# STATISTICAL ANALYSIS OF THE REPRODUCIBILITY OF THE STANDARD PENETRATION TEST: A CASE STUDY

**[1]VINICIUS R DOMINGUES, [2]BERNARDO C P E ALBUQUERQUE, [3]DARYM J FERRARI DE CAMPOS, [4]LUAN C S M OZELIM**

Consultant of PiezoGreen Solutions, Brazil
E-mail: piezogreen@gmail.com

**Abstract—** One of the most widespread survey methods of geotechnical investigation is the Standard Penetration Test - SPT. Although this test is regulated by standards in several countries, even the standardized methodology may lead to highly variable data. In this case, the data variability is not caused by sampling different stratigraphic horizons; it is caused by disturbances of three other sources, namely: human, equipment and procedure. This research, based on a case study and in a statistical analysis, isolate the influence of soil properties to clarify the relevance of these other variability sources. The analysis, based on Bootstrap resampling method, considers 8 SPT surveys executed by the same workers using the same equipment and on a pre-established soil profile. As part of the paper methodology, each blow was monitored for its height and impact velocity with a high-speed camera and a topographical target. It was possible to note that, despite the standardizations described, the data are significantly different. In the sample studied, the energy transmitted to the soil is statistically higher than the Standardized one. Consequently, parameters of soil resistance might be underestimated and, thus, engineering solutions might become more expensive than necessary. Therefore, the results show that the use of SPT may not be sufficient, depending on the risk to be assumed by stakeholders.

**Index Terms**- Bootstrap, Reproducibility of SPT, Risk in engineering, Statistical Analysis.

## I. INTRODUCTION

The knowledge of the characteristics and geotechnical parameters of soils is fundamental, leading to greater safety and efficiency in the optimization of geotechnical designs.

Currently, the most widespread type of geotechnical investigation in Brazil consists in surveys of SPT (Standard Penetration Test), standardized by a Brazilian code NBR 6484 [1]. Due to its massive use, it is considered an indispensable instrument in preliminary investigations for the foundations design. One must pay close attention to its proper interpretation and to factors that interfere its results. These factor are, mainly, the height of fall of the SPT hammer and the energy losses energy during the process of transformation of potential to kinetic energy.

The test reveals essential data, such as soil type, groundwater level and soil resistance. Due to its simplicity, robustness and fast response time, it seems reasonable to look for means that allow a more reliable evaluation of its performance and efficiency factors through a standardized procedure [2].

Following this reasoning, statistics has proved to be an analysis tool with great relevance, especially when deviations from a given procedure need to be modelled. This is the main objective of the present research: statistically model inconsistencies of the SPT results.

## II. STANDARD PENETRATION TEST

### A. The test and its efficiency

The SPT test consists in penetrating a standard sampler of 45 cm in the soil, by continuously dropping a hammer from a height of 75 cm, whose wheight is of 0,64 kN. The number of blows is recorded for each 15 cm of penetration. The penetration resistance is defined by the number of blows required for the last 30 cm of penetration ($N_{SPT}$). Thus, the test allows the evaluation of the compactness of the granular soils and the consistency of cohesive soils and even of soft rocks.

Like any ground or laboratory test, SPT is influenced by several factors, which affects its efficiency. In most cases, these factors can be classified as being results of three entities: human, equipment and procedure.

According to [3] the basic elements associated to the equipment are: hammer, rods, hole coating and sampler.

In the case of the hammer, the main factors that influence the results are the weight and the height of fall. In the hammer lifting system with manual override (the most used), the operators considerably influence the results, because if they raise the hammer more than necessary, there will probably be a greater transmission of kinetic energy to the rods, reducing the NSPT. Otherwise, the $N_{SPT}$ will increase.

For rods, the most important factors are their type and the length of the composition. For the hole coating,

influence is observed according to the ratio of the internal and external diameter of the sampler.
The human factor needs a lot of attention, as it is combined to technical procedures. It is assumed that the whole process is extremely sensitive to the operator's experience and skill, especially when the automation of the system used in the polls is precarious.
In the present paper, the influence of the operator on the drop height of the hammer shall be studied. In order to do so, the Bootstrap method was chosen as the statistical tool.

