

Abordagem Estatística para a Validação de Critérios de Aceitação de Camadas Compactadas em Pilhas e Barragens de Rejeito

Bianca Mendes de Lacerda

Engenheira Civil, Simplelab Tecnologia, Belo Horizonte, Brasil, bianca.lacerda@simplelabtech.com.br

Luiza do Bonfim Arce

Estatística, DME/UFRJ, Rio de Janeiro, Brasil, luiza@dme.ufrj.br

Tadeu Henrique Oliveira de Castro

Engenheiro Civil, Chammas Engenharia, Belo Horizonte, Brasil, tadeu@chammas.com.br

RESUMO: A pesquisa por inovações nas técnicas de construção de pilhas e barragens de rejeito vem crescendo na geotecnia, buscando a realização de uma prática mais segura, evitando possíveis novos acidentes e rompimentos futuros. A aplicação prática da estatística na construção dessas estruturas vem ganhando destaque nos últimos anos, mas ainda é um estudo de abordagem restrita. Dentre as metodologias de acompanhamento que são realizadas ao longo da obra, o controle tecnológico é destacado, sendo o processo que garante que as especificações técnicas de projeto sejam cumpridas. Atualmente, todo o processo do controle tecnológico é realizado de forma determinística, onde os parâmetros geotécnicos de controle devem estar dentro de um intervalo de valores especificados para que o ensaio realizado seja aprovado. Neste artigo, será apresentada uma abordagem estatística da aceitação dos critérios de especificação técnica para ensaios geotécnicos, levando em consideração a dispersão dos resultados obtidos para uma determinada região na estrutura, a fim de analisar o efeito destes resultados no processo do controle tecnológico. Foi realizado um estudo de caso utilizando o Geolabor, sistema de gerenciamento de dados geotécnicos, que apresenta uma solução completa de digitalização dos dados do controle tecnológico, comparando diferentes modelos estatísticos e de aprendizado de máquina, mostrando a importância das metodologias implementadas ao comparar a aderência às especificações ao se trabalhar com os parâmetros geotécnicos como variáveis aleatórias ou valores reais.

PALAVRAS-CHAVE: Geotecnia Estatística, Mineração, Pilhas e Barragens, Critérios de Aceitação, Inovação na Geotecnia

ABSTRACT: The search for innovations in the construction techniques of piles and tailings dams has been growing in geotechnics, aiming for a safer practice, avoiding potential new accidents and future breaches. The practical application of statistics in the construction of these structures has been gaining prominence in recent years, but it is still a study of restricted approach. Among the monitoring methodologies carried out throughout the work, technological control stands out, being the process that ensures that the technical specifications of the project are met. Currently, the entire process of technological control is carried out deterministically, where the geotechnical control parameters must be within a range of specified values for the test to be approved. In this article, a statistical approach to the acceptance of technical specification criteria for geotechnical tests will be presented, taking into account the dispersion of the results obtained for a specific region in the structure, in order to analyze the effect of these results on the technological control process. A case study was conducted using Geolabor, a geotechnical data management system, which provides a complete solution for digitizing technological control data, comparing different statistical and machine learning models, demonstrating the importance of the implemented methodologies when comparing adherence to specifications when working with geotechnical parameters as random variables or real values.

KEYWORDS: Statistical Geotechnics, Mining, Piles and Dams, Acceptance Criteria, Geotechnical Innovation

1 INTRODUÇÃO

O estudo da Estatística é amplamente discutido e aplicado em diversas áreas de atuação, tanto academicamente, quanto na prática. Na geotecnia, historicamente, o comportamento dos materiais é tratado de forma determinística, o que pode levar a erros significativos, devido às incertezas existentes provenientes da medição dos parâmetros geotécnicos, na realização dos ensaios e na distribuição natural dos solos (Souza, 2020). A aplicação de análises estatísticas na geotecnia da mineração está ganhando espaço e destaque nos últimos anos, mas ainda apresenta uma abordagem restrita (Flores, 2008).

