

Aplicação de Previsão Edafoclimática e Aptidão Cultural em Portugal Continental

Rubén Romero Torrado

Dissertação para a obtenção do grau de mestre em
Ciência de Dados em Agricultura, Alimentação, Floresta e Ambiente

Orientador: Rui Paulo Nóbrega Figueira

Mentor: Pompeu Pais Dias

Júri:

PRESIDENTE

Doutor Manuel Lameiras de Figueiredo Campagnolo, Professor associado com agregação do(a) Instituto Superior de Agronomia da Universidade de Lisboa.

VOGAIS

Doutor Rui Paulo Nóbrega Figueira, Professor associado do(a) Instituto Superior de Agronomia da Universidade de Lisboa, orientador; Doutor João Rui Rolim Fernandes Machado Lopes, Professor auxiliar do(a) Instituto Superior de Agronomia da Universidade de Lisboa.

Agradecimentos

Queria agradecer em primeiro lugar à minha mãe, ao meu pai e ao meu avô Romerito por me apoiarem nos momentos mais árduos da realização desta tese, sem eles não teria conseguido lidar com todos os problemas que afrontei.

Agradecer à minha namorada e aos meus amigos que entenderam que o meu afastamento temporário em relação a eles não foi uma vontade minha mas sim uma necessidade causada pelo tempo necessário para investir neste projeto.

Agradecer ao meu mentor Pompeu e o meu orientador Figueira pela ajuda e atenção prestada para ter um trabalho em condições.

Resumo

A agricultura em Portugal enfrenta elevada incerteza climática e grande variabilidade de solos. Esta incerteza tem implicações diretas na decisão sobre as culturas mais adequadas a cada região, tornando essencial dispor de ferramentas que orientem o agricultor com base em dados objetivos. Para apoiar decisões práticas sobre que cultura plantar em cada local foi criado um sistema de apoio à decisão que prevê variáveis edafoclimáticas e converte essas previsões em rankings de aptidão com indicação dos fatores limitantes. O trabalho cobre todo o ciclo de ciência de dados desde a recolha e integração de informação de 158 localizações até à modelação e à disponibilização numa interface interativa.

O modelo preditivo baseia-se em Multi-Output Random Forest e estima, em simultâneo, horas de frio, dias de geada, temperaturas, precipitação e atributos do solo. A interface permite escolher uma localização em Portugal continental e obter a aptidão de oito culturas mediterrânicas com visualizações claras e recomendações acionáveis. Os resultados mostram desempenho elevado, com coeficiente de determinação global próximo de 0,95 e especial robustez nas variáveis climáticas. As variáveis de solo apresentam menor precisão devido à maior variabilidade local, sem comprometer a utilidade do sistema. A comparação com catorze modelos, desenvolvidos individualmente para cada variável climática ou de solo, indica ligeiro ganho de exatidão pontual, mas menor capacidade de generalização, favorecendo a abordagem multioutput para utilização real.

Como medidas de melhoria do sistema desenvolvido, recomenda-se ampliar o leque de culturas e incluir variedades, enriquecer a base de dados e as covariáveis, criar uma base permanente com monitorização contínua, validar em mais regiões do país e integrar cenários futuros através de modelos mecanísticos, fortalecendo a resiliência das decisões agrícolas.

Palavras-chave: Agricultura orientada por dados, Ciência de Dados, Sistemas de Apoio à Decisão, Machine Learning, Aptidão Agrícola

Abstract

Agriculture in Portugal faces high climatic uncertainty and significant soil variability. This uncertainty directly affects decisions regarding which crops are most suitable for each region, highlighting the need for tools that can guide farmers through data-driven insights. To support practical decisions on which crop to plant in each location, a decision support system was developed that predicts edaphoclimatic variables and converts these predictions into suitability rankings, identifying the main limiting factors. The work covers the entire data science cycle—from the collection and integration of information from 158 locations to modelling and the development of an interactive interface.

The predictive model is based on a Multi-Output Random Forest, capable of simultaneously estimating chilling hours, frost days, temperature, precipitation, and soil attributes. The interactive interface allows users to select any location in mainland Portugal and obtain crop suitability results for eight Mediterranean species, presented through clear visualisations and actionable recommendations.

The results show high predictive performance, with an overall coefficient of determination close to 0.95 and particularly strong robustness for climatic variables. Soil-related variables display lower precision due to greater local variability, without compromising the system's overall usefulness. A comparison with fourteen single-output models created for each climatic and soil variables indicates slightly higher point accuracy but lower generalisation capacity, favouring the multi-output approach for real-world application.

As an improvement of the current system, it is recommended to expand the range of crops and include intra-species varieties, enrich the database and covariates, establish a permanent and continuously monitored data infrastructure, validate the model in additional regions of the country, and integrate future climate scenarios through mechanistic modelling—thereby strengthening the resilience of agricultural decision-making.

Keywords: Data-driven Agriculture, Data Science, Decision Support Systems, Machine Learning, Crop Suitability

Índice

Agradecimentos.....	I
Resumo.....	II
Abstract.....	III
Índice de Quadros.....	IV
Índice de Quadros do Anexo.....	V
Índice de Figuras.....	VI
1. Introdução.....	1
1.1 Definição do Problema e Proposta de Solução.....	1
1.2 O Papel da Ciência de Dados face aos Desafios Agrícolas.....	2
1.2.1 Aquisição e Gestão de Dados.....	2
1.2.2 Modelos Preditivos e Analíticos.....	3
1.2.3 Impacto em Sustentabilidade Ambiental, Custos e Eficiência de Recursos.....	5
1.2.4 Aplicação ao projeto.....	6
1.3 A Emergência dos Sistemas de Suporte à Decisão (SSD) na Agricultura.....	8
1.4 Arquitetura e Tendências em Sistemas de Apoio à Decisão (DSS) Agrícolas.....	10
1.5 Objetivos do trabalho.....	11
2. Materiais e métodos.....	11
2.1 Recolha e descrição dos dados.....	12
2.2. Desenvolvimento do modelo Random Forest.....	18
2.2.1 Pré-processamento das variáveis.....	19
2.2.2 A Base Teórica do Random Forest.....	19
2.2.3 As Vantagens Únicas da Regressão Multi-Output.....	20
2.2.4 Otimização dos Hiperparâmetros com GridsearchCV.....	21
2.2.5 Treino e métricas de avaliação.....	22
2.3 Comparação dos valores previstos e cálculo do índice de aptidão.....	23
3. Resultados.....	25
3.1 Análise exploratória de dados (EDA).....	25
3.2 Melhores valores de Hiperparâmetros encontrados.....	36
3.3 Métricas de avaliação e gráficos comparativos.....	36
3.4 Manual de utilização do programa.....	40
3.4.1 Estrutura.....	40
3.4.2 Guia de utilização passo a passo.....	42
4. Discussão.....	45
4.1 Discussões sobre a Análise exploratória de dados.....	46
4.2 Interpretação dos Resultados no Contexto do Problema.....	46
4.3 Comparação com a Abordagem de Modelos Individuais.....	47

4.4 Contextualização dos Resultados: Comparação com Estudos Anteriores.....	48
4.5 Interpretação da Discrepância de Desempenho.....	48
4.6 Comparação do programa com outros DSS.....	49
4.6.1 Sistemas semelhantes na literatura.....	49
4.6.2 Comparação dirigida entre o programa desenvolvido e sistemas semelhantes.....	50
4.6.3 Inovações do programa.....	51
4.6.4 Limitações e oportunidades de evolução.....	52
Recomendações.....	53
Conclusão.....	54
Anexos.....	55
Referências Bibliográficas.....	64

Índice de Quadros

Quadro 1 Conjuntos de dados usados como fonte para cada variável	18
Quadro 2 -Resumo das Variáveis Climáticas e descrição.	20
Quadro 3. Resumo das Variáveis edáficas e descrição.	21
Quadro 4 Possíveis valores para otimização de hiperparâmetros com GridsearchCV	24
Quadro 5 Resumo estatístico das Temperaturas médias mínimas e máximas.	27
Quadro 6 Resumo estatístico dos dias de geada	27
Quadro 7 Resumo estatístico das horas de frio	28
Quadro 8 Resumo estatístico da precipitação.	28
Quadro 9 Resumo estatístico das variáveis do solo.	29
Quadro 10 Melhores valores encontrados para otimizar os hiperparâmetros com GridsearchCV.	35
Quadro 11 Comparação global entre as duas abordagens, determinados a partir do conjunto de dados de teste.	36

Índice de Quadros do Anexo

Quadro A1 Métricas individuais do modelo multi-ouput implementado determinado com base no conjunto de dados de teste.	50
Quadro A2 Condições ideais e toleráveis do número de geadas para cada cultura	51
Quadro A3 Intervalos ideais e toleráveis de horas de frio entre outubro e fevereiro para cada cultura	52
Quadro A4 Intervalos ideais e toleráveis de precipitação entre outubro e abril para cada cultura	52
Quadro A5 Intervalos ideais e toleráveis de pH para cada cultura	52
Quadro A6 Condições ideais e toleráveis de temperatura normal para cada cultura	53
Quadro A7 Condições ideais e toleráveis de temperaturas extremas para cada cultura	54
Quadro A8 Texturas ideais e toleráveis para cada cultura	55
Quadro A9 Intervalos ideais e toleráveis de espessura para cada cultura	55

Índice de Figuras

Figura 1 Representação das localizações presentes no conjunto de dados	
Figura 2 Esquema da minha autoria do fluxo de trabalho para a obtenção do dataset final.	17
Figura 3. Diagramas de extremos e quartis das Variáveis do Dataset	26
Figura 4 Matriz de diagramas de dispersão de variáveis climáticas	28
Figura 5 Matriz de correlação entre as variáveis de estudo	28
Figura 6 Matriz de diagramas de dispersão de variáveis do solo	30
Figura 7 Comparação entre as métricas de desempenho entre as duas abordagens de modelação multi-output e modelos individuais.	33
Figura 8. Representação dos valores previstos vs. Valores Reais no conjunto de dados de teste para o modelo multioutput	34
Figura 9 Separador 1 do sistema de apoio à decisão - Mapa interativo.	36
Figura 10 Separador 1 do sistema de apoio à decisão – Introdução de inputs	36
Figura 11 Separador 2 do sistema de apoio à decisão – Resultados	37
Figura 12 Separador 3 do sistema de apoio à decisão - Gráficos	38
Figura 13 Separador 4 do sistema de apoio à decisão - Histórico	38

1. Introdução

A previsão edafoclimática e a avaliação da aptidão cultural são componentes fundamentais do planeamento agrícola, pois permitem antecipar o comportamento das variáveis do solo e do clima que determinam o sucesso das culturas. Tradicionalmente, estas avaliações baseiam-se em séries históricas de dados meteorológicos e em mapas edáficos de base, que são depois cruzados com requisitos agronómicos conhecidos para estimar a adequação de cada cultura a uma determinada região. Embora eficazes em contextos estáveis, estes métodos tendem a ser limitados pela resolução espacial e temporal dos dados e pela dificuldade em capturar interações complexas entre fatores ambientais. É neste ponto que a ciência de dados oferece um avanço significativo: ao integrar grandes volumes de informação de origem diversa, como sensores remotos, bases de dados geográficas e registos locais e aplicar técnicas de ciência de dados, torna possível gerar previsões mais precisas, contínuas e adaptativas. Esta abordagem não só melhora a fiabilidade das estimativas edafoclimáticas, como também permite construir sistemas de apoio à decisão mais dinâmicos, capazes de identificar oportunidades, antecipar riscos e apoiar uma agricultura mais eficiente e resiliente face à variabilidade climática.

1.1 Definição do Problema e Proposta de Solução

O problema central desta investigação reside na inadequação das metodologias tradicionais para gerar previsões edafoclimáticas multivariadas de alta precisão e resolução espacial, indispensáveis para uma determinação adaptativa da aptidão cultural em Portugal Continental. A predição de variáveis do solo e do clima é classicamente tratada de forma isolada, ignorando as relações intrínsecas e as interdependências espaciais entre elas.

A solução proposta consiste no desenvolvimento de um programa informático que integra dois componentes essenciais. Um Módulo Preditivo baseado num modelo de *Machine Learning* do tipo Multi-Output Random Forest Regressor (MORF), capaz de prever simultaneamente um conjunto de variáveis edafoclimáticas para qualquer localização em Portugal Continental,

explorando as possíveis relações existentes entre estas variáveis. Um Módulo de Avaliação Agronômica que utiliza estas previsões como *input* para calcular a aptidão das principais culturas através de um sistema de pontuação, identificando explicitamente os fatores limitantes de acordo com os requisitos ecofisiológicos da cultura.

Esta abordagem eleva o paradigma existente de sistemas de apoio à decisão no contexto nacional, passando de um foco primário na avaliação de eficiência e realocação de atividades para um sistema que incorpora uma camada de previsão e *adaptação* baseada em *Machine Learning* antes da fase de otimização, tornando as decisões mais resilientes e informadas.

1.2 O Papel da Ciência de Dados face aos Desafios Agrícolas

A agricultura moderna enfrenta um conjunto crescente de desafios, desde a imprevisibilidade da variabilidade climática até à necessidade de otimizar a produção para responder à procura global de alimentos. A complexidade destes fatores ambientais e agronômicos, que se manifestam através de interações não lineares, frequentemente supera a capacidade das metodologias analíticas convencionais (Araba, 2024). Neste contexto, a ciência de dados (data science) desempenha um papel central, integrando desde a definição do problema, recolha e gestão de dados até à sua análise avançada com métodos de inteligência artificial (IA) e *machine learning* (ML). Wolfert *et al.* (2017) realizaram uma revisão sobre as aplicações do *Big Data* na *smart agriculture* e consolidaram a ciência de dados como pilar da “agricultura inteligente”, ao integrar um grande volume de dados, modelos e governança ao longo de toda a cadeia agroalimentar.

1.2.1 Aquisição e Gestão de Dados

O ciclo da ciência de dados aplicado à agricultura engloba várias etapas. Em primeiro lugar, a aquisição e integração de dados assume importância fundamental. Estes dados podem ser

provenientes de múltiplas fontes, como bases públicas de informação meteorológica e edafológica (por exemplo, European Soil Data Centre – ESDAC, Instituto Português do Mar e da Atmosfera – IPMA), imagens de satélite (Sentinel, MODIS) e registos históricos de produção agrícola. O acesso e a interoperabilidade entre estas fontes requerem metodologias robustas de gestão de dados, incluindo normalização, tratamento de valores em falta e integração espacial e temporal (Kamilaris & Prenafeta-Boldú, 2018; Wolfert *et al.*, 2017).

O trabalho feito por Wolfert *et al.* (2017) demonstrou que a governança de dados e a interoperabilidade são condições de base para que os *pipelines* de dados conduzam a ganhos económicos e ambientais mensuráveis na exploração, do campo ao processamento e à distribuição.

O projeto desenvolvido por Kamilaris & Prenafeta-Boldú (2018) demonstrou que a preparação de dados (curadoria, *augmentation*, equilíbrio de classes) aumenta a qualidade dos modelos produzidos em tarefas de visão computacional agrícola para a detecção de doenças, contagem de frutos e classificação fenotípica. Benos *et al.* (2021) evidenciaram que a engenharia de atributos e a seleção de variáveis melhoram a robustez e a interpretabilidade de modelos para gestão de culturas, água, solo e pecuária. Estes quatro domínios somam a maioria das aplicações de ML em agricultura.

1.2.2 Modelos Preditivos e Analíticos

A utilização de modelos preditivos e analíticos como o Random Forest Regressor multi-output, tal como implementado no presente projeto, permite lidar com variáveis edafoclimáticas de forma robusta, captando relações complexas e não lineares entre os fatores do solo, clima e produção. Estes modelos caracterizam-se pela capacidade de prever múltiplas variáveis simultaneamente, explorando correlações entre saídas (outputs) que, em modelos independentes, seriam tratadas de forma isolada (Kocev *et al.*, 2013).

Nos últimos anos, a aplicação de abordagens multi-output Random Forest (MORF) tem ganho relevância em áreas que exigem a previsão conjunta de parâmetros ambientais ou agronómicos

interdependentes. O trabalho feito por Kocev et al. (2013) demonstrou que o MORF supera abordagens de regressão independentes quando existe forte correlação entre variáveis alvo, ao permitir que a partição das árvores seja influenciada por dependências entre múltiplas saídas. Esse avanço metodológico reduziu o erro preditivo em conjuntos de dados complexos, como os de produtividade agrícola e parâmetros meteorológicos combinados.

O estudo de Jeong et al. (2016) aplicou um modelo Random Forest multioutput para prever simultaneamente temperatura mínima e precipitação diária, obtendo um desempenho superior ao de redes neurais e regressões lineares múltiplas em dados climáticos coreanos. Da mesma forma, Sun et al. (2020) utilizaram um MORF para estimar de forma conjunta temperatura média, radiação solar e humidade relativa, melhorando a coerência interna das previsões meteorológicas e a fiabilidade para sistemas de apoio à decisão agrícola.

Em contextos agrícolas, Ramesh et al. (2022) aplicaram um modelo multi-output Random Forest para prever simultaneamente rendimento, evapotranspiração e consumo hídrico em culturas de arroz, demonstrando que a consideração conjunta das variáveis aumentou a precisão média em 15% face a modelos isolados. O trabalho de Nawar et al. (2021) mostrou ainda que o MORF permite capturar a interação entre atributos do solo e variáveis climáticas, obtendo previsões consistentes de produtividade em sistemas de agricultura de precisão.

No domínio da modelação edafoclimática, Hengl et al. (2018) destacaram que a implementação de *multi-output ensembles* baseados em Random Forest, tal como no *framework* SoilGrids250m, que possibilita a previsão simultânea de múltiplas propriedades do solo (pH, matéria orgânica, textura, densidade aparente), garantindo coerência espacial e física entre variáveis correlacionadas. Esse tipo de abordagem é especialmente relevante quando se pretende integrar informação climática e edáfica em mapas de aptidão agrícola.

O trabalho feito por Oliveira et al. (2023), focado em regiões mediterrânicas, reforçou que modelos multioutput podem ser aplicados à previsão simultânea de dias de geada, horas de frio e temperatura mínima média, parâmetros críticos para a seleção de espécies frutícolas adaptadas às alterações climáticas. Tal como no presente projeto, os autores salientam que o MORF não apenas melhora a precisão individual das previsões, mas também assegura maior consistência entre variáveis que partilham determinantes climáticos comuns.

Contudo, a ciência de dados não se limita ao desenvolvimento do modelo: inclui igualmente a validação cruzada, a interpretação dos resultados e a sua integração num sistema de suporte à decisão para os utilizadores finais (Chlingaryan et al., 2018; Liakos et al., 2018). Nesse sentido, a capacidade do MORF de gerar previsões simultâneas e coerentes de temperatura, precipitação, horas de frio e dias de geada reforça o seu valor como componente central de um Sistema de Suporte à Decisão (DSS) voltado à agricultura de precisão e à gestão sustentável de recursos.

O trabalho desempenhado por van Klompenburg et al. (2020) demonstrou que a qualidade da *feature engineering* e o desenho do conjunto de treino têm impacto direto no erro de previsão de rendimento, reforçando boas práticas de validação cruzada e *hold-out* temporal. Kamilaris & Prenafeta-Boldú (2018) também mostram que arquiteturas profundas, quando acompanhadas de *pre-processing* consistente e *data augmentation*, superam abordagens clássicas em múltiplas tarefas agrícolas, mas exigem rastreabilidade dos dados e *explainability* para adoção operacional.

