

MARIA RAFAELA DA SILVA NAVALHO

Licenciada em Engenharia Agronómica

MODELAÇÃO DAS PROPRIEDADES DE SOLOS AGRÍCOLAS TENDO POR BASE A UTILIZAÇÃO DO SENSOR NIR

MODELAÇÃO DAS PROPRIEDADES DE SOLOS AGRÍCOLAS TENDO POR BASE A UTILIZAÇÃO DO SENSOR NIR

MARIA RAFAELA DA SILVA NAVALHO

Licenciada em Engenharia Agronómica

Orientadora: Cristina Maria Nobre Sobral de Vilhena da Cruz Houghton,
Professora da Faculdade de Ciências da Universidade de Lisboa

Coorientadores: José Rafael Marques da Silva,
Professor Catedrático, Universidade de Évora
Maria Manuela Simões Ribeiro,
Professora Associada, Universidade NOVA de Lisboa

Júri:

Presidente: Maria Manuela Abreu da Silva,
Professor(a) Associada com Agregação, FCT-NOVA
Arguentes: Adélia Maria Oliveira Sousa,
Professor Auxiliar, Universidade de Évora
Orientador: Cristina Maria Nobre Sobral de Vilhena da Cruz
Houghton,
Professora da Faculdade de Ciências da Universidade de Lisboa

Modelação Das Propriedades De Solos Agrícolas Tendo Por Base A Utilização Do Sensor NIR

Copyright © Maria Rafaela da Silva Navalho, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

No presente trabalho não foram utilizados quaisquer conteúdos gerados por tecnologias de inteligência artificial.

AGRADECIMENTOS

Quero agradecer a todos os que fizeram parte desta caminhada.

Agradeço aos meus orientadores. À Professora Cristina da Cruz Houghton, da Faculdade de Ciências da Universidade de Lisboa, por me ter acolhido no seu grupo de trabalho, pelo conhecimento transmitido e pela simpatia e positivismo com que sempre abordou as nossas reuniões. Agradeço de igual forma ao Professor José Rafael Silva da Universidade de Évora, e à Professora Manuela Simões Ribeiro da Universidade NOVA de Lisboa pelos ensinamentos ao longo do mestrado e pela valiosa orientação nesta fase final.

A todas as pessoas do laboratório pela disponibilidade e ajuda sempre que precisei. Em especial à Inês Ferreira, pelo tempo dedicado, por me ouvir e motivar, e por estar sempre pronta para me ajudar.

Agradeço à minha família. Ao meu pai pelos conselhos, perspetiva e ânimo, à minha mãe por me incentivar sempre a ir mais além do que eu penso ser capaz, aos meus irmãos pelo carinho e preocupação incondicionais.

E a todos que não consigo mencionar aqui, mas que me apoiam na vida dando-me força e incentivo para lutar sempre pelos meus sonhos.

A todos um profundo e sincero obrigado.

Resumo

A análise de solos é tradicionalmente um processo demorado e caro. Neste contexto, a presente dissertação avalia a possibilidade de utilizar a espectroscopia de infravermelho próximo (NIR) como alternativa rápida, económica e não destrutiva às análises laboratoriais tradicionais para a avaliação da fertilidade do solo.

Foram realizadas amostragens de solo, cujos locais foram definidos com recurso ao índice de vegetação por diferença normalizada (NDVI). Foram realizadas análises laboratoriais à fertilidade e, em particular, às várias frações de carbono no solo. Em simultâneo, registaram-se espectros NIR para cada amostra. Dois métodos de modelação - *Random Forest* (RF) e regressão por mínimos quadrados parciais (PLSR) - foram aplicados para prever propriedades químicas a partir dos espectros e recorreu-se à interpolação espacial para gerar mapas de fertilidade.

O modelo RF superou consistentemente o PLSR na maioria dos parâmetros. Alcançou uma previsão elevada ($R^2 > 0,85$) para o magnésio, ferro, níquel, cálcio, sódio e matéria orgânica estimada por perda de ignição. Observou-se um bom desempenho ($0,70 < R^2 < 0,80$) para o cobalto, potássio, boro e manganês. No entanto, os modelos foram ineficazes na estimativa do fósforo, enxofre, condutividade elétrica, zinco, pH, azoto total, carbono total e para outras frações de carbono. Os mapas produzidos forneceram informação sobre a distribuição espacial para os nutrientes com melhor previsão, constituindo uma ferramenta valiosa para aplicação diferenciada da fertilização.

Palavras-chave: Espectroscopia NIR, Modelação espectral, Fertilidade do solo, Carbono do solo.

Abstract

Soil analysis is traditionally a time-consuming and costly process. In this context, the present dissertation evaluates whether near-infrared (NIR) spectroscopy can be used as a fast, economical and non-destructive alternative to traditional laboratory analyses in assessing soil fertility.

To test this, soil samples were collected at sites delineated using the Normalised Difference Vegetation Index (NDVI). Laboratory analyses focused on fertility parameters and several fractions of soil carbon were performed. In parallel, NIR spectra were recorded for each sample. Two modelling approaches—Random Forest (RF) and partial least squares regression (PLSR)—were applied to predict chemical properties from the spectra and spatial interpolation was used to generate fertility maps.

The RF model consistently outperformed PLSR across most parameters. It achieved high predictive accuracy ($R^2 > 0.85$) for magnesium, iron, nickel, calcium, sodium and organic matter estimated by loss-on-ignition. Good performance ($0.70 < R^2 < 0.80$) was observed for cobalt, potassium, boron and manganese. However, the models were ineffective at estimating phosphorus, sulphur, electrical conductivity, zinc, pH, total nitrogen, total carbon and other carbon fractions. The generated maps provided information on the spatial distribution of the best-predicted nutrients, offering a valuable tool for site-specific fertilisation.

Keywords: NIR spectroscopy, Spectral modelling, Soil fertility, Soil carbon.

Índice

1	INTRODUÇÃO.....	1
1.1	Contextualização.....	2
1.2	Fertilidade do Solo.....	3
1.3	Carbono no Solo.....	4
1.4	A Reflectometria do Infravermelho Próximo.....	8
1.4.1	A Reflectometria NIR para Avaliar a Fertilidade do Solo.....	10
1.4.2	A reflectometria NIR para determinação de Carbono.....	10
1.5	Objetivos.....	12
2	METODOLOGIA.....	13
2.1	Descrição da Área de Estudo.....	15
2.2	Definição dos Locais de Amostragem.....	16
2.3	Amostragem.....	18
2.4	Análises Químicas do Solo.....	19
2.4.1	Fertilidade do Solo.....	19
2.4.2	Teores de carbono no solo.....	20
2.5	Análise dos Solos com o Reflectómetro NIR.....	23
2.6	Modelação.....	24
2.6.1	Organização e Pré-Processamento dos Dados.....	25
2.6.2	Criação dos Modelos, Validação e Previsão.....	27

2.7	Mapas de Valores Previstos.....	29
3	RESULTADOS E DISCUSSÃO	31
3.1	Seleção dos Locais de Amostragem.....	31
3.2	Criação das Bases de Dados para Modelação	34
3.2.1	Resultados Estatísticos	34
3.3	Mapas de Previsão	38
4	CONCLUSÃO.....	40
5	BIBLIOGRAFIA.....	42
6	ANEXOS	51
6.1	Anexo I - Gráficos Temporais de NDVI	52
6.2	Anexo II - Coordenadas dos Locais de Amostragem	53
6.3	Anexo III – Pré-processamento das Variáveis Químicas.....	55
6.4	Anexo IV - Pré-processamento dos Espectros	56
6.5	Anexo V – Modelação com <i>Random Forest</i>	57
6.6	Anexo VI – Modelação com PLSR	62

Índice de Figuras

Figura 1 - Representação esquemática dos fluxos de carbono no solo.	5
Figura 2 - Esquema geral da metodologia aplicada.	14
Figura 3 - Localização geográfica à escala nacional do município de Benavente, localização da Quinta da Foz dentro dos limites do município e vista detalhada do limite da parcela de estudo.	16
Figura 4 - Representação da divisão da herdade em arrozal, tomatal e montado.	17
Figura 5 - Instrumentos utilizados na recolha de amostras.	18
Figura 6 - Diagrama sistemático das frações de carbono do solo e análises utilizadas para a sua determinação.	20
Figura 7 - Materiais utilizados na análise LOI.	21
Figura 8 - Processo de filtração das amostras com recurso à bomba de vácuo.	22
Figura 9 - Sensor NIR em utilização no laboratório.	23
Figura 10 - Esquema sistemático do processo de modelação aplicado.	24
Figura 11 - Distribuição e tipos de solo na Quinta da Foz. Fonte: Agrogos.	32
Figura 12 - Determinação dos pontos de amostragem de acordo com as classes de NDVI. a) Divisão da zona do Tomatal em três subzonas; b) Localização das amostras no Tomatal c) Localização das amostras no Montado d) Localização das amostras no Arrozal.	33
Figura 13 - Valores Reais vs. Previstos para as propriedades químicas da validação cruzada.	37
Figura 14 - Mapas com valores previstos de Na, Ni, MO_LOI, Fe e Mg para as zonas do tomatal, arrozal e montado.	39
Figura 15 - Variação do NDVI para o ano de 2023 para: a) Zona do Montado; b) Subzona 1 do Tomatal; c) Subzona 2 do Tomatal; d) Subzona 3 do Tomatal; e) Zona do Arrozal.	52

Índice de Tabelas

Tabela 1 - Balanço de carbono no solo. Fonte: soil.copernicus.org.....	6
Tabela 2 - Descrição das variáveis, respectivas unidades e metodologias de análise utilizadas.	26
Tabela 3 - Resultados estatísticos para o modelo <i>Random Forest</i> e PLSR. Legenda: RMSE – Raiz do erro quadrático médio; R^2 – Coeficiente de determinação; RPD – Razão de desempenho em relação ao desvio padrão; Cal – Resultante da calibração; CV – Resultante da validação cruzada; NC – Número de componentes;.....	36
Tabela 4 – Coordenadas e identificação das amostras.	53

Introdução

Esta dissertação intitulada “Modelação das propriedades de solos agrícolas tendo por base a utilização do sensor NIR” foi realizada no âmbito do Mestrado em Tecnologias em Agricultura de Precisão, da NOVA FCT em associação com a Universidade de Évora. Os objetivos do trabalho a desenvolver consistiram em comprovar a possibilidade de utilizar um sensor NIR para detetar propriedades do solo, criar um modelo e prever propriedades de locais não amostrados utilizando apenas a informação espectral do solo. Neste trabalho, foram realizadas análises químicas laboratoriais para quantificar a fertilidade do solo, duas técnicas para quantificar o carbono no solo: o método por perda de ignição e fumigação de amostras, em conjunto com o analisador TOC; e a espectrometria no infravermelho através do sensor NIR.

As amostras de solo foram recolhidas na Quinta da Foz (Benavente, Santarém) e a sua análise foi realizada no Centro de Ecologia Evolução e Alterações Ambientais (cE3c) da Faculdade de Ciências da Universidade de Lisboa.

O documento presente contém, no primeiro capítulo, a introdução ao tema abordado, o contexto, informações teóricas necessárias à compreensão do trabalho desenvolvido, associadas à fertilidade, ao caso específico do carbono, à espectrometria de infravermelho próximo e aos objetivos gerais a serem desenvolvidos na dissertação.

No segundo capítulo, são descritos os métodos e procedimentos realizados no processo de análise das amostras e obtenção dos resultados, que são apresentados e discutidos no capítulo seguinte.

1.1 Contextualização

A crescente pressão para assegurar um crescimento sustentável da produção agrícola, num contexto em que a procura global por alimentos é crescente, exige soluções inovadoras que garantam produtividade e a preservação do ambiente (Jin e Han, 2024). Até 2050, será necessário aumentar a produção mundial de alimentos em pelo menos 70%, o que representa uma pressão acrescida à gestão dos recursos já limitados e à pegada ambiental da atividade agrícola (Cravero e Sepúlveda, 2021).

Neste cenário, a agricultura de precisão (AP) surge como uma estratégia crucial para o futuro do setor. A *International Society of Precision Agriculture* (ISPA) define a AP como a integração da variabilidade temporal e espacial no processo de tomada de decisão, de forma a aumentar a eficiência, produtividade, rendimento, qualidade e sustentabilidade agrícolas.

O reconhecimento da variabilidade nas herdades, parcelas e linhas leva a uma utilização mais racional dos recursos, reduzindo-se o desperdício e os impactos ambientais (Majumder et al., 2021). A evolução, digitalização e utilização de tecnologias dão origem a novas formas de gestão (Karunathilake et al., 2023).

A caracterização do solo é um dos pilares fundamentais para orientar boas práticas do uso do solo, como a fertilização, as operações culturais ou os tratamentos fitossanitários (Khan, 2020). No entanto, as análises convencionais podem ser caras e demoradas, especialmente quando se trata de um elevado número de amostras (Cozzolino, 2006; McCarty et al., 2002). Alternativamente, a espectroscopia surge como uma técnica promissora, rápida, económica, não destrutiva, ecologicamente sustentável e altamente reprodutível, e que pode fornecer dados em tempo útil para serem integrados nas ferramentas de suporte à tomada de decisão.

Com mais de três décadas de equipamentos acessíveis no mercado, a utilização do infravermelho próximo (NIR), assim como outras técnicas de deteção proximal, tem demonstrado capacidade de prever uma ampla gama de propriedades do solo, embora por vezes dependam de bases de dados e metodologias específicas. Existem muitos estudos que confirmam a sua fiabilidade e boa relação entre custo e precisão para várias propriedades do solo, nomeadamente carbono orgânico e inorgânico (Morgan et al., 2009; Vohland et al., 2011), azoto (Shi et al., 2013; Zhang et al., 2016) e matéria orgânica (Fidêncio et al., 2002; Conforti et al., 2015).

1.2 Fertilidade do Solo

A fertilidade do solo, de acordo com Abbott e Murphy (2007), refere-se à capacidade de fornecer condições físicas, químicas e biológicas que sustentem a produtividade, reprodução e qualidade das plantas, exercendo impacto no bem-estar humano e animal. Inclui a fertilidade biológica, relacionada com a atividade de organismos do solo; a fertilidade química, associada à disponibilidade de nutrientes e suporte de processos biológicos e químicos; e a fertilidade física, ligada à aptidão do solo em oferecer condições estruturais sem comprometer a sua integridade.

A fertilidade química do solo, a qual foi diretamente estudada nesta dissertação, corresponde à capacidade do solo de disponibilizar, em quantidades e proporções adequadas, os nutrientes essenciais ao desenvolvimento das plantas e ao equilíbrio químico do solo, que permite a utilização dos nutrientes por parte das plantas. São necessários em maior quantidade os macronutrientes (C, H, O, N, P, K, Ca, Mg, S) e em menor os micronutrientes (Fe, Mn, Cu, Zn, Mo, B, Cl). A produtividade inicial de um solo tende a decrescer com o esgotamento ou o desequilíbrio dos nutrientes, exigindo a aplicação regular de fertilizantes. A determinação da fertilidade e das doses de fertilização adequadas deve ser realizada através de análises laboratoriais específicas, assegurando a reposição equilibrada dos elementos indispensáveis ao ciclo de vida vegetal (Loganathan, 1987). A manutenção da fertilidade do solo é a base para o uso sustentável do solo e a manutenção da sua fertilidade e requer a preservação da sua matéria orgânica, propriedades físicas e nutrientes (Du e Zhou, 2009).

A avaliação da fertilidade química do solo não é medida através de um único fator, mas sim estimada por meio de análises químicas do solo, abrangendo vários nutrientes e características químicas importantes para o crescimento das plantas. As análises realizadas podem depender do objetivo a alcançar. O teor de nutrientes (particularmente N, P e K), o teor de matéria orgânica, o pH e o COS (carbono orgânico no solo) são das principais propriedades do solo utilizadas na avaliação da fertilidade (Fainthfull, 2002; Mohkam-Singh, 2025).

O teor de P é frequentemente medido em solos calcários/neutros através do método de Olsen (NaHCO_3). O método de Mehlich-3 é amplamente usado por permitir análise simultânea de P, K, Ca, Mg e micronutrientes numa ampla gama de solos (Olsen, 1954; Mehlich, 1984). A medição do pH do solo é normalmente feita por potenciometria, utilizando um par de eletrodos para medir a diferença de potencial gerada pela atividade dos iões hidrogénio na suspensão solo/solvente (FAO, 2020b).

O teor de matéria orgânica pode ser obtido indiretamente através de vários métodos e variações desses métodos que estimam o carbono orgânico no solo, descritos na secção seguinte.

Para calcular a fertilidade global do solo a partir dos resultados de análises ao solo, vários autores sugerem agregar os resultados químicos, físicos e biológicos num índice de qualidade do solo. Este tipo de índice usa um conjunto de dados mínimo composto por propriedades biológicas, químicas e físicas e transforma-as em pontuações que, depois de somadas, fornecem uma avaliação global da fertilidade. Esta abordagem é útil porque transforma medições individuais em indicadores com informação prática que permitem definir estratégias de gestão (Chaudhry et al., 2024). A interpretação dos resultados depende dos valores de referência para a cultura e o tipo de solo, por exemplo, um solo com boa estrutura pode revelar deficiência de fósforo se a planta precisar de valores altos deste nutriente, ou pode até ser um solo rico em nutrientes, mas ter um pH demasiado ácido que reduz a disponibilidade de P (FAO, 2020a). As propriedades que limitam a disponibilidade de outras recebem um valor de importância maior, por exemplo, a matéria orgânica e o pH que afetam a disponibilidade de N, P e K e a atividade microbiana. A compactação também pode ser um fator limitante, uma vez que um solo mais denso e menos poroso retém menos ar e água e atrasa o crescimento das raízes, limitando a absorção de água e nutrientes.

A fertilidade do solo deve ser monitorizada ao longo do tempo para perceber se a fertilidade está a degradar ou a melhorar.

1.3 Carbono no Solo

Englobado na fertilidade do solo encontra-se o carbono. A quantidade de carbono acumulada no solo resulta da interação entre fatores bióticos, ambientais e o uso do solo (Post et al., 2001).

A principal entrada de carbono orgânico no solo deve-se à fixação de CO₂ atmosférico pelas plantas, através da fotossíntese, e ao transporte de parte (20-50%) deste fotossintetizado para as raízes e rizosfera e à incorporação de resíduos vegetais e animais (estrupe, restos de animais e organismos que morrem no solo). As colheitas contribuem substancialmente para os fluxos que podem ocorrer, uma vez que os seus resíduos, como raízes, caules e folhas deixados no campo após a colheita, são incorporados no solo durante a sua decomposição. Existem também alguns micro-organismos fixadores de carbono como algas e bactérias

fotossintéticas que podem contribuir para o enriquecimento do solo em carbono, principalmente nas fases iniciais da formação do solo ou em solos muito degradados. A utilização de práticas agrícolas, como a adubação verde, a rotação de culturas e a aplicação de compostos orgânicos e esterco, aumenta também a quantidade de biomassa disponível.

A principal perda de carbono do solo ocorre devido à respiração conduzida por microrganismos que decompõem a matéria orgânica, libertando CO_2 para a atmosfera (Wu et al., 2024). Existem ainda processos físicos, como a erosão e lixiviação, nos quais compostos orgânicos ou inorgânicos de carbono são transferidos para águas subterrâneas e superficiais, contribuindo para a perda de carbono armazenado no solo (Montanarella et al., 2015). A Figura 1 representa esquematicamente estes fluxos.

