Geotechnical Engineering in the XXI Century: Lessons learned and future challenges
N.P. López-Acosta et al. (Eds.)

doi:10.3233/STAL190147

Genetic Algorithm-Based Load-Settlement Curves of Driven Piles in Glacial Deposits

Jinyuan LIU[a,1] and Markus JESSWEIN[a]
[a]*Ryerson University, Toronto, Canada*

Abstract. This paper presents a genetic algorithm (GA) method to build the load-settlement curve of a driven pile in glacial deposits. The load-settlement mechanism of an axially loaded piles is a complex soil-structure interaction problem and influenced by many factors, including the pile dimensions, installation method, and ground conditions. Due to these uncertainties, pile load tests have been used for decades to verify the design assumptions in Ontario, Canada. A database of more than 168 pile load tests was collected by the Ministry of Transportation of Ontario (MTO). In this study, a simple hyperbolic curve is adopted to simulate the load-settlement curves of 36 pile compression tests in various parts of Ontario. GA, one of artificial intelligence techniques, is applied to correlate the soil and pile parameters to the two hyperbolic parameters. Based on pile and ground conditions, the proposed formulas can be used to build a load-settlement curve for the serviceability design of a pile driven in glacial deposits.

Keywords. Pile, load-settlement curve, genetic algorithm, hyperbolic curve fitting, serviceability limit state.

1. Introduction

Pile foundations have been used for centuries to support different structures, such as bridges and buildings. Yet, a pile has to meet two limit design conditions: the ultimate limit state (ULS) regards the ultimate bearing capacity of the pile, and the serviceability limit state (SLS) regards the deformation of the pile within the tolerable limit. The design involves many influential factors and uncertainties, especially from the soil conditions. For example, glacial deposits covering large portion of North America, including the province of Ontario in Canada, are known for their inconsistent material properties, which creates difficulties for engineers to accurately evaluate their properties (Barnett 1992). In addition, the SLS is significantly influenced by the installation method and pile geometry. In all, the behavior of the pile-soil interface is a complex system.

Over the last few decades, several approaches were proposed to predict the deformation of a pile for either fine-grained or coarse-grained soils. These approaches may experience variability with mixed soil profiles and high amounts of silts. Also, the relationships were usually developed for local soil conditions in various regions of the world. Soils can easily change between regions. There is limited research conducted on the serviceability design of piles in glacial deposits, in particularly in Ontario, Canada. This study is to address this urgent need.

[1] Corresponding Author, Professor, Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, Canada M5B2K3; E-mail: jinyuan.liu@ryerson.ca

As a potential solution, a rigorous statistical analysis is conducted on field and experimental data. A database of pile load tests was accumulated over four decades by the Ministry of Transportation of Ontario (MTO) and was used in this study. Both closed-ended pipe piles and H piles driven into fine- or coarse-grained soils are selected from a total of 36 pile compression tests. Pile load tests may be expensive and time consuming, but they are dependable and reliable to verify design assumptions for local soil conditions. For example, McVay et al. (2000) achieved reliable pile design methods for regions in Florida after analysing results from pile load tests.

There are a number of methods to predict the settlement of axially loaded piles, including the continuum method and load transfer method. A hyperbolic model with two parameters is adopted in this study to fit the measured load-settlement data due to its simplicity. The uncertainties within the load-settlement curves are captured by a bivariate random vector containing the hyperbolic parameters as its components. A similar method has been applied by many researchers in evaluating the SLS of geotechnical structures (Huffman and Stuedlein 2014; Phoon and Kulhawy 2008).

The procedures of this study include extracting the measured load-settlement curves from the test results; simulating the load-settlement curves using a hyperbolic model with two parameters, evaluating the dependency of two parameters; conducting a GA analysis to correlate two parameters with pile and soil parameters; and evaluating the performance of the proposed method.

2. MTO pile load test database

In this investigation, the soil measurements and pile test results are collected from MTO. The focus is on pipe piles and H piles that were tested with static axial compressive loads. Tests occurred in various regions of Ontario, but a majority of piles were located in Southern Ontario. Piles ranged in length from 3 m to 45 m. H piles were either HP310x79 or HP 310x110 with one exception case of HP 370x108 pile, while pipe piles were mainly 324 mm outside diameter pipe filled with concrete with one exception of 305 mm pipe. For the analysis, tests are selected if plunging failure or a significant amount of displacement occurred, and sites were selected that dominated in fine- or coarse-grained soils. The same approximate number of piles were driven into both soil types from the sample of tests. Many of the sites with fine-grained soils actually contained glacial deposits. A portion of the piles in the coarse-grained soils were located in waterbodies, such as rivers or lakes.

