Predicting the Side Resistance of Piles Using a Genetic Algorithm and SPT N-Values

Markus Jesswein & Jinyuan Liu
Department of Civil Engineering – Ryerson University, Toronto, ON, Canada
Minkyung Kwak
Ministry of Transportation of Ontario, Toronto, ON, Canada

ABSTRACT
A genetic algorithm (GA) was developed to predict the side resistance of piles in Ontario with the standard penetration test blowcounts (SPT N-values). Pile foundations commonly support structures by transferring loads deeper into the ground; unfortunately, the soil characteristics, such as the strength and grain-size distribution, are rarely consistent at a site. Due to the spatial variability of the soil, challenges arise to accurately predict the ultimate axial capacity of piles. This research aims to mitigate this problem by implementing genetic programming. Since the 1950s, the Ministry of Transportation of Ontario has gathered approximately 100 pile load tests in various soil conditions, and a total of 23 piles were selected for this study. These piles were either H piles or pipe piles and were subjected to extension load tests. A GA was created to correlate the side resistances to the SPT N-values, and the developed relationships were compared to existing design methods for different soil types. In all, the ultimate goal of this research was to improve local pile design in Ontario's practice.

RÉSUMÉ
Un algorithme génétique (AG) a été développé pour prédire la résistance au cisaillement des pieux en Ontario avec les compteurs d'essai de pénétration standard (valeurs N du SPT). Les fondations de pieux soutiennent généralement les structures en transférant les charges plus profondément dans le sol; Malheureusement, les caractéristiques du sol, telles que la force et la distribution granulométrique, sont rarement constantes sur un site. En raison de la variabilité spatiale du sol, il est difficile de prédire avec précision la capacité axiale ultime des piles. Cette recherche vise à atténuer ce problème en mettant en œuvre la programmation génétique. Depuis les années 1950, le ministère des Transports de l'Ontario a recueilli environ 100 essais de chargement de pieux dans diverses conditions de sol, et un total de 23 piles ont été sélectionnées pour cette étude. Ces piles étaient des piles H ou des pieux tubulaires et ont été soumises à des essais de charge d'extension. Un AG a été créé pour corréler les résistances au cisaillement aux valeurs N du SPT, et les relations développées ont été comparées aux méthodes de conception existantes pour différents types de sols. Dans l'ensemble, le but ultime de cette recherche était d'améliorer la conception locale des piles dans la pratique ontarienne.

1 INTRODUCTION

Deep foundations are designed to support bridges and buildings, but accurately predicting the capacity of a pile is a challenge due to the influences from the installation method, pile geometry, and soil properties. Especially since glacial tills are commonly found in Ontario, the site conditions are rarely consistent as the soil strength and contents can vary spatially. In addition, the standard penetration test (SPT) is unreliable because it lacks accuracy. Yet, it is commonly used in site investigations because it is a cheap and simple measurement technique and can be applied in various soil conditions, including gravel and cobble rich soils. Numerous correlations have been proposed to predict the pile capacity with SPT N-values, but many of the design methods were developed by averaging or generalizing the soil conditions. Also, a limited number of design methods have been proposed for soils with stiff clays and glacial tills, especially for Ontario.

This investigation addresses the uncertainty in design by developing a genetic algorithm (GA) to correlate SPT N-values to the side resistance of driven piles. The Ministry of Transportation of Ontario (MTO) accumulated a database with over 100 pile load tests, and results from 23 high-quality extension load tests were selected for this study. GAs can efficiently correlate several variables compared to traditional statistical approaches, and a GA was developed to predict the frictional resistance with multiple N-values and soil types along the length of a pile.

2 RESEARCH BACKGROUND

2.1 Current Design Methods

A pile subjected to axial compressive loading can experience two mechanisms: the side resistance (Q_s) and the tip resistance (Q_p). The side resistance is the friction between the soil and pile walls, while the tip resistance is developed by the strength of the soil at the pile base. As a pile reaches its maximum load, these mechanisms contribute towards the ultimate pile resistance, or capacity (Q_u):

$$Q_u = Q_s + Q_p \quad [1]$$

In this paper, the focus is on the side resistance, which depends on the unit side resistance (q_s), pile perimeter (P) and embedment length (L).

$$Q_s = q_s P L \qquad [2]$$

Usually, the side resistance is directly predicted by the average SPT N-value ($\bar{N}$) along the pile length. As shown in Table 1, many empirical correlations have been proposed for various soil conditions. Unfortunately, the average soil conditions may not accurately represent the soil behaviour.

Table 1. Existing Design Methods for Pile Side Resistance

Soil Type	Reference	Equation for q_s (kPa)
Cohesive	Shioi & Fukui (1982)	$q_s = 9.8\,\bar{N}$
Noncohesive	Meyerhof (1956)	Pipe Piles: $q_s = 1.9\,\bar{N} \leq 100$ H Piles: $q_s = 1.0\,\bar{N} \leq 100$
	Shioi & Fukui (1982)	$q_s = 1.9\,\bar{N}$
All	Brown (2001)[1]	$q_s = 1.8\,\bar{N} + 25$
	Decourt (1982)[2]	$q_s = 3.3\,\bar{N} + 9.8 \leq 170$

[1] Recommended SPT range is $3 \leq N \leq 50$ and [2] $3 \leq N \leq 15$.

