Considerations on prediction of pile lengths using statistical modelling

Considérations sur la prévision des profondeurs de pieux en employant un modèle statistique

P.E.L.Santa Maria – *Ph.D., Associate Professor, COPPE-UFRJ*
R.M.Bastos – *M.Sc., Civil Engineer, Marine Production Systems do Brasil*

ABSTRACT: This paper presents, initially, a probability-focused methodology to estimate pile lengths based on multiple linear regression model. Then, some examples of its application are provided and discussed. Finally, the conclusion, based on such applications, is that the proposed method is a useful tool to reduce the uncertainties and for risk management in deep-foundation designs.

RÉSUMÉ: Cette étude présente, premièrement, une méthodologie probabilistique pour la prévision de profondeur des pieux, basée sur un modèle de multiple regression linéaire. Alors, quelques exemples sont montrés et discutés de leur application. Finalement, la conclusion, basée sur cettes applications, donne un méthode pour réduire le risque en projects de fondations profondes.

1 INTRODUCTION

The purpose of this paper can essentially be summarized, in fact, in two words – *reduce uncertainties* – an expression closely linked to safety and reliability.

In the field of foundations, particularly regarding piles, the designer is often posed the question on how does he decide *a priori* the length of piles in regions farther from boreholes, or how does he proceed in cases where there are few boreholes? The idea very often is to resort to extrapolations, somewhat questionable, since there is in such cases no explicit methodology to help establish the degree of reliability in such estimates.

In the case of a profile of a failure load, this is generally calculated based on the information relating to the subsoil provided by standard penetration tests (SPT), which are known to be very specific. Although it is permissible to assume a small variability in subsoil conditions around a borehole, the same cannot be said for regions farther away. The question increases in importance when discussing a bored cast-in-place pile for which no information is available, compared to those taken while monitoring the execution of driven piles – set and elastic rebound, for example.

Therefore, the purpose of this paper is to propose a rational methodology based on a statistical model to help, with a reliability level considered by the designer, establish the pile length at any point in any given region, inside the boundaries of a work site, based on failure load profiles that may be obtained from a semi-empirical method of calculating a load capacity based on SPT tests.

2 FUNDAMENTALS OF THE PROPOSED METHODOLOGY

The methodology discussed herein allows, at the design stage, establishing the pile length at any point of coordinates (x, y) in the region of a work site, taking into account the variability of the foundation ground and letting the designer establish a level of reliability for his estimates.

SPT boreholes are first located in the region of the work site. Then a convenient reference base is established from which the coordinates (x, y) of the boreholes, and for any other point in the region in question, may be established. Figure 1 gives an example of the choice of this reference base, where the shaded area represents the region of the work site while the circles represent the SPT holes.

It is possible to establish the load capacity of the foundation using a semi-empirical calculation method, based on the boring logs, with the geotechnical classification of the layers crossed by the sampler, and blow count values (N-SPT). Here the Aoki-Velloso semi-empirical method was chosen, since it is widely known, used and well accepted in the Brazilian geotechnical field.

By using this method, then, a profile of failure load or ultimate strength was reached with the depth for each borehole. The criterion of the global safety factor is adopted and, according to the provision in 5.5.1 of the Brazilian Standard for Design and Building Foundations (NBR–6122, 1996), a safety factor of 2 is applied to the value calculated for the load capacity of piles or caissons without a loading test.

The pile lengths thus defined are known to be extremely pinpointed data, since they were obtained from SPT boreholes, which are just as specific. The proposed methodology uses one of the results obtained from multiple regression model – the estimate of a new reading $Y_{h(new)}$ – which, based on the ultimate strength profiles with depth, calculated by the Aoki-Velloso method, helps establish the "z" lengths of the piles at any point on coordinates (x, y), established in function of the adopted reference base.

