

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA CIVIL

Guilherme Wolkan Wendhausen Pereira

**MAPEAMENTO GEOTÉCNICO PARA FUNDAÇÕES DE USINAS
FOTOVOLTAICAS UTILIZANDO ALGORITMO DE *MACHINE LEARNING***

Florianópolis

2022

Guilherme Wolkan Wendhausen Pereira

**MAPEAMENTO GEOTÉCNICO PARA FUNDAÇÕES DE USINAS
FOTOVOLTAICAS UTILIZANDO ALGORITMO DE *MACHINE LEARNING***

Trabalho Conclusão do Curso de Graduação em Engenharia Civil do Centro Tecnológico da Universidade Federal de Santa Catarina como requisito para a obtenção do título de Bacharel em Engenharia Civil.
Orientador: Prof. Rafael Augusto dos Reis Higashi, Dr..
Coorientador: Eng. Gabriel Phelipe Nascimento Rosolem, Me..

Florianópolis

2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Pereira, Guilherme Wolkan Wendhausen
MAPEAMENTO GEOTÉCNICO PARA FUNDAÇÕES DE USINAS
FOTOVOLTAICAS UTILIZANDO ALGORITMO DE MACHINE LEARNING /
Guilherme Wolkan Wendhausen Pereira ; orientador, Rafael
Augusto dos Reis Higashi, coorientador, Gabriel Phelipe
Nascimento Rosolem, 2022.
81 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia Civil, Florianópolis, 2022.

Inclui referências.

1. Engenharia Civil. 2. Mapeamento Digital de Solos. 3.
Standard Penetration Test (SPT). 4. Random Forest. 5.
Energia Fotovoltaica. I. Higashi, Rafael Augusto dos Reis.
II. Rosolem, Gabriel Phelipe Nascimento. III. Universidade
Federal de Santa Catarina. Graduação em Engenharia Civil.
IV. Título.

Guilherme Wolkan Wendhausen Pereira

**MAPEAMENTO GEOTÉCNICO PARA FUNDAÇÕES DE USINAS
FOTOVOLTAICAS UTILIZANDO ALGORITMO DE *MACHINE LEARNING***

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título de “Engenheiro Civil” e aprovado em sua forma final pelo Curso de Engenharia Civil da Universidade Federal de Santa Catarina.

Florianópolis, 18 de março de 2022.

Prof. Liane Ramos da Silva, Dra.
Coordenadora do Curso

Banca Examinadora:

Prof. Rafael Augusto dos Reis Higashi, Dr.
Orientador
Universidade Federal de Santa Catarina

Eng. Gabriel Phelipe Nascimento Rosolem, Me.
Coorientador
Universidade Federal de Santa Catarina

Profa. Gracieli Dienstmann, Dra.
Avaliadora
Universidade Federal de Santa Catarina

Eng. Elisângela Benedet da Silva, Dra.
Avaliadora
Epagri/Ciram

Eng. Mauricio Vivan Guarnieri, Me.
Avaliador
Araxá Solar

AGRADECIMENTOS

Agradeço à minha mãe, Solange, por ter me dado condições de estudar, ter incentivado meus estudos e por nunca ter me deixado desistir. À minha irmã, Greice, por todo o apoio em casa que me permitiu me dedicar a esse trabalho. Ao meu irmão, Eduardo, por todos os conselhos profissionais.

Ao professor Higashi e ao engenheiro Gabriel, por terem aceitado orientar e terem colaborado com esse trabalho. À professora Gracieli, engenheira Elisângela e engenheiro Mauricio, por terem aceitado participar da banca.

Ao EPEC, ao professor Carlos, à equipe de geoprocessamento do LabTrans, à professora Liseane e à Araxá Solar pelas oportunidades de estágio e crescimento profissional.

Aos meus amigos do LabFSG: Dani, Fernanda, Kauê, Luís, Phelps, Ricardo e Vini, por terem deixado a graduação mais leve.

Aos meus amigos Davy, Felipe, Jhonattan e Pes, por compreenderem minhas ausências.

Agradecimentos especiais ao Pes e ao Vini pelas inúmeras vezes em que ficamos estudando na BU até ela fechar.

A todos os outros amigos da 15.1, do curso e todas as outras pessoas que contribuíram para a minha formação.

Muito obrigado!

“Eu quase que nada não sei. Mas desconfio de muita coisa.”

(João Guimarães Rosa)

RESUMO

A energia solar fotovoltaica é uma fonte de energia renovável e que não emite carbono durante a produção de energia, com expectativa de se tornar a principal fonte da matriz brasileira até 2040. A montagem dos equipamentos de usinas fotovoltaicas é um processo repetitivo e rápido. Buscando estender esse ritmo acelerado para o restante da obra, a preferência para a fundação dos painéis tem sido o uso de estacas metálicas cravadas. Sondagens SPT (*Standard Penetration Test*) auxiliam no dimensionamento das fundações e em estimativas das áreas passíveis de cravação dentro do terreno de uma futura usina. Devido às grandes dimensões das usinas, torna-se economicamente inviável realizar sondagens SPT com a mesma densidade recomendada para edificações. Através do Scorpan, modelo derivado da Pedologia, identificou-se 29 covariáveis ambientais – derivadas do modelo digital de elevação SRTM e uma imagem do satélite Landsat 8 – que representassem fatores de formação do solo. Em seguida, utilizou-se o algoritmo *random forest* para construir um modelo relacionando as covariáveis com os números de golpes de 74 ensaios SPT do terreno de uma usina de 10,2 km² no semiárido brasileiro, nas profundidades de 1, 2 e 3 metros. O modelo construído foi utilizado para criar mapas de N_{SPT}, que, por sua vez, foram utilizados para criar mapas de cravabilidade com base em um N_{SPT} limite de cravação. As variáveis mais importantes foram Altimetria, *Clay Index*, Precipitação Anual e Isotermalidade. Para a profundidade de 1 metro, o modelo apresentou R² = 0,43 e RMSE = 4,93; para a profundidade de 2 metros, R² = 0,41 e RMSE = 18,44; e para a profundidade de 3 metros, R² = 0,25 e RMSE = 17,66. Na falta de mapeamentos de N_{SPT} que avaliassem seu desempenho através do coeficiente de determinação R², os resultados foram comparados com mapeamentos de textura do solo, em especial um mapeamento também localizado no semiárido brasileiro. Conclui-se que o desempenho do mapeamento é condizente com a literatura disponível e que os mapas permitem uma boa visualização do comportamento e distribuição dos valores de N_{SPT} no terreno, auxiliando na tomada de decisões sobre as fundações.

Palavras-chave: Mapeamento digital de solos. Fundações. Energia fotovoltaica.

ABSTRACT

Photovoltaic solar energy is a renewable source that does not emit carbon during energy production, and is expected to become the main source of the Brazilian energy matrix up to 2040. The assembly of a photovoltaic power plant's equipments is a repetitive and fast process. Looking to extend this fast pace to the whole construction, the preference for the panels foundations has been the use of driven steel piles. Standard Penetration Tests (SPT) assist in the foundations design and in estimating the areas in the power plant lands that are liable to pile driving. Due to the large dimensions of the power plants, it becomes economically unfeasible to perform SPT with the same density recommended for buildings. Through Scorpan, a framework derived from Pedology, 29 environmental covariates were identified - derived from SRTM data and a Landsat 8 image - that represented soil formation factors. Then, the random forest algorithm was used to build a model associating the covariates with the number of blows of 74 SPT on the terrain of a 10.2 km² plant in the Brazilian semiarid region, at depths of 1, 2 and 3 meters. The built model was used to create N_{SPT} maps, which, in turn, were used to create drivability maps based on an N_{SPT} limit for driving piles. The most important variables were Altitude, Clay Index, Annual Precipitation and Isothermality. For a depth of 1 meter, the model showed $R^2 = 0,43$ and $RMSE = 4.93$; for a depth of 2 meters, $R^2 = 0,41$ and $RMSE = 18.44$; and for the depth of 3 meters, $R^2 = 0,25$ and $RMSE = 17.66$. In the lack of N_{SPT} mappings that evaluated its mapping performance through the determination R^2 , the results were compared with soil texture mappings, especially a also located in the Brazilian semiarid region. It is concluded that the performance of the mapping is consistent with the available literature and that the maps allow a good visualization of the behavior and distribution of N_{SPT} values on the terrain, assisting the foundations decision making process.

Keywords: Digital Soil Mapping. Foundations. Photovoltaic energy.

LISTA DE FIGURAS

Figura 1 - Relação entre células, módulos e painéis fotovoltaicos.....	18
Figura 2 - Esquema de uma usina fotovoltaica.....	19
Figura 3 - Microestaca.....	21
Figura 4 - Conjunto dos equipamentos mais comuns no SPT.....	22
Figura 5 - Esquema de execução de um ensaio SPT.....	23
Figura 6 – Prova de carga de arrancamento com tripé como elemento de reação.	24
Figura 7 – Prova de carga horizontal com máquina pesada como elemento de reação.	25
Figura 8 - Ciclos da prova de carga horizontal.....	25
Figura 9 - Ruptura de estacas sob carregamento lateral: (a) ruptura do elemento estrutural e (b) ruptura do solo.....	26
Figura 10 - Perfil de solo com horizontes claramente diferenciados pela cor.....	27
Figura 11 - Principais horizontes pedológicos.	28
Figura 12 - Erosão de encosta com deposição de colúvio.....	32
Figura 13 - Perfil de uma catena caracterizada pela perda de finos.	33
Figura 14 - Relação entre tamanho de partícula e mineralogia.	34
Figura 15 - Exemplos de variáveis que podem representar os fatores do modelo Scorpan.	35
Figura 16 - Resultado de classificação de solos no Rio Fraser, Canadá, utilizando diferentes algoritmos para a função f: (a) árvore de decisão, (b) árvore de decisão com <i>bagging</i> , (c) <i>random forest</i> , e (d) rede neural artificial.	36
Figura 17 - Exemplo de árvore de decisão representada em (a) modelo de árvore; e (b) espaços de predição em um plano cartesiano.....	37
Figura 18 - Exemplo de seleção da variável e limite do nó raiz.....	38
Figura 19 - Árvore com número de folhas igual ao número de dados e SSE igual a zero.....	39
Figura 20 - Exemplo da falta de robustez de um modelo sobreajustado.	39
Figura 21 - Localização do município de Oliveira dos Brejinhos e área de estudo.	43
Figura 22 - Localização dos ensaios SPT.....	45
Figura 23 - Distribuição espacial das variáveis do fator solo.....	49
Figura 24 – Distribuição espacial das variáveis do fator clima, precipitação anual (c1) e isothermalidade (c2).	50
Figura 25 – Distribuição espacial da variável do fator organismos.	51
Figura 26 – Distribuição espacial das variáveis do fator relevo.....	52

Figura 27 – Distribuição espacial das variáveis do fator material de origem.....	53
Figura 28 - Mapa de N_{SPT} para a profundidade de 1 metro.	59
Figura 29 - Porcentagem de área ocupada por faixa de valor de N_{SPT} para a profundidade de 1 metro.....	60
Figura 30 - Porcentagem de aumento do RMSE causado pelo embaralhamento de cada variável para 1 metro de profundidade: (a) todas as variáveis ordenadas; e (b) variáveis agrupadas por fator.....	61
Figura 31 - Mapa de N_{SPT} para a profundidade de 2 metros.	62
Figura 32 - Porcentagem de área ocupada por faixa de valor de N_{SPT} para a profundidade de 2 metros.	63
Figura 33 – Porcentagem de aumento do RMSE causado pelo embaralhamento de cada variável para 2 metros de profundidade: (a) todas as variáveis ordenadas; e (b) variáveis agrupadas por fator.....	64
Figura 34 - Mapa de N_{SPT} para a profundidade de 3 metros.	65
Figura 35 - Porcentagem de área ocupada por faixa de valor de N_{SPT} para a profundidade de 3 metros.	66
Figura 36 – Porcentagem de aumento do RMSE causado pelo embaralhamento de cada variável para 3 metros de profundidade: (a) todas as variáveis ordenadas; e (b) variáveis agrupadas por fator.....	67
Figura 37 - Área ocupada por cada faixa de N_{spt} da cravabilidade.....	68
Figura 38 - Mapas de N_{spt} máximo para cravação.	69
Figura 39 - Área acumulada do terreno para cada valor de N_{SPT} limite para cravação.....	70
Figura 40 - Pontos que não chegaram a 3 metros no sul da UFV.	73

LISTA DE QUADROS

Quadro 1 – Diâmetros limites entre areia, silte e argila, de acordo com a NBR 6502 - Solos e rochas.....	30
Quadro 2 - Variáveis utilizadas e seus respectivos fatores de formação.....	55
Quadro 3 - Coeficientes de determinação de estudos de mapeamento da textura do solo com <i>random forest</i>	71

LISTA DE TABELAS

Tabela 1 - Resumo dos ensaios SPT, com sistema de referência SIRGAS 2000 e coordenadas UTM, fuso 23, sul.....	46
Tabela 2 - Resumo dos parâmetros de <i>random forest</i> e resultados para as três profundidades.	68

LISTA DE ABREVIATURAS E SIGLAS

ABNT – Associação Brasileira de Normas Técnicas

ANEEL – Agência Nacional de Energia Elétrica

Clorpt – Modelo pedológico de fatores de formação dos solos, elaborado por Jenny (1941).

Embrapa – Empresa Brasileira de Pesquisa Agropecuária

N_{SPT} – Índice de resistência à penetração do ensaio SPT

POT – *Pull-out Test*

R^2 - coeficiente de determinação

RMSE – Raiz do erro quadrático médio (do inglês *Root Mean Squared Error*)

Scorpan – Modelo pedológico que tem como objetivo gerar funções espaciais de predição de solo, formalizado por McBratney et al. (2003).

SiBCS – Sistema Brasileiro de Classificação de Solos

SPT – *Standard Penetration Test*

SSE - Soma dos quadrados dos erros (do inglês *sum of squared errors*)

UFV – Usina Fotovoltaica

SUMÁRIO

1	INTRODUÇÃO	15
1.1	JUSTIFICATIVA	15
1.2	OBJETIVOS	16
1.2.1	Objetivo Geral.....	16
1.2.2	Objetivos Específicos	16
2	REFERENCIAL TEÓRICO	18
2.1	USINAS FOTOVOLTAICAS.....	18
2.1.1	Fundações para painéis de usinas fotovoltaicas.....	20
2.1.1.1	<i>Estaca metálica por cravação direta.....</i>	<i>20</i>
2.1.1.2	<i>Estaca metálica cravada com pré-furo.....</i>	<i>21</i>
2.1.1.3	<i>Microestaca</i>	<i>21</i>
2.1.2	Investigação do subsolo para fundações de painéis de usinas fotovoltaicas... 21	
2.1.2.1	<i>Standard Penetration Test (SPT).....</i>	<i>22</i>
2.1.2.2	<i>Provas de carga horizontal e de arrancamento ou Pull-out Test (POT)</i>	<i>24</i>
2.2	PEDOLOGIA	26
2.2.1	Formação dos solos.....	29
2.2.1.1	<i>Fatores de formação dos solos</i>	<i>30</i>
2.2.1.1.1	Modelo Clorpt.....	31
2.2.1.1.2	Modelo Scorpan.....	34
3	ÁREA DE ESTUDO	42
4	MATERIAIS E MÉTODO	44
4.1	COVARIÁVEIS AMBIENTAIS	49
5	RESULTADOS	58
5.1	MAPA DE CRAVABILIDADE	68
5.2	COMPARAÇÃO DOS RESULTADOS	70
5.3	DISCUSSÃO	72
6	CONCLUSÕES.....	74

7	SUGESTÕES PARA ESTUDOS FUTUROS.....	75
	REFERÊNCIAS	76

1 INTRODUÇÃO

A geração e o uso de energia elétrica contribuíram para o crescimento econômico e melhoria da qualidade de vida humana. Tendo sido baseado principalmente em fontes de energia emissoras de carbono, como carvão e combustíveis fósseis (DINCER; ZAMFIRESCU, 2018), tal crescimento trouxe impactos socioambientais diversos como o aumento do efeito estufa, a perda de qualidade do ar e o aumento de doenças respiratórias (WANG, 2018). Almejando principalmente diminuir as emissões de carbono, fontes alternativas e renováveis de energia foram desenvolvidas, como a biomassa, a energia eólica e a energia fotovoltaica.

A energia fotovoltaica tem como princípio o efeito fotovoltaico, no qual materiais semicondutores geram corrente elétrica ao terem elétrons deslocados pela incidência de radiação eletromagnética, como a luz do sol. O efeito foi primeiramente observado pelo físico francês Becquerel em 1839 e teve seu primeiro uso industrial apenas em 1954, com o desenvolvimento de uma célula fotovoltaica com rendimento de 6% (DINCER, 2018).

No Brasil, o marco inicial dessa forma de geração foi a Resolução Normativa Nº 482 de 17 de abril de 2012 da Agência Nacional de Energia Elétrica (ANEEL), que estabeleceu as condições para geração de energia elétrica pelo próprio usuário, com limite de potência instalada de 5 MW. No final de 2014, o Ministério de Minas e Energia contratou aproximadamente 890 MW de potência instalada na forma de energia fotovoltaica. De acordo com Guarnieri (2017), esse ato simbolizou o início efetivo da participação desta fonte na matriz energética brasileira, além de ter mostrado a necessidade do desenvolvimento tecnológico do setor no país. Em maio de 2021, o Brasil atingiu 9,4 GW de potência fotovoltaica instalada, enquanto a potência total já outorgada era de 24,7 GW. As usinas fotovoltaicas (UFV) somam 1,9% da matriz energética brasileira, com 3,4 GW (ABSOLAR, 2021). Até 2040, espera-se que essa forma de geração seja a principal fonte brasileira com 117 GW, na frente dos 111 GW das usinas hidrelétricas. O investimento no setor chegará a 97 bilhões de dólares (BLOOMBERG, 2018).

1.1 JUSTIFICATIVA

A montagem dos equipamentos de usinas fotovoltaicas é um processo repetitivo e rápido. Buscando estender esse ritmo acelerado para o restante da obra, a preferência para a fundação dos painéis tem sido o uso de estacas metálicas cravadas. Nos casos em que o solo é resistente

demais para a cravação direta ou o leito rochoso é pouco profundo, perfuratrizes são utilizadas em estacas concretadas in loco ou com pré-furos para facilitar a cravação dos perfis metálicos (GUARNIERI, 2017). Em ambos os usos da perfuratriz, o custo e a duração da execução das fundações aumentam. Caso o uso dessas fundações alternativas não tenha sido previsto, os prejuízos são ainda maiores, já que a cravação de estacas é uma atividade crítica na construção de uma usina fotovoltaica (RODRIGUES, 2021) e um atraso na obra significa atraso no retorno financeiro do empreendimento.

Por conta do baixo peso próprio dos módulos fotovoltaicos e estruturas, a principal ação nas fundações é a ação do vento (DA SILVA, 2013; ALMEIDA, 2014), causando cargas laterais e de arrancamento. A resistência do solo a esses carregamentos é medida através de provas de carga ou estimada a partir de sondagens e ensaios de solo (SOARES, 2021). Por conta das áreas das usinas, que podem chegar a centenas de hectares, é economicamente inviável realizar ensaios com a densidade com que seriam executados para fundações de outras obras correntes, como edificações.

Assim, é contínua a busca por formas mais eficientes e precisas de predição dos valores de resistência das estacas metálicas. Nesse trabalho, o potencial do uso de algoritmos de *machine learning* e covariáveis ambientais para tais fins será avaliado no terreno de uma usina fotovoltaica no semiárido baiano.

1.2 OBJETIVOS

Este capítulo estabelece o objetivo geral e os objetivos específicos deste trabalho.

1.2.1 Objetivo Geral

Realizar mapeamento geotécnico com base em algoritmo de *machine learning* para auxílio na tomada de decisão de fundações de uma usina fotovoltaica.

1.2.2 Objetivos Específicos

- i. Elaborar mapas de valores N_{SPT} em um terreno de usina fotovoltaica através de algoritmo de *machine learning*;
- ii. Quantificar as incertezas nas predições dos valores de N_{SPT} por profundidade; e

- iii. Elaborar mapas de cravabilidade de estacas metálicas em um terreno de usina fotovoltaica.

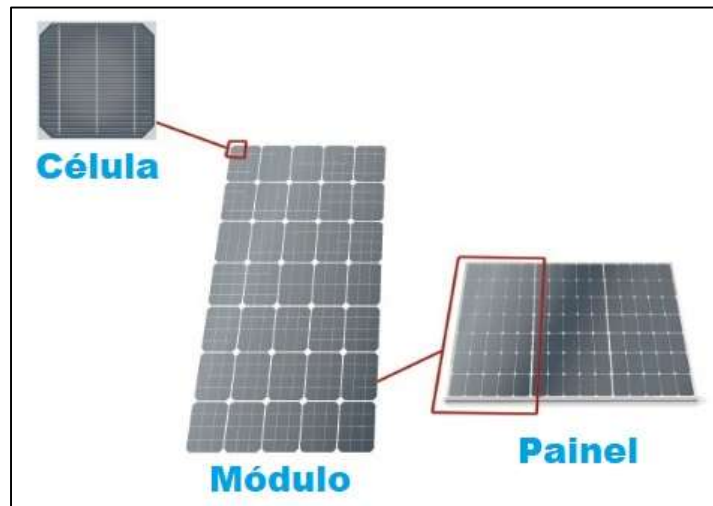
2 REFERENCIAL TEÓRICO

A primeira parte deste capítulo mostra o funcionamento básico de uma UFV, seus tipos de fundações e formas de caracterização do solo mais comuns. A segunda parte discorre sobre a Pedologia, ramo da Ciência do Solo que classifica esse material e estuda sua formação e maneiras de mapeá-lo.

2.1 USINAS FOTOVOLTAICAS

Por conta de sua baixa capacidade de geração de energia, as células fotovoltaicas são agrupadas, formando o módulo fotovoltaico. Os módulos podem ser monofaciais, quando geram energia através da radiação incidente em apenas uma face, ou bifaciais, quando ambas as faces geram energia. O agrupamento dos módulos forma o painel fotovoltaico, conforme ilustra a Figura 1.

Figura 1 - Relação entre células, módulos e painéis fotovoltaicos.

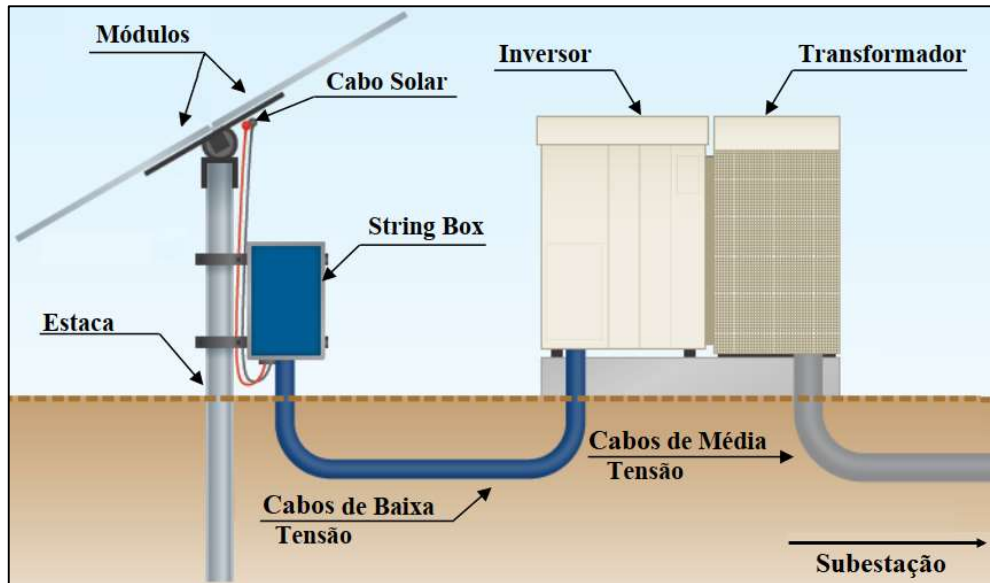


Fonte: Adaptado de Reis (2015).

Os módulos de um painel são conectados em uma série (*string*, em inglês) por cabos solares, que diferem dos cabos comuns por suportarem uma maior tensão e pela maior durabilidade de seu isolante frente ao intemperismo (VILLALVA, 2012). Os painéis são então ligados em paralelo entre si através de um quadro elétrico de combinação chamado de *string box* (GUARNIERI, 2017).

Cabos de baixa tensão enterrados no solo conectam as *string boxes* até inversores que convertem a energia de corrente contínua para corrente alternada. Em seguida, um transformador eleva a tensão da energia, que vai até a subestação por cabos de média tensão (SOARES, 2021). A subestação eleva a tensão para alta, de acordo com as linhas de transmissão. A Figura 2 resume a configuração padrão de uma UFV.

Figura 2 - Esquema de uma usina fotovoltaica.



Fonte: Adaptado de Consórcio SETA-Araxá (2020, apud Soares, 2021).

Os módulos são fixados a uma estrutura de suporte metálica, que pode ser fixa ou móvel. A estrutura móvel, chamada de *tracker*, rotaciona o painel ao longo do dia buscando manter a perpendicularidade da superfície das células fotovoltaicas em relação ao sol, o que aumenta a geração de energia. A estrutura de suporte transfere os esforços do peso próprio dos equipamentos e da ação do vento incidente nos painéis para uma estrutura de fundação. Bloss et al. (1991) afirmam que a estrutura de suporte dos painéis tem um importante papel nos custos de uma UFV, listando pontos de atenção na escolha da estrutura de forma a reduzir os custos. Dentre eles, podemos destacar:

- i. Facilidade de transporte;
- ii. Facilidade de conexão com os painéis e a estrutura de fundação;
- iii. Ausência de necessidade de ferramentas especiais para montagem;
- iv. Ausência de necessidade de mão-de-obra especializada para montagem; e
- v. Potencial para aplicar técnicas de linha de montagem.