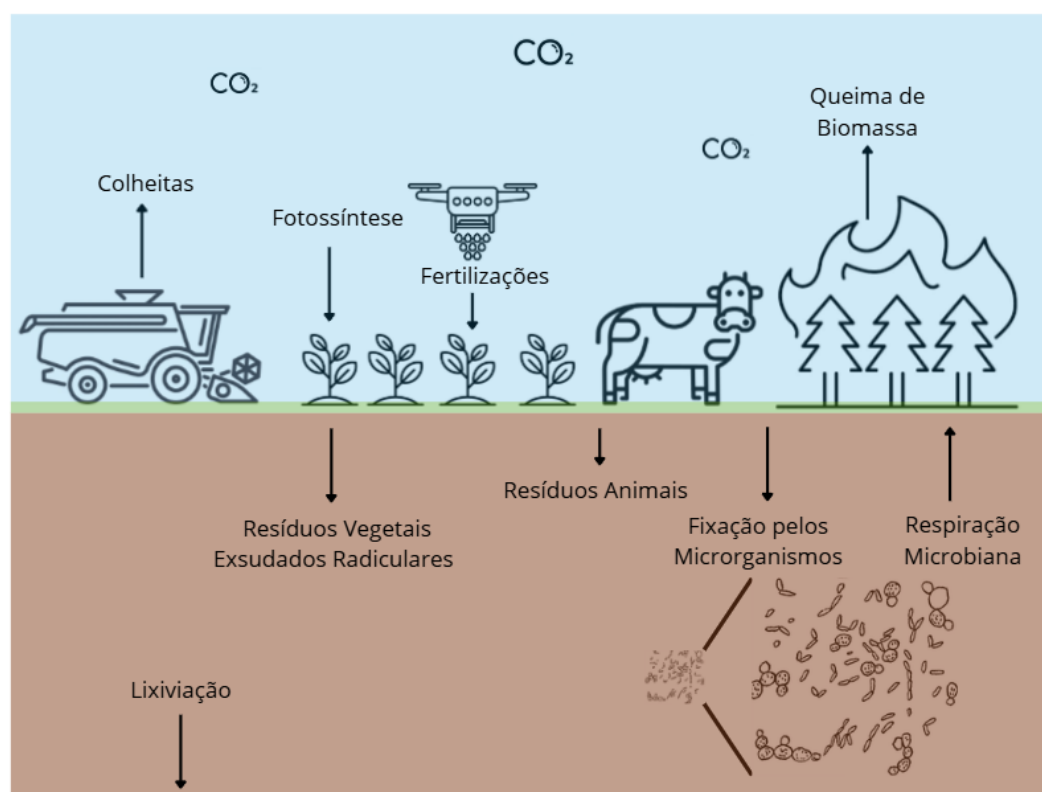


Figura 1 - Representação esquemática dos fluxos de carbono no solo.

A respiração do solo em sistemas agrícolas varia tipicamente entre 0,2 e 0,6 g de $\text{CO}_2/\text{m}^2/\text{h}$, podendo corresponder a 1,8–5 toneladas de carbono por hectare/ano. O carbono orgânico do solo apresenta uma taxa de renovação anual de cerca de 0,3–1,0 %, o que significa que, por exemplo, em 50 toneladas de C/ha, se renovam 0,15–0,5 toneladas de C/ha/ano. A utilização de culturas de cobertura pode aumentar o sequestro em aproximadamente 0,32

toneladas de C/ha/ano nas primeiras décadas, por outro lado, práticas de mau uso do solo podem resultar em perdas globais na ordem de 0,1 Pg C/ano (Tabela 1).

O valor de carbono armazenado no solo é grande comparado às variações nos fluxos anuais que nele podem ocorrer. No entanto, pequenas alterações nesses fluxos podem resultar em desequilíbrios a longo prazo, conduzindo quer a um maior sequestro, quer a maiores perdas para a atmosfera.

Tabela 1 - Balanço de carbono no solo. Fonte: soil.copernicus.org.

Fluxo	Origem	Destino	Valor típico
Entrada de carbono (resíduos vegetais, raízes, detritos)	Plantas vivas (folhas, caules, raízes)	Solo / Matéria Orgânica	~60–65 Pg C/ano
Respiração autotrófica (raízes)	Raízes vivas	Atmosfera (CO ₂)	~35–45 Pg C/ano
Respiração heterotrófica (microrganismos)	Matéria orgânica do solo	Atmosfera (CO ₂)	~50–55 Pg C/ano
Armazenamento de carbono orgânico	Acumulação no solo até 2 m de profundidade	Reservatório	~2400–3000 Pg C

Legenda: Pg: Petagrama, equivalente a 1 gigatonelada de carbono.

Globalmente, da terra habitável (superfície terrestre total excluindo corpos de água interiores) a terra dedicada à agricultura (terras aráveis, culturas permanentes e pastagens permanentes) representa 36,8% (World Bank, 2024), o que revela a importância de desenvolver práticas agrícolas adequadas para utilizar o solo como meio de sequestro e armazenamento de carbono orgânico.

O sequestro de carbono no solo pode desempenhar um papel significativo na regulação climática (Nayak et al., 2019). Uma vez que o solo armazena duas a três vezes mais carbono do que a atmosfera, um aumento relativamente pequeno do sequestro de carbono no solo repercute-se significativamente na redução da concentração de carbono na atmosfera. Utilizar o solo como reservatório de carbono é uma alternativa promissora e natural em comparação com processos tecnológicos mais dispendiosos ou menos acessíveis, como a captura direta de CO₂ na atmosfera. O aumento do conteúdo de carbono orgânico no solo melhora também a sua estrutura (Minasny et al., 2017), uma vez que contribui para a formação de agregados no solo, unindo partículas de areia, limo e argila que dão ao solo uma estrutura agregada e porosa, aumenta a infiltração de água, diminui a compactação e facilita a circulação de ar,

essencial para as raízes e os microrganismos, serve também de alimento à atividade biológica (Parvej et al., 2025). Um solo rico em carbono retém melhor a água e os nutrientes, tornando-se mais fértil e estável ao longo do tempo e menos sujeito a erosão. Participa na agregação e sorção de poluentes (Montanarella et al., 2015). Todos esses benefícios contribuem para valores maiores de produtividade.

O sequestro de carbono orgânico no solo, convertido em créditos de carbono, pode ser vendido a empresas interessadas em compensar as suas emissões de gases com efeito de estufa, e ser adquirido a agricultores que implementam boas práticas de gestão do solo que promovam a retenção de COS (Iowa State University, 2021).

No entanto, a aplicação real dos créditos de carbono tem sofrido alguns debates: leis diferentes de país para país (fiscalizações diferentes) e projetos cujo impacto é mal avaliado devido a interesses económicos (sobrestimando o impacto positivo), dificuldades em medir o impacto real (e a sua longevidade), ou substituição das emissões para outro lugar ou processo (Badgley et al., 2022; Coglianese e Giles, 2025). Muitas empresas declaram-se neutras em carbono, optando por compensar as suas emissões em vez de as reduzir (Trouwloon et al., 2023).

Um mercado de carbono com protocolos adequados, metodologias e monitorização oferece às empresas uma oportunidade valiosa de contribuírem de forma íntegra para a mitigação das alterações climáticas, investirem em projetos impactantes na geração de conhecimento, transferência de tecnologia e financiamento de regiões que normalmente não beneficiam de investimento privado (Trouwloon et al., 2023).

O bom funcionamento dos mercados assenta na estimativa fiável do teor de carbono no solo, mas para isso é essencial que exista uma padronização das técnicas e protocolos utilizados, bem como a criação de diretrizes adaptadas às diferentes escalas de amostragem e a diferentes métodos (Paustian et al., 2016). O potencial de um solo para armazenar carbono orgânico depende, entre outros fatores, da fração e estabilidade do carbono orgânico presente (Dignac et al., 2017).

O conteúdo de carbono orgânico presente no solo pode ser estimado por vários métodos. A diversidade de metodologias, bem como a complexidade dos procedimentos, dificultam a seleção de um método universal que funcione eficientemente para analisar o carbono orgânico presente em diferentes tipos de solo. A utilização de novas técnicas pode ultrapassar ou minimizar as dificuldades encontradas nas análises tradicionais.

Embora demorado e caro, a estimativa de carbono orgânico total (TOC) por oxidação química utilizando uma mistura de dicromato e ácido sulfúrico, com ou sem aplicação de calor

externo (Schwartz, 1995) é um dos métodos tradicionalmente mais utilizados. É um método bastante direto para determinar o carbono total (ou carbono orgânico, dependendo de se se remove ou ignora o carbono inorgânico, como carbonatos). A utilização de analisadores elementares baseados no método de combustão seca de Dumas (Walkley, 1934; Nelson, 1996; Heiri, 2001) em que a amostra de solo, muito finamente pulverizada, é queimada em ambiente oxidante (O_2) a altas temperaturas (tipicamente cerca de 900-1.000 °C) até oxidar todo o carbono contido e o CO_2 formado é medido por detetores sensíveis. Esta metodologia apresenta elevada precisão, mas requer equipamento caro que não está sempre disponível.

De acordo com os dados bibliográficos disponíveis, o método mais utilizado para avaliação do carbono do solo é a avaliação da perda de massa do solo por ignição (em inglês *loss-on-ignition*, LOI), em que durante o processo o COS é convertido em CO_2 . É um método simples, económico e com boa correlação com medições de COS nos mais variados tipos de solo (Chatterjee et al., 2009); no entanto, a precisão do método depende de várias características: do solo, profundidade e tamanho da amostra, e das condições de temperatura e duração do período de queima (Schulte et al., 1991). Uma das principais limitações deste método é a sobreposição de temperaturas de combustão entre carbono orgânico e inorgânico que dificulta a contabilização precisa de cada um, principalmente em solos de natureza calcária (David, 1988; Chatterjee et al., 2009).

1.4 A Reflectometria do Infravermelho Próximo

O NIR (*Near-Infrared*) ou infravermelho próximo corresponde a um intervalo específico do espectro eletromagnético onde a radiação eletromagnética (no vácuo) tem um comprimento de onda entre 750 e 2500 nanómetros. Correspondendo ao intervalo entre a extremidade vermelha da luz visível e o infravermelho médio. Quando a radiação atinge uma amostra, a radiação incidente pode ser refletida, absorvida ou transmitida e a contribuição relativa de cada fenómeno em cada comprimento de onda depende da constituição química e dos parâmetros físicos da amostra (Nicolai et al., 2007).

O NIR tem propriedades semelhantes à luz visível em termos de reflexão, absorção e transmissão em materiais, mas difere em dois pontos essenciais: na sua visibilidade para os seres humanos e na forma como interage com a matéria.

- 1) Não é visível ao olho humano porque o sistema visual humano evoluiu para detetar apenas radiação eletromagnética com comprimentos de onda entre 380 e 750 nm

(zona visível). Pode, no entanto, ser sentido na forma de calor, por exemplo, 52% da radiação solar encontra-se na região do infravermelho próximo (Cai et al., 2023).

- 2) Interações de forma diferente com a matéria porque a radiação NIR tem energia suficiente para ativar vibrações nas ligações das moléculas, como ligações de materiais orgânicos (ligações C–H, O–H ou N–H), consegue indicar a composição química de materiais ou estruturas em tecidos biológicos, mais especificamente, consegue captar ecos múltiplos e combinações de vibrações moleculares (Williams e Norris, 2001). No entanto, não é energética o suficiente para alterar ligações químicas, não sendo, por isso, perigosa nesse sentido.

É esta segunda característica que permite que seja amplamente utilizado para a caracterização de materiais orgânicos, como produtos agrícolas e alimentícios, ou em setores como o petroquímico e farmacêutico (Williams et al., 2001).

A reflectometria NIR apresenta várias vantagens sobre algumas técnicas convencionais de análise do solo por permitir uma rápida aquisição de dados, não ser dispendiosa (depois da aquisição do equipamento inicial) e, portanto, mais eficiente para um grande número de amostras (Nanni e Dematte, 2006). Permite ainda a produção de grande quantidade de dados de solo (Big data sets) de boa qualidade a baixo custo (Mouazen et al., 2007). Requer menos preparação de amostra do que, por exemplo, o MIR (Infravermelho médio) (Bellon-Maurel & McBratney, 2011).

Na agricultura, é muito utilizado como indicador da saúde das plantas. Não interage com os pigmentos fotossintéticos (como a radiação na zona do visível, onde a luz vermelha é absorvida para a fotossíntese). Ao invés disso, a reflectância do NIR na vegetação deve-se sobretudo à estrutura anatômica interna das folhas. Quando incide sobre as folhas, atravessa os tecidos internos e interage com o mesófilo esponjoso, uma camada de células com grandes espaços intercelulares preenchidos por ar. É a descontinuidade entre os índices de refração da água, da parede celular e do ar que provoca múltiplas reflexões e dispersões. O resultado é que grande parte da radiação NIR é refletida para fora da folha, enquanto outra parte é transmitida. Essa elevada reflectância no NIR é uma característica típica da vegetação saudável. É por isso que é utilizado como indicador da saúde das plantas em índices como o NDVI (*Normalised Difference Vegetation Index*). Em condições de *stress* hídrico, senescência ou danos na estrutura celular, a organização do mesófilo e a quantidade de água sofrem alterações, reduzindo a dispersão e, consequentemente, a reflectância no NIR, o que permite detetar alterações fisiológicas da vegetação a partir de imagens de satélite.

Existem outras indústrias a explorar a utilização do NIR. O artigo “NIR helping the search for life in space” (2008), por exemplo, discute a utilização do NIR no espaço para descobrir planetas com características habitáveis, na área da saúde para detetar doenças como o Alzheimer e em estudos com pessoas com esquizofrenia. No setor alimentício, as suas características estão a ser usadas para prever a frescura da carne de forma não destrutiva (Kim, Ryu, Kim, & Jo, 2025) ou para autenticar e rastrear a origem geográfica de peixes, ajudando a prevenir fraudes alimentares (Dalal, Ofano, Ruggiero, Caporale, & Adamo, 2024).

1.4.1 A Reflectometria NIR para Avaliar a Fertilidade do Solo

Existem vários artigos que confirmam a possibilidade de avaliar o conteúdo de nutrientes presentes ou disponíveis no solo utilizando o espectro do infravermelho (IV) (Chang et al., 2001; Pirie et al., 2005; Verma e Deb, 2007). Os teores de C e N do solo são os mais estudados com esta metodologia porque são os que mais reagem à radiação infravermelha. Para a previsão de outros nutrientes presentes no solo, como o P, K e micronutrientes, os resultados podem ser fiáveis e reprodutíveis só após uma calibração intensa que incorpore nos algoritmos encontrados o contexto local (Du et al., 2009). Mais recentemente, a utilização da reflectometria NIR, aliada por vezes ao espectro da zona visível, tem sido útil para desenvolver índices de fertilidade do solo a partir de espectros recolhidos em campo, permitindo monitorizar a variabilidade espacial da fertilidade com boa precisão (Munnaf e Mouazen, 2021). Pode funcionar como ferramenta para estimar com precisão a disponibilidade de nutrientes no solo, apoiando práticas de fertilização mais eficientes e sustentáveis (Recena, 2019), e em contextos muito variáveis que incluem solos secos ou encharcados como os campos de arroz (Yang et al., 2020).

1.4.2 A reflectometria NIR para determinação de Carbono

A espectroscopia no infravermelho próximo representa uma técnica promissora para a estimativa de carbono no solo, devido à sua capacidade de detetar propriedades únicas de absorção e reflectância e, por conseguinte, assinaturas espectrais presentes de vários componentes da matéria orgânica do solo na região do IV (Shepherd e Walsh, 2002).

Os primeiros estudos sobre aplicações de NIR em campo demonstraram a viabilidade de medições *in situ* do carbono do solo. Como demonstrado por Shonk et al., que em 1991 desenvolveram um sensor que apresentou uma forte correlação entre a reflectância e a matéria orgânica no solo. Anos depois, Christy et al. (2003) descreveram o desenvolvimento de um dispositivo móvel de NIR *in situ* capaz de realizar mapeamentos de carbono em tempo real.

Este desenvolvimento permitiu assim estimativas de carbono no solo de uma forma rápida, não destrutiva e adaptável ao uso em campo, embora as medições em campo frequentemente apresentem maiores erros de previsão comparadas com as medições em laboratório (Sudduth e Hummel, 1996). Algumas dificuldades relacionam-se com a humidade do solo, o tamanho das partículas e a calibração em diferentes tipos de solos.

Ulteriormente, a utilização da reflectometria NIR revelou-se uma ferramenta promissora para a estimativa de carbono orgânico do solo (COS) em grande escala. Embora os modelos espectroscópicos apresentassem erros consideráveis, especialmente em áreas com grande diversidade de solos, estes permitiram fazer previsões para estimar o COS de maneira consistente em grandes regiões. Utilizando, por exemplo, a biblioteca LUCAS (<https://esdac.jrc.ec.europa.eu/>). Estes avanços contribuíram para a construção de bibliotecas espectrais locais específicas que podem ser extrapoladas a outras áreas ou regiões (Stevens et al., 2013).

1.5 Objetivos

A literatura mostra resultados promissores tanto na utilização da reflectometria NIR para a previsão das propriedades do solo às quais se associa a fertilidade quanto ao caso específico do carbono, mas também destaca a especificidade e dificuldade no processo de calibração para cada região e tipo de solo. Existe espaço para melhorar a precisão das estimativas, acrescentando variáveis e desenvolvendo melhores modelos.

É neste sentido que esta dissertação pretende desenvolver um modelo de previsão das propriedades químicas do solo, utilizando exclusivamente dados espectrais NIR. Uma alternativa às análises laboratoriais tradicionais, que permite uma caracterização mais rápida do solo e, conseqüentemente, acelera a avaliação das necessidades do solo, otimiza o uso de fertilizantes, reduz desperdícios e minimiza a dependência de análises laboratoriais dispendiosas, demoradas e exigentes em mão de obra, tornando a agricultura mais acessível, sustentável e eficiente.

A criação de modelos com dados NIR apresenta desafios, como: a elevada dimensionalidade, a presença de ruído e a variabilidade natural dos solos, o que dificulta a generalização quando aplicada a diferentes tipos de solo.

Para superar estas limitações, serão testados modelos de regressão multivariada, nomeadamente, *Random Forest* e *Partial Least Squares Regression*. A avaliação dos modelos será realizada através de validação cruzada e métricas: RMSE, R^2 e RPD. A criação de mapas com os valores previstos permitirá a visualização da variabilidade por propriedade no solo.

Metodologia

Na presente secção descreve-se a metodologia utilizada para a obtenção dos dados que serão processados na secção seguinte. Na Figura 2 observa-se, de forma resumida, a sequência de passos realizada nesta dissertação. Começando pela seleção dos locais onde realizar a amostragem, seguida das análises realizadas às mesmas e concluindo com a utilização dos dados obtidos para a criação de mapas de previsão das propriedades.

No presente capítulo são apresentadas a caracterização do local de ensaio, a fundamentação concernente à seleção dos pontos de amostragem e a descrição do procedimento de recolha das amostras. Adicionalmente, são descritas as análises laboratoriais realizadas a cada amostra.