Borehole logs provided the results from field investigations. Soil classifications and unit weights were recorded at different depths. At some sites, the percentage of clay, silt, sand, and gravels were included in the soil distribution. The Atterberg limits were also provided. Every site varied in the extent and diversity of the site investigations. Unconfined compressive strength (UCS), unconsolidated undrained (UU) triaxial, and field vane shear (FVST) tests measured C_u at some of the sites. N-values by standard penetration tests (SPT) were commonly found, especially for sites with coarse-grained soils. Overall, a variety of soil tests occurred, but a majority of the tests were SPT.

3. Simulation of pile load-settlement curves

The primary goal of this study is to characterize the model uncertainties in modelling the load-settlement behaviour of axially loaded driven piles in glacial deposits in Ontario, Canada. A hyperbolic model with two parameters is adopted to fit the measured load-settlement data. A non-linear least-square regression analysis is conducted in SPSS (IBM 2015) to minimize the mean absolute percentage error (MAPE) and determine the hyperbolic parameters. Based on the simulated results, a statistical analysis is conducted to evaluate the performance of the hyperbolic model, including the mean value, coefficient of variation (COV), and regression coefficient, and probability distributions.

A simple method is to apply curve fitting techniques to the measured load-settlement curves. This method has been applied for reliability-based SLS design of geotechnical structures by many researchers (Phoon and Kulhawy 2008; Uzielli and Mayne 2011; Huffman and Stuedlein 2014). A simple hyperbolic curve is selected for this study to fit the measured load-settlement curves of pile load tests.

$$s = aQ/(1 - bQ) \tag{1}$$

where Q = applied load; s = pile settlement; and *a* and *b* = hyperbolic parameters, which are physically meaningful.

The reciprocals of a and b represent the initial slope and asymptote of the load-settlement curves. Although a high coefficient of determination, R^2, is obtained for each compression or tension load test, the problem is the quality of fitting at the late stages of loading, particularly for gradual yielding conditions. More research is being conducted to address this issue to reduce the errors.

A scatterplot of the two parameters, a and b, is shown in Figure 1 along with the histogram of each parameter. The Spearman correlation factor of two hyperbolic parameters is calculated as -0.2976, which indicates a negative correlation. This negative correlation can be explained as a small initial slope of the load-settlement curve (i.e., a large a value) implies a slowly decaying curve which is generally associated with a less well-defined and larger asymptote (i.e., a small *b* value). Table 1 presents the statistics of the hyperbolic parameters.

Table 1. Statistical properties of hyperbolic parameters for all 168 pile tests.

No. of piles	Hyperbolic parameter, *a*		Hyperbolic parameter, b	
	Mean	**Standard deviation**	**Mean**	**Standard deviation**
168	0.1291	0.1292	0.1486	0.0185

4. Introduction of the genetic algorithm

This study aims to correlate the soil and pile parameters with two hyperbolic parameters, *a* and *b,* using GA. GA is an optimization approach inspired by Darwin's theory of evolution (Banzhaf et al. 1998). In nature, chromosomes give an organism its attributes to survive and succeed in an environment. Through reproduction, organisms can adapt and evolve to their environment. GA represents the problem domain as a chromosome. In this case, a GA is developed to conduct symbolic regression and predict the parameters of hyperbolic curves for a given pile in a certain ground condition. For symbolic regression, chromosomes are evolvable strings or vectors within a programming language that represent mathematical expressions or functions, and the genes of a chromosome represent the components of a function: a variable, constant, or operator. GA has been recently applied in geotechnical research, including piles (Chan et al. 2009; Momeni et al. 2014).

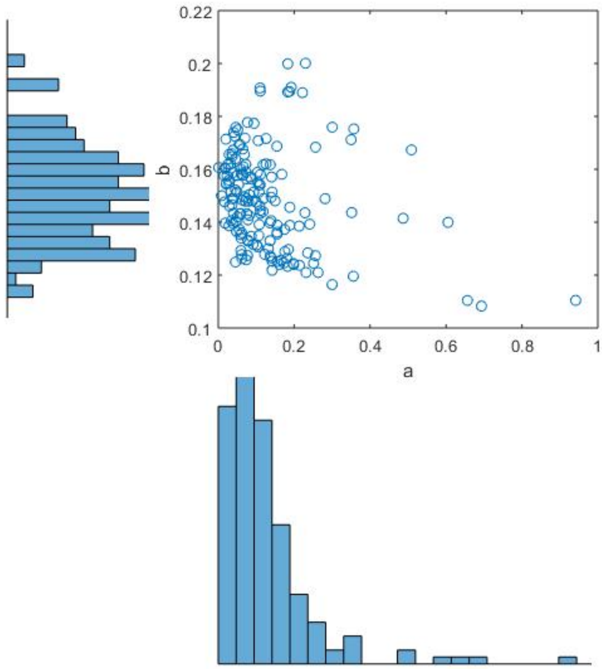

Figure 1. The random distribution of two parameters from hyperbolic curve fitting.