1.2.3 Impacto em Sustentabilidade Ambiental, Custos e Eficiência de Recursos

A literatura recente quantifica benefícios ambientais e económicos da ciência de dados quando ligada a sensores IoT, SIG/teledeteção e VRT (*Variable-Rate Technology*). Relativamente à irrigação de precisão, o trabalho feito por Zhao et al. (2023) demonstrou que a programação de rega data-driven integrada com previsão de precipitação permite poupanças adicionais de 20–50 % de água sem perda de produtividade. É necessário destacar o trabalho realizado por Sharma et al. (2025) que reportou reduções até 54 % no uso de água com sistemas sensorizados face a rega convencional, mantendo o desempenho produtivo. Revisões de 2024 e 2025 sintetizam a maturidade de sistemas de rega com IA/IoT e o seu papel na gestão sustentável da água (Lakhiar et al., 2024).

Em relação ao uso de fertilizantes azotados e fosfatados e às suas emissões associadas, o trabalho feito por Bahmutsky et al. (2024) mostrou, em revisão de ACV/custos, que a fertilização a taxa variável tende a reduzir vários impactos ambientais (ainda que não em todos os

contextos), apoiando a transição para estratégias de insumo-ótimo. A investigação efetuada por Hagn et al. (2025) em ensaios *on-farm* indicou melhorias na eficiência do azoto e no balanço N com VRA sensorizado em trigo-de-inverno, face à aplicação uniforme. O ensaio feito por Wu et al. (2024) mostrou maior rendimento com menores custos ambientais usando fertilização de precisão orientada por dados. Revisões e estudos recentes sistematizam metodologias para detecção em tempo real de azoto através de sensores, teledeteção e otimização, apontando ganhos de eficiência e redução de perdas (Ravikumar et al., 2024).

Em relação à tele detecção e aos custos operacionais, o trabalho feito por Marti-Jerez et al. (2023) demonstrou que o Sentinel-2 permite diagnosticar deficiências nutricionais de forma custo-efetiva, orientando intervenções pontuais. Revisões sobre sensores e *smart data* destacam quedas de custos operacionais por automatização da monitorização, maior oportunidade de intervenção e otimização de insumos (Soussi et al., 2024).

Para que estas oportunidades cheguem à exploração agrícola, as boas práticas de atualização de dados e modelos, a monitorização de deriva, o re-treino periódico, a interoperabilidade como APIs e catálogos FAIR são determinantes.

O trabalho feito por Liakos et al. (2018) demonstrou que a transferência para produção (interfaces e *feedback* do utilizador) é o que converte *insights* de ML em poupanças tangíveis e benefícios ambientais, sobretudo quando acoplada a VRT e teledeteção.

1.2.4 Aplicação ao projeto

A implementação de aplicações práticas representa a etapa final do ciclo da ciência de dados aplicada à agricultura. O desenvolvimento de uma interface interativa para o utilizador, como proposto neste projeto, é crucial para transformar resultados analíticos complexos em informação acionável. Interfaces gráficas e plataformas *web-based* têm vindo a consolidar-se como o elo entre o processamento avançado de dados e a tomada de decisão no terreno, tornando os modelos preditivos acessíveis a agricultores, técnicos e decisores (Liakos et al., 2018)

O trabalho feito por Antle et al. (2017) demonstrou que o sucesso de um sistema de suporte à decisão depende tanto da precisão dos modelos como da usabilidade e visualização dos resultados. A tradução de previsões complexas em métricas intuitivas, como índices de aptidão, mapas de risco ou escalas de produtividade que aumentam a compreensão e a confiança dos utilizadores, promovendo uma adoção mais ampla.

De facto, diversos estudos sublinham que a interatividade e o design centrado no utilizador são determinantes para a eficácia dos DSS agrícolas. O trabalho de Van Evert et al. (2016) desenvolveu um protótipo de interface geoespacial interativa para apoio à fertilização variável, onde os utilizadores podiam ajustar parâmetros de entrada e observar em tempo real o impacto sobre a dose recomendada. Esse tipo de *feedback* dinâmico revelou-se fundamental para o envolvimento dos agricultores no processo de decisão.

O trabalho de Beck et al. (2019), focado em modelos de previsão de pragas e doenças, implementou uma interface *web* interativa baseada em mapas e séries temporais, que permitia aos técnicos explorar cenários de risco e ajustar estratégias de intervenção. O estudo evidenciou que a visualização interativa e a manipulação de variáveis reduzem a dependência de consultores e favorecem a autonomia do utilizador final.

Exemplos de aplicações da ciência de dados na agricultura são cada vez mais frequentes. Os sistemas de recomendação baseados em dados edafoclimáticos permitem otimizar a rotação de culturas e a calendarização da sementeira. Os modelos de previsão climática têm sido usados para reduzir riscos associados a pragas e doenças (Jeong et al., 2016). Além disso, iniciativas de agricultura de precisão já utilizam dados de sensores IoT para ajustar em tempo real a irrigação e a aplicação de fertilizantes, reduzindo desperdícios e aumentando a sustentabilidade (Liakos et al., 2018).

De igual modo, Jain et al. (2020) apresentaram uma aplicação interativa baseada em *dashboard*, integrada com sensores IoT e modelos de *machine learning*, permitindo a monitorização em tempo real de parâmetros de solo e irrigação. O sistema mostrou ganhos de eficiência hídrica superiores a 25%, demonstrando que a disponibilização de previsões em interfaces acessíveis melhora diretamente o desempenho operacional.

Outro exemplo relevante é o AgroDSS desenvolvido por Kumar et al. (2021), que integra dados climáticos e edáficos num *framework* de recomendação de culturas. O sistema inclui um módulo interativo que permite ao utilizador selecionar uma localização num mapa e visualizar, de forma comparativa, as culturas mais adequadas segundo diferentes critérios (rendimento, consumo hídrico, estabilidade climática). Este tipo de abordagem combina visualização georreferenciada e recomendações personalizadas, alinhando-se diretamente com os objetivos do presente projeto.

No domínio da previsão e monitorização edafoclimática, Peter et al. (2020) desenvolveram uma interface *cloud-based* com Google Earth Engine que permitia ao utilizador explorar mapas de aptidão climática de múltiplas culturas e visualizar cenários de alteração climática. A adoção de ferramentas *web-GIS* e bibliotecas de visualização interativa, como *Leaflet* ou *Plotly Dash*, tem sido amplamente recomendada na literatura pela sua capacidade de traduzir dados complexos em representações claras e navegáveis (Wolfert et al., 2017; Kamilaris & Prenafeta-Boldú, 2018).

Assim, a interface interativa proposta neste projeto, ao traduzir previsões de variáveis edafoclimáticas em sistemas de pontuação da aptidão agrícola por cultura, permitindo ao utilizador explorar localizações, comparar resultados e selecionar as culturas mais adaptadas, enquadra-se plenamente no estado da arte dos DSS agrícolas interativos. Para além da acessibilidade, estas soluções reforçam o carácter participativo e adaptativo da ciência de dados na agricultura, facilitando a transferência de conhecimento entre investigadores e produtores e promovendo uma gestão agrícola mais sustentável e baseada em evidência (Chlingaryan et al., 2018; Jeong et al., 2016; Liakos et al., 2018; Wolfert et al., 2017; Kamilaris & Prenafeta-Boldú, 2018).

Neste enquadramento, o projeto em análise, que prevê variáveis edafoclimáticas e avalia a aptidão de culturas para localizações específicas em Portugal continental — insere-se diretamente nesta tendência global de transformação digital da agricultura, onde a interatividade e a transparência dos sistemas são essenciais para a adoção efetiva das tecnologias de *machine learning*. A sua abordagem representa um avanço significativo, ao conjugar não

apenas técnicas analíticas, mas todo o ecossistema da ciência de dados, desde a aquisição e gestão da informação até à construção de aplicações funcionais e participativas.

1.3 A Emergência dos Sistemas de Suporte à Decisão (SSD) na Agricultura

Os Sistemas de Suporte à Decisão são sistemas informáticos que auxiliam os utilizadores na tomada de decisões informadas, fornecendo dados em tempo real e recomendações práticas e assertivas (Kumar & Kumar, 2025). No contexto agrícola, os DSS combinam modelos biofísicos, bases de dados geográficas e regras de decisão para transformar dados ambientais, de mercado e de gestão em opções operacionais, reduzindo incerteza e risco.

O desenvolvimento de Sistemas de Apoio à Decisão (DSS) georreferenciados é crucial para abordar esta lacuna. Estes sistemas são concebidos para colmatar falhas de comunicação e fornecer opções de gestão alternativas que visam a melhoria do desempenho agrícola, abrangendo tanto o rendimento da colheita como indicadores de agrobiodiversidade (Marcelino et al., 2024). Entre as abordagens consolidadas destaca-se a de Minh (2024), quem realizou uma *scoping review* sobre a avaliação de aptidão de terras/culturas com integração SIG+MCDA (por exemplo, AHP, TOPSIS), que permite ponderar múltiplos critérios edafoclimáticos (textura, pH, declive, precipitação, temperaturas) e gerar classes de aptidão espacialmente explícitas. Revisões e estudos recentes mostram que SIG+MCDA aumenta a granularidade, a transparência e a reprodutibilidade da recomendação de culturas em diferentes regiões.

No plano metodológico, há duas famílias de referência para a recomendação de culturas com base em condições edafoclimáticas específicas. Fischer et al. (2021) documentaram a versão 4 das Zonas Agro-Ecológicas Globais (AEZ/GAEZ) da FAO, que integram clima, solos, topografia e níveis de insumos para estimar aptidão e potencial produtivo por cultura. Existem também modelos de nicho/aptidão climática como o EcoCrop, que calculam um índice de aptidão com base em requisitos termo-pluviométricos e pH, aplicáveis a grelhas espaciais e escalas

regionais. Estas abordagens são amplamente usadas para gerar mapas de aptidão atuais e futuros e fundamentar decisões de planejamento agrícola e de diversificação de culturas.

Sathiyamurthi et al. (2024) reforçam que a integração de critérios locais (p.ex., textura e drenagem do solo) com variáveis climáticas, através de MCDA (AHP/TOPSIS) e SIG, aumenta a precisão da recomendação e a utilidade prática do DSS para agricultores e decisores. Casos de estudo em contextos áridos como o de Bilal et al. (2025) relatam ganhos na identificação de “melhores culturas por talhão” e maior alinhamento com objetivos de sustentabilidade como eficiência no uso da água, conservação de solo e estabilidade de rendimento.

O sistema aqui proposto é um exemplo paradigmático de um SSD orientado a dados (*data-driven*). Ao permitir que o utilizador selecione uma localização num mapa interativo, preveja as variáveis edafoclimáticas associadas a essa localização e obtenha um ranking de aptidão de culturas, o programa oferece um framework completo para a gestão agrícola. Alinha-se, por um lado, com os princípios AEZ/GAEZ (integração clima-solo-gestão) e, por outro, com práticas modernas para pontuar e ordenar culturas candidatas segundo critérios locais e ponderações explícitas. Além disso, pode incorporar modelos de aptidão climática (e.g., EcoCrop ou frameworks modulares recentes como o CropSuite) para estimar janelas ótimas de sementeira, requisitos de vernalização, sensibilidade fotoperiódica e limitações térmicas, enriquecendo o ranking com informação fenológica e operacional.

Esta capacidade de fornecer informações personalizadas e localizadas, baseadas em previsões robustas, representa o valor central do projeto e alinha-o com o futuro da agricultura de precisão, onde a tecnologia é utilizada para maximizar a produtividade de forma eficiente e ambientalmente responsável. Peter et al. (2020) destacam que interfaces web/GEE com visualização interativa de nichos climáticos e aptidão de culturas aceleram a adoção por técnicos e produtores, promovendo decisões mais rápidas e transparentes, e facilitando a avaliação de cenários como o de mudança climática ou novas culturas.

1.4 Arquitetura e Tendências em Sistemas de Apoio à Decisão (DSS) Agrícolas

A utilização de DSS na agricultura é uma tendência crescente, especialmente devido à intensificação dos impactos das alterações climáticas (de Magalhães et al, 2025). Estes sistemas são concebidos para organizar e sequenciar diferentes metodologias de análise e informação. No contexto geográfico, os DSS na agricultura tipicamente utilizam sistemas de informação geográfica (GIS) para impor restrições espaciais, delimitando as áreas que são agronomicamente adequadas para cada tipo de cultura, e para ilustrar graficamente os planos de otimização propostos.

A relevância deste paradigma é confirmada por estudos de caso em Portugal, como o trabalho focado na região agrária do Ribatejo e Oeste, onde os DSS foram empregados para medir a eficiência das atividades agrícolas e para estudar a realocação de culturas. Métodos como a *Data Envelopment Analysis* (DEA) e a programação linear multiobjetivo foram utilizados para promover as atividades mais eficientes sem causar alterações drásticas no uso da terra existente (Pereira et al., 2010). A abordagem da tese proposta neste trabalho visa estender esta estrutura, utilizando os SIG apenas para a captura inicial dos dados edafoclimáticos, mas produzindo estimações baseadas em *Machine Learning*, como a primeira etapa crítica para informar a avaliação de aptidão das culturas.

Os modelos de aptidão cultural mais abrangentes tanto em área como número de culturas, como o CropSuite, demonstram a necessidade de sistemas que transcendam o foco em culturas singulares ou regiões limitadas, integrando a variabilidade climática. Tais modelos utilizam abordagens baseadas em lógica difusa (*fuzzy-logic*) e fundamentam-se na Lei do Mínimo de Liebig, permitindo a integração flexível de múltiplos fatores ambientais e socioeconómicos. Zabel et al. (2025) desenvolveram o modelo CropSuite, que por exemplo, calcula e emite resultados detalhados, incluindo aptidão do clima, do solo e da cultura, data de sementeira, bem como a identificação do(s) fator(es) limitante(s). A arquitetura proposta para a tese segue diferentes abordagens mas objetivos similares, garantindo que o DSS forneça resultados não apenas descritivos (pontuação), mas também analíticos (fator limitante).

1.5 Objetivos do trabalho

O principal objetivo do meu trabalho é desenvolver uma ferramenta inteligente capaz de estimar variáveis edafoclimáticas, como temperatura, precipitação, horas de frio e características do solo e, a partir dessas estimações, avaliar a aptidão agrícola de diferentes culturas em qualquer ponto de Portugal continental. A ideia central é criar um sistema que ajude agricultores, técnicos e consultores a escolher as culturas mais adequadas para cada local, com base em dados objetivos e atualizados, reduzindo a incerteza associada às decisões agrícolas.

Para isso, o trabalho combina diferentes etapas da ciência de dados: a recolha e integração de informação proveniente de várias fontes (meteorológicas, pedológicas), o tratamento desses dados, o desenvolvimento de um modelo de *machine learning* capaz de gerar previsões simultâneas de múltiplas variáveis e, finalmente, a criação de uma interface interativa que traduz esses resultados em rankings de aptidão cultural.

Além de produzir estimações precisas, o sistema pretende também identificar os fatores limitantes, por exemplo, a falta de horas de frio ou a falta de precipitação, que condicionam a viabilidade de determinadas culturas. Assim, mais do que um exercício técnico, o objetivo é disponibilizar uma ferramenta prática e acessível, que possa apoiar o planeamento agrícola, contribuir para uma gestão mais sustentável dos recursos e servir de base para futuras aplicações de agricultura de precisão em Portugal.

2. Materiais e métodos

Os métodos utilizados neste trabalho centraram-se na construção de um conjunto de dados robusto e representativo das condições edafoclimáticas de Portugal continental, que serviu de base ao desenvolvimento do modelo preditivo e da aplicação interativa disponível em <https://github.com/RubenRomeroTorrado/Croptimum>. Nesse repositório encontra-se também o conjunto de dados, sendo necessário pedir acesso ao autor.

2.1 Recolha e descrição dos dados

A construção do modelo baseou-se num conjunto de dados edafoclimáticos de 158 localizações em Portugal continental, representadas na figura 1, que integraram informação climática e pedológica. A abordagem inicial para a escolha dos pontos consistiu no cruzamento de um mapa climático de Portugal continental com a de um mapa edáfico obtendo um mapa final que continha as diferentes combinações possíveis dos principais tipos de solo e clima, expressos em percentagens da área do território total. Começou-se por extrair dados de 71 localizações dos principais núcleos demográficos do país, mas a fraca performance dos modelos iniciais levou a recolher mais dados, abrangendo uma maior área representada. A escolha das localizações adicionais fez-se de forma a representar uniformemente o máximo de território nacional possível e também a descoberta da carta de solos e de aptidão da terra de Entre Douro e Minho levou ao aproveitamento destes para a sua inclusão no conjunto de dados.

Para os dados climáticos, a fonte operacional foi o Google Earth Engine (GEE). A partir desta plataforma acederam-se às bases de dados originais do ERA5-Land com uma resolução horária e espacial aproximada de 9 km, da instituição ECMWF e CHIRPS Daily com resolução diária e espacial de cerca de 5 km, da instituição Climate Hazards Center, UCSB, das quais se extraíram séries horárias e diárias para o período 1991–2025, em formato CSV, mediante a delimitação prévia de cada parcela de interesse. No Quadro 1 identifica os conjuntos de dados usados como fonte para cada variável.

A criação do conjunto de dados envolveu a utilização parcial de uma ferramenta de extração e transformação de dados incluída no Microsoft Excel, para processar os dados horários, convertendo-os em valores diários para depois serem usadas as várias variáveis climáticas identificadas no Quadro 2. Ou seja, o código do Power Query permitiu contabilizar para todos os dias, as horas de frio (inferior a 7 graus Celsius) e as horas de geada (inferior a 0 graus) a partir das temperaturas, hora a hora. Também com este código foram extraídas as outras variáveis diárias como precipitação e temperaturas mínimas, máximas e médias. Com a ajuda das tabelas dinâmicas do excel, estes valores foram depois agregados em médias mensais e, posteriormente, anuais, sendo que algumas variáveis, como as horas de frio, os dias de geada

e a precipitação, passaram a ser tratadas como médias mensais sazonais, devido a haver diferentes períodos críticos destas variáveis para as culturas. É importante referir que após a extração dos dados horários do GEE, foi necessário o processamento individual de um csv para cada localização. O próprio google earth engine limitava a extração de dados a apenas 5 localizações de cada vez, que demoravam várias horas. Estes ficheiros, que foram passados a excel, continham milhares de linhas e eram muito pesados, requerendo muito tempo para processar e afetando enormemente o rendimento do pc.

Este trabalho partiu de uma base composta por 7 documentos Excel da Consulai, contendo dados climáticos diários de 7 localizações, os quais já tinham sido processados pelo código do power query e continham todas as variáveis de interesse.

A escolha desta abordagem deveu-se a vários fatores. Era o método utilizado pela empresa consultora Consulai, servindo como referência em caso de dúvida, mas também já tinha acesso aos dados mencionados de algumas localizações e aos comandos de processamento em Power Query, fornecidos pela mesma empresa. Esta resultou ser a forma mais intuitiva de calcular variáveis agroclimáticas específicas, como as horas de frio e os dias de geada.



Figura 1 Representação da minha autoria das localizações presentes no conjunto de dados. Estas foram escolhidas inicialmente com base na representatividade setorial de cada zona edafoclimática.