O principal objetivo deste trabalho é realizar um estudo de caso com dados reais, extraídos do Geolabor, a fim de construir e comparar metodologias estatísticas e de machine learning para descrever o comportamento do índice de vazios no controle tecnológico de uma pilha de rejeitos. O Geolabor é um sistema de gerenciamento geotécnico, que possui um produto completamente voltado para o controle tecnológico (de Lacerda e Chammas, 2023). É destacado por Phoon (2020) que uma das principais dificuldades da realização prática de análises estatísticas na geotecnia é a falta de informações estruturadas.

Com a utilização do Geolabor, a problemática com a falta de dados é solucionada. Será possível analisar as relações entre diferentes parâmetros geotécnicos, utilizando de uma base rica e estruturada de dados reais, para a realização de diferentes modelos estatísticos e análises críticas. Com os modelos aplicados, será possível realizar previsões, levando em consideração o efeito das possíveis incertezas que existem na medição dos parâmetros geotécnicos que ditam, em conjunto, o comportamento do material compactado.

Para o estudo de caso que será apresentado, foi selecionada uma obra onde o critério para a liberação das camadas ao longo da construção exige que o índice de vazios seja menor do que 0.60. Com as análises que serão realizadas, espera-se ser possível comparar a aderência real à especificação técnica, com a aderência calculada a partir dos valores que levam em consideração, estatisticamente, as incertezas associadas. Espera-se, também, desenvolver um melhor entendimento sobre o efeito e influência que os diferentes parâmetros geotécnicos analisados possuem entre si, separada e conjuntamente.

2 METODOLOGIA

Para entender melhor o comportamento do índice de vazios a partir de diferentes parâmetros geotécnicos obtidos no controle tecnológico de uma pilha de rejeitos, foram utilizadas duas técnicas: regressão linear múltipla e Random Forest. Para as análises estatísticas que serão realizadas neste artigo, o índice de vazios será considerado a variável resposta, e os demais parâmetros geotécnicos analisados, as variáveis explicativas.

2.1 Regressão Linear Múltipla

Seja y o vetor dos n valores de índice de vazios calculados, X a matriz de desenho, onde a primeira coluna contém apenas o valor 1 e as colunas restantes representam os $p-1$ parâmetros geotécnicos considerados, e β o vetor de coeficientes da regressão, ou seja, os parâmetros desconhecidos que medem o efeito que cada variável tem no índice de vazios. Dessa forma, a variável resposta pode ser escrita como uma combinação linear dos parâmetros geotécnicos, somada a uma quantidade aleatória que não se pode controlar, chamada de erro aleatório (ϵ). Por hipótese, os n erros aleatórios possuem distribuição normal com média 0 e variância σ^2 e são independentes entre si. Nesse contexto, supõe-se que a relação entre os parâmetros geotécnicos e o índice de vazios pode ser descrita a partir das Equações 1 a 5 a seguir.

$$y = (y_1, \dots, y_n)^T \quad (1)$$

$$\beta = (\beta_0, \dots, \beta_{p-1})^T \quad (2)$$

$$y = X\beta^T + \epsilon \quad (3)$$

$$\epsilon = (\epsilon_1, \dots, \epsilon_n)^T \quad (4)$$

$$\epsilon \sim N_n(0_n, \sigma^2 I_n) \quad (5)$$

Onde N_n indica que os n erros aleatórios possuem distribuição normal n -variada, 0_n é a matriz de zeros e I_n é a matriz identidade, ambas de tamanho $n \times n$.

Para este modelo, são realizadas algumas suposições, podendo estas apresentarem consequências para a análise dos dados: a relação entre o índice de vazios e os parâmetros geotécnicos selecionados é linear; os erros têm distribuição normal, portanto, os índices de vazios também possuem distribuição normal; os erros são independentes, portanto, os índices de vazios são independentes entre si; os erros têm variância constante σ^2 , então os índices de vazios têm variância constante σ^2 .

Dessa forma, para que o modelo utilizado seja considerado válido, é necessário verificar se os dados e o ajuste cumprem com os requisitos citados acima. Esta verificação pode ser realizada através de técnicas estatísticas, como testes de hipótese, e análises gráficas. Para a verificação da linearidade dos dados, analisou-se a significância da regressão através do teste de hipótese F , a partir do comando Summary do R; para a normalidade da distribuição dos erros, utilizou-se o teste de Shapiro Wilk (Royston, 1982); para a independência, foi realizada a análise gráfica dos resíduos, definidos como a diferença entre o valor observado e o ajustado pelo modelo; e para verificar se a variância dos erros é constante, optou-se pelo teste de Brown-Forsythe (Derrick, 2018).