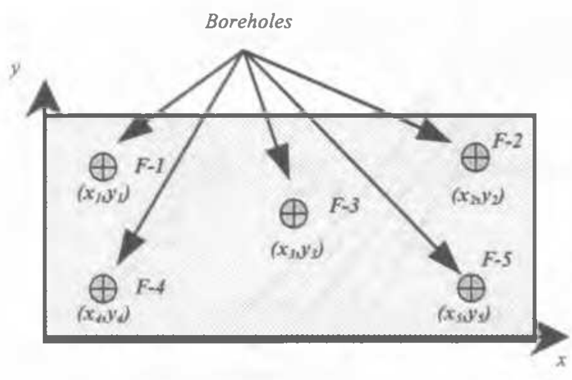

Figure 1. Illustration of choice of reference base and establishing borehole coordinates

3 CALCULATION PROCEDURES

In order to use the statistical model to estimate a new reading $Y_{h(new)}$, it is necessary to first find the regression function linking the variables in question. The idea is to use the general multiple linear regression model to establish the regression function between the dependent variable Q_f, the failure load being obtained by the Aoki-Velloso semi-empirical method and the independent variables x, y and z, where x and y are the borehole coordinates and z is the depth. Let it be known that:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1}$$

therefore, considering the case under study, it may be written as:

$$E(Q_f) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 z \quad (1)$$

which is the regression function, since there are three independent variables involved: x, y and z.

The estimator of the mean response, corresponding to the values of the independent variables $x = x_h$, $y = y_h$ and $z = z_h$, is given by:

$$\hat{Q}_{f(h)} = b_0 + b_1 x_h + b_2 y_h + b_3 z_h \quad (2)$$

where b_0, b_1 and b_2 represent the estimators of regression parameters β_0, β_1 and β_2.

The estimate interval is now considered for a new reading $Y_{h(new)}$ corresponding to the values of the independent variables $X_1, \ldots, X_{p-1}$ equal to $X_{h1}, \ldots, X_{h,p-1}$, represented by the vector $\mathbf{X}_h$. It may, then, be written as:

$$Y_{h(new)} \geq \hat{Y}_h - t(1-\alpha/2;\, n-p)\, s(Y_{h(new)})$$

$$Y_{h(new)} \leq \hat{Y}_h + t(1-\alpha/2;\, n-p)\, s(Y_{h(new)})$$

where:

$$s^2(Y_{h(new)}) = MSE(1 + X'_h (X'X)^{-1} X_h)$$

and t $(1-\alpha/2;\ n-p)$ represents the value of the varialble "t" from t-Student function with parameters $(1-\alpha/2;\ n-p)$.

In this case, the estimate interval for a new reading $Q_{f(h(new))}$ corresponding to the values of the independent variables x, y and z equal to x_h, y_h and z_h, respectively, may be written as follows:

$$Q_{f(h(new))} \geq \hat{Q}_{f(h)} - t(1-\alpha/2;\, n-p)\, s(Q_{f(h(new))})$$

$$Q_{f(h(new))} \leq \hat{Q}_{f(h)} + t(1-\alpha/2;\, n-p)\, s(Q_{f(h(new))}) \quad (3)$$

where:

$$s^2(Q_{f(h(new))}) = MSE(1 + X'_h (X'X)^{-1} X_h) \quad (4)$$

The vector $\mathbf{X}_h$, representing the group of values of the independent variables for which an estimate is desired for a new reading $Q_{f(h(new))}$, is given by:

$$X_h = \begin{bmatrix} 1 \\ x_h \\ y_h \\ z_h \end{bmatrix} \quad (5)$$

The X matrix is defined as:

$$\underset{n\times 4}{X} = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & z_n \end{bmatrix} \quad (6)$$

X'_h and X' are, thus, the transposed values of X_h and X, respectivelly.

The product X'X is given by:

$$\underset{4\times 4}{X'X} = \begin{bmatrix} n & \sum x_i & \sum y_i & \sum z_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum y_i & \sum y_i x_i & \sum y_i^2 & \sum y_i z_i \\ \sum z_i & \sum z_i x_i & \sum z_i y_i & \sum z_i^2 \end{bmatrix} \quad (7)$$

In fact, the purpose of the method is to establish length z_h of the piles based on the coordinates (x_h, y_h) of the point where it is desirable to obtain such an estimate and failure load $Q_{f(h)}$. It is not, in fact, desirable to establish an estimate interval for a new reading associated to a reliability level $(1-\alpha)$, but only to use the lower limit of this interval to obtain the depth z leading to a value of $Q_{f(h(new))}$ equal to the failure load $Q_{f(h)}$. The value of the depth, thus established, is associated to a reliability level, that is, that the probability of the failure load, $Q_{f(h)}$, being obtained at greater depths than that established through this procedure is $(\alpha/2)$.