Quadro 1 Conjuntos de dados usados como fonte para cada variável

Conjunto de dados	Variáveis extraídas	Unidades	tipo de dados
ERA5-LAND (ECMWF/ERA5_LAND/HOURLY)	Temperatura do ar a 2 metros	Graus Celcius	numérica contínua
	Precipitação total horária	mm	numérica contínua
CHIRPS (UCSB-CHG/CHIRPS/DAILY)	Precipitação acumulada diariamente	mm	numérica contínua

Quadro 2 -Resumo das Variáveis Climáticas e descrição derivadas das bases de dados originais ERA5-Land (Copernicus Climate Data Store/ECMWF) e CHIRPS Daily (Climate Hazards Center, UCSB).

Variável	Descrição	Unidades	Tipo de dados
Suma_Dias_Helada_Feb_May	Dias de geada médio entre Fev-Maio	dias	numérica contínua
Suma_Dias_Helada_Mar_May	Dias de geada médio entre Mar-Maio	dias	numérica contínua
Suma_Dias_Helada_Abr_May	Dias de geada médio entre Abr-Maio	dias	numérica contínua
Suma_Horas_Frio_Oct_Feb	Horas de frio média entre Outubro a fevereiro	horas	numérica contínua
Suma_Horas_Frio_Oct_Mar	Horas de frio média entre Outubro a Março	horas	numérica contínua
Suma_Precipitacion_Oct_Abr	Precipitação total média entre outubro e abril	milímetros	numérica contínua
Mínimo_de_Media_de_Temperatura_Minima(°C)	Temperatura mínima diária média no mês mais frio	Graus celsius	numérica contínua
Média de Média de Temperatura Mínima (°C)	Temperatura mínima diária média anual	Graus celsius	numérica contínua
Média de Média de Temperatura Máxima (°C)	Temperatura máxima diária média anual	Graus celsius	numérica contínua
Maximo_de_Media_de_Temperatura_Maxima	Temperatura máxima diária média no mês mais quente	Graus celsius	numérica contínua

No que diz respeito aos dados do solo, a recolha foi feita principalmente através da consulta das cartas digitais de solos de Portugal do Norte, Centro e Sul no QGIS (baseadas no Serviço de Reconhecimento e de Ordenamento Agrário, 1970, 1973a , 1973b). Para cada conjunto de coordenadas identificou-se o tipo de solo presente. Através de um Excel conversor da Consulai, cada tipo de solo foi depois traduzido em parâmetros específicos, como o pH e as percentagens de areia, limo e argila, tendo a variável "espessura" sido considerada inicialmente, mas posteriormente excluída. Para complementar e aumentar a densidade de dados de explorações no Noroeste de Portugal, foi também utilizada a carta de solos e de aptidão da terra de Entre Douro e Minho, da Direção Regional correspondente, em formato PDF (Direção Regional de Agricultura de Entre-Douro e Minho [DRAEDM], 1995.). Apesar de ser um método menos prático, foi considerado o mais preciso para determinar o tipo de solo em cada localização, permitindo ainda tirar partido do Excel conversor disponibilizado pela Consulai. Este excel contém informação de todos os tipos de solo no território português e os seus dados edáficos associados. No Quadro 3 apresenta-se o total de variáveis edáficas usadas no conjunto de dados final.

Quadro 3. Resumo das variáveis edáficas e descrição retiradas das cartas do solo do Serviço de Reconhecimento e de Ordenamento Agrário.

Variável	Descrição	Unidades	tipo de dados
Local	Nome da Localização	---	texto
Lat	Latitude	Graus decimais	numérica contínua
Lon	Longitude	Graus decimais	numérica contínua
Tiposolo	Tipo de solo	---	texto
Areia	Percentagem de Areia	Percentagem	numérica contínua
Limo	Percentagem de Limo	Percentagem	numérica contínua
Argila	Percentagem de Argila	Percentagem	Numérica contínua

Uma vez obtidos os dados de todas as localizações, juntaram-se manualmente no mesmo excel por localização. O fluxo de trabalho para a obtenção do dataset final pode ser melhor compreendido no esquema da figura 2.

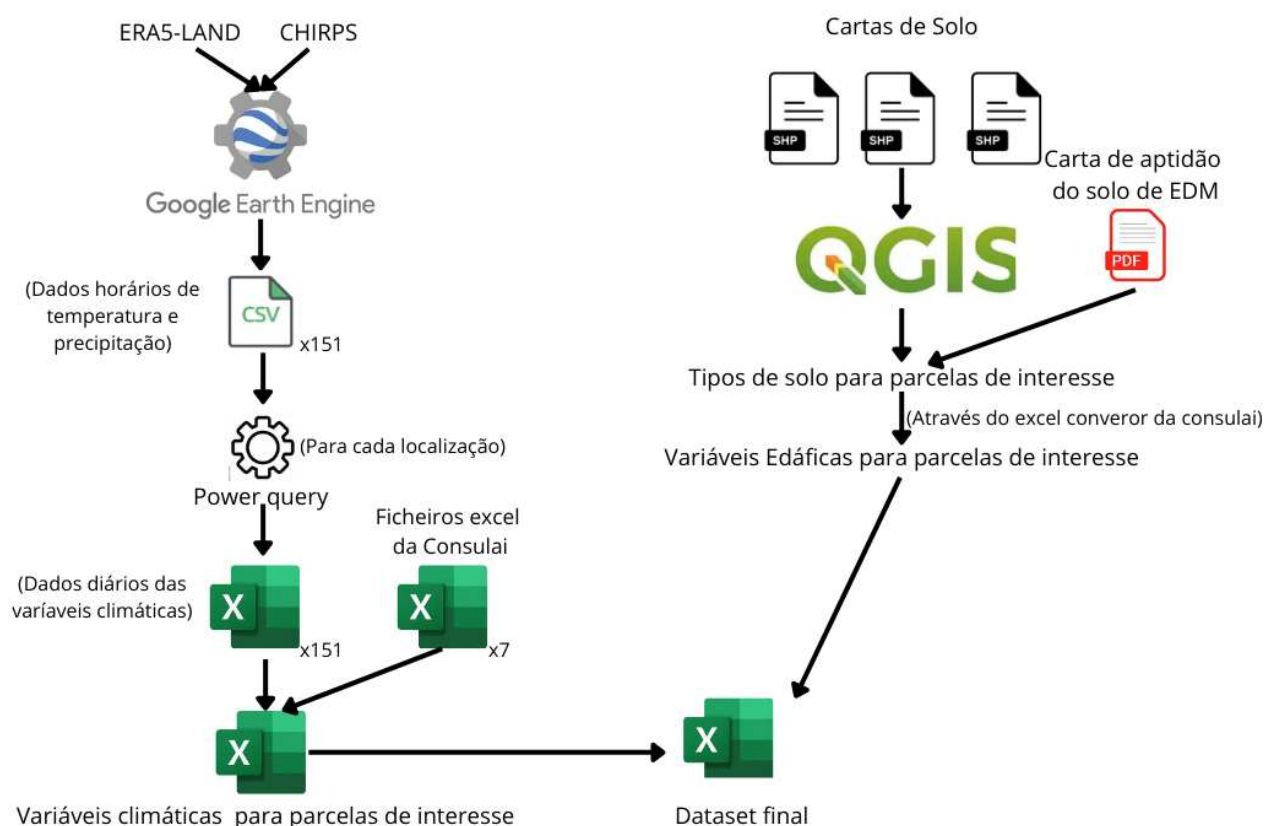


Figura 2 Esquema da minha autoria do fluxo de trabalho para a obtenção do dataset final.

Deste modo, o modelo final beneficiou de uma compilação robusta e multifonte, assegurando a representatividade tanto das condições climáticas como das características do solo nas 158 localizações.

2.2. Desenvolvimento do modelo Random Forest

Numa abordagem habitual para a criação de um modelo para suporte do sistema de apoio à decisão, seria desenvolvido um modelo para cada uma das variáveis de interesse apresentadas nos quadros 2 e 3. Por exemplo, um modelo para a estimação das horas de frio, um modelo para a estimação da percentagem de argila, etc... Neste caso, cada uma das variáveis dos quadros desempenharia o papel de *label* em cada modelo ML destinado à sua previsão, enquanto que as coordenadas seriam as *features*. No presente estudo, uma vez que a maioria das variáveis têm elevada correlação entre si, a aplicação de uma modelação com multi-output pode ser favorável, como se explicará adiante. Neste caso, as variáveis edafo-climáticas representam também o papel de features no desenvolvimento do modelo.

2.2.1 Pré-processamento das variáveis

As variáveis predictoras (X) latitude e longitude foram padronizadas utilizando um StandardScaler para centrar as suas médias em 0 e escalar os seus desvios padrão para 1. Este passo é crucial para algoritmos que dependem da distância entre pontos, mas não foi aplicado à variável alvo (y), uma vez que o Random Forest é robusto à escala das características.

2.2.2 A Base Teórica do Random Forest

A escolha do algoritmo random forest regressor para este projeto é altamente apropriada e fundamentada na literatura técnica. O Random Forest é um método de aprendizagem em conjunto (ensemble learning) que constroi um grande número de árvores de decisão durante a fase de treino e combina as suas previsões para obter um resultado. (Anco-Valdivia et al., 2025). Esta abordagem tem várias vantagens, que se revelam particularmente valiosas para conjuntos de dados de pequena dimensão, como o que foi utilizado neste estudo, (Joy, 2025).

A natureza agregadora do algoritmo torna-o intrinsecamente robusto contra o sobreajuste, um risco significativo quando o volume de dados é limitado. Ao treinar cada árvore com uma

subamostra aleatória dos dados e a um subconjunto aleatório de características, o Random Forest garante que as previsões agregadas sejam menos sensíveis a ruídos e a outliers no conjunto de dados. Além disso, o algoritmo é excepcionalmente eficaz a capturar interações complexas e não lineares entre as variáveis, uma característica essencial para modelar as relações intrincadas que existem entre o clima e o solo. A sua capacidade de fornecer uma métrica de importância das características adiciona um valor interpretativo, permitindo identificar quais as variáveis preditoras que mais influenciam o modelo (Chege, 2025).

2.2.3 As Vantagens Únicas da Regressão Multi-Output

A adoção de um modelo de regressão multi-output, em detrimento de um conjunto de modelos independentes de regressão univariada, constitui um elemento conceptual distintivo deste trabalho. A regressão multi-output é particularmente adequada a problemas em que se pretende prever simultaneamente várias variáveis contínuas potencialmente correlacionadas entre si. Ao contrário da abordagem tradicional, que treina um modelo distinto para cada variável de destino, um modelo multi-output aprende uma função vetorial do tipo:

$$f : R^p \rightarrow R^m,$$

onde p representa o número de *features* e m o número de *labels*.

No Multi Output RandomForest Regressor da scikit-learn, cada árvore do ensemble é treinada de forma a minimizar uma medida agregada de impureza, a soma dos erros quadráticos médios das m variáveis de destino:

$$\Delta I = \sum_{j=1}^m \Delta MSE_j$$

Durante o processo de construção da árvore, o critério de divisão dos nós é avaliado em função desta redução conjunta da impureza. Isto significa que cada split deve contribuir positivamente, em média, para a diminuição do erro total do vetor de respostas, e não apenas para uma variável isolada. Este mecanismo atua como uma forma de regularização estrutural, dado que o

modelo é desencorajado a produzir divisões excessivamente específicas para um único alvo e é incentivado a identificar padrões estruturais partilhados entre as variáveis.

Como consequência, o ensemble resultante tende a apresentar maior capacidade de generalização, sobretudo em contextos onde as variáveis alvo exibem dependências ou covariações relevantes. Além disso, a abordagem multi-output reduz redundâncias computacionais, na medida em que treina um único modelo em vez de vários modelos independentes, beneficiando também de uma representação interna mais coerente do espaço de resposta.

Contudo, como a otimização é realizada sobre a impureza agregada, é expectável que o desempenho de cada variável individual, quando avaliada isoladamente, possa ser ligeiramente inferior ao obtido com modelos single-output especialmente afinados para essa variável. Por esse motivo, a comparação empírica entre ambas as abordagens é necessária para determinar qual delas apresenta melhor compromisso entre desempenho individual e desempenho global.

2.2.4 Otimização dos Hiperparâmetros com GridsearchCV

Para encontrar a configuração ótima do modelo e maximizar a sua capacidade de generalização, foi utilizada a técnica de validação cruzada com pesquisa em grelha (GridSearchCV). Dada a limitada quantidade de dados, foi usado um 5-Fold Cross-Validation dentro do conjunto de treino. Isto permite um uso mais eficiente dos dados para ajustar e validar os hiperparâmetros de forma robusta.

Os hiperparâmetros otimizados através do GridSearchCV incluíram o número de árvores de decisão (estimadores) na floresta, a profundidade máxima de cada árvore na floresta, o número mínimo de amostras que são necessárias para dividir um nó interno, o número mínimo de amostras que devem estar em um nó folha (terminal) e o número de *features* a considerar ao procurar a melhor divisão. Os seus possíveis valores encontram-se no quadro 4.

Quadro 4 Possíveis valores para otimização de hiperparâmetros com GridsearchCV

Hiperparâmetro	Possíveis valores
n_estimators	(50, 100, 200, 300)
max_depth	(5, 10, 15, None)
min_samples_split	(2, 5, 10)
min_samples_leaf	(1, 2, 4)
estimator__max_features	(0,5 , 'sqrt', 'log2', None)

2.2.5 Treino e métricas de avaliação

Para avaliar o desempenho final do modelo com os melhores hiperparâmetros encontrados pelo GridSearchCV, foi realizada uma divisão do conjunto de dados para validação, reservando 20% dos dados (32 amostras) para teste (*test set*) e utilizando os 80% restantes (126 amostras) para o processo de treino e validação cruzada. O modelo final foi treinado sobre os 80% completos com os melhores hiperparâmetros e foi avaliado de forma independente e única com os 20% de teste, que não foram usados em nenhum momento durante o ajuste.

Para avaliar o desempenho da abordagem do modelo multi-output e dos modelos individuais foram calculadas métricas individuais por variável e métricas globais das abordagens. Para as métricas das variáveis individuais calculou-se a raiz do erro médio quadrado (RMSE) e o coeficiente de determinação r^2 . Para as métricas globais calculou-se o erro médio quadrado (MSE) e o (r^2).

Para cada variável j , o erro quadrático médio (MSE) foi definido como:

$$MSE_j = \frac{1}{n} \sum_{i=1}^n (y_{ij} - Y_{ij})^2,$$

sendo y_{ij} o valor real da variável j na amostra i e Y_{ij} o valor previsto da variável j na amostra i

A raiz do erro médio quadrado (RMSE) dada por:

$$RMSE_j = \sqrt{MSE_j}$$

O coeficiente de determinação individual foi expressado como:

$$R_j^2 = 1 - \frac{\sum_{i=1}^n (y_{ij} - Y_{ij})^2}{\sum_{i=1}^n (y_{ij} - \gamma_j)^2}, \text{ onde:}$$

$$\gamma_j = \frac{1}{n} \sum_{i=1}^n y_{ij}$$

Adicionalmente, foi calculado um MSE global, correspondente à média aritmética dos MSE das m variáveis dado por:

$$MSE_{global} = \frac{1}{m} \sum_{j=1}^m MSE_j$$

Também foi calculado um R^2 global, correspondente à média ponderada dos R^2 individuais das m variáveis, relativamente ao peso da sua variância definido por:

$$R_{global}^2 = 1 - \frac{\sum_{j=1}^m \sum_{i=1}^n (y_{ij} - Y_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \gamma_j)^2}$$

2.3 Comparação dos valores previstos e cálculo do índice de aptidão

A comparação dos valores previstos com os requisitos edafoclimáticos de cada cultura, representados nos quadros A2 a A9 fez-se no programa, seguindo um algoritmo desenvolvido neste trabalho. O algoritmo estabeleceu pontuações para cada cultura e requisito climático ou edáfico, atribuindo o valor 1, se o valor previsto se encontra dentro do intervalo ideal, o valor 0,5 se o valor se encontra ligeiramente abaixo ou acima desse intervalo, ou 0 se o valor previsto se encontra bastante abaixo ou acima do intervalo ideal. Para cada cultura foram criadas 8 pontuações edafoclimáticas, as quais incluem ph, geadas, horas de frio, precipitação, temperatura normal, temperaturas extremas, textura e espessura do solo. Certas pontuações

exigiram mais de um valor previsto, como por exemplo, para a pontuação da textura terem sido necessários os valores de areia, limo e argila, ou para a pontuação das geadas terem sido utilizados os valores das 3 variáveis que prevêm a soma de dias de geada. É importante referir que para a atribuição da pontuação das texturas foram definidos polígonos texturais retirados do diagrama triangular das texturas da USDA (Soil Survey Division Staff, 1993).

Para o cálculo do índice de aptidão procedeu-se à média aritmética das pontuações referidas anteriormente e estas foram apresentadas por ordem de maior a menor pontuação global. Foi decidido que as 8 condicionantes edafoclimáticas teriam o mesmo peso para determinar o índice de aptidão.

3. Resultados

Vamos apresentar os resultados da análise exploratória realizada sobre as 158 localizações incluídas no estudo, que combinam informação climática e de solo recolhida em diferentes regiões de Portugal continental. Esta análise teve como objetivo conhecer melhor a distribuição e o comportamento das variáveis antes de avançar para a modelação, permitindo identificar padrões, relações e possíveis anomalias nos dados.

3.1 Análise exploratória de dados (EDA)

Com base nos resultados estatísticos, observa-se que os 158 locais analisados, apesar de estarem numa área geograficamente concentrada em Portugal, apresentam uma alta variabilidade nas condições de solo e clima. Esta observação está em linha com o conceito de "terroir", onde fatores locais modulam as condições regionais, criando microclimas e ambientes únicos para a agricultura (Jones & Alves, 2012).

Isto sugere a existência de múltiplos microclimas e terroirs distintos no país, um facto amplamente reconhecido na literatura científica sobre a agricultura portuguesa, especialmente na viticultura. Este facto também está presente na agroecologia e é suportado por autores como Cardoso et al. (2019) que defendem que no noroeste ibérico são identificadas unidades agroecológicas distintas, revelando que mesmo regiões geográficas adjacentes abrigam terroirs climáticos e edáficos diferenciados.

O clima representado pela análise estatística é temperado, dado que os invernos são frios, com temperaturas mínimas médias presentes no quadro 5 que podem descer abaixo de 0°C (mínimo de -2.77°C), e os verões são quentes, mas não extremos (média das máximas a rondar os 30°C). A sua principal característica é a elevada variabilidade espacial, fortemente influenciada pela topografia e pela proximidade ao Atlântico. Um estudo de Jones and Alves (2012) sobre as regiões vitícolas portuguesas destaca precisamente esta diversidade, afirmando que "a

topografia complexa de Portugal e a sua linha costeira resultam em numerosos mesoclimas e microclimas, oferecendo uma vasta gama de condições para a viticultura" .

A sua análise confirma que mesmo em distâncias curtas, as variações de temperatura e insolação podem ser muito significativas, o que valida a grande amplitude observada nos dados de temperatura.

Quadro 5 Resumo estatístico das Temperaturas médias mínimas e máximas.