2.1.1 Fundações para painéis de usinas fotovoltaicas

Um sistema de fundação tem como função principal transferir carregamentos ao solo, sendo composto pelo elemento estrutural de fundação e pelo maciço de solo que o cerca (ABCP, 2014). Tanto o material que compõe a estrutura da fundação quanto o solo ao seu redor precisam resistir ao carregamento aplicado, de forma que a capacidade de carga de uma fundação é o menor valor de resistência entre os dois componentes (ALONSO, 2019; CINTRA et al., 2011).

Os esforços predominantes nas fundações de painéis fotovoltaicos são os laterais e de arrancamento devido ao vento. Assim, quando o projetista opta pelo uso de sapatas, o dimensionamento se dá pelo contrapeso do concreto, que impede a sapata de ascender ou tombar (ALMEIDA, 2014). Dependendo do peso necessário, o uso dessa alternativa é impossibilitado economicamente, seja pela distância das centrais de concretagem para a sapata concretada *in loco*, seja pelo custo elevado de transporte por conta do peso das sapatas pré-moldadas. No caso de estacas metálicas, a resistência ao arrancamento se dá pelo atrito lateral com o solo, sendo função principalmente do comprimento da estaca (CINTRA; AOKI, 2010), não precisando depender tanto do peso do material. A resistência lateral se dá pela resistência do solo e da estaca à deformação na flexão (SOARES, 2021), também não dependendo do peso do material. A leveza e a elevada resistência a tração e flexão fazem com que as estacas metálicas sejam fáceis de transportar e manipular (VELLOSO; LOPES, 2010). Bloss et al. (1991) defendem que as estacas metálicas também apresentam a vantagem da maior velocidade de execução na conexão com a estrutura de suporte. Guarnieri (2017) e Rodrigues (2021) afirmam que a estaca metálica é a solução de fundação para UFV mais encontrada no Brasil. Seus perfis mais comuns são o perfil I, o perfil C e o perfil ômega (ALMEIDA, 2014).

2.1.1.1 Estaca metálica por cravação direta

Quando a resistência do solo não é elevada demais ou o leito rochoso tem profundidade maior do que o comprimento da estaca, é possível cravá-la diretamente com o uso de pesos em queda livre, martelos hidráulicos ou vibratórios (FALCONI et al.; 2014), sendo o vibratório o mais comum em UFV (SOARES, 2021). Comparada com a estaca cravada com pré-furo e a microestaca, é a solução de execução mais rápida e de menor custo.

2.1.1.2 Estaca metálica cravada com pré-furo

Quando a resistência do solo é elevada demais para uma cravação direta, um pré-furo é executado de forma a facilitar a cravação. O furo tem diâmetro menor do que as dimensões da seção do perfil metálico, sendo o vazio ao redor concretado (STI NORLAND BRASIL, 2021), o que causa o aumento de custo e de duração da execução em relação à cravação direta.

2.1.1.3 Microestaca

Se a resistência do solo é tão alta que mesmo com o pré-furo a cravação ainda não é possível, um furo maior que a seção do perfil metálico é escavado. O perfil é posicionado no furo, que é preenchido com concreto, conforme a Figura 3. Essa solução também é utilizada quando o solo tem uma resistência tão baixa que a área lateral do perfil metálico não é grande o suficiente para mobilizar a resistência necessária, sendo realizado um aumento na seção transversal do furo.

Figura 3 - Microestaca



Fonte: Adaptado de STI Norland Brasil (2021).

2.1.2 Investigação do subsolo para fundações de painéis de usinas fotovoltaicas

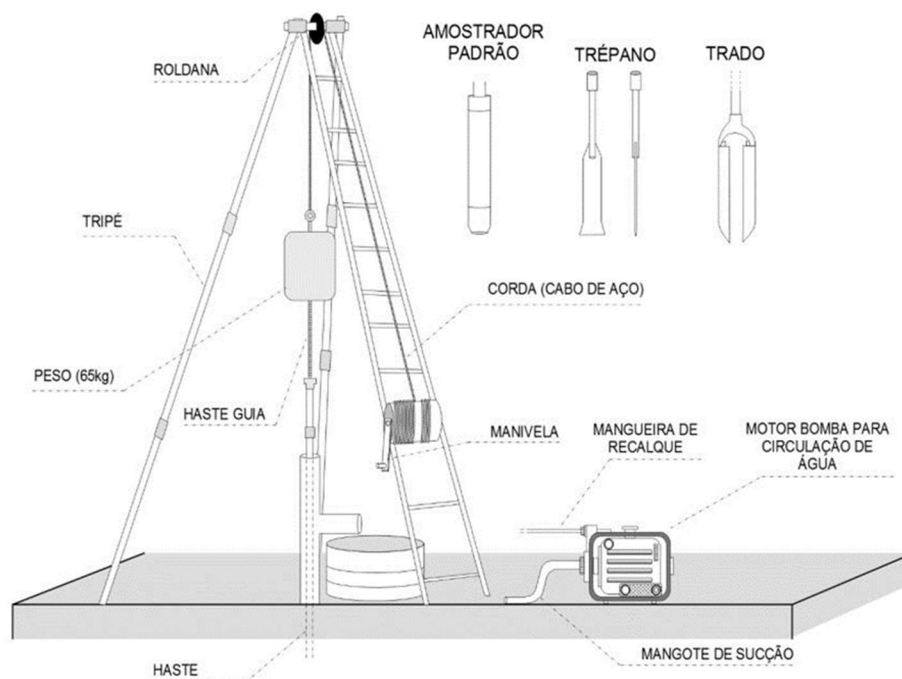
A investigação do subsolo se faz necessária para o projeto de qualquer fundação, visto que o tipo de fundação adotado depende da posição do nível d'água e dos parâmetros de resistência do solo em relação às cargas aplicadas. A seguir são apresentadas as formas de

investigação mais comuns em UFV: o *Standard Penetration Test* (SPT) e as provas de carga horizontal e de arrancamento.

2.1.2.1 Standard Penetration Test (SPT)

O SPT mede a resistência do solo à penetração de um amostrador padrão sob a carga dinâmica da queda livre de uma massa de 65 kg à altura de 75 cm, além de coletar amostras deformadas de solo. É o ensaio mais utilizado no Brasil para o projeto geotécnico de fundações (CINTRA et al., 2013) e é regulamentado no país pela NBR 6484 de 2020. Em uma UFV, o SPT tem como principais funções o reconhecimento do solo e suas variações dentro do terreno, servir como base para o projeto geotécnico das estacas e a determinação da profundidade do nível d'água e do impenetrável. A Figura 4 ilustra os equipamentos mais comumente utilizados na execução do SPT.

Figura 4 - Conjunto dos equipamentos mais comuns no SPT.

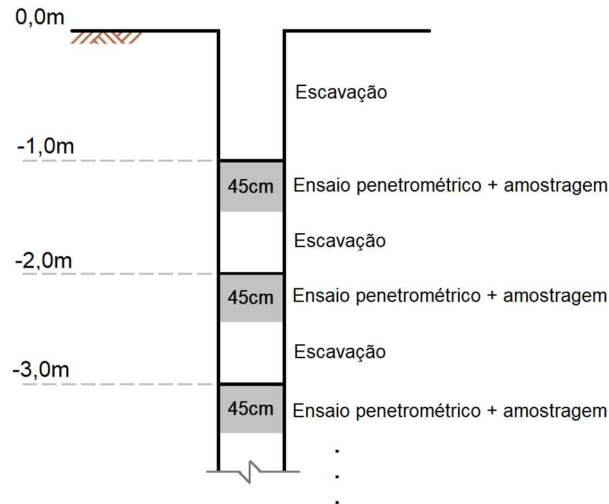


Fonte: Higashi (2016).

O ensaio começa com a escavação do primeiro metro através de um trépano. O trépano é então substituído pelo amostrador padrão, que recebe energia dos golpes do martelo até penetrar 45 cm. O amostrador é retirado com a amostra de solo, e mais 55 cm são escavados pelo trépano,

de forma a se completar um metro. Esse processo, ilustrado na Figura 5, é repetido até a profundidade definida pelo contratante ou conforme critérios estabelecidos pela NBR 6484.

Figura 5 - Esquema de execução de um ensaio SPT



Fonte: Adaptado de Cintra et al. (2013);

Os tipos de solo coletados pelo amostrador são identificados através de análise tátil-visual e de ensaios de granulometria, permitindo a elaboração de um perfil das camadas de solo do terreno naquele ponto.

A contagem do número de golpes para cravação de 45 cm do amostrador, dividida em três trechos de 15 cm cada, recebe o nome de ensaio penetrométrico (CINTRA et al., 2013). À soma do número de golpes para penetração dos dois últimos trechos de 15 cm dá-se o nome de índice de resistência à penetração (N_{SPT}). O primeiro trecho de 15cm é desconsiderado por ter sua estrutura afetada pela ação do trépano.

Além de serem utilizados para calcular a capacidade de carga de fundações e estimar peso específico, coesão e ângulo de atrito do solo, os valores de N_{SPT} permitem avaliar a cravabilidade de estacas pré-moldadas. Cintra e Aoki (2010) apresentam valores de N_{SPT} entre 25 e 55 como o limite para a cravação de estacas metálicas.

2.1.2.2 Provas de carga horizontal e de arrancamento ou Pull-out Test (POT)

O ensaio *pull-out test* (POT) consiste em ciclos de aplicação de cargas de intensidade crescente em uma estaca e a medição dos deslocamentos correspondentes, seguido de descarregamento (SOARES, 2021). Usualmente, o carregamento máximo do ensaio equivale a 200% ou 250% do valor de cálculo obtido pela análise estrutural do elemento (ALMEIDA, 2014; SOARES, 2021). No contexto das UFV, o termo *pull-out test* é muitas vezes usado para referir-se tanto a prova de carga de arrancamento quanto a horizontal.

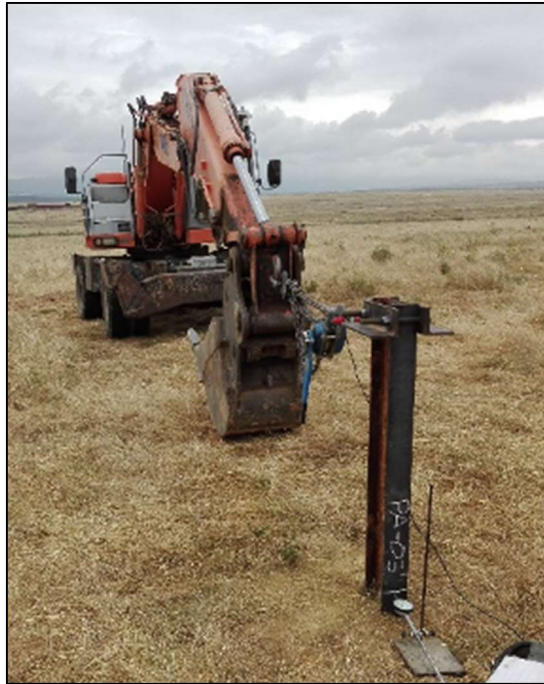
O carregamento é aplicado através de uma tralha ou um macaco hidráulico e é medido por um dinamômetro ou uma célula de carga. O deslocamento é medido por um micrômetro. O elemento de reação pode ser um tripé apoiado no solo, como mostra a Figura 6, ou uma máquina pesada, como mostra a Figura 7, ou ainda vigas de reação.

Figura 6 – Prova de carga de arrancamento com tripé como elemento de reação.



Fonte: Powerway Renewable Energy (2021).

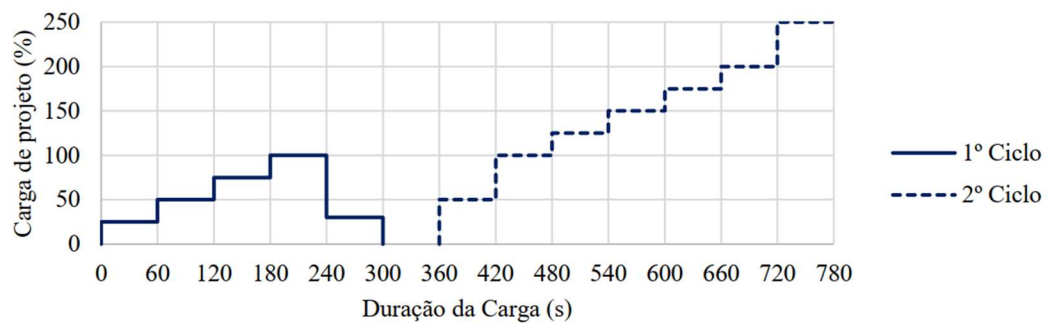
Figura 7 – Prova de carga horizontal com máquina pesada como elemento de reação.



Fonte: Elaborex (2021).

Segundo Soares (2021), a prova de carga pode ser dividida em dois ciclos: no primeiro, o carregamento final é de 100% da carga de projeto, de forma a analisar o comportamento da estaca durante a situação esperada de uso, enquanto no segundo o carregamento alcança 250% da carga de projeto, buscando identificar o coeficiente de segurança. A Figura 8 ilustra esses ciclos em um caso de carregamento horizontal.

Figura 8 - Ciclos da prova de carga horizontal



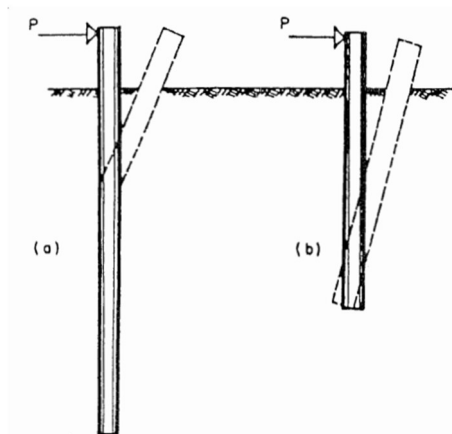
Fonte: Consórcio SETA-Araxá (2020, apud Soares, 2021).

Em solos arenosos, a resistência ao arrancamento depende principalmente do ângulo de atrito e da razão entre comprimento da estaca (sob a superfície) e o diâmetro da seção, tanto

para estacas isoladas (GAAVER, 2013) quanto para grupos (SHANKER et al., 2003). É válido assumir que a superfície de ruptura se dá ao longo do fuste da estaca (CINTRA; AOKI, 2010; SOARES, 2021).

No caso de esforços laterais, Reese e Van Impe (2011) afirmam que não há evidência suficiente para estabelecer uma relação entre diâmetro e resistência para areias, que são particularmente problemáticas. A falta de coesão faz com que o solo caia dentro do buraco deixado pelo deslocamento da estaca, não permitindo que a estaca volte à posição original mesmo no caso de um carregamento cíclico e esforços dentro do regime elástico. A Figura 9 ilustra a ruptura no elemento estrutural, mais comum em estacas compridas, e no maciço de solo, mais comum em estacas curtas (BROMS, 1964).

Figura 9 - Ruptura de estacas sob carregamento lateral: (a) ruptura do elemento estrutural e (b) ruptura do solo.



Fonte: Broms (1964)

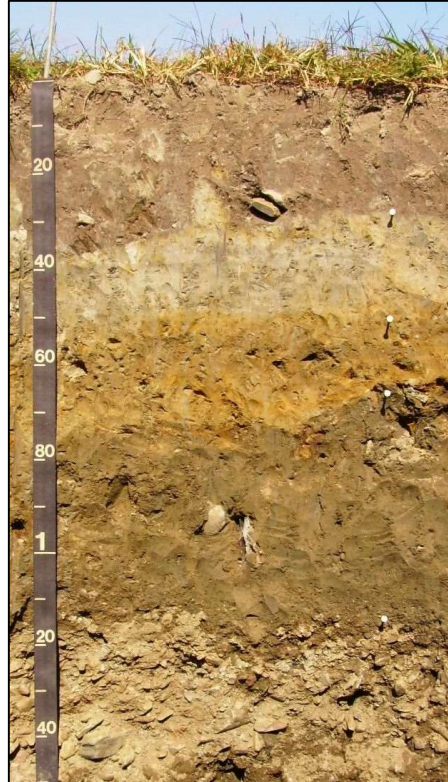
2.2 PEDOLOGIA

A Ciência do Solo envolve os estudos das propriedades físicas, químicas, biológicas e mineralógicas do solo (HUANG et al., 2012). A Pedologia é um ramo que foca na gênese, morfologia, classificação e mapeamento desse material. (PEREIRA et al., 2019). Segundo Lepsch (2010), a área teve início efetivo com o cientista russo Dokouchaev, que constatou que o solo deveria ser estudado na sua condição natural em campo, por ser um corpo dinâmico que varia geograficamente e é influenciado por diversos fatores ambientais em sua formação. O autor também afirma que não há limites bem definidos entre a Pedologia e estudos aplicados do solo como a Geotecnia.

O solo é um conjunto formado por partículas minerais, matéria orgânica, água e ar. As propriedades morfológicas que esse conjunto apresenta, como cor, textura, estrutura e

consistência, permitem dividi-lo em camadas aproximadamente paralelas à superfície, chamadas de horizontes (PEREIRA et al., 2019). A Figura 10 apresenta um perfil de solo com cinco horizontes de cores bem contrastantes.

Figura 10 - Perfil de solo com horizontes claramente diferenciados pela cor.



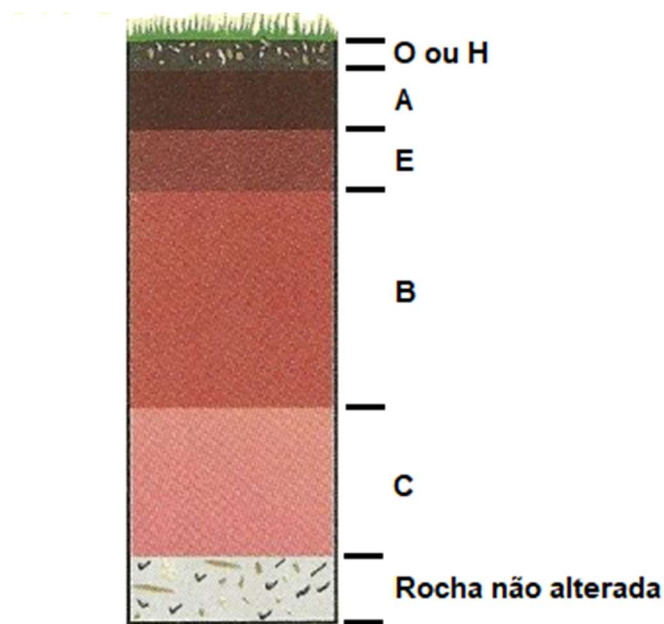
Fonte: Nesoil (2021).

Os principais horizontes, ilustrados na Figura 11, são:

- i. Horizonte O ou H: solo superficial em que predomina a matéria orgânica, diferenciado entre bem drenado (O) ou mal drenado (H). Costuma ter cores escuras;
- ii. Horizonte A: solo com presença de matéria orgânica, mas predominantemente mineral. Pode apresentar perda de material para horizontes inferiores;
- iii. Horizonte E: solo mineral mais claro, caracterizado pela perda de materiais para o horizonte inferior, principalmente argilas e óxidos de ferro (LEPSCH, 2010). Por conta disso, tende a apresentar textura arenosa (PEREIRA et al., 2019);
- iv. Horizonte B: solo mineral bem desenvolvido, isto é, tem características mais distantes das do material de origem. Pode apresentar acúmulo dos materiais perdidos pelos horizontes superiores; e

- v. Horizonte C: pode ser tanto um horizonte de sedimentos quanto um saprólito, rocha pouco alterada e que conserva significativamente as características do material de origem.

Figura 11 - Principais horizontes pedológicos.



Fonte: Adaptado de Lepsch (2010).

As propriedades físicas, químicas e morfológicas dos solos podem ser usadas para agrupá-los através de diversos sistemas de classificação. No Brasil, o sistema oficial é o Sistema Brasileiro de Classificação de Solos (SiBCS), elaborado pela Embrapa Solos, unidade da Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Atualmente, o SiBCS classifica os solos brasileiros em ordens, subordens, grandes grupos e subgrupos. As treze ordens são:

- i. Argissolos;
- ii. Cambissolos;
- iii. Chernossolos;
- iv. Espodossolos;
- v. Gleissolos;
- vi. Latossolos;
- vii. Luvisolos
- viii. Neossolos;
- ix. Nitossolos;
- x. Organossolos;

- xi. Planossolos;
- xii. Plintossolos; e
- xiii. Vertissolos.

2.2.1 Formação dos solos

Na superfície da Terra, a maioria das rochas está sob temperatura e pressão muito menores do que no ambiente em que foram formadas, encontrando-se instáveis química e fisicamente, o que as torna vulneráveis ao intemperismo (SCHAETZL; ANDERSON, 2005).

O intemperismo físico consiste na atuação de tensões que fissuram a rocha, geralmente entre os grãos de minerais, até fragmentá-la em partículas cada vez menores. As fissuras podem ser inicialmente causadas por um descarregamento da rocha, que então expande, ou pela sequência de expansão e contração em função de variações sucessivas de temperatura. Em locais em que a temperatura é baixa o suficiente, a água da chuva pode penetrar as fissuras e então congelar, aumentando de volume e causando mais tensões. O mesmo efeito é causado quando a água é salgada e evapora, formando cristais de sais. O atrito de geleiras, vento e água corrente de rios e mares também causa o intemperismo físico, desgastando a rocha (DAS, 2013).

O intemperismo químico consiste na remoção de minerais da rocha através de dissoluções ou diferentes reações químicas que os transformam em novos compostos, como íons, ácidos solúveis e argilominerais. O intemperismo químico é fortemente influenciado pelo físico, visto que uma maior fragmentação das partículas resulta em uma maior área de superfície, o que aumenta a velocidade das reações. Pelo mesmo motivo, rochas mais porosas são mais suscetíveis ao intemperismo químico.

Caso os produtos do intemperismo permaneçam no mesmo local em que foram gerados, eles tornam-se solos residuais, cuja principal característica é o aumento gradual do tamanho das partículas com a profundidade (DAS, 2013). Se os produtos são movidos para outro local por agentes como o vento e a gravidade, passam a se chamar solos transportados. O atrito das partículas durante o transporte caracteriza mais intemperismo físico, tendendo a arredondá-las. Um aspecto importante dos solos transportados é que eles não necessariamente compartilham características com a rocha do local em que se encontram.

De acordo com a NBR 6502 – Solos e Rochas, as partículas intemperizadas com diâmetros menores ou iguais a 2 milímetros já caracterizam areia. Os limites entre areia, silte e argila da NBR 6502 podem ser vistos no Quadro 1.

Quadro 1 – Diâmetros limites entre areia, silte e argila, de acordo com a NBR 6502 - Solos e rochas

Fração do solo	Diâmetro	Características principais
Areia	2,0 mm – 0,06 mm	Não coesivo e não plástico
Silte	0,06 mm – 0,002 mm	Baixa ou nenhuma plasticidade
Argila	< 0,002 mm	Apresenta coesão e plasticidade

Fonte: NBR 6502 – Solos e rochas (ABNT, 1995).

2.2.1.1 Fatores de formação dos solos

Os cientistas Dokuchaev e Hilgard perceberam, independentemente, que a mesma rocha, em diferentes locais, poderia formar diferentes solos (JENNY, 1941). O motivo seria que a formação dos solos é afetada por diferentes fatores, elencados em uma fórmula qualitativa por Dokuchaev, apresentada a seguir (JENNY, 1961):

$$S = f(\text{cl}, \text{o}, \text{p}) \text{ t}^{\circ}$$

Onde “S” é o solo, “cl” é o clima, “o” são os organismos vivos, “p” o substrato geológico (JENNY, 1961) e “t^o” é o tempo ou grau de maturação do solo, classificado como jovem, maduro ou antigo.

2.2.1.1.1 Modelo Clorpt

Baseado no trabalho de Dokuchaev e em diversas pesquisas posteriores sobre características quantitativas de solos, Jenny (1941) apresentou o modelo Clorpt, dado pela seguinte equação:

$$s = f(\text{cl}, \text{o}, \text{r}, \text{p}, \text{t}, \dots)$$

Onde “s” é uma propriedade qualquer do solo; “cl” é o fator do clima, “o” é o fator dos organismos vivos; “r” é o fator da topografia ou relevo (do inglês *relief*); “p” é o fator do material de origem (do inglês *parent material*); “t” é o fator do tempo decorrido desde o início do processo de formação do solo; e as reticências “...” indicam outros fatores que possam vir a ser adicionados, já que o autor não via motivos para acreditar que esses cinco eram os únicos. A equação permite o estudo da influência de cada fator ao analisar um grupo de solos que possuem em comum quatro dos fatores, enquanto o outro varia (SCHAETZL; ANDERSON, 2005).

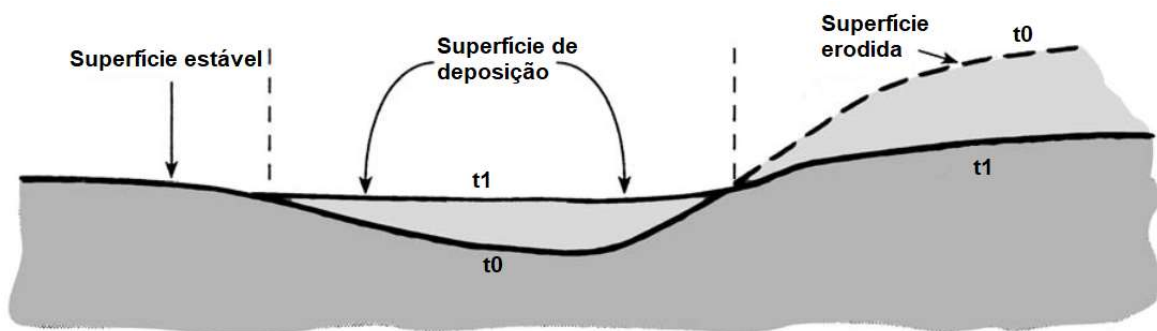
O clima é um dos fatores mais importantes (YAALON, 1983, apud SCHAETZL; ANDERSON, 2005), impactando grandemente tanto o intemperismo físico quanto o químico. Jenny (1941) ressalta que o clima é complexo ao ponto de que não é possível representá-lo por um único valor, de forma que é necessário dividi-lo em componentes, sendo os mais comuns a chuva e a temperatura. A água da chuva causa as reações de hidratação, hidrólise e dissolução, também sendo fundamental na carbonatação, por reagir com o gás carbônico da atmosfera e incidir sobre a rocha na forma de chuva ácida (SCHAETZL; ANDERSON, 2005). Ainda, funciona como meio de transporte para partículas pequenas através dos horizontes ou adicionando e removendo material no perfil, em relevos inclinados. A temperatura, além de afetar o intemperismo físico, afeta o intemperismo químico, visto que reações químicas aceleram com o aumento da temperatura. Por conta do fator clima, os solos tendem a ser menos desenvolvidos e espessos em regiões de clima temperado quando comparados a solos de regiões de clima tropical, que possuem altas temperaturas e taxas de precipitação (PEREIRA et al., 2019).

