XXII Congresso Brasileiro de Mecânica dos Solos e Engenharia Geotécnica
XI Simpósio Brasileiro de Engenheiros Geotécnicos Jovens
24 a 28 de agosto de 2026 – Brasília/DF

DOI: 10.47094/COBRAMSEG2026.2/6

Aplicações e Eficácia das Análises de Correlação Estatística na Caracterização de Propriedades Geotécnicas: Uma Revisão Sistemática

Victor Lima Cunha Borges
Discente, Universidade Federal de Goiás, Goiânia, Brasil, lima_cunha@discente.ufg.br

Larissa Andrade de Aguiar
Docente, Instituto Federal de Brasília, Samambaia, Brasil, larissa.aguiar@ifb.edu.br

Renato Resende Angelim
Docente, Universidade Federal de Goiás, Goiânia, Brasil. angelim@ufg.br

Márcia Maria dos Anjos Mascarenha
Docente, Universidade Federal de Goiás, Goiânia, Brasil, marciamascarenhaa@ufg.br

RESUMO: A caracterização precisa das propriedades geotécnicas dos solos é fundamental para projetos de engenharia civil, mas frequentemente depende de ensaios de laboratório onerosos e demorados. Este artigo apresenta uma revisão sistemática da literatura com o objetivo de compilar, caracterizar e analisar criticamente as principais aplicações de análises de correlação estatística utilizadas para predição e caracterização de propriedades geotécnicas. A metodologia seguiu o *framework* PICOC, com estratégia de busca abrangente em múltiplas bases de dados. Vinte e sete artigos foram selecionados para análise detalhada, revelando uma evolução significativa das técnicas empregadas, desde correlações lineares simples até algoritmos avançados de *Machine Learning* (ML), incluindo Random Forest, XGBoost, redes neurais e modelos Bayesianos. Os resultados destacam a superioridade dos métodos não-lineares e de ML na captura de relações complexas entre propriedades geotécnicas, com estudos demonstrando reduções substanciais em erros de predição quando comparados a correlações empíricas tradicionais. A análise individual dos estudos forneceu *insights* sobre a confiabilidade de correlações para parâmetros como resistência ao cisalhamento e compressibilidade. Conclui-se que a integração de métodos estatísticos avançados e ML representa uma mudança paradigmática na predição da caracterização geotécnica, oferecendo ferramentas mais precisas e confiáveis para engenheiros, embora a seleção adequada do método deva considerar o tamanho da amostra, a complexidade do fenômeno e a aplicabilidade regional.

PALAVRAS-CHAVE: Correlação Estatística; Propriedades Geotécnicas; Machine Learning; Caracterização de Solos; Métodos Preditivos.

ABSTRACT: Accurate characterization of geotechnical soil properties is fundamental for civil engineering projects but often relies on costly and time-consuming testing. This article presents a systematic literature review aimed at compiling, characterizing, and critically analyzing the main applications of statistical correlation analyses used for prediction and characterization of geotechnical properties. The methodology followed the PICOC framework, with a comprehensive search strategy across multiple databases. Twenty-seven articles were selected for detailed analysis, revealing a significant evolution in employed techniques, from simple linear correlations to advanced Machine Learning (ML) algorithms, including Random Forest, XGBoost, neural networks, and Bayesian models. The results highlight the superiority of non-linear and ML methods in capturing complex relationships between geotechnical properties, with studies demonstrating substantial reductions in prediction errors compared to traditional empirical correlations. Individual analysis of studies provided insights into the reliability of correlations for parameters such as shear strength and compressibility. It is concluded that the integration of advanced statistical methods and ML represents a paradigmatic shift in geotechnical characterization, offering more accurate and reliable tools for engineers,

although appropriate method selection must consider sample size, phenomenon complexity, and regional applicability.

KEYWORDS: Statistical Correlation; Geotechnical Properties; Machine Learning; Soil Characterization; Predictive Methods.

