

CAPÍTULO 2

MODELOS MISTOS NA ANÁLISE DA RESISTÊNCIA MECÂNICA DO SOLO

Data de submissão: 07/01/2025

Data de aceite: 05/03/2025

Leonardo Ribeiro

Universidade Federal de São Carlos - UFSCar, São Carlos, SP, Brasil

Murilo Mesquita Baesso

Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo - USP, Pirassununga, SP, Brasil

Jamile Raquel Regazzo

Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo - USP, Piracicaba, SP, Brasil

Cíntia Cristina Soares

Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo - USP, Piracicaba, SP, Brasil

Luís Fernando Soares Zuin

Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo - USP, Pirassununga, SP, Brasil

Daiane Aparecida Zuanetti

Universidade Federal de São Carlos - UFSCar, São Carlos, SP, Brasil

Financiamentos: Esse estudo foi financiado pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) no projeto número 2024/00413-8, Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ - código 001,

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES - Code 001 e pela Fundação de Estudos Agrários Luiz de Queiroz - FEALQ.

RESUMO: O solo é fundamental para a agricultura, sendo essencial para o desenvolvimento das plantações, e sua qualidade tem sido amplamente avaliada através da sua resistência mecânica devido à rapidez e facilidade de uso. Este estudo, portanto, investiga através dos modelos mistos de que forma os fatores como área, umidade e tipo de penetrômetro influenciam a resistência mecânica do solo. Os modelos mistos são brevemente descritos e aplicados em 1103 unidades experimentais utilizando o software R. Os resultados indicam que os fatores área, umidade e tipo de penetrômetro, bem como suas interações, são significativos para explicar a resistência média do solo. O penetrômetro de impacto apresentou as maiores resistências médias, enquanto o manual mostrou as menores. Solos encharcados exibiram menor resistência comparados a solos secos e úmidos. O modelo linear misto mostrou bom ajuste e suposições diagnósticas adequadas, identificando algumas observações atípicas nas áreas 4 e 5.

PALAVRAS-CHAVE: Análise do solo; fatores de risco e proteção; efeitos aleatórios.

MIXED MODELS IN THE ANALYSIS OF SOIL MECHANICAL STRENGTH

ABSTRACT: Soil is fundamental to agriculture, being essential for the development of crops, and its quality has been widely assessed through its mechanical resistance due to its speed and ease of use. This study, therefore, investigates through mixed models how factors such as area, moisture and type of penetrometer influence soil mechanical resistance. The mixed models are briefly described and applied in 1103 experimental units using the R software. The results indicate that the factors area, moisture and type of penetrometer, as well as their interactions, are significant in explaining the average soil resistance. The impact penetrometer showed the highest average resistances, while the manual penetrometer showed the lowest. Waterlogged soils exhibited lower resistance compared to dry and moist soils. The linear mixed model showed good fit and adequate diagnostic assumptions, identifying some atypical observations in areas 4 and 5.

KEYWORDS: Soil analysis; risk and protective factors; random effects.

INTRODUÇÃO

Em termos econômicos, a agricultura é um dos setores mais responsáveis pelo crescimento do PIB no país, sendo que ela corresponde a 21% de todas as riquezas nacionais produzidas, um quinto de todos os empregos e 43.2% das exportações brasileiras, chegando a US\$ 96.7 bilhões em 2019 (Embrapa, 2020). Com o passar dos anos, a produção alimentícia também apresentou um amplo crescimento, sendo que ao longo das últimas quatro décadas a produção de grãos exibiu um aumento de 510% (232.6 milhões de toneladas) e a produção de carnes obteve um salto de 858% (27.9 milhões de toneladas). Além dessas, outras produções também alcançaram grandes saltos de produtividade (Embrapa, 2020).

A principal matéria-prima para a agricultura é o solo que é essencial para o desenvolvimento das plantações devido à sua riqueza em nutrientes e suas funções (filtragem d'água, decomposição de resíduos, armazenamento de calor e troca de gases). Em razão dessa importância, é vital que haja um grande cuidado na conservação desse recurso para que seja possível aplicar uma agricultura produtiva e sustentável (Brasil, 2020).

O Brasil detém uma vasta diversidade de tipos de solo e cada um necessita de uma atenção especial para que a boa qualidade dessa matéria-prima seja mantida. Entretanto, o que vem acontecendo atualmente é a degradação desse recurso. De acordo com um relatório recente das Nações Unidas, quase um terço das terras cultiváveis do mundo desapareceu nas últimas quatro décadas. Também foi identificado que todo o solo superficial do mundo poderá se tornar improdutivo dentro de 60 anos se as taxas atuais de perda continuarem (Brasil, 2020).