Organismos variados podem influenciar a formação dos solos. A vegetação macroscópica pode penetrar fissuras em rochas e expandi-las com o crescimento de suas raízes, além de aumentar a porosidade do solo, facilitando a infiltração da água, da mesma forma que

animais macroscópicos como minhocas, cupins e roedores. Essa maior infiltração diminui o escoamento superficial, que é a principal fonte de erosão causada pela chuva (JENNY, 1941). As folhas também diminuem o escoamento superficial ao reter parte da água, que evapora antes de atingir o solo. Lepsch (2010) destaca que o baixo teor de finos na camada superficial dos solos do semiárido nordestino brasileiro é causado pela baixa proteção à erosão que a vegetação esparsa da caatinga fornece. Já os microrganismos influenciam principalmente o intemperismo químico. Fungos, por exemplo, quando associados a plantas, captam nutrientes ao degradar minerais primários com ácidos orgânicos (LANDEWEERT et al., 2001, apud SCHAETZL; ANDERSON, 2005), enquanto fungos associados a cupins são capazes de degradar argilominerais (JOUQUET et al., 2002, apud SCHAETZL; ANDERSON, 2005).

O fator relevo atua na formação dos solos de diversas formas, sendo geralmente o fator no qual relações quantitativas são mais facilmente encontradas (MCBRATNEY et al., 2003). Um aumento na inclinação do terreno causa um aumento na perda de material do perfil, tanto pela erosão do escoamento superficial (que ganha velocidade) quanto subsuperficialmente, carregando finos e íons solúveis. A perda de material também pode ocorrer apenas pela gravidade, formando, na base de encostas, o chamado solo coluvionar, como mostra a Figura 12. Terrenos muito inclinados tendem a ter um solo menos espesso do que terrenos planos, podendo apresentar apenas rocha exposta (LEPSCH, 2010).

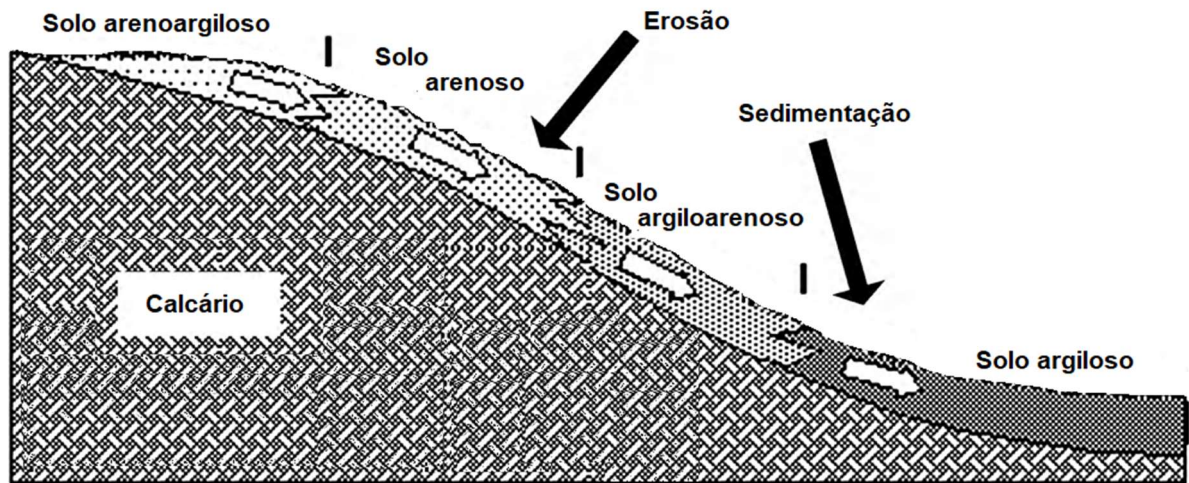
Figura 12 - Erosão de encosta com deposição de colúvio.



Fonte: Adaptado de Schaetzl e Anderson (2005).

Milne (1936, apud HUANG et al., 2012) cunhou o termo catena ao identificar que uma mesma sequência de solos pode se repetir em encostas de uma região, ilustrada na Figura 13.

Figura 13 - Perfil de uma catena caracterizada pela perda de finos.

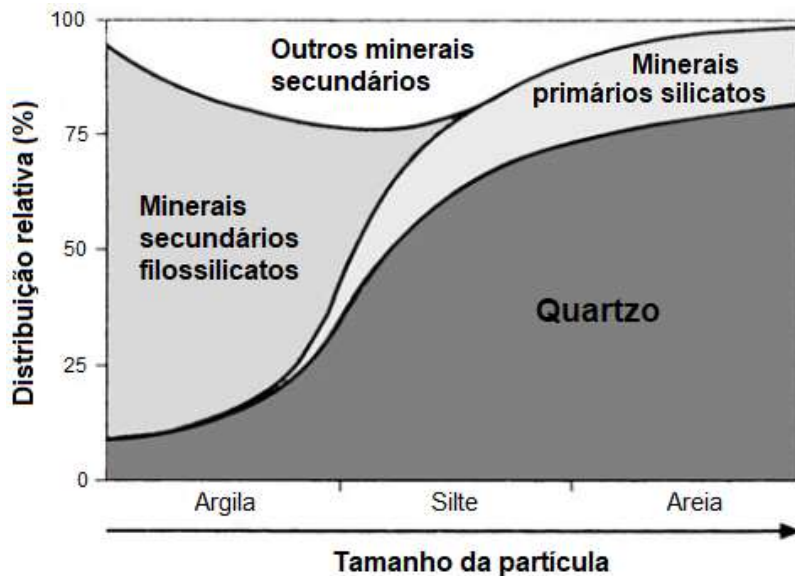


Fonte: Adaptado de Milne (1936, apud HUANG et al., 2012).

Além da inclinação, o fator relevo pode ser levado em consideração através da altitude, da posição (terço inferior, médio ou superior da encosta), orientação e curvatura da superfície, ou ainda através de outros atributos secundários derivados de modelos digitais de elevação (MCBRATNEY et al., 2003).

O fator material de origem é definido por Jenny (1941) como o estado inicial do sistema de solo, podendo ser rocha ou sedimento, consolidado ou inconsolidado. Os diferentes minerais que compõem o material de origem possuem diferentes resistências ao intemperismo, de forma que certos minerais são mais comuns em certas faixas granulométricas de solo, como mostra a Figura 14. Materiais de origem ricos em quartzo e feldspato tendem a originar solos arenosos, enquanto os ricos em olivina, piroxênios e biotita tendem a originar solos mais argilosos (PEREIRA et al., 2019). A aplicação desse fator não é simples por conta da dificuldade em definir se o perfil todo possui apenas um material de origem e, no caso de haver mais, quais seriam.

Figura 14 - Relação entre tamanho de partícula e mineralogia.



Fonte: Adaptado de Schaetzl e Anderson (2005).

O fator tempo simboliza todo o desenvolvimento do solo ocorrido desde o início do processo de formação, seja no início da alteração de uma rocha sã ou no fim da deposição de um sedimento (LEPSCH, 2010).

2.2.1.1.2 Modelo Scorpan

O modelo Scorpan é uma evolução do modelo Clorpt e tem como objetivo gerar funções espaciais de predição de solo, sendo dado pela seguinte equação (MCBRATNEY et al., 2003):

$$S = f(s, c, o, r, p, a, n)$$

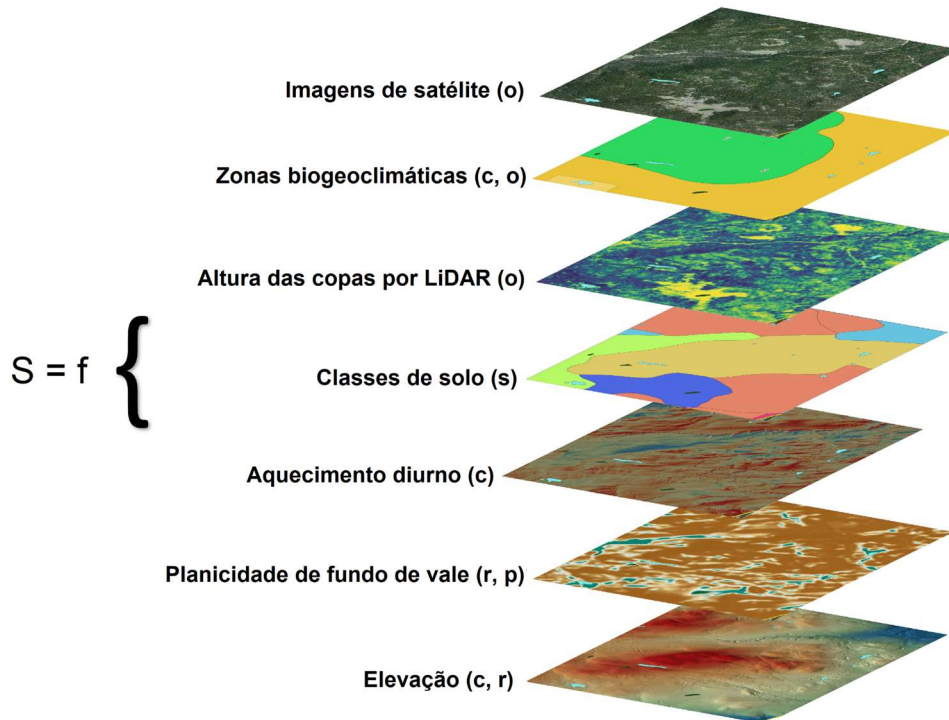
Onde “S” é a propriedade do solo que se deseja prever; “s” é outra propriedade do solo da qual haja dados em maior densidade; “c” é o fator clima; “o” é o fator organismos; “r” é o fator relevo; “a” é o fator tempo (do inglês *age*); e “n” é a posição espacial.

O fator “s” pode advir de outros mapas de solo, como classificações, e sensoriamento remoto ou proximal, como sensores multiespectrais, hiperespectrais, de radiação gama e de indução eletromagnética (HEUNG et al., 2021).

O fator “n” pode ser utilizado explicitamente levando em consideração a posição relativa das amostras, como com o uso de geoestatística; ou através de mapas de distância até acidentes geográficos, como montanhas e rios, que são particularmente interessantes para o estudo de

solos coluvionares e aluvionares, respectivamente (HEUNG et al., 2021). A Figura 15 ilustra algumas variáveis que podem ser utilizadas no mapeamento digital de solos com o modelo Scorpan.

Figura 15 - Exemplos de variáveis que podem representar os fatores do modelo Scorpan.



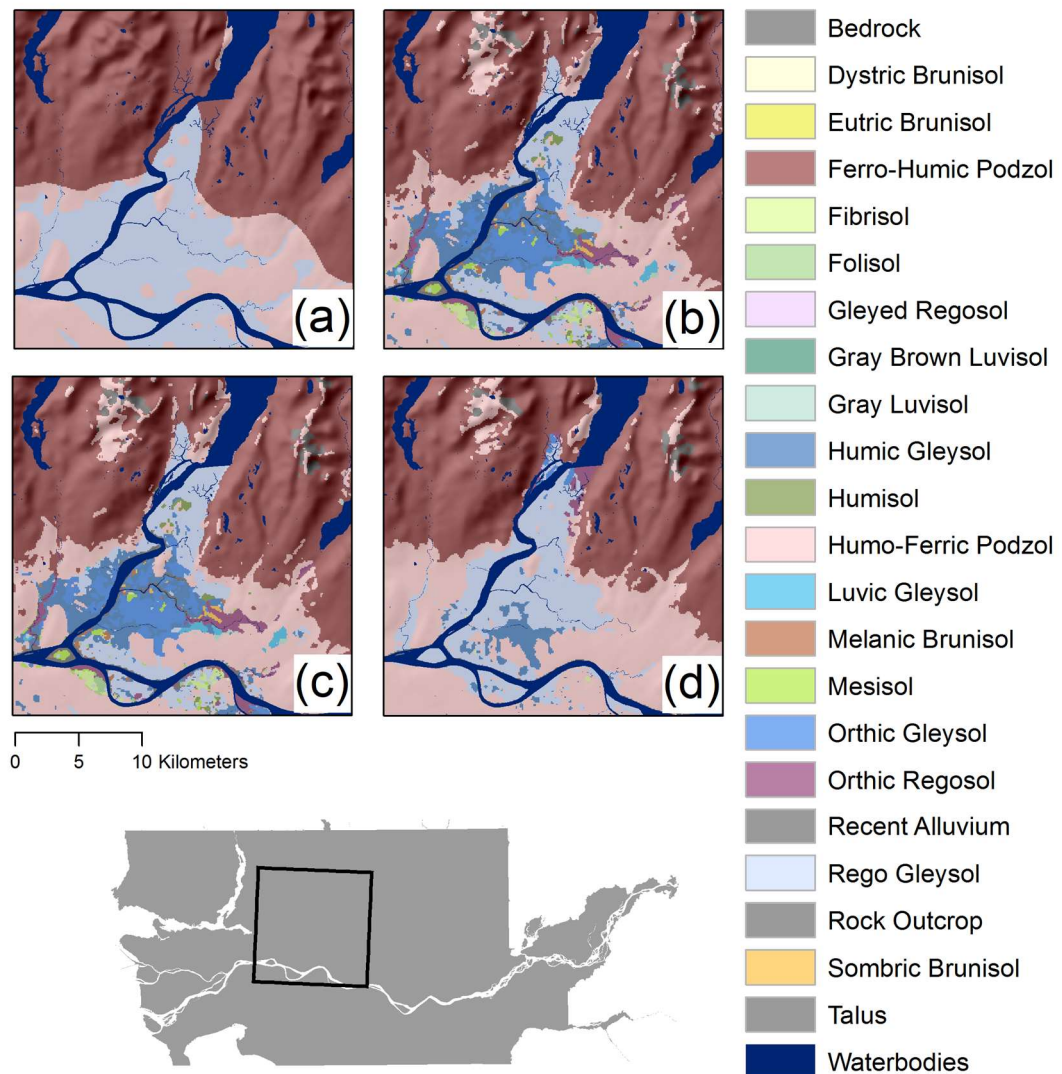
Fonte: Adaptado de Heung et al. (2021).

As principais fontes de dados para o mapeamento digital de solos são levantamentos de solos antigos e sensores em satélites, como os multiespectrais e os radares. Os sensores multiespectrais, de acordo com Boettinger et al. (2008), podem ser utilizados para representar diversos fatores, sendo particularmente úteis em regiões de clima árido e semiárido, onde a vegetação esparsa permite uma maior captação da radiação do solo. Já os radares são utilizados principalmente para adquirir informações da topografia, gerando modelos digitais de terreno.

Encontrar a função que relaciona os fatores com o solo é um problema genérico de regressão múltipla, geralmente sendo não linear (MCBRATNEY et al., 2003). Esse problema pode ser resolvido através de diversos algoritmos, como regressão linear simples e múltipla, análise de discriminante linear, modelos lineares generalizados, modelos aditivos generalizados, redes neurais artificiais, sistemas *fuzzy*, sistemas especialistas, árvores de decisão e *random forest*, entre outros (MCBRATNEY et al., 2003; KIENAST-BROWN et al.,

2017; HEUNG et al., 2021). A escolha do algoritmo pode impactar fortemente nos resultados, conforme a Figura 16, que mostra a diferença entre os resultados de quatro diferentes algoritmos em um problema de classificação de solos. Heung et al. (2021) destacam que não existe um algoritmo que sempre terá o melhor desempenho, de forma que o ideal é utilizar o máximo possível de algoritmos e comparar os resultados, de acordo com limitações de recursos computacionais, econômicos e de tempo.

Figura 16 - Resultado de classificação de solos no Rio Fraser, Canadá, utilizando diferentes algoritmos para a função f: (a) árvore de decisão, (b) árvore de decisão com *bagging*, (c) *random forest*, e (d) rede neural artificial.



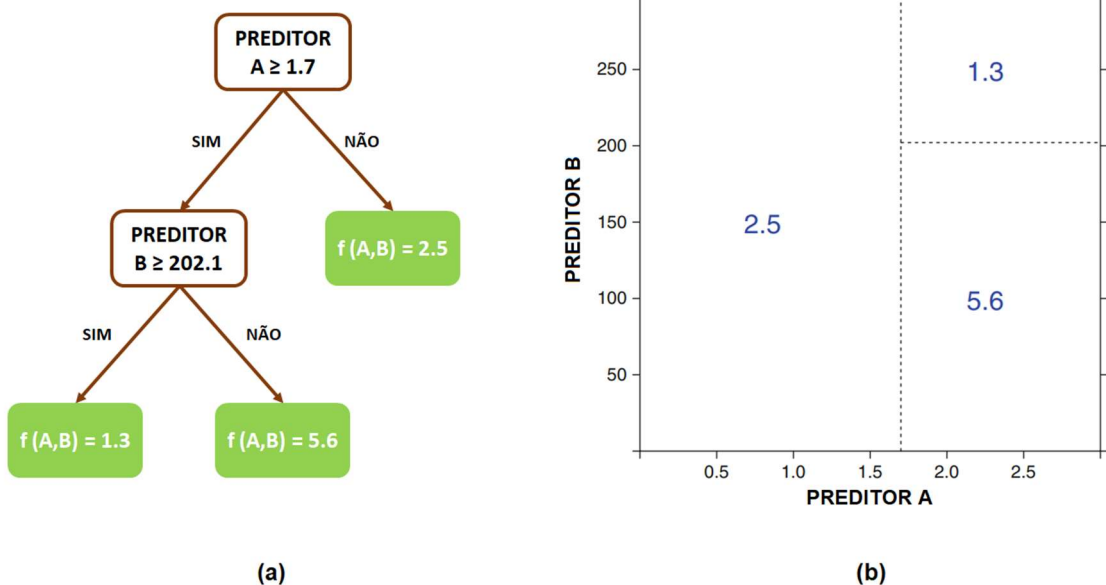
Fonte: Adaptado de Heung et al. (2021).

De acordo com McBratney et al. (2003), os modelos de árvore de decisão são amplamente utilizados para a predição de propriedades do solo por serem capazes de lidar com não

linearidades, comportamento estatístico não aditivo e tanto variáveis quantitativas quanto qualitativas; por não requererem suposições sobre o comportamento dos dados; e, principalmente, por serem de fácil interpretação, diferentemente de outros algoritmos como os modelos lineares generalizados e as redes neurais artificiais. Dentre os usos de modelos de árvore de decisão, podemos destacar o mapeamento de teores de areia, silte e argila (CHAGAS et al., 2016; HATEFFARD et al., 2019; DUCHESNE e OUIOMET, 2021; PEREIRA et al., 2021), classes de drenagem (MØLLER et al., 2019), risco de inundação (WANG et al., 2015) e suscetibilidade a deslizamentos (YEON et al., 2020).

Árvores de decisão realizam previsões particionando os dados através de uma sequência de regras. No modelo gráfico da árvore, as regras são chamadas de nós, as saídas das regras são os galhos e as previsões são as folhas (SEKULIC et al., 2020). O valor atribuído pelas folhas é a média dos valores dos dados de treinamento que se enquadram em cada sequência de regras. A Figura 17 apresenta um exemplo de árvore de decisão com duas variáveis preditoras, que, no caso do mapeamento digital de solos, poderiam ser duas covariáveis ambientais do modelo Scorpan.

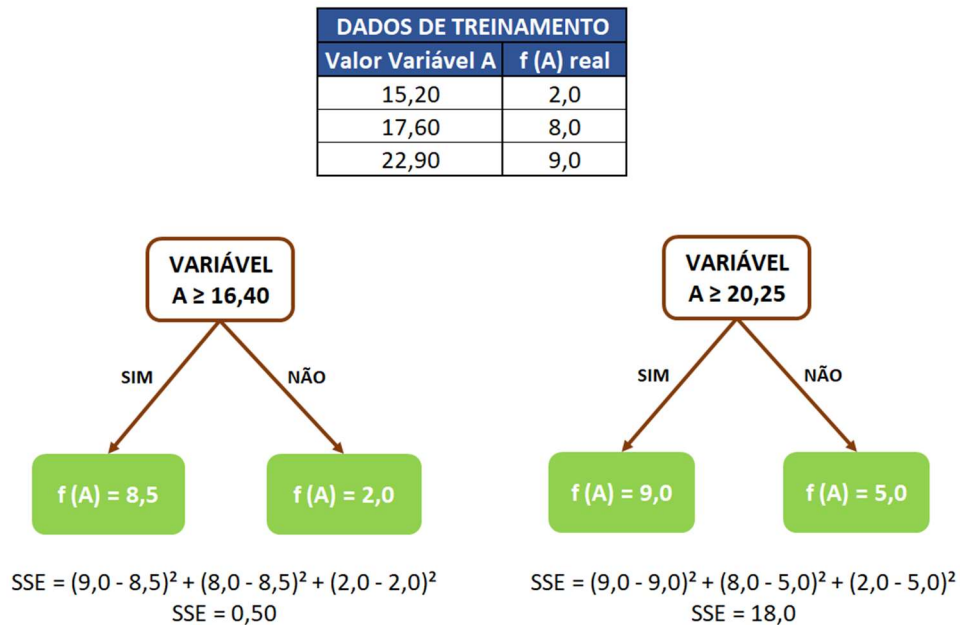
Figura 17 - Exemplo de árvore de decisão representada em (a) modelo de árvore; e (b) espaços de predição em um plano cartesiano.



Fonte: Adaptado de Kuhn e Johnson (2016).

Em cada nó, seleciona-se uma variável e um valor limite para partição, buscando minimizar a soma dos quadrados dos erros (SSE, do inglês *sum of squared errors*) (KUHN; JOHNSON, 2016). A construção da árvore começa pela definição da variável do primeiro nó, chamado de nó raiz, em que todas as variáveis e limites disponíveis devem ser testados. Os valores limites testados são as médias entre cada par de dados subsequentes. A Figura 18 mostra um exemplo desse processo, em que o limite definido para a variável A é de 16,40, por apresentar o menor SSE. Cada um dos subconjuntos resultantes da regra do nó raiz continua a ser dividido, escolhendo sempre o menor SSE.

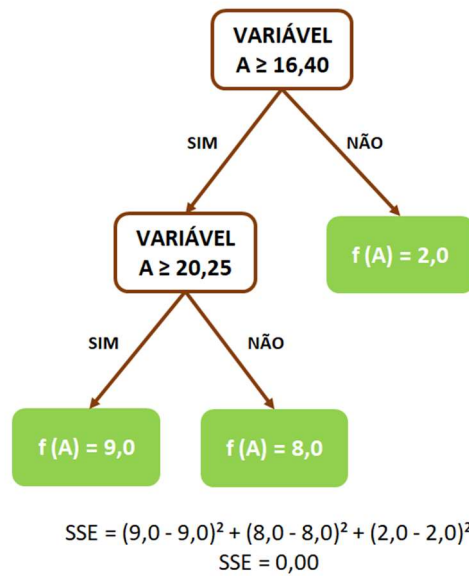
Figura 18 - Exemplo de seleção da variável e limite do nó raiz.



Fonte: Adaptado de Starmer (2021).

O principal problema das árvores de decisão é a tendência de sobreajuste, isto é, a baixa robustez do modelo, o que acarreta em aumento dos resíduos pela sensibilidade a ruídos nos dados de treinamento (SEKULIC et al., 2020). Cada subconjunto do nó raiz pode continuar sendo reduzido até o número de folhas igualar o número de dados, de forma que o SSR será zero. A Figura 19 mostra esse caso no exemplo da Figura 18.

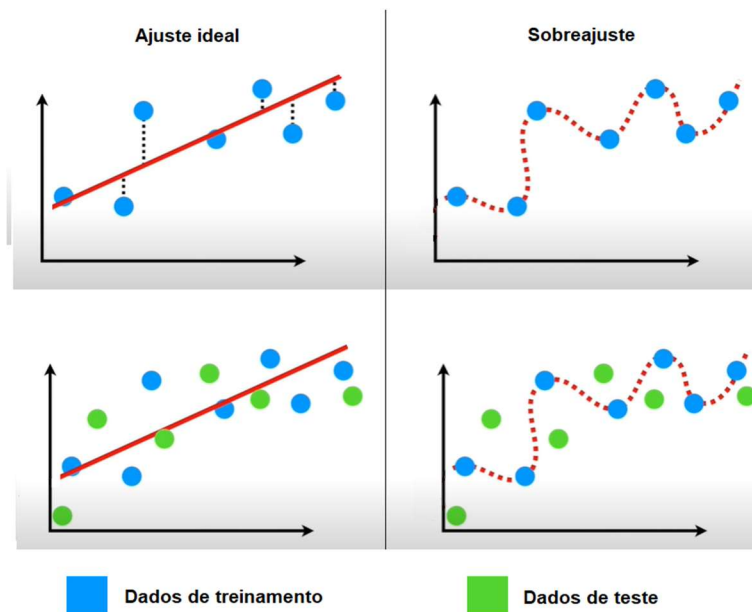
Figura 19 - Árvore com número de folhas igual ao número de dados e SSE igual a zero.