1 INTRODUÇÃO

A caracterização adequada das propriedades geotécnicas dos solos é essencial para o projeto seguro e econômico de obras de engenharia civil, incluindo fundações, estruturas de contenção, pavimentos e obras subterrâneas. Propriedades como resistência ao cisalhamento, compressibilidade, permeabilidade e rigidez dinâmica tradicionalmente requerem ensaios laboratoriais ou de campo especializados, que demandam tempo, recursos financeiros significativos e expertise técnica (Mola-Abasi & Shooshpasha, 2016; Xu et al., 2025). Em contextos onde recursos são limitados ou investigações geotécnicas abrangentes são impraticáveis, correlações estatísticas entre propriedades de fácil obtenção e parâmetros geotécnicos complexos tornam-se ferramentas valiosas para estudos e anteprojetos (Chatterjee et al., 2024; Mola-Abasi et al., 2015; Puri et al., 2018).

Historicamente, a engenharia geotécnica tem se baseado em correlações empíricas simples, frequentemente estabelecidas por meio de regressão linear (Burt et al., 2002; Chatterjee et al., 2024). Embora úteis, essas correlações tradicionais apresentam limitações significativas: assumem relações lineares que podem não capturar a complexidade do comportamento dos solos e podem apresentar baixa precisão quando aplicadas fora de suas condições de calibração (Chakraborty et al., 2020; Li et al., 2022; Wei et al., 2024).

Nas últimas duas décadas, o avanço de métodos de Machine Learning (ML) tem revolucionado o desenvolvimento de correlações geotécnicas (Chakraborty et al., 2020; Guo et al., 2025; Li et al., 2022; Puri et al., 2018). Algoritmos como Random Forest, Support Vector Machines (SVM), XGBoost e métodos Bayesianos demonstraram capacidade superior para modelar relações não-lineares complexas e fornecer predições mais precisas (Chakraborty et al., 2020; Li et al., 2022; Puri et al., 2018; Xu et al., 2025). Estudos recentes ilustram essa evolução: Chakraborty et al. (2020) demonstraram que XGBoost supera métodos tradicionais na predição de propriedades de solos; Cheng et al. (2024) introduziram o índice de comportamento do solo em correlações SPT-CPT; Xu et al. (2025) aplicaram modelos Bayesianos para reconstruir matrizes de correlação a partir de dados incompletos; e Mola-Abasi et al. (2015, 2016) utilizaram redes neurais polinomiais para aprimorar correlações tradicionais.

Diante desse cenário, o presente artigo tem como objetivo realizar uma revisão sistemática da literatura para identificar, caracterizar e analisar criticamente as principais aplicações, métodos e confiabilidade das análises de correlação estatística utilizadas para caracterização e predição de propriedades geotécnicas. A metodologia adotada segue rigorosamente os princípios de uma revisão sistemática, visando responder à questão: "Quais são as principais análises de correlação estatística aplicadas na avaliação de propriedades geotécnicas e qual sua confiabilidade para caracterização de propriedades dos solos?"

2 METODOLOGIA

Este estudo foi conduzido como Revisão Sistemática da Literatura (RSL), seguindo metodologia rigorosa para identificar, selecionar, analisar e sintetizar o conhecimento existente sobre análises de correlação estatística aplicadas à caracterização de propriedades geotécnicas. A abordagem metodológica foi alinhada com as diretrizes de revisões sistemáticas, garantindo transparência e reprodutibilidade dos resultados.

A questão de pesquisa central que guiou esta revisão foi: "Quais são as principais análises de correlação estatística aplicadas na avaliação de propriedades geotécnicas e qual sua confiabilidade para caracterização de propriedades dos solos?" Esta questão foi formulada para identificar métodos estatísticos e de *Machine Learning* empregados, avaliar sua precisão e confiabilidade, e sintetizar as melhores práticas para aplicação em contextos geotécnicos diversos.

Para estruturar a revisão e definir os critérios de inclusão e exclusão, foi utilizado o framework PICOC: População - solos e materiais geotécnicos de diferentes origens e características; Intervenção - análises de correlação estatística (Pearson, Spearman, regressão linear/não-linear) e modelos de *Machine Learning* (RF,

SVM, ANN, XGBoost); Comparação - métodos diretos de ensaio *versus* correlações indiretas, diferentes técnicas estatísticas e algoritmos de ML, modelos locais *versus* globais; *Outcomes* - precisão (R^2, RMSE, MAE), confiabilidade, aplicabilidade e vantagens/limitações das correlações para predição de propriedades geotécnicas; Contexto - projetos de engenharia geotécnica, caracterização de materiais, investigações de campo e laboratório.