Métrica	Minimo_de_Media_de Temperatura Minima	Maximo_de_Media_de Temperatura Maxima	Média de Média de Temperatura Mínima (°C)	Média de Média de Temperatura Máxima (°C)
count	158	158	158	158
mean	2.74	29.80	10.24	20.70
std	2.79	2.68	2.02	1.89
min	-2.77	23.63	6.10	17.58
25%	0.46	27.43	8.68	19.16
50%	2.58	29.76	9.89	19.86
75%	4.80	32.39	12.12	22.70
max	12.19	34.62	16.53	24.36

O número de dias de geada nos dados revela que em média existem 12 dias de geada entre Fevereiro e Maio, chegando a 40 dias em alguns locais, apresenta no quadro 6. O risco de geadas tardias é um dos principais fatores de risco climático na agricultura em Portugal. A análise de Reis e de Melo-Abreu (2017) sobre os impactos das alterações climáticas na agricultura portuguesa corrobora este ponto, identificando as geadas de primavera como uma ameaça contínua, mesmo num cenário de aquecimento global, devido à maior irregularidade

dos padrões climáticos e à possibilidade de florações mais precoces e, conseqüentemente, mais expostas a eventos de frio tardio.

Quadro 6 Resumo estatístico dos dias de geada

Métrica	Suma_Dias_Helada_Feb_May	Suma_Dias_Helada_Mar_May	Suma_Dias_Helada_Abr_May
count	158.00	158.00	158.00
mean	12.36	4.72	1.30
std	9.43	4.58	1.62
min	0.00	0.00	0.00
25%	3.66	0.65	0.00
50%	10.50	3.43	0.67
75%	19.26	7.75	2.30
max	40.58	21.00	7.88

No Quadro 7 verifica-se um número significativo de horas de frio no inverno (média de 935 a 1077 horas), o que é benéfico para culturas que necessitam de dormência, como macieiras e muitas árvores de fruto.

Quadro 7 Resumo estatístico das horas de frio

Métrica	Suma_Horas_Frio_Oct_Feb	Suma_Horas_Frio_Oct_Mar	Suma_Precipitacion_Oct_Abr
count	158.00	158.00	158.00
mean	935.19	1077.93	772.40
std	437.17	523.21	342.41
min	8.26	8.91	347.60
25%	541.24	597.46	463.58
50%	913.63	1054.67	670.24
75%	1292.13	1508.82	1044.55
max	2006.37	2365.72	1490.43

No quadro 8 podemos conferir a enorme variação na precipitação (de 347 mm a 1490 mm) reflete os padrões de distribuição da chuva em Portugal Continental, caracterizados por um gradiente acentuado de noroeste (mais húmido) para sudeais ste (mseco). Santos et al. (2002), num estudo sobre os recursos hídricos, explicam que esta variabilidade está ligada à interação entre a circulação atmosférica atlântica e o relevo do país, o que justifica a necessidade de diferentes estratégias de gestão da água, incluindo a rega em zonas mais secas.

Quadro 8 Resumo estatístico da precipitação.

Métrica	Suma_Precipitacion_Oct_Abr
count	158.00
mean	772.40
std	342.41
min	347.60
25%	463.58
50%	670.24
75%	1044.55
max	1490.43

O Quadro 9 apresenta que em média, os solos são maioritariamente arenosos (68% de areia), com menores percentagens de limo (18%) e argila (14%). Solos arenosos geralmente têm boa drenagem, mas menor capacidade de reter água e nutrientes.

Os solos tendem a ser ácidos a ligeiramente ácidos (média de pH de 5.6), com 75% das amostras a apresentar um pH inferior a 6.0. Esta variável pode ser mais facilmente analisada na figura 3, que contém os diagramas de extremos e quartis por grupo de variáveis.

A predominância de solos arenosos e de pH ácido é uma característica bem documentada de vastas áreas de Portugal, especialmente nas regiões de base granítica. Um estudo de caracterização de solos da Agência Portuguesa do Ambiente (APA) frequentemente aponta para a prevalência de Leptosolos e Cambissolos, muitos deles com textura arenosa ou franco-arenosa e reação ácida.

Quadro 9 Resumo estatístico das variáveis do solo.

Métrica	lat	lon	Areia_x	Limo_x	Argila_x	ph_x
count	158.00	158.00	158.00	158.00	158.00	158.00
mean	40.12	-7.98	67.93	17.58	14.49	5.60
std	1.40	0.56	16.87	9.41	9.68	0.85
min	37.02	-9.14	15.60	0.69	0.70	4.40
25%	38.99	-8.47	58.30	10.76	9.20	5.00
50%	40.32	-8.00	71.30	16.80	11.95	5.40
75%	41.40	-7.52	81.70	23.70	17.56	5.98
max	42.07	-6.27	98.22	47.00	60.70	8.20

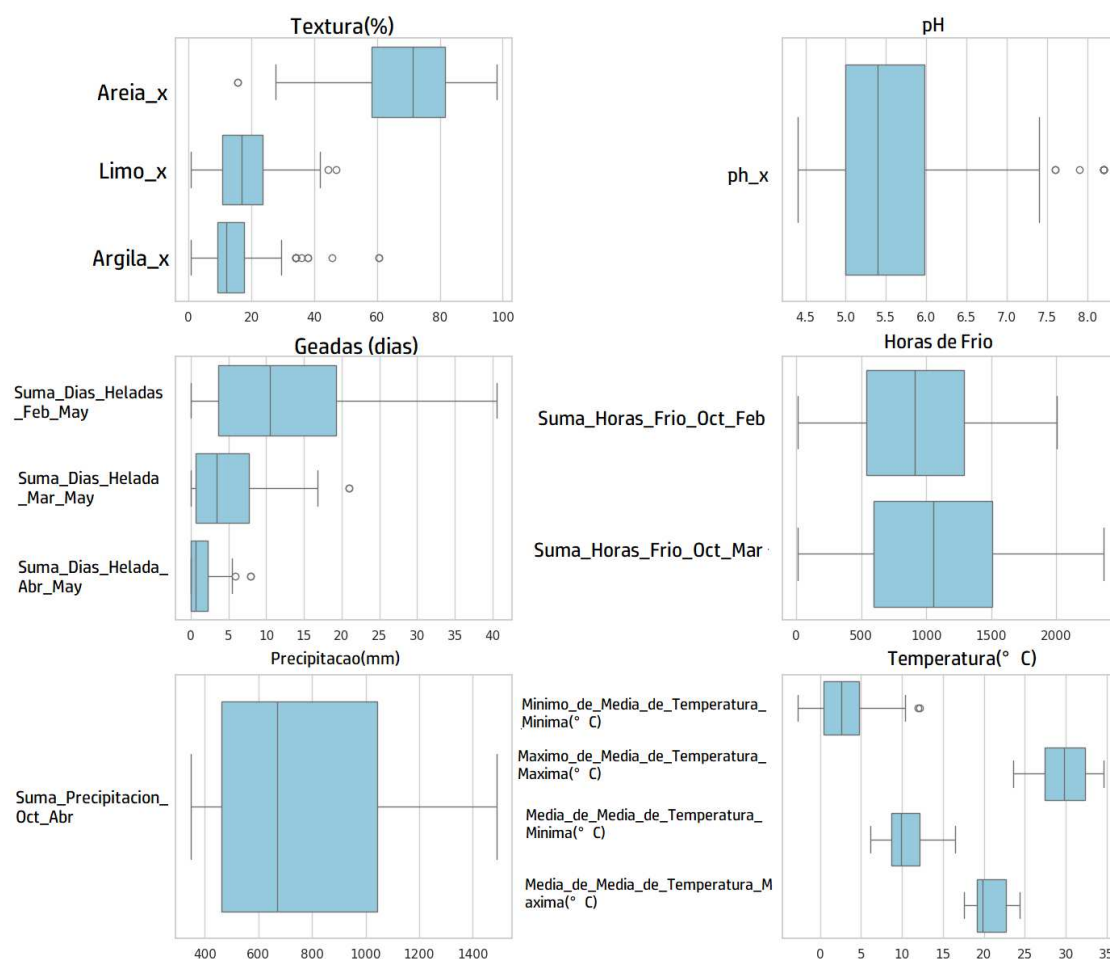


Figura 3. Diagramas de extremos e quartis das variáveis do Dataset agrupados por Textura, Geadas, Precipitação, ph, horas de frio e temperatura. Permite identificar o intervalo onde se encontram os valores de cada variável, a mediana, a amplitude interquartil, os extremos e os outliers.

A Análise Exploratória de Dados (AED) facilita a identificação de relações entre as variáveis. Em relação aos gráficos da matriz de diagramas de dispersão do clima presentes na figura 4, podemos afirmar que a variável da precipitação e da temperatura máxima média no mês mais quente têm um comportamento mais independente em relação às outras variáveis. Na matriz de

correlação presente na figura 5, estas duas variáveis também têm, na maioria, baixos coeficientes de determinação com as outras variáveis em estudo.

No entanto verifica-se que a correlação linear entre as horas de frio e os dias de geada é forte de forma positiva entre elas e de forma negativa quando correlacionadas com as variáveis de temperatura. Isto faz sentido dado que nas localizações onde existem mais horas de frio, geralmente existem mais dias de geada, mas menor é a temperatura média mínima e em menor grau a máxima.

Alguns gráficos de dispersão mostram uma relação clara, mas não uma linha reta, indicando uma relação não linear. A distribuição pode parecer mais curva ou ter um limite onde o aumento de uma variável não causa uma variação proporcional na outra, como é o caso da relação das horas de frio e dias de geada com a temperatura máxima média.

Em relação à Análise da Distribuição Individual na diagonal, muitas das variáveis apresentam distribuições Assimétricas especialmente as somas de geada e precipitação, estas parecem ser assimétricas à direita pois existe maior concentração de valores baixos com uma cauda longa em valores altos. Isso é comum para contagens ou acumulações de eventos como dias de geada e precipitação (Ye et al., 2018).

E temos também distribuições mais simétricas nas horas de frio e nas variáveis térmicas), exceto a temperatura máxima média, dado que estas variáveis tendem a ter distribuições mais “normais” em latitudes médias e altas (Zhang et al., 2022).

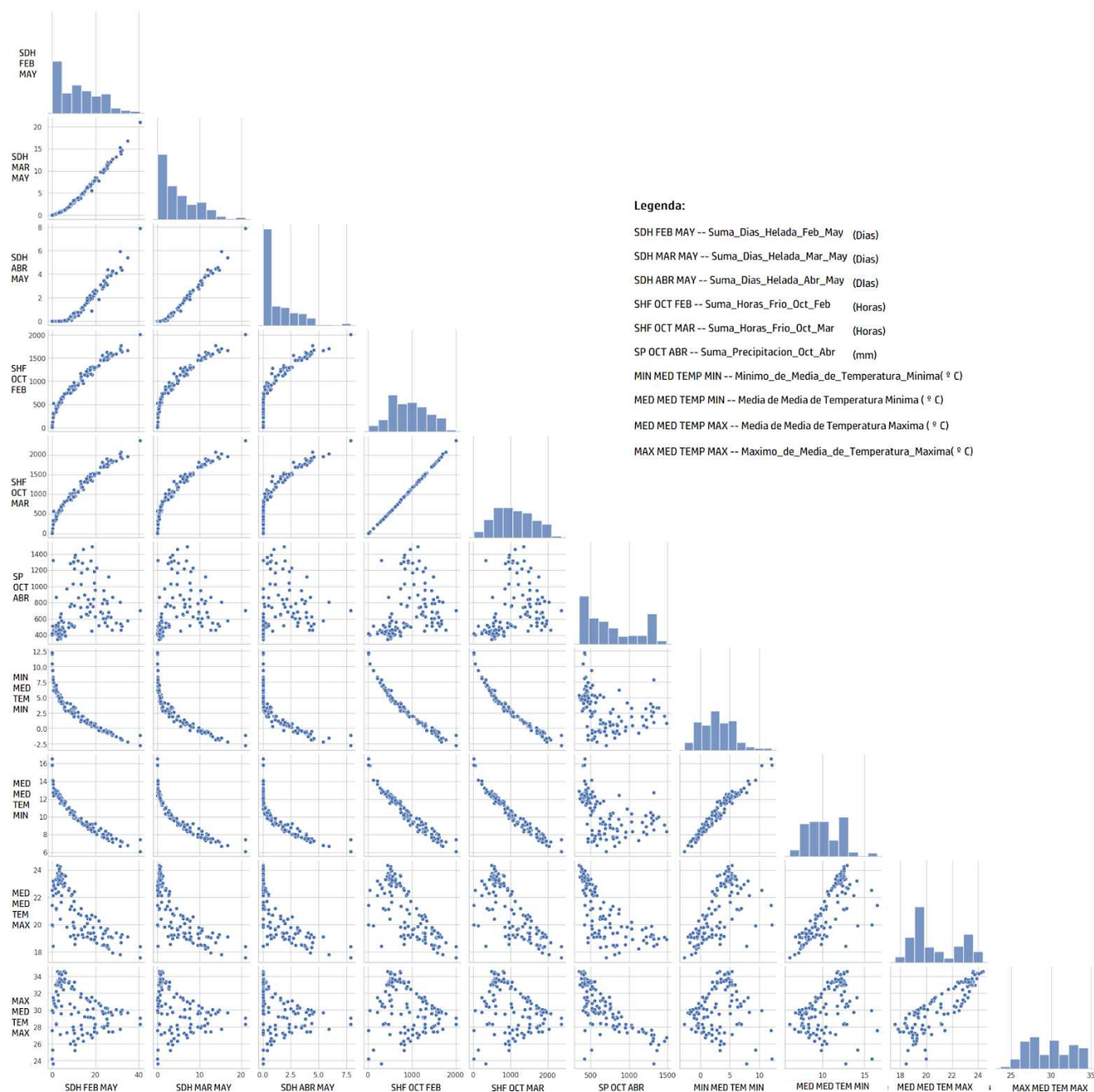


Figura 4 Matriz de diagramas de dispersão de variáveis climáticas. Observa-se de que forma as variáveis climáticas apresentam tendência entre si, a intensidade dessa tendência, a distribuição dos pontos e a simetria das distribuições. É notável a frequência de valores muito baixos nas variáveis de dias de geada e na precipitação por serem geralmente fenômenos mais raros.

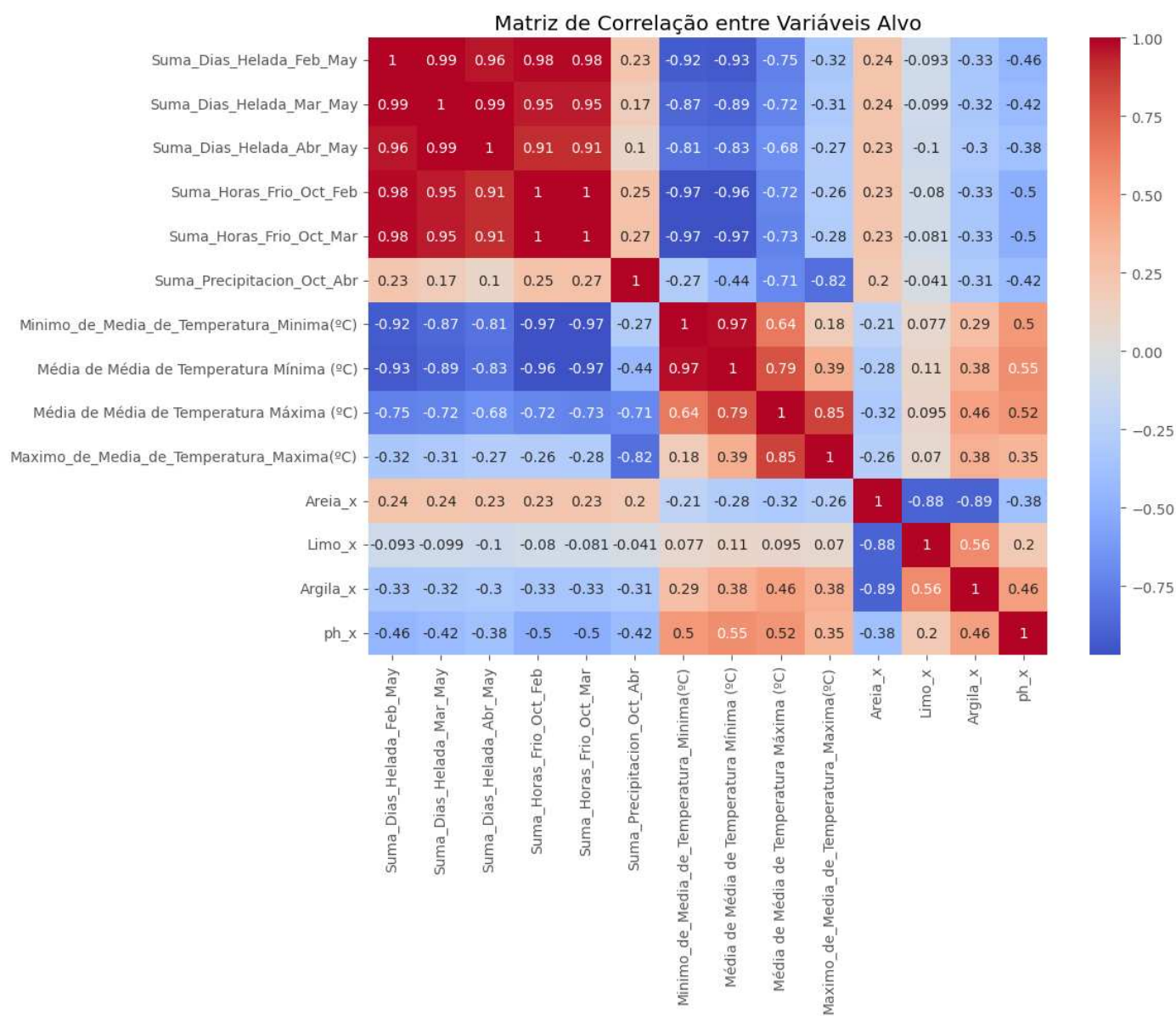


Figura 5 Matriz de correlação entre as variáveis de estudo. Esta matriz representa a correlação entre todas as variáveis, a vermelho quanto mais positiva e a azul quanto mais negativa.

A análise da matriz de variáveis do solo (figura 5) revelou padrões distintos tanto nas distribuições individuais das variáveis como nas suas relações. Os histogramas na diagonal mostram que a areia apresenta uma distribuição razoavelmente ampla, ainda que ligeiramente assimétrica para valores baixos, com a maioria das amostras a concentrar-se entre 60% e 90%. O limo, por sua vez, exhibe uma distribuição muito concentrada em valores baixos, com um pico acentuado e uma forte assimetria à direita. Padrão semelhante, mas ainda mais extremo, é observado na argila, que mostra um pico grande perto de 10%, indicando que muitas amostras possuem um teor muito baixo. Já o pH do solo distingue-se das frações granulométricas, apresentando uma distribuição concentrada em valores mais baixos, com um pico claro entre 5 e 6 mas existindo uma cauda com valores próximos de 8.

No que diz respeito às correlações, verifica-se que o aumento da percentagem de areia está fortemente associado à diminuição do limo e da argila, e vice-versa, uma relação inversa que confirma a condição fundamental de que a soma das três frações é sempre 100% (Mwendwa, 2022). Em contrapartida, o pH estabelece-se como a variável mais independente do conjunto, mostrando uma correlação fraca ou quase inexistente com os teores de areia, limo e argila, o que é visível na grande dispersão dos dados quando estas variáveis são comparadas.

A matriz de correlação ainda mostra uma fraca correlação entre as variáveis climáticas e edáficas o que é de esperar dado que a fração granulométrica e o pH do solo não é influenciado apenas por variáveis climáticas como a precipitação e temperatura média (Liu et al., 2022) mas também pela rocha mãe, idade do solo, topografia (Zhang et al., 2019), vegetação e fatores antropogénicos (NRCS, 2014).