Several references also report various influences on the capacity with the pile length (Meyerhof 1976; Vesic 1977; Poulos et al. 2001). This relationship can be due to load dissipation along the pile length; events related to the pile installation process, such as whipping; or the confinement of the effective stress with noncohesive soils.

In all, the side resistance is influenced by many variables, including the soil type, soil strength, pile geometry, and installation process. The complex and nonlinear relationship between the pile and soil can be predicted with a GA.

2.2 An Introduction to Genetic Algorithms

Genetic algorithms are 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. A GA represents the problem domain as a chromosome. In this case, a GA was developed to conduct symbolic regression and predict the side resistance. For symbolic regression, the genes of a chromosome represented the components of a function: a variable, constant, or operator.

A generic GA searches for a solution through five general steps: chromosome creation, evaluation, selection, crossover, and mutation. First, multiple attempts for a problem are made at once in a trial, or generation, by generating 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. 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.

A GA is a stochastic method. If a regression analysis was repeated with multiple trials and had the same initial conditions, the GA can complete the analysis with a similar level of fitness but provide a different solution. A GA is also data-driven and, depending on the complexity of a problem, may not find an exact solution or global minimum, but it is an efficient optimization approach, especially if aspects of a problem are unknown.

3 RESEARCH METHODOLOGY AND RESULTS

3.1 Overview of the Methodology

For driven piles in heterogeneous soils, the goal was to improve the predictability of the side resistance. A new design method was developed by following four steps. (1) Results from pile load tests and soil measurements were collected from a database by MTO. (2) For every pile, the measured side resistance was obtained from load-displacement responses that were measured at the pile top. (3) A GA was then developed to nonlinearly correlate the side resistance in heterogeneous soils with SPT N-values. (4) In the end, the accuracy of the GA was compared to existing design methods.

3.2 Testing Sites and Pile Load Tests

For this investigation, borehole logs for the soil conditions and records on the pile load tests were collected from MTO. Driven piles were either H piles or steel pipe piles, and they were subjected to static axial-tension loads to measure the side resistance. The pile properties and

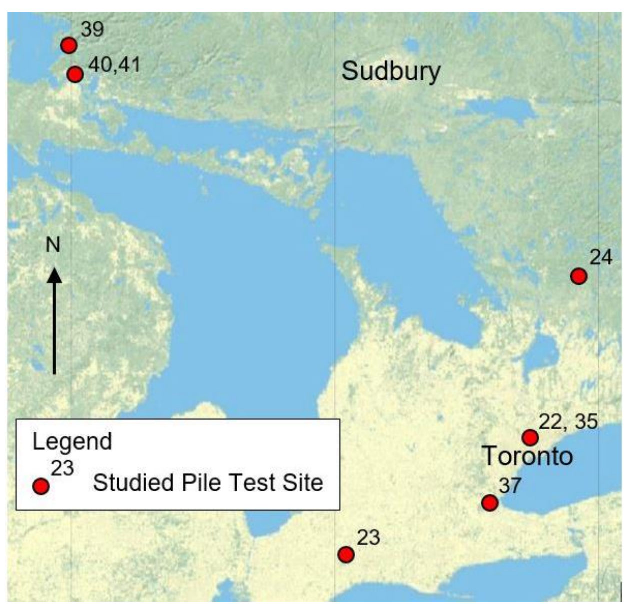

Figure 1. Location of Studied Sites

Table 2. Details on the Studied Piles

Site No.	Pile No.	Pile Type[1]	Length[2] (m)	Embedded Soil Type[3]	Q_s (kN)
22	3	324 OD Pipe	15.30	Clayey Silt	118
22	4	324 OD Pipe	30.15	Clayey Silt	340
23	2	324 OD Pipe	3.02	Silty Clay	209
23	3	HP 310x110	3.05	Silty Clay	236
24	2	324 OD Pipe	15.39	Sand	372
24	3	324 OD Pipe	22.40	Sand	401
24	4	HP 310x79	22.40	Sand	403
24	5	HP 310x79	15.39	Sand	263
35	1	HP 310x110	14.69	Layered Clayey Silt and Silty Sand	506
35	4	324 OD Pipe	14.69	Layered Clayey Silt and Silty Sand	730
35	5	HP 310x110	27.58	Layered Clayey Silt and Silty Sand	1493
37	3	HP 310x79	14.48	Sand to Silty Sand	333
37	4	HP 310x79	38.94	Sand to Silty Sand	1394
37	5	HP 310x79	31.24	Sand to Sandy Silt	420
37	6	HP 310x110	14.48	Sand to Silty Sand	383
37	7	HP 310x110	45.29	Sand to Silty Sand	1524
37	8	HP 310x110	30.92	Sand to Silty Sand	699
39	2	HP 310x110	25.50	Silty Sand; Layered Clay and Silt	614
39	3	324 OD Pipe	25.40	Silty Sand; Layered Clay and Silt	470
40	2	HP 310x110	24.50	Layered Sand and Silty Clay	598
40	3	324 OD Pipe	17.20	Sandy Silt to Sand	505
41	2	HP 310x110	19.50	Sand	1052
41	3	324 OD Pipe	16.00	Sand	664

[1] Steel H pile designations are depth (mm) by weight (kg/m). Steel pipe piles were filled with concrete before testing, and OD is the outside diameter (mm); [2] Embedment Length; [3] Classifications according to MTO standards.

ultimate side resistances are in Table 2. From the 23 piles, 9 were pipe piles, and 14 were H piles. The embedment lengths varied from 3 m to 45 m, but most of the piles were between 12 to 25 m long.