A estratégia de busca utilizou a *string*: ("statistical correlation" OR "correlation analysis") AND ("geotechnical properties" OR "soil properties") AND ("database" OR "characterization" OR "prediction") aplicada nas bases de dados *Scopus, Web of Science, Engineering Village* (EI Compendex) e *Science Direct*, sem restrições de ano de publicação, para artigos em inglês e português.

A seleção dos estudos seguiu o protocolo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) em duas etapas sequenciais. Na primeira, todos os registros foram submetidos à análise inicial baseada em título e resumo para identificar estudos potencialmente relevantes. Na segunda etapa, os artigos pré-selecionados foram lidos integralmente para confirmar elegibilidade. Os critérios de inclusão consideraram artigos que abordam correlações estatísticas ou ML para predição de propriedades geotécnicas, estudos com validação estatística clara (R^2, RMSE, MAE, validação cruzada) e publicações em periódicos revisados por pares ou anais de conferências com relevância científica. Foram excluídos estudos focados exclusivamente em aspectos ambientais ou agrícolas sem relevância geotécnica direta e registros duplicados. Para cada artigo selecionado, procedeu-se à extração sistematizada de informações: título, autores, ano, método estatístico ou algoritmo empregado, tamanho e características do banco de dados (tipo de solo, localização), métricas de validação, propriedades geotécnicas investigadas e principais resultados e inovações metodológicas.

A qualidade foi avaliada por cinco critérios pontuados de 1 a 5: definição do método (clareza na descrição da técnica), robustez do banco de dados (adequação do tamanho amostral e descrição da população), validação estatística (métricas de desempenho claras), relevância geotécnica (foco direto em propriedades de engenharia) e inovação metodológica (comparação entre métodos ou melhorias sobre correlações tradicionais). A soma totalizou máximo de 25 pontos para classificar a contribuição dos estudos.

A análise dos 27 artigos selecionados foi qualitativa, envolvendo síntese de metodologias e contribuições, e quantitativa, focando em padrões metodológicos, distribuição temporal e geográfica, métricas de desempenho reportadas e evolução das técnicas estatísticas, formando a base para a seção de resultados deste artigo.

3 RESULTADOS

Esta seção apresenta os resultados detalhados da revisão sistemática, com foco na análise bibliométrica, caracterização metodológica e síntese das principais contribuições dos 27 artigos selecionados.

3.1 Processo de Seleção dos Estudos

O processo de busca nas quatro bases de dados (*Scopus, Web of Science, Engineering Village* e *ScienceDirect*) resultou em um total de 507 registros. A distribuição inicial por base foi: *Scopus* (n = 254), Web of Science (n = 107), Engineering Village (n = 67) e ScienceDirect (n = 79).

Após a importação dos registros para o *software* de gerenciamento bibliográfico Parsif.al, foram identificados e removidos 261 artigos duplicados, resultando em 246 estudos únicos. Na etapa de triagem baseada em títulos, foram excluídos 191 artigos que claramente não atendiam aos objetivos da revisão. Posteriormente, a triagem baseada nos resumos (*abstracts*) eliminou 25 artigos adicionais, resultando em 30 estudos elegíveis para leitura completa.

Durante a etapa de leitura na íntegra, nenhum artigo foi excluído por problemas metodológicos ou falta de informações relevantes. Entretanto, foram removidos 3 artigos por estarem em idiomas que não português ou inglês, impossibilitando sua análise adequada.

Ao final do processo de triagem PRISMA, 27 artigos foram selecionados para compor a revisão sistemática e foram submetidos à análise detalhada e extração de dados.

3.2 Análise Bibliométrica

A análise bibliométrica foi conduzida com os 27 artigos incluídos, visando caracterizar o panorama das publicações sobre correlações estatísticas e modelos de aprendizado de máquina aplicados a propriedades geotécnicas. Esta análise permitiu identificar padrões de publicação, principais métodos empregados, distribuição temporal e geográfica das pesquisas, bem como as tendências emergentes na área de predição de propriedades geotécnicas de solos.