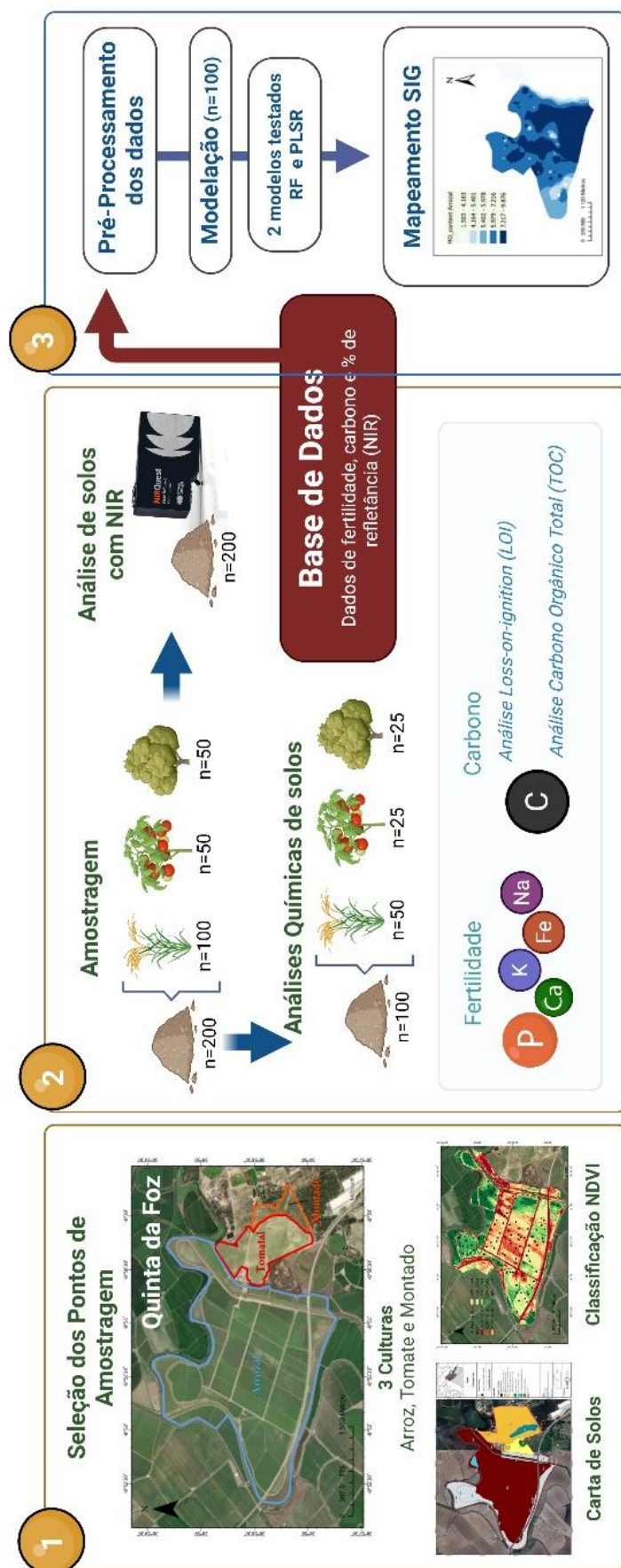


Figura 2 - Esquema geral da metodologia aplicada.

2.1 Descrição da Área de Estudo

A Quinta da Foz foi o local escolhido para o desenvolvimento do estudo e modelação das propriedades dos solos agrícolas tendo por base a utilização do sensor de reflectância NIR. Localizada, como ilustrado na Figura 3, no município de Benavente, a Quinta da Foz é uma propriedade dedicada à produção de diferentes culturas agrícolas e florestais, que abrange cerca de 650 hectares. Atualmente, a Quinta da Foz dedica a sua atividade agrícola maioritariamente à produção de arroz, com uma área de 542 hectares, sendo que possui outra área agrícola de 80 hectares, atualmente dedicada à produção sazonal de tomate e uma área de 22 hectares de montado.

Historicamente, as parcelas da Quinta da Foz tiveram outras utilizações, sendo que na zona do arrozal, para além do uso atual, já foi também produzido milho, melão, girassol e tomate. A zona de montado manteve-se sempre como tal, embora tenha funcionado também como prado natural para pastoreio de gado bovino e equino. Na zona atualmente ocupada pelo tomate, existiram prados semeados e regados, igualmente utilizados para o pastoreio de bovinos e equinos (Gomes da Silva, Comunicação pessoal).

A seleção deste local para o estudo proposto deve-se ao fato de ser um local com produções e tipos de solo diversos. Isso permite a obtenção de amostras de solo com propriedades e assinaturas espectrais muito diversas. Variação esta, fundamental para desenvolver um modelo robusto e adaptado a diferentes contextos.

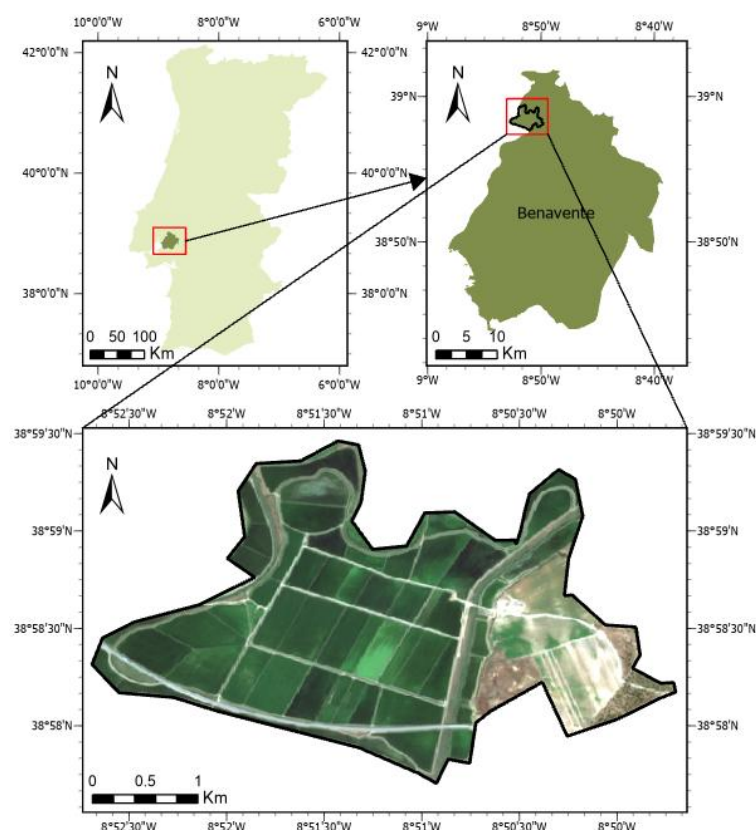


Figura 3 - Localização geográfica à escala nacional do município de Benavente, localização da Quinta da Foz dentro dos limites do município e vista detalhada do limite da parcela de estudo.

2.2 Definição dos Locais de Amostragem

O primeiro passo para a definição dos locais de amostragem foi a divisão da área da Quinta da Foz em zonas, consoante a sua utilização atual agrícola e florestal, com o objetivo de identificar variações dentro de cada uma delas. A divisão, representada na Figura 4, resultou em três zonas: arrozal (541,6 ha), tomatal (80,03 ha) e montado (21,91 ha). Recorreu-se ainda à carta de solos para realizar a caracterização de cada zona e verificar a distinção natural entre os solos dessas áreas. As três zonas - arrozal, tomatal e montado – estão inseridas em tipos de solo diferentes com gestão distinta ao longo do ano. Por este motivo, optou-se por uma amostragem dirigida. A seleção dos pontos de amostragem baseou-se em imagens de satélite com valores históricos do ano de 2023 de NDVI.

O NDVI é um indicador do vigor vegetativo, resulta da relação entre a absorção e a reflexão da radiação nas faixas do vermelho e do infravermelho próximo. Plantas saudáveis, ricas em clorofila, absorvem mais luz azul e vermelha e refletem mais luz verde e do infravermelho próximo. A variação do NDVI é amplamente utilizada para: distinguir plantas saudáveis de plantas em stress, identificar áreas cultivadas, apoiar aplicações de fertilizantes,

otimizar a gestão da rega e monitorizar o crescimento ao longo do tempo (Jena et al., 2019). Neste estudo, o NDVI foi usado para selecionar as datas de imagem e os locais de amostragem.

Através da aplicação “Smartag” da “AgroInsider” (<https://smartag.agroinsider.com/>) obtiveram-se gráficos temporais de NDVI para cada parcela (disponíveis no Anexo I). Variações abruptas no índice geralmente correspondem a dias nublados, com maior interferência no cálculo. Como exemplo, a Figura 15 a) (Anexo I), mostra a evolução do NDVI ao longo do ano de 2023 no montado, a parcela menos sujeita a oscilações.

A distribuição das amostras foi proporcional à área de cada zona, com um total de 200 amostras: 100 no arrozal, 50 no tomatal e 50 no montado. Foram então criadas classes de NDVI no ArcGIS Pro (divisão em quantis) para assegurar uma frequência equilibrada de observações por classe. A seleção dos pontos foi feita preferencialmente no centro das áreas homogêneas. No montado, considerou-se ainda a heterogeneidade do sistema, escolhendo-se pontos tanto sob a copa como em áreas descobertas. No tomatal, em cada uma das duas subzonas de menor dimensão foram feitas apenas cinco amostras, com classificação simplificada em duas classes de NDVI. No arrozal, pela maior área e número de classes, deu-se preferência a locais com melhor acessibilidade devido à dificuldade de deslocação dentro dos canteiros de arroz (no Anexo II encontram-se as coordenadas das amostras).



Figura 4 - Representação da divisão da herdade em arrozal, tomatal e montado.

2.3 Amostragem

Após definir os locais de amostragem, procedeu-se à recolha do solo nos pontos previamente marcados usando a aplicação do *Google Earth*. As coordenadas geográficas correspondentes foram importadas para um dispositivo portátil com capacidade de navegação por GPS, permitindo a navegação até cada ponto de amostragem.

Antes da recolha, a camada superficial do solo foi limpa de folhas, ervas e outros detritos com recurso a uma pequena pá, de forma a expor apenas o solo. Em seguida, um cilindro metálico foi inserido no solo até à profundidade de 10 cm, recorrendo a uma barra ou marreta caso necessário, embora a maioria dos solos fosse arenosa ou húmida, oferecendo baixa resistência. O solo recolhido foi transferido para sacos de plástico, devidamente identificados com o código da amostra (Figura 5).

Foram realizadas 200 amostras de solo: 100 no arrozal, 50 no tomatal e 50 no montado. A análise de reflectometria NIR foi feita a todas as amostras e as análises químicas apenas a metade, uma vez que o objetivo é prever a outra metade a partir dos dados de reflectância NIR.

Para as amostras destinadas às análises químicas e espectrais, a quantidade recolhida foi aproximadamente 1 kg, de forma a garantir a quantidade necessária para as análises a serem realizadas. Enquanto para as restantes amostras em que se iria realizar apenas análise espectral, a quantidade foi de cerca de 500 g.



Figura 5 - Instrumentos utilizados na recolha de amostras.

2.4 Análises Químicas do Solo

2.4.1 Fertilidade do Solo

As amostras de solo às quais se pretendia realizar análises químicas (metade, cem amostras) foram separadas em sacos identificados de pelo menos 500 g e enviadas para a empresa de análises químicas A2 (<https://www.a2analisesquimicas.com/>) um laboratório especializado em análises de solo. As propriedades químicas foram determinadas através de protocolos padronizados para análise de solos (Neves, 2015). As amostras de solo foram preparadas de acordo com a norma ISO 11464:2006, sendo homogeneizadas, secas em estufa a 40 °C durante cerca de 15 h, peneiradas numa malha de 2 mm e armazenadas em sacos de papel. O pH do solo foi determinado por potenciometria, de acordo com a norma ISO 10390:2005, após preparação de suspensões em água destilada, CaCl_2 (0,01 M) e solução tampão de hidróxido de cálcio. A condutividade elétrica (CE) do solo foi determinada por condutimetria, de acordo com a norma ISO 11265:1994, utilizando a mesma suspensão preparada para a medição do pH, com os resultados expressos em $\mu\text{S}/\text{cm}$ normalizados à temperatura ambiente. A determinação dos teores de carbono e azoto foi realizada num analisador elementar que funciona com base na combustão dinâmica da amostra, seguida de separação por cromatografia gasosa, que separa os constituintes gasosos em dois picos distintos: azoto elementar (azoto) e dióxido de carbono (carbono). Estes são depois detetados por condutividade térmica. A razão C:N foi posteriormente calculada a partir dos valores obtidos, expressa em percentagem de carbono e de azoto, e foi determinada com base na razão entre a massa do elemento de carbono e a percentagem de azoto. O teor de matéria orgânica foi obtido através da multiplicação do teor de carbono (determinado por analisador elementar, já acima descrito) pelo fator de Van Bemmelen (1,724), considerando que, no total da matéria orgânica, 58% corresponde a carbono orgânico. O teor de azoto total foi determinado por analisador elementar, já descrito acima. A determinação dos micro- e macronutrientes (Fósforo (P_2O_5), Potássio (K_2O), Cálcio (CaO), Magnésio (MgO), Enxofre, Boro, Ferro, Manganês (total), Manganês (disponível, MnAI), Cobre, Zinco, Molibdénio, Cobalto, Níquel e Sódio) foi feita a partir da técnica ICP-OES, adicionando, a duas gramas de solo, as soluções Mehlich 3 e CEC, sendo o sobrenadante posteriormente diluído (1:9) para análise por ICP-OES. Os resultados, em mg/kg, indicam o valor do micro- ou macronutriente por quilograma de solo. Para o caso do P_2O_5 , K_2O , CaO , MgO e MnAI o valor é referente à quantidade assimilável pelas plantas, para os restantes refere-se ao valor total.

2.4.2 Teores de carbono no solo

A obtenção dos teores de carbono no solo foi realizada por três métodos que mediram frações específicas de carbono (Figura 6). A obtenção do teor de carbono total foi obtida por análise elementar e permitiu calcular o valor total incluindo carbono orgânico e inorgânico. A análise LOI permitiu o cálculo do carbono orgânico das amostras (incluindo C dissolvido, resíduos frescos, C microbiano, humos e matéria orgânica mais resistente). A temperatura utilizada para causar a perda de massa (500°C) não foi suficiente para libertar o C inorgânico. A análise do TOC permitiu calcular a fração de carbono microbiano e dissolvido na solução do solo.

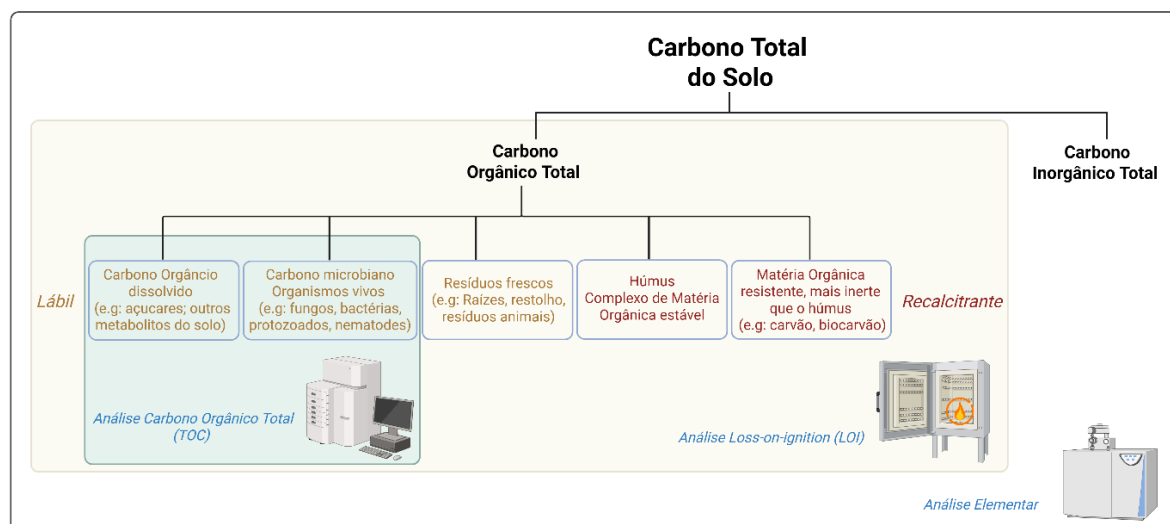


Figura 6 - Diagrama sistemático das frações de carbono do solo e análises utilizadas para a sua determinação.

2.4.2.1 Método por Perda de Ignição (LOI)

A avaliação do carbono orgânico do solo por perda de ignição foi feita com base na análise de amostras de 1 g de solo fresco. Cada amostra de solo foi pesada para um cadinho de porcelana numa balança de precisão digital (COBOS CB-Junior, erro ± 0.0005). Cada cadinho foi identificado e associado ao código da amostra (Figura 7a) e foi anotado o peso correspondente. As amostras de solo foram secas até peso constante (± 24 h) numa estufa a 45 °C com ventilação. Após ter atingido o peso constante, cada amostra foi transferida para a mufla (Nabertherm, modelo L3/11/06) onde permaneceu à temperatura de 500 °C durante 16 horas. Após esse período a mufla foi desligada, as amostras permaneceram no seu interior até estarem à temperatura de 100 °C (Figura 7b).

Quando a esta temperatura os cadinhos foram retirados com uma pinça de metal e colocados num local seco (máximo de 20% de humidade), onde foram de novo pesados (utilizando a balança utilizada para a determinação do peso seco). Todas as determinações foram feitas em triplicado. Os dados foram registados numa folha Excel com os três pesos medidos e procedeu-se ao cálculo da matéria orgânica segundo a fórmula abaixo (Robertson, S., 2011).

$$MO (\%) = \left(\frac{\text{Peso amostra seca (g)} - \text{Peso amostra queimada (g)}}{\text{Peso amostra seca (g)}} \right) \times 100$$

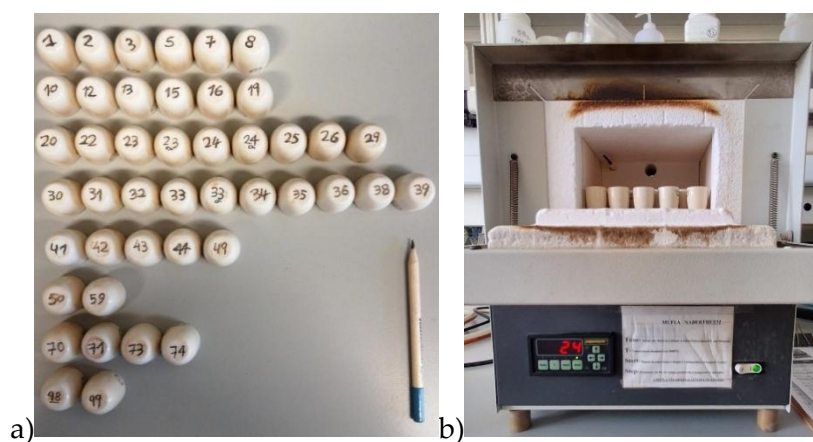


Figura 7 - Materiais utilizados na análise LOI.

