Proceedings of the XVIII ECSMGE 2024
GEOTECHNICAL ENGINEERING CHALLENGES
TO MEET CURRENT AND EMERGING NEEDS OF SOCIETY

ISBN 978-1-032-54816-6
DOI 10.1201/9781003431749-285

Geotechnical parameters estimation in iron ore tailings piles via Bayesian models

Estimation des paramètres géotechniques dans les haldes de résidus de minerai de fer via des modèles Bayésiens

B.M. de Lacerda*
Simplelab Tecnologia, Minas Gerais, Brasil

R. Chammas
Chammas Engenharia, Minas Gerais, Brasil

T.C.O. Fonseca
UFRJ, Rio de Janeiro, Brasil

T.B. dos Santos
UFOP, Minas Gerais, Brasil

**bianca.lacerda@simplelabtech.com.br*

ABSTRACT: As a result of the several events involving dams around the world, the construction of upstream dams was banned, new regulations have been established, and revisions to waste disposal norms have been made. New technologies were developed for the tailings treatment, allowing alternatives for their disposal, such as the construction of different stacking methodologies. Several measures are taken in order to assure the structures safety, such as carrying out geotechnical tests throughout the construction and technological control, which ensure that the project technical specifications are complied. Some tests must be carried out in the laboratory and require a long execution time, reducing the information number of materials physical and geomechanical parameters. In this way, important geotechnical parameters for the complete understanding of the behaviour of the tailings are restricted to a few points of the structures, and with deterministic results. Despite to work with a distribution of samples that represent the structure in a significant way, as they may be subjected to different levels of stresses, humidity, granulometry, among other conditions, using values from specific locations to assure the safety of the whole construction can lead to erroneous estimates in unsampled locations. In this article, data from piles of different locations in MG/Brazil extracted of Geolabor, an ecosystem of applications for managing geotechnical data, were used to estimate different geotechnical parameters from the technological control at non-tested points. Multiple spatial and non-spatial regressions were implemented, and the results obtained were satisfactory, showing the importance of the methodology developed and implemented with Geolabor.

RÉSUMÉ: En raison des nombreux événements impliquant des barrages dans le monde entier, la construction de barrages en amont a été interdite, de nouvelles réglementations ont été établies et des révisions des normes de traitement des déchets ont été effectuées. De nouvelles technologies ont été développées pour le traitement des résidus miniers, permettant des alternatives à leur élimination, telles que la construction de différentes méthodologies d'empilement. Plusieurs mesures sont prises pour assurer la sécurité des structures, telles que la réalisation de tests géotechniques tout au long de la construction et le contrôle technologique, qui garantissent que les spécifications techniques du projet sont respectées. Certains tests doivent être effectués en laboratoire et nécessitent un temps d'exécution long, réduisant le nombre d'informations sur les paramètres physiques et géomécaniques des matériaux. De cette manière, d'importants paramètres géotechniques pour la compréhension complète du comportement des résidus miniers sont restreints à quelques points des structures, et avec des résultats déterministes. Malgré le travail avec une distribution d'échantillons représentant la structure de manière significative, étant donné qu'ils peuvent être soumis à différents niveaux de contraintes, d'humidité, de granulométrie, entre autres conditions, l'utilisation de valeurs provenant de points spécifiques pour garantir la sécurité de toute la construction peut conduire à des estimations erronées dans les emplacements non échantillonnés. Dans cet article, des données extraites de Geolabor, un écosystème d'applications pour la gestion des données géotechniques, provenant de tas de différentes localités de MG/Brésil, ont été utilisées pour estimer différents paramètres géotechniques à partir du contrôle technologique aux points non testés. Des régressions spatiales et non spatiales multiples ont été mises en œuvre, et les résultats obtenus étaient satisfaisants, montrant l'importance de la méthodologie développée et mise en œuvre avec Geolabor.

Keywords: Piles and Dams; technological control; geostatistics; statistical geotechnics; parameter prediction.

1 INTRODUCTION

In soil study, there are uncertainties in the calculations of the physical and geomechanical materials properties, possibly associated with the limited number of samples in the structure, the tests carried out and the geological field conditions (Flores, 2008). Statistical studies in mining geotechnics have as main objective to considere these uncertainties, searching for results that are more coherent with reality. The practical applications of the statistical analyzes in mining geotechnics have been growing and gaining prominence in recent years, but they still constitute a study with a restricted approach (Ferreira, 2022).