A distribuição temporal das publicações e por base indexadora (Figura 1) demonstra um crescimento significativo no interesse científico pelo tema nas últimas duas décadas. Observa-se uma concentração de publicações a partir de 2015, com picos notáveis entre 2022-2025, refletindo a crescente adoção de técnicas de Machine Learning e análise de grandes conjuntos de dados (big data) na engenharia geotécnica. Este padrão temporal evidencia a transição de métodos estatísticos clássicos para abordagens computacionais avançadas, impulsionada tanto pela disponibilidade de dados quanto pelo desenvolvimento de ferramentas analíticas mais sofisticadas. A distribuição por base indexadora revelou que *Scopus* e *Web of Science* foram as fontes primárias mais relevantes, com vários artigos indexados exclusivamente em uma dessas bases ou em ambas simultaneamente. A base EI Compendex apresentou sobreposição significativa com as anteriores, confirmando-se como complementar mas não essencial para cobertura abrangente. Este resultado sugere que pesquisas futuras na área podem concentrar esforços em Scopus e Web of Science para máxima eficiência na busca bibliográfica.

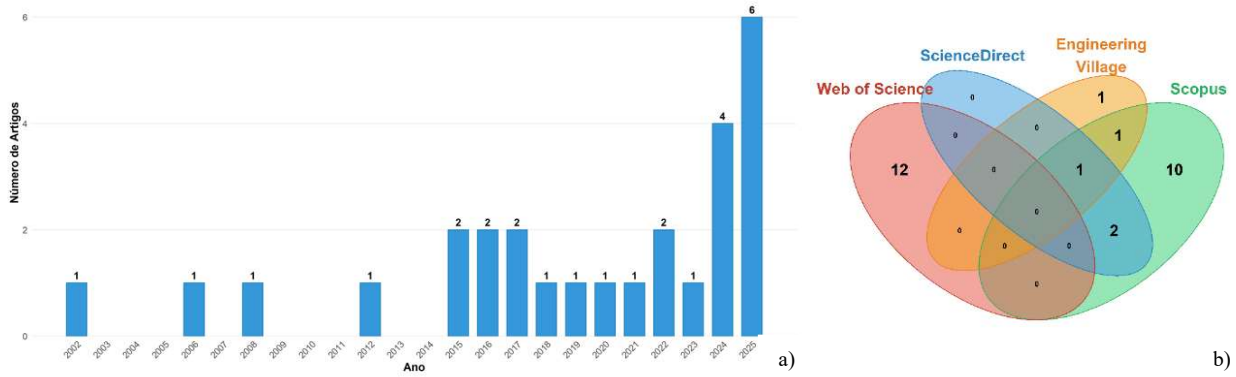


Figura 1: a) Produções por ano; b) Distribuição de artigos por base indexadora (Imagens produzidas no programa RStudio).

A distribuição geográfica das pesquisas (Figura 2) revela uma representação global significativa, com estudos conduzidos em diversos continentes. Destacam-se contribuições da Ásia (China, Índia, Irã, Turquia, Sri Lanka) (Chakraborty et al., 2020; Chatterjee et al., 2024; Cheng et al., 2024; Mola-Abasi et al., 2015; Mola-Abasi & Shooshpasha, 2016; Xu et al., 2025; Şahin & Gündüz, 2025), Europa (Chipre, Itália, Espanha, Croácia) (Micó et al., 2006; Tzampoglou et al., 2025; Viti & Tanganelli, 2017), América do Sul (Peru) (Araujo & Ruiz, 2025), América do Norte (Canadá e EUA) (Biswas & Si, 2012; Burt et al., 2002) e África (Quênia e Camarões) (Ng et al., 2024; Yannick et al., 2022). Esta diversidade geográfica demonstra a relevância universal do tema e a necessidade de desenvolver correlações adaptadas a diferentes contextos geológicos, climáticos e práticas de engenharia locais.

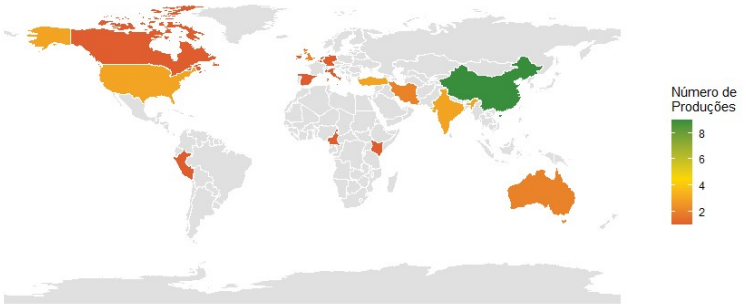


Figura 2. Mapa de distribuição de artigos por país (Imagem produzida no programa RStudio).