Figure 1 shows the locations of the sites. The soils were generally compact or stiff, but some loose sands were found. Noncohesive soils were common in the database. For the pipe piles, two were mainly in cohesive soils, four dominated in noncohesive soils, and the remaining had mixed soil conditions. One H pile was fully embedded in cohesive soils, while nine H piles were in noncohesive soils. Borehole logs contained the soil type, SPT N-values, and occasionally, the unit weights at the sites. SPT N-values were corrected according to CGS (2006) for a hammer efficiency of 60% and the overburden conditions.

3.3 Measured Side Resistance (Q_s)

The load-displacement response was measured with dial gauges at the top of the piles during the load tests. The failure load was determined by the criteria from De Beer (Fellenius 1980). De Beer recommends plotting both axes of the load-displacement curves with a log-scale, and the failure load is indicated by the largest change in slope on the plot (Fellenius 1980).

3.4 Construction of the Genetic Algorithm

3.4.1 Introduction

Each pile was divided into 50 segments to consider the varying side resistance along their length. The variables in the analysis included the corrected SPT N-values (N), the soil type (S), effective stress (σ'), and pile slenderness ratio ($(L-z)/D$). N-values were corrected for the hammer efficiency (N_{60}) for cohesive soils, and the overburden correction was also applied ($(N_{60})_1$) for noncohesive soils. The soil type was a binary variable equal to 1 for noncohesive soils or 2 for cohesive soils. The slenderness ratio was modified to determine the side resistance at any depth. It was composed of the embedment length (L), depth to the centre of a pile segment (z), and maximum width or diameter of a pile (D). The side resistance of a H pile will be influenced if the soil creates an unplugged, fully plugged, or partially plugged condition. If a pile is assumed to be fully plugged, illogical results may be provided, especially for noncohesive soils. Since the actual perimeter of the pile is known, H piles were assumed to be unplugged for the analysis. The database was divided for H piles and pipe piles, and the GA performed 5 trials for each pile type (a total of 10 runs) to regress the variables and test results.

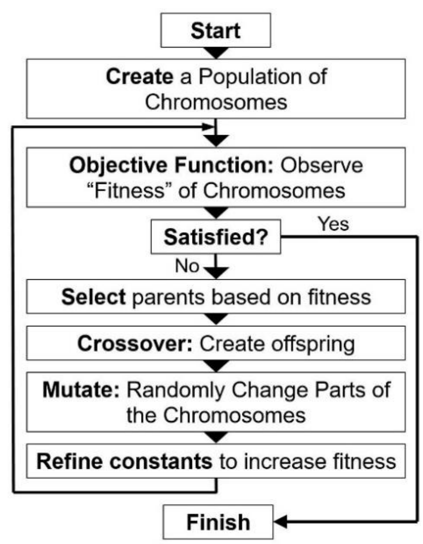

Figure 2. Process of the Genetic Algorithm

As displayed in Figure 2, the GA in this investigation evolved the chromosomes with the following steps: creation, evaluation, selection, crossover, mutation, and constant refinement. The GA was created with Matlab (Mathworks 2017) and applies the Multi Expression Programming (MEP) technique (Oltean & Dumitrescu, 2002) to encode and evaluate the chromosomes. MEP was based on the activation of programs or code with integers and can efficiently encode or decode functions compared to other techniques (Oltean & Dumitrescu, 2002). Table 3

Table 3. Settings for the Genetic Algorithm

Parameter	Parameter Setting
Number of generations	60
Population size	2000
Function set	$+, -, \times, \div$, power, exponential, logarithmic, hyperbolic tangent
Chromosome length	20
Fitness function	Mean Squared Error (MSE)
Mutation rate (%)	10
Crossover rate (%)	90
Crossover type	Uniform with brood recombination
Population size for brood crossover	4
Brood crossover rate (%)	50
Population size for brood constant refinement	50
Tournament selection size	2
Initial operator likelihood (%)	30
Initial variable likelihood (%)	40
Initial constant likelihood (%)	30

shows the settings of the GA.