2.4.2.2 TOC

Para esta análise, foram pesados 5 g de solo fresco em recipientes novos ou previamente lavados com ácido clorídrico (HCl 10%) e devidamente identificados. As amostras não fumigadas foram transferidas diretamente para tubos *Falcon* de 50 ml, enquanto as amostras destinadas à fumigação foram pesadas em gobelets de 50 ml, identificados a lápis.

A fumigação foi realizada na câmara de fumos do laboratório num exsicador de vidro no qual foi previamente colocado papel humedecido no fundo. No interior do exsicador, foram colocados os gobelets com as amostras de solo e um gobelet contendo aproximadamente 50 ml de clorofórmio e esferas de vidro (para evitar trasbordo do clorofórmio e contaminação das amostras). Nas zonas de contato entre a tampa e a base do exsicador, aplicou-se lubrificante de silicone para vácuo, de modo a facilitar a posterior remoção da tampa. Com o exsicador fechado, foi conectada uma bomba de vácuo, que foi acionada até o clorofórmio entrar em ebulição. Após cerca de 30 segundos de ebulição, fechou-se a torneira do exsicador e cobriu-se o mesmo com papel de alumínio, de forma a impedir a entrada de ar e luz.

Passadas 24 horas, o papel de alumínio foi retirado, a torneira aberta para permitir a entrada de ar e a tampa levantada. O gobelet com clorofórmio e o papel humedecido foram removidos, o exsicador foi novamente fechado e ligado à bomba de vácuo para eliminar resíduos de clorofórmio da atmosfera interior. Posteriormente, as amostras de solo foram transferidas para tubos *Falcon* de 50 ml devidamente identificados.

O procedimento seguinte foi comum às amostras fumigadas e não fumigadas. Os tubos foram cheios com solução de K_2SO_4 0,5 M até ao volume de 50 ml. As amostras foram agitadas até à completa desagregação do solo, recorrendo a um agitador *vortex* sempre que necessário, e colocadas num agitador horizontal durante 30 minutos. Seguidamente, foram centrifugadas durante 20 minutos a 3500 rpm, após pesagem prévia para garantir o equilíbrio da centrífuga.

Após a centrifugação, as amostras foram filtradas com recurso à bomba de vácuo (Figura 8), o filtrado foi recolhido em tubos *Falcon* novos e identificados, ficando assim preparadas para análise num analisador de TOC.

A determinação do carbono microbiano foi realizada com base na diferença entre os valores obtidos em amostras fumigadas e não fumigadas (equação abaixo). Parte-se do pressuposto de que a fumigação é a única variável entre as duas amostras, pelo que a diferença entre os conteúdos de carbono da amostra fumigada e não fumigada pode ser atribuída ao carbono libertado pela lise das células na presença de clorofórmio, que na sua maioria corresponde a carbono microbiano. Este carbono adicional passa a pertencer à fração medida de carbono orgânico total, permitindo, assim, a estimativa do teor de carbono microbiano no solo.

$$C_{\text{Microbiano}} (\text{mg/L}) = \text{TOC após fumigação} - \text{TOC}$$



Figura 8 - Processo de filtração das amostras com recurso à bomba de vácuo.

2.5 Análise dos Solos com o Reflectómetro NIR

A detecção da reflectância no infravermelho foi realizada com o sensor NIRQUEST+2.5 (Ocean Insight) (<https://www.oceanoptics.com>) (Figura 9). As amostras de solo para análise foram dispostas em caixas de Petri de igual dimensão (9 cm de diâmetro), , sendo realizadas três réplicas por amostra. Tendo em conta a sensibilidade do sensor, confirmou-se que: as caixas estavam totalmente preenchidas e a superfície nivelada, não existiam vibrações externas (por exemplo, provenientes de centrífugas ou agitadores em funcionamento), não existiam correntes de ar direccionadas para a mesa de medição. Sempre que se verificava uma deslocação geral da curva espectral, ou aproximadamente a cada 30 minutos, o sensor era novamente calibrado.

Para cada amostra, foram então utilizadas três caixas de Petri, realizando-se três medições por caixa. Em cada medição, o sensor NIR registou automaticamente três leituras, resultando num total de 27 ficheiros por amostra.



Figura 9 - Sensor NIR em utilização no laboratório.

2.6 Modelação

Na presente secção descreve-se o processo de tratamento dos dados obtidos pelas metodologias anteriores, o tipo de algoritmos e validação testados que levaram à previsão das propriedades químicas. Na Figura 10 está representada a sequência de passos utilizados para obter os valores previstos que serão utilizados na criação dos mapas de previsão. As bases de dados resultantes das análises anteriores para as propriedades químicas e dados NIR podem ser consultadas no seguinte link: https://drive.google.com/drive/folders/15ckR-zIVV8RbnJQ1FYjxSf_QDm0-o4Hia?usp=sharing.

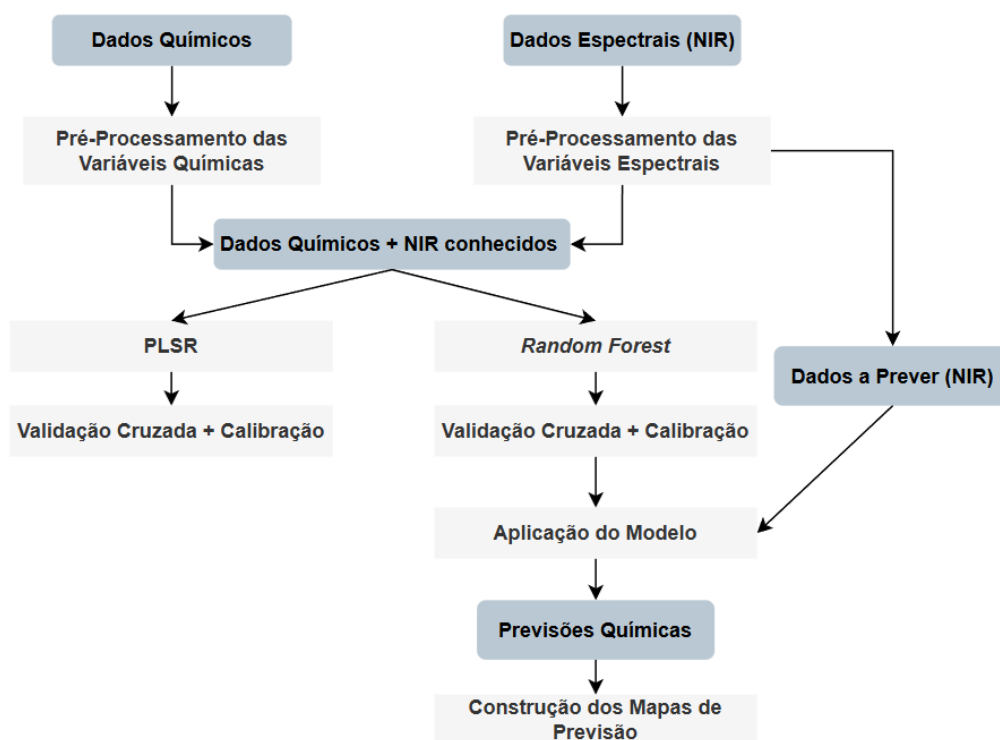


Figura 10 - Esquema sistemático do processo de modelação aplicado.

2.6.1 Organização e Pré-Processamento dos Dados

Os dados obtidos foram organizados numa tabela com as propriedades laboratoriais do solo (variável resposta ou dependente) e os comprimentos de onda NIR (variável preditiva ou independente). Foram criadas três novas variáveis: **ID**, **ID_Amostra**, variáveis usadas para identificação dos registos; e **Cultura**, variável descritiva do tipo de cultura. 0-Montado, 1-Tomate, 2-Arroz. Para os dados NIR, criou-se ainda uma variável binária: **Analisada** com valores 0 - se a amostra não tiver sido analisada, apenas foi feita a leitura NIR (amostras às quais se pretende prever os valores das propriedades químicas); ou 1 - se tiver sido analisada para fertilidade e carbono.

Todas as variáveis utilizadas na criação do modelo são numéricas. Os *scripts* desenvolvidos são disponibilizados em anexo e foram criados utilizando a linguagem *Python*.

2.6.1.1 Variáveis Químicas

Os dados das análises de fertilidade foram transcritos para uma tabela Excel, com os valores medidos por amostra. A partir da multiplicação do valor da razão carbono-azoto ($R_{azao_C_N}$) e da percentagem de azoto total ($Azoto_T$), obteve-se uma nova variável: a percentagem de carbono total ($Carbono_T$).

Os resultados da análise de LOI, após o cálculo da MO, foram obtidos pela média das três repetições de cada amostra (MO_LOI). Os valores do carbono microbiano que se pretendiam avaliar através do analisador de TOC, após cálculo, também foram adicionados à base de dados (C_microb).

Apesar do foco principal de utilizar o analisador TOC ser calcular o carbono microbiano do solo, na mesma análise, foram calculadas outras frações de carbono necessárias para o cálculo do carbono microbiano por isso incluíram-se também na análise na tentativa de perceber se seria possível prevêê-las. Foram retirados os resultados do carbono orgânico total (TOC), do carbono inorgânico (IC) e do carbono total das amostras fumigadas (F_TC). Estas frações referem-se ao carbono solúvel e extraível, não ao carbono total presente no solo.

Resultando uma tabela completa contendo todos os dados sobre as propriedades do solo e as variáveis utilizadas, bem como a sua unidade e origem, apresentadas na Tabela 2.

Tabela 2 - Descrição das variáveis, respetivas unidades e metodologias de análise utilizadas.

<i>Variável</i>	<i>Unidade</i>	<i>Método de análise</i>
MO_LOI	%	LOI
MO	%	Análises de Fertilidade
C_microb	mg/L	TOC
TOC	mg/L	TOC
IC	mg/L	TOC
F_TC	mg/L	TOC
pH	-	Análises de Fertilidade
CE	mS/cm	Análises de Fertilidade
Razao_C_N	%	Análises de Fertilidade
Carbono_T	%	Cálculo
Azoto_T	%	Análises de Fertilidade
P (P ₂ O ₅)	mg/Kg	Análises de Fertilidade
K (K ₂ O)	mg/Kg	Análises de Fertilidade
Ca (CaO)	mg/Kg	Análises de Fertilidade
Mg (MgO)	mg/Kg	Análises de Fertilidade
MnAI	mg/Kg	Análises de Fertilidade
S, B, Fe, Mn, Cu, Zn, Co, Ni, Na	mg/Kg	Análises de Fertilidade

Previamente à aplicação dos modelos, foram realizados vários passos de preparação dos dados. Nos dados laboratoriais, foi analisada a percentagem de valores abaixo do limite de deteção. A variável Molibdénio foi removida, uma vez que todos os valores se encontravam no limite mínimo (<0,2) (sem variabilidade e poder preditivo, o modelo não consegue aprender de uma variável constante). Para os restantes casos, substituiu-se o valor pelo limite de deteção, para poder utilizar esses valores no modelo.

Realizou-se ainda uma análise exploratória que permitiu identificar a distribuição das propriedades e a remoção de *outliers* (no Anexo III pode ser encontrado o script produzido para produzir estes resultados).

2.6.1.2 Variáveis Espectrais

Os 27 ficheiros por amostra resultantes das leituras com o sensor NIR contêm os gráficos das leituras espectrais, foram posteriormente tratados e foi feita a média de forma a obter uma única curva espectral representativa por amostra.

Os dados espectrais continham inicialmente 512 colunas, correspondentes aos comprimentos de onda entre 894,248 nm e 2497,776 nm. Começou-se por agrupar e calcular a média para as medições do NIR por amostra, uma vez que para cada amostra se realizaram 3 repetições. O segundo passo foi visualizar as curvas de reflectância, estes valores devem estar entre 0 e 100, alguma anomalia pode significar erros de medição ou condições inadequadas, como elevada humidade. Após definir o intervalo a trabalhar, foram removidos os comprimentos de onda fora desse intervalo, reduzindo o total para 312. Aplicou-se de seguida médias móveis para suavizar as curvas e reduzir o ruído. No final, foram retidos 298 comprimentos de onda. O *script* correspondente à análise exploratória dos dados até à sua exportação pode ser encontrado no Anexo IV.

2.6.2 Criação dos Modelos, Validação e Previsão

Na realização deste trabalho, foram seleccionados dois algoritmos para prever as propriedades químicas do solo a partir dos dados espectrais: o *Random Forest* (RF) e a regressão *Partial least squares regression* (PLSR). Ambos os modelos são amplamente utilizados em dados espectrais (Vestergaard et al., 2021; Hao et al., 2023). Pela especificidade do tipo de dados, é necessário um modelo capaz de lidar com muitas variáveis correlacionadas (pois existe uma forte correlação entre espectros “vizinhos”), que consiga capturar relações lineares e não lineares e que seja robusto para dados com ruído. O PLSR permite reduzir a dimensionalidade e é eficaz quando as relações espectro-variável são lineares. O RF permite detetar relações complexas entre comprimentos de onda e pode encontrar padrões não lineares.

Para ambos os modelos – RF e PLSR, foi aplicada validação cruzada (CV), uma técnica estatística para avaliar a capacidade de generalização de um modelo, ou seja, como ele se comporta quando recebe novos dados que não viu durante o treino, para garantir que os resultados fossem consistentes e não dependessem apenas de uma divisão dos dados. Particularmente adequada para uso onde poucos dados estão disponíveis, na tentativa de aproveitar ao máximo as amostras existentes (Soriano-Disla et al., 2013).

Durante a CV a base de dados é dividida em cinco partes. A cada repetição o modelo é treinado com 4 partes e testado na quinta parte, repete-se o processo 5 vezes, cada vez com uma parte diferente a servir como teste, sucessivamente até que todas as partes tenham sido usadas para teste uma vez. O conjunto de treino é o grupo de dados onde o modelo realmente “aprende” e o conjunto de teste é usado para avaliar o desempenho final do modelo,

garantindo que ele consiga generalizar bem e fazer previsões para dados que não viu durante o treino.

Em cada repetição, 80% dos dados participam “no treino” e 20% participam “no teste”. No fim, juntam-se todas as previsões e calculam-se métricas (R^2_{CV} , RMSE_{CV} e RPD_{CV}) e com base nelas definem-se os hiperparâmetros ótimos. Após estas alterações, o modelo mais aprimorado é treinado para todas as amostras analisadas e guardado. Fazem-se as previsões nas mesmas amostras usadas para treinar e calculam-se as métricas (R^2_{cal} , RMSE_{cal} e RPD_{cal}).

Neste caso foram calculados o Coeficiente de Determinação (R^2), *Root Mean Squared Error* (RMSE) e *Ratio of Performance to Deviation* (RPD) para ambos os modelos para perceber o desempenho do modelo no geral e para cada propriedade. Estas métricas são as utilizadas geralmente para avaliar o desempenho (exatidão e precisão) quando se usam este tipo de variáveis (Soriano-Disla et al., 2013).

O R^2 mede a proporção da variabilidade explicada pelo modelo (valores $>0,9$ são excelentes, $>0,7$ bons a muito bons). É importante ter valores altos de coeficiente R^2 de calibração; no entanto, ter apenas esse valor alto pode indicar que o modelo está demasiado específico para os dados utilizados (*overfitting*). Por isso, também é importante ter um R^2 alto na validação cruzada, que reflete a capacidade de generalização, uma vez que esse valor demonstra o comportamento do modelo em dados que ele não conhece. O RMSE representa o erro médio nas unidades da variável real; quanto menor, melhor, embora a comparação entre variáveis seja limitada pelos diferentes intervalos de valores.

A razão de desempenho em relação ao desvio padrão (RPD) é outro indicador utilizado. Estatisticamente, um RPD < 2 é considerado insuficiente, um valor entre 2 e 2,5 representa uma previsão aproximada, um valor entre 2,5 e 3 pode ser classificado como bom e um RPD > 3 indica uma excelente previsão quantitativa (Zhang et al., 2017).

No caso do modelo PLSR, calcula-se também o número de componentes ideais na validação cruzada (o RF não trabalha com componentes, mas sim com os comprimentos de onda originais), escolhendo um equilíbrio entre simplicidade e capacidade preditiva. Poucos componentes podem não captar toda a informação e relações entre variáveis e levar a um modelo fraco (*underfitting*) e um número excessivo pode resultar num modelo que aprendeu a funcionar especificamente para o conjunto de dados utilizado e perder a capacidade de generalização, de prever dados que nunca tenha “visto” (*overfitting*). Os resultados obtidos foram guardados para comparar o desempenho dos dois modelos e a capacidade de previsão para cada

propriedade. Cada modelo treinado foi guardado individualmente, o que permite a sua reutilização futura sem necessidade de novo treino.

Por fim, os modelos foram aplicados a novas amostras. As previsões das variáveis químicas foram organizadas num ficheiro, juntamente com os identificadores, a informação da cultura correspondente e a localização. Esta análise foi feita em *Python*, utilizando as bibliotecas: *pandas*, *numpy*, *scikit-learn* e *joblib*.

2.7 Mapas de Valores Previstos

A gestão localizada apresenta três benefícios principais: maior eficiência no uso de fatores de produção, melhoria das margens económicas da produção agrícola e redução dos riscos ambientais. (Yasrebi et al., 2008). Em conjunto com o conhecimento das necessidades nutricionais da cultura, permite planear uma fertilização diferenciada e aplicar exatamente o que cada zona do campo necessita (Aggelopoulou et al., 2011).

Deste modo e com os valores previstos, utilizando o ArcGIS Pro, procedeu-se à criação dos mapas geográficos previstos. Estes foram criados utilizando interpolação recorrendo à ferramenta de IDW (*Inverse Distance Weighted*). Este método pressupõe que a influência da variável mapeada diminui com a distância da sua localização amostrada. Neste caso, num dado local, o solo tem tendência a ser mais parecido com o solo situado num local mais próximo do que com um situado num local mais afastado.

A interpolação é utilizada para criar uma superfície *raster* (contínua) a partir de pontos (neste caso, os valores reais e previstos das propriedades químicas).

Utilizando o sistema de coordenadas GCS_WGS_1984. Foram criados mapas específicos para cada uma das três zonas para várias propriedades. A simbologia foi alterada para cinco classes e para cores diferentes de forma a representar a distribuição geográfica da variação de cada propriedade em cada zona.

A variabilidade natural dos solos resulta principalmente da interação entre geologia (rocha-mãe), topografia, clima e uso do solo. A procura por informação mais precisa sobre a distribuição espacial dos solos tem aumentado (Yasrebi et al., 2008) e possibilita a identificação de padrões que seriam dificilmente reconhecidos apenas por observações diretas ou análises pontuais. Permite a definição de zonas homogêneas de gestão (Management Zones), que possibilitam intervenções localizadas e racionais, adaptadas às características de cada área da parcela. Além disso, a interpretação destes mapas pode ainda ser relacionada com fatores

ambientais e práticas de utilização do solo anteriores. Pode ser também utilizada para comparar com mapas futuros e analisar a alteração de uma zona ao longo do tempo.

Resultados e Discussão

3.1 Seleção dos Locais de Amostragem

Foram selecionados três locais para amostragem: arrozal (541,6 ha), tomatal (80,03 ha) e montado (21,91 ha). A carta de solos, Figura 11, evidencia uma distinção clara entre duas zonas principais.