Em termos de colaborações internacionais e análise de frequência dos termos-chave (Figura 3), identificaram-se parcerias entre instituições de diferentes países em diversos estudos, particularmente em pesquisas envolvendo grandes bancos de dados ou validação de modelos em múltiplos contextos geográficos (Ng et al., 2024; Xu et al., 2025; Şahin & Gündüz, 2025). Esta tendência colaborativa reflete a natureza cada vez mais interdisciplinar e globalizada da pesquisa em geotecnia, facilitada por plataformas digitais de compartilhamento de dados e cooperação científica.

A análise de frequência dos termos-chave mais relevantes identificou os conceitos centrais que permeiam as pesquisas na área. Para esta análise, aplicou-se a técnica de *stemming* (redução de palavras à raiz comum), permitindo agrupar variações morfológicas de um mesmo termo, como "correlation" e "correlations", que foram unificadas como "correlation". Os termos mais frequentes foram "soil" (17 ocorrências), "correlation" (8 ocorrências), "analysis" e "model" (7 ocorrências cada), "compression" e "statistical" (4 ocorrências cada), confirmando que a predição de propriedades do solo por meio de correlações estatísticas e modelagem é o tema unificador central. Termos com frequência moderada incluem "test", "clay", "distribution", "heavy metal", "learning", "penetration", "prediction" e "settlement" (3 ocorrências cada), evidenciando o foco em propriedades mecânicas fundamentais, ensaios de campo e aplicações em contaminação do solo.

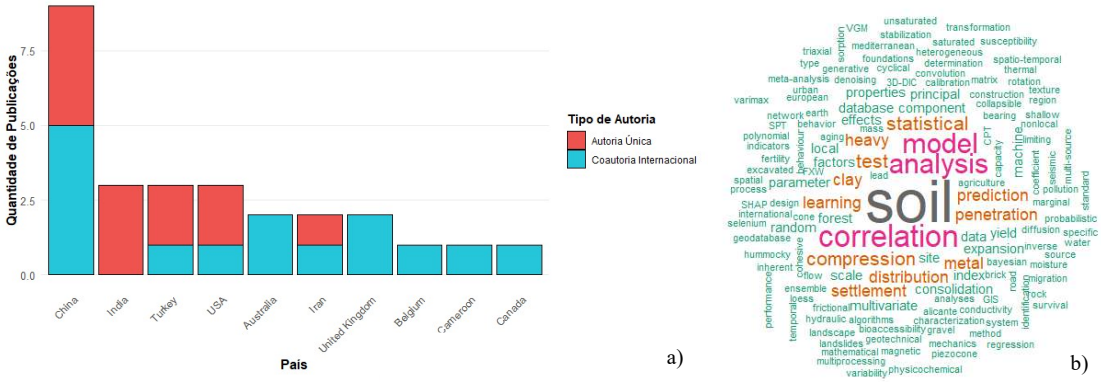


Figura 3: a) Gráfico de colaborações internacionais; b) Nuvem de palavras-chave (Imagens produzidas no programa RStudio).

Termos metodológicos específicos revelam a amplitude de técnicas empregadas: "machine learning" e "random forest" (2 ocorrências cada), "principal component analysis", "multivariate", "bayesian", "SHAP" (SHapley Additive exPlanations) e "ensemble" (presentes na base de dados). Esta diversidade metodológica demonstra a evolução da área, transitando de análises de correlação bivariada simples para modelagens multivariadas complexas e algoritmos de aprendizado de máquina interpretáveis. A presença significativa de termos relacionados a ensaios *in situ* ("SPT", "CPT", "cone penetration", "standard penetration") e propriedades físico-químicas ("physicochemical", "texture", "moisture") indica a integração entre dados de campo, laboratório e características intrínsecas do solo.