Literature authors, such as Azevedo (2019), Badaró (2022) and Ferreira (2022), discuss different ways of applying statistics in mining geotechnics, these being, respectively: The study of outliers for identification discrepant readings in piezometers when monitoring dams; The use of statistical analyzes to compare the performance of different methods of obtaining moisture for the technological control of filtered waste disposal structures; And the comparison of deterministic and probabilistic analyzes to determine the safety factor in the study of dam stability. In this article, a case study will be carried out with real data from iron ore tailings piles, applying statistical techniques to predict geotechnical variables from technological control.

As emphasized by Phoon (2020), one of the greatest difficulties in carrying out statistical analyzes in geotechnics is the small amount of data available. This problem is not caused by the lack of soil studies in geotechnical constructions, but by the lack of a database that can be used for future analyses, enabling relationships between different parameters, types of materials and interpretations of data in relation to the regions under study.

Despite being recognized for years in the literature, the importance of working with a data management system has been gaining prominence in recent years, and there are currently many mining companies that carry out geotechnical processes on physical paper, writing results for digitizing information (de Lacerda and Chammas, 2023). In order to meet this need, Geolabor was developed, an intelligent geotechnical data management system, aimed at the mining field. Acting in all engineering processes, from planning, operation, data processing and approval of results, Geolabor builds a structured database, automatically updated with its use for field investigations, laboratory and technological control.

By using Geolabor data, it is possible to compile data from different structures to carry out a complete statistical analysis. In this article, data from two iron ore tailings piles from different locations in Minas Gerais/Brazil will be used to estimate geotechnical parameters of technological control, via Bayesian geostatistical models. Totaling more than two thousand samples collected, the main objective of this article is to predict the missing values of geotechnical variables obtained in tests carried out in the technological control of the structures under study.

Spatial and non-spatial analyzes will be carried out in order to understand the importance of taking into account spatial effects for this type of material. In that way, this work validates the possibility of using data from different piles together, resulting in analyzes with a greater quantity of observed data and enabling, for example, the future estimation of the material's physical and geomechanical parameters obtained from triaxial tests in the laboratory and carried out less frequently than those resulting from other tests from technological control.

2 MISSING DATA

The three main types of missing data are Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR), where: MCAR applies when there is no need to treat missing data, since the data observed in the sample are representative of the population; MAR applies to cases where the missing data is not random, however, there is information in the other variables in the sample that can be used to complete the data; and MNAR, which will deal with samples where the reason for the existence of missing data is related to the variable of interest, resulting in the possibility of biased analyzes (Little and Rubin, 2019).

In this paper, the missing data will be characterized as MAR, since the geotechnical variables can be explained by the other variables observed from the same material. The analyzes will be carried out in R, by using the averages of the nearest neighbors; the MissForest packages (Stekhoven and Bühlmann, 2012), a non-parametric method that allows interactive effects and no linear, working with data of mixed types; and MICE package (Van Bureen and Karin, 2011), a method that works with imputation under the multivariate normal and log-linear distribution, specifying a multivariate distribution for the missing data and imputing the values from MCMC (Markov Chains Monte Carlo).

3 ANALYSIS

In this article, predictions were made of the following geotechnical parameters, obtained from HILF tests:

Ground Moisture (w_{med} – %); Maximum Dry Density (γ_{dmax} – g/cm³); Optimum Moisture (w_{opt} – %); Degree of Compaction (Dc – %); Moisture Degree (Dw – %). As the work is being carried out with real information, the missing data is, in fact, unknown. Thus, in order to enable compare the predicted values with the real ones, ramdom points were removed, for subsequent verification of results. Table 1 presents the number of points analyzed, along with the percentage of missing data for each variable studied in each pile.

Table 1. Missing data percentage of each covariate.

Pile	Pile 1	Pile 2	Pile 1 and Pile 2
Total Samples	1823	655	2478
Removed Samples	182	57	239
Missing w_{med} (%)	9.98	20.15	12.67
Missing γ_{dmax} (%)	10.09	9.47	9.93
Missing w_{opt} (%)	10.04	9.47	9.89
Missing Dc (%)	10.20	21.83	13.66
Missing Dw (%)	10.04	20.15	13.59

The histograms of each geotechnical variable observed in Pile 1 and Pile 2 together are shown in Figure 2.