3.2 Random Forest

O conceito geral de Random Forest foi apresentado por Breiman (2001). Este é um método ensemble, que combina múltiplos modelos de aprendizado de máquinas para obter métodos mais robustos e resultados de maior qualidade. Em especial, este tipo de modelo utiliza árvores de decisão, um tipo de modelo de aprendizado que particiona os dados de acordo com os padrões observados, de forma a obter grupos e subgrupos dispostos em uma estrutura hierárquica, em que haja homogeneidade dentro dos grupos e heterogeneidade entre eles.

O problema do uso dessas árvores é a grande variância das estimativas, além do alto risco de overfitting, ou seja, de que o modelo aprenda demais sobre um determinado conjunto de dados e não consiga generalizar as informações para observações que não estão ali. Diante desta situação, foi desenvolvido o Random Forest, metodologia que utiliza um conjunto de árvores de decisão para minimizar os riscos de haver um overfitting através da utilização da média do ajuste de cada árvore da floresta (Hastie et al, 2009).

Um ponto positivo do Random Forest é que não existem pressuposições a serem feitas a partir dos dados, vantagem que o torna popular e amplamente difundido na área. Por outro lado, embora o modelo traga uma fácil aderência a qualquer tipo de dado, ainda existem limitações: devido à natureza da floresta, não é possível estabelecer uma relação direta entre as variáveis explicativas e a variável resposta, como no caso do modelo de regressão múltipla. Além disso, não há uma interpretação direta para os resultados, como há no modelo de regressão, em que β_i representa o efeito do i -ésimo parâmetro geotécnico no índice de vazios. Ainda assim, o Random Forest pode ser uma boa alternativa para previsões.

3 ESTUDO DE CASO

3.1 Análise exploratória

A base de dados reais utilizada para a construção do estudo de caso que será apresentado neste artigo é composta pelos parâmetros geotécnicos apresentados na Tabela 1, sendo exibidos os resultados provenientes dos ensaios realizados nas primeiras 5 amostras, das 1030 analisadas. Todas as amostras foram coletadas em uma mesma obra e para o mesmo tipo de material. Não existem dados faltantes na base de dados para a realização das análises. Os parâmetros geotécnicos utilizados e suas respectivas nomenclaturas e unidades são: Índice de Vazios (e), Grau de Compactação ($GC - \%$), Massa Específica Seca Máxima ($\gamma_{D, \text{máx}} - g/cm^3$), Umidade ($w - \%$), Umidade Ótima ($w_{ot} - \%$), Desvio de Umidade ($dw - \%$), Massa Específica dos Grãos ($\gamma_s - g/cm^3$), e Massa Específica Aparente In Situ ($\gamma_D - g/cm^3$).

Tabela 1 - Apresentação dos parâmetros geotécnicos utilizados no estudo de caso.

| e | GC | γ_D , máx. | w | wot | dw | γ_s | γ_D |
|------|---------|-------------------|--------|--------|--------|------------|------------|
| 0.48 | 100.80% | 2.07 | 10.30% | 11.10% | -0.80% | 3.08 | 2.09 |
| 0.6 | 98.90% | 1.852 | 10.00% | 13.40% | -3.40% | 2.93 | 1.83 |
| 0.51 | 100.40% | 2.017 | 11.40% | 11.40% | 0.00% | 3.06 | 2.03 |
| 0.5 | 100.50% | 2.021 | 10.90% | 11.30% | -0.40% | 3.05 | 2.03 |
| 0.58 | 99.00% | 1.876 | 11.60% | 13.50% | -1.90% | 2.94 | 1.86 |

O histograma da variável resposta está apresentado na Figura 1 a seguir. A fim de avaliar a distribuição deste parâmetro, em vermelho está representada a curva de densidade do índice de vazios e, em azul, a curva de densidade da distribuição normal calculada a partir da média e desvio padrão dos dados.