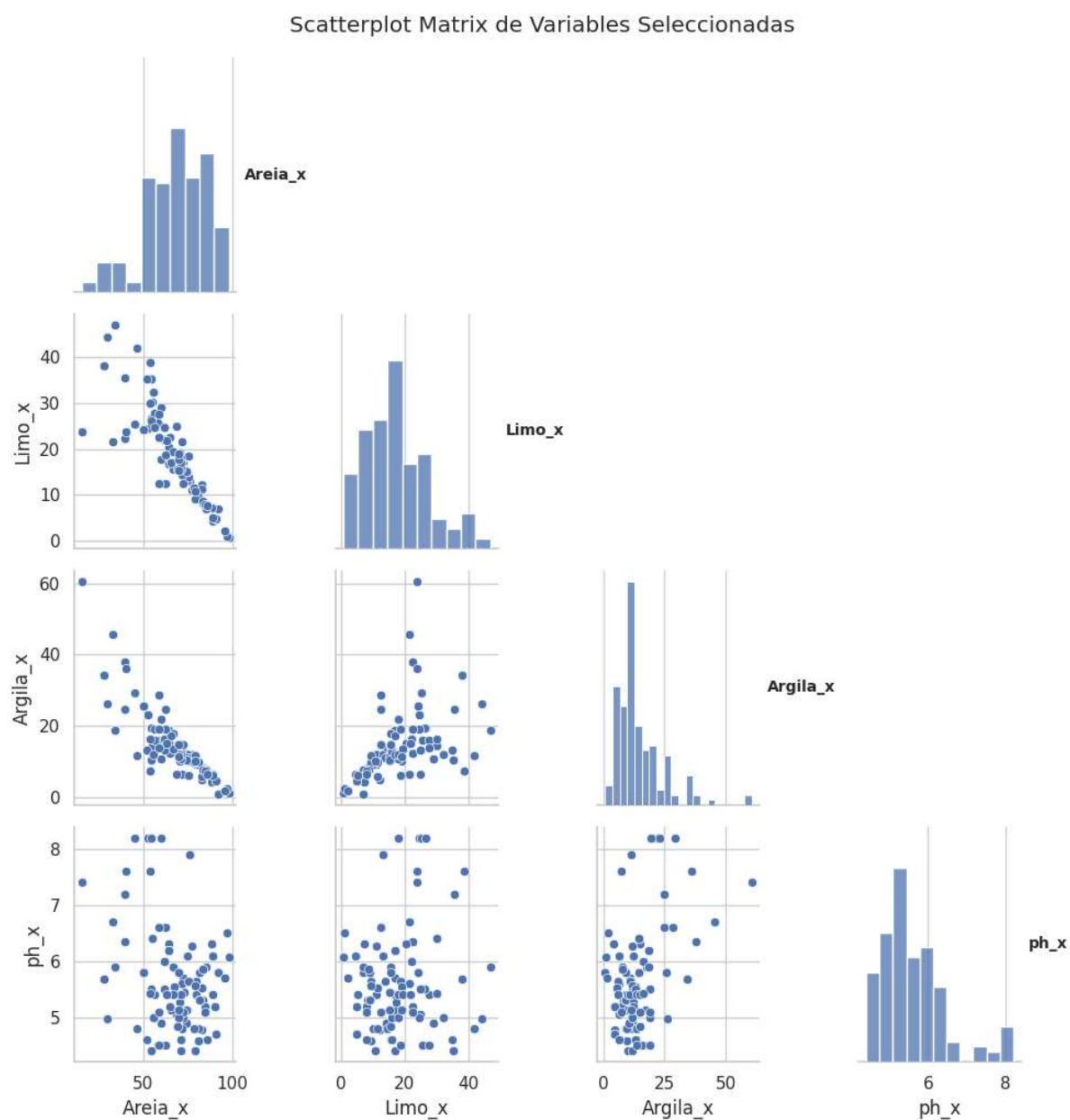


Figura 6 Matriz de diagramas de dispersão de variáveis do solo .Nesta figura observa-se de que forma as variáveis edáficas apresentam tendência entre si, a intensidade dessa tendência, a distribuição dos pontos e a simetria das distribuições. É notável a ausência de tendência entre o ph e as texturas.

3.2 Melhores valores de Hiperparâmetros encontrados

No Quadro 10 estão representados os melhores valores que foram encontrados para afinar os hiperparâmetros do modelo com GridsearchCV.

Quadro 10 Melhores valores encontrados para otimizar os hiperparâmetros com GridsearchCV.

Hiperparâmetro	Valor Otimizado
n_estimators	300
max_depth	10
min_samples_split	5
min_samples_leaf	1
estimator__max_features	None

3.3 Métricas de avaliação e gráficos comparativos

Esta secção tem como objetivo apresentar os principais resultados obtidos a partir da implementação do modelo multioutput Random Forest com a abordagem dos modelos individuais, otimizado através de GridSearchCV, para a previsão de variáveis edafoclimáticas. Serão exibidos os outputs principais, incluindo métricas de desempenho do modelo no Quadro 11 e A1. Na figura 7 é apresentado o valor da diferença da métrica de performance do modelo

multi-output com o modelo individual. Na figura 8 se apresentam os diagramas de dispersão entre os valores previstos e valores reais, para o conjunto de dados de teste

Quadro 11 Comparação global entre as duas abordagens, determinados a partir do conjunto de dados de teste.

Modelo usado	Métrica	Valor	Observação
Modelo multioutput Random Forest Regressor	R ² global (Coeficiente de Determinação)	0.95	Percentagem de variância da variável alvo explicada pelo modelo.
	Erro Quadrático Médio global (MSE)	2261.5 3	Média dos quadrados dos erros. Mais sensível a outliers.
14 modelos random forest regressors	R ² global (Coeficiente de Determinação)	0.99	Percentagem de variância da variável alvo explicada pelo modelo.
	Erro Quadrático Médio global (MSE)	734.45	Média dos quadrados dos erros. Mais sensível a outliers.

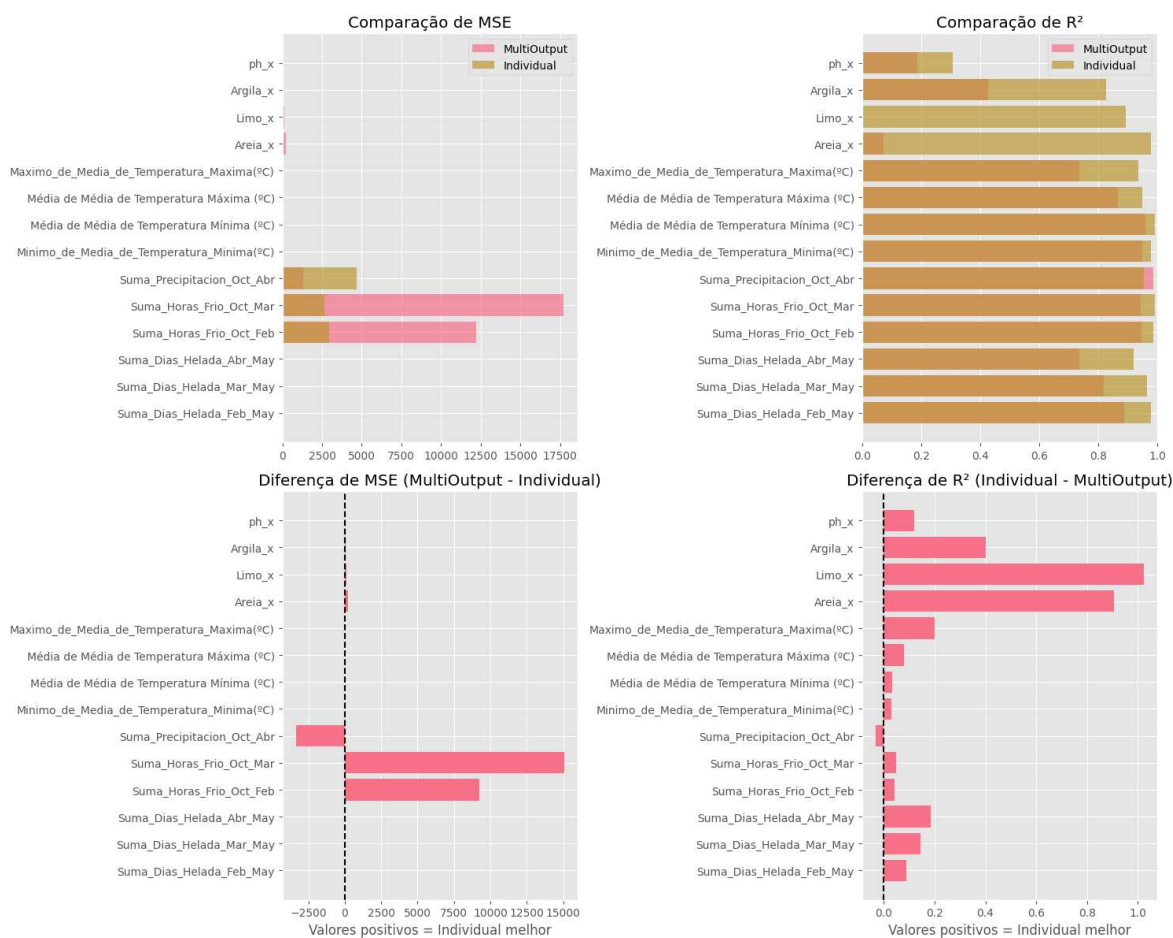


Figura 7 Comparação entre as métricas de desempenho entre as duas abordagens de modelação multi-output e modelos individuais.

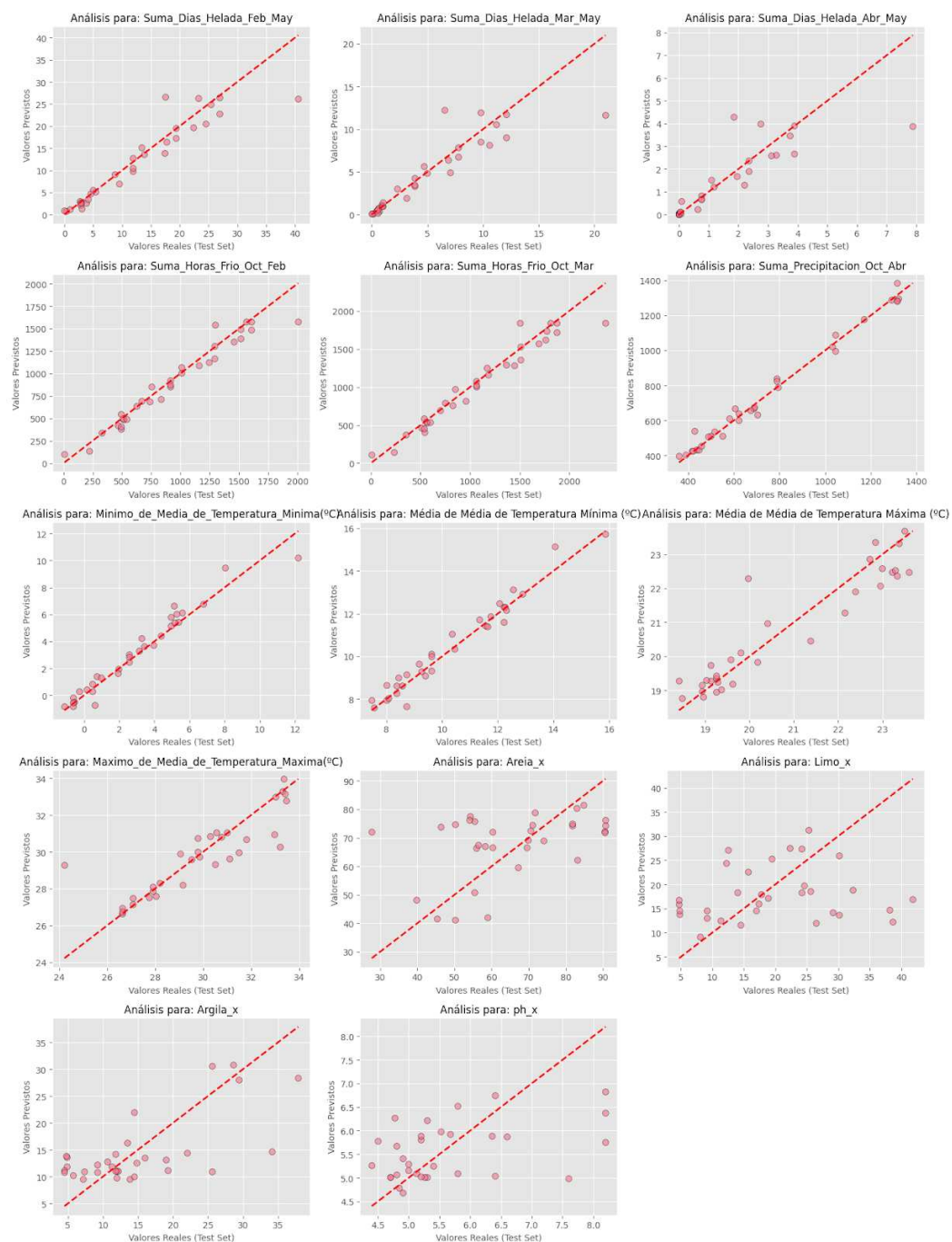


Figura 8. Representação dos valores previstos vs. Valores Reais no conjunto de dados de teste para o modelo multioutput. Nota-se o maior ajuste dos pontos à reta nas variáveis climáticas, indicando um erro mínimo.

De forma geral os valores previstos coincidem com os valores reais do conjunto de teste, ao estarem próximos da reta, nomeadamente para as variáveis climáticas. No entanto, nas variáveis edáficas existe uma maior distância dos pontos à reta, indicando um maior erro de previsão.

3.4 Manual de utilização do programa

Constitui um objectivo principal deste trabalho o desenvolvimento de um sistema de apoio à decisão baseado em expert knowledge que identifique as culturas de maior aptidão a uma localidade de interesse. O programa desenvolvido é um Sistema de Apoio à Decisão (DSS) georreferenciado que avalia a aptidão agronómica para oito culturas: oliveira (*Olea europaea*), amendoeira (*Prunus dulcis*), videira (*Vitis vinífera*), noqueira (*Juglans regia*), medronheiro (*Arbutus unedo*), pistácio (*Pistacia vera*), macieira (*Malus domestica*) e pereira (*Pyrus communis*). O sistema trabalha sobre previsões edafoclimáticas multivariadas e devolve uma ordenação da aptidão, através de pontuações por parâmetro e fatores limitantes, com visualização gráfica e histórico de consultas na sessão.

3.4.1 Estrutura

Apresenta-se de seguida a estrutura da interface gráfica e sequência típica de utilização da aplicação.

O programa contém um painel informativo e 4 menus/separadores:

- Separador 1- Mapa interativo (Figura 9):
 - Mapa com diferentes camadas espaciais (mapa normal, mapa topográfico, mapa de satélite)
 - Permite seleccionar a localização no mapa para obter as coordenadas do ponto ou introduzi-las manualmente
 - Permite introduzir inputs do utilizador como a espessura do solo e o sistema da cultura (Figura 10).
 - Guardar e avaliar o local para as diferentes culturas
- Painel esquerdo informativo:
 - Contém um minimapa que facilita a confirmação da localização
- Separador 2 – Resultados (Figura 11)
 - Resultados das variáveis previstas
 - *Ranking* de Culturas por Aptidão por ordem de maior pontuação global, apresentando as pontuações dos parâmetros avaliados
 - Visualizar as pontuações das culturas individualmente
 - Recomendação Final
- Separador 3 – Gráficos (Figura 12)
 - Visualização gráfica das pontuações

- o Produtividade Potencial Comparativa
- o Composição do Solo
- Separador 4 – Histórico (Figura 13)
 - o Registo das pesquisas temporário na sessão com apresentação dos resultados principais e possibilidade de carregar a pesquisa

3.4.2 Guia de utilização passo a passo

Passo 1: Selecionar a visualização do mapa preferido e clicar no ponto do mapa pretendido ou introduzir as coordenadas manualmente.



Figura 9 Separador 1 do sistema de apoio à decisão - Mapa interativo.

Passo 2: Introduzir a espessura do solo e o sistema de cultura pretendido

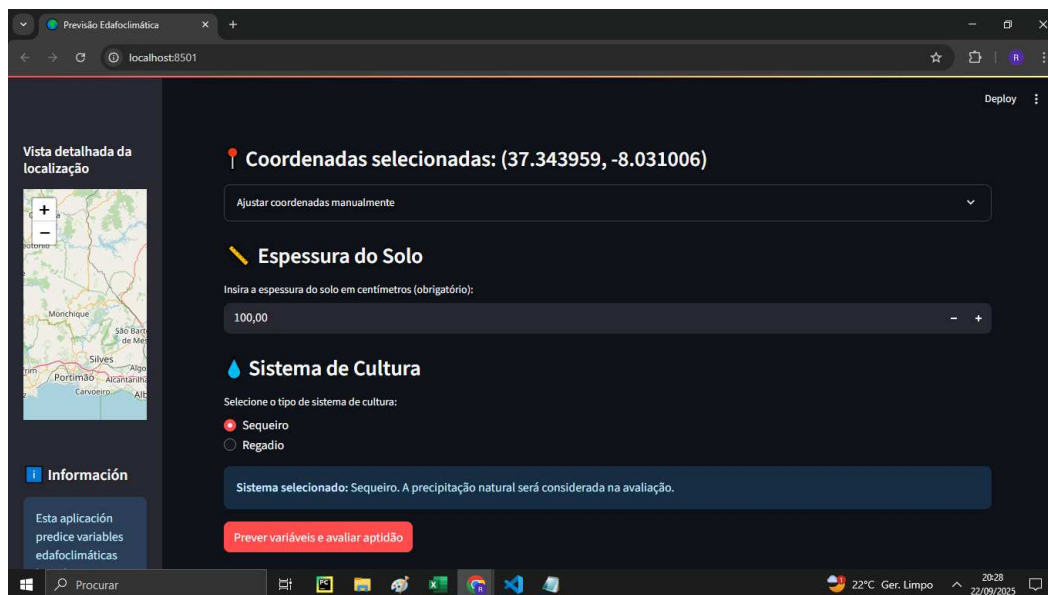


Figura 10 Separador 1 do sistema de apoio à decisão – Introdução de inputs

Passo 3: Clicar no botão de prever variáveis.

Passo 4: Abrir o separador 2 e deslizar para baixo para visualizar os resultados e o ranking das culturas. Opcional: Visualizar cultura em pormenor.

