

PhD

3rd
CYCLE

FCUP
2024

U.PORTO

Advanced methodologies for the diagnosis of
agronomic processes based on systems biology for
precision agriculture

Renan Tosin

FC

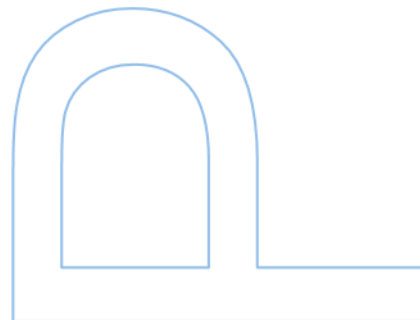
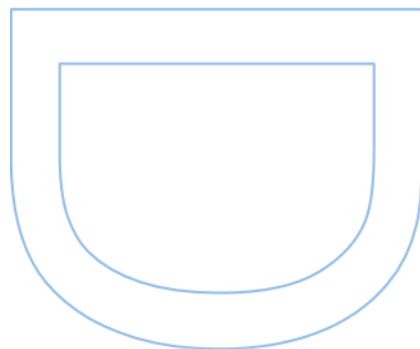
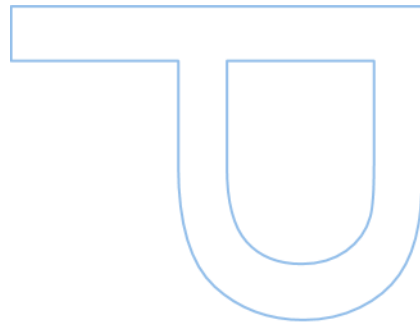
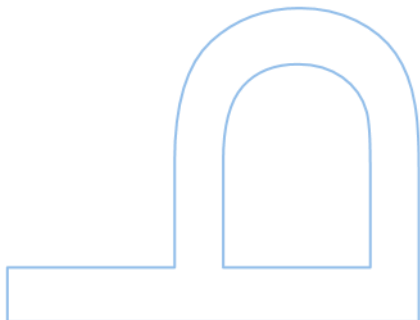


Advanced methodologies for the diagnosis of agronomic processes based on systems biology for precision agriculture

Renan Tosin

Doctoral Program in Agrarian Sciences
Department of Geosciences, Environment and Spatial Plannings
Faculty of Sciences of the University of Porto

2024



Advanced methodologies for the diagnosis of agronomic processes based on systems biology for precision agriculture



Renan Tosin

Thesis carried out as part of the Doctoral Program in
Agrarian Sciences
Department of Geosciences, Environment and Spatial Plannings
2024

Supervisor

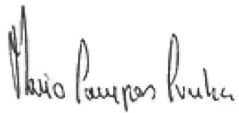
Mário Manuel de Miranda Furtado Campos Cunha, Associate
Professor with habilitation, Faculty of Sciences, University of Porto
Researcher at Institute for Systems and Computer Engineering,
Technology and Science (INESC-TEC)

Co-supervisor

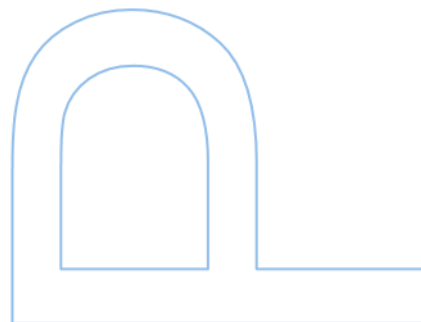
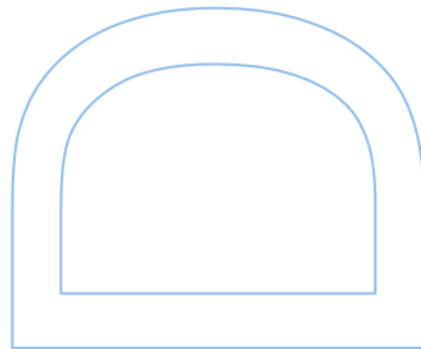
Rui Miguel da Costa Martins, Senior Researcher/Principal
Researcher at Institute for Systems and Computer Engineering,
Technology and Science (INESC-TEC)

Todas as correções determinadas
pelo júri, e só essas, foram
efetuadas.

A/O Orientadora(o),



Porto, 06 / 06 / 2024



"The greatest use of life is to spend it for something that will outlast it."

William Jame

Previous note

Renan Tosin acknowledges Fundação para a Ciência e Tecnologia (FCT) PhD research grant reference SFRH/BD/145182/2019.

This project received backing from the following projects:

- National Funds, through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology), funded this work within the project 'OmicBots: High-Throughput Integrative Omic-Robots Platform for a Next Generation Physiology-based Precision Viticulture,' with reference PTDC/ASP-HOR/1338/2021.
- SpecTOM - Non-Invasive spectroscopy tomographic technology.
- National Funds provided by the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology), through the project 'Metbots-Metabolomic Robots with Self-learning Artificial Intelligence for Precision Agriculture' (Reference PTDC/EEI-ROB/31124/2017). Additionally, it was supported by FEDER funds through the COMPETE 2020 Programme under project number POCI-01-0145-FEDER-031124.
- Wine-spectra. This project results from a protocol between the Faculty of Sciences of the University of Porto (FCUP), the Association for the Development of Viticulture Duriense (ADVID), and the companies Real Companhia Velha and Symington Family Estates.

This research was conducted at the facilities of the Department of Geosciences, Environment, and Spatial Planning of the Faculty of Sciences of Porto University, TRIBE - Robotics and IoT Laboratory for Precision Agriculture and Forestry of the Centre of Robotics in Industry and Intelligent Systems (CRIIS), affiliated with the Researcher at the Institute for Systems and Computer Engineering, Technology, and Science (INESC-TEC), Integrative Biology and Biotechnology Laboratory (iB2), Quinta dos Aciprestes (Real Companhia Velha), Quinta do Vale dos Cavalos (Poças Junior) and Association for the Development of Duriense Viticulture – ADVID.

The thesis contains case studies presented in chapters, each corresponding to submitted scientific manuscripts or in preparation for future publication. Acknowledging the collaborative nature of these studies, the student recognises the involvement of the supervisor, Mario Cunha (FCUP), and co-supervisor Rui Martins (INESC TEC). Renan Tosin emphasises personal participation in the writing, revision, design, sample collection, data analysis, discussion of results, and the creation and editing of illustrations.



REAL COMPANHIA VELHA



Acknowledgements

Renan Tosin sincerely thanks his supervisor, Mario Cunha, for the wealth of knowledge taught throughout this work. His continuous guidance and solid attention were fundamental to completing this thesis. Renan acknowledges the exhaustive effort and contributions made by Mario Cunha, emphasising the indispensable role played in the entire process.

Special gratitude is extended to the co-supervisor, Rui Martins, for his unique perceptions, particularly in the coding aspect of this work. Renan also thanks Rui Martins for his critical perspective during the scientific dissemination of the research.

Acknowledgement is extended to the Department of Geosciences, Environment, and Spatial Planning at the Faculty of Sciences of Porto University. Renan thanks the entire department staff, including Fernanda Silva, Didi, Paulo Ferreira, Joana Damásio, Monica Alves, Cristina Vilas Boas, Paulo Ferreira, and the department director, Prof. Dr. Joaquim Esteves, for their support in the development of this work, and Paula Marques, from the Post-Graduation Office.

Gratitude is expressed to the TRIBE and CRIIS teams for their support throughout the research, particularly in addressing hardware issues. Special mention goes to Filipe Santos for consistent support from the outset of Renan's PhD. Gratitude is also extended to Pedro Moura and Sandro Magalhães for their assistance with hardware challenges.

Renan Tosin acknowledges the ADVID team, including Catarina Barbosa, Igor Gonçalves, Ana Morais, and Leonor Pereira, for their support during fieldwork.

Thanks are given to colleagues who provided solid support during both positive and challenging moments: Mafalda Pereira, Igor Portis, Diva Cabral, Maling Wang, Telmo Vieira, Cátia Rodrigues, and Pedro Aguiar. A special acknowledgement is reserved for Isabel Pôças, Renan's former co-supervisor in his master's dissertation, who continues to support his academic journey.

Renan expresses gratitude to therapists Lídia Oliveira and Ofélia Castro for revealing unique perceptions and a different perspective on confrontation problems, proving essential to completing this work.

Special admiration is directed towards friends Leandro, Nayara, Francisco, Adriano, and Fred for their unwavering emotional support during moments of anxiety.

Lastly, Renan expresses profound gratitude to his mother for her constant support during challenging times, never allowing him to fail or abandon his goals.

Resumo

A agricultura de precisão (AP) emergiu como uma ferramenta crucial na agricultura moderna, visando otimizar a produtividade, qualidade e competitividade das colheitas, enquanto reduz o uso de recursos e minimiza o impacto ambiental. Comparado aos métodos tradicionais de monitorização de culturas, que dependem da observação visual do fenótipo e da amostragem destrutiva no campo, a AP oferece uma alternativa mais eficiente e precisa. Estes métodos tradicionais são intensivos em mão-de-obra e propensos a erros, potencialmente resultando em decisões agronómicas imprecisas, como a programação de rega ou o momento da colheita. Avanços recentes em espectroscopia, inteligência artificial (IA) e biologia de sistemas apresentam oportunidades promissoras para melhorar as práticas de monitorização e gestão agrícola.

Existe uma lacuna na compreensão dos processos fisiológicos das plantas sob diferentes condições ambientais. A necessidade de uma abordagem genótipo, ambiente e práticas culturais, é essencial para preencher esta lacuna na compreensão dos processos fisiológicos das plantas em várias condições ambientais. A investigação atual frequentemente carece de uma ferramenta de modelação abrangente combinada com informações espectrais para aplicar estas metodologias em diferentes culturas e ambientes, dificultando sua transferibilidade entre diferentes anos e locais. Esta tese visa integrar informações hiperespectrais com IA e biologia de sistemas, permitindo explorar como a fisiologia/metabolismo das plantas é determinada pelo genótipo x ambiente x práticas culturais, o que é crucial para entender melhor como os perfis de metabólitos dependem desses fatores.

Em resposta a estes desafios, esta tese visa desenvolver e validar metodologias avançadas para diagnosticar processos agronómicos com base nos princípios da biologia de sistemas para a AP. Especificamente, a tese visa: i) fornecer explicações fisiológicas para o estado de água da videira; ii) apresentar métodos de baixo custo com preparação mínima de laboratório baseados em espectroscopia para avaliar a composição de nutrientes em pequenas quantidades de tecido vegetal; iii) utilizar técnicas de espectroscopia baseadas nas zonas visíveis e no infravermelho próximo (Vis-NIR) para quantificar nutrientes nos tecidos vegetais; iv) desenvolver e validar uma metodologia não destrutiva para avaliar a maturação interna da fruta através do diagnóstico dos níveis de maturação na pele, polpa e sementes, usando uvas e tomates como estudos de caso; v) mapear a maturação dos tecidos internos de acordo com

processos bioquímicos, como variação de açúcar e pigmento para fornecer gestão específica do local; e vi) combinar espectroscopia com biologia de sistemas para fornecer explicações fisiológicas sobre culturas no campo, usando o exemplo do estudo de caso do uso de redes de sombreamento em vinhas.

O Capítulo 1 da tese estabelece o contexto temático da investigação, destacando a importância da AP e as lacunas existentes na compreensão dos processos fisiológicos das plantas sob diferentes condições ambientais. Este capítulo apresenta os desafios enfrentados pela agricultura moderna e a necessidade de desenvolver métodos mais eficazes para monitorar e gerir culturas. O Capítulo 2 delinea os conceitos-chave e teorias que servirão de base para a investigação. Os fundamentos da espectroscopia, IA e biologia de sistemas são apresentados, enfatizando suas aplicações potenciais na AP. Além disso, apresenta as metodologias e abordagens metodológicas empregadas ao longo da tese, fornecendo um quadro teórico para orientar a investigação.

O Capítulo 3 apresenta o primeiro estudo de caso, que compara dados Vis-NIR acoplados com algoritmos de *machine learning* e *self-learning artificial intelligence* (SL-AI) para prever o potencial de água foliar ao amanhecer em um conjunto de dados composto por diferentes cultivares de videira, localizações e anos. Este estudo de caso fornece indicações fisiológicas da fisiologia da folha da videira desde o pintor até a colheita, fornecendo uma melhor percepção para avaliar os parâmetros ecofisiológicos da vinha no campo, como o estado de água e a resposta dos pigmentos foliares a diferentes estresses abióticos, o que é essencial para otimizar o planejamento do sistema de rega. O SL-AI produziu resultados mais robustos no modo de regressão ($R^2 = 0,97$, erro percentual médio absoluto (MAPE) = 18,30%) e no modo de classificação (precisão de 86,27%) em comparação com as Redes Neurais Artificiais com Análise de Componentes Principais (ANN-PCA) (modo de regressão: $R^2 = 0,85$, MAPE = 43,64%; modo de classificação: precisão de 72,37%) e Quadrados Mínimos Parciais (PLS) (modo de regressão: $R^2 = 0,94$, MAPE = 28,76%; modo de classificação: precisão de 73,79%). O SL-AI também revelou uma relação causal entre o estado de água e os comprimentos de onda associados a clorofilas, xantofilas e água.

O Capítulo 4 aborda a caracterização nutricional de tecidos vegetais, apresentando dois estudos de caso considerando duas metodologias diferentes. O estudo de caso II utiliza laser induced breakdown spectroscopy (LIBS) para avaliar a película, polpa e semente de duas cultivares de *Vitis vinifera* em diferentes estágios de maturação, mostrando diferenças significativas ($p < 0,05$) nas concentrações de cálcio, magnésio e nitrogênio entre os tecidos e cultivares de uva. O estudo de caso III aborda a escassez de dados

na AP baseada em fisiologia, introduzindo a ampliação de dados através da hibridação de espectroscopia baseada no modo de covariância (CovM), utilizando técnicas como PLS, mínimos quadrados parciais locais (LocPLS) e SL-AI para avaliar a capacidade de expansão da base de conhecimento. O uso de conjuntos de dados sintéticos prevê com precisão a composição foliar real, especialmente para macronutrientes (N, P, K, Ca, Mg e S) em folhas de uva ($R \sim 0,607 - 0,939$ e $SE \sim 0,043-0,048\%$) e folhas de maçã ($R \sim 0,514-0,721$ e $SE \sim 0,019-0,127\%$); e micronutrientes (Fe, Mn, Zn, Cu e B) em folhas de uva ($R \sim 0,658 - 0,91$ e $SE \sim 0,875-3,978$ ppm) e folhas de maçã ($R \sim 0,525-0,76$ e $SE \sim 8,892-37,89$ ppm). Quando os dados reais não são suficientemente representativos, a precisão e confiabilidade das análises são comprometidas. Técnicas de ampliação de dados podem ajudar a contornar este problema fornecendo uma maior diversidade de dados sintéticos que refletem uma gama mais ampla de cenários possíveis, assim aumentando a robustez e generalização dos modelos desenvolvidos.

O Capítulo 5 apresenta uma nova técnica para avaliar a evolução da maturação dos frutos. Este capítulo introduz uma patente que propõe um método e dispositivo inovadores para caracterização tomográfica não invasiva de amostras de tecido de tomate e uva. Esta patente representa um avanço na tecnologia da AP, oferecendo um método mais eficiente e preciso para a composição de tecidos sem amostragem invasiva ou análise laboratorial. Como prova de conceito, este capítulo apresenta o estudo de caso IV, reconstruindo os espectros de película, polpa e semente de tomate em vários estágios de maturação usando dados hiperespectrais, fornecendo alta precisão na previsão do conteúdo de sólidos solúveis (SSC), clorofila, licopeno e força de punção. A reconstrução espectral de tecidos internos de tomate, incluindo pele, polpa e sementes, teve MAPE de 30,44%, 5,37% e 6,42%, e coeficientes de correlação de Pearson de 0,85, 0,98 e 0,99 para pele, polpa e semente, respectivamente. Parâmetros de qualidade como SSC, clorofila, licopeno e força de punção foram avaliados e modelados com PLS em conjuntos de dados originais e reconstruídos, demonstrando $R^2 > 0,84$. Esta abordagem oferece uma visão empírica da maturação do tomate em tecidos internos, destacando a dinâmica da clorofila e do licopeno durante o processo de maturação.