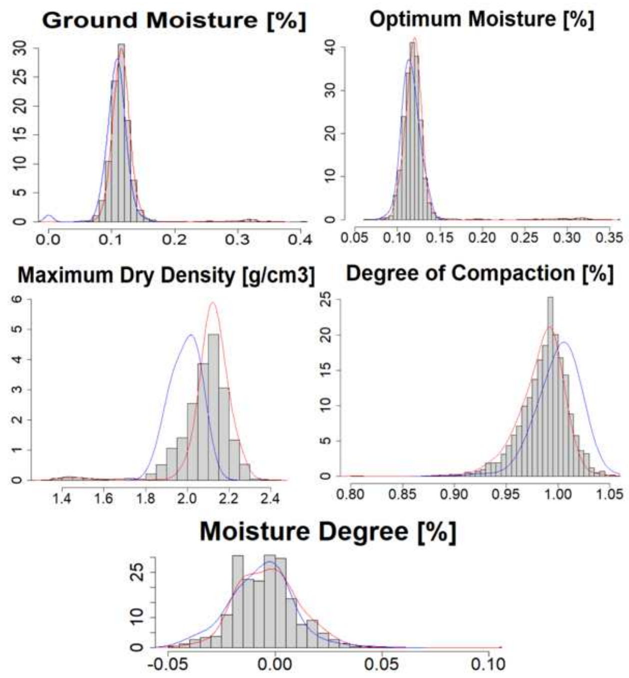

Figure 1. Histograms of each geotechnical variable observed in Pile 1 and Pile 2 together, along with the density curves of Pile 1, in red, and Pile 2, in blue.

Figure 2 presents the two-by-two relationships between the geotechnical variables observed in Pile 1, in red, and in Pile 2, in blue.

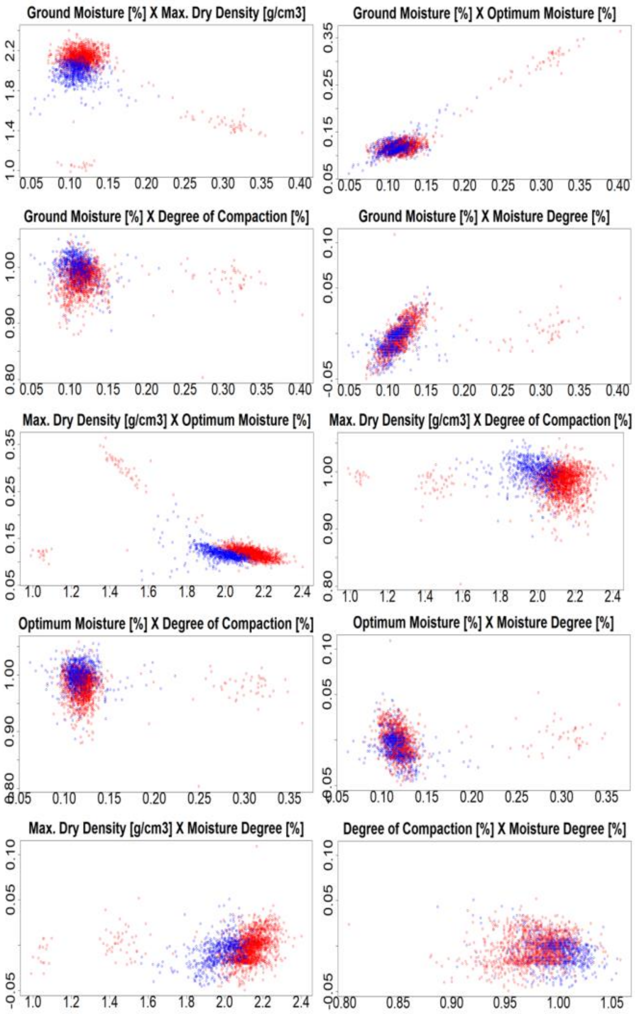

Figure 2. Two-by-two relations between the variables in analyse (red are referred to Pile 1 and blue to Pile 2).

It can be observed, from the histograms presented in Figure 2 and the graphics presented in the Figure 3, that the geotechnical variables present a similar behavior, with a slightly greater discrepancy in values for the Maximum Dry Density. This similarity is due to the same type of material being studied, despite coming from different mineral activities, and arranged in different structures.

Table 2 contains the percentage errors of the predictions made from the missing data imputation. The percentages of points where the error obtained was less than 5% and 10% are also presented. The results are obtained by comparing the value predicted by the above methodologies with the real values that were initially artificially removed for verification. An1 is refer to Close Neighbouhood; An2 to MissForest R Package; and An3 to MICE R Package. All results in Table 1 are given as percentages (%).

Table 2. Errors from analysis, with the best results in bold.