À esquerda, coincidente com a zona do arrozal, predominam os para-aluviosolos: solos hidromórficos com pouca drenagem, sem horizonte eluvial, textura pesada com elevado teor de argila, tanto no horizonte superficial (A1) como nos inferiores. Estes solos caracterizam-se pela receção periódica de sedimentos, apresentando geralmente boa fertilidade e aptidão agrícola. Na mesma zona encontram-se ainda áreas com solos halomórficos (salinos de salinidade moderada, formados de aluviões e com textura pesada), e em menor proporção, surgem ainda solos incipientes (aluviosolos modernos não calcários, de textura pesada), solos jovens de origem aluvional, suscetíveis a encharcamento, com fertilidade elevada, adequados para culturas exigentes em nutrientes e água.

Na Figura 11, à direita, na zona correspondente ao montado, predominam regossolos psamíticos (solos incipientes), de textura arenosa, soltos, com horizontes pouco diferenciados e, por vezes, apenas um delgado horizonte superficial com pequena acumulação de matéria orgânica. Têm aptidão essencialmente florestal, embora possam ser utilizados em agricultura de regadio intensivo.

Na restante área (Figura 11, à direita) encontra-se a zona do tomatal, onde se observam planossolos, caracterizados pela presença de horizontes superficiais leves, sobre camadas mais argilosas compactas e pouco permeáveis, e solos litólicos não húmicos, pouco evoluídos e de arenitos grosseiros (DGADR, 2025).

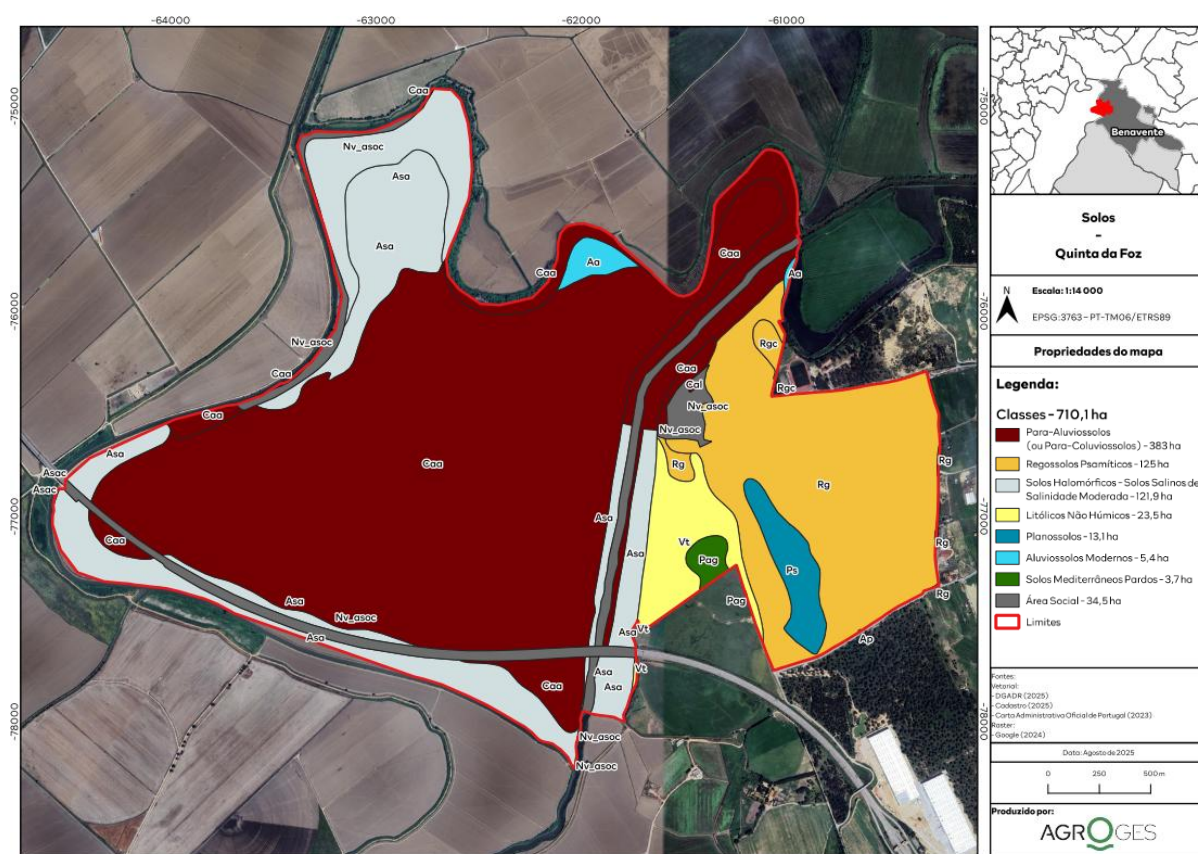


Figura 11 - Distribuição e tipos de solo na Quinta da Foz. Fonte: Agrogres.

Procedeu-se à análise da variação do índice vegetativo NDVI espacialmente e temporalmente, descrita na secção 2.2, para que se escolhessem as localizações das amostras do solo. Isto porque o NDVI reflete a saúde das plantas, influenciada pelo solo no qual as plantas se desenvolvem. A localização resultante das amostras pode ser observada na Figura 12.

Para a escolha dos locais de amostragem no montado, foram utilizadas as imagens de 27 de agosto de 2023, data em que o NDVI atingia valores mínimos após o verão. Nesta altura, no montado, o crescimento da vegetação reflete sobretudo as características do solo.

No tomatal, as imagens de satélite revelaram heterogeneidade, sendo que a 17 de agosto parte da área apresentava ausência quase total de cobertura vegetal, enquanto outras zonas exibiam valores altos de NDVI (Figura 12 a) (ver também Figura 15 b, c e d do Anexo I). Por este motivo, subdividiu-se a parcela em três subzonas. A seleção dos pontos foi feita no último período de declínio do NDVI em cada subzona: final de julho (subzona 1) e 1 de setembro (subzonas 2 e 3). No arrozal, com crescimento a partir de junho e colheita em outubro, o momento escolhido foi o final de setembro, após o ciclo de crescimento.

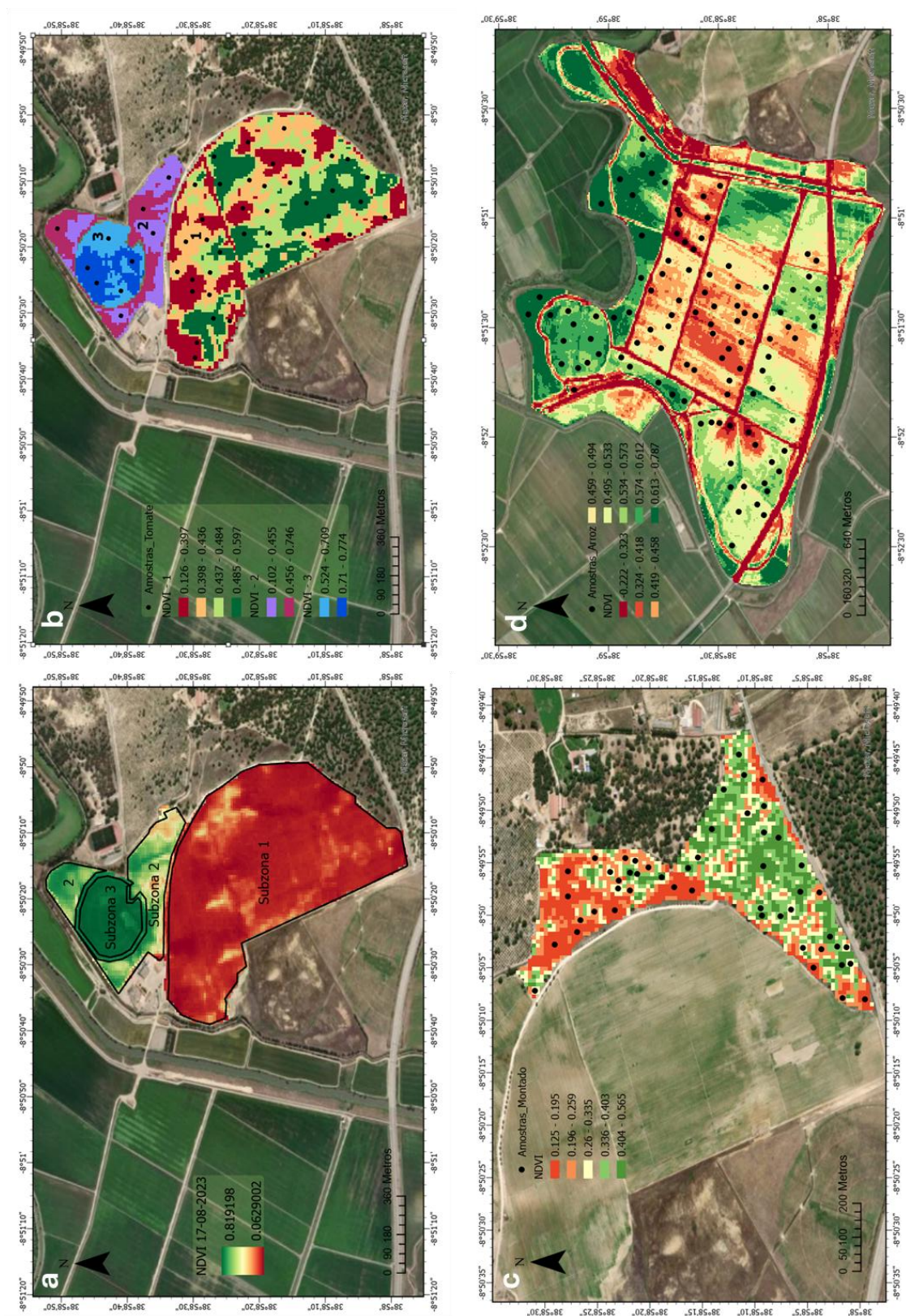


Figura 12 - Determinação dos pontos de amostragem de acordo com as classes de NDVI. a) Divisão da zona do Tomatal em três subzonas; b) Localização das amostras no Tomatal c) Localização das amostras no Montado d) Localização das amostras no Arrozal.

3.2 Criação das Bases de Dados para Modelação

As bases de dados pré-processadas, como descrito na secção 2.6, resultaram em duas tabelas a serem utilizadas nos modelos.

3.2.1 Resultados Estatísticos

A aplicação dos modelos de RF e PLSR permitiu avaliar a capacidade da reflectometria NIR para prever parâmetros de fertilidade do solo. De forma geral, os resultados evidenciam que o modelo RF apresentou melhor desempenho do que o PLSR, pelo que foi esse que se utilizou depois para prever os valores das variáveis químicas para as amostras às quais apenas se fez a análise NIR. Os resultados estatísticos obtidos para os dois modelos podem ser observados na Tabela 3.

Em termos de fiabilidade, ambos os modelos apresentaram valores elevados de R^2 para algumas das propriedades estudadas (Tabela 3). Os modelos RF e PLSR apresentaram desempenhos excelentes para Mg, MO_LOI, Fe, Ni, Ca, Na, Co e K com valores R^2 de calibração e validação cruzada superiores a 0,8, confirmando a elevada correlação destas variáveis com a informação espectral NIR. Para estas variáveis, também se apresentaram valores de RPD de calibração elevados (>2.5). Estes valores estão em conformidade com diversos estudos já realizados nestas propriedades (Viscarra Rossel et al., 2006)

Para as variáveis como B, Mn e MnAI, o desempenho foi eficaz (R^2 entre 0,7 e 0,8; RPD ≈ 2), sugerindo que a técnica pode ser aplicada com confiança moderada, ainda que sujeita a calibrações adicionais. Mesmo assim, para ambos os valores de Mn e MnAI, o desempenho foi melhor do que, por exemplo, os resultados obtidos, respetivamente, 0,55 e 0,64, na revisão feita por Soriano-Disla et al. em 2013. Para o B, os valores foram semelhantes aos obtidos por Towett et al. em 2015, mesmo que estes tenham usado uma metodologia diferente (infravermelho-médio e fluorescência total por raios-X).

Por outro lado, as restantes variáveis (Tabela 3) apresentaram fraca capacidade de previsão, com valores de R^2 de validação cruzada abaixo de 0,7. A literatura aponta que o NIR é mais eficaz na estimativa de parâmetros que possuem bandas de absorção específicas (Stenberg et al., 2010). Por outro lado, a previsão de algumas propriedades do solo torna-se mais difícil devido à ausência de uma “assinatura espectral” direta ou porque a sua expressão no espectro depende de relações indiretas com outras propriedades (Soriano-Disla et al., 2013). A ausência de bandas de absorção diretas para propriedades como o pH, por exemplo, faz com

que a sua previsão seja baseada em correlações com a matéria orgânica ou o teor de argila, por este motivo e para manter a robustez da calibração, existe a necessidade de calibrar os modelos de forma específica para diferentes contextos.

O carbono microbiano (e as outras variáveis relativas à análise com analisador TOC) não revelaram ser promissores na detecção através do NIR. Apesar dos resultados na fase de calibração até serem bons no caso do modelo com RF (Tabela 3), os valores da validação não foram correspondidos. O estudo realizado por Heinze et al. (2013), revelou que o desempenho depende do tipo de solo e da correlação com a matéria orgânica. Este ensaio difere no método de quantificação de carbono microbiano e no intervalo espectral (inclui a zona do visível), o que pode significar que testar diferentes métodos pode ser crucial para a obtenção de valores de previsão confiáveis, assim como o método de análise e a forma química analisada.

Em 2023, Jia et al. testaram vários modelos (PLSR, RF, SVM, CNN) para os nutrientes N, P e K. Os resultados atribuíram melhor desempenho ao modelo SVM.

Este não foi utilizado na presente metodologia, no entanto, comparando com os resultados dos mesmos modelos produzidos nesta dissertação, a previsão para o P revelou-se melhor com PLSR e ligeiramente melhor com RF, já o K apresentou previsões muito mais promissoras nos dois modelos avaliados no presente estudo; o N mostrou resultados mais fracos relativamente ao obtidos por Jia et al. (2023).

O exemplo do estudo referido anteriormente, reforça a necessidade de testar modelos não lineares mais avançados, como *Artificial Neural Network* (ANN) (Afandi et al., 2016) ou *Support Vector Machines* (SVMs), que têm demonstrado melhor desempenho em problemas de elevada complexidade.

A análise conjunta dos resultados demonstra que a espectroscopia NIR, aliada a modelos de *machine learning*, constitui uma ferramenta promissora para a caracterização rápida da fertilidade dos solos, permitindo reduzir o recurso a análises laboratoriais tradicionais. Contudo, o sucesso da previsão depende do tipo de nutriente avaliado, da qualidade da calibração e da representatividade da base de dados utilizada. Este trabalho mostra assim que algumas variáveis revelaram ser promissoras (Mg, MO_LOI, Fe, Ni, Ca, Na, Co e K), enquanto outras (Cu, S, CE, Zn, pH, MO, variáveis medidas através do analisador de TOC, percentagem de azoto e carbono e a razão entre estas duas) requerem ainda investigação adicional e aumento da diversidade de amostras para atingir níveis de fiabilidade aceitáveis.

Variável	Modelo																
	RF						PLSR										
	RMSE	Cal	R ²	R ²	RPD	CV	RMSE	Cal	R ²	R ²	RPD	CV	RMSE	Cal	RPD	CV	NC
Mg	53.85	139.825	0.991	0.939	10.571	4.071	127.587	147.217	0.949	0.932	4.462	3.867	5				
MO_LOI	0.406	1.019	0.984	0.897	7.863	3.131	0.969	1.163	0.907	0.866	3.291	2.742	5				
Fe	39.279	111.127	0.987	0.896	8.81	3.114	61.387	118.662	0.968	0.881	5.637	2.916	8				
Ni	0.156	0.406	0.983	0.887	7.773	2.99	0.254	0.456	0.956	0.858	4.774	2.667	8				
Ca	210.462	534.074	0.979	0.866	6.974	2.748	483.488	581.598	0.89	0.841	3.036	2.524	5				
Na	12.645	37.163	0.983	0.851	7.648	2.602	30.305	35.64	0.901	0.863	3.191	2.713	5				
Co	0.148	0.419	0.981	0.847	7.272	2.569	0.398	0.475	0.862	0.803	2.705	2.263	5				
K	21.596	60.273	0.98	0.842	7.055	2.528	53.284	61.681	0.876	0.834	2.859	2.47	5				
B	0.121	0.314	0.961	0.736	5.077	1.956	0.29	0.311	0.776	0.742	2.124	1.977	4				
Mn	17.856	50.295	0.965	0.725	5.397	1.916	43.282	46.132	0.796	0.769	2.227	2.089	4				
MnAl	68.529	195.966	0.964	0.71	5.333	1.865	165.294	175.986	0.793	0.766	2.211	2.077	4				
Cu	0.449	1.163	0.927	0.511	3.722	1.438	0.401	1.179	0.942	0.498	4.173	1.419	12				
S	22.088	58.129	0.927	0.496	3.727	1.416	40.726	60.018	0.753	0.463	2.022	1.372	7				
P	42.446	122.491	0.931	0.425	3.826	1.326	95.607	114.758	0.65	0.495	1.699	1.415	5				
CE	0.031	0.086	0.916	0.355	3.467	1.252	0.079	0.079	0.454	0.448	1.36	1.36	2				
Zn	1.329	3.43	0.883	0.219	2.934	1.137	2.672	3.16	0.526	0.337	1.459	1.234	5				
pH	0.303	0.815	0.879	0.126	2.889	1.075	0.532	0.72	0.627	0.316	1.646	1.215	6				
Azoto_T	0.024	0.063	0.869	0.077	2.779	1.046	0.056	0.059	0.263	0.176	1.171	1.107	2				
TOC	0.721	2.152	0.889	0.011	3.018	1.011	1.648	1.952	0.42	0.186	1.32	1.114	5				
MO	0.52	1.405	0.86	-0.027	2.681	0.992	0.991	1.238	0.489	0.202	1.406	1.125	5				
Carbono_T	0.311	0.826	0.85	-0.055	2.599	0.979	0.671	0.767	0.304	0.09	1.205	1.054	4				
Razao_CN	0.812	2.254	0.848	-0.167	2.581	0.93	1.896	2.076	0.174	0.009	1.106	1.01	4				
IC	0.169	0.459	0.837	-0.201	2.489	2.489	0.395	0.41	0.111	0.044	1.066	1.028	2				
C_microb	2.666	8.089	0.867	-0.225	2.757	0.909	7.233	7.459	0.021	-0.042	1.016	0.985	1				
F_TC	2.996	9.226	0.852	-0.407	2.611	0.848	7.774	8.146	0.001	-0.097	1.006	0.96	1				

Tabela 3 - Resultados estatísticos para o modelo *Random Forest* e PLSR. Legenda: RMSE – Raiz do erro quadrático médio; R² – Coeficiente de determinação; RPD – Razão de desempenho em relação ao desvio padrão; Cal – Resultante da calibração; CV – Resultante da validação cruzada; NC – Número de componentes;

Foram elaborados gráficos de previsão vs. valor real para as propriedades do solo com melhor desempenho (Figura 13). É possível observar que, para certas variáveis (Mg, MO_LOI, Fe, Ni, Ca, Na, Co, K e B), os pontos se distribuem em duas nuvens distintas. Essa separação ocorre devido à inclusão de amostras de dois tipos de solo muito diferentes (arrozal comparado com o montado e tomatal), o que cria dois “grupos” com características químicas e espectrais desiguais.