Posteriormente, a prova de conceito da patente foi estendida para uvas, apresentando o estudo de caso V, mostrando que os espectros reconstruídos de tecidos internos de uvas Loureiro e Vinhão do pós-verão à colheita não diferiram significativamente dos originais ($p < 0,001$). Loureiro apresentou métricas estatísticas melhores (uva total: $r = 0,99$; pele: $r = 1$; polpa: $r = 1$; semente: $r = 0,94$) do que Vinhão (uva total: $r = 0,92$; pele:

$r = 0,92$; polpa: $r = 0,95$; semente: $r = 0,95$). Usando SL-AI, os parâmetros de qualidade foram previstos com sucesso para ambas as uvas: SSC (%; MAPE < 13%), força de punção (N; MAPE < 29%), teor de clorofila (a.u.; MAPE < 29%), e teor de antocianina (a.u.; MAPE < 17%, apenas Vinhão).

O Capítulo 6 apresenta o Estudo de Caso VI, que investiga a dinâmica da maturação da uva e a variabilidade do terreno dentro da vinha da Região Vinícola do Douro. Utilizando um método semelhante à tomografia com base em dados Vis-NIR coletados de um sensor de ponto de medição, este estudo cria mapas metabólicos para examinar o fenótipo da uva em diferentes níveis de rega (sem rega, 30% Etc e 60% Etc) ao longo de dois anos consecutivos (2021 e 2022). Estes mapas metabólicos representam as interações bioquímicas que ocorrem durante a maturação da uva, fornecendo percepções sobre os processos metabólicos subjacentes e facilitando o refinamento de estratégias de gestão mais precisas. Medições não destrutivas foram realizadas para avaliar parâmetros físico-químicos como SSC, força de punção (PF), volume, relação peso/volume, clorofila e antocianina na pele, polpa e semente em seis datas após o pós-verão. Através de uma análise comparativa entre *random forest* (RF) e SL-AI, SL-AI demonstra uma robustez superior dentro do conjunto de dados destrutivo (SL-AI – $R^2 \geq 0,87$, MAPE $\leq 37,55\%$; RF – $R^2 \geq 0,86$, MAPE $\leq 46,20\%$). Além disso, simulações baseadas em variáveis biofísicas para objetivos de vinificação contribuem para o avanço das práticas vitivinícolas, oferecendo percepções sobre as respostas da videira às características ambientais e do terreno. Considerando os princípios de gestão específica do local, são implementadas estratégias de alocação de recursos e agronómicas adaptadas para atender às necessidades específicas ao longo da vinha. Esta abordagem melhora a eficiência da produção de vinho através de práticas de viticultura de precisão (PV), permitindo que as intervenções sejam calibradas e adaptadas com precisão para otimizar a maturação da uva em diferentes seções da vinha.

O Capítulo 7 apresenta como a biologia de sistemas pode ser incorporada na AP para abordar a falta de explicação dos processos fisiológicos. Esta demonstração é apresentada através do estudo de caso VII, descrevendo o impacto das redes de sombreamento na fotossíntese da videira. Este estudo explora as respostas da videira ao sombreamento induzido por cortinas artificiais, empregando análises bioquímicas, dados hiperespectrais e técnicas de biologia de sistemas. Realizado na Região Vinícola do Douro, em Portugal, o sombreamento durante o pós-pintor reduziu o estresse hídrico nas videiras sombreadas em comparação com as não sombreadas. A análise

hiperespectral revelou níveis de clorofila a-b 150% mais altos nos tratamentos sombreados, enquanto os níveis de ROS foram 51,22% mais altos nos tratamentos não sombreados. O modelo SL-AI previu com precisão a clorofila a (R^2 : 0,92, MAPE: 24,39%), clorofila b (R^2 : 0,96, MAPE: 17,61%) e níveis de ROS (R^2 : 0,76, MAPE: 52,17%). Simulações in silico identificaram fenótipos distintos para videiras sombreadas e não sombreadas, destacando o papel da redutase do ferredoxina do cloroplasto. A integração destas técnicas melhora a compreensão das respostas da videira ao estresse ambiental, com futuras pesquisas focadas em mecanismos moleculares para refinar as práticas vitícolas para uma produção sustentável de uvas.

Esta tese visa melhorar as práticas de AP através da integração de informações hiperespectrais, IA e biologia de sistemas. Metodologias avançadas para diagnosticar processos agronômicos são desenvolvidas abordando lacunas na compreensão dos processos fisiológicos das plantas em diferentes ambientes. A eficácia da combinação de espectroscopia, IA e biologia de sistemas na melhoria das práticas de monitorização e gestão de culturas é demonstrada através de uma série de estudos de caso, que vão desde a previsão do estado de água da videira até a avaliação da maturação da fruta e a integração da biologia de sistemas na AP. Esta tese contribui para uma compreensão mais profunda de como o genótipo, ambiente e práticas culturais influenciam a fisiologia e o metabolismo das plantas, oferecendo insights valiosos para uma gestão mais precisa e eficaz dos recursos agrícolas.

Os resultados desta tese destacam o potencial da integração de informações hiperespectrais, IA e biologia de sistemas na AP. No entanto, há uma necessidade de focar em áreas adicionais de pesquisa e melhoria. A padronização e validação de metodologias em diferentes culturas e condições ambientais são cruciais para aumentar sua aplicabilidade e transferibilidade. Além disso, é essencial elevar o nível de prontidão tecnológica (TRL) dessas abordagens para adoção na prática agrícola. Investigações futuras devem se concentrar em refinar os protocolos de aquisição e modelação de dados e explorar algoritmos de machine learning, técnicas de IA e técnicas de análise espectral para aumentar a precisão diagnóstica. Investigar os efeitos a longo prazo e a escalabilidade da implementação dessas metodologias em ambientes agrícolas reais é crucial para sua adoção generalizada e impacto na sustentabilidade agrícola. Esforços contínuos de investigação interdisciplinar são necessários para explorar totalmente o potencial da integração de informações hiperespectrais, IA e biologia de sistemas na AP, visando melhorar a gestão de culturas e promover a sustentabilidade.

Palavras-chave: agricultura de precisão, tomografia, espectroscopia, maturação de frutas, SSC, força de perfuração, fotossíntese.

Abstract

Precision agriculture (PA) has emerged as a crucial tool in modern farming, aiming to optimise crop productivity, quality, and competitiveness while reducing resource usage and minimising environmental impact. Compared to traditional crop monitoring methods, which rely on visual phenotype observation and destructive sampling in the field, PA offers a more efficient and accurate alternative. These traditional methods are labour-intensive and prone to errors, potentially resulting in inaccurate agronomic decisions such as irrigation scheduling or harvest timing. Recent advances in spectroscopy, artificial intelligence (AI), and systems biology present promising opportunities to enhance agricultural monitoring and management practices.

There is a gap in understanding plant physiological processes under different environmental conditions. To address the need for a genotype, environment, and cultural practices approach, it is essential to fill this gap in understanding plant physiological processes under various environmental conditions. Current research often lacks a comprehensive modelling tool combined with spectral information to apply these methodologies across different crops and environments, hindering their transferability between different years and locations. This thesis aims to integrate hyperspectral information with AI and systems biology, allowing exploration of how plant physiology/metabolism is determined by genotype x environment x cultural practices, which is crucial for better understanding how metabolite profiles depend on these factors.

In response to these challenges, this thesis aims to develop and validate advanced methodologies for diagnosing agronomic processes based on systems biology principles for PA. Specifically, the thesis aims to: i) provide physiological explanations for the vine's water status; ii) present low-cost methods with minimal laboratory preparation based on spectroscopy to assess nutrient composition in small amounts of plant tissue; iii) use spectroscopy techniques based on the visible and near-infrared (Vis-NIR) zones to quantify nutrients in plant tissues; iv) develop and validate a non-destructive methodology to assess fruit internal ripening by diagnosing levels of ripeness in the skin, pulp, and seeds, using grapes and tomatoes as case studies; v) map the maturation of internal tissues according to biochemical processes, such as sugar and pigment variation to provide site-specific management (SSM); and vi) combine spectroscopy with systems biology to provide physiological explanations about crops in the field, using the example case study of using shading curtains in vineyards.

Chapter 1 of the thesis establishes the thematic context of the research, highlighting the importance of PA and the existing gaps in understanding plant physiological processes under different environmental conditions. This chapter presents the challenges modern agriculture faces and the need to develop more effective methods for monitoring and managing crops. Chapter 2 outlines the key concepts and theories that will serve as foundations for the research. The fundamentals of spectroscopy, AI, and systems biology are presented, emphasising their potential applications in PA. Additionally, it presents the methodologies and methodological approaches employed throughout the thesis, providing a theoretical framework to guide the research.

Chapter 3 presents the first case study, which benchmarks Vis-NIR data coupled with machine learning algorithms and self-learning artificial intelligence (SL-AI) to predict leaf water potential at dawn in a dataset composed of different grapevine cultivars, locations, and years. This case study provides physiological indications of grapevine leaf physiology from post-veraison to harvest, providing a better perception to evaluate the ecophysiological parameters of vineyards in the field, such as the water status and the response of leaf pigments to different abiotic stresses, which is essential for optimising the irrigation system planning. SL-AI yielded more robust results in regression mode ($R^2 = 0.97$, mean absolute percentage error (MAPE) = 18.30%) and classification mode (accuracy of 86.27%) compared to Artificial Neural Networks with Principal Component Analysis (ANN-PCA) (Regression mode: $R^2 = 0.85$, MAPE = 43.64%; classification mode: accuracy of 72.37%) and Partial Least Square (PLS) (Regression mode: $R^2 = 0.94$, MAPE = 28.76%; classification mode: accuracy of 73.79%). SL-AI also revealed a causal relationship between water status and wavelengths associated with chlorophylls, xanthophylls, and water.

Chapter 4 addresses the nutritional characterisation of plant tissues, presenting two case studies considering two different methodologies. Case study II uses laser-induced breakdown spectroscopy (LIBS) to assess the skin, pulp, and seed of two *Vitis vinifera* cultivars at different ripening stages, showing significant differences ($p < 0.05$) in calcium, magnesium, and nitrogen concentrations among grape tissues and cultivars. Case study III addresses data scarcity in physiology-based PA, introducing data augmentation through spectroscopy hybridisation based on covariance mode (CovM), using techniques like PLS, local partial least squares (LocPLS), and SL-AI to evaluate the knowledge base expansion capability. The use of synthetic datasets accurately predicts actual leaf composition, especially for macronutrients (N, P, K, Ca, Mg, and S) in grape leaves ($R \sim 0.607 - 0.939$ and $SE \sim 0.043-0.048\%$) and apple leaves ($R \sim 0.514-0.721$).

and SE ~ 0.019-0.127%); and micronutrients (Fe, Mn, Zn, Cu, and B) in grape leaves ($R \sim 0.658 - 0.91$ and SE ~ 0.875-3.978 ppm) and apple leaves ($R \sim 0.525-0.76$ and SE ~ 8.892-37.89 ppm). When real data is not sufficiently representative, the accuracy and reliability of analyses are compromised. Data augmentation techniques can help bypass this issue by providing a greater diversity of synthetic data reflecting a wider range of possible scenarios, thus enhancing the robustness and generalisation of the developed models.

Chapter 5 presents a new technique for evaluating fruit ripening evolution. This chapter introduces a patent proposing an innovative method and device for non-invasive tomographic characterisation of tomato and grape tissue samples. This patent represents an advancement in PA technology, offering a more efficient and accurate method for tissue composition without invasive sampling or laboratory analysis. As a proof of concept, this chapter presents case study IV, reconstructing tomato skin, pulp, and seed spectra at various ripening stages using hyperspectral data providing high precision in predicting soluble solids content (SSC), chlorophyll, lycopene, and puncture force. The spectral reconstruction of tomato internal tissues, including skin, pulp, and seeds, had MAPE of 30.44%, 5.37%, and 6.42%, and Pearson correlation coefficients of 0.85, 0.98, and 0.99 for skin, pulp, and seed, respectively. Quality parameters such as SSC, chlorophyll, lycopene, and puncture force were evaluated and modelled with PLS in original and reconstructed datasets, demonstrating $R^2 > 0.84$. This approach offers an empirical view of tomato ripening in internal tissues, highlighting the dynamics of chlorophyll and lycopene during the ripening process.

Subsequently, the patent proof of concept was extended to grapes, presenting case study V, showing that reconstructed spectra of internal tissues of Loureiro and Vinhão grapes from post-veraison to harvest did not differ significantly from the originals ($p < 0.001$). Loureiro presented better statistical metrics (total grape: $r = 0.99$; skin: $r = 1$; pulp: $r = 1$; seed: $r = 0.94$) than Vinhão (total grape: $r = 0.92$; skin: $r = 0.92$; pulp: $r = 0.95$; seed: $r = 0.95$). Using SL-AI, quality parameters were successfully predicted for both grapes: SSC (%; MAPE < 13%), puncture force (N; MAPE < 29%), chlorophyll content (a.u.; MAPE < 29%), and anthocyanin content (a.u.; MAPE < 17%, only Vinhão).

Chapter 6 introduces Case Study VI, which investigates the dynamics of grape maturation and terrain variability within the Douro Wine Region vineyard. Employing a tomography-like method based on Vis-NIR data collected from a point-of-measurement sensor, this study creates metabolic maps to scrutinise grape phenotype across varying irrigation levels (no irrigation, 30% Etc, and 60% Etc) over two consecutive years (2021

and 2022). These metabolic maps represent the biochemical interactions occurring during grape maturation, providing perceptions into the underlying metabolic processes and facilitating the refinement of more precise management strategies. Non-destructive measurements were conducted to assess physicochemical parameters such as soluble solids content (SSC), puncture force (PF), volume, weight/volume ratio, chlorophyll, and anthocyanin in the skin, pulp, and seed across six post-veraison dates. Through a comparative analysis between random forest (RF) and SL-AI algorithms, SL-AI demonstrates superior robustness within the destructive dataset (SL-AI – $R^2 \geq 0.87$, $MAPE \leq 37.55\%$; RF – $R^2 \geq 0.86$, $MAPE \leq 46.20\%$). Moreover, simulations based on biophysical variables for winemaking goals contribute to advancing viticulture practices, offering insights into grapevine responses to environmental and terrain characteristics. Embracing site-specific management (SSM) principles, tailored resource allocation and agronomic strategies are implemented to address specific needs throughout the vineyard. This approach enhances wine production efficiency through precision viticulture (PV) practices, enabling interventions to be precisely calibrated and tailored to optimise grape maturation across different vineyard sections.

Chapter 7 presents how systems biology can be incorporated into PA to address the lack of explanation of physiological processes. This demonstration is presented through case study VII, describing the impact of shading nets on vine photosynthesis. This study explores grapevine responses to shading induced by artificial curtains, employing biochemical analyses, hyperspectral data, and systems biology techniques. Conducted in Portugal's Douro Wine Region, shading during post-veraison reduced water stress in shaded vines compared to unshaded ones. Hyperspectral analysis revealed 150% higher chlorophyll a-b levels in shaded treatments, while ROS levels were 51.22% higher in unshaded treatments. The SL-AI model accurately predicted chlorophyll a (R^2 : 0.92, $MAPE$: 24.39%), chlorophyll b (R^2 : 0.96, $MAPE$: 17.61%), and ROS levels (R^2 : 0.76, $MAPE$: 52.17%). In silico simulations identified distinct phenotypes for shaded and unshaded vines, highlighting the role of chloroplast ferredoxin Reductase. Integrating these techniques enhances understanding of grapevine responses to environmental stressors, with future research focusing on molecular mechanisms to refine viticultural practices for sustainable grape production.

This thesis aims to enhance PA practices by integrating hyperspectral information, AI, and systems biology. Advanced methodologies for diagnosing agronomic processes are developed by addressing gaps in understanding plant physiological processes in different environments. The effectiveness of combining spectroscopy, AI, and systems

biology in improving crop monitoring and management practices is demonstrated through a series of case studies, ranging from predicting vine water status to assessing fruit ripening and integrating systems biology into PA. This thesis contributes to a deeper understanding of how genotype, environment, and cultural practices influence plant physiology and metabolism, offering valuable insights for more precise and effective management of agricultural resources.

The results of this thesis highlight the potential of integrating hyperspectral information, AI, and systems biology in PA. However, there is a need to focus on further research and improvement areas. Standardisation and validation of methodologies across different crops and environmental conditions are crucial to enhance their applicability and transferability. Additionally, it is essential to elevate the technology readiness level (TRL) of these approaches for adoption in agricultural practice. Future investigations should focus on refining data acquisition and modelling protocols and exploring machine learning algorithms, AI techniques, and spectral analysis techniques to increase diagnostic accuracy. Investigating the long-term effects and scalability of implementing these methodologies in real agricultural environments is crucial for their widespread adoption and impact on agricultural sustainability. Continuous interdisciplinary research efforts are necessary to fully exploit the potential of integrating hyperspectral information, AI, and systems biology in PA, aiming to improve crop management and promote sustainability.