The GA in this investigation searches for a solution through six general steps: chromosome creation, evaluation, selection, crossover, mutation, and constant refinement. First, a population of functions or chromosomes were created randomly. Multiple attempts for a problem are made at once in a trial, or generation, by having a population of chromosomes with different attributes. The performance or fitness of a single chromosome is measured by an objective function. From the population of chromosomes, potential parents are selected for the creation of offspring, which are new functions. Typically, during selection, a preference is given to chromosomes with a higher fitness. The population size remains constant throughout every generation, and the previous chromosomes, or at least a majority, are replaced by new offspring. The population then evolves through several generations by reproduction mechanisms, such as crossover and mutation. Since the creation of the chromosomes is random, the constants within the functions are refined by randomly changing the values through several attempts, and the set of constants that provide a better fitness are kept for the next generation. If a regression analysis is repeated with multiple trials with the same initial conditions, GA can complete the analysis with a similar level of fitness but provides a different solution since GA is a stochastic method.

The GA is created with MATLAB® (Mathworks 2017) and applies the Multi Expression Programming (MEP) technique (Oltean & Dumitrescu, 2002) to encode and evaluate the chromosomes. MEP is based on the activation of programs or code with integers and can efficiently encode or decode functions compared to other techniques.

The fitness function is a crucial component of GA to rank a population of potential solutions. For regression, it compares the predicted values (P_R) to the measured values (M) with the root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum(P_R - M)^2} \tag{2}$$

where n is the number of analyzed piles. A lower RMSE indicates a better fit between the measured and predicted values.

Generally, care is needed to ensure that illogical errors do not occur during evaluation. Division operators may be protected by simply returning the numerator if a denominator of zero is found (Banzhaf et al. 1998), but Oltean & Dumitrescu (2002) recommend mutating division into a variable or constant. Other operators are protected and transformed as suggested by Brameier & Banzhaf (2007). More details about GA procedures can be found in Jesswein et al. (2018).

4.1. Input parameters used in GA analysis

The variables in this analysis include the pile type (P_T); unplugged pile perimeter (P); pile embedment length (L); average SPT N-values (N_{avg}) along a pile side; SPT N-value at the pile base (N_B); average SPT N-value (N_{Top}) along the top 3 m of the pile; average SPT N-value (N_{Bot}) along the bottom 3 m of the pile; the dominating soil type (S_T); pile modulus (E_P) multiplied by the cross-sectional area of the pile (A_P); unplugged area of the pile base (A_b); and the load (Q_{max}) applied to the pile to achieve plunging failure or a significant amount of displacement. N-values are not corrected. S and P_T are binary variables. S is equal to 1 for noncohesive soils or 2 for cohesive soils, while P_T is equal to 1 for pipe piles and 2 for H piles. Since the actual dimensions of the piles are known, H piles are assumed to be unplugged for the analysis.

The GA performed 3 trials (runs) to regress the 11 variables with each of the 2 hyperbolic parameters. Each trial will evolve 2000 chromosomes for 200 generations. The other main settings of the GA include: chromosome length, 45; population size for brood crossover, 5; crossover type, uniform with brood recombination; and function sets, +, -, ×, ÷, power, exponential, logarithmic, hyperbolic tangent.

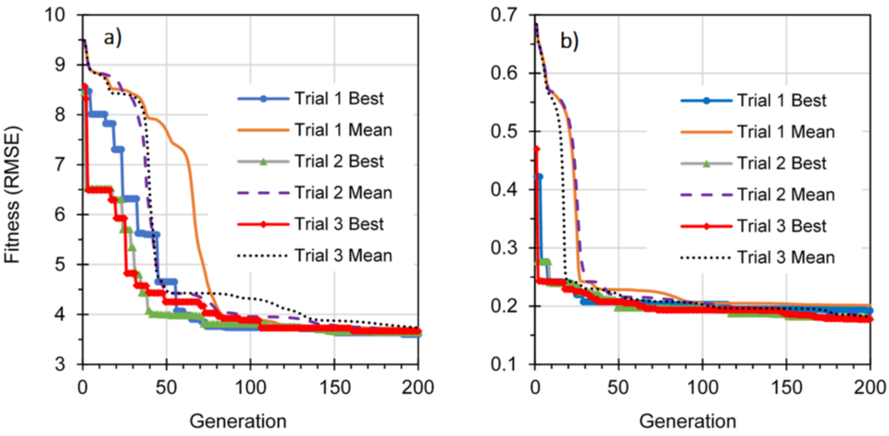

Figure 2. Fitness performance of GA for hyperbolic parameters, *a* and *b*.