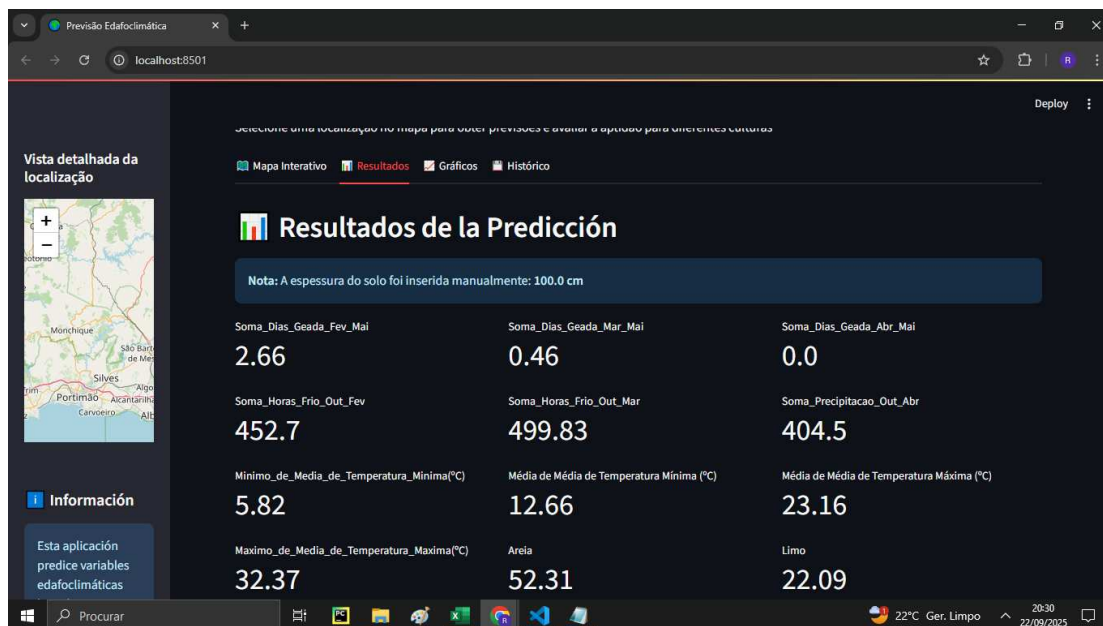


Figura 11 Separador 2 do sistema de apoio à decisão – Resultados

Passo 5: Fazer o mesmo no separador 3 para visualizar os gráficos

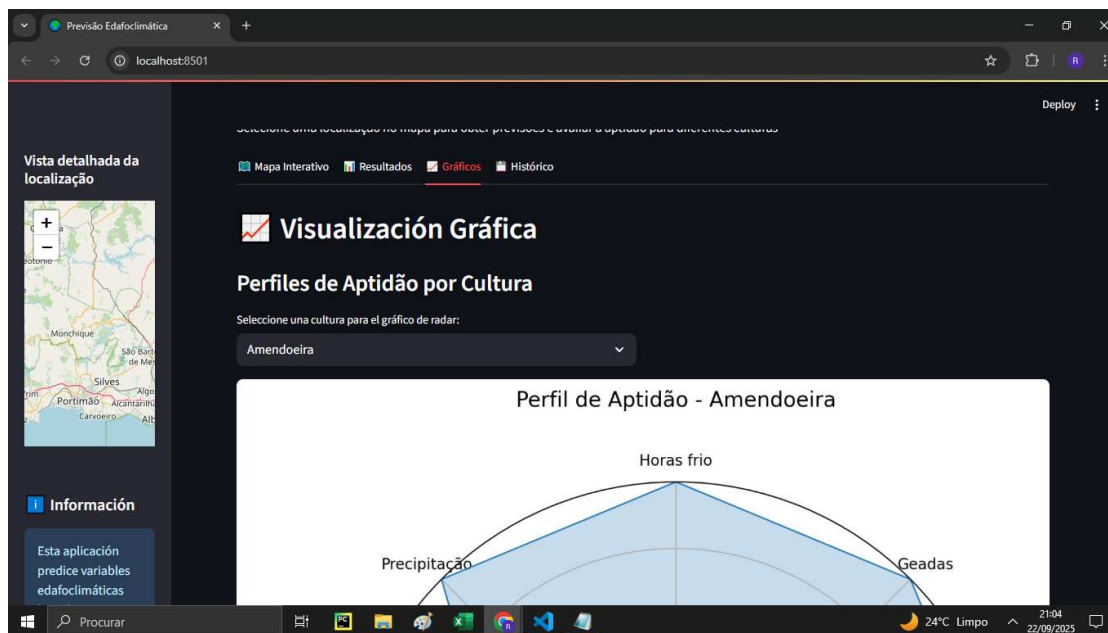


Figura 12 Separador 3 do sistema de apoio à decisão - Gráficos

Passo 6: Caso faça mais de uma consulta pode ir ao histórico para retomar alguma das pesquisas anteriores, desde que não feche a sessão.

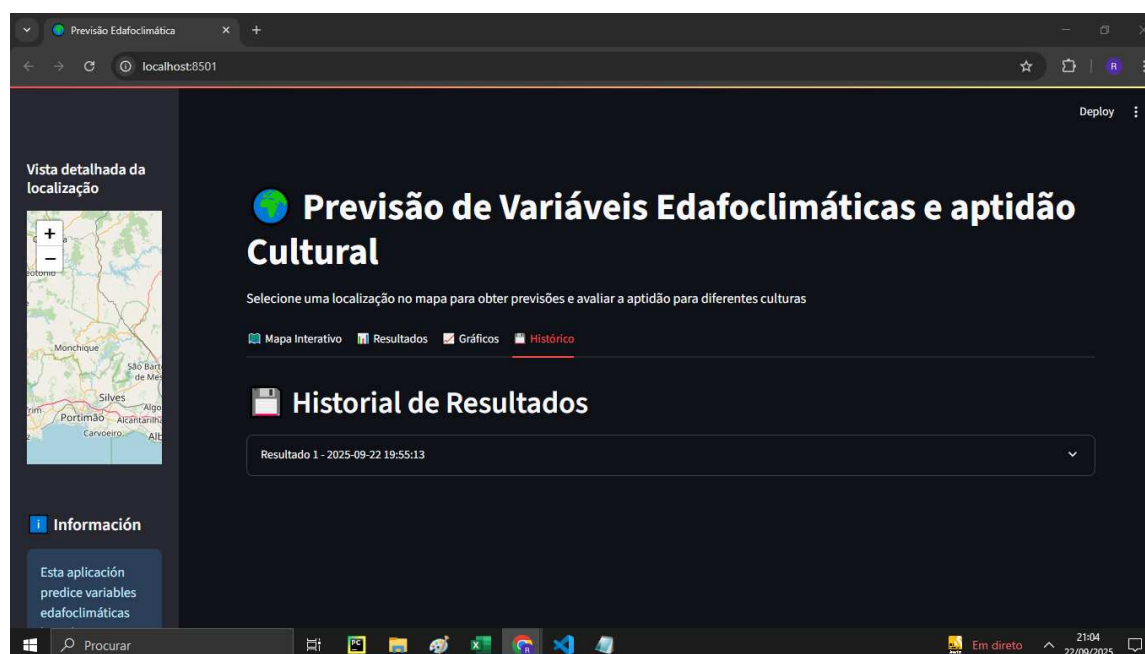


Figura 13 Separador 4 do sistema de apoio à decisão - Histórico

4. Discussão

O sistema de apoio à decisão desenvolvido pretende ser uma ferramenta de apoio ao especialista, que na sua atividade de consultor, tem de produzir recomendações sobre a aptidão de culturas a uma localização. Nessa recomendação, o consultor tem de ter em consideração as condições edafoclimáticas do local, e confrontá-las com os limites definidos para cada cultura. A aplicação desenvolvida pretende facilitar essa tarefa, prevendo os valores dos vários parâmetros edafoclimáticos, e calculando um índice de aptidão para cada cultura, apresentado através da ordenação das culturas com maior aptidão.

O desenvolvimento deste sistema partiu da definição de 158 localizações, para os quais se determinou o valor de 14 parâmetros (10 climáticos e 4 pedológicos), usados para calcular um modelo de previsão multi-output, com base no algoritmo Random Forest Regressor. O número de amostras utilizado é pequeno, o que pode configurar uma limitação deste trabalho. No

entanto não foi possível adicionar mais amostras ao conjunto de dados devido à natureza lenta e manual de obtenção e processamento dos dados. Este requereu de muitas horas para delimitar parcelas, descarregar os dados climáticos do GEE e o seu posterior processamento em excel, que foi feito individualmente para cada localização. Pretendeu-se implementar o mesmo sistema de obtenção de dados utilizado na Consulai. No entanto verificou-se que este procedimento revelou ser limitante, pois não explora o potencial de computação que o GEE permite, podendo ser muito mais eficiente se implementado sobre aquela plataforma para a determinação das variáveis agregadas de clima, do que determinado através do software excel. Este é um dos pontos fundamentais de melhoria do sistema desenvolvido no futuro.

4.1 Discussões sobre a Análise exploratória de dados

A grande variação de condições edafoclimáticas significa que a região a nível local não é homogénea. Um vale pode ser ideal para uma cultura sensível à geada que não sobreviveria numa encosta próxima. O principal desafio é a gestão do risco de geadas tardias. Por exemplo, agricultores em locais com maior incidência (valores mais altos em `Suma_Dias_Helada_Mar_May`) podem precisar de investir em sistemas anti-geada. A acidez do solo pode exigir correção (calagem) dependendo da cultura. A textura arenosa e a enorme variação na precipitação tornam a gestão da rega um fator crítico para o sucesso, com necessidades muito diferentes de local para local.

4.2 Interpretação dos Resultados no Contexto do Problema

Os resultados obtidos demonstram, que o modelo `multioutput random forest regressor` foi capaz de aprender padrões complexos a partir de um conjunto de dados limitado para prever as variáveis edafoclimáticas em estudo.

O coeficiente de determinação global (R^2) de aproximadamente 0,95 no conjunto de dados de teste é considerado um resultado bastante consistente em relação aos objectivos do sistema de apoio à decisão, especialmente dado o reduzido número de observações. As raízes dos erros

quadrados médios (RMSE) apontam na maioria para valores toleráveis relativamente à variável descrita.

Apesar disto a performance do modelo não foi tão boa na previsão das variáveis edáficas, o que pode ser explicado pela menor correlação entre as variáveis do solo entre si e com as variáveis climáticas. Isto não é limitante para a determinação da aptidão cultural uma vez que a maioria das culturas admitem várias texturas e as próprias texturas do solo admitem intervalos suficientemente amplos de valores de areia limo e argila. O mesmo acontece com o valor de pH ideal.

4.3 Comparação com a Abordagem de Modelos Individuais

A comparação entre o desempenho do modelo multi-output e o conjunto de 14 modelos individuais é essencial para validar a opção metodológica adotada. Os resultados obtidos mostram que, apesar de o modelo multi-output apresentar um R^2 global de 0,95, os modelos individuais atingem um valor ligeiramente superior (0,98), bem como um MSE global inferior. Estes resultados refletem a maior capacidade dos modelos single-output se ajustarem de forma mais específica a cada variável de destino.

Contudo, esta ligeira vantagem em desempenho individual está alinhada com o comportamento esperado e amplamente descrito na literatura sobre *multi-target regression* (Borchani et al., 2015; Pineda et al., 2020). Modelos individuais tendem a otimizar exclusivamente o erro da variável alvo correspondente, o que aumenta o risco de captura de ruído e sobreajuste, especialmente em datasets de dimensão moderada (Dietterich, 2000; Caruana, 1997). No caso presente, o ganho marginal de R^2 dos modelos individuais pode refletir precisamente essa sensibilidade ao ruído e à variabilidade idiossincrática de cada variável.

Por contraste, o modelo multi-output aplica um critério de divisão conjunto, minimizando a impureza agregada das diversas variáveis-alvo. Este mecanismo atua como uma forma de regularização estrutural, forçando o modelo a privilegiar padrões partilhados e evitando divisões excessivamente específicas para uma única variável (Zhang & Yeung, 2014; Kocev et al., 2013). A literatura é clara em demonstrar que abordagens multi-output tendem a produzir

modelos mais robustos, particularmente quando existe correlação, redundância ou dependência entre as variáveis de destino (Spyromitros-Xioufis et al., 2016).

Além disso, treinar múltiplos modelos independentes implica ausência total de partilha de informação entre tarefas, desperdício computacional e maior suscetibilidade a inconsistências entre previsões (Caruana, 1997; Borchani et al., 2015). Em problemas ambientais e geoespaciais, onde as variáveis climáticas apresentam relações estruturais fortes, a literatura aponta que modelos multi-task capturam dependências úteis que melhoram a capacidade de generalização (Zhang & Yang, 2018; Madjarov et al., 2012).

Assim, a conclusão de que o pequeno ganho de precisão dos modelos individuais não justifica a perda de generalização relativa à abordagem multi-output é suportada por evidência teórica e empírica. A opção pelo modelo multi-output é consistente com princípios de regularização, redução de variância e aprendizagem multi-tarefa, amplamente reconhecidos como vantajosos em cenários reais onde a robustez é prioritária face a ganhos marginais de precisão.

Por fim, os resultados confirmam que as features selecionadas, latitude e longitude, capturam informação espacial suficiente para explicar grande parte da variabilidade climática e, em menor grau, das variáveis edáficas, respeitando o comportamento esperado da estrutura espacial subjacente.

4.4 Contextualização dos Resultados: Comparação com Estudos Anteriores

Para contextualizar o desempenho do modelo, é útil compará-lo com estudos similares na área. Um estudo sobre a previsão da aptidão de culturas para o deserto indiano utilizando Random Forest alcançou valores de R^2 entre 0,79 e 0,86 para a cultura do trigo (Moharana et al., 2024). A obtenção de um R^2 global de 0,95 neste projeto demonstra um nível de desempenho comparável e, em muitos casos, superior. Além disso, outros trabalhos na área da agricultura de precisão e previsão de aptidão cultural também validam a utilização de modelos ensemble como o Random Forest para lidar com as complexas relações entre fatores ambientais e produtividade.

4.5 Interpretação da Discrepância de Desempenho

A diferença de desempenho entre as previsões das variáveis do clima e do solo é um resultado importante que deve ser interpretado. As variáveis climáticas dependem fortemente da localização geográfica (latitude e longitude), enquanto as variáveis do solo, como o pH e a textura, são muito mais influenciadas por fatores locais e, por isso, não podem ser explicadas apenas pelas coordenadas.

A análise exploratória dos dados do solo já tinha mostrado que o pH tinha pouca correlação com os outros atributos, indicando uma variabilidade que o modelo não conseguiu captar. Os resultados confirmaram que a previsão das percentagens de limo, foram até piores do que o esperado ao acaso, com valores negativos de R^2 . Este resultado reflete de forma honesta as limitações dos dados, e não do algoritmo, e serve como base sólida para melhorar os dados e orientar futuras investigações.

A utilização de Google Earth Engine para a obtenção de dados climáticos e de QGIS para adquirir dados de solos de Portugal continental, constitui um exemplo claro de uma integração que combina fontes de dados de sensorização remota com dados de processamento manual. Isto não é uma limitação, mas sim uma necessidade pragmática. A dependência de dados localmente curados, como as cartas de solo e os comandos de processamento de dados fornecidos pela Consulai, demonstra que, para um DSS ser verdadeiramente eficaz e específico para um país como Portugal, é fundamental integrar dados globais e acessíveis com conhecimento especializado e informações de campo, superando as lacunas de dados que ainda existem para necessidades regionais. Esta combinação de fontes de dados, tanto digitais como analógicas, confere um valor e um realismo adicionais à metodologia.

4.6 Comparação do programa com outros DSS

4.6.1 Sistemas semelhantes na literatura

Foram analisados sistemas com objetivos semelhantes aos do meu programa, de forma a compreender as diferentes abordagens existentes para a avaliação da aptidão agrícola e o apoio à decisão. Estes sistemas representam várias gerações de ferramentas desenvolvidas para integrar dados de clima, solo e gestão, com o objetivo comum de otimizar a produção e reduzir a incerteza nas escolhas agronômicas. Alguns baseiam-se em modelos mecânicos detalhados, outros em metodologias multicritério ou em lógica difusa, refletindo a diversidade de estratégias utilizadas para traduzir informação complexa em recomendações práticas.

CropSuite (v1.0, 2025) é um modelo de aptidão de culturas open-source que utiliza lógica difusa e se baseia na Lei do Mínimo de Liebig, integrando fatores ambientais e socioeconômicos. Este programa considera sistemas de rega e sequeiro, vernalização, limiares letais e fotoperíodo, incluindo ainda downscaling climático para alta resolução (Zabel et al., 2025). O DSSAT, por sua vez, é um programa de simulação mecânica (*process-based*) para mais de 45 culturas, que simula o crescimento, o desenvolvimento e o rendimento em função da dinâmica solo-planta-atmosfera. Este sistema inclui utilitários de solos, meteorologia e gestão, e possui um longo histórico de uso científico e operacional (Hoogenboom et al., 2019). O AquaCrop, desenvolvido pela FAO, é um modelo centrado na relação entre a água e o rendimento das culturas herbáceas, amplamente utilizado e validado, com recentes extensões e integrações de tecnologias IoT (Puig et al., 2025). Já o APSIM é uma plataforma de simulação modular e extensível, com uma forte base em investigação e mais de 30 anos de evolução, amplamente adotada para avaliar o impacto do clima, da gestão e do solo em diferentes sistemas agrícolas (Keating, 2024). Paralelamente, as abordagens baseadas em SIG e lógica difusa ou AHP representam uma linha de trabalhos que realizam a adequação de terras através da integração de múltiplos critérios, muitas vezes sem previsão explícita via *machine learning*, mas com metodologias de decisão multicritério como AHP e *fuzzy logic* (Abuzaid et al., 2025).

4.6.2 Comparação dirigida entre o programa desenvolvido e sistemas semelhantes

O programa desenvolvido insere-se no universo dos Sistemas de Apoio à Decisão (DSS) agrícolas, mas distingue-se pelo seu núcleo data-driven baseado em um modelo *Multi-Output Random Forest Regressor* (MORF). Esta abordagem contrasta com outros modelos de referência na literatura, que recorrem predominantemente a paradigmas process-based, fuzzy-logic.

O CropSuite (Zabel et al., 2025) representa uma das propostas mais recentes na área da avaliação de aptidão agrícola mas está focado para o continente africano e para culturas típicas do seu clima, como a cultura do café, banana, sésamo, papaia, tabaco e cana-de-açúcar. Embora robusto na modelação da aptidão, não incorpora uma camada de previsão de variáveis edafoclimáticas por *machine learning*, trabalhando sobretudo sobre inputs climáticos e edáficos previamente estimados ou downscaled.

Por outro lado, os modelos mecanísticos, como o DSSAT e o APSIM (Holzworth et al., 2014), centram-se na simulação detalhada do crescimento e rendimento das culturas com base na dinâmica solo-planta-atmosfera. São sistemas extremamente completos, mas que requerem parametrizações extensivas e conhecimento técnico especializado, além de dependerem de séries climáticas e dados de solo fornecidos pelo utilizador. O AquaCrop, desenvolvido pela FAO, segue uma lógica semelhante, com foco específico na resposta das culturas à disponibilidade hídrica (Steduto et al., 2009).

Comparativamente, o programa aqui descrito oferece uma solução intermédia: a robustez preditiva de um modelo de *machine learning* multivariado, aliada a uma interface interativa que devolve *rankings* de culturas e fatores limitantes. A sua curva de utilização é menos exigente do que a dos sistemas mecanísticos, e o seu valor acrescentado reside em prever diretamente variáveis edafoclimáticas, algo que não está presente em sistemas como CropSuite, DSSAT ou AquaCrop.

4.6.3 Inovações do programa

O primeiro aspecto inovador é a utilização de um modelo multioutput de *machine learning* para previsão simultânea de variáveis edafoclimáticas. Esta abordagem permite captar correlações entre fatores climáticos e do solo, resultando em previsões mais consistentes e adaptadas ao contexto espacial (Taghizadeh-Mehrjardi et al., 2016; Padarian et al., 2020).