Fonte: Adaptado de Starmer (2021).

Entretanto, dificilmente esse modelo terá um bom desempenho ao ser aplicado em um novo conjunto de dados, conforme ilustra a Figura 20. Um modelo sobreajustado apresenta pouco ou nenhum erro em relação aos dados que foram utilizados para criá-lo, mas um erro alto ao ser comparado com novos dados.

Figura 20 - Exemplo da falta de robustez de um modelo sobreajustado.



Fonte: Adaptado de Starmer (2021).

Uma forma de amenizar o problema do sobreajuste é através da poda da árvore (do inglês *pruning*), que consiste na remoção de nós segundo algum processo, como definir um valor mínimo de redução do SSR ou um número mínimo de dados por nó para que a divisão ocorra. (KUHN; JOHNSON, 2016).

Random forest é o algoritmo de *machine learning* mais utilizado no mapeamento digital de solos (KHALEDIAN; MILLER, 2020). Tem como base a árvore de decisão, mas utiliza o conceito de *bootstrap aggregating* para contornar o problema de sobreajuste e aumentar a precisão do modelo (HASTIE et al., 2009). *Bootstrapping* é o processo de reamostrar aleatoriamente o conjunto de dados original, criando novos conjuntos. Por ser aleatório, pode haver repetição nas amostras dentro de um mesmo conjunto *bootstrap*, de forma que parte dos dados – aproximadamente um terço, de acordo com Kuhn e Johnson (2016) – não aparecerá em nenhum conjunto, sendo chamado de dados *out-of-bag*. Esses dados são utilizados como conjunto de teste para aferir a precisão do modelo através do erro *out-of-bag*, dispensando a necessidade de um conjunto de testes adicional (BREIMAN, 2001, apud KUHN; JOHNSON, 2016). A partir de cada conjunto *bootstrap* é criada uma árvore de decisão, sendo que, ao particionar cada nó, apenas um subconjunto aleatório das variáveis preditoras disponíveis é considerado. O valor da predição para um determinado ponto será a média dos valores preditos por todas as árvores. Segundo Khaledian e Miller (2020), *random forests* são menos sensíveis a conjuntos de dados pequenos do que algoritmos como redes neurais, por exemplo.

O número de árvores e o tamanho do subconjunto aleatório de variáveis preditoras são parâmetros que devem ser definidos pelo usuário. Breiman (2001, apud KUHN; JOHNSON, 2016) recomenda um terço do número de variáveis disponíveis para o subconjunto aleatório, enquanto o crescimento do número de árvores apenas aumenta a precisão do modelo (HASTIE et al., 2009), embora tenda a estabilizar após um certo ponto. O tamanho do conjunto *bootstrap* também pode ser otimizado (PROBST et al., 2019).

Para avaliar a importância de cada variável em um modelo de *random forest*, os pontos *out-of-bag* de cada árvore são utilizados. Os valores da variável em análise são embaralhados entre os pontos, enquanto as outras variáveis permanecem inalteradas. O modelo é, então, aplicado nos pontos e o erro é calculado. Assim, a variável mais importante para o modelo é aquela cujo embaralhamento causa o maior aumento do erro, em relação ao erro do conjunto *out-of-bag* original (LIAW; WIENER, 2002).

3 ÁREA DE ESTUDO

A área de estudo deste trabalho é uma UFV de aproximadamente 10,2 km² e 8.960 *trackers* de 19 estacas, totalizando 170.240 estacas. As fundações adotadas foram estacas metálicas de perfil I e C com comprimento enterrado variando entre 1,4 m e 2,8 m. Além disso, foram empregados em torno de um milhão de módulos fotovoltaicos, que somam uma potência instalada de 475 MWp.

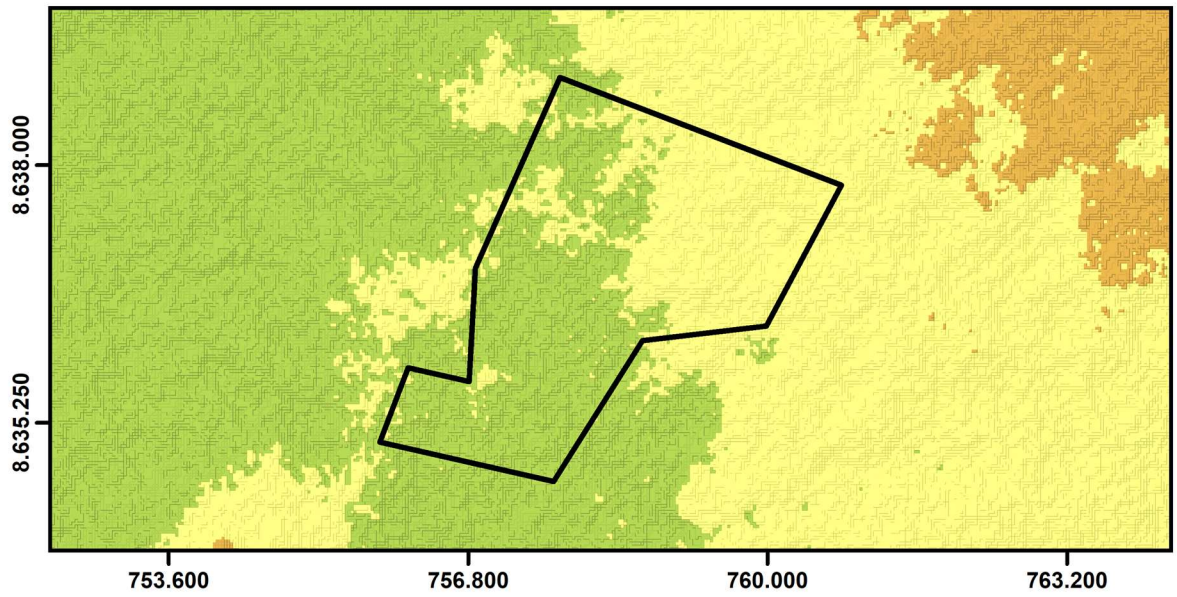
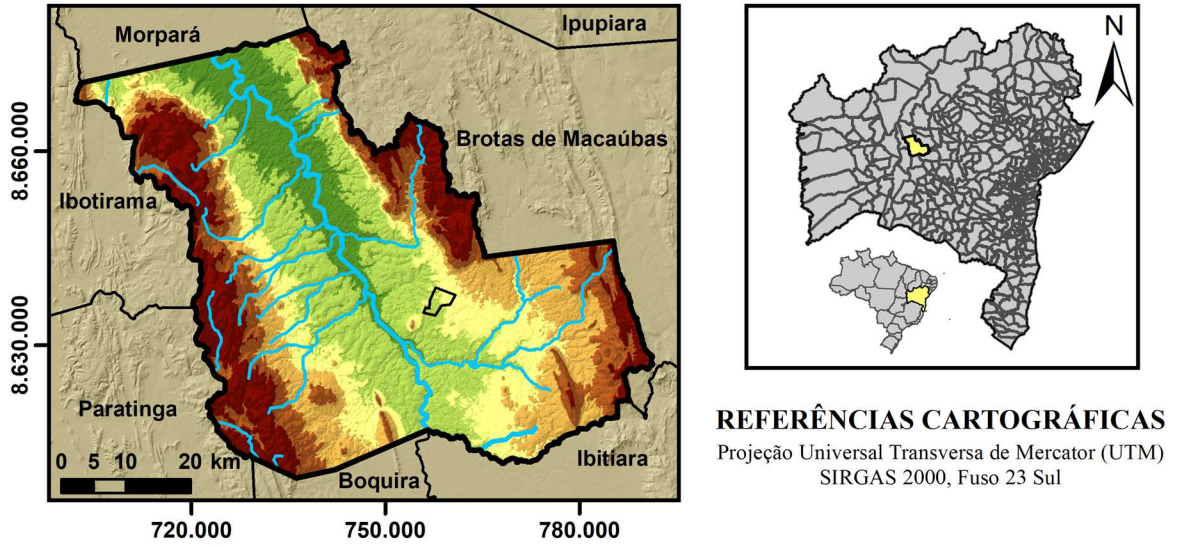
A UFV fica localizada no município de Oliveira dos Brejinhos, no estado da Bahia, conforme Figura 21. O município possui uma área de 3.315 km², 21.839 habitantes (IBGE, 2010) e é atravessado pelo Rio Paramirim, um afluente da margem direita do Rio São Francisco.




Segundo Consórcio SETA-Araxá (2020, apud SOARES, 2021), as formações geológicas da região contam com quartzitos, micaxistos, arenitos, pelitos, conglomerados e gnaisses. Quanto aos solos, a UFV está inserida totalmente na unidade de mapeamento LAd6 (IBGE, 2022). Esta unidade é formada por uma associação de latossolos e argissolos. O latossolo é um tipo de solo com baixo teor de matéria orgânica, poroso, bem drenado e com perfis profundos (AGÊNCIA EMBRAPA DE INFORMAÇÃO TECNOLÓGICA, 2022). O argissolo também tem baixo teor de matéria orgânica, porém possui drenagem e profundidade variáveis (EMBRAPA, 2006).

Pela classificação de Köppen, o clima é semiárido quente (ALVARES et al., 2014), caracterizado por temperatura média alta; forte insolação e baixa nebulosidade; intensa evaporação e baixa umidade relativa do ar; e chuva escassa, cuja concentração em curtos períodos de tempo pode causar enchentes em certas épocas e secas em outras (CARVALHO, 2003).

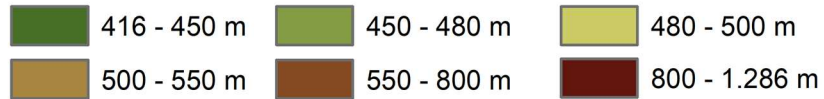
O bioma da região é a caatinga, caracterizada pela vegetação baixa, decídua e esparsa, com pouca grama. De acordo com Andrade-Lima (1981), a vegetação da caatinga é altamente correlacionada com aspectos geopedológicos. Isso é demonstrado em regiões de temperaturas e umidade muito parecidas, em que a vegetação é de caatinga quando as rochas são originárias do escudo cristalino pré-cambriano brasileiro, e de cerrado quando as rochas são da bacia do Rio Parnaíba. Outro exemplo dessa relação é a presença de espécimes dos gêneros *Pilosocereus*, *Poeppigia*, *Dalbergia* e *Piptadenia* coincidir com as rochas sedimentares da série da Serra do Cipó.

Figura 21 - Localização do município de Oliveira dos Brejinhos e área de estudo.



 Hidrografia  Área de estudo  Municípios vizinhos

HIPSOMETRIA



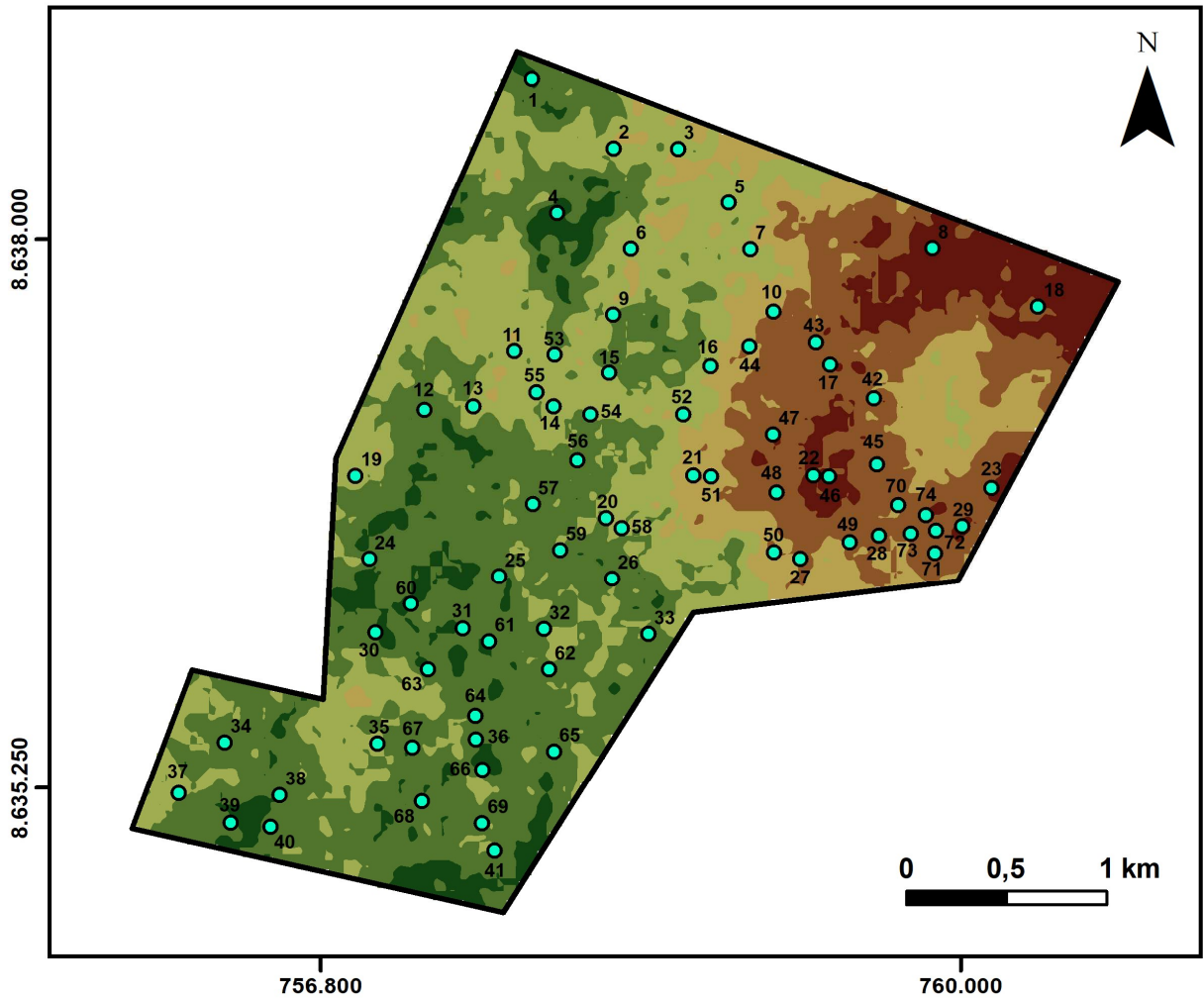
Fonte: Elaborada pelo autor.

4 MATERIAIS E MÉTODO

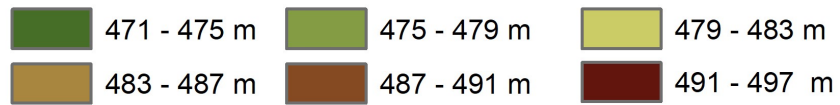
Inicialmente, 90 relatórios de ensaios SPT executados em janeiro de 2020 foram disponibilizados por uma empresa envolvida no projeto da UFV. Os ensaios foram contratados até a profundidade de 4 metros. O autor não esteve envolvido e também não teve acesso aos critérios e premissas utilizados na definição da quantidade e disposição dos ensaios. Dos 90 relatórios, 16 foram descartados por apresentarem inconsistências nos critérios de parada ou por não atenderem a NBR 6484, restando 74 ensaios. Destes, 38 ensaios não atingiram a profundidade de 4 metros. Por conta do baixo número de pontos restantes para 4 metros, combinado com a incerteza associada aos critérios de parada e o fato de que a estaca mais longa utilizada para esse projeto foi de 2,8 metros, optou-se por não trabalhar com essa profundidade.

Dentre os 74 ensaios, 12 não atingiram a profundidade de 3 metros. Portanto, foram utilizados 74 pontos para as profundidades de 1 e 2 metros, e 62 pontos para a profundidade de 3 metros. A Figura 22 mostra a disposição das sondagens na área de estudo, segundo a identificação da Tabela 1.

Figura 22 - Localização dos ensaios SPT



HIPSOMETRIA



REFERÊNCIAS CARTOGRÁFICAS

Projeção Universal Transversa de Mercator (UTM)
SIRGAS 2000, Fuso 23 Sul

Fonte: Elaborada pelo autor.

A Tabela 1 resume os ensaios SPT utilizados.

Tabela 1 - Resumo dos ensaios SPT, com sistema de referência SIRGAS 2000 e coordenadas UTM, fuso 23, sul.

Identificação	N _{SPT_1m}	N _{SPT_2m}	N _{SPT_3m}	Coord. X	Coord. Y
1	15	30	59	757856	8638800
2	6	10	25	758264	8638460
3	8	12	15	758588	8638450
4	25	37	45	757983	8638130
5	7	12	28	758840	8638180
6	10	17	24	758350	8637950
7	8	18	53	758948	8637950
8	7	15	24	759860	8637950
9	27	66	88	758263	8637620
10	9	19	79	759065	8637630
11	17	25	19	757769	8637440
12	18	28	41	757318	8637140
13	9	18	39	757565	8637160
14	9	12	27	757966	8637160
15	12	21	27	758243	8637330
16	12	60	82	758749	8637360
17	6	9	19	759347	8637370
18	8	11	24	760388	8637660
19	12	28	80	756974	8636810
20	21	75	-	758228	8636600
21	7	8	61	758663	8636810
22	8	11	16	759264	8636820
23	10	17	22	760155	8636750
24	17	28	55	757043	8636400
25	16	29	57	757693	8636310
26	28	75	-	758258	8636300
27	8	14	25	759199	8636400
28	8	17	25	759591	8636510
29	7	15	20	760008	8636560
30	21	85	-	757074	8636020
31	17	23	70	757510	8636050
32	13	26	60	757916	8636040
33	11	80	-	758440	8636020

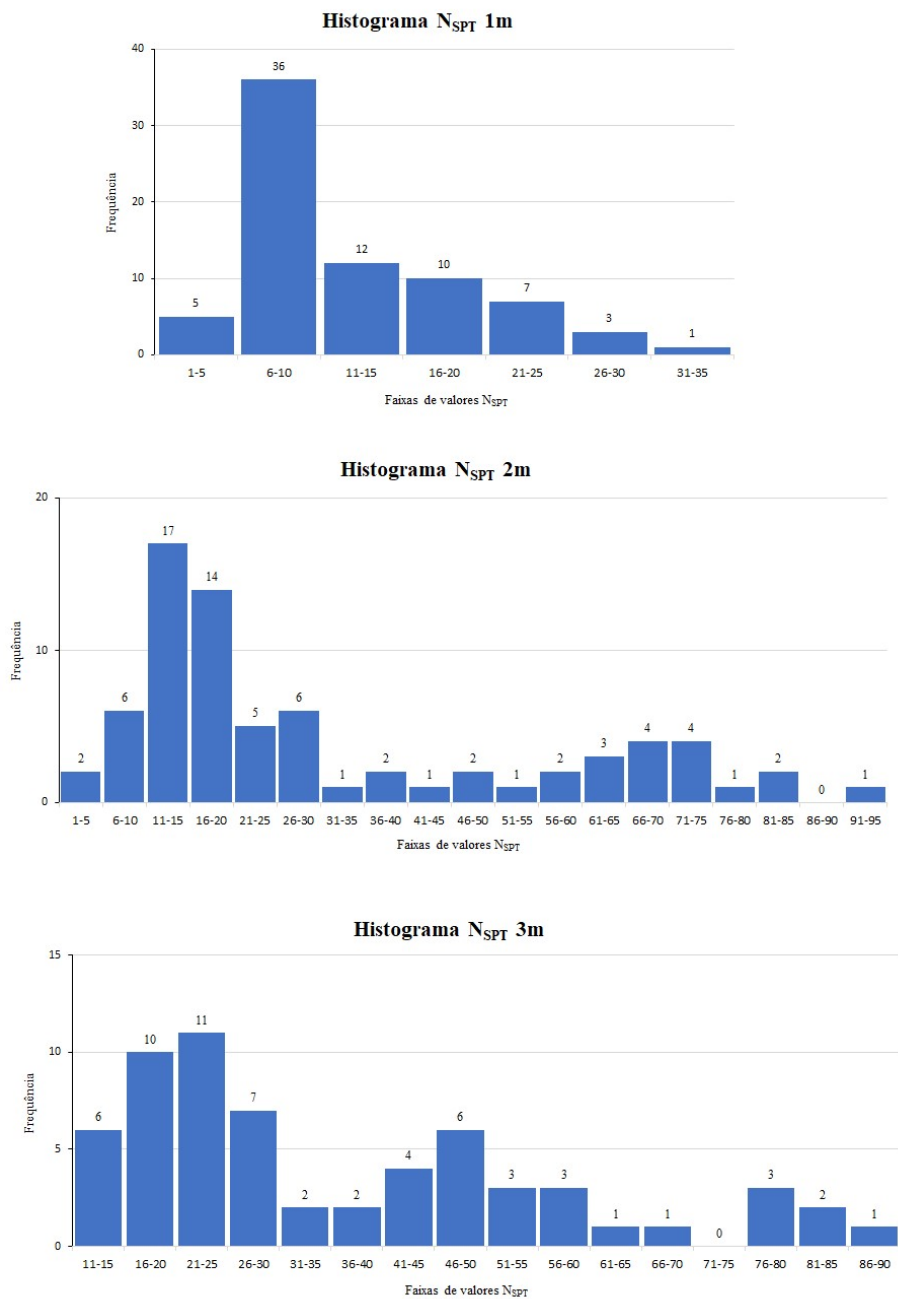
34	21	41	80	756320	8635480
35	7	81	-	757084	8635470
36	16	70	-	757576	8635490
37	15	23	23	756090	8635220
38	16	20	32	756595	8635210
39	31	61	50	756350	8635070
40	18	31	42	756550	8635050
41	27	62	81	757671	8634930
42	8	9	30	759566	8637200
43	8	13	20	759277	8637480
44	8	13	22	758943	8637460
45	8	17	17	759582	8636870
46	6	11	22	759342	8636810
47	7	15	20	759062	8637020
48	6	11	22	759080	8636730
49	6	13	54	759445	8636480
50	8	19	49	759068	8636430
51	8	19	50	758751	8636810
52	9	25	35	758613	8637120
53	8	17	19	757969	8637420
54	8	12	27	758149	8637120
55	8	15	-	757879	8637230
56	21	68	26	758084	8636890
57	11	72	-	757861	8636670
58	7	17	-	758305	8636550
59	8	37	-	757999	8636440
60	16	54	50	757251	8636170
61	22	74	-	757642	8635980
62	17	57	30	757942	8635840
63	11	92	50	757338	8635840
64	11	50	50	757575	8635610
65	25	66	42	757967	8635430
66	15	61	-	757610	8635340
67	8	19	36	757258	8635450
68	11	50	14	757307	8635180
69	7	19	15	757607	8635070
70	3	4	16	759688	8636670
71	5	7	15	759872	8636420

72	5	12	15	759877	8636540
73	5	8	16	759752	8636520
74	3	5	15	759828	8636620

Fonte: Elaborada pelo autor.

A Figura 23 mostra os histogramas de distribuição dos valores dos ensaios, por profundidade.

Figura 23 - Histograma de valores NSPT utilizados, por profundidade.



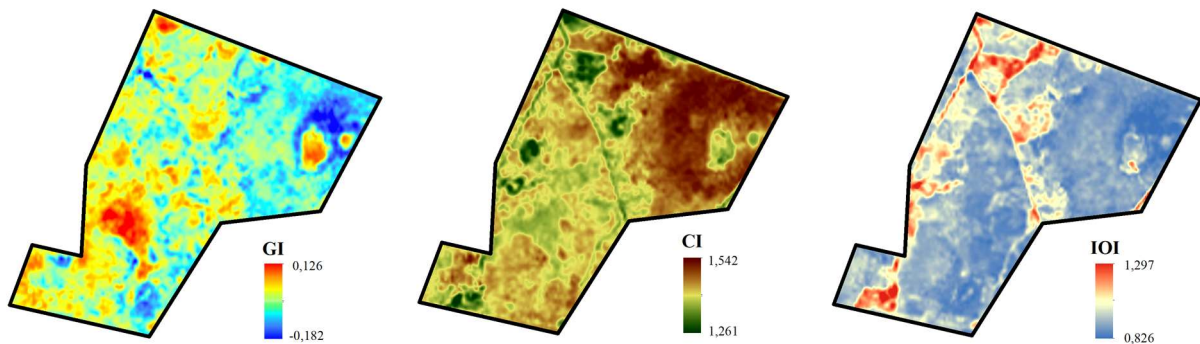
Fonte: elaborada pelo autor.

4.1 COVARIÁVEIS AMBIENTAIS

Buscou-se dados utilizados em estudos de mapeamento anteriores para representar cada fator do modelo Scorpan. Apenas dados de livre acesso foram considerados. Todos os mapas e informações espaciais foram convertidos para o sistema de referência SIRGAS2000 e projetados no sistema de coordenadas UTM, Fuso 23, hemisfério sul. Para os índices de sensoriamento remoto, foi utilizada uma imagem do satélite Landsat 8, cortesia do *U.S. Geological Survey*, de 30 de janeiro de 2019, correspondente ao período chuvoso da região.

O fator solo foi expresso através do Índice de Gipsita (GI, do inglês *Gypsum Index*), do Índice de Argila (CI, do inglês *Clay Index*) (TAGHIZADEH-MEHRJARDI et al., 2016) e do Índice de Óxidos Ferro (IOI, do inglês *Iron Oxide Index*) (CHAGAS et al., 2010), índices de sensoriamento remoto que indicam a presença de cada um destes elementos no solo. O GI é dado pela diferença normalizada das bandas de infravermelho de ondas curtas 1 (banda 6) e infravermelho próximo (banda 5); o CI é dado pela razão das bandas de infravermelho de ondas curtas 1 e 2 (bandas 6 e 7); e o IOI é dado pela razão da banda vermelha (banda 4) e banda azul (banda 2). A Figura 24 ilustra a distribuição das variáveis do fato Solo na UFV.