3.4.2 Creating Chromosomes

For a given problem, a potential solution is represented by a chromosome. In this investigation, the chromosomes with MEP were linear entities, or arrays, and represented a function for the unit side resistance of a pile. The genes, or entries within the arrays, were divided into two components. The first part indicated the activation of a function component, such as a variable, constant, or operator. The second part links the action of the operators. Figure 3 shows an example to decode an MEP chromosome. The first row of the chromosome in the figure has negative integers to represent the operators. In this example, -1 is for addition. Positive integers designated the activation of variables and constants, which are represented in general by a and b in the figure. A constant or variable must be the first entry within the chromosome to prevent illogical errors during evaluation (Oltean & Dumitrescu, 2002). The last two rows indicate the locations, or column numbers, for the operators to be performed. During evaluation, the result of every gene is stored, and the operators are applied to the results from

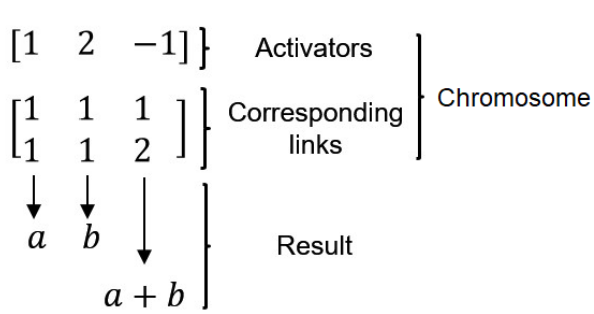

Figure 3. Example of Decoding the Chromosome

previous portions of the chromosome. Since variables and constants are numerical values, their corresponding links in the chromosome are meaningless.

The GA in this investigation was capable of simple arithmetic, but it also included power, logarithmic, and hyperbolic tangent operators because they represent nonlinear relations. However, the possible combinations for a GA to search can also increase exponentially if many operators are added (Banzhaf et al. 1998).

The population of chromosomes were initially created randomly, but a probability was assigned for the likelihood of occurrence for the operators, constants, and variables. The chromosomes were given a maximum length of 20 genes.

3.4.3 Evaluation of the Chromosomes

The goal of the GA in this study was to find the function with the best fitness to represent the unit side resistance of a pile. During evaluation, the changing shear strength from the soil conditions were considered by dividing the pile into several segments. The GA predicted the unit side resistance for each layer, and the total side resistance of a pile was found by the summation of the side resistances on the pile segments. Load-displacement responses were measured at the top of the piles during testing. Thus, the fitness function compared the predicted total side resistance (Q_{pred}) to the measured side resistance (Q_s) with the mean squared error (MSE):

$$MSE = \frac{1}{n}\sum(Q_{pred} - Q_s)^2 \qquad [3]$$

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

Generally, care is needed to ensure that illogical errors do not occur during evaluation. Examples include dividing by zero or taking the logarithm of a negative value. 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 were protected and transformed as suggested by Brameier & Banzhaf (2007).

3.4.5 Selection of the Parents

A pair of parents were selected for mating using tournament selection. For each parent, a number, or tournament size, of chromosomes were randomly sampled, and the chromosome from this group with the highest fitness became a parent. Tournament selection was repeated until the new population size matched the original size.

3.4.6 Crossover and Mutation

The activation of crossover was assigned a probability. If crossover was chosen to not occur, the parents were copied and sent for mutation. Otherwise, uniform crossover was applied with brood recombination (Figure 4). Uniform crossover randomly distributes the genes from the parents

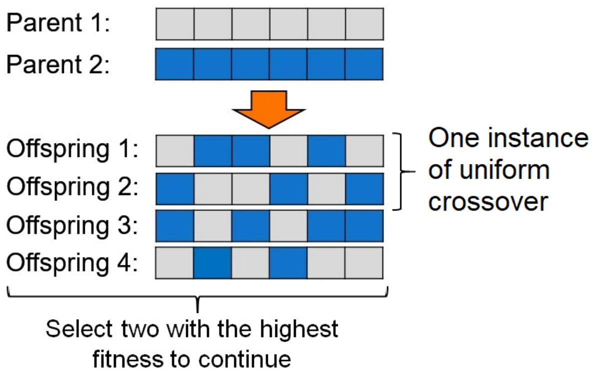

Figure 4. Example of Uniform Crossover with Brood Recombination

to the offspring. Brood recombination was inspired by organisms having a litter of offspring (Tackett 1994), and it attempts to extract the best attributes from two parents. Two parents performed crossover several times to create a subpopulation (N_s) of new chromosomes. The two offspring with the best fitness continued for mutation. Ifevery pair of parents experienced brood recombination, the total number of created offspring would be N_s multiplied by the population size, and the computational effort would be significantly increased (Banzhaf et al. 1998). Thus, the chance of brood recombination was assigned a probability.

For mutation, a selected gene would be transformed randomly into a different component type. For example, an addition operator could become a constant. The chance of mutation was set to be low at 10 %. After mutation, the resulting offspring replaced the worst chromosomes from the original population if they developed a higher fitness.

3.4.7 Constant Refinement

For symbolic regression, the constants are either evolutional or non-evolutional (Banzhaf et al. 1998). Non-evolutional constants are kept the same throughout a generation, but the GA can apply operations to manipulate the value of a constant within a function. For evolutionary constants, optimization techniques, such as the Levenberg-Marquardt algorithm (Marquardt 1963) or Nelder-Mead simplex method (Nelder and Mead 1965), can be applied. These methods are mathematically complex and usually iterative. They may take numerous trials to terminate on a potential solution, especially with several variables. Since the population of chromosomes may be large, the computational effort should be minimized. Brood recombination was applied in this GA as a simple approach to refine the values of the constants. For every chromosome, the values of the constants were randomly changed for 50 attempts. The values that provided the best fitness were kept as the new constants.