Em segundo lugar, o programa disponibiliza um *ranking* de aptidão cultural com identificação explícita dos fatores limitantes, permitindo que o utilizador compreenda não apenas qual a cultura mais adequada, mas também quais os parâmetros ambientais que restringem outras opções. Este aspecto aumenta a interpretabilidade do sistema e aproxima-se de modelos fuzzy, mas com a vantagem de estar suportado por previsões geradas por *machine learning*.

Uma das inovações do programa prende-se com a generalização espacial do modelo. O desempenho deste depende da diversidade e qualidade das covariáveis utilizadas, que fez com que fosse necessário realizar validações cruzadas em diferentes regiões de Portugal, reduzindo o risco de sobreajuste e aumentando a robustez das previsões (Padarian et al., 2020).

Outro ponto diferenciador é a interface gráfica georreferenciada, que integra num único fluxo interativo: a seleção do local, a previsão de variáveis, a avaliação da aptidão e a visualização de resultados em gráficos e histórico. Esta experiência “end-to-end” aproxima o sistema das necessidades reais de agricultores e técnicos, tornando-o mais acessível do que soluções que exigem parametrizações complexas, como DSSAT ou APSIM (Holzworth et al., 2014).

Por último, a especialização em culturas mediterrânicas permanentes como a oliveira, amendoeira ou videira torna o sistema altamente relevante para o contexto agrícola português, oferecendo respostas adaptadas a espécies de elevado valor económico e estratégico.

4.6.4 Limitações e oportunidades de evolução

Apesar do seu carácter inovador, o programa apresenta algumas limitações que abrem espaço para evolução futura.

Uma das principais limitações refere-se ao número restrito de culturas consideradas. O sistema contempla atualmente oito espécies lenhosas mediterrânicas: oliveira (*Olea europaea*), amendoeira (*Prunus dulcis*), videira (*Vitis vinifera*), noqueira (*Juglans regia*), medronheiro (*Arbutus unedo*), pistácio (*Pistacia vera*), macieira (*Malus domestica*) e pereira (*Pyrus communis*). Embora esta seleção seja altamente relevante para o contexto agrícola português, a limitação pode reduzir o potencial de utilização em cenários de diversificação agrícola ou em regiões que procurem introduzir culturas emergentes ou tropicais, como a Pitaia, o Abacate e a Manga. Estudos prévios sublinham que a utilidade de DSS aumenta proporcionalmente à abrangência de culturas incluídas, o que lhes confere maior flexibilidade e aplicabilidade prática (Holzworth et al., 2014). Apesar de ter poucas culturas, o programa pode ser facilmente ampliado se forem introduzidas novas culturas com os requisitos de apoio à recomendação, sem que isso implique uma alteração do modelo. Só seria necessário alterar o modelo se as variáveis em que este é baseado não forem significativas para as culturas em causa.

Para além disso, o programa não diferencia atualmente variedades infraespecíficas dentro das culturas analisadas. Esta é uma oportunidade importante de evolução, já que diferentes variedades de uma mesma espécie podem ter requisitos ecofisiológicos distintos, em particular no que respeita a tolerância térmica, necessidades de frio, resistência a pragas ou produtividade sob diferentes sistemas de manejo. A literatura destaca que a escala varietal é crítica na avaliação da aptidão agrícola, sobretudo em culturas de grande importância económica, como a videira e a macieira, onde a escolha da variedade pode ser tão ou mais determinante que a escolha da espécie (White et al., 2011; Fraga et al., 2016).

Finalmente, a ausência de simulação de cenários futuros de clima e manejo limita a sua aplicabilidade em projeções de longo prazo. Modelos *process-based* como DSSAT ou APSIM possuem vantagem neste domínio, pois permitem explorar o impacto de alterações climáticas ou de estratégias de irrigação e fertilização alternativas. A integração do presente sistema com simulações de cenários futuros potenciais, utilizando, por exemplo, os outputs do MORF como inputs iniciais para modelos mecanísticos representam uma oportunidade relevante para o futuro desenvolvimento do programa.

Recomendações

Apesar da sua utilidade prática, o programa apresenta limitações que podem ser superadas através de estratégias de evolução que consolidem o seu papel como ferramenta profissional e sustentável. Atualmente, contempla apenas oito espécies lenhosas mediterrânicas, o que restringe a sua aplicação em cenários de maior diversificação agrícola. Além disso, a introdução de variedades intraespecíficas surge como uma oportunidade relevante, dado que diferentes cultivares apresentam requisitos ecofisiológicos distintos, sobretudo no que respeita à tolerância térmica, às necessidades de frio ou à resistência a pragas, aspetos críticos em culturas como a videira e a macieira.

Outra recomendação é o alargamento da base de dados com novas amostras, de modo a aumentar a robustez e a capacidade de generalização do modelo. Paralelamente, a inclusão de novas variáveis preditoras poderá permitir a captura de mecanismos ainda não representados, reduzindo o erro de previsão e aumentando o poder explicativo. Também se torna essencial validar o sistema em diferentes regiões de Portugal para evitar sobreajuste e reforçar a sua robustez em contextos geográficos e temporais distintos.

No plano da infraestrutura, recomenda-se a criação de uma base de dados permanente que permita ao sistema evoluir de um protótipo estático para uma ferramenta dinâmica, capaz de aprender continuamente com novos dados. Em complemento, a implementação de um pipeline de monitorização da performance asseguraria a deteção precoce de desvios do modelo, garantindo que o sistema se mantém preciso face a alterações climáticas ou de uso do solo.

Finalmente, uma oportunidade de evolução reside na integração com modelos mecanísticos que permitem simular cenários futuros de clima e de manejo agrícola. Ao conjugar as previsões rápidas do modelo atual com a capacidade de simulação de longo prazo destes sistemas, seria possível construir uma solução híbrida mais completa e resiliente.

Conclusão

O presente trabalho demonstrou que a integração entre ciência de dados e agricultura pode gerar ferramentas práticas e de elevado valor para a tomada de decisão no setor agrícola português. Através da recolha e integração de dados climáticos e edáficos de 158 localizações em Portugal continental, foi possível desenvolver um modelo preditivo robusto, baseado em *machine learning*, capaz de estimar simultaneamente múltiplas variáveis edafoclimáticas. O sistema resultante traduz essas previsões em *rankings* de aptidão cultural para diferentes espécies mediterrânicas, oferecendo uma visão clara e acessível sobre quais culturas são mais adequadas para cada local e quais fatores limitam o seu desenvolvimento. Esta abordagem representa um avanço face às metodologias tradicionais, ao permitir análises dinâmicas e adaptativas, ajustadas à realidade climática e territorial do país.

Os resultados do modelo, com um coeficiente de determinação global próximo de 0,95, comprovam a sua fiabilidade e utilidade prática, especialmente no que se refere à previsão das variáveis climáticas. As previsões das variáveis de solo apresentaram menor precisão, refletindo a maior variabilidade local e a limitação dos dados disponíveis. Ainda assim, a abordagem multi-output revelou-se vantajosa por garantir maior consistência entre variáveis correlacionadas e melhor capacidade de generalização, elementos essenciais para um sistema de apoio à decisão com aplicação real. O sistema cumpre, assim, o seu propósito principal: apoiar o agricultor e o técnico agrícola na escolha informada de culturas, reduzindo a incerteza e promovendo uma gestão mais eficiente dos recursos naturais.

O programa distingue-se também pela sua interface interativa e pelo carácter participativo que imprime ao processo de decisão. Ao disponibilizar previsões e recomendações personalizadas num ambiente web acessível, o sistema aproxima a ciência de dados do utilizador final, tornando o conhecimento técnico e analítico mais transparente e utilizável. Essa tradução da complexidade em informação prática é um passo fundamental para a adoção de tecnologias de agricultura de precisão, especialmente num contexto de alterações climáticas e crescente

pressão sobre a sustentabilidade produtiva. Além disso, o programa preenche uma lacuna existente no contexto português, ao combinar previsões edafoclimáticas com uma avaliação sistemática da aptidão agrícola, focada em culturas de grande relevância económica.

Apesar das suas limitações, como o número restrito de culturas e a ausência de variedades infraespecíficas ou simulações de cenários futuros, o trabalho estabelece uma base sólida para o desenvolvimento de sistemas de apoio à decisão mais completos e inteligentes. As linhas futuras de evolução incluem o alargamento da base de dados, a integração de novas variáveis e a incorporação de modelos mecanísticos para explorar cenários de clima e manejo agrícola a longo prazo. Em suma, este projeto reforça a importância de combinar dados, tecnologia e conhecimento agronómico para criar soluções que tornem a agricultura mais previsível, sustentável e resiliente, contribuindo de forma concreta para o futuro da agricultura de precisão em Portugal.

Anexos

Quadro A1 Métricas individuais do modelo multi-ouput implementado determinado com base no conjunto de dados de teste.

Abordagem	Multioutput		Individuais	
	RMSE	R ²	RMSE	R ²
Suma_Dias_Helada_Feb_May	3,314	0.88842	1,48	0.9776
Suma_Dias_Helada_Mar_May	2,087	0.8189	0,927	0.9642
Suma_Dias_Helada_Abr_May	0,908	0.73577	0,499	0.9201
Suma_Horas_Frio_Oct_Feb	110,782	0.94574	54,24	0.9870
Suma_Horas_Frio_Oct_Mar	133,59	0.94460	51,51	0.9918
Suma_Precipitacion_Oct_Abr	37,294	0.98614	68,196	0.9536
Minimo_de_Media_de_Temperatura_Minima(°C)	0,662	0.94980	0,423	0.9795
Média de Média de Temperatura Mínima (°C)	0,415	0.95962	0,19	0.9915
Média de Média de Temperatura Máxima (°C)	0,672	0.86703	0,42	0.9485

Abordagem	Multioutput		Individuais	
	RMSE	R ²	RMSE	R ²
Maximo_de_Media_de_Temperatura_Maxima	1,247	0.73754	0,61	0.9371
Areia_x	15,596	0.06825	2,635	0.98
Limo_x	10,971	-0.14	3,33	0.89
Argila_x	6,629	0.43306	3,66	0.8275
ph_x	0,965	0.18782	0,89	0.3076

Quadro A2 Condições ideais e toleráveis do número de geadas para cada cultura

Cultura	Condição Ideal 1	Condição Ideal 2	Condição Tolerável 1	Condição Tolerável 2
Oliveira	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 5	Suma_Dias_Helada_Abr_May ≤ 0.2	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 7	Suma_Dias_Helada_Abr_May ≤ 0.3
Amendoeira	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 4	Suma_Dias_Helada_Abr_May ≤ 0.3	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 6	Suma_Dias_Helada_Abr_May ≤ 0.4
Macieira	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 17	Suma_Dias_Helada_Abr_May ≤ 0.2	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 20	Suma_Dias_Helada_Abr_May ≤ 0.8
Pistácio	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 20	Suma_Dias_Helada_Abr_May ≤ 0.13	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 22	Suma_Dias_Helada_Abr_May ≤ 0.29
Medronheiro	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 21	Suma_Dias_Helada_Abr_May ≤ 1.9	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 23	Suma_Dias_Helada_Abr_May ≤ 2.2
Pereira	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 12	Suma_Dias_Helada_Abr_May ≤ 0.5	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 13	Suma_Dias_Helada_Abr_May ≤ 0.7
Videira	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 7	Suma_Dias_Helada_Abr_May ≤ 0.1	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 8	Suma_Dias_Helada_Abr_May ≤ 0.3
Nogueira	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 17	Suma_Dias_Helada_Abr_May ≤ 0.3	Suma_Dias_Helada_Feb_May-Suma_Dias_Helada_Mar_May ≤ 19	Suma_Dias_Helada_Abr_May ≤ 0.4

Quadro A3 Intervalos ideais e toleráveis e os seus respetivos limites de horas de frio entre outubro e fevereiro para cada cultura

Intervalo	Ideal		Tolerável	
	Limite inferior	Limite superior	Limite inferior	Limite superior
Oliveira	250	---	220	250
Amendoeira	333	---	300	333
Macieira	800	1200	700	1300
Pistácio	600	1000	500	1200
Medronheiro	0	2500	---	---
Pereira	600	900	200	1500
Videira	250	800	100	1200
Nogueira	400	1200	350	1300

Quadro A4 Intervalos ideais e toleráveis e os seus respetivos limites de precipitação entre outubro e abril para cada cultura

Intervalo	Ideal		Tolerável	
	Limite inferior	Limite superior	Limite inferior	Limite inferior
Oliveira	400	---	370	400
Amendoeira	350	---	300	350
Macieira	600	---	500	600
Pistácio	600	---	400	600
Medronheiro	400	---	250	400
Pereira	550	---	350	550
Videira	300	---	225	300
Nogueira	700	---	400	700

Quadro A5 Intervalos ideais e toleráveis e os seus respetivos limites de pH para cada cultura

Intervalo	Ideal		Tolerável	
	Limite inferior	Limite superior	Limite inferior	Limite inferior
Oliveira	6.0	7.5	5.7	7.8
Amendoeira	6.0	8.5	5.8	8.7
Macieira	5.5	6.5	5.3	6.8
Pistácio	6.0	8.0	5.5	8.5
Medronheiro	5.6	6.5	5.2	6.9
Pereira	6.0	7.0	5.5	7.5
Videira	6.0	6.5	5.5	7.0
Nogueira	6.0	7.0	5.4	8.4

Quadro A6 Condições ideais e toleráveis de temperatura normal para cada cultura

Cultura	Condição Ideal 1	Condição Ideal 2	Condição Tolerável 1	Condição Tolerável 2
Oliveira	Média de Média de Temperatura Mínima (°C) ≥ 15	Média de Média de Temperatura Máxima (°C) ≤ 30	Média de Média de Temperatura Mínima (°C) ≥ 13.5	Média de Média de Temperatura Máxima (°C) ≤ 33
Amendoeira	Média de Média de Temperatura Mínima (°C) ≥ 12	Média de Média de Temperatura Máxima (°C) ≤ 22	Média de Média de Temperatura Mínima (°C) ≥ 10.8	Média de Média de Temperatura Máxima (°C) ≤ 24.2
Macieira	Média de Média de Temperatura Mínima (°C) ≥ 7	Média de Média de Temperatura Máxima (°C) ≤ 20	Média de Média de Temperatura Mínima (°C) ≥ 6.3	Média de Média de Temperatura Máxima (°C) ≤ 22
Pistácio	Média de Média de Temperatura Mínima (°C) ≥ 6	Média de Média de Temperatura Máxima (°C) ≤ 26	Média de Média de Temperatura Mínima (°C) ≥ 5.4	Média de Média de Temperatura Máxima (°C) ≤ 28.6
Medronheiro	Média de Média de Temperatura Mínima (°C) ≥ 6	Média de Média de Temperatura Máxima (°C) ≤ 27	Média de Média de Temperatura Mínima (°C) ≥ 5.4	Média de Média de Temperatura Máxima (°C) ≤ 29.7
Pereira	Média de Média de Temperatura Mínima (°C) ≥ 6	Média de Média de Temperatura Máxima (°C) ≤ 26	Média de Média de Temperatura Mínima (°C) ≥ 5.4	Média de Média de Temperatura Máxima (°C) ≤ 28.6
Videira	Média de Média de Temperatura Mínima (°C) ≥ 10	Média de Média de Temperatura Máxima (°C) ≤ 26	Média de Média de Temperatura Mínima (°C) ≥ 9	Média de Média de Temperatura Máxima (°C) ≤ 28.6
Nogueira	Média de Média de Temperatura Mínima (°C) ≥ 6	Média de Média de Temperatura Máxima (°C) ≤ 26	Média de Média de Temperatura Mínima (°C) ≥ 5.4	Média de Média de Temperatura Máxima (°C) ≤ 28.6

Quadro A7 Condições ideais e toleráveis de temperaturas extremas para cada cultura

Cultura	Condição Ideal 1	Condição Ideal 2	Condição Tolerável 1	Condição Tolerável 2
Oliveira	Minimo de Media de Temperatura Minima(°C) ≥ 15	Maximo de Media de Temperatura Maxima ≤ 30	Minimo de Media de Temperatura Minima(°C) ≥ 13.5	Maximo de Media de Temperatura Maxima ≤ 33
Amendoeira	Minimo de Media de Temperatura Minima(°C) ≥ 12	Maximo de Media de Temperatura Maxima ≤ 22	Minimo de Media de Temperatura Minima(°C) ≥ 10.8	Maximo de Media de Temperatura Maxima ≤ 24.2
Macieira	Minimo de Media de Temperatura Minima(°C) ≥ 7	Maximo de Media de Temperatura Maxima ≤ 20	Minimo de Media de Temperatura Minima(°C) ≥ 6.3	Maximo de Media de Temperatura Maxima ≤ 22
Pistácio	Minimo de Media de Temperatura Minima(°C) ≥ 6	Maximo de Media de Temperatura Maxima ≤ 26	Minimo de Media de Temperatura Minima(°C) ≥ 5.4	Maximo de Media de Temperatura Maxima ≤ 28.6
Medronheiro	Minimo de Media de Temperatura Minima(°C) ≥ 6	Maximo de Media de Temperatura Maxima ≤ 27	Minimo de Media de Temperatura Minima(°C) ≥ 5.4	Maximo de Media de Temperatura Maxima ≤ 29.7
Pereira	Minimo de Media de Temperatura Minima(°C) ≥ 6	Maximo de Media de Temperatura Maxima ≤ 26	Minimo de Media de Temperatura Minima(°C) ≥ 5.4	Maximo de Media de Temperatura Maxima ≤ 28.6
Videira	Minimo de Media de Temperatura Minima(°C) ≥ 10	Maximo de Media de Temperatura Maxima ≤ 26	Minimo de Media de Temperatura Minima(°C) ≥ 9	Maximo de Media de Temperatura Maxima ≤ 28.6
Nogueira	Minimo de Media de Temperatura Minima(°C) ≥ 6	Maximo de Media de Temperatura Maxima ≤ 26	Minimo de Media de Temperatura Minima(°C) ≥ 5.4	Maximo de Media de Temperatura Maxima ≤ 28.6

Quadro A8 Texturas ideais e toleráveis para cada cultura

Cultura	Textura Ideal	Textura toleravel
Oliveira	Franco arenoso; Franco argilo arenoso; Franco; Franco argilosa; Franco argilo limosa; Franco limosa	Textura ideal com margem $\pm 20\%$
Amendoeira	Franco arenoso; Franco.	Textura ideal com margem $\pm 20\%$
Macieira	Franco arenoso; Franco.	Textura ideal com margem $\pm 20\%$
Pistácio	Franco arenoso; Franco.	Textura ideal com margem $\pm 20\%$
Medronheiro	Franco arenoso; Franco; Franco limosa.	Textura ideal com margem $\pm 20\%$
Pereira	Franco; Franco limosa; Argilo limosa.	Textura ideal com margem $\pm 20\%$
Videira	Franco arenoso; Franco; Franco argilosa; Argilosa.	Textura ideal com margem $\pm 30\%$
Nogueira	Franco arenoso; Franco; Argilo limosa; Franco limosa; Franco argilo limosa.	Textura ideal com margem $\pm 20\%$

Quadro A9 Intervalos ideais e toleráveis e os seus respetivos limites de espessura para cada cultura

Intervalo	Ideal		Tolerável	
	Limite inferior	Limite superior	Limite inferior	Limite superior
Oliveira	110	---	80	109
Amendoeira	80	---	60	79
Macieira	60	---	40	59
Pistácio	30	---	20	29
Medronheiro	60	---	40	59
Pereira	30	---	20	29
Videira	60	---	35	59
Nogueira	100	---	80	99

Referências Bibliográficas

A

Abuzaid, A. S., Abbas, H. H., Mostafa, M. A., El Ghonamy, Y. K., Rebouh, N. Y., & Shokr, M. S. (2025). A comprehensive crop suitability assessment under modern irrigation system in arid croplands. *PLoS ONE*, *20*(6), e0326183. <https://doi.org/10.1371/journal.pone.0326183>

Agência Portuguesa do Ambiente. (s.d.). *Sistema Nacional de Informação de Solos (SNISolo)*. Obtido em 2025, de <https://snisolo.apambiente.pt/>

Anco-Valdivia, J., Valencia-Félix, S., Espinoza Vigil, A. J., Anco, G., Booker, J., Juarez-Quispe, J., & Rojas-Chura, E. (2025). A methodology based on random forest to estimate precipitation return periods: A comparative analysis with probability density functions in Arequipa, Peru. *Water*, *17*(1), 128.