Com o desenvolvimento mecânico da agricultura, o número de máquinas de grande porte intensificou-se sobre o solo das lavouras (Hussein *et al.*, 2023). Por conta disso, acentuou-se uma das principais causas da degradação desse recurso, a compactação. O excesso de manipulação do solo por uso constante de máquinas agrícolas e pisoteio de animais pesados faz com que, devido à pressão causada, ocorra uma diminuição do seu volume não saturado, causando assim, a expulsão do ar do solo e, consequentemente, um aumento de densidade.

A compactação leva ao aumento da resistência mecânica do solo no crescimento das raízes das plantas (Colombi e Keller, 2019). Devido ao aumento de densidade causada pela pressão no solo, seu rompimento pelas raízes das plantas na profundidade fica prejudicado e, em muitos casos, a planta não consegue crescer sua raiz totalmente, além de ocorrer a diminuição do ar no solo. Isso pode causar uma morte nas raízes por asfixia e também ter um baixo acesso à água e nutrientes, ocasionando no não crescimento da planta (Akhmetov *et al.*, 2024).

Uma das técnicas mais utilizadas para a avaliação do solo é a análise da sua resistência mecânica à penetração. Essa técnica é a preferida para esse processo pela rapidez na verificação e pelo fácil manuseamento da ferramenta utilizada: o penetrômetro. Menezes (2018) relata um projeto e resultados conduzidos com o objetivo de comparar os diferentes tipos de penetrômetros, além de investigar a influência de características texturais, estrutura e umidade nas leituras da resistência mecânica do solo à penetração. O pesquisador realiza uma análise de variância (ANOVA) para a obtenção de seus resultados considerando independência entre os dados. Porém, os dados são dependentes, por se tratar de medições em diferentes profundidades no mesmo solo e local, e a metodologia adotada pode não ser adequada. Um dos modelos mais tradicionais e flexíveis para analisar dados dependentes de medidas repetidas ou longitudinais é o modelo misto (Singer *et al.*, 2018; Zhang, 2015).

O modelo misto é um modelo de regressão que considera seus coeficientes de maneira fixa e aleatória. Pressupor esses tipos de coeficientes permite uma correlação entre as variáveis respostas de uma mesma unidade experimental, ou seja, uma dependência nos dados. Também permite variâncias heterogêneas para diferentes observações e, consequentemente, uma melhor precisão nos resultados e fácil interpretabilidade (Visalli *et al.*, 2024).

Dessa maneira, este trabalho tem como objetivo comparar a performance dos penetrômetros e também verificar a influência de características como umidade, tipo de solo e profundidade nas leituras da resistência mecânica do solo à penetração por meio da utilização do modelo misto. O conjunto de dados a ser analisado é o citado acima e disponível em Menezes (2018).

Este relatório está organizado como a seguir. A Seção 2 apresenta e descreve as principais características dos modelos mistos, metodologia estatística a ser utilizada nesse estudo. A Seção 3 mostra como foram coletados os dados, as variáveis a serem trabalhadas no decorrer do estudo e, por último, exibe uma análise inicial para os efeitos principais, algumas interações duplas e triplas. A Seção 4 apresenta os principais resultados e, finalmente, na Seção 5 apresentamos as conclusões e estudos futuros.

MODELOS MISTOS

Alguns modelos estatísticos assumem que os dados não possuem correlação entre si, ou seja, há independência entre eles, e também, que os erros são provenientes de uma distribuição $Normal(0, \sigma^2)$, em que σ^2 é um número real positivo e constante. Porém, o trabalho tem como intuito analisar medições em 8 camadas de profundidade diferentes em determinados solos e locais. Conduzir esse experimento, nessas condições, não nos garante que os dados são independentes pois são longitudinais em relação às profundidades consideradas nas medições e esses apresentam uma certa correlação entre as camadas, principalmente as adjacentes. Além disso, podemos ter heterocedasticidade nas medições por causa dos níveis de profundidade e compactações do solo observadas em cada nível. Nesse cenário, a princípio, o modelo misto apresenta as propriedades adequadas para descrever o conjunto de dados e também possui parâmetros de fácil interpretação.

O modelo misto é um modelo de regressão que apresenta efeitos fixos e aleatórios. Os efeitos fixos são vistos como parâmetros desconhecidos, entretanto, comuns a todas as observações e que impactam na média da variável resposta. Já os aleatórios, são vistos como uma variável aleatória podendo resultar em qualquer valor para cada unidade amostral, além de impactarem diretamente na variância da variável de interesse. O modelo, para cada unidade amostral, pode ser representado pela Equação (1) como:

$$\mathbf{Y}_i^{m_i \times 1} = \mathbf{X}_i^{m_i \times p} \times \boldsymbol{\beta}^{p \times 1} + \mathbf{Z}_i^{m_i \times q} \times \mathbf{b}_i^{q \times 1} + \mathbf{e}_i^{m_i \times 1}, \text{ para } i = 1, \dots, n, \quad (1)$$

