Geotechnique*, **29**(4), 361-393.
Samui, P. and Sitharam, T.G. (2010), “Site characterization model using articial neural network and kriging”, *Int. J. Geomech.*, **10**(5), 171-180.
Sekulic, S. and Kowalski, B.R. (1992), “MARS: a tutorial”, *J. Chemometr.*, **6**, 199-216.
Tomlinson, M.J. (1971), *Some effects of pile driving on skin friction, behaviour of piles*, Institution of Civil Engineers, London, 107-I 14.
Vidoli, F. (2011), “Evaluating the water sector in Italy through a two stage method using the conditional robust nonparametric frontier and multivariate adaptive regression splines”, *Eur. J. Oper. Res.*, **212**, 583-595.
Yang, C.C., Prasher, S.O., Lacroix, R. and Kim, S.H. (2004), “Application of multivariate adaptive regression splines (mars) to simulate soil temperature”, *Trans. ASAE*, **47**(3), 881-887.
Zaman, M., Solanki, P., Ebrahimi, A. and White, L. (2010), “Neural network modeling of resilient modulus using routine subgrade soil properties”, *Int. J. Geomech.*, **10**(1), 1-12.
Zhang, G., Xiang, X. and Tang, H. (2011), “Time series prediction of chimney foundation settlement by neural networks”, *Int. J. Geomech.*, **11**(3), 154-158.

CC