## III. STATISTICAL TOOLS – BOOTSTRAP

During engineering practice, statistical methods are not constantly used to interpret geotechnical investigations, thus it can be said that this science has still been treated as deterministic.
In the present research, a Bootstrap resampling method, introduced by [4,5], is used. The method, based on available data, allows the quantification of uncertainties by calculating the standard errors and confidence intervals, as well as performing hypothesis tests of interest.
By considering that the sample is representative of the population from which the former has been drawn and that the observations are independent and identically distributed, the Bootstrap method is capable of estimating the sampling distribution of a given statistic (for example the mean and variance of the population).
Besides, Bootstrap sampling is equivalent to sampling with replacement from the empirical probability distribution function.

### B. Confidence Interval (CI)

By using the bootstrap method we can obtain approximate confidence intervals of 1-2α of probability coverage. There are different ways of estimating the CI. For further information one may refer to [6]. In the case of the present paper, the bias-corrected and accelerated (BCa) method shall be used.

### C. Hypothesis Testing

In tests of hypothesis involving two datasets (X and Y), there is a possibility that both are considered as being drawn from the same distribution or not. Let one consider the test where the null hypothesis is that X and Y are equal, with the same distribution and, on the other hand, the alternative hypothesis is that X and Y are drawn from different distributions.
Since the intention is to know whether there is equality of distributions, one can calculate a bilateral test by assuming, for example, that the test statistics is the difference between means, no matter whether it is negative or positive.
With the p-value calculated and a certain level of confidence established, one can accept the null hypothesis or not. In [6] onemay find details on how to calculate this p-value by means of a Bootstrap algorithm.

## IV. RESULTS AND ANALYSIS

In order to analyze the reproducibility of the Standard Penetration Test, the data from [7] was taken, where the authors executed 8 tests in a soil with a stratigraphic profile of soft clay (7.5 to 8m) followed by smaller layers of clay, sand, silt and clay again. It is important to state that the tests were executed by the same team of workers (one experienced driller and three assistants) using a manual equipment. For all the 8 sequences, each blow was monitored for its height and impact velocity with a high-speed camera and a topographical target [7].

To be able to compare the tests, and assuming the heights may not follow a normal distribution, an statistical analysis based on the Bootstrap method was carried out. For that analysis the null hypothesis was that all the tests follow the same distribution.
The first step to analyze the reproducibility of the Standard Penetration Test was to group all the drop heights in the same sample (assuming the null hypothesis is true) and the mean and its confidence interval (CI) was calculated. After that, the CI for the mean drop height of each sequence was determined individually. The number of blows in each test as well as its mean's Confidence Interval can be seen in Table 1.

**Table 1 - Quantity of blows, mean drop height and confidence intervals for each test**

| Sequence | Blow Counts | Mean Drop Height (cm) | Confidence Interval (cm) |
|---|---|---|---|
| S1 | 4 | 87 | 80 – 92 |
| S2 | 2 | 71 | 66 – 71 |
| S3 | 7 | 77 | 73 – 81 |
| S4 | 15 | 87 | 84 – 90 |
| S5 | 3 | 69 | 65 – 71 |
| S6 | 2 | 68 | 66 – 68 |
| S7 | 17 | 69 | 68 – 71 |
| S8 | 19 | 83 | 80 – 86 |
| Total | 69 | 79 | 77 – 81 |

Figure 1 shows the histogram of the Probability Distribution Function (PDF) of the drop heights and its mean's confidence interval.

As it can be seen from Table 1 and Figure 1, the confidence interval for mean of the sample lies between 77 and 81 cm, which cannot be considered in accordance to Brazilian Standards, which states that

the drop height must be of 75 cm. In practice, this would lead to a conservative and more expensive design, since a greater height means that a smaller number of blows is needed to penetrate the same distance.