em que \mathbf{Y}_i é o vetor aleatório com as variáveis respostas da i -ésima unidade amostral com dimensão ($m_i \times 1$) e m_i é o número de observações nessa unidade amostra; $\boldsymbol{\beta}$ é o vetor de efeitos fixos do modelo com dimensão ($p \times 1$); \mathbf{X}_i é a matriz de planejamento associada a $\boldsymbol{\beta}$ com dimensão ($m_i \times p$), e contém as observações das covariáveis associadas a efeitos fixos da i -ésima unidade amostral; \mathbf{b}_i é o vetor dos efeitos aleatórios associados à i -ésima unidade amostral com dimensão ($q \times 1$), tradicionalmente assumimos que $\mathbf{b}_i \sim Normal_q(\mathbf{0}, \mathbf{G})$ e \mathbf{G} é a matriz de covariâncias para os efeitos aleatórios; \mathbf{Z}_i é a matriz de planejamento associada a \mathbf{b}_i com dimensão ($m_i \times q$), e contém as covariáveis associadas a efeitos aleatórios da i -ésima unidade amostral; e \mathbf{e}_i é o vetor dos erros aleatórios da i -ésima unidade amostral com dimensão ($m_i \times 1$), geralmente assumimos que $\mathbf{e}_i \sim Normal_{m_i}(\mathbf{0}, \mathbf{R}_i)$, $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ e \mathbf{I} é a matriz identidade.

Frequentemente, adotamos a independência entre \mathbf{b}_i e \mathbf{e}_i para obtermos a média e variância do vetor de variáveis respostas facilmente. A normalidade nos erros e nos efeitos aleatórios também é tradicional por conta de ser uma distribuição muito conhecida, possuir propriedades desejáveis e termos um domínio algébrico sobre ela. Sob essas suposições, a esperança marginal de \mathbf{Y}_i é $E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}$ e sua variância marginal é $V(\mathbf{Y}_i) = \boldsymbol{\Omega}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{mr}$.

Podemos então notar que a variância marginal é uma soma de dois termos. O primeiro modela a dispersão dos perfis individuais de resposta em torno de um perfil médio definido pela parte fixa do modelo, enquanto o segundo termo da soma é relacionado com a dispersão dos valores observados em torno dos perfis individuais, isto é, a variabilidade das observações dentro de cada perfil (Singer *et al.*, 2018).

Como, para esse estudo, foram coletadas medições em 8 camadas de profundidade diferentes de solo, espera-se ter uma correlação (associação) entre as resistências observadas nessas camadas. Em termos do problema em questão, $m_i = 8$ e o q , número de efeitos aleatórios, será definido mediante análise descritiva dos dados e discutido mais à frente. Também é necessário definir a estrutura da matriz \mathbf{G} , ou seja, definir como as variâncias e covariâncias entre os efeitos aleatórios se comportam. Existem várias estruturas possíveis para \mathbf{G} visando estimar um modelo mais adequado mesmo com uma grande quantidade de parâmetros. Daremos mais detalhes da estrutura de covariância escolhida no decorrer do texto. Estruturas de covariância diferentes da homocedástica e independente também podem ser assumidas para os erros aleatórios.

Outro ponto a se destacar é a escolha dos efeitos fixos e aleatórios. Essa decisão não é simples, nem única e depende de outros fatores, tais como o tipo de estudo, objetivo do pesquisador e contexto dos dados. Alguns itens a serem considerados são:

- os efeitos que parecem ser constantes para toda a população são considerados fixos e aqueles que devem variar entre as unidades amostrais são aleatórios;
- colocar muitos efeitos aleatórios no modelo faz com que aumente a dificuldade para estimá-los. Por conta disso, muitos estudos consideram apenas o intercepto como efeito aleatório;
- é interessante ajustar diversos modelos e, via critério de seleção e análise de diagnóstico, escolher aquele que parece ser mais adequado para os dados trabalhados; e
- uma covariável pode apresentar efeitos tanto fixos quanto aleatórios simultaneamente. Assim, em geral, atribuímos efeitos fixos a todas as covariáveis, enquanto algumas delas também são designadas para ter efeitos aleatórios.