3.5 Results from the Genetic Algorithm

The GA analyzed both pipe piles and H piles separately with 2000 chromosomes for 60 generations. The plots in Figures 5 and 6 show the average and lowest MSE within the population of chromosomes for pipe and H piles. For

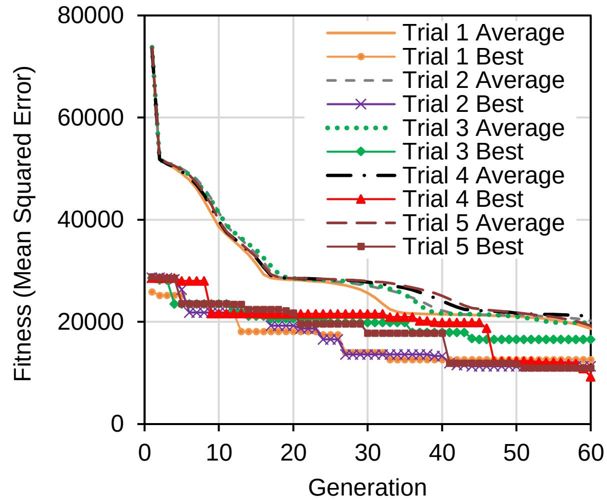

Figure 5. Fitness Performance with Pipe Piles

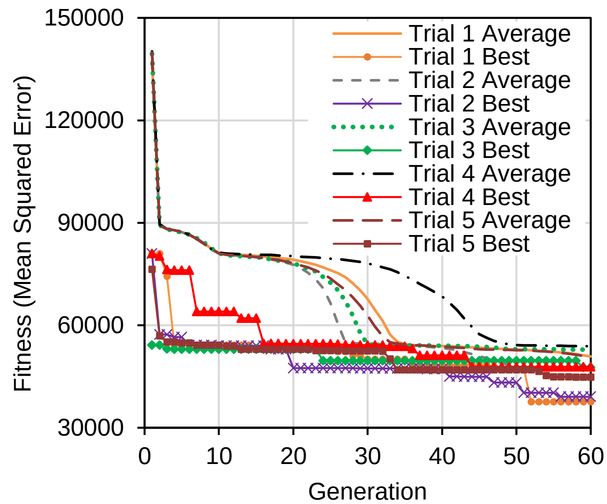

Figure 6. Fitness Performance with H Piles

each of the 5 trials, the analysis usually terminated with a similar MSE. As a better links were made, the brood recombination during crossover and constant refinement resulted in sudden drops in the best fitness throughout the generations.

At the end of the 5 trials, the function with the best fitness (BF) was collected for each pile type. The MSE by itself may seem misleading since the pipe piles had lower side resistances on average than the H piles. Yet, as shown in Figures 7 and 8, the best-fit function for pipe and H piles had a reasonably good R^2, and results were also mainly within ± 25 % of the 1:1 line. For pipe piles, the function with the lowest MSE had a R^2 of 0.73:

$$q_s = \exp(2.20/N^2)\,[8.27 + (0.71\,N)/(L-z)] \qquad [4]$$

The function with the best fit for unplugged H piles had a R^2 of 0.82:

$$q_s = \frac{[16.9(L-z)/D + 17.3]}{\tanh[71.2(L-z)^2/D^2 + 71.2(L-z)/D]} \qquad [5]$$

For pipe piles, the developed function had a predicted to measured resistance ratio (Q_{pred}/Q_s) of 1.09 on average, and it overestimated the resistance of the piles dominating in cohesive soils. The overestimation likely resulted since a limited number of piles were fully embedded in clay. In general, the GA rarely considered the soil type for H piles because noncohesive soils dominated the sites. While observing the applied variables in the final generation, functions frequently contained the slenderness ratio for H piles, and Equation 5 does not apply any other variable. This result may indicate the unreliability of SPT or the soil plugging and installation effects of H piles. Equation 5 typically underestimated the side resistance with an average Q_{pred}/Q_s of 0.94. Equations 4 and 5 may also tend to be more conservative for piles with higher resistances or longer lengths.

The function with the lowest fitness, as demonstrated

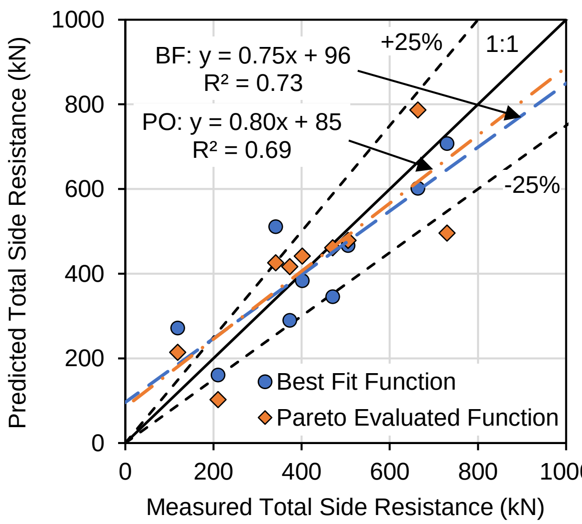

Figure 7. Comparison of Measured and Predicted Side Resistances by GA for Pipe Piles