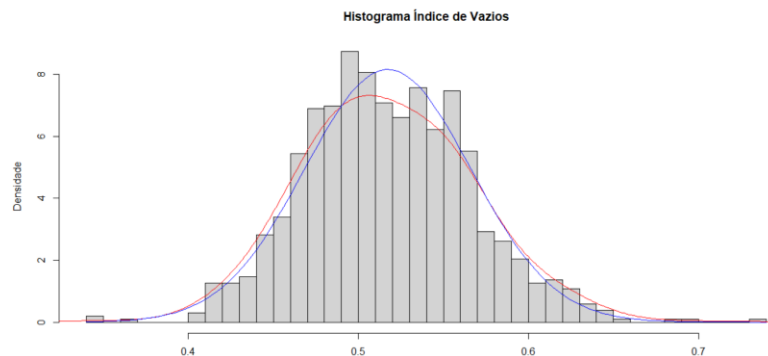


Figura 1 – Histograma da variável resposta, índice de vazios, junto às curvas de densidade da distribuição normal (azul) e dos dados (vermelha).

É possível observar, a partir da Figura 1, que o índice de vazios não segue perfeitamente uma distribuição conhecida, estando próximo da normal, porém com uma variação em torno da média. Verifica-se, também, que existem dois pequenos picos nas caudas dos dados, que não são identificados na distribuição normal. Essas particularidades dos dados podem afetar os resultados das análises que serão realizadas.

Para entender a relação entre os parâmetros geotécnicos analisados, será apresentada a matriz de correlações na Tabela 2 a seguir. Quanto mais próximo de 1 ou -1, maior é a correlação existente entre as variáveis, analisadas separadamente.

Tabela 2 – Matriz de correlação entre os parâmetros geotécnicos analisados.

| | e | GC | γ_D | w | γ_s | dw | wot | γ_D , máx. |
|-------------------|-------|-------|------------|-------------|------------|-------------|-------------|-------------------|
| e | 1.00 | -0.46 | -0.64 | 0.30 | 0.33 | 0.07 | 0.34 | -0.39 |
| GC | -0.46 | 1.00 | 0.38 | -0.22 | -0.05 | -0.32 | 0.16 | -0.21 |
| γ_D | -0.64 | 0.38 | 1.00 | -0.09 | 0.51 | 0.21 | -0.45 | 0.82 |
| w | 0.30 | -0.22 | -0.09 | 1.00 | 0.22 | 0.78 | 0.28 | 0.04 |
| γ_s | 0.33 | -0.05 | 0.51 | 0.22 | 1.00 | 0.33 | -0.17 | 0.57 |
| dw | 0.07 | -0.32 | 0.21 | 0.78 | 0.33 | 1.00 | -0.38 | 0.41 |
| wot | 0.34 | 0.16 | -0.45 | 0.28 | -0.17 | -0.38 | 1.00 | -0.58 |
| γ_D , máx. | -0.39 | -0.21 | 0.82 | 0.04 | 0.57 | 0.41 | -0.58 | 1.00 |

A partir dos resultados apresentados na Tabela 2, é possível observar que, no geral, as correlações dois a dois entre os parâmetros geotécnicos avaliados não são muito altas, estando, em módulo, em torno de 0.4 para a maioria das variáveis. É interessante observar, como destacado na Tabela 2, que as correlações entre o desvio de umidade e o índice de vazios; a umidade e a massa específica seca máxima; e a umidade ótima e o grau de compactação são muito baixas ou quase nulas. Este é um comportamento que vem sendo observado em campo, quando se trata de rejeitos, diferente do que acontece com solos naturais. As curva de compactação dos rejeitos são praticamente paralelas, e a variação do índice de vazios aparenta ser mais dependente da energia de compactação do que dos parâmetros físicos de compactação.

Dos parâmetros geotécnicos que irão compor a matriz de covariáveis explicativas, são conhecidas duas relações diretas entre os dados, apresentadas nas Equações 6 e 7.

$$dw = w - wot \tag{6}$$

$$GC = \frac{\gamma_D}{\gamma_{D,m\acute{a}x}} \times 100 \tag{7}$$

Para a construção correta dos modelos estatísticos, as variáveis explicativas precisam, por hipótese, ser independentes. Desta forma, serão retirados das análises os parâmetros geotécnicos Desvio de Umidade e Massa Específica Aparente Seca In Situ. A fim de verificar que, de fato, não existem problemas de multicolinearidade entre as variáveis restantes, será calculado o VIF, Fator de Inflação de Variância (Fox e Monette, 1992), que indica multicolinearidade para valores acima de 10, calculado a partir da diagonal da inversa da matriz de correlações entre as covariáveis explicativas. Os resultados do VIF calculado estão apresentados na Tabela 3 a seguir.