Vimos anteriormente no Modelo (1), o modelo misto para as medidas associadas a apenas uma das perfurações (unidade amostral) feitas no solo com diferentes profundidades observadas. Singer *et al.* (2018) apresenta o modelo compactado, dessa vez com todas as perfurações em que foram coletadas as amostras do estudo e também com a suposição de que \mathbf{b} é independente de \mathbf{s} . O modelo completo é definido por

$$\underset{N \times 1}{\mathbf{Y}} = \underset{N \times p}{\mathbf{X}} \times \underset{p \times 1}{\boldsymbol{\beta}} + \underset{N \times nq}{\mathbf{Z}} \times \underset{nq \times 1}{\mathbf{b}} + \underset{N \times 1}{\mathbf{e}}, \quad (2)$$

em que $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_n^T)^T$ com dimensão $(N \times 1)$, $N = m_1 + m_2 + \dots + m_n$, contém as respostas das n unidades amostrais; $\boldsymbol{\beta}$ é o vetor de efeitos fixos do modelo com dimensão $(p \times 1)$; $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$ com dimensão $(N \times p)$, é a matriz de planejamento associada a $\boldsymbol{\beta}$ e contém as covariáveis associadas a efeitos fixos; $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)^T$, com dimensão $(nq \times 1)$, é o vetor dos efeitos aleatórios associados às n unidades amostrais, $\mathbf{b} \sim \text{Normal}_{nq}(\mathbf{0}, \Gamma)$ em que $\Gamma = \mathbf{I}_n \otimes \mathbf{G}$ e \otimes representa o produto de Kronecker; $\mathbf{Z} = \bigoplus_{i=1}^n \mathbf{Z}_i$, com dimensão $(N \times nq)$, é a matriz de planejamento associada a \mathbf{b} e contém as covariáveis associadas a efeitos aleatórios, em que \bigoplus representa a soma direta de matrizes; e $\mathbf{e} = (\mathbf{e}_1^T, \dots, \mathbf{e}_n^T)^T$, com dimensão $(N \times 1)$, é o vetor dos erros aleatórios das n unidades com dimensão $(m_i \times 1)$, $\mathbf{e} \sim \text{Normal}_N(\mathbf{0}, \mathbf{R})$ com $\mathbf{R} = \bigoplus_{i=1}^n \mathbf{R}_i$.

Existem diversos métodos de estimação do modelo apresentado em (2), o método de máxima verossimilhança (MV) e o de máxima verossimilhança restrita (MVR) os mais tradicionais. O método de máxima verossimilhança restrita foi proposto por Patterson e Thompson (1971). O método consiste em maximizar a verossimilhança de uma transformação linear ortogonal dos dados com o objetivo de estimar termos da variância. A transformação utilizada é $\mathbf{Y}^t = \mathbf{U}^T \mathbf{Y}$ em que, no geral, usamos $\mathbf{U} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ e com dimensão $(N \times N)$ para se obter a verossimilhança restrita. Além disso, temos $E(\mathbf{Y}^t) = \mathbf{0}$, $\mathbf{U}^T \mathbf{X} = \mathbf{0}$ e $\mathbf{Y}^t \sim \text{Normal}_N(\mathbf{0}, \mathbf{U}^T \Omega \mathbf{U})$, ou seja, a distribuição marginal de \mathbf{Y}^t não depende de $\boldsymbol{\beta}$. Mais detalhes do processo de estimação podem ser encontrados em Zhang (2015) e Singer et al. (2018).

Os preditores dos efeitos aleatórios podem ser encontrados por meio da distribuição conjunta de \mathbf{Y} e \mathbf{b} dada por

$$f(\mathbf{y}, \mathbf{b}) = f(\mathbf{y} | \mathbf{b})f(\mathbf{b}), \quad (3)$$

em que $\mathbf{Y} | \mathbf{b} \sim \text{Normal}_N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R})$ e $\mathbf{b} \sim \text{Normal}_{nq}(\mathbf{0}, \Gamma)$.

No software R, um modelo misto pode ser estimado por MV ou MVR usando o pacote *nlme* de Pinheiro et al. (2023). Geralmente, se a convergência for alcançada, os métodos apresentam soluções muito semelhantes.

O EXPERIMENTO E OS DADOS

O experimento, por meio do qual os dados analisados foram extraídos, foi realizado na Faculdade de Zootecnia e Engenharia de Alimentos (FZEA) da Universidade de São Paulo (USP), localizada no município paulista de Pirassununga. Cinco áreas quadradas distribuídas na Prefeitura do Campus USP Fernando Costa (PUSP-FC), com lados de 5 metros, diferentes e classificadas como latossolo vermelho eutroférico, ricas em óxido ferroso, altamente férteis e com histórico de cultivo foram utilizadas com o intuito de avaliar a resistência mecânica do solo à penetração em MPa (variável resposta de interesse).

Os dados foram coletados entre outubro de 2016 e março de 2017. Nessa coleta, em cada área, foram utilizados três níveis de umidade (S - seco, U - úmido e E - encharcado) para verificar a resistência mecânica do solo em situações diferentes do dia a dia. Além disso, três tipos diferentes de penetrômetros (I - impacto, M - manual e A - automático) foram usados nas medições em 8 camadas de profundidade distantes em 5 cm entre 0 e 40 cm. O experimento apresenta 30 repetições por combinação de área, penetrômetro, umidade do solo e camada de profundidade, totalizando 10800 observações.