Keywords: precision agriculture, tomography-like, spectroscopy, fruit maturation, SSC, puncture force, photosynthesis.

Publications and communications

Publications in international peer-reviewed journals

The following outputs were reached in the context of the present PhD thesis:

Published

Scientific research articles

Tosin, R., Cunha, M., Monteiro-Silva, F., Santos, F., Barroso, T., and Martins, R. (2024). Bi-directional hyperspectral reconstruction of cherry tomato: diagnosis of internal tissues maturation stage and composition. *Frontiers in Plant Science*. doi: <https://doi.org/10.3389/fpls.2024.1351958>. Q1: PLANT SCIENCES – SCIE 27/239; JIF: 5.6; percentile: 88.9.

Tosin, R., Monteiro-Silva, F., Martins, R., and Cunha, M. (2024). A New Approach for Element Characterization of Grapevine Tissue with Laser-Induced Breakdown Spectroscopy. *Horticulturae* 10, 82. doi: <https://doi.org/10.3390/horticulturae10010082>. Q1: HORTICULTURE - SCIE 6/36; JIF: 3.1; percentile: 84.7.

Tosin, R., Monteiro-Silva, F., Martins, R., and Cunha, M. (2023). Precision maturation assessment of grape tissues: Hyperspectral bi-directional reconstruction using tomography-like based on multi-block hierarchical principal component analysis. *Biosystems Engineering* 236, 147-159. doi: <https://doi.org/10.1016/j.biosystemseng.2023.10.011>. Q1: AGRICULTURE, MULTIDISCIPLINARY - SCIE 8/58 and AGRICULTURAL ENGINEERING – SCIE 4/14; JIF: 5.1; percentile: 81.1.

Tosin, R., Martins, R., Pôças, I., and Cunha, M. (2022). Canopy VIS-NIR spectroscopy and self-learning artificial intelligence for a generalised model of predawn leaf water potential in *Vitis vinifera*. *Biosystems Engineering* 219, 235-258. doi: <https://doi.org/10.1016/j.biosystemseng.2022.05.007>. Q1: AGRICULTURE, MULTIDISCIPLINARY - SCIE 8/58 and AGRICULTURAL ENGINEERING – SCIE 4/14; JIF: 5.1; percentile: 81.1.

Conference proceedings (indexed in Scopus/WoS)

Tosin, R., Martins, R., and Cunha, M. (2023). Tomography-like for hyperspectral bi-directional grape tissue reconstruction based on machine learning: Implications for diagnosis composition and precision maturation monitoring. *BIO Web of Conferences* 68. doi: <https://doi.org/10.1051/bioconf/20236801017>

Tosin, R., Monteiro-Silva, F., Martins, R., and Cunha, M. (2023). "LIBS-Based Analysis of Elemental Composition in Skin, Pulp, and Seeds of White and Red Grape Cultivars", in: *Csac 2023*). doi: <https://doi.org/10.3390/CSAC2023-14918>

Submitted or in preparation

Combining spectral sensing and system biology for leveraging precision viticulture: a case study unveiling the impact of shade curtains on grape physiology

Metabolic maps for precision grape maturation: integrating physicochemical parameters using sensing high-throughput tomography-like method to define site-specific management

Spectral data augmentation for dealing with data scarcity in leaf nutrient uptake quantification

Patent

Martins, R.C., Santos, F., Cunha, M., Monteiro-Silva, F., **Tosin, R.**, Magalhães, S., and Pereira, M.R. (2023). *Method and device for non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues* [Online]. Available: <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2023126532& cid=P12-LK2EKH-19952-1> [Accessed 29/01/2024].

Conference proceedings (others)

Santos-Campos, M., **Tosin, R.**, Rodrigues, L., Gonçalves, I., Barbosa, C., Martins, R., Santos, F., and Cunha, M. (2023). "Enhancing Grape Brix Prediction in Precision Viticulture: A Benchmarking Study of Predictive Models Using Hyperspectral Proximal Sensors". doi: <https://doi.org/10.3390/IECAG2023-15914>

Cunha, M., **Tosin, R.**, Novo, H., Teixeira, J., Fontes, N., Graça, A., and Pôças, I. (2020). Integração da reflectância hiperespectral e do conteúdo de pigmentos da folha para estimativa do potencial hídrico foliar de base da videira na Região do Douro. *SibEH 2020. II Simpósio Ibérico de Engenharia Hortícola 2020*.

Publications in national technical magazines

Tosin, R., Martins, R.C., Santos, F., and Cunha, M. (2022). Sensores proximais hiperespectrais e algoritmos computacionais para fenotipagem digital de parâmetros fisiológicos da videira. *Rev. Assoc. Port. Hortic. 2022* 146, 24-26.

Oral communications in congress

Tosin, R., Martins, R., and Cunha, M. (2023). Tomography-like for hyperspectral bi-directional grape tissue reconstruction based on machine learning: Implications for diagnosis composition and precision maturation monitoring. *44th World Congress of Vine and Wine*.

Posters in scientific conferences

Tosin, R., Silva, F., Reis Pereira, M., Testa, V., Jorge, P., Ferreira, M., Santos, F., Magalhães, S., Moura, P., Aguiar, A., Boaventura, J., Cunha, M., and Martins, R.(2021). Laser induced breakdown spectroscopy for monitoring Vitis-vinifera nutritional uptake and grape maturity evaluation. *11th Euro-Mediterranean Symposium on Laser-Induced Breakdown Spectroscopy 2021*.

Awards

Better communication

Cunha, M., **Tosin, R.**, Novo, H., Teixeira, J., Fontes, N., Graça, A., and Pôças, I. (2020). Integração da reflectância hiperespectral e do conteúdo de pigmentos da folha para estimativa do potencial hídrico foliar de base da videira na Região do Douro. *SibEH 2020. II Simpósio Ibérico de Engenharia Hortícola 2020*.

Prize

Winner in the category “Agro-indústria 4.0” of the “8ª edição do Prémio Empreendedorismo e Inovação CA” with the project SpectOM - Non-Invasive spectroscopy tomographic technology.

Scientific Award by University of Porto, Fundação Amadeus Dias and Santander-Universities Grant with the SpectOM - Non-Invasive spectroscopy tomographic technology.

Table of Contents

List of tables.....	xxv
List of supplementary tables	xxvii
List of figures.....	xxix
List of supplementary figures.....	xxxv
List of graphical abstracts	xxxvii
List of abbreviations.....	xxxix
1. Introduction.....	5
2. Conceptual framework	17
2.1. Precision agriculture	17
2.1.1. Precision agriculture challenges	18
2.1.2. Spectral optical sensors.....	19
2.2. Data collection and modelling in precision agriculture.....	22
2.3. Non-destructive monitoring of internal and external fruit maturation tissues	23
2.3.1. Fruit maturation dynamics: critical considerations for precision agriculture	24
2.3.2. Fruit maturation: a thorough analysis	25
2.3.3. Tomato maturation: biochemical and biophysical transformations	26
2.3.4. Grape maturation: biochemical and biophysical transformations.....	27
2.3.5. Nutrient and environmental influence on grape and tomato maturation	28
2.3.6. Tomography-like method: assessing fruit maturation	29
2.4. System biology applied in precision agriculture	32
2.4.1. Leveraging molecular biology and bioinformatics for enhanced crop	
phenotyping	35
3. Case study I: study on predawn leaf water potential and its impact on post-veraison pigments.	41
Canopy VIS-NIR spectroscopy and self-learning artificial intelligence for a generalised	
model of predawn leaf water potential in <i>Vitis vinifera</i>	43
Highlights	43
Abstract.....	43
Keywords	44
3.1. Introduction	44
3.2. Material and Methods	57
3.2.1. Test site	57
3.2.2. Predawn leaf water potential.....	57
3.2.3. Spectroscopy methodology.....	58
3.2.4. Self-learning artificial intelligence	58
3.2.4.1. Benchmark on the remote sensing-based previsions of Ψ	58
3.2.4.2. Performance of SL-AI against other modelling techniques	59
3.2.5. VIS-NIR regions of interest in modelling Ψ_{pd}	61
3.3. Results	61

3.3.1.	Data characterisation	61
3.3.2.	Performance of SL-AI against other modelling techniques	62
3.3.3.	Physiological insights in modelling Ψ_{pd}	68
3.3.4.	Benchmark on the remote sensing-based previsions of Ψ	69
3.4.	Discussion	71
3.5.	Conclusion	74
	Acknowledgments	74
	CRedit author statement	74
	Supplementary material	76
	Figures	76
	Tables.....	78
4.1.	Case study II: LIBS for nutrient identification in small biological tissues.	83
	A New Approach for Element Characterisation of Grapevine Tissue with Laser-Induced Breakdown Spectroscopy	85
	Abstract.....	85
	Keywords	86
4.1.1.	Introduction.....	86
4.1.2.	Materials and Methods	88
4.1.2.1.	Test Site.....	88
4.1.2.2.	Characterisation of the Grapes	89
2.3.	LIBS Instrumentation	90
2.4.	Element Identification	91
2.5.	Statistical Analysis.....	91
4.1.3.	Results	91
4.1.4.	Discussion	96
4.1.5.	Conclusions.....	100
	Author Contributions.....	100
	Funding:	100
	Data Availability Statement:.....	101
	Acknowledgments	101
	Conflicts of Interest:.....	101
	Disclaimer/Publisher's Note:.....	101
4.2.	Case study III: quantitative Vis-NIR for nutrient identification in apple and grape leaves.	103
	Spectral Data Augmentation for Dealing with Data Scarcity in Leaf Nutrient Uptake Quantification	105
	Highlights	105
	Abstract.....	105
	Keywords	106
4.2.1.	Introduction.....	106

4.2.2.	Materials and Methods	109
4.2.2.1.	Experimental Design.....	109
4.2.2.2.	Sampling	109
4.2.2.3.	Spectroscopy.....	110
4.2.2.4.	Element quantification	111
4.2.2.5.	Spectral Data Augmentation	111
4.2.2.6.	Benchmarking.....	112
4.2.2.7.	Validation.....	113
4.2.3.	Results and Discussion	114
4.2.3.1.	Spectral characterisation	114
4.2.3.2.	Significance and Representativity	117
4.2.3.3.	Nutrient quantification	119
4.2.3.4.	Bias-Variance Analysis	123
4.2.3.5.	Data Augmentation Analysis	126
4.2.4.	Conclusions.....	130
	CRedit authorship contribution statement.....	130
	Declaration of Competing Interest	130
	Acknowledgements	131
	Supplementary Material	132
	Tables.....	132
5.1.	Patent: Tomography-like method for internal fruit tissue assessment. ...	141
	Method and device for non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues	143
	Inventors	143
	Applicants	143
	Abstract.....	143
5.1.1.	Description	143
5.1.1.1.	Technical field.....	143
5.1.1.2.	Background	143
5.1.2.	General description	144
5.1.3.	Detailed description	147
5.1.4.	Claims	152
	Acknowledgments	154
5.2.	Case study IV: Tomato case study: identifying tissue changes during maturation.....	155
	Bi-directional hyperspectral reconstruction of cherry tomato: diagnosis of internal tissues maturation stage and composition.....	157
	Abstract.....	157
	Keywords	158
5.2.1.	Introduction.....	158

5.2.2.	Material and methods	162
5.2.2.1.	Sampling and tomato properties	162
5.2.2.2.	Spectroscopy.....	163
5.2.2.3.	Hierarchical latent structures reconstruction	164
5.2.2.4.	Association and bi-directionality.....	165
5.2.2.5.	Validation.....	166
5.2.2.6.	Prediction of tomato quality.....	167
5.2.3.	Results	170
5.2.3.1.	Tomato tissue reconstruction	170
5.2.3.2.	Quality parameters evaluation	173
5.2.4.	Discussion	175
5.2.5.	Conclusion.....	179
	Conflict of Interest	180
	Author Contributions.....	180
	Acknowledgements	180
	Supplementary Material	181
	Figures	181
5.3.	Case study V: Grape case study: comparing cultivars, tracking quality changes during maturation.....	185
	Precision maturation assessment of grape tissues: hyperspectral bi-directional reconstruction using tomography-like based on multi-block hierarchical principal component analysis.....	187
	Highlights	187
	Abstract.....	187
	Keywords	188
5.3.1.	Introduction.....	188
5.3.2.	Material and Methods	191
5.3.2.1.	Test site.....	191
5.3.2.2.	Grape measurements	191
5.3.2.3.	Spectral measurements.....	193
5.3.2.4.	Bi-directional spectral reconstruction	194
5.3.2.5.	Validation of spectral reconstruction	196
5.3.2.6.	Predictive modelling of SSC, puncture force and pigments.....	197
5.3.2.7.	Data characterisation.....	198
5.3.3.	Results	198
5.3.3.1.	Grape characterisation.....	198
5.3.3.2.	Bi-directional spectral reconstruction	200
5.3.3.3.	Spectral patterns of grapes composition during maturation.....	203
5.3.3.4.	Modelling grape SSC, puncture force and pigments using spectral data	205

5.3.3.5. Biophysical parameters along the maturation	206
5.3.4. Discussion	208
5.3.5. Conclusion.....	211
CRedit authorship contribution statement.....	212
Declaration of interests.....	212
Acknowledgements	212
Supplementary material	213
Algorithms.....	213
Figures	215
Table	223
6. Case study VI: Douro wine region case study: metabolic maps for tissue evolution and maturation.	227
Metabolic maps for precision grape maturation: integrating physicochemical parameters using sensing high-throughput tomography-like method to define site-specific management	229
Abstract.....	229
Keywords	229
6.1. Introduction	229
6.2. Material and Methods.....	233
6.2.1. Test site.....	233
6.2.2. Vineyard's spatiotemporal variability based on satellite imagery	234
6.2.3. Experimental design of grape sampling	234
6.2.4. Grape measurements	236
6.2.4.1. Spectra measurement.....	236
6.2.4.2. Physicochemical parameters measurements	236
6.2.4.3. Spectra processing	237
6.2.5. Tomography-like method	237
6.2.5.1. Identifying internal tissues of non-destructive grapes	239
6.2.6. Modelling and mapping physicochemical parameters	239
6.2.6.1. Decision maps criteria based on site-specific management	240
6.3. Results	240
6.3.1. Meteorology during the maturation period.....	240
6.3.2. Vineyard spatiotemporal variability	241
6.3.3. Tomography-like of grapes	243
6.3.4. Estimating the physicochemical variables based on the destructive dataset	244
6.3.5. Predicting and mapping the physicochemical variables on the non-destructive dataset.....	245
6.3.5.1. The dynamic of chlorophyll content in the skin, pulp and seed.....	246
6.3.5.2. The dynamic of anthocyanin content in the skin, pulp and seed.....	247
6.3.5.3. Assessment of physicochemical parameters to harvest	249

6.3.5.4. Site-specific management.....	250
6.4. Discussion.....	251
6.5. Conclusion	254
Acknowledgements	255
Supplementary material	256
Supplementary figures.....	256
Supplementary tables.....	258
7. Case Study VII: Combining Spectral Sensing and Systems Biology for Leveraging Precision Viticulture: Unravelling Photosynthetic Pathways in the Douro Wine Region - Assessing the Impact of Shade Treatment on Grape Physiology	265
Combining spectral sensing and system biology for leveraging precision viticulture: a case study unveiling the impact of shade curtains on grape physiology	267
Abstract.....	267
Keywords	268
7.1. Introduction	268
7.2. Material and Methods.....	274
7.2.1. Test Site and Sampling.....	274
7.2.2. Phenotype characterisation	274
7.2.3. Biochemical analysis	275
7.2.4. Spectral biochemical modelling approach.....	275
7.2.5. <i>In silico</i> simulations.....	276
7.3. Results and Discussion	277
7.3.1. Phenotype characterisation	277
7.3.2. Modelling biochemical variables	280
7.3.3. <i>In silico</i> simulations.....	281
7.3.4. Innovation for precision viticulture	286
7.4. Conclusions.....	288
Acknowledgements	289
8. Discussion	294
8.1. Chapter 3: assessing ecophysiological plant parameters in the field	294
8.2. Chapter 4: methodology for conducting a nutritional characterisation of various plant organs	296
8.3. Chapter 5: assessing the evolution of maturation in fruits.....	298
8.4. Chapter 6: performing tomography-like assessments in the field to evaluate maturation	299
8.5. Chapter 7: the relationship between precision agriculture and systems biology	301
8.6. Enhancing consumer awareness and mitigating food waste: a holistic perspective on advanced methodologies for precision agriculture rooted in systems biology	302
8.7. Perspectives and challenges	304