4.2. Results from the genetic algorithm

Two parameters, *a* and *b,* are analyzed by GA separately in three trials each. Figure 2 shows the average and lowest RMSE within the population of chromosomes for each trial. For each of the three trials, the analysis usually is terminated with a similar RMSE. As a better links are made, the brood recombination during crossover and constant refinement results in sudden drops in the best fitness throughout the generations. The RMSE for the best function generally stabilizes after 100 generations.

At the end of the 3 trials, the function with the best fitness is collected as follows:

$$a = 0.001\left[\ln(N_{avg}) + \frac{S_T^{P_T}}{P_T + 9.67 - N_{Top}} + S_T\right] \quad (3)$$
$$(\mathrm{R}^2 = 0.856, \mathrm{RMSE} = 3.60)$$

$$b = \frac{(A_b + 0.965)\cdot(956 - 9.41\cdot L)}{1000\cdot Q_{max}} \quad (4)$$
$$(\mathrm{R}^2 = 0.946, \mathrm{RMSE} = 0.177)$$

From the selection of variables, GA usually regresses the most influential factors. For parameter a, the commonly regressed variables, such as N_{avg}, S_T, and L, can be related to the pile side resistance or soil shear modulus along the pile. For a hyperbolic load-displacement response, it is logical that Q_{max} would be inversely proportionate to *b*, but other variables are included to describe the variability of the system as some piles did not perfectly conform to the hyperbolic relationship. Q_{max} and L are consistently found in each of the 3 best functions for parameter b and have negative or inverse relationships. N_{avg} and N_{Top} are also regressed with b, and their presence may indicate the varying soil stiffness. Since the piles had similar dimensions and stiffness, E_pA_p, is rarely chosen by GA. Instead, L is commonly found and will likely be more related to the pile stiffness and soil resistance.

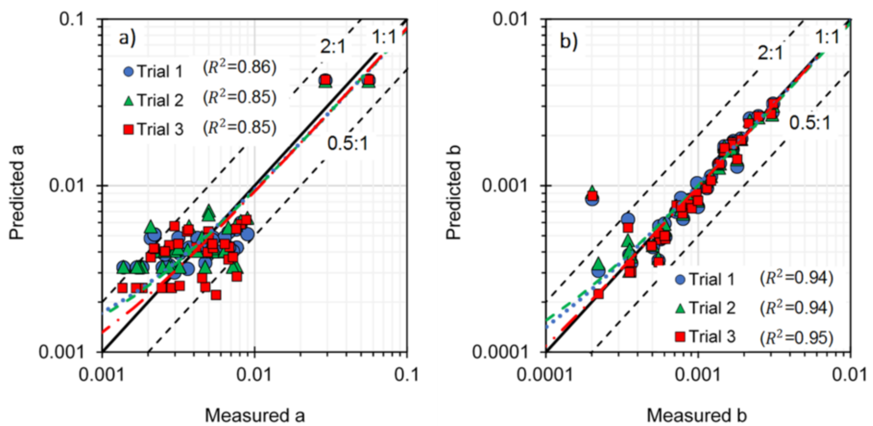

Figure 3. Comparison of simulated and predicted hyperbolic parameters, *a* and *b*.

The predictions for a and b by GA is compared with the simulated results and shown in Figure 3. Although a low MAPE and high R^2 are achieved with the simulated hyperbolic curves, not every load-displacement plot, particularly from mainly end-bearing piles, conform perfectly to the hyperbolic relationship. The lower values of b have the greatest variability in Figure 3b and are usually found to be end-bearing piles. These piles generally do not have a pure plunging failure and experienced a more gradual yielding slope on the load-displacement curve. In Figure 3a, two outliners have significantly higher values for a compared to the remaining tests, and these outliners are pipe piles in weak clayey silts. The other piles are generally surrounded by stiffer soils. The poor predictions for a can be improved with a larger sample size, and they indicate the importance of applying the correct yielding load-displacement response.