Apesar de inicialmente termos um experimento balanceado, foram descartadas todas as réplicas com pelo menos uma camada com resistência do solo à penetração igual ou superior a 6.5 MPa e também as medições do penetrômetro de impacto com resultado igual a 0. Com isso, o experimento totaliza 8824 observações ($N = 8824$) em 1103 unidades amostrais.

Os fatores e covariáveis disponíveis para a análise são, então:

- Área: local onde o experimento é realizado (1 a 5);
- Umidade: nível de quantidade de água no solo (seco, úmido ou encharcado);
- Penetrômetro: ferramenta utilizada nas medições da resistência mecânica do solo à penetração (impacto, manual ou automático); e
- Profundidade (em centímetros): nível de profundidade na qual ocorreu a medição em relação ao solo (5, 10, 15, 20, 25, 30, 35 e 40).

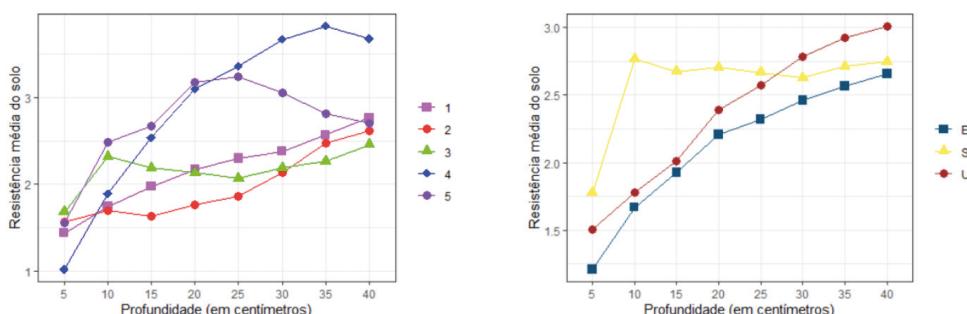


Figura 1: Gráfico de linhas para a resistência mecânica média do solo à penetração entre as interações duplas envolvendo as profundidades (5 a 40) com as áreas (1 a 5) e umidades (E, S e U), respectivamente.

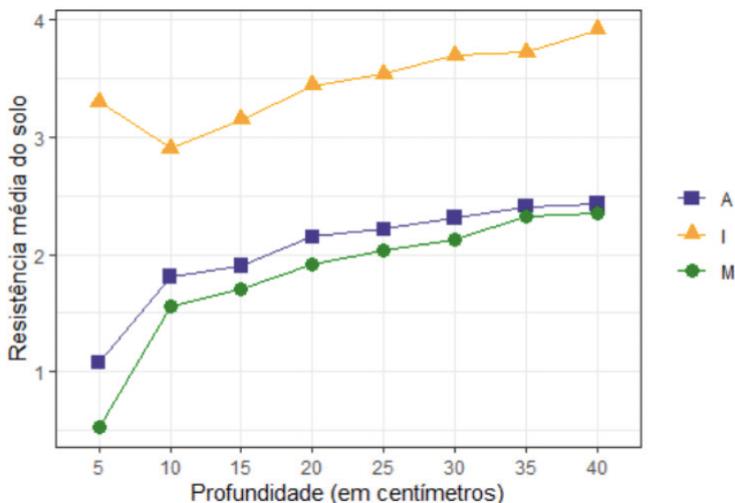


Figura 2: Gráfico de linhas para a resistência mecânica média do solo à penetração entre as interações duplas envolvendo as profundidades (5 a 40) com os penetrômetros (A, I e M).

As Figuras 1 e 2 apresentam uma análise gráfica dos dados. Pela Figura 1, temos um indicativo de que, em média, as áreas 1, 2 e 3 apresentam um comportamento parecido conforme a profundidade (em centímetros) aumenta e aparentemente linear ou levemente de forma logarítmica. A área 4 possui uma resistência mecânica média do solo à penetração maior em maiores profundidades e a área 5 possui resistência mecânica média maior do que 1, 2 e 3 a partir de 15 cm.

Ainda na Figura 1, temos um indicativo de que o solo seco (S) inicialmente apresenta uma resistência mecânica média bem maior que os outros tipos de umidade, porém, essa resistência mecânica média fica constante a partir de 10 centímetros de profundidade e, por conta disso, a partir dos 25 centímetros o solo seco apresenta resistência mecânica média parecida com os demais e, também, nos pontos de maior profundidade possui valores menores que o solo úmido (U). Já os solos úmido (U) e encharcado (E) possuem o mesmo comportamento crescente e linear na resistência mecânica média do solo à penetração com o aumento da profundidade, sendo o solo úmido o que sempre apresenta resistência mecânica média maior.