9. Conclusion	310
References	314
A. Technical paper.....	346
Sensores proximais hiperespectrais e algoritmos computacionais para fenotipagem digital de parâmetros fisiológicos da videira em viticultura de precisão	348
Viticultura de precisão fisiológica	348
Sensores proximais hiperespectrais em viticultura de precisão	350
Fenotipagem digital	351
Perspetivas dos sensores proximais na vinha	352
Agradecimentos.....	352

List of tables

Table 1-1. Organisation of scientific questions, corresponding responses, and the presented case studies.	11
Table 3-1. Bibliography review of the prediction of Ψ in a regressive approach.	48
Table 3-2. Bibliography review of the prediction of Ψ in a classification approach.	55
Table 3-3. Statistical results of predawn leaf water potential (Ψ_{pd} , MPa) for the different test sites, irrigation regime and cultivars among the three years of study.	62
Table 3-4. Confusion matrix of the classification of predawn leaf water potential (Ψ_{pd}) in the three classes established (Low: $\Psi_{pd} \geq -0.25$ MPa, Medium: -0.25 MPa $> \Psi_{pd} \geq -0.5$ MPa and High: $\Psi_{pd} < -0.5$ MPa) using the Self-Learning Artificial Intelligence, Artificial Neural Networks with Principal Component Analysis and Partial Least Square.	67
Table 4.1-1. Sampling dates of the Loureiro and Vinhão cultivars.	89
Table 4.1-2. The average LIBS spectra intensity peaks of Ca, Mg, and N and water content (% w/w) for each tissue on the three assessment dates (S1, S2, and S3) for the Loureiro and Vinhão cultivars.	93
Table 4.2-1. The F-test p-value of RWD prediction: PLS cross-validated with RWD (PLS-RWD), PLS optimised with SSD (PLS-SSD→RWD), Loc-PLS optimised with SSD (LocPLS-SSD→RWD), and SLAI optimised with SSD (SLAI-SSD U RWD). Non-significant quantitative predictions are marked in underlined bold format (p-values > 0.005).	118
Table 5.2-1. Mean square error (MSE), mean absolute percentage error (MAPE), p-value and Pearson’s correlation coefficient (r) for the reconstruction of tomato spectra and decomposition of the entire tomato spectra into skin, pulp and seeds spectra.	172
Table 5.2-2. Reconstruction benchmark of the tomato, skin, pulp and seed in predicting the soluble solid content (SSC), chlorophyll, lycopene and puncture force. The original dataset exclusively predicted SSC and puncture force for the entire tomato. Chlorophyll and lycopene content were predicted solely in the reconstructed dataset, as these pigments were estimated using real data, rendering their prediction in the original dataset nonsensical.	174
Table 5.3-1. Comparison of volume (cm ³), soluble solids content (SSC), puncture force (PF), anthocyanin (Ant), and chlorophyll (Chl) in total grape (TO), skin (SK), pulp (PU), and seed (SE) of Loureiro and Vinhão cultivars across three assessment dates from days after veraison.	199
Table 5.3-2. Statistics for the spectral reconstruction compared with the original spectra of the total grape and the skin, pulp and seed spectra for the Loureiro and Vinhão cultivar in the 12 assessment dates.	200
Table 5.3-3. Anthocyanin (Ant), chlorophyll (Chl), SSC (%) and puncture force (PF -N) were predicted with the self-learning artificial intelligence (SL-AI) algorithm for the original and reconstructed dataset for the grape tissues of Loureiro and Vinhão cultivar for 12 assessment dates.	205
Table 5.3-4. The statistical description of the reconstructed spectral dataset of the total grape, skin, pulp and seed using self-learning artificial intelligence (SL-AI) in predicting chlorophyll (Chl – a.u.) for the skin, pulp and seed of Loureiro cultivar and the predicted values for the soluble solid content (SSC - %) and puncture force (PF – N) for the total grape in the 05, 24 and 47 days after veraison (DAV).	208
Table 6-1. Field sampling dates and corresponding proximal satellite imagery were selected to represent each sampling date in 2021.	234
Table 6-2. The number of grapes used in destructive and non-destructive measurements during the evaluation dates in 2021 and 2022.	235
Table 6-3. Statistical metrics of the bi-directional spectral reconstruction of the grape’s tissues in 2021 and 2022.	243

Table 6-4. Benchmark of self-learning artificial intelligence (SL-AI) and random forest (RF) in predicting the physicochemical variables of the destructive dataset (n=357) encompassing the data of 2021 (n=191) and 2022 (n=166).	245
Table 7-1. Enzymes and metabolic pathways in grapevines related by light and shading conditions. The table is categorised into two sections: unshaded and shaded. Each section contains columns indicating metabolic pathways, enzyme expression (whether upregulated or downregulated), enzyme function, and relevant references.....	270
Table 7-2. Comparison of shaded and unshaded conditions for chlorophyll a (mg/gFM), chlorophyll b (mg/gFM), chlorophyll a-b (mg/gFM) and ROS O2 (ABS/gFM).	280
Table 7-3. The top 10 reactions ranked through the PCA loadings contribute to the leave's phenotype in shaded and unshaded conditions and the Flux Balance Analysis (FBA) in a.u. for each simulation.	286

List of supplementary tables

Table S 3-1. The number of grapevines sampled per test site, cultivar, and irrigation conditions.	78
Table S 3-2. Dates of predawn leaf water potential measurement and number of each cultivar assessed in each test site in the sampling dates.	78
Table S 3-3. Irrigation events and irrigation amounts (L/Plant/day) per test site and irrigation treatment (adapted from Pôças et al. (2020)).	78
Table S 3-4. Models tested with Self-learning Artificial Intelligence (SL-AI), Artificial Neural Networks with Principal Component Analysis (ANN-PCA) and Partial Least Square Regression (PLS) according to the training, validation, full dataset, year (2014, 2015 and 2017), Cultivar (TN: Touriga Nacional; TF: Touriga Franca; TB: Tinta Barroca) and Test Site (TS 1: Test Site 1 - Quinta dos Aciprestes); TS 2: Test Site 2 - Quinta do Ataíde).	79
Table S 4.2-1. Fertilisation experimental design N-P levels (kg/ha) in Vines.	132
Table S 4.2-2. Apple tree fertilisation experimental design.	132
Table S 4.2-3. Nutrient concentrations dataset characteristics of grapevine and apple tree.	132
Table S 4.2-4. Leaf nutrient macronutrient quantification benchmarks in the bibliography.	133
Table S 4.2-5. Leaf nutrient micronutrient quantification benchmarks in bibliography.	134
Table S 4.2-6. Results benchmarks for PLS, LocPLS and SLAI for real-world data and a combination of real-world data and synthetic samples data in grapevine and apple leaves.	135
Table S 4.2-7. Results benchmark for real-world data vs synthetic samples data for PLS, Loc-PLS and SLAI models in grapevine leaves.	136
Table S 4.2-8. Results benchmark for real-world data vs synthetic samples data for PLS, Loc-PLS and SLAI models in apple leaves.	137
Table S 5.3-1. Comparison of volume (cm ³), soluble solids content (SSC - %), puncture force (PF - N), anthocyanin (Ant - a.u.), and chlorophyll (Chl - a.u.) in total grape (TO), skin (SK), pulp (PU), and seed (SE) of Loureiro and Vinhão cultivars across three assessment dates from days after veraison.	223
Table S 6-1. The statistical mean value of the normalised difference index (NDVI) for 2021 in the six sampling dates S1 (29/07/2021 - 05 days after veraison (DAV)), S2 (05/08/2021 - 12 DAV), S3 (12/08/2021 - 19 DAV), S4 (19/08/2021 - 26 DAV), S5 (26/08/2021 - 33 DAV), and S6 (02/09/2021 - 40 DAV).	258
Table S 6-2. Comparison of predawn leaf water potential (PLWP - MPa) across irrigation treatments in 2021 and 2022 in the six post-veraison dates. For 2021, S1 (29/07/2021 - 05 days after veraison (DAV)), S2 (05/08/2021 - 12 DAV), S3 (12/08/2021 - 19 DAV), S4 (19/08/2021 - 26 DAV), S5 (26/08/2021 - 33 DAV), and S6 (02/09/2021 - 40 DAV) and for 2022: S1 (28/07/2022 - 03 DAV), S2 (04/08/2022 - 17 DAV), S3 (11/08/2023 - 24 DAV), S4 (18/08/2022 - 31 DAV), S5 (25/08/2022 - 31 DAV), and S6 (01/09/2022 - 38 DAV).	259
Table S 6-3. Results of SL-AI in predicting soluble solids content (SSC - %), puncture force (PF - N), Volume (VOL - cm ³), weight per volume (W/VOL - g/cm ³), anthocyanin content in the total grape (ANT-TO - a.u.), skin (ANT-SK - a.u.), pulp (ANT-PU - a.u.) and seed (ANT-SE - a.u.), chlorophyll content in the total grape (CHL-TO - a.u.), skin (CHL-SK - a.u.), pulp (CHL-PU - a.u.) and seed (CHL-SE - a.u.) and the predawn leaf water potential (MPa). Based on the non-destructive dataset (n= 720).	260
Table S 6-4. Results of Random Forest in predicting soluble solids content (SSC - %), puncture force (PF - N), Volume (VOL - cm ³), weight per volume (W/VOL - g/cm ³), anthocyanin content in the total grape (ANT-TO - a.u.), skin (ANT-SK - a.u.), pulp (ANT-PU - a.u.) and seed (ANT-SE - a.u.), chlorophyll content in the total grape (CHL-TO -	

a.u.), skin (CHL-SK – a.u.), pulp (CHL-PU – a.u.) and seed (CHL-SE – a.u.) and the predawn leaf water potential (MPa). Based on the non-destructive dataset (n= 720).261

List of figures

Figure 1-1. Scheme of the current precision agriculture (PA) what is proposed in this thesis.	5
Figure 1-2. Structure of the thesis and organisation of the content to be presented.	9
Figure 2-1. Illustration showcasing various sensors highlighting diverse scales, targeted precision levels, and spectral details.	20
Figure 2-2. Basic principle of the tomography-like method.	31
Figure 2-3. Schematic representation of how system biology can be incorporated in the context of precision agriculture.	34
Figure 3-1. Regressions with Self-learning Artificial Intelligence (SL-AI) in the three years of study (i). Data was performed separately according to the years (a: 2014, b: 2015 and c: 2017), cultivars (d: TN: Touriga Nacional, e: TF: Touriga Franca and f: TB: Tinta Barroca) and test site (g: TS 1: Quinta dos Aciprestes, h: TS 2: Quinta do Ataíde).	64
Figure 3-2. Regressions with Artificial Neural Network with Principal component analysis (ANN-PCA) in the three years of study (i). Data was performed separately according to the years (a: 2014, b: 2015 and c: 2017), cultivars (d: TN: Touriga Nacional, e: TF: Touriga Franca and f: TB: Tinta Barroca) and test site (g: TS 1: Quinta dos Aciprestes, h: TS 2: Quinta do Ataíde).	65
Figure 3-3. Regressions with Partial Least Square (PLS) in the three years of study (i). Data was performed separately according to the years (a: 2014, b: 2015 and c: 2017), cultivars (d: TN: Touriga Nacional, e: TF: Touriga Franca and f: TB: Tinta Barroca) and test site (g: TS 1: Quinta dos Aciprestes, h: TS 2: Quinta do Ataíde).	66
Figure 3-4. Principal Component Analysis (PCA) with the loadings for the test site 2 (Quinta do Ataíde) in the cultivar Touriga Franca in 2015. ‘a’, ‘d’, ‘g’ and ‘j’ represent the PCA with the two first Principal Components (PC 1 and PC 2) with the variance explained in each component. No filled circles mean irrigation treatment (I) and filled squares the non-irrigated treatment (NI). ‘b’, ‘e’, ‘h’ and ‘k’ represent the loadings extracted from PC 1 (Y-axis) per wavelengths (X-axis) corresponding to the group separation presented in ‘a’, ‘d’, ‘g’ and ‘j’. ‘c’, ‘f’, ‘i’ and ‘l’ represent the mean reflectance according to the irrigation regime. The zones of the spectrum were identified according to chlorophyll (Chl), Xanthophyll (Xan) and Water (Wat) zones of the spectrum.	68
Figure 3-5. Benchmark of estimation of water potential (Ψ) based on optical devices and in a regression (‘a’) and classification (‘b’) approach, n: dataset size, c: number of classes of Ψ , PLS: Partial Least Square Regression, SLR: Single Linear Regression Model, MPLS: Modified Partial Least Square Regression, MLR: Multiple Linear Regression, GBM: Generalised Boosted Model, BRNN: Bayesian regularisation-feed-forward neural network, B-MARS: Bagged Multivariate Adaptive Regression Splines, GAM: Generalised Additive Model, ANN: Artificial Neural Network Regression, RF and RF 2: Random Forest, GPR: Gaussian Process Regression, VH-GPR: Variational Heteroscedastic Gaussian Process Regression, BT: Bagging Trees, XGBOOST: Extreme Gradient Boosting, OLR: Ordinal Logistic Regression, SVM: Support Vector Machines, SL-AI: Self-learning artificial intelligence, ANN-PCA: Artificial Neural Network with Principal Component Analysis and SVC: Support Vector Regression. In the regressive approach, squared dots, and in the classification approach, grey bars represent the results performed with the dataset presented in this work. *Due to the similarity performance and numerous variable selection methods, the best result of each methodology tested by Wei et al. (2021) is represented in this figure.	70
Figure 4.1-1. Sample preparation, pellet making, and LIBS analysis of the skin, pulp, and seeds of the Loureiro and Vinhão cultivars.	89
Figure 4.1-2. Experimental setup for Laser-Induced Breakdown Spectroscopy (LIBS), showcasing laser incidence, spectral data acquisition, and manual adjustments for focus. Adapted from Ribeiro et al. (2021).	90

Figure 4.1-3. The average signal intensities of the Ca (■), Mg (▲), and N (●) in the skin, pulp, and seed tissues of the Loureiro cultivar (a–c) and the Vinhão cultivar (d–f) 47 days after veraison (S1). The sample size for each group is specified as n = 20 for Loureiro and n = 16 for Vinhão.	92
Figure 4.1-4. Principal component analysis of Loureiro cultivar across three assessment dates (S1(■), S2 (●), S3(▲)); a-c represent calcium (Ca) concentrations in skin, pulp, and seeds, while d-f show magnesium (Mg) concentrations, and g-i illustrate nitrogen (N) concentrations in the corresponding tissues.	95
Figure 4.1-5. Principal component analysis of Vinhão cultivar across three assessment dates (S1(■), S2 (●), S3(▲)); a-c represent calcium (Ca) concentrations in skin, pulp, and seeds, while d-f show magnesium (Mg) concentrations, and g-i illustrate nitrogen (N) concentrations in the corresponding tissues.	96
Figure 4.2-1. Nutrient quantification using visible near-infrared (Vis-NIR) spectroscopy: (a) Apple and Grapevine leaf sampling for spectral and inductively coupled plasma-optical emission spectroscopy (ICP-OES) analysis; (b) leaf spectral measurement; and (c) data augmentation for exploratory analysis including synthetic spectroscopy dataset (SSD) generation from SSD using hybridisation and validation schemes (1 to 4) to determine SSD representativity of real-world data (RWD).....	110
Figure 4.2-2. Real-world data (RWD) and synthetic sample generation by hybridisation: (a) RWD polygon defining the existing scarce dataset characteristics: edges represent the multi-dimensional boundaries of RWD in the feature space where data can be interpolated, and natural correlations are observed (green area), and non-observable characteristics by extrapolation of convex face of the RWD (red area); (b) synthetic data generated by hybridisation of RWD sample edges preserving the information of the RWD in interpolation (◦) and extrapolation (◻) areas.	112
Figure 4.2-3. Spectra information and nutrient composition relationships: (a) grapevine leaf spectra: (■) real-world data (RWD) and (■) synthetic spectroscopy dataset (SSD); (b) Principal component analysis (PCA) scores and loadings of grape leaf nutrient composition; (c) grapevine leaf spectral PCA scores and nutrient gradient vector; (d) apple tree leaf spectra; (d) PCA scores and loadings apple tree leaf nutrient composition; and (e) grapevine leaf spectral PCA scores and nutrient gradient vector; where (•) are RWD samples and (•) are SSD samples for grapevine and apple tree leads.	115
Figure 4.2-4. State-of-the-art figures of merit (Tables A4 and A5): leaf nutrient quantification figures of merit: (a) Correlation vs Standard error for macronutrients – nitrogen (N) (■), phosphorus (P) (■), potassium (K) (■), calcium (Ca) (■), magnesium (Mg) (■), sulfur (S) (■); (b) Correlation vs Standard error for macronutrients – iron (Fe) (■), manganese (Mn) (■), zinc (Zn) (■), copper (Cu) (■), boron (B) (■); (c) histogram of number of samples used in partial least square (PLS) model; and (d) histogram of number of latent variable (LV) used in PLS model.	120
Figure 4.2-5. Nutrient quantification in grapevine: (a) nitrogen, (b) phosphorus, (c) potassium, (d) calcium, (e) magnesium, (f) sulfur, (g) iron, (h) manganese, (i) zinc, (j) copper and (k) boron - where (•) are RWD samples and (•) are SSD samples for grapevine leaves.....	122
Figure 4.2-6. Nutrient quantification in apple tree leaves: (a) nitrogen, (b) phosphorus, (c) potassium, (d) calcium, (e) magnesium, (f) sulfur, (g) iron, (h) manganese, (i) zinc, (j) copper and (k) boron - where (•) are RWD samples and (•) are SSD samples for apple tree leaves.....	127
Figure 4.2-7. The standard error (SE) and correlation coefficient (R) benchmarks between PLS-RWD, and SLAI-SSD→RWD are for I. grapevine leaves and II—apple leaves.	128
Figure 4.2-8. Correlation vs Standard errors Benchmarks: (a) grapevine leaf macronutrients; (b) grapevine micro-nutrients; (c) apple macro-nutrients; and (c) apple micro-nutrients; where: ■ PLS-RWD, • PLS-SSD→RWD, ◆ PLS-RWD→SSD, ▲PLS-RWD+SSD, × LocPLS-SSD→RWD,+ LocPLS-RWD+SSD, ◦ SLAI-SSD→ RWD, ◻ SLAI-	