Error	w_{med}	γ_{dmax}	w_{opt}	Dc	Dw
Pile 1: Error <= 5%	31.87	89.56	58.79	94.51	4.4
	89.01	**92.86**	**89.01**	**96.7**	26.92
	85.16	84.62	86.81	88.46	**37.36**
Pile 1: Error <= 10%	57.14	99.45	84.07	**100**	6.59
	92.31	**100**	**96.7**	99.45	40.66
	91.21	98.9	95.05	99.45	**50**
Pile 1: Average Error	10.4	2.46	6.23	1.81	170.26
	2.4	**1.96**	**2.21**	**1.64**	**41.72**
	2.81	2.74	2.67	2.2	47.74
Pile 2: Error <= 5%	26.92	75.44	36.84	97.83	9.43
	73.08	**96.49**	**80.7**	**100**	**22.64**
	36.54	85.96	49.12	95.65	11.32
Pile 2: Error <= 10%	46.15	98.25	64.91	100	11.32
	92.31	**100**	**91.23**	**100**	**37.74**
	63.46	98.25	73.68	100	13.21
Pile 2: Average Error	12.76	3.12	8.4	1.43	142.45
	3.41	**1.95**	**3.57**	**1.05**	**35.09**
	9.17	2.92	6.2	1.88	133.09
Piles 1 and 2: Error <= 5%	31.47	75.21	50	**96.92**	5.56
	86.64	**81.93**	**86.97**	96.48	**35.04**
	83.62	64.71	74.37	88.11	32.48
Pile 1 and Pile 2 Together: Error <= 10%	61.21	94.96	76.47	**99.56**	5.56
	94.4	**97.06**	**98.32**	**99.56**	**50.43**
	92.24	94.12	91.6	99.12	**50.43**
Pile 1 and Pile 2 Together: Average Error	10.32	3.82	7.01	1.79	123.69
	2.37	**2.87**	**2.05**	**1.74**	**35.7**
	4.22	4.3	3.92	2.53	45.41

4 CONCLUSIONS AND NEXT STEPS

Based on the results presented in Table 2, the methodology that presented the best results in most analyzes was the MissForest R package. It is possible to state that the implemented methodologie is mostly satisfactory, achieving predictions with percentage errors lower than 4% for all geotechnical variables analyzed, but to the Moisture Degree. The error quality analysis corroborates this statement, by presenting at least 73% of results with a percentage error of less than 5% and at least 91% than 10%. The domain of the moisture deviation variable is very close to zero, which may be affecting the predictions or the calculation of errors. This in-depth analysis will be carried out later.

So far, no multivariate imputation methodology that takes spatial effects into account has been implemented, nor has the existence and influence of potential outliers been analyzed. Analyzing the results presented, it is possible to verify that the reduction in error quality is not significant when data from the same material and different locations are combined in the analyses. Thus, the development of the work so far suggests that it is feasible and makes practical sense to combine information from different stacks to make predictions of values that would not be possible in a single structure due to insufficient data. Following the development, the same analyses presented in this article will be performed with the geotechnical variables obtained in the CPN, GPS, MES, and CIUsat tests, considering outliers and spatial effects.

ACKNOWLEDGEMENTS

The authors are thankful for the support of the Institute of Mathematics and Statistics at UFRJ and the team at Simplelab Tecnologia, for the Geolabor data.

REFERENCES

Azevedo, A. G. (2019). Statistical analysis to identify discrepant piezometry readings: case study in an earthen dam. Available in: <monografias.ufop.br/>.

Badaró, J. V. O. (2022). Statistical analysis as a decision-making tool: case study in determining moisture in the technological control of a filtered waste disposal structure. Available in: <ufop.br>.

de Lacerda, B. M. and Chammas, S. C. (2023). The Importance of Geotechnical Data Management in Pile and Tailings Dam Works. SNGB, Dams Week, Foz do Iguaçú. Available in: <https://cbdb-web>.

Ferreira, P. M. (2022). Assessment of the reliability of the safety factor obtained by probabilistic analysis in relation to deterministic analysis in mining dams. Available in: <monografias.ufop.br/>.

Flores, E. A. F. (2008). Probabilistic analyses of slope stability considering the spatial variability of the soil. Rio de Janeiro. Available in: <abge.org.br/>.

Little, Roderick JA; Rubin, Donald B. Statistical analysis with missing data (2019). Vol. 973.

Phoon, K. K. (2020). *The story of statistics in geotechnical engineering*. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 14(1), 3-25. https://doi.org/10.1080/17499518.2019.1700423.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest non-parametric missing value imputation for mixed-type data. Bioinformatics, 112-118. https://doi.org/10.1093/bioinformatics/btr597.

Van Bureen, Stef; G.O., Karin. *Mice: Multivariate imputation by chained equations in R*. Journal of statistical software (2011). v. 45, p. 1-6. https://doi.org/10.18637/jss.v045.i03.