3.3 Caracterização Metodológica e Análise por Propriedade Geotécnica

A análise dos 27 artigos revelou cinco categorias metodológicas principais. Correlações Lineares Tradicionais (8 estudos) empregam regressão linear e correlações de Pearson/Spearman, caracterizando-se por simplicidade mas apresentando R^2 moderado (0,46-0,75) e limitações para relações não-lineares (Araujo & Ruiz, 2025; Güner & Özgan, 2025). Regressão Não-Linear e Polinomial (6 estudos) demonstram melhorias substanciais, com Güner & Özgan (2025) reportando aumento de R^2 de 0,46 para 0,90 na predição de recalques. Machine Learning Supervisionado (11 estudos) apresenta desempenho superior consistente, com Chakraborty et al. (2020) demonstrando que XGBoost captura não-linearidades melhor que regressão tradicional, e Li et al. (2022) mostrando superioridade do AdaBoost sobre modelos empíricos clássicos. Análise Multivariada (5 estudos) utiliza PCA e geoestatística para identificar padrões em dados complexos (Lu et al., 2023; Ng et al., 2024). Métodos Bayesianos (3 estudos) destacam-se pela capacidade de quantificar incertezas e reconstruir dados incompletos (Xu et al., 2025; Zou et al., 2017).

Para resistência ao cisalhamento, Puri et al. (2018) demonstraram que M5 Model Tree e SVM superaram regressão linear para predição de ϕ e c, enquanto Mola-Abasi et al. (2015) utilizaram GMDH para correlacionar Vs (velocidade da onda de cisalhamento) com CPT, reduzindo drasticamente o erro. Em compressibilidade, Chatterjee et al. (2024) provaram que Pc (tensão de pré-adensamento) é o melhor preditor individual para Cc (índice de compressão), e Xu et al. (2025) aplicaram modelos Bayesianos para reconstruir matriz de correlação completa a partir de dados incompletos. Para correlações SPT-CPT, Cheng et al. (2024) desenvolveram modelo de curva-S com índice Ic que superou 10 modelos existentes, especialmente para solos mistos. Em propriedades térmicas e hidráulicas, Li et al. (2022) mostraram que AdaBoost superou modelos empíricos para condutividade térmica. Aplicações emergentes incluem uso de IA generativa para correlacionar imagens com propriedades mecânicas (Guo et al., 2025) e SHAP para explicar influência de variáveis em predições (Tzampoglou et al., 2025).

A análise comparativa revelou superioridade consistente de ML sobre regressão linear, importância de modelos não-lineares mesmo sem ML, e valor da validação rigorosa. Estudos com maior impacto (pontuação $\geq$24) compartilham bancos de dados robustos, técnicas avançadas e validação estatística rigorosa (Chakraborty et al., 2020; Cheng et al., 2024; Xu et al., 2025; Zou et al., 2017). Limitações de amostras pequenas ($n<50$) resultaram em R^2 inferior e maior risco de overfitting (Biswas & Si, 2012; Güner & Özgan, 2025). Correlações desenvolvidas localmente superaram modelos globais genéricos, destacando a importância da calibração regional (Cheng et al., 2024; Zou et al., 2017).

4 CONSIDERAÇÕES FINAIS

Esta revisão sistemática da literatura demonstrou a evolução significativa das análises de correlação estatística aplicadas à caracterização de propriedades geotécnicas. A análise detalhada dos 27 artigos selecionados revelou uma transição clara de métodos estatísticos clássicos para abordagens computacionais avançadas que incorporam *Machine Learning*, análise multivariada e modelagem de incertezas.

Os resultados evidenciam que algoritmos de ML como XGBoost, Random Forest, SVM e AdaBoost consistentemente superam correlações empíricas tradicionais na predição de propriedades geotécnicas complexas, capturando relações não-lineares que métodos convencionais não conseguem representar adequadamente. Estudos demonstraram reduções substanciais em erros de predição, com aumentos em R^2 de 0,46 para 0,90 ao transitar de modelos lineares para cúbicos ou baseados em ML (Güner & Özgan, 2025).

Identificou-se que a contextualização regional permanece essencial, com correlações desenvolvidas para condições específicas superando modelos globais genéricos (Zou et al., 2017). Simultaneamente, métodos Bayesianos demonstraram potencial para maximizar o uso de bancos de dados incompletos (Xu et al., 2025), problema comum em geotecnia.