RWD+SSD. Macronutrients: nitrogen (N) (■), phosphorus (P) (■), potassium (K) (■), calcium (Ca) (■), magnesium (Mg) (■), sulfur (S) (■). Micronutrients: iron (Fe) (■), manganese (Mn) (■), zinc (Zn) (■), copper (Cu) (■), boron (B) (■).....	129
Figure 5.1-1. Schematic representation of an embodiment of a device, robotic platform and metabolic quantification according to the disclosure.....	147
Figure 5.1-2. Schematic representation of an embodiment of 3D internal spectroscopy reconstruction method using hierarchical latent spaces.	148
Figure 5.1-3. Schematic representation of an embodiment of macroscopic tomography in grapes: measurements made with a robotic device and HR4000 spectrometer to the hole grape; correspondence between spectral variance space of the robotic device and HR4000 spectrometer and its relationship to the tomography fusion space, obtained by using the information from skin, pulp and seed variance sub-spaces.....	149
Figure 5.1-4. Schematic representation of an embodiment of macroscopic tomography reconstruction: 'in- vivo' measurement of two different grapes at veraison and near harvest time, corresponding 'in- silica' reconstruction of the skin, pulp and seed tissues spectra and estimated relevant properties of each tissue.....	150
Figure 5.2-1. Internal tomato tissues and the spectroscopy system used to obtain spectral information from the entire tomato and the respective tissues (skin, pulp and seed).	163
Figure 5.2-2. Representation of the bi-directional reconstruction process and decomposition of the spectral information. T is the superset latent space; T24 the superset latent information of the dimension of the subsets (T1 , T2 , T3); U is the feature space of each subset; U8 the superset latent information of the dimension of the subsets Y is the tomato spectra; P118t is the contribution of each subset (P18 , P28 , P38); X1 , X2 and X3 spectra of the skin, pulp and seed, respectively.....	165
Figure 5.2-3. Spectral bi-directionality in reconstructing tomato components: skin, pulp, and seeds, at different maturation levels. The numbers in the box indicate the observations from the dataset for each maturation stage: ■ (53) for green, ▲ (102) for turning, and ● (5) for red. a) and b) represent the feature space for the total and total reconstructed spectra, respectively. c), d) and e) define the feature space of the skin, pulp and seed, respectively.....	171
Figure 5.2-4. Spectral signatures of the tomato and the internal tissues. In the red line (□) is the original spectral signature, and in the blue dashed line (□) is the reconstructed spectral. Figures a to c are tomatoes subjected to LED light. Figure a) is a tomato in the green stages of maturation; b) tomato in the turning stages of maturation; c) tomato in the red stage of maturation. The p-value < 0.001 indicates no significant difference between the original and reconstructed spectra. r is Pearson's correlation.	172
Figure 5.2-5. Spectral information obtained from tomatoes at different maturation stages and internal tissues and the dynamics of lycopene and chlorophyll during maturation. Chlorophyll content was empirically quantified, with light green representing lower levels and dark green representing higher levels. Lycopene levels were also quantified, with dark red representing lower levels and light red representing higher levels. a), b) and c) are red, turning and green tomatoes over the light source, respectively. The green spectra signature (□) represents the spectral signature of the skin, the orange spectral signature (□) represents the spectral signature of the pulp, and the red spectra signature (□) represents the spectral signature of the seed.	175
Figure 5.3-1. Comparison of spectral information acquisition methods for grapes. (a) Proximal sensing device measurement obtains spectral information until a particular depth within the grape, α is the conical field-of-view. This method allows for the assessment of the grape's spectral characteristics from its surface to a certain internal depth. (b) Point-of-measurement enables the acquisition of spectral data from main internal grape tissues. This method captures spectral information from specific points within the grape's internal structure, providing detailed information into its inner composition and characteristics.	189

Figure 5.3-2. A transversal cut of the grape of Vinhão (a) and Loureiro (b) cultivar to show the light transmittance and morphological description of the tissues and details of the seeds. The figure shows the distinct light transmittance patterns within the grapes and provides the structural characteristics of the seeds, showing their optical characteristics..... 194

Figure 5.3-3. Representation of the spectral reconstruction and decomposition of the grape and the internal tissues. T is the superset latent space; T_i the superset latent information of the dimension of the subsets (T_1, T_2, T_3); U is the feature space of each subset; U_i the superset latent information of the dimension of the subsets of the total grape spectra; P_iT is the contribution of the SK, PU and SE of each subset (P_1T, P_2T, P_3T)..... 195

Figure 5.3-4. Comparative analysis of soluble solids content (SSC - %) levels between Loureiro (blue line) and Vinhão (orange line) cultivars, based on the total rainfall (grey bars; mm) on the sampling dates. 200

Figure 5.3-5. Comparison of original and reconstructed (Rec.) (grey background) spectral signatures for grape cultivars (Loureiro e Vinhão) for total grape and its tissues at different stages of maturation: 5 (red line), 24 (blue line), and 47 (brown line) days after veraison (DAV). For Loureiro, it includes the original (a, e, i, m) and reconstructed (b, f, j, n) spectral signatures of the total grape, skin, pulp, and seed, respectively. Likewise, Vinhão provides the same grape tissues' original (c, g, k, o) and reconstructed (d, h, l, p) spectral signatures..... 202

Figure 5.3-6. Principal component analysis (PCA) illustrating variations of original and reconstructed (Rec.) (light grey background pattern) spectral data for Loureiro and Vinhão cultivars over time (5 (green square), 24 (red circle), and 47 (yellow triangle) days after veraison) and differentiation of grape tissues. For Loureiro, it includes the original (a, e, i, m) and reconstructed (b, f, j, n) spectral signatures of the total grape, skin, pulp, and seed, respectively. Likewise, Vinhão provides the same grape tissues' original (c, g, k, o) and reconstructed (d, h, l, p) spectral signatures..... 203

Figure 5.3-7. Temporal progression of maturation in the Vinhão cultivar, presented through the reconstructed spectral dataset of the total grape, skin, pulp and seed at 5, 24, and 47 days after veraison (DAV). Anthocyanin (Ant.) and chlorophyll (Chl.) in the skin, pulp, and seed (expressed in arbitrary units, a.u.), and predicted values of soluble solids content (SSC - %) and puncture force (PF - N) for the total grape. Reconstructed spectra of the skin (brown line), pulp (green line), and seed (orange line) for the respective DAVs. A t-test compared the observed (Table 5.3-1) with the predicted values of Ant., Chl., SSC, and PF using the reconstructed datasets through self-learning artificial intelligence (SL-AI). * indicates statistical significance at $p < 0.05$, and "ns" denotes non-significance..... 207

Figure 6-1. Description of the sampling approach employed in 2021 and 2022. In 2021, 20 grapes per irrigation line were selected for spectral analysis, with 32 samples randomly chosen from the 120 collected per date for subsequent assessment of internal tissues. S4 is the fourth set of grapevines in the row with a lower altitude, whereas S9 is the ninth sector with a higher altitude. 235

Figure 6-2. Exemplification of the tomography-like method. P_{it} is the contribution of the SK, PU and SE of each subset (P_1T, P_2T and P_3T). The variables ($n=3648$) refer to each principal component (PC) of the spectral data. Sample ($n=357$) refers to the destructive measurements of grapes that were used to construct the knowledge base..... 238

Figure 6-3. Daily mean temperature and total precipitation between 1 July and 30 September 2021 and 2022. In 2021 S1 (29/07/2021), S2 (05/08/2021), S3 (12/08/2021), S4 (19/08/2021), S5 (26/08/2021), and S6 (02/09/2021). In 2022, S1 (28/07/2022), S2 (04/08/2022), S3 (11/08/2022), S4 (18/08/2022), S5 (25/08/2022), and S6 (01/09/2022). 241

Figure 6-4. Test site characterisation regarding the irrigation treatments (NI: non-irrigated, Etc 30%, Etc 60%) and altitude. 241

Figure 6-5. The normalised difference index (NDVI) based on Sentinel imagery mapped across the three assessment dates in 2021. S1 (05 days after veraison – DAV), S3 (19 DAV) and S5 (33 DAV).	242
Figure 6-6. Mean of the predawn leaf water potential of the irrigation treatment (NI: non-irrigation, Etc 60% and Etc 30%), regarding the vigour zones of the parcel (H: high vigour, L: low vigour) in 2021 (a) and 2022 (b).	243
Figure 6-7. Chlorophyll content in the skin (CHL SK – I. a), b), and c)), in the pulp (CHL PU – II. d), e) and f)), and the seed (CHL SE – III. G), h), i)) maps for Touriga Nacional cultivar across three post-veraison sampling dates (S3, S4 and S6) in 2021 to represent the prediction of the grapes that were measured using only spectral data. S3 (19 days after veraison (DAV), a), d) and g)), S4 (26 DAV, b), e) and h)) and S6 (40 DAV, c), f) and i)). NI: non-irrigated, 30% Etc: 30% of evapotranspiration and 60% Etc: 60%. ...	247
Figure 6-8. Anthocyanin content in the skin (ANT SK – I. a), b), and c)), in the pulp (ANT PU – II. d), e) and f)), and the seed (ANT SE – III. G), h), i)) maps for Touriga Nacional cultivar across three post-veraison sampling dates (S3, S4 and S6) in 2021 to represent the prediction of the grapes that were measured using only spectral data. S3 (19 days after veraison (DAV), a), d) and g)), S4 (26 DAV, b), e) and h)) and S6 (40 DAV, c), f) and i)). NI: non-irrigated, 30% Etc: 30% of evapotranspiration and 60% Etc: 60%. ...	248
Figure 6-9. Altitude a), normalised difference vegetation index (NDVI - b), solid soluble content (SSC - c), puncture force (PF – d), volume (vol – e), weight per volume (W/VOL – f), chlorophyll in the skin (CHL SK - g), chlorophyll in the pulp (CHL PU – h), chlorophyll in the seed (CHL SE – i), anthocyanin in the skin (ANT SK – j), anthocyanin in the pulp (ANT PU – k), anthocyanin in the seed (ANT SE – l) at 40 days after veraison. NI: non-irrigated, 30% Etc: 30% of evapotranspiration and 60% Etc: 60%.	250
Figure 6-10. Decision maps regarding different goals in the maturation. Model 1 stands for no weight in the soluble solids content (SSC), anthocyanin in the skin (ANT SK) and chlorophyll in the pulp (CHL PU). Model 2 attributes 60% to SSC, 30 % to ANT SK, and 10 % to CHL PU, and Model 3 stands for 30% SSC, 60 % to ANT SK and 10 for CHL PU.	251
Figure 7-1. Representation of an integrated methodology combining genomics, metabolomics, and systems biology approaches aimed to establish connections between laboratory experiments and field observations. Additionally, the figure demonstrates the integration of sensors for detecting molecular components and monitoring the physiological state of plants to feed the system biology.	273
Figure 7-2. Predawn Leaf Water Potential Variation in Grapevine Leaves under unshaded and shaded conditions. * Statistically significant ($p < 0.05$) according to t-test.	278
Figure 7-3. Mean spectra from leaves vines exposed to unshaded conditions and shaded conditions.	279
Figure 7-4. Results of the use of foliar spectral data combined with self-learning artificial intelligence (SL-AI) in modelling: a) chlorophyll a (mg/gFM), b) chlorophyll b (mg/gFM) and c) reactive oxygen species (ROS – ABS/gFM). N= number of samples considered. Coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE - %).....	281
Figure 7-5. Photorespiration cycle under shaded a) and unshaded b) conditions, showcasing the Photorespiration III with the objective reactions in the flux balance analyses of Peroxisomal Catalase and Peroxisomal Glycolate Oxidase. The diagram illustrates the reactions involved, offering a dynamic regulation of photorespiratory metabolism in response to environmental light conditions. The arrows' size indicates each pathway's intensity under the respective condition.	283
Figure 7-6. Panel a) shows the phenotype spaces for the three Flux Balance Analysis (FBA) models under different photorespiration conditions. For unshaded conditions, FBA models include FBA_P_CT_U (Peroxisomal Catalase Reaction as objective), FBA_P_96_U (Peroxisomal Glycolate Oxidase Reaction), and FBA_B3_U (Peroxisomal	

Catalase Reaction as objective and Peroxisomal Glycolate Oxidase Reaction as objective). Similarly, shaded conditions are represented by FBA_P_CT_S, FBA_P_96_S and FBA_B3_S, respectively. Panel b) demonstrates the phenotype spaces resulting from Monte Carlo (MC) simulations, illustrating the range of variations in chlorophyll and reactive oxygen species (ROS) levels assessed in laboratory experiments. MC simulations encompass MC_B3 Peroxisomal Catalase Reaction as objective and Peroxisomal Glycolate Oxidase Reaction as objective), MC_P_96 Peroxisomal Glycolate Oxidase Reaction as objective) and MC_P_CT (Peroxisomal Catalase Reaction as objective). In the case of MC_P_96 and MC_P_CT, 1 represents higher ROS (unshaded condition) and 10, lower ROS (shaded condition). MC_B3 did not result in a valid MC simulation. 285

Figure 8-1. Omicbots: integrating in situ, in vitro, and in silico approaches for precision agriculture..... 305

Figure A-1. Utilização do Metbots na medição do potencial hídrico foliar de base nas vinhas do Douro: a) obtenção dos espectros das folhas antes de serem retiradas e medidas na câmara de pressão (Scholander), antes do amanhecer, b) espectro apresentado pelo Metbots e c) O Metbots (indicado com a seta vermelha) acoplado num robô com outros sensores..... 350

Figure A-2. Relação entre dados observados e estimados com base em sensores hiperespectrais de proximidade e diferentes técnicas de modelação dos seguintes parâmetros biofísicos da vinha na região do Douro. a) PHB utilizando o algoritmo Self-learning artificial intelligence (SL-AI) num dataset com 847 observações recolhidas em 2014, 2015 e 2017; b) modelação com Partial Least Square Regression (PLS) do °Brix em Touriga Nacional (n=192) no ano de 2021 em três diferentes datas (S1: 29/07/2021, S2: 5/08/2021, S3: 12/08/2021, S4: 19/08/2021, S5: 26/08/2021, S6: 02/09/2021); c) modelação das clorofilas ab através do algoritmo Support Vector Machine (SVM) num dataset (n=105) com três cultivares (Touriga Nacional, Touriga Franca e Tinta Roriz) em quatro datas (S1: 03/08/2018, S2: 17/08/2018; S3: 31/08/2018; S4: 12/09/2018). R²: coeficiente de determinação, RMSE: Erro quadrático médio, MAPE: Erro médio absoluto previsto. 351