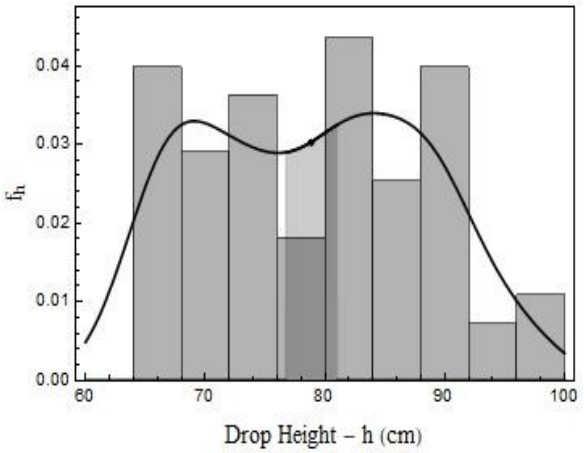



**Figure 1. Histogram of the probability distribution and Confidence Interval of the mean (shaded) of the Drop Height.**

Besides, Table 1 shows that there might be 2 different distributions, the first one of the S2, S5, S6 and S7 and the second of the S1, S4 and S8. This can also be viewed by the bimodal shape of the histogram in Figure 1. It should be stated that, besides the statistical relevance of tests S1, S2, S5 and S6 might be questionable due to the small number of blow counts, these results are already a sign that, even with the same team of workers, equipment and stratigraphic profile, the test may not be reproductible.

Figure 2 shows the histogram of the Cumulative Distribution Function (CDF) of the drop heights. The almost constant 45 degree slope of the CDF shows that one should expect a high standard deviation for the sample or even a constant distribution. This would lead to conclude that the manual execution of Standard Penetration Test is not trustworthy once every drop height can be achieved with equal probability.

Lastly, the samples' distributions were estimated with the SmoothKernelDistribution function of the software Mathematica [8]. Figure 3 presents the estimated densities for each of the 8 sequences of SPT and their means' confidence intervals. A pairwise comparison of the tests was performed, assuming all of them belong to the same distribution (null hypothesis). The p-Values for each comparison can be seen in Table 2.

**Table 2. The p-values of the Hypothesis tests**

| Test | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 |
|---|---|---|---|---|---|---|---|---|
| S1 | 1,00 | 0,05 | 0,04 | 0,92 | 0,02 | 0,03 | 0,00 | 0,30 |
| S2 | 0,05 | 1,00 | 0,16 | 0,01 | 0,54 | 0,42 | 0,50 | 0,02 |
| S3 | 0,04 | 0,15 | 1,00 | 0,00 | 0,04 | 0,05 | 0,00 | 0,03 |
| S4 | 0,93 | 0,01 | 0,00 | 1,00 | 0,00 | 0,00 | 0,00 | 0,08 |
| S5 | 0,02 | 0,55 | 0,04 | 0,00 | 1,00 | 0,88 | 0,66 | 0,00 |
| S6 | 0,03 | 0,42 | 0,05 | 0,00 | 0,88 | 1,00 | 0,54 | 0,01 |
| S7 | 0,00 | 0,50 | 0,00 | 0,00 | 0,66 | 0,55 | 1,00 | 0,00 |
| S8 | 0,30 | 0,02 | 0,04 | 0,07 | 0,00 | 0,01 | 0,00 | 1,00 |

As tests S1, S2, S5 and S6 have little blow counts, their statistical significance is weak, therefore pairwise comparisons with those samples are questionable and may not be accurate. For that reason, all the comparisons with those samples are shaded in Table 2.