Por último, na Figura 2, temos um indicativo que, em média, o penetrômetro de impacto (I) possui resistência mecânica média do solo à penetração muito maior do que os outros dois em qualquer camada de profundidade. Já os penetrômetros automático e manual (A e M, respectivamente) possuem comportamentos médios semelhantes conforme aumenta a profundidade.

RESULTADOS

Um modelo misto com apenas o intercepto como efeito aleatório ($q = 1$ e \mathbf{Z}_i se resume a um vetor coluna do valor 1) e estrutura de covariância autoregressiva entre os erros aleatórios foi, então, ajustado para modelar o logaritmo da resistência mecânica do solo à penetração adicionando 1 unidade de MPa para as 1103 unidades experimentais do banco de dados. Outros modelos considerando a profundidade também como efeito aleatório e outras estruturas de covariância para os erros foram ajustados, mas esse se mostrou o mais adequado aos dados.

Utilizando o método de estimação de máxima verossimilhança restrita (MVR), os resultados em relação aos efeitos fixos podem ser vistos na Tabela 1.

Pela Tabela 1, temos evidências de que, com um nível de significância de 5% e utilizando o teste de Wald, apenas o efeito principal para a categoria Área 4 e a interação dupla Área 5 e Prof. 30 não foram significativos para explicar o logaritmo da variável resposta. Em contrapartida, os demais efeitos foram fortemente significativos, visto que cada efeito possui um valor-p praticamente igual a 0.

A análise da resistência mecânica à penetração revelou padrões distintos nas diferentes áreas de estudo. As áreas 1, 2 e 3 apresentaram menor resistência média, enquanto a área 5 é a que apresenta maior resistência, em média. Em relação aos métodos de medição, o penetrômetro manual exibe os valores mais baixos de resistência mecânica em média, enquanto o método de impacto registra os valores mais elevados. Quanto às condições do solo, observou-se que os solos encharcados apresentam as menores resistências médias, seguidos pelos solos úmidos e secos. Além disso, a resistência média aumenta com a profundidade. A interação entre área 4 e 5 e profundidade influenciam significativamente os resultados, explicando por que o efeito principal da área 4 não foi estatisticamente significativo. Nas profundidades iniciais, o comportamento médio da área 4 assemelha-se ao das áreas 1, 2 e 3, sugerindo uma complexidade adicional nos padrões observados.

	Estimativa	Erro padrão	Graus de liberdade	Estatística teste	Valor-p
Intercepto	0.4872	0.0162	7710	30.1211	$<2 \times 10^{-16}$
Área 4	0.0251	0.0206	1096	1.2164	0.2241
Área 5	0.2413	0.0194	1096	12.4221	$<2 \times 10^{-16}$
Penet. Impacto	0.5017	0.0151	1096	33.2354	$<2 \times 10^{-16}$
Penet. Manual	-0.1139	0.0136	1096	-8.3736	$<2 \times 10^{-16}$
Umid. Seco	0.2438	0.0146	1096	16.6773	$<2 \times 10^{-16}$
Umid. Úmido	0.1300	0.0137	1096	9.5013	$<2 \times 10^{-16}$
Profundidade	0.0155	0.0004	7710	39.0851	$<2 \times 10^{-16}$
Área 4 e Prof. 20	0.1361	0.0168	7710	8.0997	6.35×10^{-16}

Área 4 e Prof. 25	0.1655	0.0219	7710	7.5590	4.53×10^{-14}
Área 4 e Prof. 30	0.2027	0.0250	7710	8.1115	5.77×10^{-16}
Área 4 e Prof. 35	0.1982	0.0272	7710	7.2877	3.47×10^{-13}
Área 4 e Prof. 40	0.1068	0.0289	7710	3.6919	0.0002
Área 5 e Prof. 20	0.0937	0.0156	7710	5.9949	2.13×10^{-9}
Área 5 e Prof. 25	0.0576	0.0204	7710	2.8208	0.0048
Área 5 e Prof. 30	-0.0410	0.0234	7710	-1.7534	0.0796
Área 5 e Prof. 35	-0.1640	0.0255	7710	-6.4345	1.31×10^{-10}
Área 5 e Prof. 40	-0.2614	0.0272	7710	-9.6063	$< 2 \times 10^{-16}$

Tabela 1: Estimativas pontuais e o valor da estatística e do valor-p para o teste Wald de significância de cada efeito.