List of supplementary figures

Figure S 3-1. The study area's location identifies the Test Sites 1 and 2 in the Douro Wine Region, Northeast Portugal, and respective the sub-regions: Baixo Corgo, Cima Corgo and Douro Superior.....	76
Figure S 3-2. The overall accuracy of the predicted predawn leaf water potential (Ψ_{pd}) based on the three classes of Ψ_{pd} used by Pôças et al. (2020). "a", "b" and "c" are the results performed with Self-learning artificial intelligence (SL-AI), Artificial Neural Network with Principal Component Analysis (ANN-PCA) and Partial Least Square (PLS) respectively, according to the year (2014, 2015, 2017), cultivar (TB: Tinta Barroca, TF: Touriga Franca and TN: Touriga Nacional) and test site (TS 1- Quinta dos Aciprestes, TS 2 – Quinta do Ataíde).	77
Figure S 5.2-1. Gradient of tomatoes (a) in different stages of maturation used in the experiment and the tomography-like system used to obtain the spectral information of the tomato and the internal tissues (skin, pulp and seed). In b), the spectrometer (HR4000), the fibre probe, and the light source with the tomato are described; c) it demonstrates how the spectral information is obtained; d) shows the aliquots of the skin, pulp, and seeds considered for measurements.....	181
Figure S 5.2-2. Regression plots for the soluble solid content (SSC) in the original (a) and reconstructed (b) spectra considering the whole tomato. Circled observations represent the validation dataset.	181
Figure S 5.2-3. Regression plots for the chlorophyll (a.u.) in the reconstructed spectra considering the whole tomato (a) and the skin (b), pulp (c) and seed (d). Circled observations represent the validation dataset.	182
Figure S 5.2-4. Regression plots for the lycopene (a.u.) in the reconstructed spectra considering whole tomato (a) and the skin (b), pulp (c) and seed (d). Circled observations represent the validation dataset.	183
Figure S 5.2-5. Regression plots for the puncture force (N) in the original (a) and reconstructed (b) spectra considering the whole tomato. Circled observations represent the validation dataset.	184
Figure S 5.3-1. Climatic data from the veraison (14/08/2020) until the last sampling date (14/10/2020). Green dotted point represents the sampling dates.....	215
Figure S 5.3-2. Comparison of chlorophyll (a.u.) levels predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Loureiro grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a), skin (b), pulp (c) and seed (d). Reconstructed spectra for the total grape (e), skin (f), pulp (g) and seed (h).	216
Figure S 5.3-3. Comparison of soluble solids content predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Loureiro grape. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).	217
Figure S 5.3-4. Comparison of puncture force (N) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Loureiro grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).....	218
Figure S 5.3-5. Comparison of anthocyanin levels (a.u.) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a), skin (b), pulp (c) and seed (d). Reconstructed spectra for the total grape (e), skin (f), pulp (g) and seed (h).	219

Figure S 5.3-6. Comparison of chlorophyll levels (a.u.) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a), skin (b), pulp (c) and seed (d). Reconstructed spectra for the total grape (e), skin (f), pulp (g) and seed (h).	220
Figure S 5.3-7. Comparison of soluble solids content predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).	221
Figure S 5.3-8. Comparison of puncture force (N) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).	222
Figure S 6-1. Regression plots by applying self-learning artificial intelligence (SL-AI) in estimating: a) Soluble solid content (SSC - %), b) Puncture force (N), c) Volume (cm ³), d) Weight/volume (g/cm ³), e) Chlorophyll (a.u.) total in the grape, f) Chlorophyll (a.u.) in the skin, g) Chlorophyll (a.u.) in the pulp, h) Chlorophyll (a.u.) in the seed, i) Anthocyanin (a.u.) total in the grape, j) Anthocyanin (a.u.) in the skin, k) Anthocyanin (a.u.) in the pulp, and l) Anthocyanin (a.u.) in the seed.	256
Figure S 6-2. Regression plots by applying random forest (RF) in estimating: a) Soluble solid content (SSC - %), b) Puncture force (N), c) Volume (cm ³), d) Weight/volume (g/cm ³), e) Chlorophyll (a.u.) total in the grape, f) Chlorophyll (a.u.) in the skin, g) Chlorophyll (a.u.) in the pulp, h) Chlorophyll (a.u.) in the seed, i) Anthocyanin (a.u.) total in the grape, j) Anthocyanin (a.u.) in the skin, k) Anthocyanin (a.u.) in the pulp, and l) Anthocyanin (a.u.) in the seed.	257

List of graphical abstracts

Graphical abstract. 3-1. Grafical abstract case study I (Tosin et al., 2022).....	41
Graphical abstract. 4.2-1. . Graphical abstract case study III (Martins et al., Under review).....	103
Graphical abstract. 5.1-1. Graphical abstract of the patent (Martins et al., 2023).	141
Graphical abstract. 5.2-1. Graphical abstract case study IV (Tosin et al., 2024a).....	155
Graphical abstract. 5.3-1. Graphical abstract case study V (Tosin et al., 2023b).....	185
Graphical abstract. 7-1. Graphical abstract case study VII.	265

List of abbreviations

ANN	artificial neural networks
AI	artificial intelligence
ANN	artificial neural networks
ANN-PCA	artificial neural networks with principal component analysis
ANOVA	analysis of variance
ARlopt	anthocyanin reflectance index optimised
B	boron
BA	band area
B-MARS	bagged multivariate adaptive regression splines
BRNN	bayesian regularization-feed-forward neural network
BT	blind testing
BT	bagging trees
Ca	calcium
Chl	chlorophyll
CovM	covariance mode
CSWI	crop water stress index
Cu	copper
CV	cross-validation
DA	data augmentation
DAV	days after veraison
DL	deep learning
DoE	design of experiments
DOY	day of the year
DS	data scarcity
DVI	difference vegetation index
EVI	enhanced vegetation index
Fe	iron
GAM	generalised additive model
GBM	generalised boosted model
GC	gas chromatography
GI	green index
GNDVI	green normalised difference vegetation index
GPR	gaussian process regression
HO	hold-out samples
HPLC	high-performance liquid chromatography
HPLS	hierarchical partial least square
I	irrigation
ICP-OES	inductively coupled plasma optical emission spectroscopy
IoT	internet of things
ISPA	International Society of Precision Agriculture
K	potassium
KNN	k-nearest neighbour
LIBS	Laser-Induced breakdown spectroscopy
LIDAR	light detection and ranging
local-PLS	local partial least squares

LocPLS	local partial least squares
LOOCV	leave-one-out cross-validation
LV	latent variable
M5	m5 tree
MAE	mean absolute error (mpa)
MAPE	mean absolute predicted error (%)
MBD	maximum band depth
Mg	magnesium
MHPCA	multi-block hierarchical principal component analysis
MLR	multiple linear regression
Mn	manganese
MPa	megapascal
MPLS	modified partial least square regression
MSAVI	modified soil adjusted vegetation index
MSC-log	logarithm multiplicative scattering correction
MSE	mean standard error
N	nitrogen
Na	data non-available
NCS	total nitrogen, carbon, and sulfur
NDGI	normalised difference greenness vegetation index
NDGIopt	normalised difference greenness vegetation index optimised
NDRE	normalised difference red edge index
NDVI	normalised difference vegetation index
NI	no irrigation
NIPALS	nonlinear iterative partial least squares
NIR	near-infrared
NIST	National Institute of Standards and Technology
nm	nanometre
Nobs	dataset size
NPClopt	normalised pigment chlorophyll index optimised
NPK	nitrogen, phosphorous and potassium
NRI	normalised reflectance index
NRIgreen_LUT	normalised ratio index optimised for the lutein content in the green zone of the spectrum
NRIgreen_LWC	normalised ratio index optimised for the leaf water content in the green zone of the spectrum
OLR	ordinal logistic regression
OSAVI	optimal soil adjusted vegetation index
OSCAR	Optical Science Center for Applied Research at Delaware State University
P	phosphorus
PA	precision agriculture
PC	principal component
PCA	principal component analysis
PCR	principal component regression
PF	puncture force
PLS	partial least squares
PLS	partial least square
POM	point-of-measurement

PRESS	predicted error sum of squares
PRI	photochemical reflectance index
PRI_CI2opt	photochemical reflectance index with chlorophyll index optimised
PRInorm	photochemical reflectance index normalised
PU	pulp
PV	precision viticulture
p-value	probability of null hypothesis test
PWI	plant water index
R	correlation coefficient
R ²	coefficient of determination
RDVI	renormalised difference vegetation index
RF	random forest
RGI	relative greenness index
RGRI	red green ratio index
RMSC	robust multiplicative scattering correction
RMSE	root mean square error (mpa)
ROI	regions of interest
RWD	real-world data
RWD+SSD	model using rwd and ssd during cv and bt
RWD→SSD	model using rwd during cv, ssd in bt
S	sulfur
SAVI	soil-adjusted vegetation index resistant index
SE	standard error
SE	standard error
SE	seeds
SIPi	structure intensive pigment index
SK	skin
SLAI	self-learning artificial intelligence
SL-AI	self-learning artificial intelligence
SLR	single linear regression model
SPVlopt1	spectral polygon vegetation index optimised 1
SPVlopt2	spectral polygon vegetation index optimised 2
SR	simple ratio index
SSC	soluble solids content
SSD	synthetic spectroscopy data
SSD→RWD	model using ssd during cv, rwd in bt
SVC	support vector regression
SVD	singular value decomposition
SWIR	short-wave infrared
T	spectral information
TB	Tinta Barroca
TF	Touriga Franca
TN	Touriga Nacional
TO	total grape
TRL	technology readiness level
TS	test site
U	leaf nutrient composition
UV-Vis-NIR	ultraviolet-visible-near-infrared

VARlopt	visible atmospherically resistant index optimised
VH-GPR	variational heteroscedastic gaussian process regression
VI	spectral vegetation index
Vis-NIR	visible near-infrared
VIVC	Vitis international variety catalogue
WABI	water balance index
Wat	water
X	spectra matrix
Xan	xanthophyll
XGBoost	extreme gradient boosting
Y	leaf composition matrix
Zn	zinc
Ψ	water potential
Ψ_{pd}	predawn leaf water potential (Mpa)

SECTION I

--- CHAPTER 1

INTRODUCTION

1. Introduction

PA can be defined by the International Society of Precision Agriculture (ISPA) as a management strategy that utilises temporal, spatial, and individual data, integrated with other relevant information, to enhance decision-making processes and optimise resource utilisation, productivity, and sustainability in agricultural production.

As the foundation of human civilisation, agriculture has been continuously shaped by technological advancements and responses to ecological and social demands (Fuentes et al., 2023). Today's agricultural sector confronts challenges from a rapidly growing global population and escalating environmental pressures. These challenges demand heightened productivity, quality, and sustainability during the elaborate interaction of factors such as climate change and the extensive use of agrochemicals (e.g. fertilisers, pesticides) (Lo Presti et al., 2023). In PA, the persistent search for enhanced methodologies drives the core of this thesis (Figure 1-1), which provides the connection of hyperspectral data analysis and computational algorithms to unravel agronomic processes, with a distinct emphasis on investigating the physiological parameters of crops through a systems biology lens.

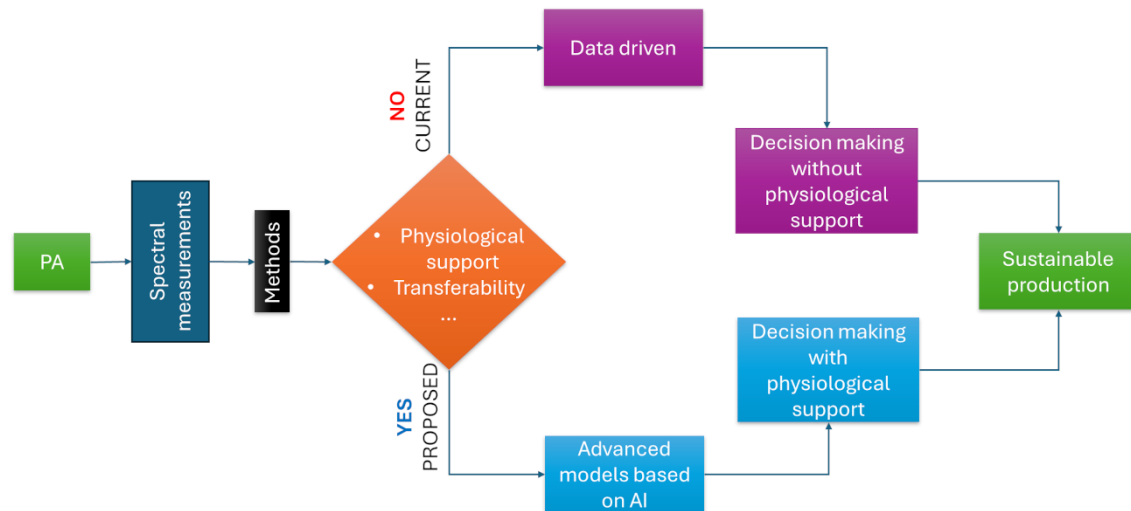


Figure 1-1. Scheme of the current precision agriculture (PA) what is proposed in this thesis.

PA predominantly trusts data-driven approaches relying on vegetation indices (e.g., normalised difference vegetation index - NDVI) and machine learning algorithms (e.g., support vector machine - SVM) and exhibits challenges concerning the transferability of the model across years and locations (Tilly et al., 2015). Thus, a gap emerges concerning providing physiological explanations for observed phenotypes. The core of this limitation lies in the models' incapacity to account for the physiological mechanism inherent in leaves and fruits, particularly when subjected to diverse edaphoclimatic

conditions. Moreover, the transferability of existing models is evident when confronted with edaphoclimatic variations (e.g., temperature, drought) (Wagenaar et al., 2021). The physiological properties of leaves and fruits, responsive to changes in soil and climate, remain a blind spot in conventional PA methodologies.

Among the various methodologies employed to gather data for informed decision-making in PA, satellite-based and drone-based approaches represent two prominent strategies, each offering distinct advantages and limitations. Satellite-based PA, exemplified by platforms like Sentinel-2, provides a macroscopic view of agricultural landscapes, offering extensive coverage over large areas (Sassu et al., 2021). This approach proves advantageous for broad-scale assessments and monitoring, facilitating insights into regional trends and spatial variability (Cogato et al., 2019). However, satellite imagery may lack the fine-grained resolution required for plant-specific analyses, limiting its utility in detecting variations at the canopy level. Furthermore, temporal and spectral constraints inherent in satellite data acquisition may pose challenges for real-time monitoring and analysis, particularly in dynamic agricultural environments (Helman et al., 2018).

Conversely, drone-based PA offers the flexibility to capture detailed canopy spectra and high-resolution imagery, enabling precise assessments at the field level. Drones equipped with multispectral or hyperspectral cameras can provide fine-grained perceptions into crop health, nutrient status, and stress detection (Romero et al., 2018; Matese & Di Gennaro, 2018). Despite these advantages, factors such as flight altitude, camera spectral resolution, and environmental conditions may influence data quality and consistency (Sassu et al., 2021). Additionally, operational constraints such as battery life and flight regulations impose practical limitations on drone-based data collection (Sassu et al., 2021). Thus, while satellite-based PA offers broad-scale coverage and temporal consistency, drone-based approaches provide detailed, field-level insights, complementing the other in the search for PA.

Proximal hyperspectral sensors emerge as effective tools, providing detailed perceptions of target areas. This capability facilitates modelling approaches, incorporating chemometrics (e.g., Mishra et al., 2023) and machine learning (e.g., Fernandes et al., 2015) for pigment and chemical compound assessments (e.g., Yu et al., 2021). Operating in both active and passive modes, these sensors, reliant on external radiation, introduce considerations related to light conditions during measurements. Point-of-Measurement (POM) devices, equipped with hyperspectral capabilities, can offer perceptions of the physiological complexities of leaves and fruits (Martins et al., 2022).

This emphasis on internal properties departs from conventional PA practices that rely heavily on surface-level data.

In PA, methodological challenges such as model transferability and limitations in accounting for physiological mechanisms pose barriers to accurate and reliable data interpretation. Addressing these challenges is crucial for enhancing the robustness and applicability of PA methodologies. As an interdisciplinary approach, systems biology takes the lead in unravelling the complexities of biological systems. It examines components and interactions as an integrated whole, recognising that biological events arise from dynamic relationships among various molecular, cellular, and environmental factors (Trewavas, 2006). Integrating systems biology with spectral data represents a transformative approach in PA, offering perceptions into agronomic processes and supporting more effective decision-making. Hyperspectral data provides the potential to estimate biophysical parameters like chlorophyll (Agati et al., 2008), carotenes (Van Wittenberghe et al., 2024), anthocyanin (Liu et al., 2019b), flavonoids (Falcioni et al., 2023), and reactive oxygen species (ROS) (Herman et al., 2018), along with physiological parameters like leaf water potential (Santesteban et al., 2019), lycopene, and β -carotene (Tilahun et al., 2018), essential to feed systems biology models, and represents an alternative to conventional wet-lab procedures, which often entail high costs, extensive time investment, and limited adaptability for analysing small tissue samples (Power et al., 2019). This combination can provide high-resolution spectral information and incorporates systems-level analyses, offering a holistic understanding of how crops engage with their environment (Pazhamala et al., 2021). Such a broad approach enriches the perception of physiological processes, acting as a transformative bridge that connects traditional agricultural practices with cutting-edge scientific methodologies. Integrating systems biology and spectral data in PA signifies a paradigm shift, facilitating a deeper understanding of plant responses to environmental conditions. This integration serves to decipher the language of plants encoded in spectral signatures, ultimately advancing the precision and sustainability of modern agriculture. By incorporating genomics, transcriptomics, and metabolomics data (Soubeyrand et al., 2018; Bramley et al., 2017), this transformative approach transcends traditional agricultural norms, can offer an understanding of crop physiology and environmental interactions. By leveraging advanced analytical techniques, such as omics technologies, it is possible to investigate into the molecular mechanisms governing plant responses to environmental inducements, thereby optimising agricultural practices for enhanced productivity and sustainability.