Tabela 3 – Fator de Inflação para as variáveis explicativas que serão utilizadas no estudo de caso.

| GC | w | γ_s | wot | $\gamma_{D, m\acute{a}x}$ |
|------|------|------------|------|---------------------------|
| 1.14 | 1.29 | 1.65 | 1.80 | 2.35 |

Como é observado na Tabela 3, todos os valores obtidos para o VIF são menores do que 10, indicando que não existem mais problemas com multicolinearidade nos dados. Sendo assim, estes serão os parâmetros geotécnicos que irão compor a matriz de covariáveis para a realização do estudo de caso. As estatísticas de resumo para os dados que serão utilizados neste trabalho estão apresentadas na Tabela 4.

Tabela 4 – Estatísticas de resumo para os parâmetros geotécnicos que serão analisados neste trabalho.

| Variável | Média | Desvio Padrão | Mínimo | 1° Quartil | 2° Quartil | 3° Quartil | Máximo |
|---------------------------|---------|---------------|--------|------------|------------|------------|---------|
| e | 0.517 | 0.049 | 0.34 | 0.484 | 0.514 | 0.551 | 0.732 |
| GC | 100.24% | 2.04% | 88.70% | 99.10% | 100.40% | 101.50% | 108.40% |
| w | 11.10% | 1.33% | 4.83% | 10.29% | 11.14% | 11.93% | 16.39% |
| γ_s | 3.056 | 0.089 | 2.803 | 2.987 | 3.054 | 3.113 | 3.381 |
| wot | 11.54% | 0.90% | 8.10% | 10.90% | 11.50% | 12.10% | 14.30% |
| $\gamma_{D, m\acute{a}x}$ | 2.011 | 0.067 | 1.81 | 1.97 | 2.021 | 2.061 | 2.176 |

Por fim, para a modelagem, a base foi dividida em 80% de treino e 20% de teste. A seguir, tem-se os resultados para as técnicas utilizadas.

3.2 Resultados

A fim de visualizar a relação que os parâmetros geotécnicos possuem em conjunto, foi calculado um esquema de árvore de decisão aleatória (Breiman, et al, 1984), que está apresentado na Figura 2.

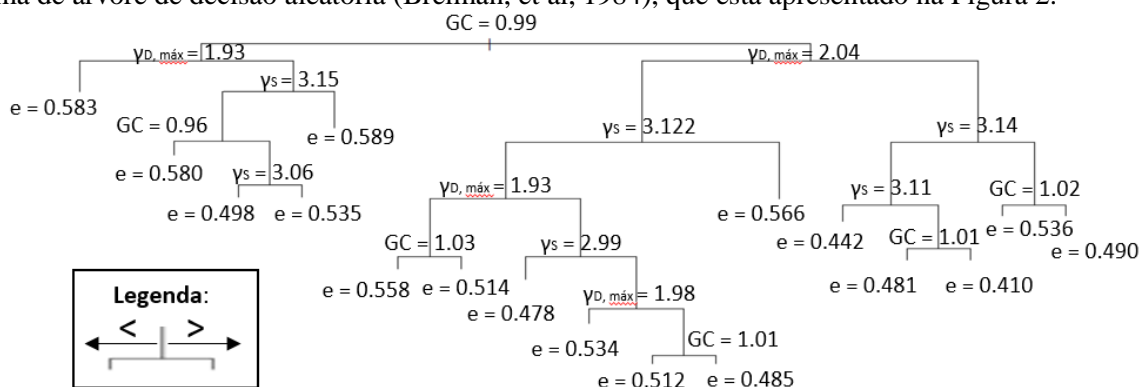


Figura 2 – Relação de árvore aleatória, da variação do índice de vazios com os intervalos das explicativas.