5. Application GA-driven load-settlement curve simulation

The pile displacement is predicted with the hyperbolic parameters from Equations (3) and (4) and compared with measured values. The predictions tend to follow the 1:1 line with an average predicted to measured ratio of 1.38, but the R^2 is low at 0.28 and the MAPE is 79.7 %. Variability is mainly found with very small displacements (< 1 mm) or displacements greater than 20 mm. The accuracy of the large displacements, in particular, will largely depend on the ability to capture the correct yielding behaviour of the pile.

6. Conclusions and discussions

A total of 36 load-settlement curves of piles tested in glacial deposits are studied with a hyperbolic relationship and correlate the soil and pile parameters with the two hyperbolic parameters using genetic algorithm, one of artificial intelligence techniques. Two formulas are proposed from the GA-based analysis to develop the load-settlement of driven piles in glacial deposits. Based on the comparison between predicted and measured settlement, it is found GA is able to achieve reasonable accuracy, particularly in the displacement ranging from 1 mm and 20 mm.

GA is able to achieve accurate predictions for parameter b with Q_{max} and L likely the most influential factors. Parameter a is likely influenced by the soil conditions around the pile shaft, but the limited ability to predict parameter a may indicate the importance to estimating the soil-pile stiffness at the early stage of loading. The hyperbolic relationship is likely appropriate for a particular failure mechanism experienced by the piles, but the combined interaction between the side and tip resistance adds to the challenge to predict the load-displacement response.

The limitations of this research are based on the accuracy of the data acquired from field and pile load tests. Every site varied in the frequency and use of SPT. In the analysis, the pile types need to be separated to reduce the variability, or the GA can be modified to better handle categorical variables. Mixed soil profiles can be considered with AI methods, such as fuzzy logic. In the future, the authors wish to investigate other curve fitting formulas in addition to hyperbolic curves for the load-displacement response. Since some measurements have a limited reliability or generalized description of the soil, such as N_{avg}, it may be appropriate to use other AI techniques and predict the hyperbolic parameters with a probable range instead of a static value. The piles in this study had similar section sizes. More study is required to investigate the impact of pile sections on the findings.

Acknowledgements

The presented research was made possible with funding by the Ministry of Transportation of Ontario with the Highways Infrastructure Innovations Funding Program. In addition, the authors would like to thank Mr. David Staseff and Ms. Minkyung Kwak from MTO for sharing the data of pile load tests.

References

W. Banzhaf, P. Nordin, R.E. Keller, F.D. Francone. *Genetic Programming: An Introduction*, Morgan Kaufmann Publishers Inc., 1988.

P.J. Barnett. *Chapter 21: Quaternary Geology of Ontario*. In *Geology of Ontario*. Special Vol. 4, Part 2, 1011-1088, 1992.

M. Brameier, W. Banzhaf. *Linear Genetic Programming*, 2007.

C.M. Chan, L.M. Zhang, J.T. Ng. Optimization of pile groups using hybrid genetic algorithms, *Journal of Geotechnical and Geoenvironmental Engineering* 2009. 135 (4).

J.C. Huffman, A.W. Stuedlein. Reliability-based serviceability limit state design of spread footings on aggregate pier reinforced clay. *Journal of Geotechnical and Geoenvironmental Eng.* 2014. 140 (10).

IBM. 2015. SPSS (Version 23.0) [Software]. IBM, Armonk, NY.

M. Jesswein, J. Liu, M. Kwak Predicting the side resistance of piles using a genetic algorithm and SPT N-values, *2018 GeoEdmonton CGS Conference*, AB, Canada.

Mathworks. 2017. Matlab [Computer software]. Natick, MA, USA.

M.C. McVay, B. Birgisson, et al. Load and resistance factor design (LRFD) for driven piles using dynamic methods – a Florida perspective, *Geotechnical Testing Journal (2002)*, 23(*1*): 55-66.

E. Momeni, R. Nazir, J.D. Armaghani, H. Maizir. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. Measurement (2014), 54, 122-131.

K.K. Phoon, F. H. Kulhawy. Serviceability-limit state reliability-based design. *Reliability-based design in geotechnical engineering: Computations and applications* (2008), 344–383.

M. Oltean, D. Dumitrescu. *Multi Expression Programming.* 2002.

M. Uzielli, P.W. Mayne. Serviceability limit state CPT-based design for vertically loaded shallow footings on sand. *Geomech. Geoeng.* 2011. 6 (2): 91–107.