In the face of contemporary agricultural challenges, such as the need to improve production sustainability (European Commission, 2021) while contending with the impacts of climate change (Lozano-Elena et al., 2022), integrating systems biology with spectral data provides a powerful toolset. This approach enhances the ability to mitigate the adverse effects of climate variability and enables the development of more resilient and adaptive agricultural systems. By elucidating the relationships between genetic factors, biochemical processes, and environmental signs, it is possible to unlock new avenues for optimising crop performance, resource utilisation, and environmental stewardship in the pursuit of sustainable agriculture.

Each objective of this thesis addresses specific gaps in the current literature while contributing to the advancement of PA. Firstly, exploring physiological explanations concerning the water status of grapevines addresses a gap in understanding the mechanisms governing plant hydration, particularly in dynamic environmental conditions. Secondly, introducing cost-effective spectroscopy techniques for nutrient evaluation in plant tissue fills a void in the accessibility of advanced analytical methods, democratising the process of nutrient assessment in agriculture. Thirdly, developing non-invasive techniques for assessing internal fruit ripeness levels enhances the capacity to monitor crop maturation accurately, offering perceptions of fruit quality and harvest optimisation. Finally, integrating systems biology with spectral data represents a paradigm shift in PA, providing a holistic understanding of plant-environment interactions and offering novel insights into crop physiology. Therefore, each objective addresses existing gaps and propels the field forward by introducing innovative methodologies and perspectives.

This thesis is organised into three sections and nine chapters, as Figure 1-2 illustrates. The first section comprises two chapters. After laying the foundational concepts in this introduction (Chapter 1), it proceeds to explore deeper into the PA concepts. Chapter 2 presents the conceptual framework, exploring aspects such as PA methodologies, optical sensors, data collection, modelling techniques, fruit maturation dynamic, system biology integration into PA, and the essential challenges. As the transition goes to the subsequent chapters, this sets the stage for an in-depth examination of the leaves of grapes and apples and an analysis of fruit maturation, particularly in tomatoes and grapes. It also explores innovative non-destructive methods for characterising grape tissues and integrating systems biology, bridging theoretical understanding and practical applications in PA.

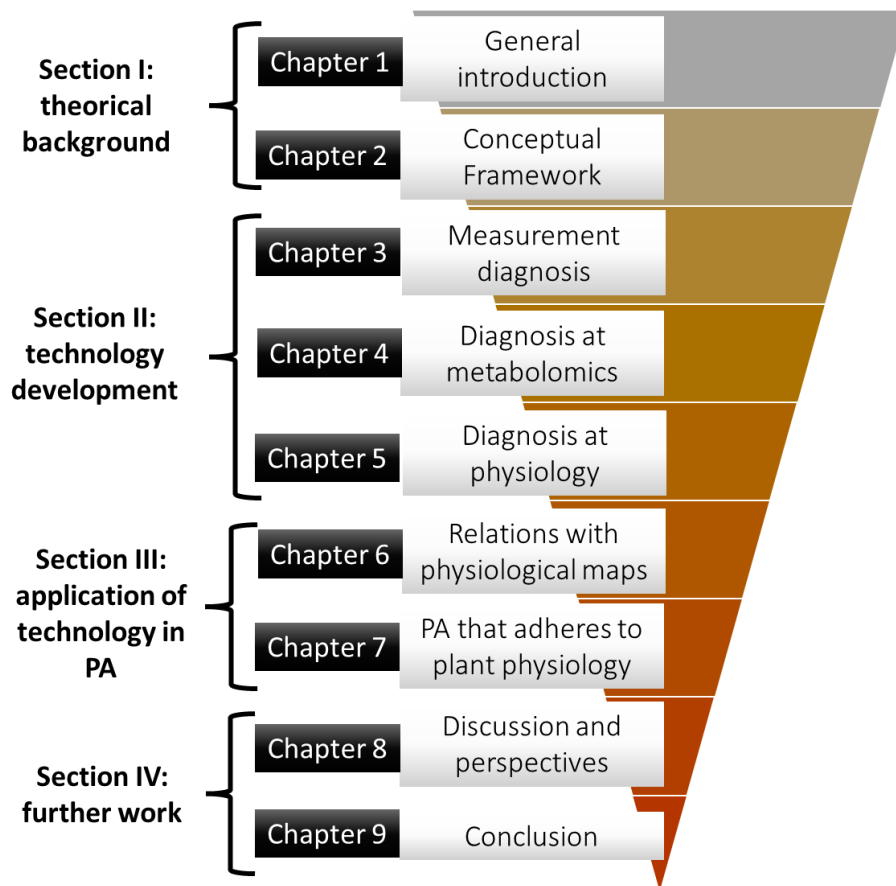


Figure 1-2. Structure of the thesis and organisation of the content to be presented.

The second section of this thesis, as outlined in Table 1-1, comprises three chapters, one patent, and seven case studies. Chapter 3 addresses the question: “What methods are used to assess ecophysiological plant parameters in the field?” This question is confronted by investigating predawn leaf water potential and its impact on post-veraison foliar pigments through hyperspectral data, machine learning, and artificial intelligence algorithms.

In Chapter 4, two case studies aim to answer the question: “What methodologies are employed to conduct nutritional characterisation of various plant organs?” Case study II answers this question and utilises laser-induced breakdown spectroscopy (LIBS) for nutrient identification in small biological tissues of grapes. In contrast, the subsequent case study (III) employs quantitative Vis-NIR spectroscopy for nutrient identification in apple and grape leaves through data augmentation.

Chapter 5 explores the question: “What methods are utilised to evaluate the progression of maturation in fruits?” The answer to this question is addressed through the development of a patent that utilises hyperspectral data to bi-directionally reconstruct the internal tissues of fruits, creating a tomography-like approach. This chapter includes two

case studies: Case Study IV analyses tissue changes and biophysical variables and pigments changes during tomato maturation, while Case Study V examines the same biophysical characteristics and pigments changes during grape maturation, considering two grape cultivars.

The thesis then moves to Chapter 6, which addresses the question: “What methods are employed to conduct tomography-like assessments in the field for assessing maturation?” This question is answered by presenting a novel approach termed metabolic maps, which integrate various physicochemical parameters (e.g., pigments, sugar content) to define site-specific management strategies through the tomography-like method presented.

Finally, Chapter 7 explores the question: “What is the relationship between Precision Agriculture and systems biology?” This question directs to a case study by employing system biology techniques coupled with laboratory and hyperspectral measurements in characterising the phenotype of the use of shade curtains, by simulating the virtual phenotype through ROS mechanisms.

Table 1-1. Organisation of scientific questions, corresponding responses, and the presented case studies.

Section II	Chapter 3	What methods are used to assess ecophysiological plant parameters in the field?	Study on predawn leaf water potential and its impact on post-veraison pigments.
			This question is answered with case study I (paper I) , which aims to investigate the prediction of predawn leaf water potential and reveal the relationship between grapevine hydric status and absorbance bands related to chlorophyll, xanthophylls, and water, offering perceptions into identifying and stress conditions.
	Chapter 4	What methodologies are employed to conduct nutritional characterisation of various plant organs?	LIBS for nutrient identification in small biological tissues.
			This question is answered with case study II (paper II) , which focuses on the elemental characterisation of grapevine tissues using laser-induced breakdown spectroscopy (LIBS) over traditional chemical methods for element quantification in grapes, which are expensive and impractical for widespread commercial use. The primary objective is to discern potential variations in element concentrations among different grape tissues, exploring their impact on the metabolic constitution and quality parameters. The study compares mineral contents across three post-veraison times for white (Loureiro) and red (Vinhão) grape cultivars.
			Quantitative nutrient identification in apple and grape leaves through Vis-NIR.
			This question is answered by case study III (paper III) : The study introduces a novel approach called data augmentation in spectroscopy. This research addresses data scarcity in physiology-based precision agriculture, specifically in quantifying nutrient uptake in plant leaves using Vis-NIR spectroscopy. This study features the importance of experimental design in field experiments to ensure a representative knowledge base for effective data augmentation in spectroscopy, ultimately contributing to precision fertilisation in agriculture.
	Chapter 5	What methods are utilised to evaluate the progression of maturation in fruits?	Non-destructive method in assessing internal fruit tissue.
			This question originated the creation of a patent proposing a technique and apparatus for the non-invasive tomographic analysis of a sample containing various distinct tissues. Specifically, the invention involves a computer-based approach for characterising the metabolites in a fruit sample without invasive measures. The sample in question consists of previously described tissues, and the proposed method utilises a multidimensional latent structure model of these distinct tissues for non-invasive tomographic metabolite characterisation.
			Tomato case study: identifying tissue changes during maturation.
			The concept proof of the patent goes first to case study iv (paper iv) : monitoring the maturation of tomatoes through the tomography-like method. Quality parameters such as soluble solid content, chlorophyll, lycopene, and puncture force were successfully assessed. The study showcases the proposed approach's reliability, early indication, and operational simplicity, emphasising its

			potential for precision fruit maturation assessment with implications for various agronomic and food chain applications. Grape case study: comparing cultivars, tracking quality changes during maturation. The concept proof of the tomograph-like extends to the case study V (paper v): this paper uses the tomography-like method for precision maturation assessment of grape tissues. The study includes two grape cultivars, Loureiro (white) and Vinhão (red), measured at various dates after veraison. Biophysical parameters such as soluble solids content, puncture force, chlorophyll, and anthocyanin are predicted. The method presents its ability to deal with different grape cultivars and explain the various dynamics of maturation compared to different grape cultivars.
Section II	Chapter 6	What methods are employed to conduct tomography-like assessments in the field for assessing maturation?	Douro wine region case study: metabolic maps for tissue evolution and maturation. This representation of metabolic maps of the internal tissues is demonstrated in case study VI (paper VI): This study explores grape maturation dynamics in the Douro Wine Region using a tomography-like method based on Vis-NIR data. Metabolic maps generated from non-destructive measurements reveal variations in physicochemical parameters across different irrigation levels, offering insights into grapevine responses to environmental factors and enhancing precision viticulture practices.
	Chapter 7	What is the relationship between Precision Agriculture and systems biology?	Combining spectral sensing and systems biology for leveraging precision viticulture: unravelling photosynthetic pathways in the Douro wine region - assessing the impact of shade treatment on grape physiology Case study VII (paper VII): Investigating the impact of shade curtains on grape physiology, this study employs a multidisciplinary approach combining biochemical analyses, hyperspectral data, and system biology techniques. Results show reduced water stress and increased chlorophyll concentrations in shaded vines, highlighting the importance of integrating spectral sensing and systems biology for understanding grapevine responses to environmental stressors and optimising viticultural practices.

Finally, Section III includes Chapter 8, which discusses the results and future works, and Chapter 9, which presents the main conclusions of the thesis.

The research conducted in this thesis aligns with and contributes to specific goals outlined in global sustainability initiatives, such as the 2030 agenda for sustainable development (UN DESA, 2016), Farm to Fork (European Commission, 2020), and the Green Deal (European Commission, 2021). By leveraging advanced analytical techniques and innovative methodologies, this research aims to enhance resource efficiency, promote sustainable production practices, and minimise environmental impact (European Commission, 2021). Specifically, developing and validating non-invasive techniques for assessing crop maturation contribute to Goal 12 of the 2030 agenda, aiming to ensure sustainable consumption and production patterns (UN DESA, 2016). Furthermore, by elucidating the relationships between genetic factors, biochemical processes, and environmental signals, this research supports the objectives of the Farm to Fork Strategy, which seeks to improve food systems and promote sustainability (European Commission, 2020). Additionally, integrating systems biology with spectral data facilitates a deeper understanding of plant-environment interactions, aligning with the goals of the Green Deal to reduce environmental impact and foster competitiveness in agriculture (European Commission, 2021). Ultimately, this research has the potential to contribute to the transition towards a circular and environmentally friendly agrifood system, as outlined in these sustainability initiatives.

CHAPTER 2

CONCEPTUAL FRAMEWORK

2. Conceptual framework

This conceptual framework serves as a roadmap for the practical implementation and advancement of PA techniques, particularly in monitoring fruit maturation. By drawing from the case studies outlined in this thesis, this framework explains the integration of PA methodologies into agricultural practices through tangible examples. These case studies encompass a range of applications, including the influence of predawn leaf water potential on post-veraison pigments, the utilisation of LIBS for nutrient identification in small biological tissues, and quantitative nutrient identification in apple and grape leaves using Vis-NIR spectroscopy. Additionally, patented technology for non-destructive assessment of internal fruit tissue further demonstrates the practical application of PA techniques in fruit maturation assessment.

Moreover, this conceptual framework explores data analysis techniques employed in PA, emphasising the significance of data augmentation, machine learning, and artificial intelligence in interpreting datasets obtained from spectral optical sensors. These analytical tools facilitate informed decision-making processes in crop management, enhancing agricultural productivity and efficiency.

Aligned with the forward-thinking principles of PA, a subsection of this framework explores future directions and emerging technologies within the field. It examines potential advancements in sensing technologies and data analytics, highlighting their potential to improve agricultural practices and lead innovation in PA.

Furthermore, environmental sustainability emerges as a paramount consideration within this conceptual framework. A dedicated subsection addresses sustainability challenges and solutions within PA, including strategies for resource optimisation, conservation practices, climate resilience, and ecosystem management. By supporting environmentally sustainable approaches, this framework aims to mitigate the ecological footprint of agricultural activities and promote the long-term viability of food production systems.

2.1. Precision agriculture

PA has transformed with the integration of artificial intelligence (AI) and machine learning (ML) algorithms, leading to advancements in agricultural data analysis and decision-making processes (Duncan et al., 2021). These technological advances align closely with the overarching objectives of PA, which include cost reduction, economic optimisation, and minimising environmental impact (Pandey & Pandey, 2023). The targeted and prompt approaches to information and resource management within PA

resonate with the core aims of this thesis, which seeks to propose advanced methodologies for diagnosing agronomic processes in PA.

Regarding spatial-temporal representation, current PA practices often rely on maps based on vegetation indices to discern various physicochemical parameters, canopy hydric status (Fernández-Novales et al., 2021), nutrient (Sharifi, 2020), health (Yang, 2020), SSC (Fernandez-Novales et al., 2019b) and yield potential (Sun et al., 2017a). However, these maps often lack a physiological explanation, which limits their utility in providing deeper perceptions into crop physiology.

Crop physiology is essential for advancing PA methodologies, particularly in explaining the metabolic contents of internal tissues and their implications for crop metabolism. The development of metabolic maps has emerged as a tool for spatial and temporal visualisation of biochemical parameters within biological tissues, thereby offering physiological perceptions into crop development. Recent studies by Yu et al. (2021) and Shmulevitz et al. (2024) have demonstrated progress in this domain by employing metabolic mapping techniques to delineate molecular-level spatial variability in grape berry maturation and vineyard vigour, leveraging sensors and gene expression analysis. Despite these advancements, maps describing variations in inner tissue physiology remain relatively vague, underscoring the necessity to integrate metabolic mapping into existing research methodologies. Such integration holds immense potential for enhancing the understanding of crop physiology and enabling informed decision-making in management practices, mainly through site-specific management (SSM) tailored to specific production goals.

2.1.1. Precision agriculture challenges

While PA holds promise for optimising agricultural practices, it copes with challenges hindering its effectiveness in accurately diagnosing agronomic processes (Lu et al., 2020). Among these challenges is the limited ability of existing data-driven approaches to provide detailed physiological explanations for observed phenotypes, particularly in dynamic agricultural environments characterised by diverse edaphoclimatic conditions (Bwambale et al., 2023; Martins et al., 2022). These limitations emphasise the need for innovative methodologies that can overcome the constraints of current techniques and offer precise diagnostic capabilities for enhancing the robustness and applicability of PA practices.