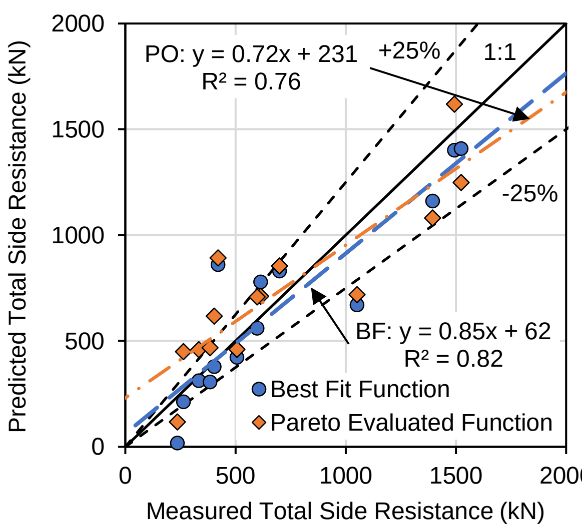

Figure 8. Comparison of Measured and Predicted Side Resistances by GA for H Piles

with Equation 5, is not always practical, beneficial, or appropriate. Another function for each pile type was then selected by Pareto optimization (PO). The results from the final generations of the 5 trials were pooled together to create a population of 10000 functions. These functions were then graphically evaluated by their fitness and complexity. The complexity is the number of components in a function, and the Pareto front was created in Figures 9 and 10 by finding the best fitness for each complexity. In general, a lower complexity, or a shorter function, results in a higher MSE, but a longer function can have several operations to create a better fitness. The orange square markers are points on the Pareto front, and the blue circles are the remaining results. Any point along the Pareto front can be a potential solution; thus, the preferred solution mainly relies on the tolerable error and judgement of the investigator (Smits & Kotanchek 2005).

For pipe piles, the improvement of the MSE is low for a complexity between 9 to 15. The function on the Pareto front with 9 components was selected since shorter functions had a significantly higher MSE. The corresponding function is below (R^2 = 0.69):

$$q_s = [(0.87\ \mathrm{N})/(\mathrm{L} - \mathrm{z}) + 12.4]/\mathrm{S} \qquad [6]$$

The Pareto front in Figure 10 was linear for H piles, and the sudden increase in fitness at a complexity of 15 may be due to the volatile nature of brood recombination. A preference was given to a function containing several variables. The selected function initially had a complexity of 9 but was simplified to the following (R^2 = 0.76):

$$q_s = \mathrm{N} \cdot \mathrm{D}/[2.86\ (L - z)] + 14.6 \qquad [7]$$

The fitness of Equations 6 and 7 is displayed in Figures 7 and 8. The functions from the Pareto optimization have a small difference in R^2 compared to the functions with the lowest MSE. Since Equation 6 included the soil type, it bears more information on the soil conditions than Equation 4, and it had a slightly better average Q_{pred}/Q_s of 1.06. Equations 7 and 5 did not include the soil type, and Equation 7 tends to overestimate compared to Equation 5 with an average Q_{pred}/Q_s of 1.16. The effective stress was

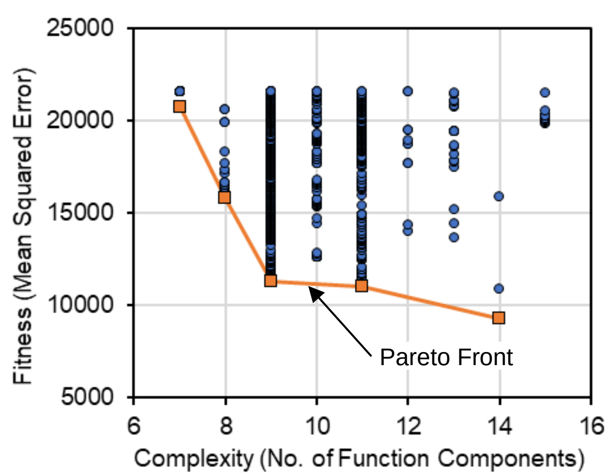

Figure 9. Pareto Front from GA results for Pipe Piles

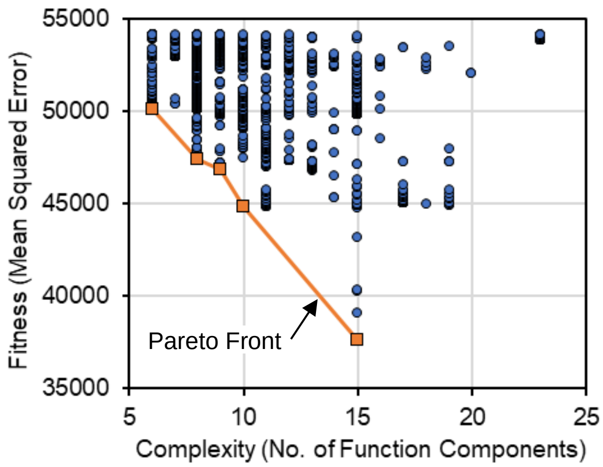

Figure 10. Pareto Front from GA results for H Piles

not included in any of the functions from the GA.