A partir da Figura 2, é possível observar que, para intervalos de dados muito próximos, existe uma diferença considerável do índice de vazios obtido, dependendo dos outros parâmetros geotécnicos, o que destaca a complexidade das relações entre os dados, quando analisados em conjunto. Dentre as cinco variáveis explicativas, a maior influência para a variação dos valores do índice de vazios parece ser devido ao grau de compactação, massa específica dos grãos e massa específica seca máxima, uma vez que a umidade ótima e a umidade do aterro não aparecem na árvore gerada. Esta conclusão corrobora com o que foi analisado

Com o objetivo de comparar diferentes modelos com os dados disponíveis, além de utilizar de duas diferentes metodologias, regressões lineares múltiplas e Random Forest, foi trabalhado com diferentes combinações de parâmetros geotécnicos compondo a matriz de variáveis explicativas e transformações na variável explicativa. A Tabela 5 apresenta o resumo dos resultados obtidos em todas as análises realizadas.

Tabela 5 – Resumo dos resultados obtidos nos diferentes modelos estatísticos construídos.

| Análise | Modelo | R ² | RMSE | Shapiro-Wilk p-valor | Resíduos Indep. |
|-----------|---|----------------|--------------|-------------------------|--------------------|
| 1 | $e \sim GC + w + \gamma_S + wot + \gamma_{D, \text{máx}}$ | 0.998 | 0.002 | 2.20E-16 | Sim |
| 2 | $e \sim GC + w + wot + \gamma_{D, \text{máx}}$ | 0.513 | 0.034 | 3.43E-07 | Não |
| 3 | $e \sim GC + \gamma_{D, \text{máx}} + \gamma_S$ | 0.998 | 0.002 | 2.20E-16 | Sim |
| 4 | $\ln(e) \sim GC + w + wot + \gamma_{D, \text{máx}}$ | 0.475 | 0.065 | 3.58E-04 | Não |
| 5 | $\ln(e) \sim GC + \gamma_{D, \text{máx}} + \gamma_S$ | 0.997 | 0.005 | 2.20E-16 | Sim |
| 6 | $\exp(e) \sim GC + w + wot + \gamma_{D, \text{máx}}$ | 0.472 | 0.056 | 9.30E-09 | Não |
| 7 | $\exp(e) \sim GC + \gamma_{D, \text{máx}} + \gamma_S$ | 0.995 | 0.006 | 2.20E-16 | Sim |
| 8 | $\ln(e) \sim GC + w + wot + \gamma_{D, \text{máx}} - 1$ | 0.989 | 0.072 | 0.004 | Sim |
| 9 | $\ln(e) \sim GC + \gamma_{D, \text{máx}} + \gamma_S - 1$ | 0.997 | 0.035 | 3.35E-08 | Não |
| 10 | $\ln(e) \sim GC + \gamma_S + wot + \gamma_{D, \text{máx}}$ | 0.998 | 0.030 | 0.0010 | Não |
| 11 | RF: $e \sim GC + w + \gamma_S + wot + \gamma_{D, \text{máx}}$ | 0.925 | 0.014 | - | - |
| 12 | RF: $e \sim GC + w + wot + \gamma_{D, \text{máx}}$ | 0.412 | 0.036 | - | - |
| 13 | RF: $\ln(e) \sim GC + w + \gamma_S + wot + \gamma_{D, \text{máx}}$ | 0.920 | 0.015 | - | - |
| 14 | RF: $\ln(e) \sim GC + w + wot + \gamma_{D, \text{máx}}$ | 0.409 | 0.036 | - | - |

Na Tabela 5, a presença do “-1” na descrição do modelo indica que o intercepto não foi considerado para a regressão. Como mencionado anteriormente, a distribuição dos dados do índice de vazios não segue uma distribuição perfeitamente normal e, por este motivo, nenhuma das regressões lineares múltiplas realizadas (análises 1 – 10) apresentou resíduos com distribuição normal (shapiro-wilk p-valor > 0.05). Sendo assim, com um R² alto, Raiz do Erro Quadrático Médio (RMSE) baixo, resíduos independentes e mais próximos da normalidade do que nas outras análises, o melhor modelo para este tipo de metodologia foi o de número 8, como destacado na Tabela 5. A partir dos resultados apresentados, é possível observar que a aplicação da transformação logarítmica da variável resposta e a retirada do intercepto levaram a uma melhora do modelo.