O desvio padrão para o intercepto aleatório foi de 2.6832×10^{-5} e dos erros aleatórios 0.2876. Como os efeitos aleatórios impactam diretamente na variância da variável resposta, o objetivo é que não tenha um desvio padrão muito próximo de zero (porque nesse caso os valores preditos dos efeitos são todos praticamente nulos) e não ser muito inferior ao desvio dos erros, pois assim ele pode não ser relevante para explicar a associação entre as observações. Aqui, observa-se que o desvio padrão do intercepto aleatório é bem baixo em relação ao desvio dos erros e isso talvez tenha acontecido por conta da estrutura autoregressiva de covariância escolhida para os erros. Durante o ajuste do modelo, foram realizados vários testes na busca de modelos diferentes e, talvez, mais adequados. Apesar de não serem mostrados aqui, um modelo sem a estrutura autoregressiva na covariância dos erros aleatórios foi estimado, o desvio padrão do intercepto aleatório aumentou consideravelmente, mas outras métricas de qualidade ficaram piores. De qualquer maneira, o intercepto aleatório (apesar de apresentar baixo valor de desvio padrão estimado) ainda é relevante para esse modelo.

As estimativas dos parâmetros da matriz de covariância autoregressiva para os erros aleatórios foi $\sigma^2 = 0.083$ (desvio padrão de 0.2876 como acima) e coeficiente de correlação de 0.696 entre erros de medidas adjacentes. A estimativa da correlação evidencia uma correlação positiva moderada a forte entre os erros de medições adjacentes. Na Tabela 2 temos, nas linhas e nas colunas, as profundidades em que foram feitas as medições das observações sendo 1 a profundidade de 5 centímetros e 8 a de 40 centímetros. Observa-se que **R**, apresenta maiores covariâncias entre os erros de profundidades mais próximas, quando observado linha a linha, e menores covariâncias entre erros de medições mais distantes. Esse comportamento era o desejado na escolha da estrutura autoregressiva para a matriz de covariâncias.

	1	2	3	4	5	6	7	8
1	0.0827	0.0576	0.0401	0.0279	0.0194	0.0135	0.0094	0.0065
2	0.0576	0.0827	0.0576	0.0401	0.0279	0.0194	0.0135	0.0094
3	0.0401	0.0576	0.0827	0.0576	0.0401	0.0279	0.0194	0.0135
4	0.0279	0.0401	0.0576	0.0827	0.0576	0.0401	0.0279	0.0194
5	0.0194	0.0279	0.0401	0.0576	0.0827	0.0576	0.0401	0.0279
6	0.0135	0.0194	0.0279	0.0401	0.0576	0.0827	0.0576	0.0401
7	0.0094	0.0135	0.0194	0.0279	0.0401	0.0576	0.0827	0.0576
8	0.0065	0.0094	0.0135	0.0194	0.0279	0.0401	0.0576	0.0827

Tabela 2: Estimativa da matriz \mathbf{R}_r

Para verificar a adequabilidade do modelo ajustado e a veracidade das suposições assumidas em relação aos erros e efeitos aleatórios, várias análises de diagnóstico foram realizadas. Entre elas, análise dos resíduos condicionais padronizados via gráficos de dispersão e histograma, gráfico quantil-quantil da distribuição normal e histograma para os resíduos com confundimento mínimo padronizados, histograma e boxplot dos valores preditos dos efeitos aleatórios, análise de pontos atípicos, entre outras. Aqui não mostramos no detalhe essas análises, mas elas podem ser observadas em Ribeiro (2024). No geral, as suposições do modelo ajustado estão satisfeitas e, apesar de ter observações atípicas, elas representam 0,75% da quantidade de observações e geralmente acontecem em combinação de áreas e profundidades que, pela análise descritiva, já mostravam um comportamento diferente do comportamento médio das demais observações.

Um modelo linear generalizado misto (GLMM), assumindo distribuição gama para a variável resposta na sua escala original (sem a transformação logarítmica) foi ajustado para comparação e apresentou resultados semelhantes aos apresentados por aqui, com maior complexidade de interpretação e conclusão.

DISCUSSÃO

Nesse trabalho, apresentamos a metodologia de modelos mistos, além de toda sua estrutura matemática e de estimação para o uso em uma aplicação de dados de solo. Essa técnica de modelagem foi utilizada por conta das características das observações feitas em cada perfuração do solo, que se tratam de dados longitudinais medidos em níveis de profundidades distintos porém na mesma posição do solo.

O modelo linear misto foi ajustado e, através dele, concluímos que os fatores: área, umidade e penetrômetro e as interações duplas entre às áreas 4 e 5 com as profundidades de 20 a 40, com exceção da interação entre área 5 e profundidade 30, foram significativos para explicar o logaritmo da resistência mecânica do solo à penetração.