Figure 2 helps clarify the process. It must be remembered, however, that the curves that represent, on plane $Q_f - z$ shown in the figure, the lower and upper limits of the estimate interval, are merely a sketch to facilitate understanding. The estimate intervals, described so far, are drawn up on specific points, that is, for a single set of values of independent variables called vector $\mathbf{X}_h$. Note that the broken lines in the sketch represent the upper and lower limits of the reliability interval of the mean response $E(Q_{f(h)})$, and the unbroken lines in the sketch are the upper and lower limits of the estimate interval of the new reading $Q_{f(h(new))}$.

Consider that coordinates (x_h, y_h) of the point where the depth z_h is to be calculated, are fixed. Figure 2 shows that this is the depth leading to $Q_{f(h(new))} = Q_{f(h)}$ = lower limit of the estimate interval, for those fixed coordinates (x_h, y_h), $Q_{f(h)}$ being the failure load of the foundation. It therefore follows that the depth z_h is obtained directly from solving the following equation:

$$\hat{Q}_{f(h)} - t(1-\alpha/2;\, n-p)\, s(Q_{f(h(new))}) = Q_{f(h)} \quad (8)$$

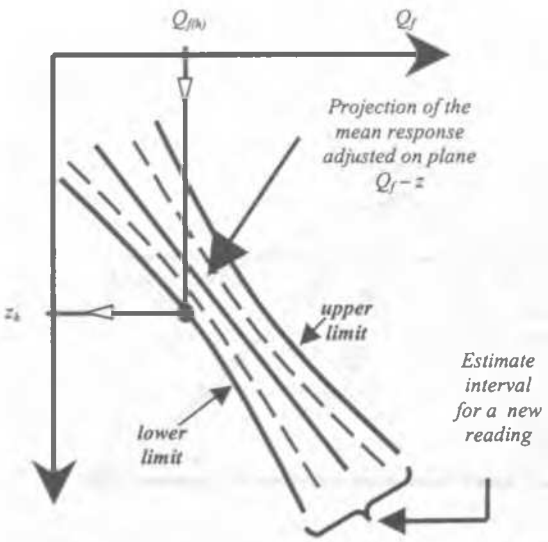

Figure 2. Establishing the depth z_h using the proposed methodology

4 APPLICATIONS OF THE METHODOLOGY

Some examples are given below of the proposed methodology. Fourteen works built in two neighborhoods of the city of Rio de Janeiro, Brazil – Botafogo and Tijuca – were chosen from an available database, each with its respective SPT reports over a total of eighty-four holes.

The coordinates (x_h, y_h) were first established from a convenient reference base of the points corresponding to the boreholes available in each of the fourteen work sites.

Once the coordinates (x_h, y_h) of the points where the depths z_h will be calculated are established, then the ultimate strength profiles with the depth are calculated, based on the boring logs, using the Aoki-Velloso method. For the examples herein, it was decided to choose a bored cast-in-place pile with the following characteristics:

- diameter: 700 mm
- area: 0.3848 m^2
- perimeter: 2.20 m
- load on pile, for compressive stress on the concrete equal to 30 kgf/cm^2: 1150 kN (115 tf)

In fact, a global safety coefficient of 2 was applied to the value calculated for the load capacity using the Aoki-Velloso semi-empirical method in the examples herein in order to obtain the working load.

The next step consists of using the failure load capacity profiles with depth, for each SPT hole, and with coordinates (x_i, y_i) of these holes, establish a regression function of the kind presented in (1), as well as the estimator of the mean response and the mean square error.

Once the mean adjusted responses and the values of the Mean Standard Error were obtained for each project, then the depth z_h is calculated at the points of coordinates (x_h, y_h). Thus, depths z_h were established in each of the fourteen work sites, at the points whose coordinates (x_h, y_h) are the same as those of the SPT holes. In fact, any other points could have been selected in the area covering each project. This choice was made only to facilitate the comparison between the values of the depths calculated using the "statistical method", taking into consideration the variability of the information relating to the site, and the values obtained merely by directly using the Aoki-Velloso method.

The results obtained for each project are first shown in the table 1 below: the regression function, $\hat{Q}_f = b_0 + b_1 x + b_2 y + b_3 z$, the correlation coefficient, *r*, and the mean standard error, *MSE*. It must be mentioned that the reliability level associated with depth estimates, based on the proposed methodology, was chosen as: $(1-\alpha) = 95\%$.