Mesmo com os avanços apresentados, recomenda-se que os dados e correlações obtidos por meio dessas metodologias sejam aplicados com critério e parcimônia na prática geotécnica. O solo é um material de natureza intrinsecamente heterogênea e variável, cujo comportamento é condicionado por fatores intrínsecos como mineralogia e por fatores extrínsecos relacionados ao processo de investigação. Correlações estatísticas e modelos de ML, por mais robustos que sejam, são desenvolvidos a partir de conjuntos de dados específicos e podem não capturar adequadamente a variabilidade natural de perfis geológicos distintos daqueles utilizados

no treinamento. Portanto, sua utilização deve ser acompanhada de análise crítica por parte do engenheiro, validação com dados locais sempre que possível e consciência das limitações inerentes a qualquer modelo preditivo.

Em suma, a integração de métodos estatísticos avançados e *Machine Learning* representa uma mudança paradigmática na caracterização geotécnica, oferecendo ferramentas mais precisas e confiáveis para engenheiros. A seleção adequada do método deve considerar o tamanho amostral disponível, a complexidade do fenômeno e a aplicabilidade regional. O futuro da área aponta para maior adoção de IA explicável (Tzampoglou et al., 2025; Wei et al., 2024), integração de múltiplas fontes de dados (Guo et al., 2025; Ng et al., 2024) e desenvolvimento de plataformas abertas de compartilhamento de dados geotécnicos.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa e produtividade em pesquisa da coautora M.M.A. Mascarenha (Processo: 304529/2025-7).

REFERÊNCIAS BIBLIOGRÁFICAS

Araujo, W.; Ruiz, G. (2025). *Preliminary Analysis of the Identification of Key Parameters in Expansive Clay Soil.* In: International Conference on Civil Structural and Transportation Engineering (ICCSTE 2025), 10., London. *Proceedings...* London: Avestia Publishing. Paper No. 310. DOI: 10.11159/iccste25.310.

Biswas, A.; Si, B. C. (2012). *Identifying effects of local and nonlocal factors of soil water storage using cyclical correlation analysis. Hydrological Processes,* v. 26, n. 24, p. 3670–3677. DOI: 10.1002/hyp.8459.

Burt, R. et al. (2002). *Phosphorus characterization and correlation with properties of selected benchmark soils of the United States. Communications in Soil Science and Plant Analysis,* v. 33, n. 1-2, p. 117–141. DOI: 10.1081/CSS-120002382.

Chakraborty, P. et al. (2020). *Spatial structure, parameter nonlinearity, and intelligent algorithms in constructing pedotransfer functions from large-scale soil legacy data. Scientific Reports,* v. 10, art. 13576. DOI: 10.1038/s41598-020-72018-2.

Chatterjee, S. et al. (2024). *Estimation of Compression Index of Cohesive-Frictional Soil in Southwest Delhi District Using Statistical Correlations. Journal of the Geological Society of India,* v. 100, n. 6, p. 841–851. DOI: 10.17491/jgsi/2024/173914.

Cheng, X. et al. (2024). *A new SPT-CPT correlation by statistical analysis based on a database from Anhui, China. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards,* v. 18, n. 4, p. 838–853. DOI: 10.1080/17499518.2024.2379965.

Guo, Q.-M. et al. (2025). *Correlation of excavated soil multi-source heterogeneous data using multimodal diffusion model. Acta Geotechnica.* DOI: 10.1007/s11440-025-02690-z.

Güner, A. B. S.; Özgan, E. (2025). *Statistical Analysis of Soil Parameters Affecting the Bearing Capacity and Settlement Behaviour of Gravel Soils. Applied Sciences,* v. 15, n. 10, art. 5271. DOI: 10.3390/app15105271.

Li, K.-Q.; Liu, Y.; Kang, Q. (2022). *Estimating the thermal conductivity of soils using six machine learning algorithms. International Communications in Heat and Mass Transfer,* v. 136, art. 106139. DOI: 10.1016/j.icheatmasstransfer.2022.106139.

Liu, D. et al. (2016). *Spatial distribution of soil magnetic susceptibility and correlation with heavy metal pollution in Kaifeng City, China. Catena,* v. 139, p. 53–60. DOI: 10.1016/j.catena.2015.11.004.

Lu, X. et al. (2023). *A global meta-analysis of the correlation between soil physicochemical properties and lead bioaccessibility. Journal of Hazardous Materials,* v. 453, art. 131440. DOI: 10.1016/j.jhazmat.2023.131440.