O modelo RF consegue diferenciar bem estes grupos, mas pode não conseguir que o modelo descreva bem a relação entre espectro e propriedade dentro de cada grupo.

Recomenda-se, se se pretender continuar a explorar o potencial do NIR para avaliar propriedades do solo, aumentar o número de amostras. Podia ser interessante incluir solos com propriedades intermédias, de forma a preencher o espaço espectral entre os dois grupos (FAO, 2024).

Na eventualidade de ser possível realizar um maior volume de amostras, pode-se considerar calibrar modelos separados por tipo de solo.

Caso se pretenda utilizar um modelo numa região diferente ou atualizar a área no futuro, é recomendado acrescentar cerca de 20 % de amostras do novo conjunto ao conjunto de calibração (Purushothaman et al., 2024).

Idealmente, as amostras utilizadas para fazer previsão devem ser congêneres a futuras amostras incógnitas. Quando as amostras de calibração são demasiado restritas, pode ser mais difícil ter boas previsões para amostras que, por características específicas do local, se afastam dos valores médios (Soriano-Disla et al., 2013).

Um modelo que inclua uma diversidade maior de solos permite avaliar propriedades em diferentes regiões; por outro lado, um modelo construído para uma herdade específica pode alcançar elevada precisão local, mas provavelmente terá um desempenho fraco quando aplicado a solos de regiões muito diferentes.

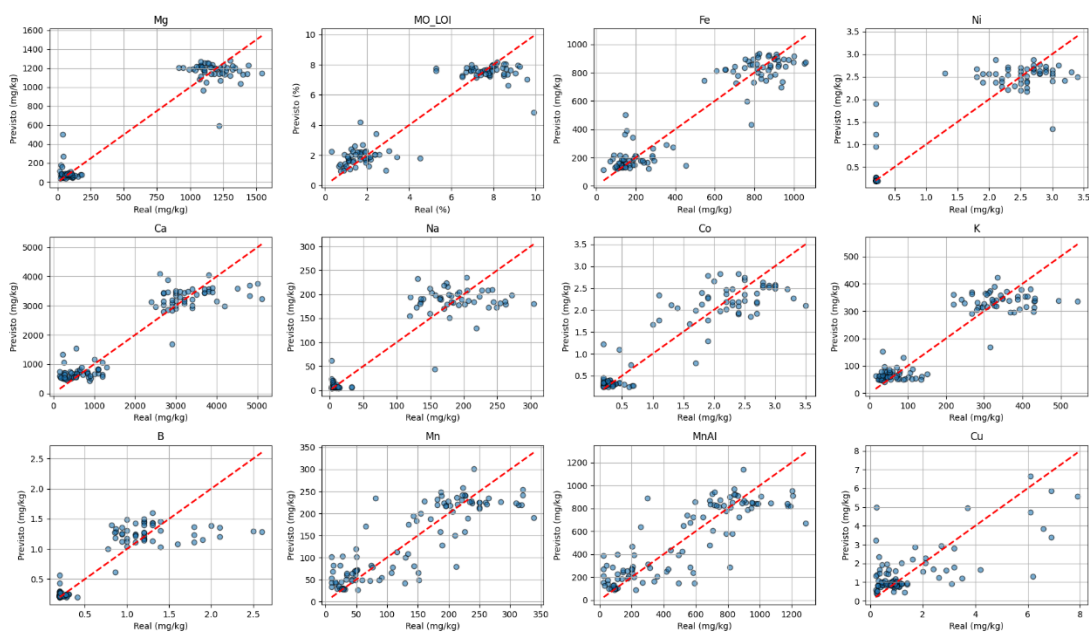


Figura 13 - Valores Reais vs. Previstos para as propriedades químicas da validação cruzada.

3.3 Mapas de Previsão

Na Figura 14 observam-se os mapas de previsão obtidos para as cinco propriedades com melhores métricas de calibração (Mg, Fe, MO_LOI, Ni, Na) nas zonas do montado, arrozal e tomatal.

A construção destes mapas realizou-se utilizando sistemas de informação geográfica (SIG) interpolando, através da ferramenta IDW, as previsões obtidas, permitindo representar a variação espacial das propriedades de forma contínua. Desta forma, é possível visualizar claramente as áreas com maior e menor concentração de cada elemento, avaliar a distribuição de fertilidade e definir zonas de gestão. Estes mapas podem ser ferramentas valiosas para fazer um melhor planeamento agrícola e gestão ambiental, podendo ser utilizados, por exemplo, para fazer aplicação variável de fertilizantes. Permitem identificar zonas com limitações ou potencialidades de produção agrícola e podem ser usados como guia para amostragem futura e para alertar para riscos de deficiência ou toxicidade de elementos (Alomar et al., 2022).

No entanto, é preciso realçar que um mapa digital de propriedades do solo é sempre uma generalização da realidade, muito dependente da precisão e qualidade de todos os passos necessários até à sua obtenção. Seria recomendável fazer validações quer através dos valores de NDVI quer em campo, verificar se zonas diferentes no mapa correspondem também a diferenças na variação do índice vegetativo e a variações reais na resposta das culturas a nível de produção, qualidade e crescimento das plantas, respetivamente. Para poder assim assegurar uma ferramenta confiável de apoio à decisão para o agricultor (Abdu et al., 2023).

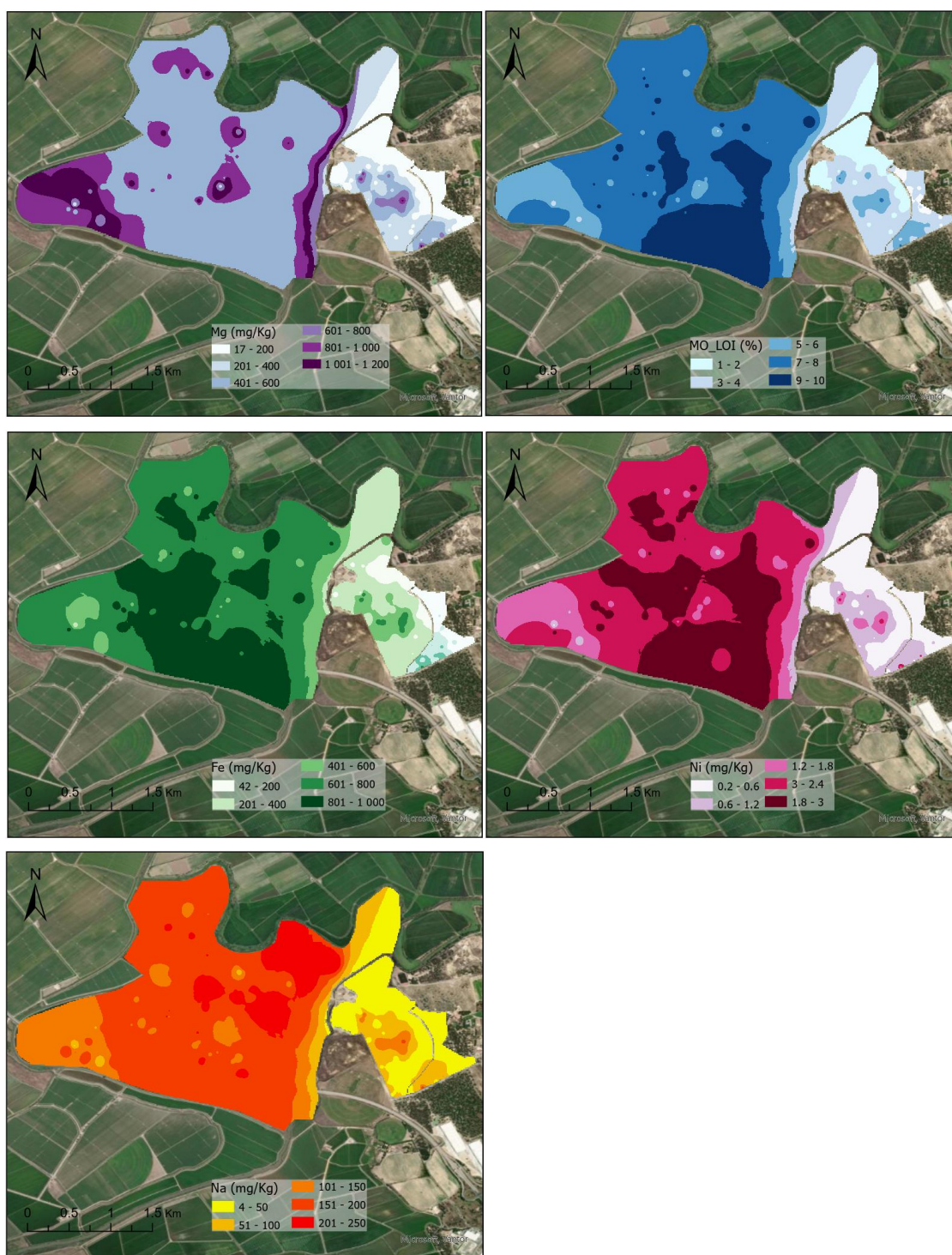


Figura 14 - Mapas com valores previstos de Na, Ni, MO_LOI, Fe e Mg para as zonas do tomatal, arrozal e montado.

Conclusão

A problemática central que fundamenta esta dissertação pretende responder à necessidade de aumentar a produtividade agrícola de maneira sustentável, conservando os interesses dos agricultores e o meio ambiente. Possibilitar a produção de alimentos para todos hoje e sempre. A agricultura de precisão alinhou-se à espectroscopia nesta dissertação para tentar, mesmo que infimamente, contribuir para a resolução deste desafio. Utilizou-se o NIR como ferramenta rápida, não destrutiva e de baixo custo, para estimar várias propriedades do solo.

O objetivo desta dissertação foi o estudo da possibilidade de utilização do espectro de reflectância dos solos na zona do NIR para a medição e contabilização de propriedades do solo. Propriedades essas que, de outra forma, apenas podem ser contabilizadas com processos demorados e custosos. As tarefas propostas e realizadas consistiram em recolher amostras, analisá-las e criar um modelo que as utilizasse para prever propriedades desconhecidas de solos. Os resultados obtidos permitem explorar o potencial da utilização do NIR para contabilizar algumas propriedades do solo, mas também revelam limitações no tipo de propriedades que é possível prever.

Os modelos apresentaram resultados promissores ($R^2 > 0,85$; $RPD > 2,5$) para prever vários nutrientes (magnésio, ferro, níquel, cálcio e sódio) e o carbono orgânico calculado através do método por perda de ignição. Mostraram prever de forma eficaz ($R^2 > 0,70$; $RPD \approx 2$) o cobalto, potássio, boro, manganês e manganês AI. De forma menos precisa, mostraram pouca ou praticamente nenhuma relação com o cobalto, enxofre, fósforo, condutividade elétrica, zinco, pH, percentagem de azoto total e de carbono, relação carbono-azoto, matéria orgânica calculada nas análises de fertilidade, carbono inorgânico, carbono microbiano, carbono orgânico total medido pela análise TOC e carbono total das amostras fumigadas medido pela análise TOC. O que revela que nem todas as variáveis químicas são sensíveis ao NIR ou que os dados disponíveis foram insuficientes.

A construção de mapas de fertilidade mostrou, para as propriedades com melhor desempenho, o potencial de utilizar o NIR para desenvolver mapas de previsão onde se visualiza

a distribuição espacial entre as zonas (arrozal, tomatal e montado) e a possibilidade de atuar conforme esta informação, fornecendo uma ferramenta prática para os agricultores.

A utilização de imagens NDVI para definir uma amostragem inteligente permitiu que o número limitado de amostras que era possível fazer fosse bem aproveitado e que existisse uma representação de todas as áreas e da sua variação. Contribuiu também para um melhor modelo, o pré-processamento específico às amostras de solo e curvas espectrais.

Esta dissertação contribui com um fluxo repetível aplicável a outro local, no entanto, existem limitações que influenciam o sucesso da previsão e que devem ser reconhecidas: o número de amostras que influencia a capacidade de calibrar e construir um modelo robusto, a variabilidade natural dos solos, a dependência de protocolos laboratoriais repetitivos e demorados e a cobertura geográfica limitada à Quinta da Foz e aos seus solos.

Quando ultrapassadas estas limitações, os mapas de previsão conseguem contribuir para aumentar a eficiência do uso de fertilizantes e reduzir desperdícios, aumentando a produção com menor impacto ambiental e reduzindo custos.

Também a monitorização e medição do carbono orgânico através do NIR, que garante precisão das medições, podem fornecer dados que podem ser utilizados na contabilização de sequestro de carbono.

Idealmente, no futuro, seria recomendado fazer amostragens mais extensas e diversas, abrangendo diferentes tipos de solos e épocas do ano, incluir outras bandas do espectro (visível e MIR) e testar algoritmos mais complexos. Seria também útil desenvolver bibliotecas espectrais regionais (FAO, 2024).

5 Bibliografia

Abbott, L. K., & Murphy, D. V. (2007). What is soil biological fertility? In L. K. Abbott & D. V. Murphy (Eds.), *Soil biological fertility: A key to sustainable land use in agriculture* (pp. 1–15). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6619-1_1

Abdu, A., Laekemariam, F., Gidago, G., Kebede, A., & Getaneh, L. (2023). Variability analysis of soil properties, mapping, and crop test responses in Southern Ethiopia. *Heliyon*, 9(3), e14013. <https://doi.org/10.1016/j.heliyon.2023.e14013>

Afandi, S. D., Herdiyeni, Y., Prasetyo, L. B., Hasbi, W., Arai, K., & Okumura, H. (2016). Nitrogen content estimation of rice crop based on near infrared (NIR) reflectance using artificial neural network (ANN). *Procedia Environmental Sciences*, 33, 63–69. <https://doi.org/10.1016/j.proenv.2016.03.057>

Aggelopoulou, K. D., Pateras, D., & Fountas, S. (2011). Soil spatial variability and site-specific fertilization maps in an apple orchard. *Precision Agriculture*, 12(1), 118–129. <https://doi.org/10.1007/s11119-010-9161-x>

Alomar, S., Mireei, S. A., Hemmat, A., Masoumi, A. A., & Khademi, H. (2022). Prediction and variability mapping of some physicochemical characteristics of calcareous topsoil in an arid region using Vis-SWNIR and NIR spectroscopy. *Scientific Reports*, 12(1), 8435. <https://doi.org/10.1038/s41598-022-12276-4>

Badgley, G., Chay, F., Chegwidan, O. S., Hamman, J. J., Freeman, J., & Cullenward, D. (2022). California's forest carbon offsets buffer pool is severely undercapitalized. *Frontiers in Forests and Global Change*, 5, 930426. <https://doi.org/10.3389/ffgc.2022.930426>

Bellon-Maurel, V., & McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils—Critical review and research perspectives. *Soil Biology and Biochemistry*, 43(7), 1398–1410. <https://doi.org/10.1016/j.soilbio.2011.02.019>

Cai, Y., Luo, F., Guo, Y., Guo, F., Shi, W., & Yang, S. (2023). Near-infrared light driven ZnIn₂S₄-based photocatalysts for environmental and energy applications: Progress and perspectives. *Molecules*, 28(5), 2142. <https://doi.org/10.3390/molecules28052142>

Chang, C. W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R., Jr. (2001). Near-infrared reflectance spectroscopy—Principal components regression analysis of soil properties. *Soil Science Society of America Journal*, 65(2), 480–490. <https://doi.org/10.2136/sssaj2001.652480x>

Chatterjee, A., Lal, R., Wielopolski, L., Martin, M. Z., & Ebinger, M. H. (2009). Evaluation of different soil carbon determination methods. *Critical Reviews in Plant Sciences*, 28(3), 164–178. <https://doi.org/10.1080/07352680902776556>

Chaudhry, H., Vasava, H. B., Chen, S., Saurette, D., Beri, A., Gillespie, A., & Biswas, A. (2024). Evaluating the soil quality index using three methods to assess soil fertility. *Sensors*, 24(3), 864. <https://doi.org/10.3390/s240308>

Christy, C. D., Drummond, P., & Laird, D. A. (2003). An on-the-go spectral reflectance sensor for soil (Paper No. 031044). In *Proceedings of the 2003 ASAE Annual Meeting*. American Society of Agricultural and Biological Engineers. <https://doi.org/10.13031/2013.13716>

Coglianese, C., & Giles, C. (2025). Third-party auditing cannot guarantee carbon offset credibility (U of Penn Law School, Public Law Research Paper No. 25-28). SSRN. <https://doi.org/10.2139/ssrn.5345783>

Conforti, M., Castrignanò, A., Robustelli, G., Scarciglia, F., Stelluti, M., & Buttafuoco, G. (2015). Laboratory-based Vis–NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content. *Catena*, 124, 60–67. <https://doi.org/10.1016/j.catena.2014.09.006>

Cozzolino, D., & Morón, A. (2006). Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *Soil and Tillage Research*, 85(1), 78–85. <https://doi.org/10.1016/j.still.2004.12.006>

Cravero, A., & Sepúlveda, S. (2021). Use and adaptations of machine learning in big data—Applications in real cases in agriculture. *Electronics*, 10(5), 552. <https://doi.org/10.3390/electronics10050552>

Dalal, N., Ofano, R., Ruggiero, L., Caporale, A. G., & Adamo, P. (2024). What the fish? Tracing the geographical origin of fish using NIR spectroscopy. *Current Research in Food Science*, 7, 100789. <https://doi.org/10.1016/j.crfs.2024.100789>

David, M. B. (1988). Use of loss-on-ignition to assess soil organic carbon in forest soils. *Communications in Soil Science and Plant Analysis*, 19(13), 1593–1599. <https://doi.org/10.1080/00103628809368037>

Dignac, M. F., Derrien, D., Barré, P., Barot, S., Cécillon, L., Chenu, C., Chevallier, T., Freschet, G. T., Garnier, P., Guenet, B., & Hedde, M. (2017). Increasing soil carbon storage: Mechanisms, effects of agricultural practices, and proxies. *Agronomy for Sustainable Development*, 37(1), 14. <https://doi.org/10.1007/s13593-017-0421-2>

Du, C., & Zhou, J. (2009). Evaluation of soil fertility using infrared spectroscopy – A review. In E. Lichtfouse (Ed.), *Climate change, intercropping, pest control and beneficial microorganisms* (Sustainable agriculture reviews, Vol. 2). Springer. https://doi.org/10.1007/978-90-481-2716-0_16

Faithfull, N. T. (2002). *Methods in agricultural chemical analysis* (pp. 57–104). CABI Publishing.