Utilizando o Random Forest (análises 11 – 14), por sua vez, não houve uma diferença significativa ao considerar a variável resposta com ou sem o logaritmo aplicado, como é esperado, uma vez que, nesta metodologia, não é realizada uma regressão linear, e o aprendizado de máquina vai funcionar da mesma forma para as duas situações. Observa-se que não foi possível construir um modelo eficiente sem utilizar a massa específica dos grãos como parte das variáveis explicativas, como aconteceu no modelo 8. Dessa forma, o melhor modelo para o Random Forest foi aquele que apresentou o maior R² e menor RMSE, sendo este o modelo 11, destacado na Tabela 5. Apesar das análises anteriores sugerirem uma baixa influência da umidade e da umidade ótima, nas duas metodologias analisadas, os melhores modelos contam com o efeito destes parâmetros geotécnicos para explicar o índice de vazios de forma mais eficiente.

Apesar de terem alcançado ajustes bons, com R² consideravelmente altos, os melhores modelos destacados não apresentaram uma boa capacidade para ajustar valores da variável resposta nas caudas da distribuição. A quantidade de amostras analisadas que apresentam índice de vazios maior do que 0.60 é de apenas 52 em 1030 observados. Dessa forma, os modelos podem estar interpretando esses pontos como outliers, além de não possuírem informações suficientes para representar o comportamento da relação e efeito

das variáveis explicativas na variável resposta para estes valores. As análises realizadas até este momento destacam a influência dos diferentes parâmetros geotécnicos do controle tecnológico no resultado do índice de vazios, principalmente do grau de compactação e da massa específica seca máxima (além da massa específica dos grãos, que apresenta relação direta e conhecida com o índice de vazios).

Além da maior compreensão do comportamento e relação conjunta dos parâmetros geotécnicos em rejeitos, o outro objetivo deste estudo de caso seria comparar a alteração da aderência à especificação técnica, uma vez que se levasse em consideração uma variação nos dados que explicam o índice de vazios. Como não foi possível, para este conjunto de dados, modelar os valores acima do especificado, este objetivo não foi concluído. Observando os valores muito baixos dos desvios padrões apresentados na Tabela 2, é possível afirmar que os resíduos analisados neste estudo de caso compõem um material homogêneo. Será realizada, então, uma análise simplificada a partir do desvio padrão, a fim de ressaltar a importância desse tipo de análise e incentivar futuros trabalhos mais aprofundados.

O índice de vazios é obtido a partir da relação entre a massa específica dos grãos e a massa específica seca aparente in situ, apresentada na Equação 8.

$$e = \frac{\gamma_s}{\gamma_D} - 1 \quad (8)$$

Dessa forma, serão realizados três cálculos: considerando o desvio padrão da massa específica dos grãos; da massa específica seca aparente in situ; e do índice de vazios. A partir das duas primeiras análises, será possível calcular a variação da aderência à especificação técnica, ao considerar que podem haver incertezas nas medições de γ_s e γ_D , respectivamente. Já na terceira análise, além de considerar as mesmas possíveis incertezas das anteriores, ainda é levado em consideração, indiretamente, as incertezas associadas à obtenção outros parâmetros geotécnicos que influenciam no índice de vazios, como o grau de compactação e a massa específica seca máxima.

Para a realização das análises, serão considerados, apenas, os casos mais críticos quanto ao efeito no resultado do índice de vazios, ou seja, o desvio padrão de sua respectiva variável será acrescido ao γ_s , decrescido de γ_D e acrescido no índice de vazios. Os resultados obtidos estão apresentados na Tabela 6 a seguir.