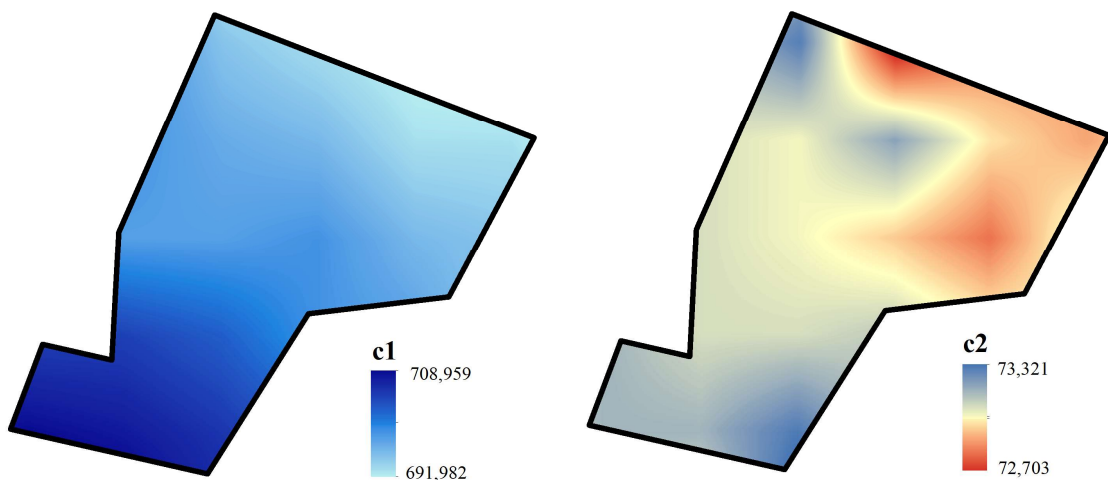
Figura 24 - Distribuição espacial das variáveis do fator solo.



Fonte: elaborada pelo autor.

Para o fator clima, foram utilizadas as variáveis Precipitação Anual e Isotermalidade (razão da variação máxima diurna da temperatura e variação máxima anual) do conjunto de dados WorldClim (FICK; HIJMANS, 2017), utilizado por Dharumarajan et al. (2021). O WorldClim disponibiliza a média dos dados de 1970 a 2000 em resolução de aproximadamente 1 km², que foi reamostrada para a resolução de aproximadamente 30 m das outras variáveis. A Figura 25 mostra a distribuição espacial das variáveis de clima na área de estudo.

Figura 25 – Distribuição espacial das variáveis do fator clima, precipitação anual (c1) e isotermalidade (c2).

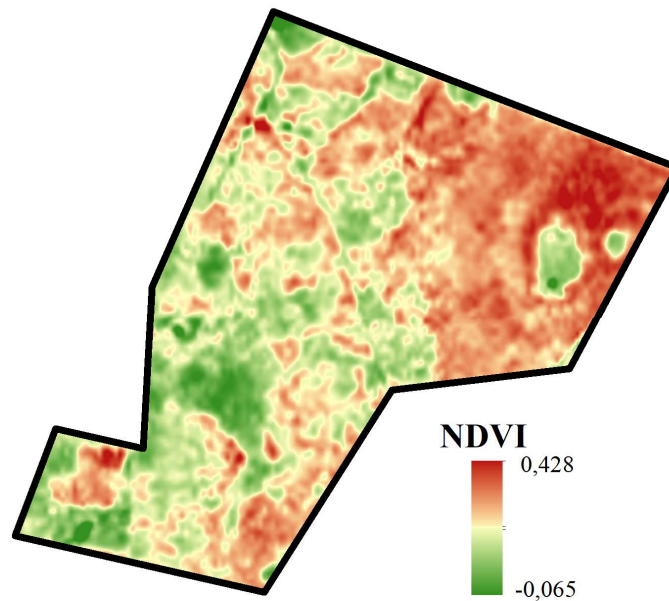


Fonte: elaborada pelo autor.

O fator organismos foi representado pelo NDVI (do inglês *Normalized Difference Vegetation Index*), que é a diferença normalizada entre as bandas de infravermelho e vermelho de um sensor multiespectral. Normalizado com valores entre -1 e 1, o NDVI indica vegetação com valores próximos de 1, sendo que, quanto maior o valor, mais saudável é a planta. O NDVI é a forma mais comum de representar o fator organismos (BOETTINGER et al., 2008) e, segundo Chagas et al. (2010), diversos estudos já atestaram a correlação desse índice com

propriedades do solo. Para o cálculo do índice, foram utilizadas as bandas 5 e 4 das imagens do satélite Landsat 8. O mapa de NDVI da área de estudo pode ser visto na Figura 26.

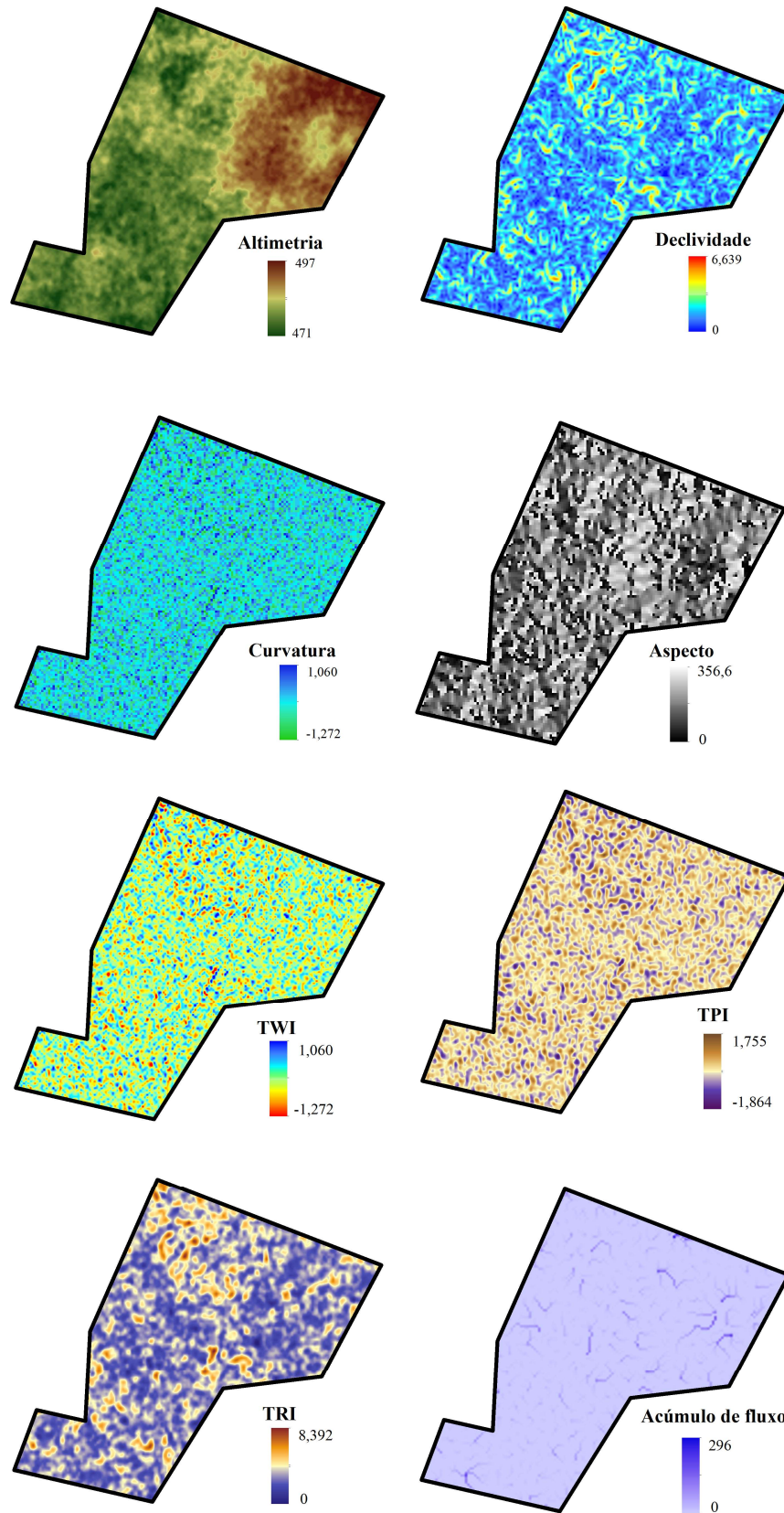
Figura 26 – Distribuição espacial da variável do fator organismos.



Fonte: elaborada pelo autor.

A base do fator relevo foi o modelo digital de terreno do SRTM (do inglês *Shuttle Radar Topography Mission*) (FARR et al., 2017), de resolução aproximada de 30 m. Dele, foram extraídas as variáveis altimetria, declividade, curvatura, aspecto (ângulo de orientação das faces do terreno em relação ao norte), o Índice Topográfico de Umidade (TWI, do inglês *Topographic Wetness Index*), o Índice de Posição Topográfica (TPI, do inglês *Topographic Position Index*), o Índice de Aspereza do Terreno (TRI, do inglês *Terrain Ruggedness Index*) (DHARUMARAJAN et al., 2017), e o acúmulo de fluxo (MØLLER et al., 2019). O TWI indica a relação da topografia com a drenagem através do logaritmo da razão entre área de contribuição e declividade (SØRENSEN et al., 2006). O TPI é a diferença da altitude de uma célula (ou pixel) e a média de suas células vizinhas, enquanto o TRI é a média da altitude das diferenças uma célula e suas células vizinhas (QGIS ASSOCIATION, 2022). O acúmulo de fluxo é o número de células de maior altitude das quais uma célula recebe escoamento superficial (ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, 2022). A Figura 27 ilustra as variáveis de relevo.

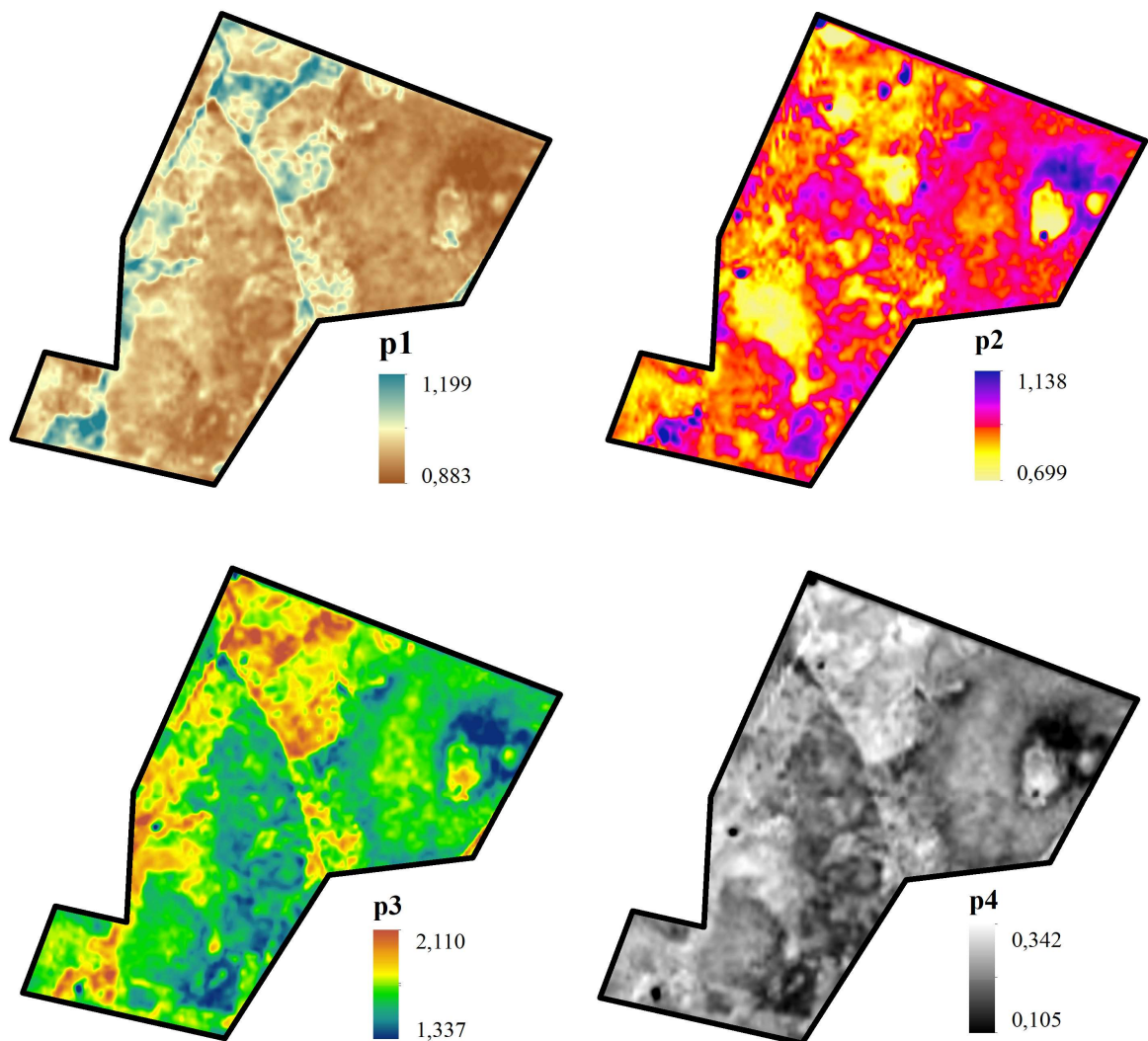
Figura 27 – Distribuição espacial das variáveis do fator relevo.



Fonte: elaborada pelo autor.

O material de origem foi considerado a partir de quatro índices de sensoriamento remoto não nomeados. Foram calculadas a razão das bandas vermelho e verde (bandas 4 e 3 no Landsat 8) para acentuar carbonatos e a razão das bandas vermelho e infravermelho de onda curta 2 (bandas 4 e 7 no Landsat 8) para acentuar íons ferro (COLE et al., 2007), além da razão das bandas infravermelho de onda curta 1 e azul (bandas 6 e 2 no Landsat 8) para contrastar feições geológicas (MWANIKI et al., 2015). Além disso, calculou-se a diferença normalizada das bandas infravermelho de onda curta 1 e verde (bandas 6 e 3 no Landsat 8) para indicar a presença de calcário (LEVI; RASMUSSEN, 2014). Para esse trabalho, as variáveis do material de origem foram nomeadas p1, p2, p3 e p4, respectivamente. As suas distribuições espaciais podem ser vistas na Figura 28.

Figura 28 – Distribuição espacial das variáveis do fator material de origem.



Fonte: elaborada pelo autor.

Também foram utilizadas de maneira direta as bandas de 1 a 11 da imagem de satélite Landsat 8, que tem seu uso validado por trabalhos como o de Chagas et al. (2016) e Taghizadeh-Mehrjardi et al. (2016). As bandas podem ser correlacionadas indiretamente com os fatores solo, clima, organismos e material de origem (BOETTINGER et al., 2008).

Não foi possível encontrar dados que representassem explicitamente o fator tempo. Também não foi possível gerar dados do fator espacial, que aparece apenas implicitamente através do caráter espacial de todas as outras covariáveis.

Diferentes combinações de covariáveis foram testadas. O teste inicial, que utilizou apenas as bandas da imagem do satélite Landsat 8 de forma direta, obteve um R^2 de 0,18 para 1 metro, por exemplo. O mapeamento aerogeofísico do Atlas Aerogeofísico do Estado da Bahia (TELES, 2019) e as outras variáveis do WorldClim (FICK; HIJMANS, 2017) também foram testadas, mas eliminadas do modelo, por piorarem as métricas de desempenho. Isso pode ser explicado por uma falta de correlação com a resistência à penetração do SPT ou por ruído nos dados. O teste que incluiu esses dados resultou em R^2 de 0,36, 0,35 e 0,19 para as três profundidades.

O Quadro 2 resume as covariáveis ambientais utilizadas.

Quadro 2 - Variáveis utilizadas e seus respectivos fatores de formação.

Fator	Variável	Descrição	Referência
S	GI	Landsat 8, (Banda 6 - Banda 5) / (Banda 6 + Banda 5)	Taghizadeh-Mehrjardi et al. (2016)
	CI	Landsat 8, (Banda 6) / (Banda 7)	
	IOI	Landsat 8, (Banda 4 / Banda 2)	Chagas et al. (2010)
C	c1	Média da precipitação anual entre 1970 e 2000.	Dharumarajan et al. (2021)
	c2	Média da isothermalidade anual entre 1970 e 2000.	
O	NDVI	Landsat 8, (Banda 5 - Banda 4) / (Banda 5 + Banda 4)	Boettinger et al. (2008)
R	Altimetria	Modelo digital de elevação com resolução de 30m	Dharumarajan et al. (2017)
	Declividade	Variação da altitude	
	Curvatura	Curvatura do terreno	
	Aspecto	Ângulo em relação ao norte	
	TWI	Índice Topográfico de Umidade	
	TPI	Índice de Posição Topográfica	
	TRI	Índice de Aspreza Topográfica	
	Acúmulo de fluxo	Número de células das quais a célula em análise recebe fluxo superficial.	Møller et al. (2019)
P	p1	Landsat 8, (Banda 4) / (Banda 3),	Cole et al. (2007)
	p2	Landsat 8, (Banda 4) / (Banda 7)	
	p3	Landsat 8, (Banda 6) / (Banda 2)	Mwaniki et al. (2015)
	p4	Landsat 8, (Banda 6 - Banda 3) / (Banda 6 + Banda 3)	Levi e Rasmussen (2014)
S, C, O, P	l8_b1	Banda 1 de uma imagem do satélite Landsat 8	Chagas et al. (2016) Taghizadeh-Mehrjardi et al. (2016)
	l8_b2	Banda 2 de uma imagem do satélite Landsat 8	
	l8_b3	Banda 3 de uma imagem do satélite Landsat 8	
	l8_b4	Banda 4 de uma imagem do satélite Landsat 8	
	l8_b5	Banda 5 de uma imagem do satélite Landsat 8	
	l8_b6	Banda 6 de uma imagem do satélite Landsat 8	
	l8_b7	Banda 7 de uma imagem do satélite Landsat 8	
	l8_b8	Banda 8 de uma imagem do satélite Landsat 8	
	l8_b9	Banda 9 de uma imagem do satélite Landsat 8	
	l8_b10	Banda 10 de uma imagem do satélite Landsat 8	
	l8_b11	Banda 11 de uma imagem do satélite Landsat 8	

Fonte: Elaborada pelo autor.

A Figura 29 apresenta as correlações de Spearman entre os valores de NSPT de cada profundidade e as covariáveis ambientais. Em grandes conjuntos de dados, variáveis de baixa correlação são eliminadas antes da construção do modelo por se considerar que contribuirão pouco enquanto aumentarão o tempo de processamento consideravelmente. No caso deste trabalho, a construção do modelo para cada profundidade demorou alguns minutos, de forma que não foi necessário eliminar nenhuma variável com a justificativa de acelerar o processamento.

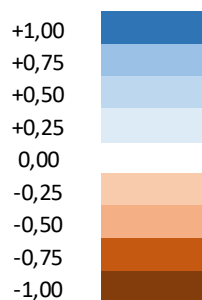
Figura 29 - Correlação de Spearman entre valores de NSPT e covariáveis ambientais.

	NSPT_2m		NSPT_3m							
NSPT_1m	0,88	0,68								
NSPT_2m			0,75							

	gi	ci	ioi	c1	c2	ndvi	altimetria	declividade	curvatura	aspecto
NSPT_1m	0,33	-0,68	0,13	0,49	0,73	-0,50	-0,73	0,09	-0,02	-0,01
NSPT_2m	0,22	-0,65	0,01	0,57	0,78	-0,40	-0,75	0,08	-0,11	-0,01
NSPT_3m	0,17	-0,53	-0,03	0,45	0,69	-0,32	-0,65	0,14	-0,20	0,06

	twi	tpi	tri	acum_fluxo	p1	p2	p3	p4	l8_b1	l8_b2
NSPT_1m	-0,01	0,07	0,15	0,13	0,22	0,08	-0,13	-0,15	0,30	0,30
NSPT_2m	0,05	0,05	0,13	0,22	0,10	0,10	-0,24	-0,26	0,17	0,18
NSPT_3m	0,03	-0,05	0,11	0,21	0,05	0,07	-0,22	-0,23	0,08	0,08

	l8_b3	l8_b4	l8_b5	l8_b6	l8_b7	l8_b8	l8_b9	l8_b10	l8_b11
NSPT_1m	0,11	0,18	-0,22	0,01	0,16	0,19	0,10	0,39	0,38
NSPT_2m	-0,01	0,06	-0,26	-0,13	0,02	0,07	0,02	0,28	0,26
NSPT_3m	-0,03	0,00	-0,30	-0,15	-0,04	0,00	-0,06	0,23	0,20



Fonte: elaborada pelo autor.

O algoritmo *random forest* foi o utilizado para construir o modelo de regressão múltipla entre covariáveis ambientais e SPT, através da linguagem de programação R (R CORE TEAM, 2021) e seus pacotes *randomForest* (LIAW; WIENER, 2002), *ModelMap* (FREEMAN;

FRESCINO, 2009) e *terra* (HIJMANS, 2022). Um modelo foi construído para cada uma das três profundidades. Foram otimizados os parâmetros número de árvores (*ntree*, no *ModelMap*), tamanho do subconjunto aleatório de variáveis preditoras (*mtry*, no *ModelMap*) e tamanho do conjunto *bootstrap* (*samplesize*, no *ModelMap*). Depois disso, os modelos foram aplicados nos mapas das covariáveis para gerar os mapas de N_{SPT} .

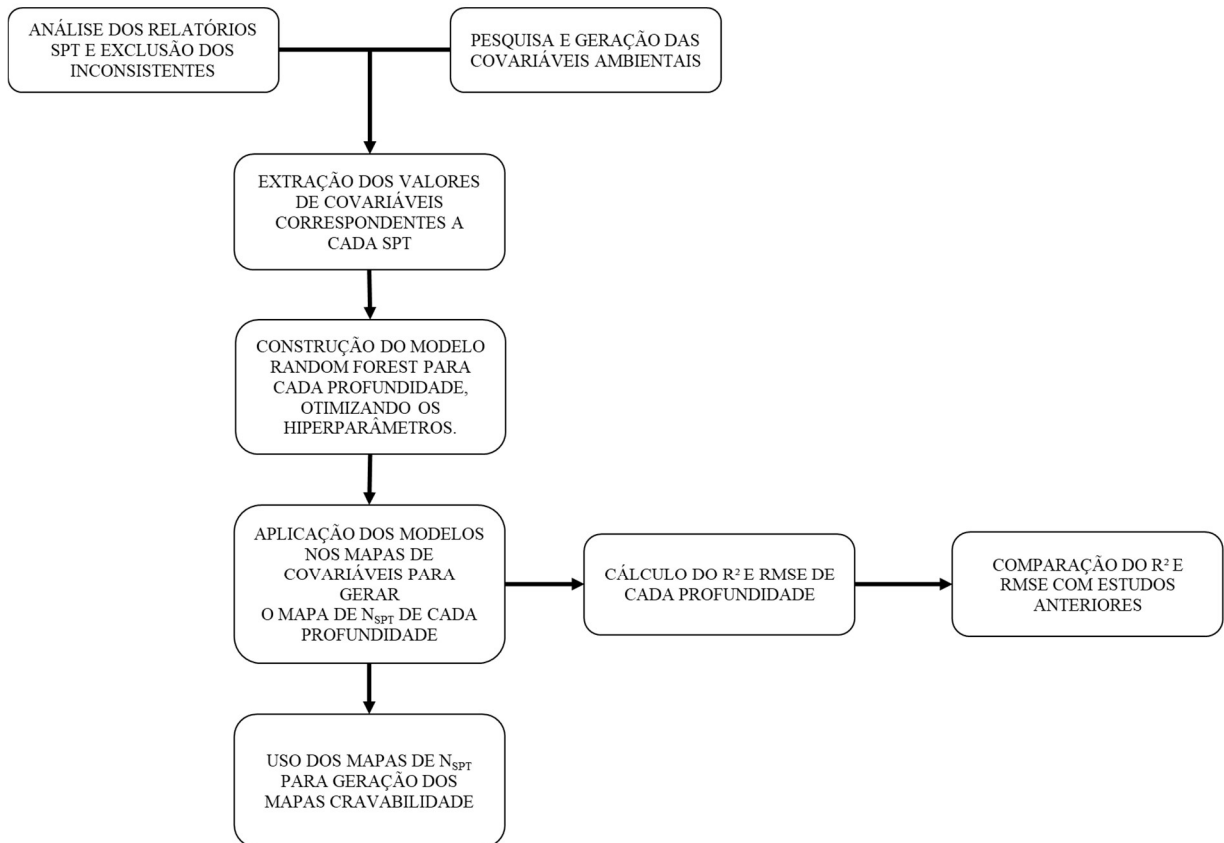
O desempenho dos modelos foi avaliado através dos conjuntos de dados *out-of-bag*, com métricas recomendadas por Khaledian e Miller (2020), visando a padronização e comparação de resultados de mapeamento digital de solos: a raiz do erro quadrático médio (RMSE, do inglês *Root Mean Squared Error*) e o coeficiente de determinação (R^2). O RMSE permite a avaliação do erro na mesma unidade da variável alvo, além de possibilitar a comparação com outros algoritmos de *machine learning* que usem derivadas para minimizar a função erro. Como ponto negativo, é sensível à escala dos dados, dificultando a comparação em áreas ou camadas de solo de diferentes resistências. Para esses casos, pode ser utilizado o R^2 , que leva em consideração a variância dos dados para avaliar o erro.

Os resultados foram então comparados com estudos o mais semelhantes possível. A importância de cada variável foi verificada através do aumento do RMSE que o embaralhamento dos valores da variável causa.