Fidêncio, P. H., Poppi, R. J., de Andrade, J. C., & Cantarella, H. (2002). Determination of organic matter in soil using near-infrared spectroscopy and partial least squares regression. *Communications in Soil Science and Plant Analysis*, 33(9–10), 1607–1615. <https://doi.org/10.1081/CSS-120004299>

Food and Agriculture Organization of the United Nations. (2020a). *Acid soils*. FAO Soils Portal. <https://www.fao.org/soils-portal/soil-management/management-of-some-problem-soils/acid-soils/en/>

Food and Agriculture Organization of the United Nations. (2020b). *Soil testing methods manual – Soil Doctors Global Programme: A farmer-to-farmer training programme*. FAO. <https://doi.org/10.4060/ca2796en>

Food and Agriculture Organization of the United Nations. (2024). *A primer on soil analysis using visible and near-infrared (Vis-NIR) and mid-infrared (MIR) spectroscopy* (Revision). FAO. <https://doi.org/10.4060/cb9005en>

Hao, X., Chen, Z., Yi, S., & Liu, J. (2023). Application of improved stacking ensemble learning in NIR spectral modeling of corn seed germination rate. *Chemometrics and Intelligent Laboratory Systems*, 243, 105020. <https://doi.org/10.1016/j.chemolab.2023.105020>

Heiri, O., Lotter, A. F., & Lemcke, G. (2001). Loss on ignition as a method for estimating organic and carbonate content in sediments: Reproducibility and comparability of results. *Journal of Paleolimnology*, 25(1), 101–110. <https://doi.org/10.1023/A:1008119611481>

Jena, J., Misra, S. R., & Tripathi, K. P. (2019). Normalized difference vegetation index (NDVI) and its role in agriculture. *Agriculture and Food: E-Newsletter*, 1(12), 387–389.

Jia, X., Fang, Y., Hu, B., Yu, B., & Zhou, Y. (2023). Development of soil fertility index using machine learning and visible–near-infrared spectroscopy. *Land*, 12(12), 2155. <https://doi.org/10.3390/land12122155>

Jin, T., & Han, X. (2024). Robotic arms in precision agriculture: A comprehensive review of the technologies, applications, challenges, and future prospects. *Computers and Electronics in Agriculture*, 221, 108938. <https://doi.org/10.1016/j.compag.2024.108938>

Karunathilake, E. M. B. M., Le, A. T., Heo, S., Chung, Y. S., & Mansoor, S. (2023). The path to smart farming: Innovations and opportunities in precision agriculture. *Agriculture*, 13(8), 1593. <https://doi.org/10.3390/agriculture13081593>

Khan, H., Farooque, A. A., Acharya, B., Abbas, F., Esau, T. J., & Zaman, Q. U. (2020). Delineation of management zones for site-specific information about soil fertility characteristics through proximal sensing of potato fields. *Agronomy*, 10(12), 1854. <https://doi.org/10.3390/agronomy10121854>

Kim, H. J., Ryu, J., Kim, G., & Jo, C. (2025). Enhancement of non-destructive chicken freshness prediction using Vis/NIR spectroscopy through wavelength selection and data augmentation. *LWT*, 221, 117602. <https://doi.org/10.1016/j.lwt.2025.117602>

Loganathan, P. (1987). *Soil quality considerations in the selection of sites for aquaculture* (Chapter 6). Food and Agriculture Organization of the United Nations. <https://www.fao.org/4/ac172e/AC172E06.htm>

Lourenço, J. M. (2021). *The NOVAthesis LaTeX template user's manual*. NOVA University Lisbon. https://github.com/joaomlourenco/novathesis_word/raw/master/novathesis_word-FINAL-EN.pdf

Majumder, D., Akhter, J., Mandal, A., et al. (2021). Precision input management for minimizing and recycling of agricultural waste. In R. Jha, V. K. Singh, A. Parihar, & A. K. Gupta (Eds.), *Input use efficiency for food and environmental security* (pp. 567–603). Springer. https://doi.org/10.1007/978-981-16-5199-1_19

McCarty, G. W., Reeves, J. B., Reeves, V. B., Follett, R. F., & Kimble, J. M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal*, 66(2), 640–646. <https://doi.org/10.2136/sssaj2002.6400a>

Mehlich, A. (1984). Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Communications in Soil Science and Plant Analysis*, 15(12), 1409–1416. <https://doi.org/10.1080/00103628409367568>

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., & O'Rourke, S. (2017). Soil carbon 4 per mille. *Geoderma*, 292, 59–86. <https://doi.org/10.1016/j.geoderma.2017.01.002>

Mohkam-Singh, Y., & Nunes, M. R. (2025). A 30-year overview of soil fertility, soil quality, and soil health research in Florida. *Ecological Indicators*, 176, 113660. <https://doi.org/10.1016/j.ecolind.2025.113660>

Montanarella, L., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yemefack, M., Aulakh, M., Yagi, K., Hong, S.-Y., Vijarnsorn, P., Zhang, G.-L., Arrouays, D., Black, H., Krasilnikov, P., Sobocká, J., Alegre, J., Henríquez, C., Mendonça Santos, M., Taboada, M., & McKenzie, N. (2015). *The status of the world's soil resources (Main report)*. Food and Agriculture Organization of the United Nations. <https://www.fao.org/3/i5199e/i5199e.pdf>

Morgan, C. L., Waiser, T. H., Brown, D. J., & Hallmark, C. T. (2009). Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. *Geoderma*, 151(3–4), 249–256. <https://doi.org/10.1016/j.geoderma.2009.04.010>

Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., & Ramon, H. (2007). On-line measurement of some selected soil properties using a VIS-NIR sensor. *Soil & Tillage Research*, 93(1), 13–27. <https://doi.org/10.1016/j.still.2006.03.009>

Munnaf, M. A., & Mouazen, A. M. (2021). Development of a soil fertility index using on-line Vis-NIR spectroscopy. *Computers and Electronics in Agriculture*, 188, 106341. <https://doi.org/10.1016/j.compag.2021.106341>

Nanni, M. R., & Demattê, J. A. M. (2006). Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Science Society of America Journal*, 70(2), 393–407. <https://doi.org/10.2136/sssaj2003.0285>

Nayak, A. K., Gangwar, B., Shukla, A. K., Mazumdar, S. P., Kumar, A., Raja, R., Kumar, A., Kumar, V., Rai, P. K., & Mohan, U. (2012). Long-term effect of different integrated nutrient management on soil organic carbon and its fractions and sustainability of the rice–wheat system in the Indo-Gangetic Plains of India. *Field Crops Research*, 127, 129–139. <https://doi.org/10.1016/j.fcr.2011.11.011>

Nayak, A. K., Rahman, M. M., Naidu, R., Dhal, B., Swain, C. K., Nayak, A. D., Tripathi, R., Shahid, M., Islam, M. R., & Pathak, H. (2019). Current and emerging methodologies for estimating carbon sequestration in agricultural soils: A review. *Science of the Total Environment*, 665, 890–912. <https://doi.org/10.1016/j.scitotenv.2019.02.125>

Nelson, D. W., & Sommers, L. E. (1996). Total carbon, organic carbon, and organic matter. In D. L. Sparks (Ed.), *Methods of soil analysis: Part 3 – Chemical methods*. Soil Science Society of America; American Society of Agronomy. <https://doi.org/10.2136/sssabookser5.3.c34>

Neves, A. (2017). *Técnicas de análise laboratorial para solos, matéria vegetal e corretivos orgânicos* [Dissertação de mestrado, Universidade do Minho]. Universidade do Minho.

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99–118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>

Olsen, S. R., Cole, C. V., & Watanabe, F. S. (1954). *Estimation of available phosphorus in soils by extraction with sodium bicarbonate* (USDA Circular No. 939). U.S. Government Printing Office.

Parvej, M. R., Brandt, D., Myers, R., Nelson, K., Singh, G., & Reinbott, T. (2025, July). *Soil organic carbon: A foundational indicator of soil health*. University of Missouri Extension.

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., & Smith, P. (2016). Climate-smart soils. *Nature*, 532(7597), 49–57. <https://doi.org/10.1038/nature17174>

Pirie, A., Singh, B., & Islam, K. (2005). Ultraviolet, visible, near-infrared and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Australian Journal of Soil Research*, 43(6), 713–721. <https://doi.org/10.1071/SR04182>

Post, W. M., Izaurrealde, R. C., Mann, L. K., & Bliss, N. (2001). Monitoring and verifying changes of organic carbon in soil. *Climatic Change*, 51(1), 73–99. <https://doi.org/10.1023/A:1017514802028>

Purushothaman, N. K., Garg, K. K., Venkataradha, A., Anantha, K., Singh, R., Jat, M., & Das, B. S. (2024). Applicability of calibrated diffuse reflectance spectroscopy models across spatial and temporal boundaries. *Geoderma*, 449, 117012. <https://doi.org/10.1016/j.geoderma.2024.117012>

Recena, R., Fernández-Cabanás, V. M., & Delgado, A. (2019). Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. *Geoderma*, 337, 368–374. <https://doi.org/10.1016/j.geoderma.2018.09.049>

Robertson, S. (2011). Direct estimation of organic matter by loss on ignition: Methods. Simon Fraser University, Geography and Soil Science Lab.

Schulte, E. E., Kauffmann, C., & Peter, J. B. (1991). The influence of sample size and heating time on soil weight loss-on-ignition. *Communications in Soil Science and Plant Analysis*, 22(1–2), 159–168. <https://doi.org/10.1080/00103629109368402>

Schwartz, V. (1995). Fractionated combustion analysis of carbon in forest soils—New possibilities for the analysis and characterization of different soils. *Fresenius Journal of Analytical Chemistry*, 351(6), 629–631. <https://doi.org/10.1007/BF00323339>

Segnini, A., Santos, L. M., Silva, W. T. L., Martin-Neto, L., Borato, C. E., Melo, W. J., & Bolonhezi, D. (2008). Estudo comparativo de métodos para a determinação da concentração de carbono em solos com altos teores de Fe (Latossolos). *Química Nova*, 31(1), 94–97. <https://doi.org/10.1590/S0100-40422008000100020>

Shepherd, K. D., & Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal*, 66(3), 988–998. <https://doi.org/10.2136/sssaj2002.9880>

Shi, T., Cui, L., Wang, J., Fei, T., Chen, Y., & Wu, G. (2013). Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. *Plant and Soil*, 366(1–2), 363–375. <https://doi.org/10.1007/s11104-012-1436-8>

Shonk, J. L., Gaultney, L. D., Schulze, D. G., & Van Scoyoc, G. E. (1991). Spectroscopic sensing of soil organic matter content. *Transactions of the ASAE*, 34(5), 1978–1984. <https://doi.org/10.13031/2013.31826>

Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., Macdonald, L. M., & McLaughlin, M. J. (2013). The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*, 49(2), 139–186. <https://doi.org/10.1080/05704928.2013.811081>

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107, 163–215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7)

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLoS ONE*, 8(6), e66409. <https://doi.org/10.1371/journal.pone.0066409>

Sudduth, K. A., & Hummel, J. W. (1996). Geographic operating range evaluation of a NIR soil sensor. *Transactions of the ASAE*, 39(5), 1599–1604. <https://doi.org/10.13031/2013.27674>

Towett, E. K., Shepherd, K. D., Sila, A., Aynekulu, E., & Cadisch, G. (2015). Mid-infrared and total X-ray fluorescence spectroscopy complementarity for assessment of soil properties. *Soil Science Society of America Journal*, 79(5), 1375–1385. <https://doi.org/10.2136/sssaj2014.11.0458>

Trouwloon, D., Streck, C., Chagas, T., & Martinus, G. (2023). Understanding the use of carbon credits by companies: A review of the defining elements of corporate climate claims. *Global Challenges*, 7(1), 2200158. <https://doi.org/10.1002/gch2.202200158>

Verma, S. K., & Deb, M. K. (2007). Direct and rapid determination of sulphate in environmental samples with diffuse reflectance Fourier transform infrared spectroscopy using KBr substrate. *Talanta*, 71(3), 1546–1552. <https://doi.org/10.1016/j.talanta.2006.07.056>

Vestergaard, R.-J., Vasava, H. B., Aspinall, D., Chen, S., Gillespie, A., Adamchuk, V., & Biswas, A. (2021). Evaluation of optimized preprocessing and modeling algorithms for prediction of soil properties using VIS-NIR spectroscopy. *Sensors*, 21(20), 6745. <https://doi.org/10.3390/s21206745>

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131(1–2), 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>

Vohland, M., Besold, J., Hill, J., & Fründ, H. C. (2011). Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma*, 166(1), 198–205. <https://doi.org/10.1016/j.geoderma.2011.08.001>

Walkley, A., & Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, 37(1), 29–38. <https://doi.org/10.1097/00010694-193401000-00003>

Williams, P., & Norris, K. (2001). *Near-infrared technology in the agricultural and food industries* (2nd ed.). American Association of Cereal Chemists.

Wu, H., Cui, H., Fu, C., Li, R., Qi, F., Liu, Z., Yang, G., Xiao, K., & Qiao, M. (2024). Unveiling the crucial role of soil microorganisms in carbon cycling: A review. *Science of the Total Environment*, 909, 168627. <https://doi.org/10.1016/j.scitotenv.2023.168627>

Yang, M., Mouazen, A., Zhao, X., & Guo, X. (2020). Assessment of a soil fertility index using visible and near-infrared spectroscopy in the rice paddy region of southern China. *European Journal of Soil Science*, 71(4), 615–626. <https://doi.org/10.1111/ejss.12907>

Yasrebi, J., Saffari, M., Fathi, H., Karimian, N., Emadi, M., & Baghernejad, M. (2008). Spatial variability of soil fertility properties for precision agriculture in southern Iran. *Journal of Applied Sciences*, 8(9), 1642–1650. <https://doi.org/10.3923/jas.2008.1642.1650>

Zhang, Y., Li, M., Zheng, L., Zhao, Y., & Pei, X. (2016). Soil nitrogen content forecasting based on real-time NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 29–36. <https://doi.org/10.1016/j.compag.2016.03.016>

Zhang, L., Yang, X. M., Drury, C. F., Chantigny, M., Gregorich, E., Miller, J., Bittman, S., Reynolds, W., & Yang, J. (2017). Infrared spectroscopy prediction of organic carbon and total nitrogen in soil and particulate organic matter from diverse Canadian agricultural regions. *Canadian Journal of Soil Science*, 98(1), 77–90. <https://doi.org/10.1139/cjss-2017-0070>

Referências de Páginas Web:

Iowa State University Extension and Outreach. *Measuring soil organic carbon: A crucial Iowa resource*. Iowa State University Extension and Outreach. <https://store.extension.iastate.edu/product/16546> (Acedido a 17 de novembro de 2024).

World Bank. (n.d.). *Agriculture, forestry, and fishing, value added (% of GDP)*. World Bank. <https://data.worldbank.org/indicator/AG.LND.AGRI.ZS> (Acedido a 12 de novembro de 2024). <https://smartag.agroinsider.com/> (Acedido a 22 de novembro de 2024).

Direção-Geral de Agricultura e Desenvolvimento Rural. *Nota explicativa da Carta dos Solos de Portugal e da Carta de Capacidade de Uso do Solo*. DGADR. <https://www.dgadr.gov.pt/pt/nota-explicativa> (Acedido a 14 de agosto de 2025).

A2 Análises Químicas. *Métodos de análise de solos*. <https://www.a2analisesquimicas.com/> (Acedido a 17 de agosto de 2025).

NIR helping the search for life in space. (2008). *NIR News*, 19(3), 3–15. <https://doi.org/10.1255/nirn.1075> (Acedido a 18 de agosto de 2025).

International Society of Precision Agriculture. (2024). *Precision agriculture definition*. International Society of Precision Agriculture. <https://www.ispag.org/resources/definition> (Acedido a 17 de junho de 2025).

6 Anexos

6.1 Anexo I - Gráficos Temporais de NDVI

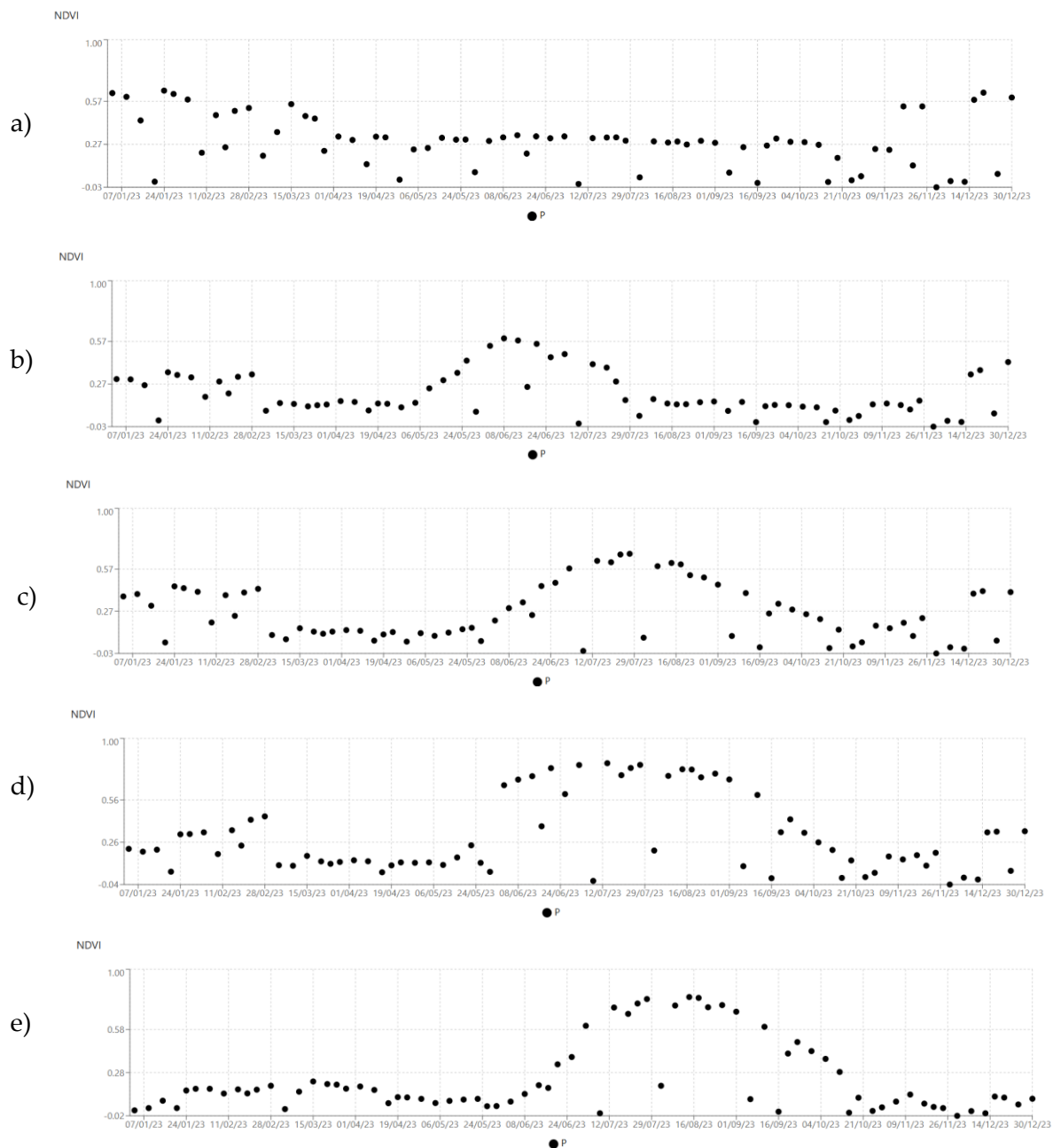


Figura 15 - Variação do NDVI para o ano de 2023 para: a) Zona do Montado; b) Subzona 1 do Tomatal; c) Subzona 2 do Tomatal; d) Subzona 3 do Tomatal; e) Zona do Arrozal.

6.2 Anexo II - Coordenadas dos Locais de Amostragem

Tabela 4 – Coordenadas e identificação das amostras.