3.6 Performance of Existing Design Methods

The side resistances of the piles were calculated with design methods that were intended for both cohesive and noncohesive soils: Shioi and Fukui (1982), Decourt (1982), and Brown (2001). N-values were corrected and limited as mentioned by the references, and H piles were assumed to be fully plugged as suggested by Brown (2001). The results of the predictions are provided in Figures 11 to 13.

The three existing design methods mainly overestimated the side resistance and gave erratic results. Especially for the pipe piles, a logical linear relationship with a decent fitness could not be established between the measured and predicted values. The approach by Brown (2001) had the worst performance with an average Q_{pred}/Q_s of 2.51 and 2.97 for pipe and H piles, respectively. The method by Decourt (1982) overestimated the side resistance by 2.40 times on average for both pile types,

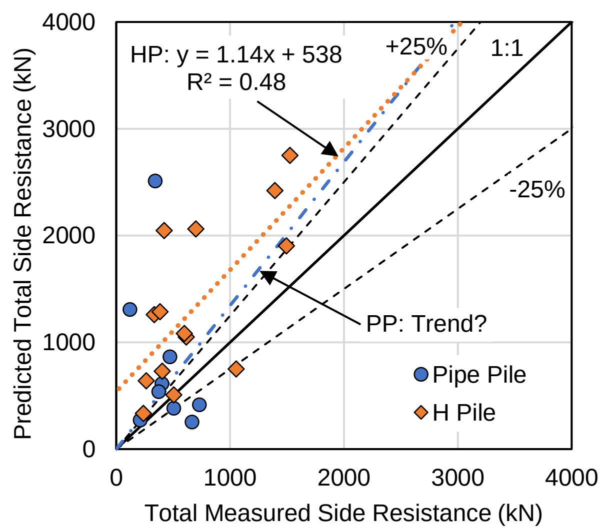

Figure 11. Comparison of Measured and Predicted Side Resistances by Shioi and Fukui (1982)

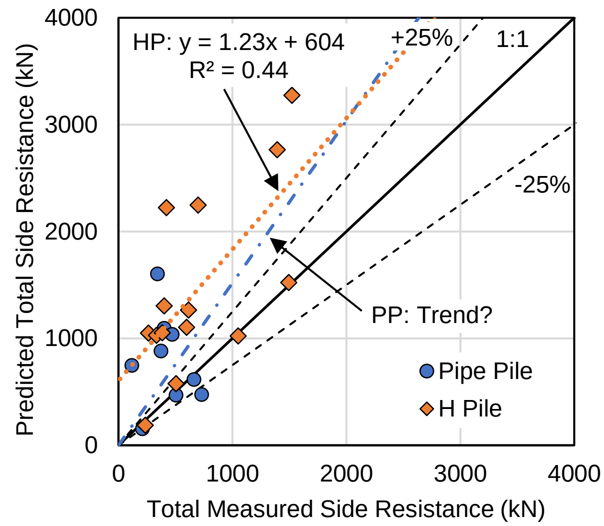

Figure 12. Comparison of Measured and Predicted Side Resistances by Decourt (1982)

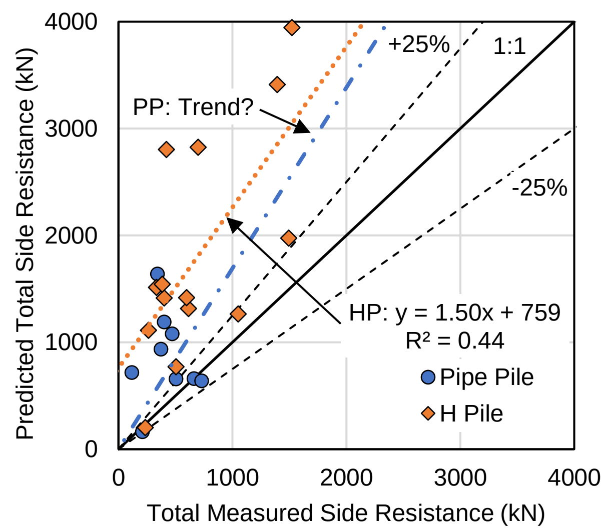

Figure 13. Comparison of Measured and Predicted Side Resistances by Brown (2001)

and it gave the best results among the existing methods. The greatest over predictions occurred with piles in clays or very stiff soils.

4 CONCLUSIONS AND DISCUSSIONS

This preliminary investigation demonstrated the capability of a simple GA to predict the side resistance of 23 piles with SPT N-values. Although a small sample size was analyzed, the GA was given more detail on the soil measurements by dividing the piles into segments. This GA was then tailored to consider heterogenous soil conditions, and the correlated functions were refined with Pareto optimization.