A partir dos resultados, mapas de cravabilidade foram elaborados com base no limite de N_{SPT} entre 25 e 55 para estacas metálicas (CINTRA; AOKI, 2010). Para cada profundidade, são comparados, pixel a pixel, os valores da profundidade em análise e das anteriores, adotando-se o maior. Como os autores não comentam ou citam referências com o motivo da variação do valor limite, uma forma de dividir o mapa seria considerar apenas N_{SPT} maior ou menor que 25. Essa abordagem mais conservadora tem, por um lado, maior chance de evitar o uso inesperado de estacas escavadas. Por outro, pode aumentar muito os custos da obra ao indicar uma quantidade de estacas escavadas maior que o necessário. Assim, é interessante buscar um meio termo que minimize os custos do uso de estacas escavadas, tanto esperados quanto inesperados. Os valores foram então separadores em menores que 25, 25 a 35, 35 a 45, 45 a 55 e maiores que 55, de forma a poder se analisar o comportamento de um possível limite de cravação do terreno.

A Figura 30 resume o método deste trabalho através de um fluxograma.

Figura 30 - Fluxograma do método



Fonte: elaborada pelo autor.

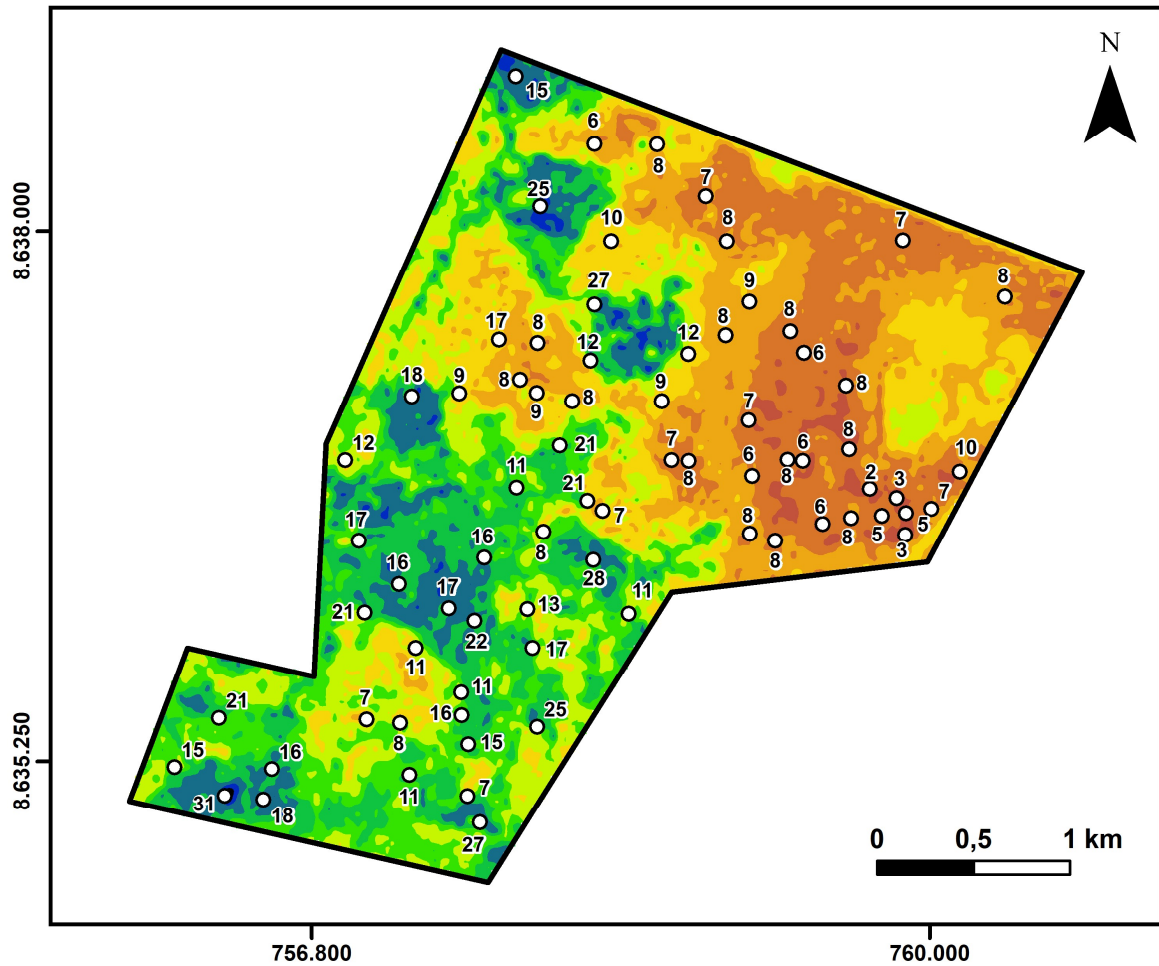
5 RESULTADOS

A otimização dos parâmetros na profundidade de 1 metro resultou nos seguintes valores:

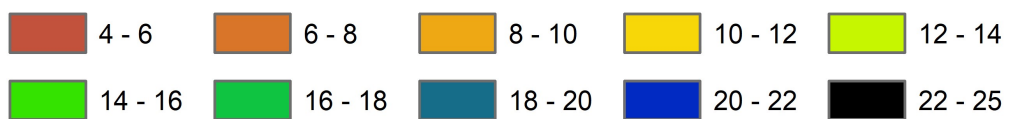
- i. $mtry = 29$;
- ii. $samplesize = 25$;
- iii. $ntrees = 2500$.

O modelo, que pode ser visualizado na Figura 31, teve $R^2 = 0,43$ e $RMSE = 4,93$. Cada ponto apresenta o valor de N_{SPT} do ensaio.

Figura 31 - Mapa de N_{SPT} para a profundidade de 1 metro.



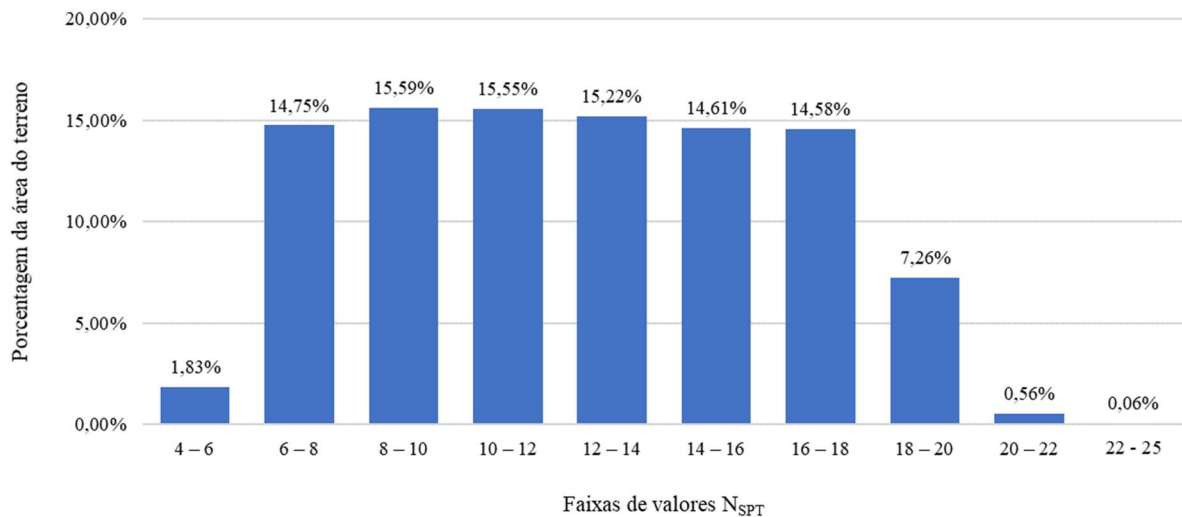
INTERVALOS DE N_{SPT}



Fonte: elaborada pelo autor.

As porcentagens de áreas ocupadas por cada faixa de valor de N_{SPT} estão disponíveis na Figura 32. É interessante notar que mais de 90% da área da UFV tem N_{SPT} entre 6 e 18 nessa profundidade.

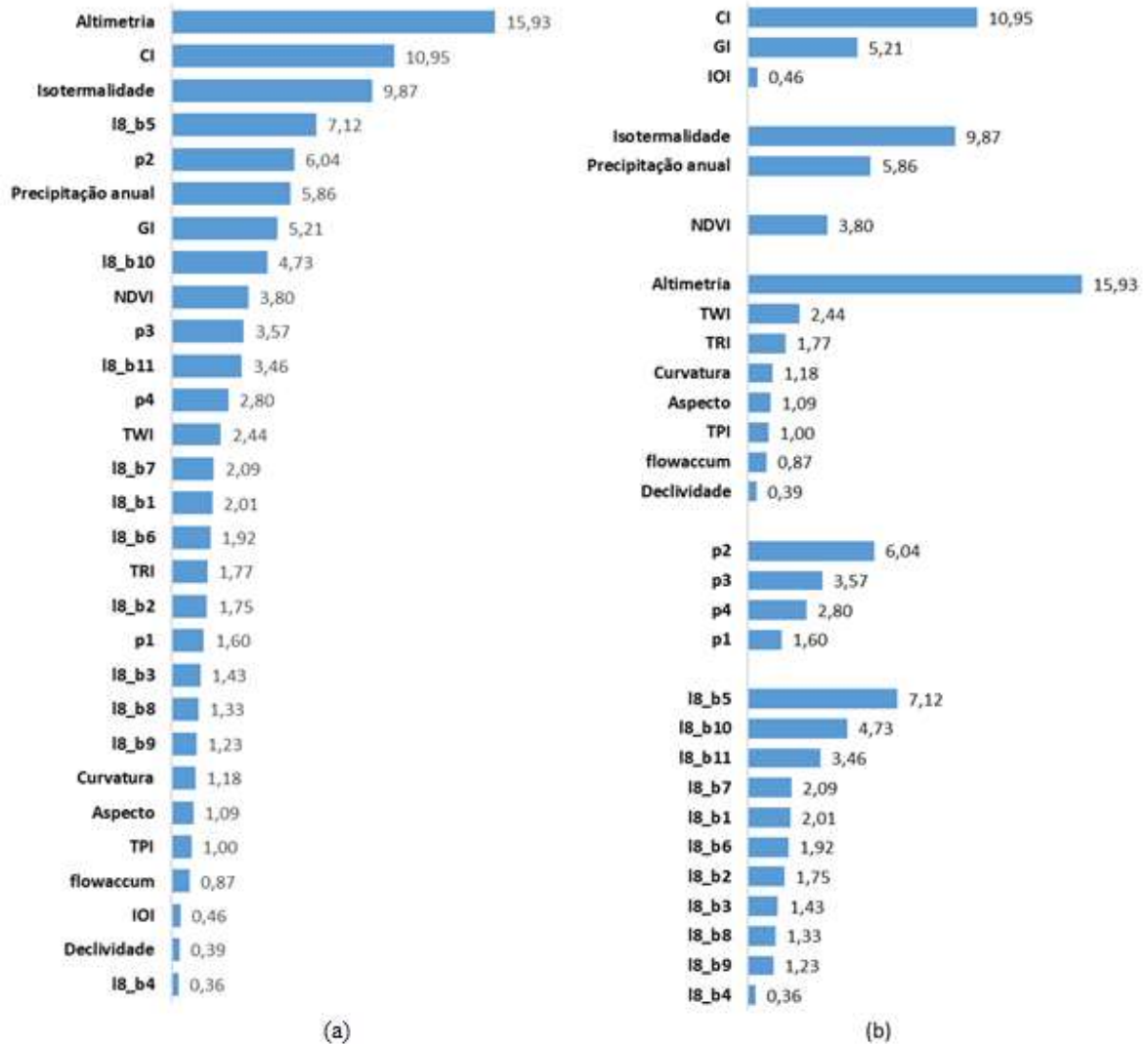
Figura 32 - Porcentagem de área ocupada por faixa de valor de N_{SPT} para a profundidade de 1 metro.



Fonte: elaborada pelo autor.

A Figura 33 apresenta a importância de cada variável no modelo de 1 metro de profundidade. A variável mais importante foi a Altimetria, muito mais que todas as outras variáveis do fator relevo. Isso pode ser explicado pelo fato de o terreno ter uma variação de Altimetria considerável, mas que acontece de forma muito suave. Também é interessante notar, comparando o mapa de Altimetria da Figura 27 com o mapa de N_{SPT} da Figura 31, que os pontos mais altos do terreno parecem coincidir com as menores resistências. O *Clay Index* (CI) teve uma alta importância, assim como no estudo de Chagas et al. (2010), que mapeou classes de solo. A variável p_2 também teve uma importância considerável, da mesma forma que no trabalho de Chagas et al. (2016), que mapeou textura com *random forest* em uma área no norte do semiárido baiano. Além destas, também se destacaram as variáveis isothermalidade e banda 5 da imagem Landsat 8 (infravermelho próximo).

Figura 33 - Porcentagem de aumento do RMSE causado pelo embaralhamento de cada variável para 1 metro de profundidade: (a) todas as variáveis ordenadas; e (b) variáveis agrupadas por fator.



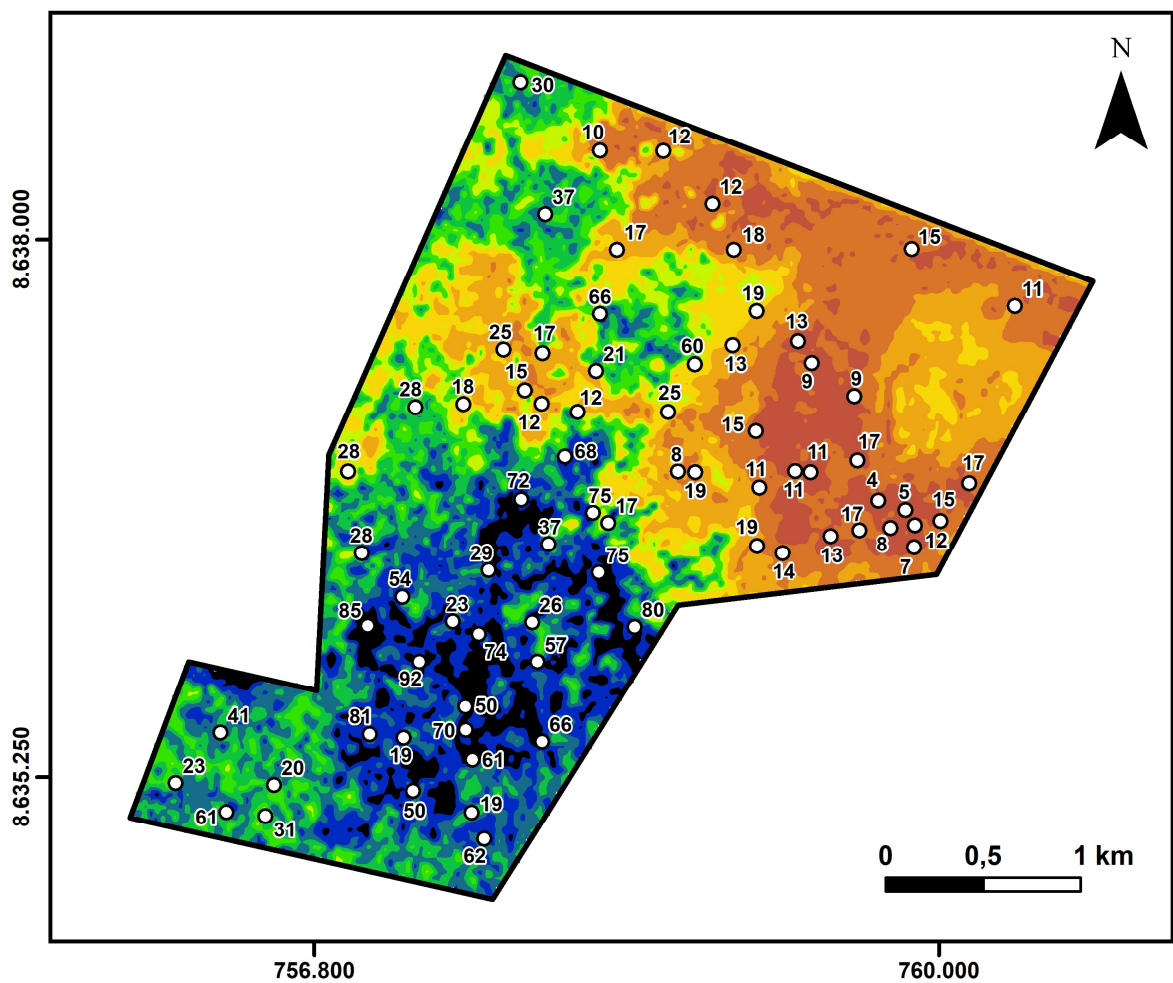
Fonte: elaborada pelo autor.

Para a profundidade de 2 metros, a otimização dos parâmetros resultou nos seguintes valores:

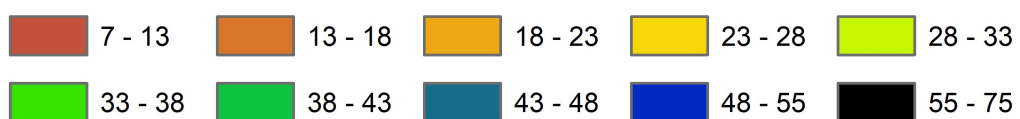
- i. $mtry = 18$;
- ii. $samplesize = 23$;
- iii. $ntrees = 2500$.

O modelo, que pode ser visualizado na Figura 34, teve $R^2 = 0,41$ e $RMSE = 18,44$. Cada ponto apresenta o valor de N_{SPT} do ensaio.

Figura 34 - Mapa de N_{SPT} para a profundidade de 2 metros.



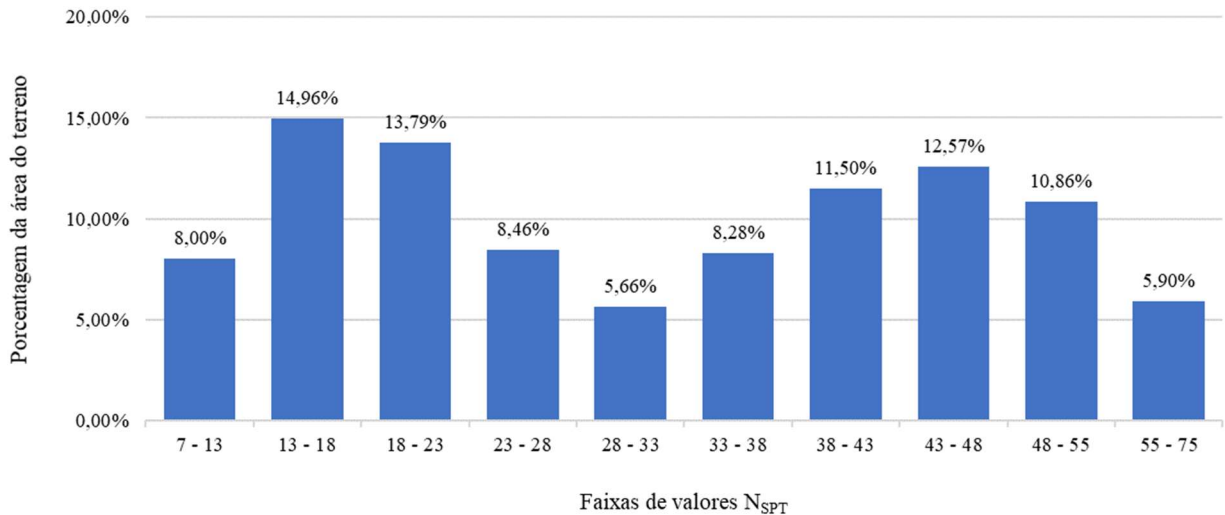
INTERVALOS DE N_{SPT}



Fonte: elaborada pelo autor.

É interessante notar que, por conta do algoritmo *random forest* utilizar médias, o maior valor predito pelo modelo foi 75, apesar de haverem pontos com valores consideravelmente mais altos como 85 e 92. A Figura 35 mostra a área de cada faixa de valor de N_{SPT} .

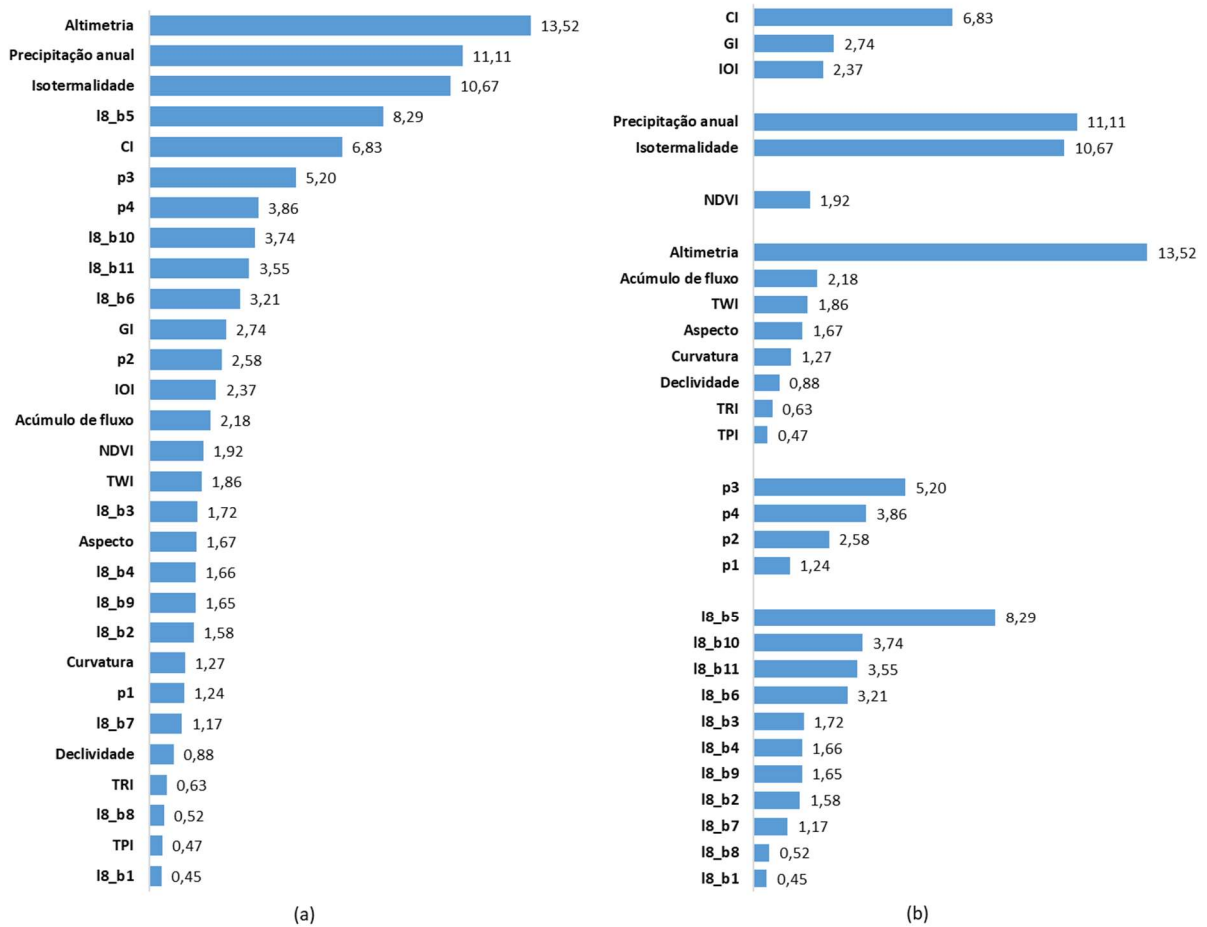
Figura 35 - Porcentagem de área ocupada por faixa de valor de N_{SPT} para a profundidade de 2 metros.



Fonte: elaborada pelo autor.

A Figura 36 apresenta a importância de cada variável no modelo de 2 metros de profundidade. A variável mais importante foi a Altimetria, assim como para 1 metro. A distribuição de valores N_{SPT} parece seguir o mesmo padrão da camada anterior, apenas aumentando em intensidade. A precipitação anual tem um aumento considerável de importância, enquanto o CI cai. A imagem de satélite em infravermelho próximo tem um aumento de importância leve, mas suficiente para ultrapassar o CI. Nas variáveis de material de origem, p3 e p4 passam a ter mais influência no modelo do que p2.

Figura 36 – Porcentagem de aumento do RMSE causado pelo embaralhamento de cada variável para 2 metros de profundidade: (a) todas as variáveis ordenadas; e (b) variáveis agrupadas por fator.



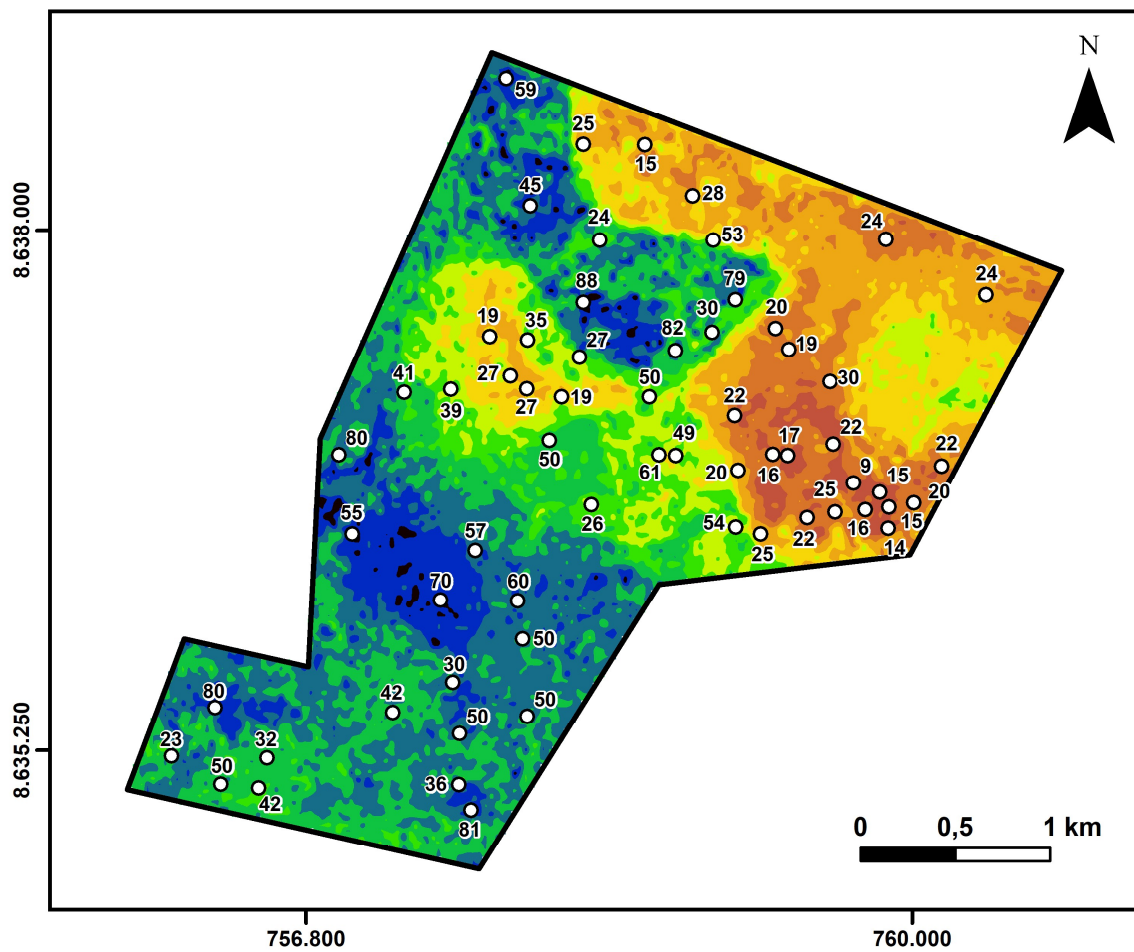
Fonte: elaborada pelo autor.