ID	Latitude	Longitude	ID	Latitude	Longitude	ID	Latitude	Longitude
1	38.96709	-8.83466	38	38.97398	-8.83344	75	38.97389	-8.83915
2	38.96922	-8.83201	39	38.97298	-8.83237	76	38.97416	-8.84172
3	38.97114	-8.83197	40	38.96709	-8.83555	77	38.97492	-8.8427
4	38.96875	-8.83124	41	38.9742	-8.83206	78	38.97283	-8.83842
5	38.9675	-8.83386	42	38.97472	-8.83412	79	38.97212	-8.83992
6	38.96921	-8.83331	43	38.97421	-8.83377	80	38.9693	-8.83764
7	38.9682	-8.83274	44	38.97313	-8.83322	81	38.97413	-8.83999
8	38.97061	-8.83114	45	38.97442	-8.83283	82	38.97527	-8.83869
9	38.97285	-8.83226	46	38.96658	-8.83553	83	38.96763	-8.83594
10	38.97202	-8.83239	47	38.9679	-8.83466	84	38.96842	-8.83836
11	38.96724	-8.83413	48	38.96924	-8.82984	85	38.97193	-8.83896
12	38.96823	-8.832	49	38.97108	-8.83272	86	38.97344	-8.84206
13	38.96921	-8.8312	50	38.97158	-8.83259	87	38.97558	-8.83997
14	38.96981	-8.82885	51	38.97421	-8.83511	88	38.97452	-8.83869
15	38.97009	-8.82993	52	38.97305	-8.83734	89	38.9692	-8.83507
16	38.96923	-8.83043	53	38.97266	-8.83601	90	38.97429	-8.84076
17	38.97257	-8.83229	54	38.97263	-8.83452	91	38.97669	-8.83836
18	38.96959	-8.82951	55	38.97116	-8.83397	92	38.9772	-8.83746
19	38.96931	-8.83312	56	38.97162	-8.83538	93	38.98061	-8.83812
20	38.96873	-8.83334	57	38.9703	-8.83492	94	38.97597	-8.83594
21	38.97528	-8.83528	58	38.96846	-8.83522	95	38.97809	-8.84171
22	38.97303	-8.83272	59	38.97503	-8.84081	96	38.97852	-8.83848
23	38.97322	-8.83216	60	38.97098	-8.83634	97	38.97952	-8.83983
24	38.97265	-8.83189	61	38.97197	-8.83739	98	38.97797	-8.84061
25	38.9709	-8.83183	62	38.97465	-8.83774	99	38.97756	-8.83944
26	38.96962	-8.83064	63	38.9749	-8.84056	100	38.97906	-8.84049
27	38.96846	-8.83319	64	38.9719	-8.83839	105	38.98082	-8.85051
28	38.96705	-8.8342	65	38.96928	-8.83862	106	38.98624	-8.85674
29	38.96819	-8.83418	66	38.96693	-8.8377	107	38.98366	-8.85759
30	38.9669	-8.83461	67	38.97196	-8.83624	110	38.97884	-8.847
31	38.96769	-8.8342	68	38.97385	-8.83625	111	38.97986	-8.85087
32	38.96773	-8.83274	69	38.97485	-8.84354	112	38.97945	-8.85263
33	38.97244	-8.83205	70	38.97033	-8.83834	122	38.97858	-8.86254
34	38.97268	-8.8328	71	38.96794	-8.83693	123	38.98351	-8.85711
35	38.97288	-8.83184	72	38.96913	-8.8366	124	38.98546	-8.85869
36	38.9736	-8.83181	73	38.97018	-8.83707	125	38.98	-8.85353
37	38.97371	-8.83325	74	38.9749	-8.83628	126	38.97345	-8.8694

Anexo II - Coordenadas dos Locais de Amostragem (Continuação)

ID	Latitude	Longitude	ID	Latitude	Longitude	ID	Latitude	Longitude
128	38.97734	-8.8631	186	38.97268	-8.8626	239	38.97417	-8.85781
129	38.97925	-8.86203	187	38.97256	-8.8618	240	38.97353	-8.85595
130	38.97081	-8.87162	188	38.97174	-8.85998	241	38.97353	-8.85595
131	38.9726	-8.86308	189	38.96977	-8.86763	242	38.97193	-8.85829
132	38.96982	-8.85756	190	38.96989	-8.86894	243	38.97304	-8.85889
133	38.97057	-8.85721	191	38.97365	-8.86731	244	38.97156	-8.8587
141	38.97891	-8.84563	193	38.97687	-8.86547	245	38.9699	-8.85776
142	38.97834	-8.8474	194	38.97108	-8.87235	246	38.97059	-8.85737
143	38.98094	-8.85646	195	38.97153	-8.87171	247	38.97758	-8.86457
144	38.98122	-8.85741	197	38.97344	-8.86876	249	38.97293	-8.85846
145	38.98557	-8.86099	200	38.97302	-8.86326	253	38.97856	-8.85078
146	38.98336	-8.85906	201	38.98145	-8.85909	254	38.98087	-8.85752
147	38.98433	-8.8616	202	38.97927	-8.86141	255	38.97535	-8.8571
148	38.9861	-8.86197	203	38.97846	-8.85969	257	38.97126	-8.86679
149	38.98594	-8.85948	204	38.97945	-8.85392	262	38.97268	-8.86078
150	38.98462	-8.85675	207	38.97671	-8.85413	263	38.97547	-8.85736
151	38.97659	-8.86509	209	38.97433	-8.85535	264	38.97234	-8.85966
154	38.98134	-8.86089	212	38.97369	-8.86369	267	38.97597	-8.8485
155	38.98402	-8.85699	213	38.97296	-8.87001	268	38.97579	-8.84995
156	38.98358	-8.86056	219	38.9685	-8.85374	269	38.97388	-8.86439
159	38.98035	-8.84399	222	38.98141	-8.86016	271	38.97422	-8.85819
162	38.96985	-8.85479	223	38.97441	-8.85547	274	38.97159	-8.8576
163	38.9708	-8.85632	224	38.9759	-8.85173	275	38.9752	-8.86116
164	38.96988	-8.85497	226	38.96767	-8.85388	276	38.9729	-8.86567
166	38.97204	-8.87021	228	38.97451	-8.84738	277	38.97376	-8.86556
168	38.97028	-8.87044	230	38.98017	-8.85538	278	38.97673	-8.85257
169	38.97322	-8.86642	231	38.97738	-8.85656	279	38.97744	-8.85224
171	38.9857	-8.85729	233	38.97789	-8.8575	280	38.97814	-8.85155
178	38.98274	-8.8604	236	38.97527	-8.86144	293	38.98516	-8.8619
181	38.9725	-8.85633	237	38.97466	-8.86172	300	38.97815	-8.84856
184	38.97005	-8.85408	238	38.97464	-8.86155			

6.3 Anexo III – Pré-processamento das Variáveis Químicas

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_excel("C:/Users/maria/OneDrive/Área de Trabalho/AP/Tese/Mo-
delo/quim_originais.xlsx") # ALTERAR CAMINHO QUANDO NECESSÁRIO
print(f"Data shape: {df.shape}")

# Remover a coluna "Mol"
df.drop(columns=["Mol"], inplace=True)

# Substituir todos os valores que estão abaixo do limite de detecção pelo li-
mite de detecção
df = df.applymap(
    lambda x: float(x.replace("<", "")) if isinstance(x, str) and
x.startswith("<") else x
)

# Remove outliers extremos e substitui por NaN
df_limpo = df.copy()
num_cols = df_limpo.select_dtypes(include="number").columns

for col in num_cols:
    z = df_limpo.groupby("Cultura")[col].transform(lambda x: (x - x.mean()) /
x.std())
    df_limpo.loc[z.abs() > 3, col] = np.nan

df_limpo.to_csv(r"C:/Users/maria/OneDrive/Área de Trabalho/AP/Tese/Mo-
delo/quim_p_modelo.csv", index=False) # ALTERAR CAMINHO QUANDO NECESSÁRIO
```


6.4 Anexo IV - Pré-processamento dos Espectros

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:/Users/maria/OneDrive/Área de Trabalho/AP/Tese/Mo-
delo/NIR_originais.csv", delimiter=";") # ALTERAR CAMINHO QUANDO NECESSÁRIO
print(f"Data shape: {df.shape}")

nir_cols = [col for col in df.columns if str(col).replace('.', '').isdigit()]

# Remover espectros nir negativos e "noise"
keep_cols = [col for col in nir_cols
              if not (1350 <= float(col) <= 1430) and float(col) <= 2000 and
float(col) >= 980]
df_clean = df[keep_cols].copy()

# Aplicar médias moveis
def moving_average(spectra, window_size=5):
    """Apply moving average smoothing and return both smoothed data and valid
    wavelengths"""
    smoothed = spectra.rolling(window=window_size, axis=1, center=True,
min_periods=1).mean()
    return smoothed

wavelengths = df_clean.columns.astype(float)

df_smoothed = moving_average(df_clean)

assert len(wavelengths) == df_smoothed.shape[1], "Dimension mismatch!"

df_final = df_smoothed.copy()
df_final = df_final.join(df[['ID', 'Cultura', 'Analisada']])

output_path = "C:/Users/maria/OneDrive/Área de Trabalho/AP/Tese/Mo-
delo/NIR_p_modelo.csv" # ALTERAR CAMINHO QUANDO NECESSÁRIO

df_final.to_csv(output_path, index=False)
```

6.5 Anexo V – Modelação com *Random Forest*

```
import pandas as pd
import numpy as np
import joblib
from sklearn.model_selection import KFold, cross_val_predict
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score

nir_all = pd.read_csv(r"C:\Users\maria\OneDrive\Área de Trabalho\AP\Tese\Mo-
delo_TESE\NIR_p_modelo.csv", sep=",") # ALTERAR CAMINHO QUANDO NECESSÁRIO
quim = pd.read_csv(r"C:\Users\maria\OneDrive\Área de Trabalho\AP\Tese\Mo-
delo_TESE\quim_p_modelo.csv", sep=",") # ALTERAR CAMINHO QUANDO NECESSÁRIO

data = pd.merge(nir_all, quim, on="ID")

nir_cols = nir_all.columns.drop(["ID", "Cultura", "Analisada"])
quim_cols = quim.columns.drop(["ID", "ID_Amostra", "Cultura"])

# Validação cruzada (5-fold) + calibração
print("\n Validação cruzada (5-fold) + métricas de calibração:")
results_all = []
cv = KFold(n_splits=5, shuffle=True, random_state=42)

for var in quim_cols:
    y = data[var]
    mask = ~y.isnull()
    X_var = data.loc[mask, nir_cols].values
    y_var = y[mask].values

    rf_cv = RandomForestRegressor(
        n_estimators=100,
        max_depth=None,
        random_state=42,
        n_jobs=-1
    )
    y_cv = cross_val_predict(rf_cv, X_var, y_var, cv=cv)
    rmse_cv = np.sqrt(mean_squared_error(y_var, y_cv))
    r2_cv = r2_score(y_var, y_cv)
    rpd_cv = np.std(y_var, ddof=1) / rmse_cv

    rf_cal = RandomForestRegressor(
        n_estimators=100,
        max_depth=None,
        random_state=42,
        n_jobs=-1
```

```

)
rf_cal.fit(X_var, y_var)
y_cal = rf_cal.predict(X_var)
rmse_cal = np.sqrt(mean_squared_error(y_var, y_cal))
r2_cal = r2_score(y_var, y_cal)
rpd_cal = np.std(y_var, ddof=1) / rmse_cal

results_all.append((var, rmse_cal, r2_cal, rpd_cal,
                    rmse_cv, r2_cv, rpd_cv))
print(f"{var}: RMSEcal={rmse_cal:.3f}, R²cal={r2_cal:.3f},
RPDcal={rpd_cal:.3f}, "
      f"RMSEcv={rmse_cv:.3f}, R²cv={r2_cv:.3f}, RPDcv={rpd_cv:.3f}")

# Guardar resultados finais
df_results = pd.DataFrame(
    results_all,
    columns=["Variável", "RMSE_Cal", "R²_Cal", "RPD_Cal",
            "RMSE_CV", "R²_CV", "RPD_CV"]
)
df_results.to_csv("resultados_RF.csv", sep=";", index=False)
print("\n Resultados de calibração, validação e RPD guardados em resulta-
dos_RF.csv")

print("\nA treinar modelos finais...")
data_train = data[data["Analisada"] == 1]

models = {}
for var in quim_cols:
    y = data_train[var]
    mask = ~y.isnull()
    X_var = data_train.loc[mask, nir_cols].values
    y_var = y[mask].values

    rf = RandomForestRegressor(
        n_estimators=500,
        max_depth=None,
        random_state=42,
        n_jobs=-1
    )
    rf.fit(X_var, y_var)

    models[var] = rf
    joblib.dump(rf, f"modelo_RF_{var}.pkl")

```

```

print("Modelos finais treinados e guardados.")

# Previsão das variáveis químicas
data_novo = nir_all[nir_all["Analisada"] == 0]
X_new = data_novo[nir_cols].values

preds = {}
for var, model in models.items():
    preds[var] = model.predict(X_new)

df_preds = pd.DataFrame(preds)
df_preds.insert(0, "ID", data_novo["ID"].values)
df_preds.insert(1, "Cultura", data_novo["Cultura"].values)

df_preds.to_csv("previcoes.csv", sep=";", index=False)

print("\n Primeiras previsões:")
print(df_preds.head())
print("\n Previsões guardadas em previcoes.csv")

# Visualização reais vs. previstas
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import KFold, cross_val_predict

vars_plot = [
    "Mg", "MO_LOI", "Fe", "Ni", "Ca", "Na", "Co", "K",
    "B", "Mn", "MnAI", "Cu", "S", "P", "CE", "Zn",
    "ph", "Azoto_T", "TOC", "MO", "Carbono_T", "Razao_C_N", "IC", "C_microb",
    "F_TC"
]

units = {
    "Mg": "mg/kg",
    "MO_LOI": "%",
    "Fe": "mg/kg",
    "Ni": "mg/kg",
    "Ca": "mg/kg",
    "Na": "mg/kg",
    "Co": "mg/kg",
    "K": "mg/kg",
    "B": "mg/kg",
    "Mn": "mg/kg",
    "MnAI": "mg/kg",
    "Cu": "mg/kg",
    "S": "mg/kg",
    "P": "mg/kg",

```

```

"CE": "mS/cm",
"Zn": "mg/kg",
"ph": "",
"Azoto_T": "%",
"TOC": "mg/L",
"MO": "%",
"Carbono_T": "%",
"Razao_C_N": "%",
"IC": "mg/L",
"C_microb": "mg/L",
"F_TC": "mg/L"
}

cv = KFold(n_splits=5, shuffle=True, random_state=42)

fig, axes = plt.subplots(7, 4, figsize=(18, 24))
axes = axes.flatten()

for i, var in enumerate(vars_plot):
    try:
        y = data[var]
        mask = ~y.isnull()
        X_var = data.loc[mask, nir_cols].values
        y_var = y[mask].values

        rf = RandomForestRegressor(
            n_estimators=100,
            max_depth=None,
            random_state=42,
            n_jobs=-1
        )

        y_pred = cross_val_predict(rf, X_var, y_var, cv=cv)

        ax = axes[i]
        ax.scatter(y_var, y_pred, alpha=0.6, edgecolor="k")
        ax.plot([y_var.min(), y_var.max()],
                [y_var.min(), y_var.max()],
                "r--", lw=2, label="1:1")
        ax.set_xlabel(f"Real ({units.get(var, '')})")
        ax.set_ylabel(f"Previsto ({units.get(var, '')})")
        ax.set_title(f"{var}")
        ax.grid(True)
    except Exception as e:
        print(f"Erro com {var}: {e}")

for j in range(len(vars_plot), len(axes)):

```

```
fig.delaxes(axes[j])  
  
plt.tight_layout()  
plt.show()
```

6.6 Anexo VI – Modelação com PLSR

```
import pandas as pd
import numpy as np
import joblib
from sklearn.model_selection import KFold, cross_val_predict
from sklearn.cross_decomposition import PLSRegression
from sklearn.metrics import mean_squared_error, r2_score

nir_all = pd.read_csv(r"C:\Users\maria\OneDrive\Área de Trabalho\AP\Tese\Mo-
delo_TESE\NIR_p_modelo.csv", sep=",") # ALTERAR CAMINHO QUANDO NECESSÁRIO
quim = pd.read_csv(r"C:\Users\maria\OneDrive\Área de Trabalho\AP\Tese\Mo-
delo_TESE\quim_p_modelo.csv", sep=",") # ALTERAR CAMINHO QUANDO NECESSÁRIO
data = pd.merge(nir_all, quim, on="ID")

nir_cols = nir_all.columns.drop(["ID", "Cultura", "Analisada"])
quim_cols = quim.columns.drop(["ID", "ID_Amostra", "Cultura"])

# Validação cruzada (5-fold)
print("\n Validação cruzada (5-fold):")
results_cv = []
cv = KFold(n_splits=5, shuffle=True, random_state=42)
max_comp = 20

for var in quim_cols:
    y = data[var]
    mask = ~y.isnull()
    X_var = data.loc[mask, nir_cols].values
    y_var = y[mask].values

    best_r2cv, best_rmsecv, best_comp, best_rpd_cv = -np.inf, None, None, None

    for n_comp in range(1, min(max_comp, X_var.shape[1]) + 1):
        pls = PLSRegression(n_components=n_comp)
        y_cv = cross_val_predict(pls, X_var, y_var, cv=cv)

        rmse_cv = np.sqrt(mean_squared_error(y_var, y_cv))
        r2_cv = r2_score(y_var, y_cv)
        rpd_cv = np.std(y_var, ddof=1) / rmse_cv

        if r2_cv > best_r2cv:
```

```

        best_r2cv, best_rmsecv, best_comp, best_rpd_cv = r2_cv, rmse_cv,
n_comp, rpd_cv

    data_train = data[data["Analisada"] == 1]
    y_train = data_train[var]
    mask_train = ~y_train.isnull()
    X_train = data_train.loc[mask_train, nir_cols].values
    y_train = y_train[mask_train].values

    pls_final = PLSRegression(n_components=int(best_comp))
    pls_final.fit(X_train, y_train)

    y_cal = pls_final.predict(X_train).ravel()
    rmse_cal = np.sqrt(mean_squared_error(y_train, y_cal))
    r2_cal = r2_score(y_train, y_cal)
    rpd_cal = np.std(y_train, ddof=1) / rmse_cal

    joblib.dump(pls_final, f"modelo_PLSR_{var}.pkl")

    results_cv.append((var, rmse_cal, r2_cal, rpd_cal,
                        best_rmsecv, best_r2cv, best_rpd_cv, best_comp))

    print(f"{var}: RMSEcal={rmse_cal:.3f}, R²cal={r2_cal:.3f},
RPDcal={rpd_cal:.3f}, "
          f"RMSEcv={best_rmsecv:.3f}, R²cv={best_r2cv:.3f},
RPDcv={best_rpd_cv:.3f}, Comp={best_comp}")

df_results_cv = pd.DataFrame(
    results_cv,
    columns=["Variável", "RMSE_Cal", "R²_Cal", "RPD_Cal",
             "RMSE_CV", "R²_CV", "RPD_CV", "N_Componentes"]
)

df_results_cv.to_csv("resultados_PLSR.csv", sep=";", index=False)
print("\n Resultados de calibração, validação e RPD guardados em resulta-
dos_PLSR.csv")

```