For both pipe and H piles, a function was initially selected from two different criteria: the best fitness and Pareto optimization. Equations 6 and 7 were determined from the Pareto evaluation, and they are recommended over the functions from the best fitness. Thus, the following equation is proposed for pipe piles (R^2 = 0.69):

$$q_s = [(0.87\ \mathrm{N})/(\mathrm{L} - \mathrm{z}) + 12.4]/\mathrm{S}\ \leq 100\ kPa \qquad [8]$$

The function below is suggested for unplugged H piles (R^2 = 0.76):

$$q_s = \mathrm{N} \cdot \mathrm{D}/[2.86\ (L - z)] + 14.6\ \leq 100\ kPa \qquad [9]$$

For these two functions, it is suggested, like Meyerhof (1976), to limit the unit side resistance to 100 kPa. From the studied piles, the measured unit side resistance did not surpass this value.

Both Equations 8 and 9 were directly proportionate to the SPT N-values and indicate a higher unit side resistance with stiffer soils. They also apply the inverse of the slenderness ratio. This variable was commonly applied by the GA, and it can indicate in the equations that the soil disturbance is lower towards the pile base. The side resistance could also be higher towards the pile base during pull-out because every pile had an over-sized base plate or reinforcement base plate. Since the sites dominated in the noncohesive soils, the common use of the slenderness ratio could also indicate the influence of the effective stress, but it is difficult to evaluate without results from fully instrumented piles.

The results from the GA were more accurate compared to the existing design methods. Yet, the existing design methods solely relied on the SPT N-values and were intended for weaker soils. Cohesive soils were the main cause of overestimation, but Equation 6 from the GA assumes cohesive soils have a lower side resistance than noncohesive soils. Shioi and Fukui (1982) received the opposite result. This investigation did not have many piles in stiff undrained clays; thus, Equations 6 and 7 may be more appropriate for noncohesive soils and drained clays.

Although the findings heavily rely on the extent of the site investigations and pile load tests, the performance of the existing methods demonstrates a need in Ontario for locally-developed design methods for the pile capacity. The GA gained practical functions with multiple variables and soil measurements along the piles. It can likely provide more accurate results with advanced soil testing, such as the cone penetration test, or data from fully instrumented piles load tests. The GA was also efficient at considering nonlinear relationships, which would be difficult to achieve with traditional statistics. In all, machine learning techniques can help address uncertainty in geotechnical engineering and offer better design methods for future infrastructure projects in Ontario.

ACKNOWLEDGEMENTS

The presented research was made possible with funding and a master's scholarship from the National Sciences and Engineering Research Council of Canada and support and resources provided by the Ministry of Transportation of Ontario. The authors would like to thank Mr. David Staseff from MTO for sharing the database of pile load tests. The authors also appreciate the help from Mr. Andy Lai, Mr. Filipe Batista, Ms. Chantel Yung, Ms. Maeeda Khan, and Ms. Maribel Castro with the collection of data.

REFERENCES

Banzhaf, W., Nordin, P., Keller, R.E., and Francone, F.D. 1998. *Genetic programming: an introduction*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Brameier, M. and Banzhaf, W. 2007. *Linear genetic programming*, Springer, New York, NY, USA.

Brown, R. P. 2001. *Predicting the ultimate axial resistance of single driven piles* (PhD dissertation), University of Texas, Austin, TX, USA.

Canadian Geotechnical Society (CGS). 2006. *Canadian Foundation Engineering Manual*, 4th ed., CGS, Richmond, BC, Canada.

Decourt, L. 1982. Predictions of bearing capacity based exclusively on N values of the SPT, *In Proceedings of the 2nd European Symposium on Penetration Testing*, Amsterdam, 1: 29-34.

Fellenius, B. H. 1980. The analysis of results from routine pile load tests, *Ground engineering*, 13(6): 19-31.

Marquardt, D.W. 1963. An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11(*2*): 431-441.

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

Meyerhof, G.G. 1956. Penetration tests and bearing capacity of cohesionless soils, *Journal of the Soil Mechanics and Foundations Division*, ASCE, 82(*1*): 1-19.

Meyerhof, G.G. 1976. Bearing capacity and settlement of pile foundations, *Journal of the Geotechnical Engineering Division*, ASCE, 102(*3*): 197-228.

Nelder, J.A. and Mead, R. 1965. A simplex method for function minimization, *The Computer Journal*, 7(*4*): 308-313.

Oltean, M. and Dumitrescu, D. 2002. *Multi Expression Programming*, Babes-Bolyai University, Romania.

Poulos, H.G., Carter, J.P., and Small, J.C. 2001. Foundations and retaining structures -research and practice, *In Proceedings of the 15th International Conference on Soil Mechanics and Geotechnical Engineering*, Istanbul, Istanbul, Turkey, 4: 2527-2606.

Shioi, Y. and Fukui, J. 1982. Application of N-value to design foundations in Japan, *In Proceedings of the 2nd European Symposium on Penetration Testing*, Amsterdam, 1: 159-164.

Smits, G.F. and Kotanchek, M. 2005. Pareto-front exploitation in symbolic regression, In O'Reilly et al. (Eds.), *Genetic programming theory and practice II*, Springer, Boston, MA, USA.

Tackett, W. A. 1994. *Recombination, selection, and the genetic construction of computer programs* (PhD dissertation), University of Southern California, Los Angeles, CA, USA.

Vesic, A.S. 1977. *Design of pile foundations, synthesis of highway practice 42*, Transportation Research Board, Washington, DC, USA.