Em 3 metros de profundidade, os valores otimizados dos parâmetros foram:

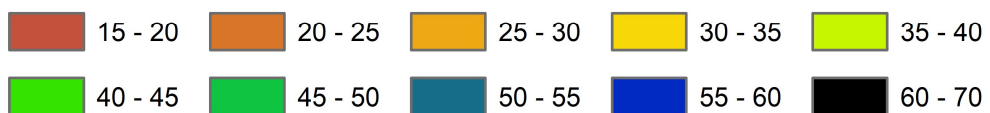
- i. $mtry = 18$;
- ii. $samplesize = 21$;
- iii. $ntrees = 2500$.

O modelo, que pode ser visualizado na Figura 37, teve $R^2 = 0,25$ e $RMSE = 17,66$. Cada ponto apresenta o valor de N_{SPT} do ensaio.

Figura 37 - Mapa de N_{SPT} para a profundidade de 3 metros.

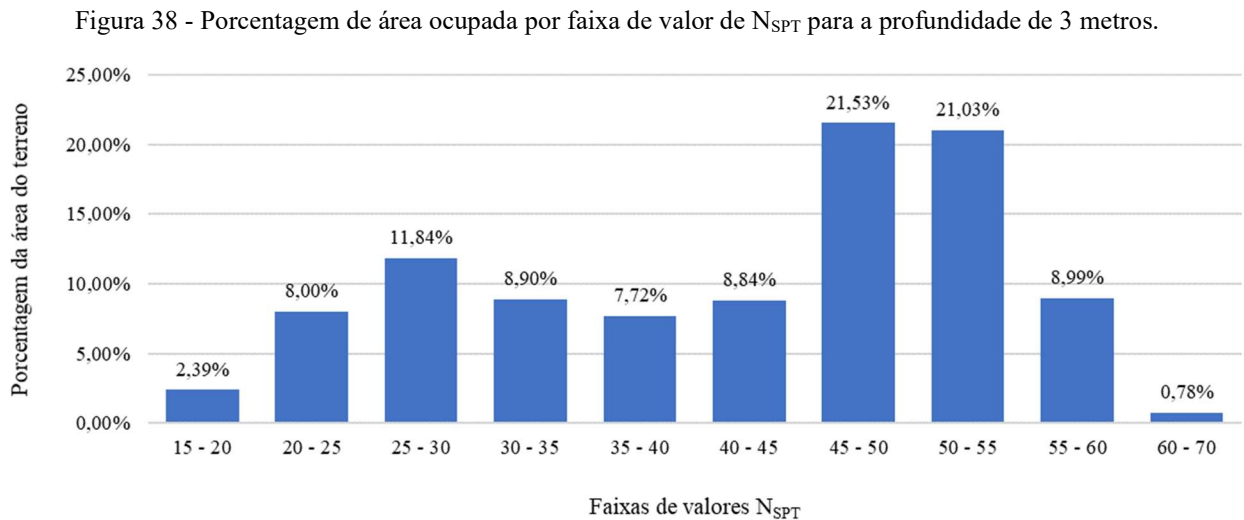


INTERVALOS DE N_{SPT}



Fonte: elaborada pelo autor.

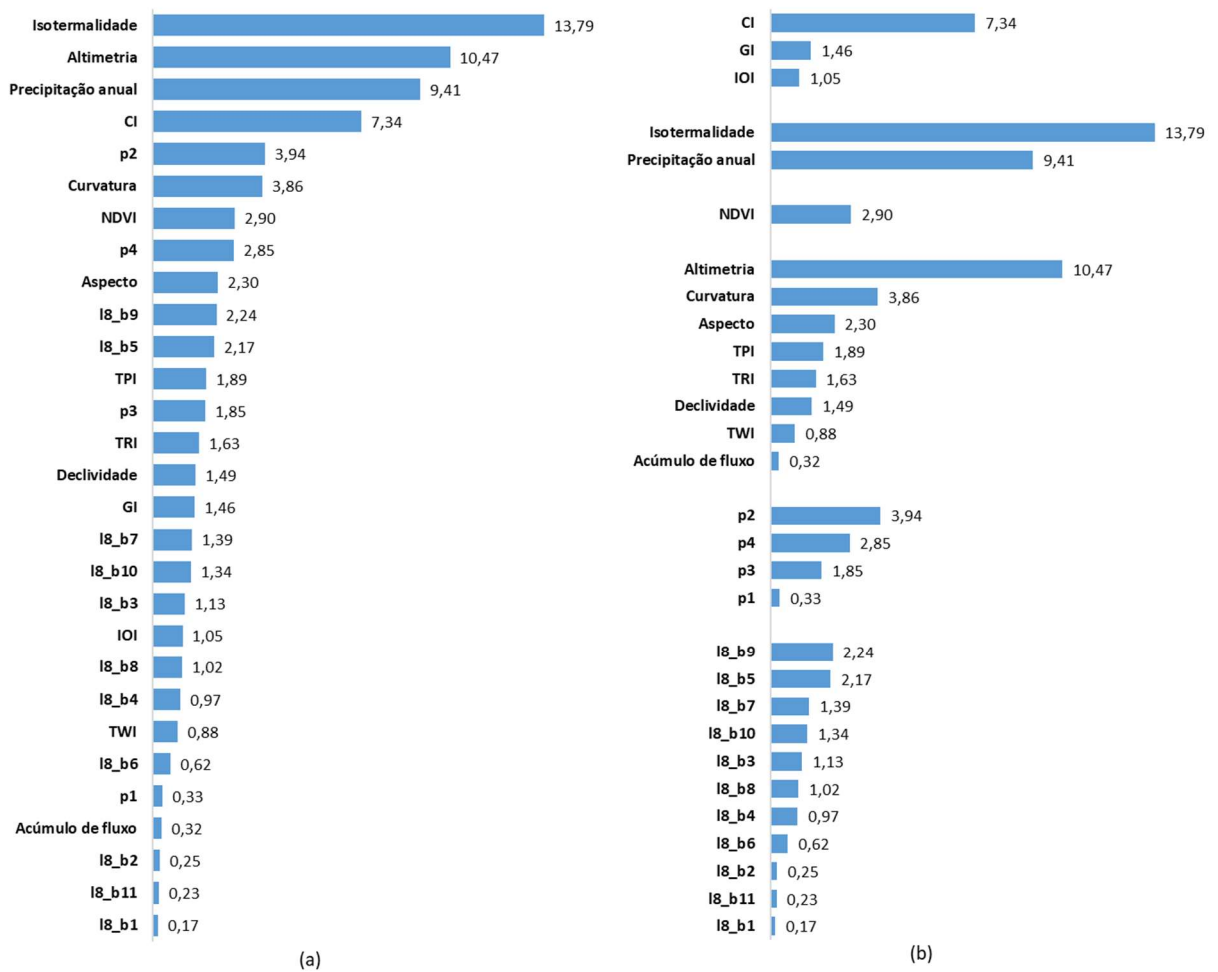
Percebe-se, comparando a Figura 34 e a Figura 37, que a tendência de menor resistência para os pontos de maior altimetria permanece. Através da Figura 38, que mostra a área de cada faixa de valor de N_{SPT} , é possível notar que mais da metade da área da UFV possui valor maiores que 45 aos 3 metros de profundidade.



Fonte: elaborada pelo autor.

A Figura 39 mostra a importância de cada variável no modelo de 3 metros de profundidade. A principal mudança em relação à camada anterior é a Isotermalidade ultrapassar a Altimetria como a mais influente. Além disso, a variável p2 volta à primeira posição dentro do fator material de origem.

Figura 39 – Porcentagem de aumento do RMSE causado pelo embaralhamento de cada variável para 3 metros de profundidade: (a) todas as variáveis ordenadas; e (b) variáveis agrupadas por fator.



Fonte: elaborada pelo autor.

Um resumo dos valores utilizados para os parâmetros e dos resultados das métricas de desempenho pode ser visto na Tabela 2. É importante notar que, embora o RMSE para 2 metros seja mais de 3 vezes maior que o de 1 metro, o R^2 é praticamente igual. Isso acontece porque o RMSE é sensível a escala dos dados, e todos os valores de 1 metro são menores do que 60% dos valores de 2 metros. A terceira camada teve valor de RMSE próximo ao da segunda, já que a faixa de N_{SPT} é semelhante, porém quase metade do R^2 .

Tabela 2 - Resumo dos parâmetros de *random forest* e resultados para as três profundidades.

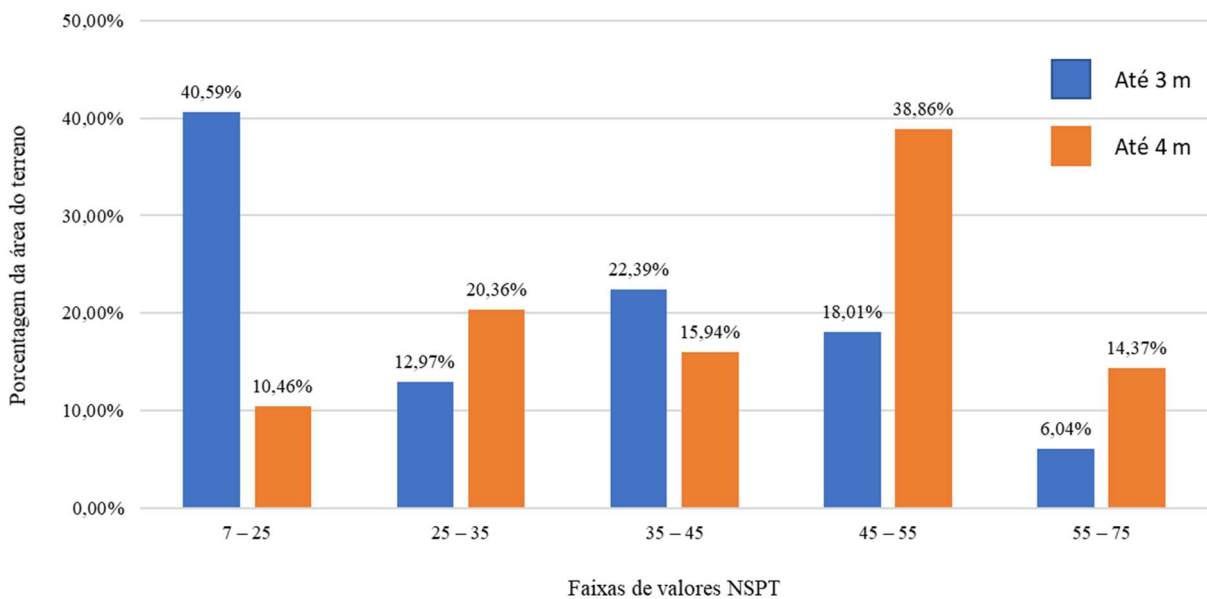
Profundidade	Pontos	n tree	m try	sampsize	R ²	RMSE
1 metro	74	2500	29	25	0,43	4,93
2 metros	74	2500	18	23	0,41	18,44
3 metros	62	2500	18	21	0,25	17,66

Fonte: elaborada pelo autor.

5.1 MAPA DE CRAVABILIDADE

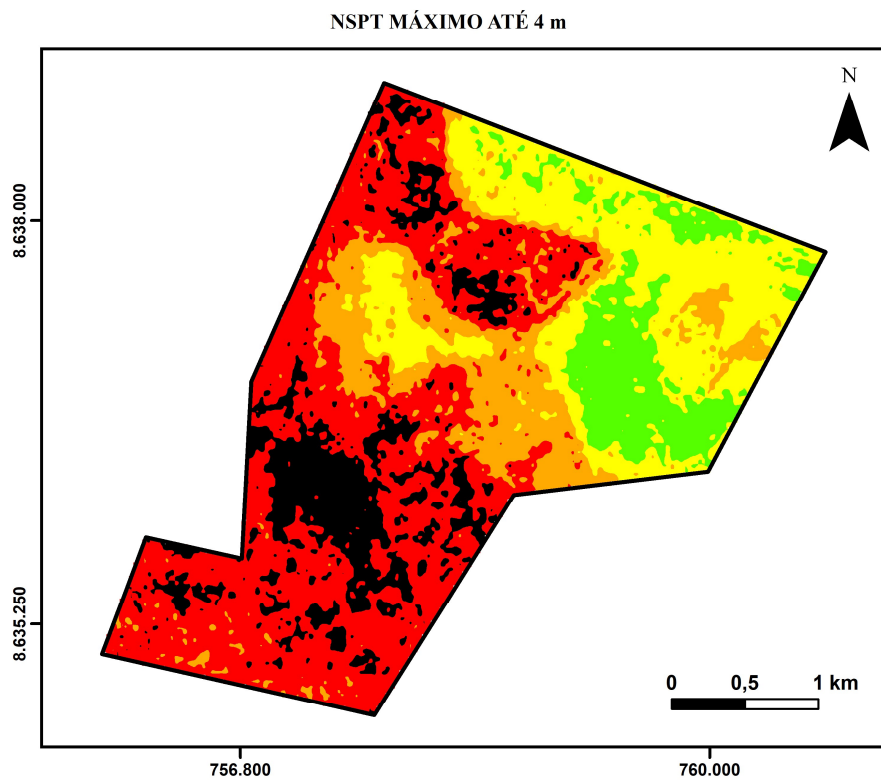
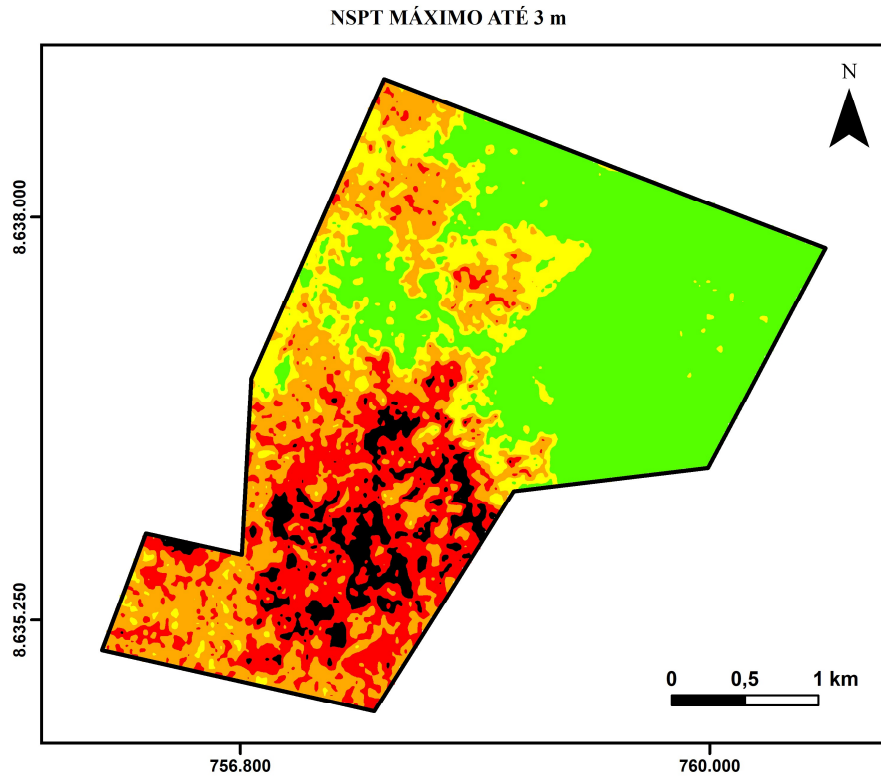
Para a profundidade de até 2 metros (primeiro valor do ensaio penetrométrico), é possível verificar pela Figura 31 que não há nenhum valor maior que 25, de forma que toda a área da UFV é considerada cravável. Para a cravabilidade até 3 metros, foi utilizado o maior entre o primeiro e segundo valor do SPT. Para a cravabilidade até 4 metros, foi utilizado o maior entre os três valores do SPT. Ambos os mapas podem ser vistos na Figura 41. As zonas verdes, que tem valor menor que 25, são craváveis segundo Cintra e Aoki (2010). Já as regiões pretas, com valores maiores que 55, não são passíveis de cravação direta. A Figura 40 apresenta a área de cada faixa de valor da Figura 41.

Figura 40 - Área ocupada por cada faixa de N_{spt} da cravabilidade.



Fonte: elaborada pelo autor.

Figura 41 - Mapas de Nspt máximo para cravação.



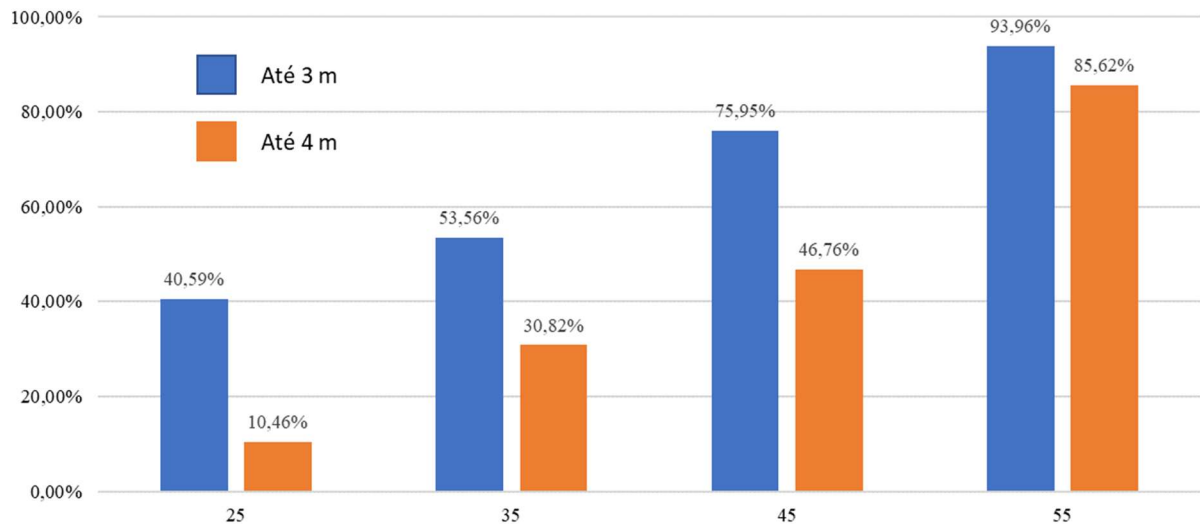
INTERVALOS DE NSPT



Fonte: elaborada pelo autor.

A Figura 42 mostra a área acumulada para quatro possibilidades de N_{SPT} limite para cravação. É interessante notar que uma mudança do valor mais conservador de 25 para 35 quase triplica a área passível de cravação até 4 metros.

Figura 42 - Área acumulada do terreno para cada valor de N_{SPT} limite para cravação.



Fonte: elaborada pelo autor.

Para a definição de um limite para a obra em análise, pode-se testar a cravação de estacas no terreno, selecionando pontos em cada área de N_{SPT} até 55.

5.2 COMPARAÇÃO DOS RESULTADOS

O trabalho de Balasubramani e Dodagoudar (2020) utilizou krigagem ordinária e simples para mapear o N_{SPT} em sondagens que alcançaram até 30 metros, porém não apresentou R^2 , apenas RMSE. Além disso, não separou os dados por profundidades, exibindo apenas um valor global para cada modelo: 3,09 para a krigagem ordinária e 4,65 para a simples. O artigo de Sitharam e Samiu (2007), precursor do citado anteriormente, também utilizou krigagem ordinária e simples sem apresentar R^2 e agrupando todas as profundidades no RMSE, porém apresentou resultados para apenas duas sondagens, e não do modelo como um todo. O primeiro furo, de 8 metros, obteve 2,94 e 4,65 para as krigagens ordinária e simples, respectivamente. O segundo furo, de 9 metros, obteve valores próximos, de 3,10 e 4,54. Apesar de nenhum dos dois trabalhos terem utilizado covariáveis ambientais, ainda podem ser enquadrados no modelo

Scorpan, como um caso em que apenas o fator espacial foi utilizado (MCBRATNEY et al., 2003).

Comparados a esses estudos, os RMSE dos mapeamentos deste trabalho estão na mesma ordem de grandeza. Tal comparação só é válida caso a ordem de grandeza dos resultados dos ensaios seja semelhante, o que é difícil de afirmar.

Pela impossibilidade de comparar valores N_{SPT} sem o viés de escala do RMSE, mapeamentos de textura também foram analisados. Diversos problemas de geotecnia dependem da análise da tensão efetiva no solo, isto é, a tensão nos pontos de contato entre as partículas sólidas (DAS, 2013). No caso da previsão de capacidade de carga de estacas, tanto os métodos teóricos quanto os semiempíricos utilizam coeficientes que variam em função do tamanho de partícula do solo analisado (CINTRA; AOKI, 2010), como é o caso dos métodos semiempíricos Aoki-Velloso, Décourt-Quaresma e o de Teixeira (1996, apud CINTRA; AOKI, 2010). Portanto, é justificada a comparação dos resultados com trabalhos de mapeamento de textura. O Quadro 3 apresenta resultados de mapeamentos de textura na superfície com *random forest*, compilados por Chagas et al. (2016). Destes, destaca-se o próprio trabalho de Chagas et al. (2016), cuja área de estudo também se localiza no semiárido baiano.

Quadro 3 - Coeficientes de determinação de estudos de mapeamento da textura do solo com *random forest*.

Estudo	R ² (%)		
	Areia	Silte	Argila
Ließ et al. (2012), apud Chagas et al. (2016)	0,30	0,26	0,43
Akpa et al. (2014), apud Chagas et al. (2016)	0,48 – 0,49	0,26 – 0,27	0,53 – 0,56
Vaysse e Lagacherie (2015), apud Chagas et al. (2016)	0,33 – 0,55	0,23 – 0,29	0,31 – 0,35
Chagas et al. (2016)	0,44	0,08	0,46

Fonte: Adaptado de Chagas et al. (2016).

É possível verificar que os resultados de R² deste trabalho são condizentes com mapeamentos de textura com *random forest* encontrados na literatura.

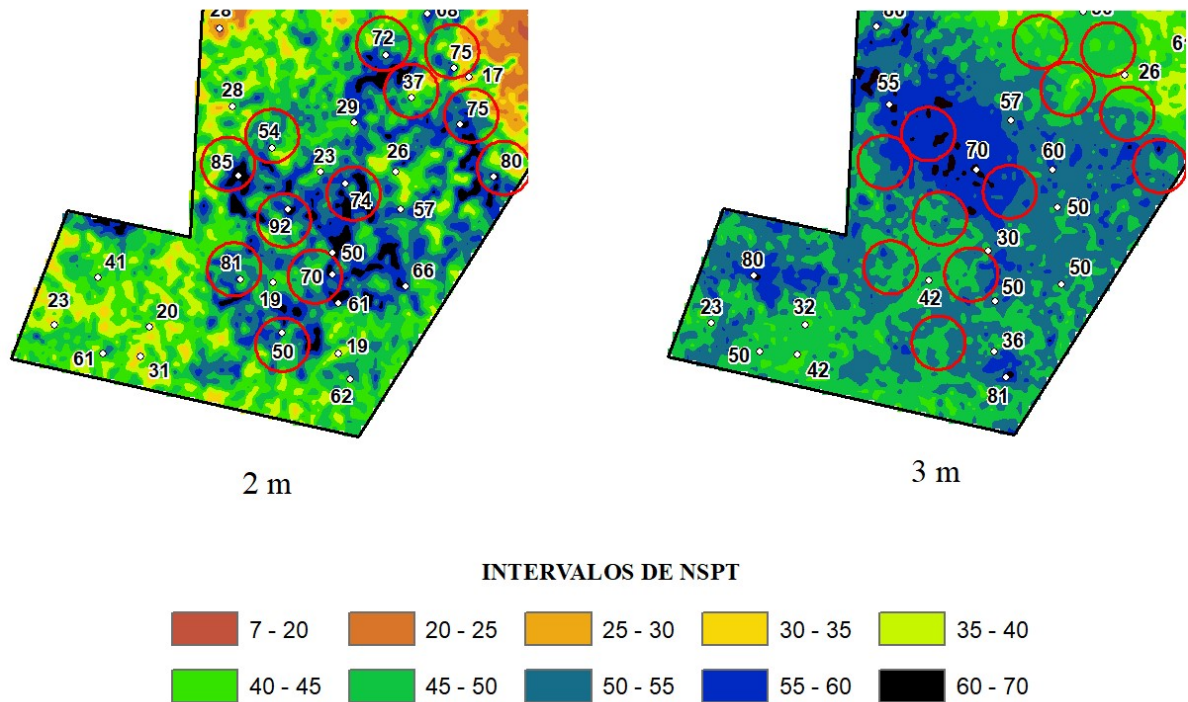
5.3 DISCUSSÃO

O fato de a variável Altimetria ter tido a maior importância em 1 e 2 metros e a segunda maior importância em 3 metros parece corroborar com a afirmação de Schaetzel e Anderson (2005) de que, geralmente, a variabilidade de solos em distâncias pequenas é melhor explicada pelo fator relevo; o que, combinado com a grande importância da Precipitação Anual em 1 e 2 metros e da Isotermalidade em todas as profundidades, vai ao encontro da afirmação de Jenny (1941) de que clima e relevo são, provavelmente, os fatores que agem de forma mais intensa nos solos. Entretanto, de acordo com Wadoux et al. (2020), as correlações encontradas no mapeamento digital de solos não podem ser usadas para inferir diretamente mecanismos de formação de solos, mas apenas usadas para formular hipóteses que devem ser analisadas posteriormente com estudos aprofundados.