Tabela 6 – Comparação das aderências à especificação técnica ($e < 0.6$) ao levar em consideração o desvio padrão das variáveis γ_s , γ_D e índice de vazios.

| | γ_s | γ_D | e |
|-------------------------|------------|------------|--------|
| Desvio Padrão | 0.0796 | 0.0713 | 0.0522 |
| Nova Aderência | 91.61% | 85.33% | 88.22% |
| Aderência Original | 95.00% | 95.00% | 95.00% |
| Diferença de Aderências | 3.39% | 9.67% | 6.78% |

4 CONCLUSÕES

Neste artigo, foi realizada uma análise estatística completa, com uma revisão teórica e aplicação do conhecimento, para o entendimento dos parâmetros geotécnicos em obra de pilha de rejeitos. Este tipo de estudo ainda é pouco difundido na geotecnia de mineração, principalmente quando se trata de rejeitos, sendo de extrema importância para compreender melhor o comportamento destes materiais e ajudar a garantir a qualidade das construções. Apesar dos modelos iniciais não terem sido suficientes para analisar com completude a influência das incertezas dos parâmetros geotécnicos na aderência à especificação técnica, a partir destas ficou clara a complexidade de suas correlações, destacando a importância da realização de novos modelos aplicados para este tipo de controle.

Nas análises com o cálculo do desvio padrão, foi possível verificar uma alteração considerável na aderência à especificação técnica ao somar/subtrair este valor às observações de cada variável, principalmente quando a incerteza está associada a massa específica seca aparente in situ, destacando a influência das possíveis incertezas de medição. Esta foi uma análise mais simplificada, para um material homogêneo, com o intuito de ressaltar a necessidade de um cuidado e da realização de análises críticas ao aprovar valores muito próximos ao especificado para a liberação das camadas. A partir destas análises, também é ressaltada a importância de incorporar os estudos estatísticos nas obras de pilhas e barragens de rejeitos.

Para futuros trabalhos, sugere-se que sejam realizadas análises com diferentes modelos para a mesma base de dados, considerando a possibilidade de associar a variável resposta a um modelo de mistura, a fim de rastrear as múltiplas modas que existem na distribuição; realizar regressões bayesianas, assumindo distribuições a priori para as variáveis explicativas; aplicar modelos não paramétricos para explicar o índice de vazios; e aplicar diferentes metodologias de machine learning para comparação de resultados e eficiência.

É sugerido, também, que sejam realizadas análises com uma nova base de dados, onde sejam conhecidos mais pontos que estejam acima da especificação técnica de projeto. Nesse caso, ainda será possível aplicar modelos lineares generalizados da família binomial para modelar a probabilidade do índice de vazios estar fora do limite da especificação técnica. Além dessas possibilidades, fica como sugestão de trabalhos práticos estudos que comparem os resultados estatísticos de amostras de rejeitos e de solos naturais com distribuição granulométrica similar, a fim de entender as particularidades do comportamento destes materiais.

AGRADECIMENTOS

Os autores gostariam de agradecer ao apoio da equipe da Simplelab Tecnologia, pelo incentivo e disponibilização dos dados para a realização do estudo de caso apresentado neste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) Classification and Regression Trees. Wadsworth.
- Breiman, L. (2001). "Random forests", *Mach. Learn.*, vol. 45, no. 1, pp. 5-32.
- de Lacerda, B. M. and Chammass, S. C. (2023). A Importância da Gestão de Dados Geotécnicos em Obras de Pilhas e Barragens de Rejeitos. SNGB, Dams Week, Foz do Iguaçu. Available in: <<https://cbdb-web>>
- Derrick, B; Ruck, A; Toher, D; White, P (2018). "Tests for equality of variances between two samples which contain both paired observations and independent observations". *Journal of Applied Quantitative Methods*. 13 (2): 36–47.
- Flores, E. A. F. (2008). Análises probabilísticas da estabilidade de taludes considerando a variabilidade espacial do solo. Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro.
- Fox, J. and Monette, G. (1992) Generalized collinearity diagnostics. *JASA*, 87, 178–183.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). Random Forests. In: *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, NY. Doi: /10.1007/978-0-387-84858-7_15
- 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. Doi: /10.1080/17499518.2019.1700423
- Royston, P. (1982). An extension of Shapiro and Wilk's WW test for normality to large samples. *Applied Statistics*, 31, 115–124. Doi: 10.2307/2347973.
- Sousa, D. H. S. (2020). Obtenção de modelos probabilísticos para parâmetros geotécnicos: estudo de caso. Disponível em: <https://repositorio.ufop.br/handle/123456789/12113>