Pelas estimativas associadas aos penetrômetros de impacto e manual, podemos constatar que os três tipos da ferramenta são diferentes em relação à resistência mecânica média observada. O penetrômetro de impacto é o que apresenta, em média, os maiores valores de resistência mecânica. Esta observação está alinhada com estudos anteriores que destacam a influência e importância da seleção do equipamento nas avaliações de compactação do solo. Zhang *et al.* (2024), por exemplo, discutem a necessidade de estudos relacionados à compactação do solo e a escolha das ferramentas necessárias, visto que ela afeta a absorção de nutrientes pelas culturas, dificulta o crescimento das raízes, causa crescimento atrofiado das culturas, leva à descoloração das folhas e reduz a altura das plantas, afetando, em última análise, o rendimento das culturas. Em adição, o impacto da umidade do solo na resistência foi evidente em nosso estudo, com solos secos exibindo a maior resistência.

Além disso, o modelo apresentou um bom ajuste para o logaritmo da resistência do solo à penetração, uma vez que pela análise de diagnóstico, observamos que suas suposições foram satisfatoriamente atendidas. Observações e perfurações atípicas e discrepantes também foram identificadas, mas elas representam uma quantidade muito pequena dentro do conjunto de dados analisado. Geralmente, elas estão presentes na área 4 e 5 e seria interessante avaliar se essas áreas possuem, de fato, alguma característica diferente ou se tivemos algum problema de medição nelas.

Pesquisas futuras devem considerar a expansão do estudo para várias regiões geográficas e tipos de solo, o que ajudaria a validar e ampliar as descobertas deste trabalho em diferentes condições ambientais. Além disso, explorar o impacto de vários fatores, como a compactação do solo de diferentes tipos de máquinas agrícolas e sua interação com a umidade e resistência do solo, forneceria resultados mais profundos. Incorporar tecnologias emergentes, como sensores de umidade em tempo real e modelos preditivos, poderia aprimorar ainda mais a pesquisa, permitindo análises mais dinâmicas e precisas das condições do solo.

REFERÊNCIAS

- Akhmetov, A., Akhmedov, S., e Ishchanov, J. (2024). Investigating the impact of speed and tire pressure of a wheel tractor on soil properties: A case study in Northeastern Uzbekistan. *AgriEngineering*, 6(3), 2067.
- Brasil, C. (2020). O solo: patrimônio essencial da agricultura. <https://croplifebrasil.org/noticias/o-solo-patrimonio-essencial-da-agricultura/>. Acessado em: 13-07-2023.
- Colombi, T. e Keller, T. (2019). Developing strategies to recover crop productivity after soil compaction – A plant eco-physiological perspective. *Soil and Tillage Research*, 191, 156–161.
- Embrapa (2020). *VII Plano Diretor da Embrapa 2020-2030*. Embrapa Informação Tecnológica, Brasília, DF, 1^a edição.

Hussein, S. A., Hussein, M. A., e Al-Rijabo, S. A. (2023). Mechanization assessment of soil compaction induced by traffic of farm machinery. In *IOP Conference Series: Earth and Environmental Science*, volume 1214, page 012051. IOP Publishing.

Menezes, T. A. V. (2018). *Comparação entre três penetrômetros na avaliação da resistência mecânica do solo à penetração de um latossolo vermelho eutroférrico*. Ph.D. thesis, Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP.

Patterson, H. D. e Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545–554.

Pinheiro, J., Bates, D., DebRoy, S., Deepayan, S., authors, E., Heisterkamp, Siem Willigen, B. V., Ranke, J., e Team, R. C. (2023). *Linear and Nonlinear Mixed Effects Models*. R Core Team (R-core at R-project.org), <https://svn.r-project.org/R-packages/trunk/nlme/>. Version 3.1-162.

Ribeiro, L. (2024). *Comparação de penetrômetros e identificação de fatores na resistência mecânica do solo à penetração via modelos mistos*. Trabalho de conclusão de curso (Graduação em Estatística), Universidade Federal de São Carlos, São Carlos, <https://repositorio.ufscar.br/handle/ufscar/19294>.

Singer, J. M., Nobre, J. S., e Rocha, F. M. (2018). Análise de dados longitudinais. São Paulo, SP. Universidade de São Paulo - Departamento de Estatística.

Visalli, A., Montefinese, M., Viviani, G., Finos, L., Vallesi, A., e Ambrosini, E. (2024). ImeEEG: Mass linear mixed-effects modeling of EEG data with crossed random effects. *Journal of Neuroscience Methods*, **401**, 109991.

Zhang, B., Jia, Y., Fan, H., Guo, C., Fu, J., Li, S., Li, M., Liu, B., e Ma, R. (2024). Soil compaction due to agricultural machinery impact: A systematic review. *Land Degradation & Development*, **35**(10), 3256–3273.

Zhang, X. (2015). A tutorial on restricted maximum likelihood estimation in linear regression and linear mixed-effects model. <https://xiuming.info/docs/tutorials/reml.pdf>.