Figure 3 presents a comparative plot which shows all results obtained. The pile lengths obtained from the Aoki-Velloso method are marked on the abscissa axis, and those calculated using the proposed methodology on the ordinate axis.

The broken line in figure 3 establishes these abscissa values that are equal to the ordinate values and has a slope of 45°.

It is found that, in every case studied, the depths established using the statistical method are greater than those established by the Aoki-Velloso method, at the points corresponding to the holes. This was expected, since the statistical method, unlike the Aoki-Velloso method, considers the variability of the ultimate strength values along the lengths obtained in the various holes. The more "uniform" the results are, or the smaller their variability, the closer the values of the lengths established by both methods. It is found that the unbroken line, in figure 3, is merely the adjusted value of the mean response to a simple regression of *Y* in *X*, where *Y* is the length obtained by the statistical method and *X* the length established by the Aoki-Velloso method. It is noted, however, that the passage of the line adjusted by the origin was forced, and with this, the following equation for the regression function was obtained:

$$\hat{Y} = 1.3455\,X \qquad (9)$$

Table 1. Results obtained for each project

WORK SITE	REGRESSION FUNCTION	R	MSE
Alexandre de Gusmão 9	$\hat{Q}_f = 10.8+4.4x+1.3y+17.1z$	0.87	1317.6
Alfredo Pinto 58	$\hat{Q}_f = 76.5+2.2x-0.9y+19.6z$	0.67	9968.2
Alzira Brandão 349	$\hat{Q}_f = 64.9+0.2x+0.3y+14.5z$	0.51	19725.9
Antônio Basílio 409	$\hat{Q}_f = -370.1-13.4x-0.5y+71.2z$	0.56	26607.4
Antônio Basílio 613	$\hat{Q}_f = -652.6+80.7x-48.8y+87.7z$	0.51	23262.9
Araújo Pena 15	$\hat{Q}_f = 195.9-1.5x-7.1y+12.1z$	0.59	10641.6
Bambina 60	$\hat{Q}_f = -1131.9+52.3x+17.8y+17.2z$	0.83	5839.5
Dona Mariana 138	$\hat{Q}_f = -270.8-1.3x+2.0y+15.3z$	0.64	8885.8
Maria Amália 101	$\hat{Q}_f = -14.1-7.7x-2.3y+36.9z$	0.52	31210.2
Mariz e Barros 1001	$\hat{Q}_f = 31.2+3.3x+0.7y+13.0z$	0.55	7440.8
Professor Gabizo 107	$\hat{Q}_f = 54.1-15.2x-4.6y+56.3z$	0.63	10017.2
Professor Gabizo 161	$\hat{Q}_f = -117.4+26.7x-3.0y+23.3z$	0.80	5009.8
Santa Sofia 134	$\hat{Q}_f = 40.4+0.8x-1.1y+18.1z$	0.74	4897.0
Visconde de Ouro Preto 39	$\hat{Q}_f = -283.6-2.0x-0.7y+17.5z$	0.78	7419.5

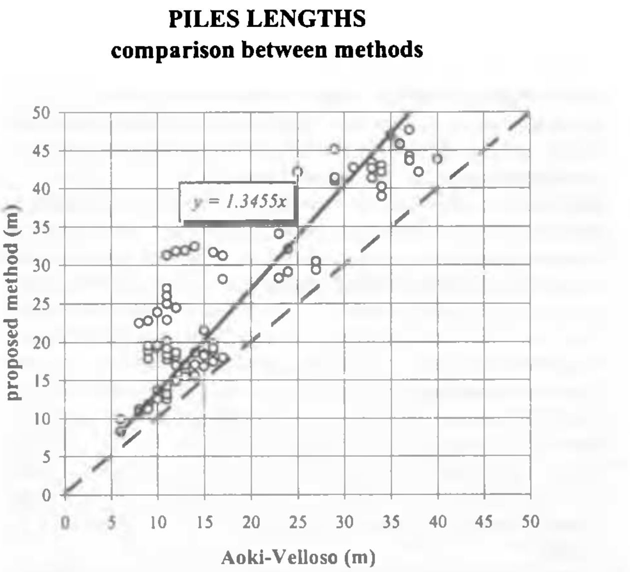

Figure 3. Comparison between the pile lengths obtained using both methods.