Antle, J. M., Jones, J. W., & Rosenzweig, C. (2017). Next generation agricultural system data, models, and knowledge products: Introduction. *Agricultural Systems*, *155*, 179–185.

Araba, B. (2024). Predictive modeling of crop yields in agriculture using random forest algorithm. *Global Scientific Journals*, *12*(10), 1911-1915.

B

Bahmutsky, S., Grassauer, F., Arulnathan, V., & Pelletier, N. (2024). A review of life cycle impacts and costs of precision agriculture for cultivation of field crops. *Sustainable Production and Consumption*, *52*, 347–362. <https://doi.org/10.1016/j.spc.2024.09.012>

Beck, H. E., McFarlane, J. C., Kumar, A., & Singh, R. P. (2019). Interactive decision-support tools for crop disease management based on machine learning and web visualization. *Computers and Electronics in Agriculture*, *162*, 300–312. <https://doi.org/10.1016/j.compag.2019.04.015>

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758. <https://doi.org/10.3390/s21113758>

Bilal, H., Lahlou, F.-Z., & Al-Ansari, T. (2025). Land suitability assessment and self-sufficiency evaluation for fodder crop production in a hyper arid environment coupling GIS-based multi-criteria decision making and optimization. *Ecological Modelling*, 501, 111021. <https://doi.org/10.1016/j.ecolmodel.2024.111021>

Borchani, H., Varando, G., Bielza, C., & Larrañaga, P. (2015). *A survey on multi-output regression*. WIREs Data Mining and Knowledge Discovery. <https://doi.org/10.1002/widm.1157>

C

Cardoso, A. S., Alonso, J., Rodrigues, A. S., Araújo-Paredes, C., Mendes, S., & Valín, M. I. (2019). Agro-ecological terroir units in the North West Iberian Peninsula wine regions. *Applied Geography*, 107, 51–62. <https://doi.org/10.1016/j.apgeog.2019.04.008>

Caruana, R. (1997) Multitask Learning. *Machine Learning*, 28, 41–75 . <https://doi.org/10.1023/A:1007379606734>

Chege, S. (2025). *How to utilize random forest for geospatial precision agriculture, step-by-step guide*. Towards AI. <https://pub.towardsai.net/how-to-utilize-random-forest-for-geospatial-precision-agriculture-step-by-step-guide-0cc5df934b5c>

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>

D

de Magalhães, L. P., Sais, A. C., & Rossi, F. (2025). Research on Using Ensemble Models to Assess the Impacts of Climate Change on Agriculture Production: A Review. *AgriEngineering*, 7(7), 219. <https://doi.org/10.3390/agriengineering7070219>

Dietterich, T. G. (2000). *Ensemble Methods in Machine Learning*. In *Multiple Classifier Systems* (pp. 1–15). Springer. https://doi.org/10.1007/3-540-45014-9_1

Dietterich, T. (2014). *Would a random forest with multiple outputs be possible/practical?* [Online forum post]. Cross

Validated. <https://stats.stackexchange.com/questions/28968/would-a-random-forest-with-multiple-outputs-be-possible-practical>

Direção Regional de Agricultura de Entre-Douro e Minho. (1995). *Carta dos solos e carta da aptidão da terra de Entre-Douro e Minho: Volume de anexos* [Relatório técnico]. Agroconsultores & Geometral.

F

Fischer, G., Nachtergaele, F. O., van Velthuisen, H. T., Chiozza, F., Franceschini, G., Henry, M., Muchoney, D., & Tramberend, S. (2021). *Global Agro-Ecological Zones v4 – Model documentation*. FAO. <https://doi.org/10.4060/cb4744en>

Fraga, H., García de Cortázar Atauri, I., Malheiro, A. C., & Santos, J. A. (2016). Modelling climate change impacts on viticultural yield, phenology and stress conditions in Europe. *Global Change Biology*, 22(11), 3774–3788. <https://doi.org/10.1111/gcb.13382>

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Shukla, S., Husak, G. J., Rowland, J. D., Harrison, L., Hoell, A., & Michaelsen, J. (2015).

The Climate Hazards InfraRed Precipitation with Station data (CHIRPS) version 2.0.

Data set. Climate Hazards Center, University of California, Santa Barbara.

<https://doi.org/10.15780/G2RP4Q>

G

Ganati, B. A., & Sitote, T. M. (2025). Predicting land suitability for wheat and barley crops using machine learning techniques. *Scientific Reports*, 15, 15879. <https://doi.org/10.1038/s41598-025-99070-0>

H

Hagn, L., Mittermayer, M., Kern, A., Kimmelmann, S., Maidl, F.-X., & Hülsbergen, K.-J. (2025). Effects of Sensor-Based, Site-Specific Nitrogen Fertilizer Application on Crop Yield, Nitrogen Balance, and Nitrogen Efficiency. *Sensors*, 25(3), 795. <https://doi.org/10.3390/s25030795>

Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangquan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R.,

MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, *12*(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., ... & Keating, B. A. (2014). APSIM—evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, *62*, 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>

Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White, J. W., Asseng, S., Lizaso, J. I., Moreno, L. P., Pavan, W., Ogoshi, R., Hunt, L. A., Tsuji, G. Y., & Jones, J. W. (2019). The DSSAT crop modeling ecosystem. En K. J. Boote (Ed.), *Advances in Crop Modeling for a Sustainable Agriculture* (pp. 173–216). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/AS.2019.0061.10>

J

Jain, M., Singh, R., & Ghosh, A. (2020). Smart irrigation decision support using IoT and data analytics dashboards. *Sustainability*, *12*(14), 5674.

Jeong, J. H., Park, H. S., & Kim, M. K. (2016). Simultaneous prediction of temperature and precipitation using multi-output Random Forest models. *Environmental Modelling & Software*, *79*, 301–310.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, *11*(6), e0156571. <https://doi.org/10.1371/journal.pone.0156571>

Jones, G. V., & Alves, F. (2012). Impact of climate change on wine production: A global overview and regional assessment in the Douro Valley of Portugal. *International Journal of Global Warming*, *4*(3/4), 383–406. <https://doi.org/10.1504/IJGW.2012.049448>

Joy, A. P. (2025). *Random forest: Why ensemble learning outperforms individual models*. Skillcamper. <https://www.skillcamper.com/blog/random-forest-why-ensemble-learning-outperforms-individual-models>

K

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>

Keating, B. A. (2024). APSIM's origins and the forces shaping its first 30 years of evolution: A review and reflections. *Agronomy for Sustainable Development*, 44, 24. <https://doi.org/10.1007/s13593-024-00959-3>

Kocev, D., Vens, C., Struyf, J., & Džeroski, S. (2013). Tree ensembles for predicting structured outputs. *Pattern Recognition*, 46(3), 817–833. <https://doi.org/10.1016/j.patcog.2012.09.023>

Kocev, D., Vens, C., Struyf, J., & Džeroski, S. (2013). *Tree ensembles for predictive clustering and multi-target prediction*. In *Information Sciences* (Vol. 268, pp. 21–35). <https://doi.org/10.1016/j.ins.2013.05.036>

Kumar, A., & Kumar, A. (2025). A decision-support system perspective of machine learning in agriculture: A review. *International Journal of Sustainable Studies, Technologies, and Assessments*, 1(1).

Kumar, R., Sharma, P., & Singh, A. (2021). AgroDSS: A web-based decision support system for crop selection under edaphoclimatic variability. *Environmental Modelling & Software*, 145, 105198.

L

Lakhiar, I. A., Yan, H., Zhang, C., Wang, G., He, B., Hao, B., Han, Y., Wang, B., Bao, R., Syed, T. N., Chauhdary, J. N., & Rakibuzzaman, M. (2024). A Review of Precision Irrigation Water-Saving Technology under Changing Climate for Enhancing Water Use Efficiency, Crop Yield, and Environmental Footprints. *Agriculture*, 14(7), 1141. <https://doi.org/10.3390/agriculture14071141>

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>

Liu, Y., Wu, X., Wu, T., Zhao, L., Li, R., Li, W., Hu, G., Zou, D., Ni, J., Du, Y., Wang, M., Li, Z., Wei, X., & Yan, X. (2022). Soil texture and its relationship with environmental factors on the Qinghai–Tibet Plateau. *Remote Sensing*, 14(15), 3797.

M

Madjarov, G., Kocev, D., Gjorgjevikj, D., & Džeroski, S. (2012). *An extensive experimental comparison of methods for multi-label learning*. *Pattern Recognition*, 45(9), 3084–3104.

<https://doi.org/10.1016/j.patcog.2012.03.004>

Marcelino, S. M., Gaspar, P. D., Paço, A., Lima, T. M., Monteiro, A., Franco, J. C., Santos, E. S., Campos, R., & Lopes, C. M. (2024). Decision Support System for the Assessment and Enhancement of Agrobiodiversity Performance. *Sustainability*, 16(15), 6519. <https://doi.org/10.3390/su16156519>

Marti-Jerez, K., Català-Forner, M., Tomàs, N., Murillo, G., Ortiz, C., Sánchez-Torres, M. J., Vitali, A., & Lopes, M. S. (2023). Agronomic performance and remote sensing assessment of organic and mineral fertilization in rice fields. *Frontiers in Plant Science*, 14, 1230012. <https://doi.org/10.3389/fpls.2023.1230012>

Moharana, P. C., Yadav, B., Malav, L., & Biswas, H. (2024). Machine learning-based crop suitability prediction: An emerging technique for sustainable agricultural production in the desert region of India. *Communications in Soil Science and Plant Analysis*. Advance online publication. <https://doi.org/10.1080/00103624.2024.2419994>

Muñoz Sabater, J. (2019). *ERA5-Land hourly data from 1950 to present [Data set]*. Copernicus Climate Change Service (C3S), ECMWF. <https://doi.org/10.24381/cds.e2161bac>

Mwendwa, S. (2022). Revisiting soil texture analysis: Practices towards a more accurate Bouyoucos method. *Heliyon*, 8(5), e09395.

N

Natural Resources Conservation Service. (2014). *Soil health - guides for educators*. U.S. Department of Agriculture.

Nawar, S., Corstanje, R., Halcro, G., Mulla, D., & Mouazen, A. M. (2021). Predicting soil and crop parameters using Random Forest regression and multi-output ensembles. *Geoderma*, 383, 114793.

O

Oliveira, M., Santos, R., & Fonseca, A. (2023). Aplicação de modelos multioutput para previsão de variáveis térmicas e fenológicas em fruteiras mediterrânicas. *Revista de Engenharia Agrícola e Ambiental*, 27(4), 102–115.

P

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Using deep learning for digital soil mapping. *Soil*, 6(1), 79–91. <https://doi.org/10.5194/soil-6-79-2020>

Pereira, H., Dias, L. C., & Alves, M. J. (2010). Assessing the Efficiency and Distribution of Agricultural Activities Using DSS: A Case Study for a Portuguese Region. In B. Manos, N. Matsatsinis, K. Paparrizos, & J. Papathanasiou (Eds.), *Decision Support Systems in Agriculture, Food and the Environment: Trends, Applications and Advances* (pp. 288–316). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-61520-881-4.ch014>

Peter, B. G., Messina, J. P., Lin, Z., & Snapp, S. S. (2020). Crop climate suitability mapping on the cloud: A geovisualization application for sustainable agriculture. *Scientific Reports*, 10(1), 15487. <https://doi.org/10.1038/s41598-020-72384-x>

Pineda, R., Armenta, A., & Salcedo-Sanz, S. (2020). *Multi-output tree ensemble methods: A review*. *Information Fusion*, 56, 1–13.

<https://doi.org/10.1016/j.inffus.2019.09.017>

Puig, F., Garcia-Vila, M., Soriano, M. A., & Rodríguez-Díaz, J. A. (2025). AquaCrop-IoT: A smart irrigation platform integrating real-time images and weather forecasting. *Computers and Electronics in Agriculture*, 235, 110372. <https://doi.org/10.1016/j.compag.2025.110372>

R

Ramesh, K., Singh, D., & Pandey, A. (2022). Data-driven multi-output regression models for crop yield and water use prediction under varying climate scenarios. *Agricultural Water Management*, 268, 107692.

Ravikumar, S., Vellingiri, G., Sellaperumal, P., Pandian, K., Sivasankar, A., & Sangchul, H. (2024). Real-time nitrogen monitoring and management to augment N use efficiency and ecosystem sustainability—A review. *Journal of Hazardous Materials Advances*, 16, 100466. <https://doi.org/10.1016/j.hazadv.2024.100466>

Reiniger, B. (2022). *Difference between multioutputregressor(randomforestregressor) versus randomforestregressor with n_outputs* [Online forum post]. Stack Overflow. <https://stackoverflow.com/questions/74155460/difference-between-multioutputregressorrandomforestregressor-versus-randomfo>

Reis, F. and de Melo-Abreu, J.P. (2017). Impacts of Climate Change on Portuguese Agriculture. In *Climate Change and Agriculture* (eds P.S.S. Yadav, R.J. Redden, J.L. Hatfield, A.E. Hall and S.A. Mohanty). <https://doi.org/10.1002/9781119057591.ch12>

S

Santos, F. D., Forbes, K., & Moita, R. (Eds.). (2002). *Climate change in Portugal: Scenarios, impacts and adaptation measures - The SIAM Project*. Gradiva.

Sathiyamurthi, S., Sivasakthi, M., Saravanan, S., Gobi, R., Praveen kumar, S., & Karuppanan, S. (2024). Assessment of crop suitability analysis using AHP-TOPSIS and geospatial techniques: A case study of Krishnagiri District, India. *Environmental and Sustainability Indicators*, 24, 100466. <https://doi.org/10.1016/j.indic.2024.100466>

scikit-learn developers. (2025). *Scikit-learn: Machine Learning in Python* (Version 1.5) [Computer software]. <https://scikit-learn.org/stable/>

Segal, M. R., & Xiao, Y. (2011). Multivariate random forests. *WIRES Data Mining Knowl Discov*, 1(1), 80-87.

Spyromitros-Xioufis, E., Ntoutsis, I., Papadopoulos, S., & Kompatsiaris, I. (2016). *Multi-target regression via input/output space structural regularization*. *IEEE Transactions on Neural Networks and Learning Systems*, 27(2), 427–439.

<https://doi.org/10.1109/TNNLS.2015.2416208>

Serviço de Reconhecimento e de Ordenamento Agrário. (1970). *Carta dos solos de Portugal: Classificação e caracterização dos solos de Portugal. Volume I: Classificação e caracterização morfológica dos solos* (Boletim No. 14). Secretaria de Estado da Agricultura.

Serviço de Reconhecimento e de Ordenamento Agrário. (1973a). *Carta dos solos de Portugal: Classificação e caracterização dos solos de Portugal. Volume II: Dados analíticos das unidades pedológicas* (Boletim No. 16). Secretaria de Estado da Agricultura.

Serviço de Reconhecimento e de Ordenamento Agrário. (1973b). *Carta dos solos de Portugal. Agronomia Lusitana*, 33(1-4).

Sharma, V., Kaur, G., Sreethu, S., Chhabra, V., & Kashyap, R. (2025). Smart irrigation systems in agriculture: An overview. *Computers and Electronics in Agriculture*, 239(Part B), 111008.

Soussi, A., Zero, E., Sacile, R., Trincherio, D., & Fossa, M. (2024). Smart sensors and smart data for precision agriculture: A review. *Sensors*, 24(8), 2647. <https://doi.org/10.3390/s24082647>

Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/10.2134/agronj2008.0139s>

Sun, Q., Li, X., & Zhang, Y. (2020). Multi-output Random Forest for meteorological variable prediction in agricultural DSS. *Computers and Electronics in Agriculture*, 175, 105588.

T

Taghizadeh-Mehrjardi, R., Nabiollahi, K., & Kerry, R. (2016). Digital soil mapping using machine learning algorithms in Iran. *Soil Systems*, 1(1), 1–15. <https://doi.org/10.3390/soilsystems1010001>

V

Van Evert, F. K., Booij, R., Jukema, J. N., ten Berge, H. F. M., Uenk, D., Meurs, E. J. J., & van der Schans, D. A. (2016). A geospatial web-based decision support system for variable-rate fertilization in arable farming. *Computers and Electronics in Agriculture*, *127*, 87–96. <https://doi.org/10.1016/j.compag.2016.05.001>

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. <https://doi.org/10.1016/j.compag.2020.105709>

W

White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, *124*(3), 357–368. <https://doi.org/10.1016/j.fcr.2011.07.001>

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, *153*, 69–80. <https://doi.org/10.1016/j.agsy.2017.01.023>

Wu, P., Wu, Q., Huang, H., Xie, L., An, H., Zhao, X., Wang, F., Gao, Z., Zhang, R., Bangura, K., Xue, J., Sun, M., Zhang, P., Jia, Z., & Gao, Z. (2024). Global meta-analysis and three-year field experiment shows that deep placement of fertilizer can enhance crop productivity and decrease gaseous nitrogen losses. *Field Crops Research*, *307*, 109263. <https://doi.org/10.1016/j.fcr.2023.109263>

Y

Ye, L., Hanson, L. S., Ding, P., Wang, D., & Vogel, R. M. (2018). The probability distribution of daily precipitation at the point and catchment scales in the United States. *Hydrology and Earth System Sciences*, *22*(12), 6519–6531.

Z

Zabel, F., Knüttel, M., & Poschlod, B. (2025). CropSuite v1.0 – a comprehensive open-source

crop suitability model considering climate variability for climate impact assessment. *Geoscientific Model Development*, 18(4), 1067–1087. <https://doi.org/10.5194/gmd-18-1067-2025>

Zhang, B., Linz, M., & Chen, G. (2022). Interpreting observed temperature probability distributions using a relationship between temperature and temperature advection. *Journal of Climate*, 35(3), 705–724.

Zhang, Y., & Yang, Q. (2018). *A survey on multi-task learning*. *IEEE Transactions on Knowledge and Data Engineering*, 34(12), 5586–5609.

<https://doi.org/10.1109/TKDE.2021.3070203>

Zhang, Y., & Yeung, D.-Y. (2014). *A regularization approach to learning task relationships in multitask learning*. *ACM Transactions on Knowledge Discovery from Data*, 8(3), 1–31.

<https://doi.org/10.1145/2538028>

Zhang, Y.-Y., Wu, W., & Liu, H. (2019). Factors affecting variations of soil pH in different horizons in hilly regions. *PLoS ONE*, 14(6).

Zhao, H., Di, L., Guo, L., Zhang, C., & Lin, L. (2023). An Automated Data-Driven Irrigation Scheduling Approach Using Model Simulated Soil Moisture and Evapotranspiration. *Sustainability*, 15(17), 12908. <https://doi.org/10.3390/su151712908>