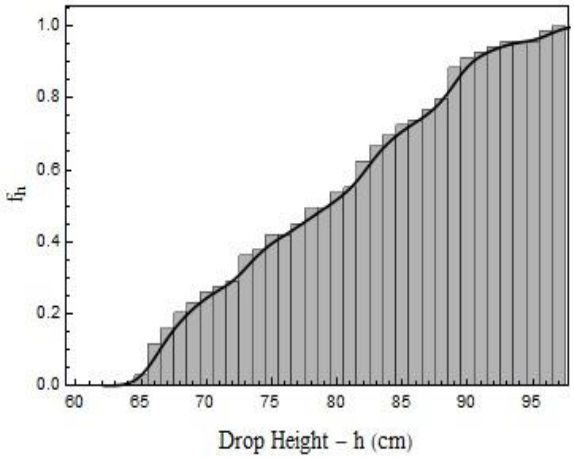



**Figure 2. Cumulative Distribution of the Drop Height.**

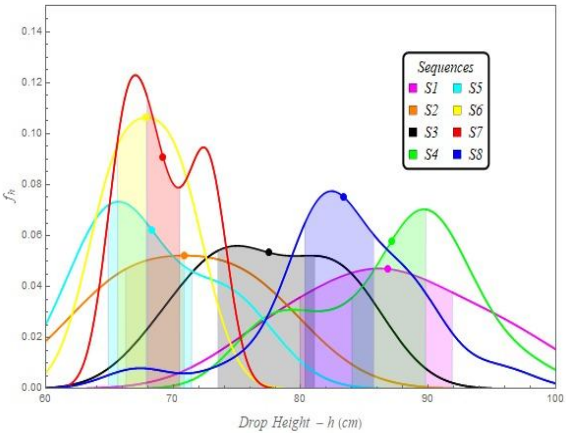



**Figure 3. Samples distributions estimated with the smooth kernel technique and their means' confidence intervals (shaded).**

For the other samples (S3, S4, S7, and S8) it can be seen that solely tests S4 and S8 may belong to the same distribution, with small p-values. Again, this raises questions about the reproducibility of the Standard Penetration Test since all the samples were executed by the same team, with the same equipment, in the same stratigraphic profile.

It should be emphasized, however, that the samples used in this study are not big and were executed by just one team. A more robust study comparing several teams and mechanical equipment should be carried on in different stratigraphic profiles with more samples per team in order to be able to state the reproducibility of Standard Penetration Test.

## CONCLUSION

A case study, presented in this paper, showed the errors intrinsic to the SPT methodology. Eight sequences of test which were executed by the same workers with the same equipment on a specific soil profile. It was noted that, despite the standardizations used in the test, the data were significantly different. The sample analyzed showed that the energy transmitted to the soil is statistically higher than the Brazilian Standard suggests. Consequently, parameters of soil resistance might be underestimated and, thus, engineering solutions might become more expensive than necessary. Therefore, the use of SPT may not be sufficient, depending on the risk to be assumed by stakeholders.

## REFERENCES

[1] ABNT-Brazilian Association of Technical Codes. "NBR 6484: Soil - Simple recognition probes with SPT - Test method." Rio de Janeiro. RJ, 2001.

[2] H. B. Seed, K. Tokimatsu, L.F. Hader, and R.M. Chung, "Influence of SPT Procedures in Soil Liquefation Resultance Evaluations". Journal of Geotechnical Engineering. ASCE. Vol. 111, n. 12. 1985.

[3] A. Palacios, "Theory and Measurements of Energy Transfer During Standard Penetration Test Sampling". (Postdoctoral thesis). University of Florida,Gainesville, USA. 1977.

[4] B. Efron, "Bootstrap method: another look at the Jackknife". The Analysis of Statistics 7 (1). 1-26, 1979

[5] B. Efron, "The Jackknife, The Bootstrap and Other Resampling Plans". Society for Industrial and Applied Mathematics CBMS-NSF Monographs 38. USA, 1982.

[6] B. Efron & R. Tibshirani, "An Introduction to the Bootstrap." Champman & Hall, United Kingdom, 436p., 1993.

[7] E. S. Silva, "Medição de Altura de Queda e Velocidade de Impacto do Martelo no SPT" (undergraduation's thesis). Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, 2015.

[8] WRI "Mathematica, Version 11.2", Wolfram Research Inc.,Champaign, IL., 2018.

★★★