As the statistical method takes into consideration the variability of the results, its application always leads to greater length values than those established by the Aoki-Velloso method. It is known that, as mentioned above, a safety factor of 2 was applied to the values of the failure load capacity. Therefore, from the way the calculations were made, the length calculated based on the proposed methodology is affected both by the safety factor applied to the failure load and by the effect from the consideration of the variability of results, which is inherent to the actual statistical method. It is suggested, then, that a reduction factor be applied to the lengths calculated by this method, so that the mean adjusted line, shown in figure 3, coincides with the line sloping at 45°. This procedure would help attenuate the effect caused by the safety factor 2, which has been conceived to be used in calculation methods in which the variability is not considered. Therefore, some lengths calculated by the statistical method would be less than those calculated by the Aoki-Velloso method. This means that, in cases where the results are more "uniform" or the variability is less, the lengths calculated by the statistical method would be less than those calculated only by directly using the semi-empirical method. Therefore, the following must be considered for the adjusted line to coincide with the line sloping at 45°:

- by using the expression (9), the tangent of the sloping angle δ of the adjusted line is easily obtained:

$$tg(\delta) = \frac{\hat{Y}}{X} = 1.3455 \tag{10}$$

- in order that the adjusted line has the same slope as the line at 45°, only multiply the value of the tangent, obtained from (10), by the reduction factor (*fr*) given by:

$$fr = \frac{tg(45^\circ)}{tg(\delta)} = \frac{1}{1.3455} \approx 0.75 \tag{11}$$

So, the reduction factor to be applied to the lengths calculated by the statistical method, when adopting this suggestion, is equal to 0.75. It should be borne in mind that this is merely a suggestion, based on the analyses made for one adopted kind of pile and level of reliability, and for the 87 holes available. It is important to understand this procedure as a possible "direction" to be followed.

5 CONCLUSIONS

1. It was found that, in all cases in this study, the pile lengths obtained using the proposed methodology were greater than those established by only the direct use of the Aoki-Velloso semi-empirical method. This is due to the probabilistic approach inherent to this methodology, which, *ipso facto*, considers the variability of the ultimate strength values along the depth, calculated for the various holes.
2. Bearing in mind that a global safety factor 2 was applied to the failure load capacity values established by the Aoki-Velloso method, it is concluded that, in addition to including the effect of the variability of the results, the pile length calculated using the proposed methodology is also affected by that safety factor. It was then suggested that a reduction factor be applied to the lengths calculated with basis on the proposed methodology, in order to attenuate the influence of the latter, which was designed for use in deterministic calculation methods. Therefore, the value of 0.75 was reached for the length reduction factor, based on the examples studied herein. It is important to emphasize, however, that this is only a suggested procedure, based on a function of the studies made herein.
3. Although the methodology proposed here is still in the early stages of analysis, based on statistical concepts, it has proven to be a useful tool to estimate pile lengths at any point in a given region where works are planned. The steps taken in preparing the design are practically the same as normally taken in foundation engineering practice, except for just one more simple calculation sequence. Another relevant aspect is the possibility of associating the length estimate at a reliability level, and facilitating risk management by the designer. It may therefore be said that this methodology is a step towards reducing the uncertainties involved in deep foundation design.

ACKNOWLEDGEMENTS

The work was carried out at Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE/UFRJ) under a grant from Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors are grateful to Flávia C.M. de Santa Maria, Thais M.G. de Siqueira, Roberto C.D. d'Oliveira for having kindly prepared and revised the final layout of this paper and also to L.B. de Santa Maria, Adriana Bandeira dos Santos and Rosa Marques Bastos for the relevant suggestions and continuous support.

REFERENCES

ABNT, 1996. *Foundation Design and Execution*, NBR-6122, Rio de Janeiro.

BASTOS, R. M. 2000. *Considerations on Prediction of Pile Lenghts Using Statistical Modelling* . M.Sc. dissertation, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

BECKER, D. E. 1996, "Eighteenth Canadian Geotechnical Colloquium: Limit States Design for Foundations. Part I. An Overview of the Foundation Design Process", *Canadian Geotechnical Journal*, vol. 33, pp. 956-983.

NETER, J., WASSERMAN, W. 1974. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, 1st ed., Homewood, Richard D. Irwin, Inc.