Micó, C. et al. (2006). *Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis. Chemosphere,* v. 65, n. 5, p. 863–872. DOI: 10.1016/j.chemosphere.2006.03.016.

Mola-Abasi, H.; Dikmen, U.; Shooshpasha, I. (2015). *Prediction of shear-wave velocity from CPT data at Eskisehir (Turkey), using a polynomial model. Near Surface Geophysics,* v. 13, n. 2, p. 155–168. DOI: 10.3997/1873-0604.2015010.

Mola-Abasi, H.; Shooshpasha, I.; Ebrahimi, A. (2016). *Prediction of the compression index of saturated clays (Cc) using polynomial models. Scientia Iranica. Transactions A: Civil Engineering,* v. 23, n. 2, p. 500–507. DOI: 10.24200/sci.2016.2134.

Ng, W. et al. (2024). *Exploring mid-infrared spectral transfer functions for the prediction of multiple soil properties using a global dataset. Soil Science Society of America Journal,* v. 88, n. 4, p. 1234–1247. DOI: 10.1002/saj2.20697.

Puri, N.; Prasad, H. D.; Jain, A. (2018). *Prediction of Geotechnical Properties using Machine Learning Techniques. Procedia Computer Science,* v. 125, p. 509–517. DOI: 10.1016/j.procs.2017.12.066.

Şahin, O. G.; Gündüz, O. (2025). *Derivation of soil hydraulic properties (SHPs) using a Physics-Based inverse calibration method and International soil moisture network database. Journal of Hydrology,* v. 660, art. 133445. DOI: 10.1016/j.jhydrol.2025.133445.

Tzampoglou, P. et al. (2025). *Correlation Between Geotechnical Indexes and Landslide Occurrence in Southwestern Cyprus Using GIS and Machine Learning. Geotechnical and Geological Engineering,* v. 43, art. 43. DOI: 10.1007/s10706-024-03031-9.

Viti, S.; Tanganelli, M. (2017). *Effects of soil characterization on the seismic input.* In: International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2017), 6., Rhodes. *Eccomas Proceedia...* p. 1007–1017. DOI: 10.7712/120117.5472.17353.

Wei, W. et al. (2024). *Characterization and modeling of exogenous selenite aging in soils using machine learning and traditional data analysis. Geoderma,* v. 451, art. 117074. DOI: 10.1016/j.geoderma.2024.117074.

Whetton, R. L. et al. (2021). *Modelling the Influence of Soil Properties on Crop Yields Using a Non-Linear NFIR Model and Laboratory Data. Soil Systems,* v. 5, n. 1, art. 12. DOI: 10.3390/soilsystems5010012.

Xiao, H.; Böttcher, J.; Utermann, J. (2015). *Evaluation of field-scale variability of heavy metal sorption in soils by scale factors — Scaling approach and statistical analysis. Geoderma,* v. 241–242, p. 115–125. DOI: 10.1016/j.geoderma.2014.11.012.

Xu, J. et al. (2025). *Development of a database on multivariate soil properties for collapsible loess in Xi'an, China. Bulletin of Engineering Geology and the Environment,* v. 84, art. 245. DOI: 10.1007/s10064-025-04265-4.

Yannick, T. L. et al. (2022). *Statistical analysis of Nkoulou soils properties and suitability for earthen constructions. Heliyon,* v. 8, n. 5, e11141. DOI: 10.1016/j.heliyon.2022.e11141.

Zhang, C. et al. (2008). *Identifying sources of soil inorganic pollutants on a regional scale using a multivariate statistical approach: Role of pollutant migration and soil physicochemical properties. Environmental Pollution,* v. 151, n. 3, p. 470–476. DOI: 10.1016/j.envpol.2007.04.017.

Zhu, Y.; Medina-Cetina, Z.; Pineda-Contreras, A. R. (2022). *Spatio-Temporal Statistical Characterization of Boundary Kinematic Phenomena of Triaxial Sand Specimens. Materials,* v. 15, n. 6, art. 2189. DOI: 10.3390/ma15062189.

Zou, H. et al. (2017). Multivariate correlation analysis of seismic piezocone penetration (SCPTU) parameters and design properties of Jiangsu quaternary cohesive soils. Engineering Geology, v. 228, p. 11–38. DOI: 10.1016/j.enggeo.2017.07.005.