O NDVI não teve uma importância muito expressiva, ao contrário do que se esperaria por conta da afirmação de Andrade-Lima (1981) a respeito da interação dos solos e vegetação da caatinga. Entretanto, é possível que outras variáveis representativas da vegetação pudessem ter um impacto maior no modelo.

Na metade sul da UFV, é visível uma queda de resistência entre 2 e 3 metros, aparentemente um viés de sobrevivência (CZEISLER et al., 2021): ensaios com resistência muito alta para 2 metros, mas que não alcançaram os 3 metros – por terem um número de golpes muito alto para os primeiros 15 centímetros, por exemplo – deixam de afetar o modelo na última profundidade. Apenas os ensaios com resistência menor, os “sobreviventes”, permanecem, diminuindo a resistência na região. A Figura 43 destaca essa zona, com a mesma escala de cor para o N_{SPT} . Pontos na metade inferior da imagem tiveram crescimento de 20, 31 e 19 para 32, 42 e 42, respectivamente, enquanto na metade superior da imagem aconteceram aumentos 28, 29 e 23 para 55, 57 e 70, respectivamente, o que causa uma mancha azul escuro na predição (valores entre 55 e 60). Esse efeito também pode afetar os mapas de N_{SPT} limite para cravação.

Figura 43 - Pontos que não chegaram a 3 metros no sul da UFV.



Fonte: elaborada pelo autor.

Também é importante notar que, apesar de contar com pontos com número de golpes como 81, 85 e 90 aos 2 metros e 80 e 81 aos 3 metros, o modelo não prediz valores maiores que 70 para nenhuma das profundidades, já que o *random forest* funciona com médias. No caso dos valores mais baixos, isso é um problema por poder levar a superestimação da resistência, caso os pontos mais fracos do terreno não tenham sido amostrados nas sondagens. Para valores mais baixos, pode levar a erros de predição de cravabilidade: o solo pode ser mais resistente que a predição e não permitir a penetração da estaca. Deste modo, as incertezas dos mapas de cravabilidade produzidos seriam melhor avaliadas se comparadas com a solução adotada na execução das fundações dos painéis. Contudo, o autor não teve acesso a um controle da cravação da obra.

6 CONCLUSÕES

O modelo Scorpan, através do algoritmo *random forest*, foi utilizado para mapear os números de golpes do ensaio SPT para as profundidades de 1, 2 e 3 metros no terreno de uma usina fotovoltaica. As primeiras duas camadas tiveram desempenho equivalente, enquanto a terceira teve um desempenho inferior. As variáveis de maior importância foram a Altimetria, o *Clay Index*, a Precipitação Anual e a Isotermalidade. Os mapas permitem uma boa visualização do comportamento e distribuição dos valores de N_{SPT} no terreno, auxiliando na tomada de decisões. Não foram encontrados estudos que mapeassem o N_{SPT} e utilizassem métricas que permitissem uma comparação adequada, como o coeficiente de determinação (R^2). Portanto, os resultados foram comparados com mapas de textura que utilizaram o mesmo algoritmo, sendo um do semiárido baiano (CHAGAS et al., 2016), assim como a área de estudo desse trabalho. Como as profundidades de 1 e 2 metros tiveram valores muito próximos aos de Chagas et al. (2016), que os autores consideraram satisfatórios, estendeu-se essa avaliação para esse trabalho. Já a avaliação profundidade de 3 metros variou entre boa (quando comparada com o pior resultado de Chagas et al. (2016), o silte) e ruim (quando comparada com o melhor resultado, a argila), o que não teria muito impacto no caso específico desse projeto, em que o maior comprimento de estaca foi de 2,7 metros.

A partir dos mapas de N_{SPT} foram elaborados mapas de cravabilidade com base nos valores limites indicados por Cintra e Aoki (2010). Mesmo na estimativa mais conservadora, o mapa indica que mais de 40% da área da UFV são passíveis de cravação direta até os 3 metros de profundidade. Caso se tratasse de uma nova obra, uma área maior poderia ser considerada através da validação do mapa com testes de cravação, cuja viabilidade econômica deveria ser verificada.

7 SUGESTÕES PARA ESTUDOS FUTUROS

A seguir são elencadas sugestões para estudos futuros, de forma a avançar o tema da pesquisa:

- i. Utilizando os mesmos pontos e área de estudo, repetir o mapeamento com outros algoritmos, como krigagem, co-krigagem, árvore de decisão, redes neurais e etc., de forma a se permitir uma comparação adequada através do RMSE;
- ii. Utilizar os mesmos pontos e área de estudo com algoritmos que usem algum tipo de equação nas folhas - como o *cubist* (KUHN et al., 2014), em vez da média do *random forest*, de forma a se contornar o problema de valores máximos e mínimos;
- iii. Repetir o mapeamento com variáveis de relevo de maior resolução, derivadas de um levantamento topográfico ou aerofotogramétrico, por exemplo;
- iv. Utilizar o Scorpan, com *random forest* e outros algoritmos, para mapear valores de N_{SPT} em outras áreas, dentro e fora do semiárido baiano, para comparação através do R^2 . Isso deve ser feito preferencialmente com uma amostragem que leve em conta as covariáveis ambientais, como através da amostragem pelo hipercubo latino condicionado (MINASNY; MCBRATNEY, 2006);
- v. Testar a cravação de estacas de forma a calibrar o valor limite de N_{SPT} de Cintra e Aoki (2010), com análise econômica de risco considerando as previsões e o custo de cada alternativa;
- vi. Verificar formas de se lidar com o viés de sobrevivência causado por pontos que não alcançam todas as profundidades analisadas, como, por exemplo, repetir o valor do metro anterior.

REFERÊNCIAS

AGÊNCIA EMBRAPA DE INFORMAÇÃO TECNOLÓGICA. **Latossolos**. Disponível em: https://www.agencia.cnptia.embrapa.br/gestor/bioma_caatinga/arvore/CONT000g5twggzh02wx5ok01edq5s5v9cimz.html. Acesso em: 03 fev. 2022.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL). **Resolução Normativa nº 482, de 17 de abril de 2012**. Estabelece as condições gerais para o acesso de microgeração e minigeração distribuída aos sistemas de distribuição de energia elétrica, o sistema de compensação de energia elétrica, e dá outras providências. Brasília, 2012.

ALMEIDA, Patrícia Lopes Brandão. **Análise e Dimensionamento de Fundações de Painéis Fotovoltaicos Inseridos em Parques Solares**. 2014. 216 f. Dissertação (Mestrado) - Curso de Mestrado Integrado em Engenharia Civil, Universidade do Porto, Porto, 2014.

ALONSO, Urbano Rodriguez. **Dimensionamento de fundações profundas**. 3. ed. São Paulo: Blucher, 2019.

ALVARES, Clayton Alcarde et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, Berlin, v. 22, n. 6, p. 711-728, jan. 2014.

ANDRADE-LIMA, Dárdano de. The caatingas dominium. **Revista Brasileira de Botânica**, São Paulo, v. 4, n. 1, p. 149-153, jan. 1981.

ASSOCIAÇÃO BRASILEIRA DE CIMENTO PORTLAND (ABCP). **Manual de Estruturas de Concreto**: Fundação. 2014. Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=16761-setec-orientacoes-sob. Acesso em: 31 ago. 2021.

ASSOCIAÇÃO BRASILEIRA DE ENERGIA SOLAR FOTOVOLTAICA (ABSOLAR). **Panorama da solar fotovoltaica no Brasil e no mundo**. Disponível em: <https://www.absolar.org.br/mercado/infografico/>. Acesso em: 11 ago. 2021.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6484**: Solo - Sondagem de simples reconhecimento com SPT - método de ensaio. Rio de Janeiro, 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6052**: Solos e rochas. Rio de Janeiro, 1995.

BALASUBRAMANI, Divya Priya; DODAGOUDAR, G. R.. Modelling the spatial variability of Standard Penetration Test data for Chennai City using kriging and product-sum model. **Geomechanics And Geoen지니어ing**: An International Journal, Londres, v. 17, n. 1, p. 1-15, jan. 2020.

BLOOMBERG. **O Brasil deve dobrar sua capacidade atual de energia renovável instalada para 316 GW em 2040**. 2018. Disponível em:

<https://www.bloomberg.com.br/blog/o-brasil-deve-dobrar-sua-capacidade-atual-de-energia-renovavel-instalada-para-316-gw-em-2040/>. Acesso em: 08 ago. 2021.

BLOSS, W. et al. Photovoltaic Power Stations: supporting structures. In: WINTER, C.-J.; SIZMANN, Rudolf L.; VANT-HULL, Lorin L. (ed.). **Solar Power Plants: fundamentals, technology, systems, economics**. Heidelberg: Springer, 1991. p. 300-304.

BOETTINGER, J.L. et al. Landsat Spectral Data for Digital Soil Mapping. In: HARTEMINK, Alfred E. et al (ed.). **Digital Soil Mapping with Limited Data**. Amsterdam: Elsevier, 2008. p. 193-202.

BROMS, Bengt B.. Lateral resistance of piles cohesive soils. **Journal Of The Soils Mechanics And Foundations Division: Proceedings of the American Society of Civil Engineers**, Reston, v. 2, n. 3825, p. 27-63, mar. 1964.

COLE, N.J. et al. Pedogenic Understanding Classification Methodology for Mapping Soils, Powder River Basin, Wyoming, USA. In: LAGACHERIE, P. et al (ed.). **Developments in Soil Science**. 31. ed. Amsterdam: Elsevier, 2007. p. 377-388.

CHAGAS, César da Silva et al. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 5, p. 497-507, maio 2010.

CHAGAS, César da Silva et al. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. **Geoderma**, Amsterdam, v. 139, n. 1, p. 232-240, abr. 2016.

CINTRA, José Carlos A.; AOKI, Nelson. **Fundações por estacas: projeto geotécnico**. São Paulo: Oficina de Textos, 2010.

CINTRA, José Carlos A. et al. Filosofias de projeto. In: CINTRA, José Carlos A. et al. **Fundações diretas**. São Paulo: Oficina de Textos, 2011. p. 13-20.

CINTRA, José Carlos A. et al. **Fundações: ensaios estáticos e dinâmicos**. São Paulo: Oficina de Textos, 2013.

CARVALHO, Paulo Ernani Ramalho. Introdução. In: CARVALHO, Paulo Ernani Ramalho et al. **Espécies Arbóreas Brasileiras**. Colombo: Embrapa Informação Tecnológica, 2003. p. 21-56.

CZEISLER, Mark É. *et al.* Uncovering survivorship bias in longitudinal mental health surveys during the COVID-19 pandemic. **Epidemiology And Psychiatric Sciences**, Cambridge, v. 45, n. 30, p. 1-10, maio 2021.

DA SILVA, Ana Margarida Escrivão Cardoso. **Acção do vento em seguidores solares**. 2013. 230 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia Civil, Universidade Nova de Lisboa, Lisboa, 2013.

DAS, Braja M.. **Fundamentos de Engenharia Geotécnica**. 7. ed. São Paulo: Cengage Learning, 2013.

DHARUMARAJAN, S. *et al.* Predicting and Mapping of Soil Hydraulic Properties in Karnataka. **Journal Of The Indian Society Of Remote Sensing**, Nova Delhi, v. 7, n. 49, p. 1623-1631, jun. 2021.

DHARUMARAJAN, S. *et al.* Spatial prediction of major soil properties using Random Forest techniques: a case study in semi-arid tropics of south india. **Geoderma Regional**, Amsterdam, v. 10, n. 1, p. 154-62, jan. 2017.

DINCER, Ibrahim; BICER, Yusuf. Photonic Energy Production. In: DINCER, Ibrahim (ed.). **Comprehensive Energy Systems**. Cambridge: Elsevier, 2018. p. 707-754.

DINCER, Ibrahim; ZAMFIRESCU, Calin. Environmental Dimensions of Energy. In: DINCER, Ibrahim (ed.). **Comprehensive Energy Systems**. Cambridge: Elsevier, 2018. p. 49-100.

DUCHESNE, Louis; OUMET, Rock. Digital mapping of soil texture in ecoforest polygons in Quebec, Canada. **Peerj**, Corte Madera, v. 2, n. 1, p. 3-4, jun. 2021. Disponível em: <https://peerj.com/articles/11685/>. Acesso em: 21 dez. 2021.

ELABOREX. **Pull out test – ensayos plantas fotovoltaicas**. Disponível em: <https://www.elaborex.es/en/actividades-de-calidad-en-la-construccion/pull-out-test-ensayos-plantas-fotovoltaicas/?reload=908856>. Acesso em: 11 set. 2021.

EMBRAPA. **Sistema brasileiro de classificação de solos**. Rio de Janeiro: Embrapa-Spi, 2006.

ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE. **How Flow Accumulation works**. Disponível em: <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-analyst/how-flow-accumulation-works.htm>. Acesso em: 15 fev. 2022.

FARR, Tom G. *et al.* THE SHUTTLE RADAR TOPOGRAPHY MISSION. **Reviews Of Geophysics**, Nova Iorque, v. 2, n. 45, p. 1-33, maio 2017.

FICK, Stephen E.; HIJMANS, Robert J.. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal Of Climatology**, Londres, v. 12, n. 37, p. 4302-4315, mar. 2017.

FREEMAN, Elizabeth; FRESCINO, Tracey. ModelMap: Modeling and Map production using Random Forest and Stochastic Gradient Boosting. **USDA Forest Service: Rocky Mountain Research Station**, Ogden, v. 1, n. 1, p. 1-10, jan. 2009.

GAAVER, Khaled E. et al. Uplift capacity of single piles and pile groups embedded in cohesionless soil. **Alexandria Engineering Journal**, [S.L.], v. 52, n. 3, p. 365-372, set. 2013. Elsevier BV. <http://dx.doi.org/10.1016/j.aej.2013.01.003>.

GUARNIERI, Maurício Vivan. **Usinas solares fotovoltaicas com seguimento em um eixo no Brasil**: aspectos da construção, custos e expectativa de desempenho. 2017. 184 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Civil, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2017.

HASTIE, Trevor et al. Random Forest. In: HASTIE, Trevor et al. **The Elements of Statistical Learning**: data mining, inference and prediction. 2. ed. Nova Iorque: Springer, 2009. p. 587-603.

HATEFFARD, Fatemeh et al. Assessing the performance of decision tree and neural network models in mapping soil properties. **Journal Of Mountain Science**, Chengdu, v. 16, n. 8, p. 1833-1847, jun. 2019.

HIGASHI, Rafael Augusto dos Reis. **Metodologia de Uso e Ocupação dos Solos de Cidades Costeiras Brasileiras Através de SIG com Base no Comportamento Geotécnico e Ambiental**. 2006. 398 f. Tese (Doutorado) - Curso de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2006.

HIGASHI, R. A. R. **Notas de aulas**. UFSC - Universidade Federal de Santa Catarina. Curso de graduação em Engenharia Civil. Florianópolis, 2016.

HIJMANS, Robert J.. **Terra: Spatial Data Analysis.**: r package version 1.5-12. R package version 1.5-12. 2022. Disponível em: <https://CRAN.R-project.org/package=terra>. Acesso em: 16 fev. 2022.

HEUNG, Brandon et al. Digital Soil Mapping. In: KRZIC, Maja et al (ed.). **Digging Into Canadian Soils**: an introduction to soil science. Pinawa: Canadian Society Of Soil Science, 2021. p. 533-577. Disponível em: <https://openpress.usask.ca/soilscience/chapter/digital-soil-mapping/>. Acesso em: 16 dez. 2021.

HUANG, Pan Ming et al (ed.). **Handbook of Soil Sciences**: properties and processes. Boca Raton: Taylor and Francis, 2012.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Brasileiro de 2010**. Rio de Janeiro: IBGE, 2012.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Informações Gerais sobre SD23 - LAd6**. Rio de Janeiro: IBGE, 2022.

JENNY, Hans. Derivation of State Factor Equations of Soils and Ecosystems. **Soil Science Society Of America Journal: Division V—Soil Genesis, Morphology, and Classification**, Fitchburg, v. 25, n. 5, p. 385-388, out. 1961.

JENNY, Hans. **Factors of Soil Formation: A System of Quantitative Pedology**. Nova Iorque: McGraw-Hill, 1941.

KHALEDIAN, Yones; MILLER, Bradley A.. Selecting appropriate machine learning methods for digital soil mapping. **Applied Mathematical Modelling**, Amsterdam, v. 81, n. 1, p. 401-418, maio 2020.

KIENAST-BROWN, Suzann et al. Digital Soil Mapping. In: USDA Soil Science Division Staff. **Soil Survey Manual**. Washington, D.C: Government Printing Office, 2017. p. 295-346.

KUHN, Max; JOHNSON, Kjell. Regression Trees and Rule-Based Models. In: KUHN, Max; JOHNSON, Kjell. **Applied Predictive Modeling**. Nova Iorque: Springer, 2016. p. 173-218.

Kuhn, Max; et al. Cubist: Rule-and Instance-based Regression Modeling, R package version 0.0.18; CRAN: Vienna, Austria, 2014.

LEPSCH, Igo F.. **Formação e Conservação dos Solos**. 2. ed. São Paulo: Oficina de Textos, 2010.

LEVI, Matthew R.; RASMUSSEN, Craig. Covariate selection with iterative principal component analysis for predicting physical soil properties. **Geoderma**, Amsterdam, v. 220, n. 219, p. 46-57, jan. 2014.

LIAW, A.; WIENER, M.. Classification and Regression by randomForest. **R News**, Auckland, v. 3, n. 2, p. 18-22, dez. 2002. Disponível em: https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf. Acesso em: 16 fev. 2022.

MØLLER, Anders Bjørn et al. Prediction of soil drainage classes in Denmark by means of decision tree classification. **Geoderma**, Amsterdam, v. 352, n. 1, p. 314-329, out. 2019.

MCBRATNEY, Alex B. et al. On Digital Soil Mapping. **Geoderma**, Amsterdam, v. 117, n. 1, p. 3-52, nov. 2003.

MINASNY, Budiman; MCBRATNEY, Alex B.. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers & Geosciences**, Amsterdam, v. 32, n. 9, p. 1378-1388, nov. 2006.

MOURA, Magna Soelma Beserra de et al. Clima e água de chuva no Semi-Árido. In: BRITO, Luiza Teixeira de Lima et al (ed.). **Potencialidades da Água da Chuva no Semi-Árido Brasileiro**. Petrolina: Embrapa Semi-Árido, 2007. p. 37-62.

MWANIKI, M. W. et al. A comparison of Landsat 8 (OLI) and Landsat 7 (ETM+) in mapping geology and visualising lineaments: a case study of central region Kenya. **The International Archives Of The Photogrammetry, Remote Sensing And Spatial Information Sciences**, Berlim, v. 3, n. 7, p. 897-903, maio 2015.

NESOIL. **New England Soil Profiles: raypol (like) series**. Disponível em: <http://nesoil.com/images/raypol.htm>. Acesso em: 25 set. 2021.

PEREIRA, Luís Flávio et al. Modelling and mapping superficial soil texture through machine learning and limited legacy data. In: PEDOMETRICS BRAZIL, 2., 2021, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: Embrapa, 2021. p. 5-8.

PEREIRA, Marcos Gervasio et al. Formação e caracterização de solos. In: TULLIO, Leonardo et al (org.). **Formação, Classificação e Cartografia dos Solos**. Ponta Grossa: Atena, 2019. p. 1-21.

POWERWAY RENEWABLE ENERGY. **Powerway H-profile piles pull-out test**. Disponível em: https://www.youtube.com/watch?v=9KNiJ6yAI2k&ab_channel=Powerway. Acesso em: 11 set. 2021.

R CORE TEAM. **R: A language and environment for statistical computing**. 2021. Disponível em: <https://www.R-project.org/>. Acesso em: 16 fev. 2022.

PROBST, Philipp et al. Hyperparameters and tuning strategies for random forest. **Wires Data Mining Knowledge Discovery**, Hoboken, v. 9, n. 3, p. 1-15, jan. 2019.

QGIS ASSOCIATION. **QGIS 3.16: geographic information system api documentation**. Geographic Information System API Documentation. Disponível em: https://docs.qgis.org/3.16/en/docs/server_manual/index.html. Acesso em: 15 fev. 2022.

REESE, Lymon C.; VAN IMPE, William. **Single Piles and Pile Groups Under Lateral Loading**. Boca Raton: Taylor And Francis, 2011.

REIS, Pedro. **Como funcionam as células solares fotovoltaicas**. 2015. Disponível em: <https://www.portal-energia.com/como-funcionam-celulas-solares-componentes-operacoes/>. Acesso em: 30 ago. 2021.

RODRIGUES, Luiza Faggion. **Análise do Planejamento de Atividades Complexas na Construção de Usinas Fotovoltaicas**. 2021. 115 f. TCC (Graduação) - Curso de Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2021.

ROSOLEM, Gabriel Phelipe Nascimento. **Análise da Vulnerabilidade a Deslizamentos da Ocupação Urbana na Bacia do Rio Comprido – Joinville/SC**. 2018. 170 f. Dissertação (Mestrado) - Curso de Pós-Graduação em Engenharia de Transportes e Gestão Territorial, Universidade Federal de Santa Catarina, Florianópolis, 2018.

SØRENSEN, R. et al. On the calculation of the topographic wetness index:: evaluation of different methods based on field observations. **Hydrology And Earth System Sciences**, Göttingen, v. 1, n. 10, p. 101-112, fev. 2006.

SCHAETZL, Randall J.; ANDERSON, Sharon. **Soils: Genesis and Geomorphology**. Nova Iorque: Cambridge University Press, 2005.

SEKULIC, Aleksandar et al. Random Forest Spatial Interpolation. **Remote Sensing**, Basiléia, v. 12, n. 10, p. 200-229, maio 2020.

SHANKER, K. et al. Uplift Capacity of Pile Groups Embedded in Sands: predictions and performance. **Soils And Foundations**, [S.L.], v. 46, n. 5, p. 605-612, out. 2006. Elsevier BV. <http://dx.doi.org/10.3208/sandf.46.605>.

SITHARAM, T. G.; SAMUI, P.. Geostatistical modelling of spatial and depth variability of SPT data for Bangalore. **Geomechanics And Geoengineering: An International Journal**, Londres, v. 4, n. 2, p. 307-316, dez. 2007.

SOARES, Gabryel Gomercindo. **Análise e Dimensionamento Geotécnico de Fundações em Estacas Metálicas Curtas para Trackers em Usinas Fotovoltaicas no Brasil**. 2021. 160 f. TCC (Graduação) - Curso de Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2021.

STARMER, Joshua. **Machine Learning Fundamentals: Bias and Variance**. Disponível em: <https://www.youtube.com/watch?v=EuBBz3bI-aA>. Acesso em: 23 dez. 2021.

STI NORLAND BRASIL. **Tipos de cravação**: adaptação ao terreno. adaptação ao terreno. Disponível em: https://www.linkedin.com/posts/sti-norland-brasil_tipos-de-crava%C3%A7%C3%A3o-sti-norland-activity-6788571778093514752-gi-w/. Acesso em: 03 set. 2021.

TAGHIZADEH-MEHRJARDI, R. et al. Predicting and mapping of soil particle-size fractions with adaptive neuro-fuzzy inference and ant colony optimization in central Iran. **European Journal Of Soil Science**, London, v. 1, n. 1, p. 1-19, abr. 2016.

TELES, Marcos Rogério Lima (org.). **Atlas aerogeofísico do estado da Bahia**. Salvador: CPRM, 2019.

THIESEN, Stephanie. **Aplicação de Ferramenta SIG para Mapeamento Geotécnico e Cartas de Aptidão para Fundação a partir de Ensaio SPT: Um Estudo de Caso em Blumenau/SC**. 2016. 207 f. Dissertação (Mestrado) - Curso de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2016.

VELLOSO, Dirceu de Alencar; LOPES, Francisco de Rezende. **Fundações: fundações profundas**. São Paulo: Oficina de Textos, 2010.

VILLALVA, Marcelo Gradella. **Energia Solar Fotovoltaica: conceitos e aplicações - sistemas isolados e conectados à rede**. 2. ed. São Paulo: Editora Érica, 2012.

WADOUX, Alexandre M. J.-C. et al. A note on knowledge discovery and machine learning in digital soil mapping. **European Journal Of Soil Science**, Londres, v. 71, n. 1, p. 133-136, jan. 2020.

WANG, Zhaoli et al. Flood hazard risk assessment model based on random forest. **Journal Of Hydrology**, Amsterdam, v. 527, n. 1, p. 1130-1141, ago. 2015.

WANG, Zhihua. Energy and Air Pollution. In: DINCER, Ibrahim (ed.). **Comprehensive Energy Systems**. Cambridge: Elsevier, 2018. p. 909-949.

YEON, Young-Kwang et al. Landslide susceptibility mapping in Injae, Korea, using a decision tree. **Engineering Geology**, Amsterdam, v. 116, n. 3, p. 274-283, set. 2010.