PA heavily relies on data-driven approaches, particularly emphasising data analysis from optical sensors (Kok et al., 2021). While these methods excel in modelling biophysical crop indicators like sugar content and water status, they often fail to provide detailed

physiological explanations for observed phenotypes (Lu et al., 2020). This deficiency arises from the limitations of data-driven techniques in accurately capturing the complex physiological phenotypes in agricultural fields. Environmental conditions (e.g., soil type, drought, high temperature), agronomic practices (e.g., irrigation, fertilisation), and the diverse geographical sources of data further complicate the efficacy and transferability of these methods, imposing constraints on their applicability across different regions (Bwambale et al., 2023; Martins et al., 2022).

One of the primary limitations of current data-driven approaches in PA is their lack of physiological explanation. These approaches are often based on conditions prevailing during assessment, varying across years, locations, and cultivars. Consequently, the resulting models may not adequately account for the underlying physiological mechanisms driving crop responses to environmental inducements (Tantalaki et al., 2019). This limitation challenges the robustness and generalisability of data-driven models, hindering their effectiveness in guiding PA practices across diverse contexts. Considering these challenges, it is imperative to acknowledge the limitations of data-driven approaches in PA and understand their implications for complex agri-food systems. This thesis seeks to address this gap by emphasising the importance of physiological adherence in PA and proposing innovative methodologies to enhance the precision and applicability of agricultural technologies across diverse contexts.

2.1.2. Spectral optical sensors

In PA, the electromagnetic spectrum has become a fundamental tool for investigating crop biophysical indicators using diverse optical sensors (Fuentes et al., 2023; Yandun Narvaez et al., 2017). These sensors, classified primarily based on their number of bands and spectral range, include multispectral (e.g., satellites) and hyperspectral sensors (e.g., spectroradiometers) commonly utilised in PA. Hyperspectral sensors, with their higher number of bands resulting from shorter bandwidths, offer superior spectral resolution compared to their multispectral counterparts (Mariotto et al., 2013). Figure 2-1 underscores a pivotal distinction: the broader scale of data captured by satellite imagery contrasts with the finer precision offered by proximal sensors. This difference underscores the trade-off between scale and precision in agricultural data collection. While multispectral data from satellites yield valuable insights, the narrower bandwidth of hyperspectral sensors contributes to heightened precision, reflecting the capabilities of each sensor type in capturing crop biophysical indicators.

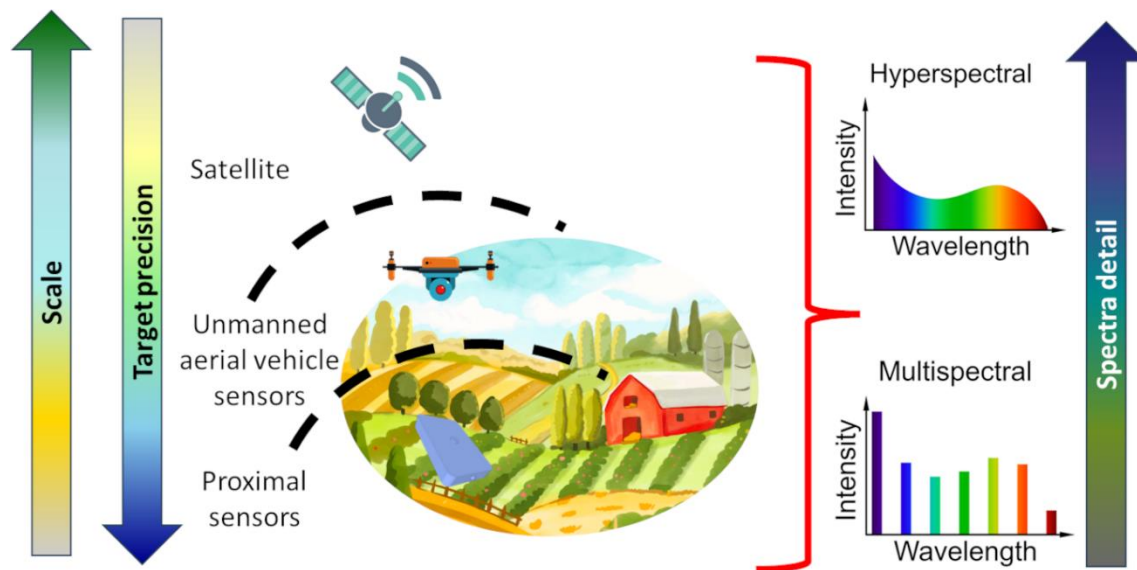


Figure 2-1. Illustration showcasing various sensors highlighting diverse scales, targeted precision levels, and spectral details.

Despite the widespread adoption of optical sensors in PA, their effectiveness is affected by lighting conditions, with overcast days hindering satellite imagery acquisition and necessitating computational processes to address these challenges. Similar issues arise with cameras mounted on drones and proximity sensors. The mitigation of lighting constraints is overcome through proximity sensors emitting artificial light during measurements, which enhance precision and enable the exploration of inner vegetation tissues.

Optical systems have shown promise in estimating biophysical parameters of fruits, particularly in Vis-NIR spectrum associated with pigment concentrations in leaves and fruits, as well as maturation stages (dos Santos Costa et al., 2019).

Traditional methods for estimating biophysical parameters in crops are often labour-intensive and time-consuming, such as the pressure chamber method for estimating leaf water potential and refractometers for sugar accumulation estimation (Pôças et al., 2017; Matthews et al., 2009). In contrast, optical sensors offer a non-destructive and expedited alternative for such estimations (dos Santos Costa et al., 2019; Pôças et al., 2017).

The near-infrared spectrum has been extensively studied for fruit maturation assessment, with handheld devices commonly used. Handheld spectroradiometers, coupled with data mining techniques, have demonstrated potential in estimating various parameters related to grape maturation (dos Santos Costa et al., 2019). Additionally, spectrograph cameras have been utilised to obtain hyperspectral data from grape seeds and assess phenolic compounds during maturation (Rodríguez-Pulido et al., 2013).

Chlorophyll fluorescence is a proxy of phenolic contents in leaves and anthocyanins in grape berries (Barthod et al., 2007; Schreiber et al., 1995). However, distinguishing signals from different morphological parts of grape berries is crucial due to variations in pigment concentrations and polyphenols across tissues (Sparrow et al., 2015; Jordao et al., 2012).

Devices like the Multiplex® offer multi-parametric fluorescence sensing capabilities, facilitating the estimation of chlorophyll and flavanol content in vines (Agati et al., 2013; Baluja et al., 2012b; Ben Ghazlen et al., 2010). Similarly, the Dualex FLAV™ and Dualex ANTH™ show promise to predict phenolic and anthocyanins in individual grapes and estimate nitrogen content in vines (Cerovic et al., 2015; Cerovic et al., 2008).

Various optical methods based on Vis-NIR have enabled the development of robotic techniques for grape bunch identification and yield prediction (Pérez-Zavala et al., 2018). Additionally, sensors capable of obtaining spectral signatures of inner tissues have opened avenues for estimating biochemical compounds in these tissues (Martins, 2019). The application of optical sensors extends beyond agriculture into various scientific domains such as medicine (Wencel et al., 2018), food science (Sadani et al., 2019), and agriculture (Mishra et al., 2023). The integration of AI and spectral data has led to the development of novel optical devices capable of spectral-based diagnosis, with promising applications in food detection (Liu et al., 2018), cancer detection (Talari et al., 2019), and citrus phenotyping (Ampatzidis & Partel, 2019).

While current sensors enable the characterisation of crop biophysical and biochemical indicators, few integrate bioinformatics resources with systems biology for cause-effect analysis of cultural physiology. The research team of this thesis has patented a sensor capturing spectral information across various spectra, coupled with AI for spectral analysis, allowing the detection and quantification of compounds in biological solutions and systems (Martins, 2019; Martins, 2018; Sousa et al., 2014). Preliminary results demonstrate the potential of these tools in quantifying SSC (Santos-Campos et al., 2023) and water potential (David et al., 2023) and mapping metabolites related to key physiological processes in grapevine physiology.

In PA, non-destructive optical proximity sensors are commonly employed to infer the biophysical properties of different plant structures through spectrum analysis (Martins et al., 2022). However, estimating the biochemical properties of a plant structure based on a single spectrum without considering the interference of various substructures has limitations. This work proposes an innovative approach that integrates the spectral properties of each constituent of a given plant structure (e.g., fruit) to obtain more

accurate spectra and determine the desired biophysical/biochemical properties effectively.

2.2. Data collection and modelling in precision agriculture

Precision agriculture relies on effective data collection and modelling to drive informed agronomic decision-making. Spatio-temporal models, essential in PA, provides perceptions into crop variability across space and time, helping the identification of the changeability in crop yield and other agricultural parameters, enabling targeted interventions tailored to specific areas and times, allocating resources more efficiently, reducing waste and environmental impact serving as decision support systems, guiding farmers on when and where to implement actions to maximise yield and minimise costs facilitating continuous monitoring and evaluation of crop performance and management practices, allowing for adjustments as needed (Norberto et al., 2023). Also, spectroscopy, especially in the Vis-NIR range, facilitates non-destructive and high-throughput data acquisition, supporting various agricultural applications (Martins et al., 2022). However, ensuring the reliability of acquired data demands the development of robust models and validation methods.

Traditional analytical methods like chemical assays, high-performance liquid chromatography (HPLC), and inductively coupled plasma optical emission spectroscopy (ICP-OES) are often impractical for analysing small biological tissue samples and present challenges for large-scale and frequent sampling across diverse plant organs (Lu et al., 2022). Despite their accuracy, these methods are time-consuming and costly, limiting their widespread adoption in PA. Therefore, there is a pressing need for cost-effective and efficient alternatives to generate augmented datasets for training models, thereby enhancing spectral analysis techniques' robustness and generalisation capabilities.

Data augmentation, commonly used in machine learning and data analysis, involves artificially increasing the diversity and size of a dataset by creating synthetic data samples (Tan et al., 2024). By introducing new samples with different characteristics, data augmentation mitigates overfitting and improves model robustness (Barroso et al., 2022). Although commonly used in computer vision, data augmentation can also enhance non-imagery techniques in PA, such as spectroscopy, for agricultural applications like nutrient quantification.

A promising approach for data augmentation in PA is the covariance mode (CovM), which maximises the extraction of unique constituent information while minimising interference (Martins et al., 2022; Barroso et al., 2022). CovM represents groups of

samples with similar levels of target information, allowing for the isolation of spectral variance related to target compounds. For instance, CovM can be utilised for data augmentation in spectroscopy, generating new samples that cover spectral gradients and provide covariance information between spectral features and target biological tissues like leaves.

Modelling in PA encompasses both parametric and non-parametric approaches, with ML techniques adept at handling complex data relationships (Goussev et al., 2019). Examples of parametric methods include support vector machines (SVM) for estimating soybean leaf area foliar (Yuan et al., 2017), random forest (RF) for predicting predawn leaf water potential (Loggenberg et al., 2018), and artificial neural networks (ANN) for estimating SSC in grapes (Poblete-Echeverría et al., 2020). Furthermore, ML algorithms such as k-nearest neighbours (K-NN) have been employed for citrus tristeza virus detection in mandarin oranges (Hazarika et al., 2020), and decision tree algorithms for predicting crop yield (Gonzalez-Sanchez et al., 2014).

Advancements in machine learning algorithms, such as the SL-AI algorithm based on CovM principles, offer opportunities to address dependencies and relationships between variables in data, ensuring model robustness and enhancing predictive accuracy in PA (Martins, 2018). SL-AI has successfully quantified sugar and acids (Martins et al., 2022), demonstrating its potential to overcome the limitations of traditional methods and pave the way for more efficient and reliable data collection and modelling in precision agriculture.

2.3. Non-destructive monitoring of internal and external fruit maturation tissues

This section delves into the dynamics of grape and tomato maturation, offering insights into a non-destructive method for assessing the internal tissues of fruits. It emphasises the significance of understanding fruit maturation in PA, explicitly focusing on grapes and tomatoes due to their economic importance and widespread cultivation, particularly in Portugal (Mourao & Domingues Martinho, 2017; Schwarz et al., 2014). The section explores the multi-layered nature of fruit maturation, encompassing biochemical and biophysical transformations, environmental influences, and nutrient interactions. It highlights its role in optimising agricultural practices and evaluating fruit quality in PA.

Spectral optical sensors represent a foundation of PA, offering versatile means of monitoring multiple parameters in crop fields and facilitating high-throughput data collection (Grosskinsky et al., 2018). Recent innovations in spectral sensing technology, such as LIBS, have opened up new possibilities for on-field elemental composition characterisation with minimal sample preparation and reduced operational costs.

(Andrews et al., 2023a; Senesi et al., 2019). Similarly, hyperspectral Vis-NIR proximal devices offer capabilities in characterising fruit maturation and estimating key biochemical and biophysical parameters, making them invaluable tools for advancing PA methodologies (Oduntan et al., 2022; Ciaccheri et al., 2018).

Spectral optical sensors offer a versatile means of monitoring multiple parameters in crop fields, facilitating data collection. In trait phenotyping, which involves systematically assessing various plant characteristics, spectral optical sensors can provide precise and high-throughput measurements directly within field conditions. This multidimensional evaluation encompasses morphological (Oduntan et al., 2022), phenological (Sassu et al., 2021), physiological (Grosskinsky et al., 2018), biochemical (Ecartot et al., 2013), and genetic features (Kumar et al., 2015), enabling a deeper understanding of plant behaviour and informing agricultural practices and research decision-making.

Hyperspectral Vis-NIR proximal devices represent another advancement in spectral optical sensing. These devices can model various crop biophysical factors, providing detailed information about target objectives. Vis-NIR spectroscopy has demonstrated capabilities in characterising fruit maturation, estimating key parameters such as lycopene (e.g., Ciaccheri et al., 2018), β -carotene (e.g., Tilahun et al., 2018), water potential (e.g., Pôças et al., 2020), acids (e.g., Huang et al., 2018a), SSC (e.g., Mishra et al., 2023), pH levels (e.g., Giovenzana et al., 2015), colour attributes (e.g., Clément et al., 2015), vitamin C content (e.g., Azadshahraki et al., 2018), firmness (e.g., Najjar & Abu-Khalaf, 2021), phenols (e.g., Rouxinol et al., 2022), malic acid (e.g., Torres et al., 2015), citric acid (e.g., Ding et al., 2016).

Moreover, Vis-NIR spectroscopy is a powerful tool for quantifying plant tissue nutrients by measuring leaves' spectral reflectance and employing chemometrics/artificial intelligence (Martins et al., 2022). Vis-NIR spectroscopy enables the quantification of essential macronutrients like nitrogen (N) (e.g., Ulissi et al., 2011), phosphorus (P) (e.g., Acosta et al., 2023), potassium (K) (e.g., Prananto et al., 2020), calcium (Ca) (e.g., Santoso et al., 2018), and magnesium (Mg) (e.g., Oliveira & Santana, 2020). In addition to macronutrients, Vis-NIR spectroscopy facilitates the quantification of essential micronutrients such as iron (Fe), manganese (Mn), zinc (Zn), copper (Cu) (van Maarschalkerweerd & Husted, 2015).

2.3.1. Fruit maturation dynamics: critical considerations for precision agriculture

The diversity and adaptability of grape varieties provide fertile ground for investigating maturation dynamics, given their success in diverse environmental conditions (van

Wyngaard et al., 2021). Furthermore, grapes and tomatoes, as watery fruits, offer perceptions of the complex influences of environmental factors on fruit maturation processes.

Analysing grape and tomato maturation holds implications for agricultural research for several reasons. Examining biochemical (e.g., SSC, chlorophyll) and biophysical (e.g., texture) transformations during maturation provides insights into nutritional content, sensory attributes, and overall quality. Secondly, understanding the role of macronutrients and micronutrients in fruit maturation informs PA strategies to optimise crop development (Dhaliwal et al., 2022; Broadley et al., 2012). Thirdly, environmental factors influence pigment dynamics, flavour development, and fruit quality, underscoring the importance of considering these factors in PA (Palliotti & Poni, 2015; Bertin et al., 2015). Lastly, exploring agronomic practices tailored to the unique maturation characteristics of grapes and tomatoes offers promising avenues for enhancing crop quality and yield while minimising resource inputs, aligning with the goals of PA (Brunori et al., 2019; Roberts & Mattoo, 2018).

2.3.2. Fruit maturation: a thorough analysis

Understanding fruit maturation is fundamental for optimising harvest timing and ensuring desired fruit quality. It represents a multi-layered process influenced by various factors, including climatic conditions, physiological changes within the plant, and environmental factors. Fruit maturation marks the stage at which fruits reach their optimal size and quality, indicating their readiness for consumption. Distinguishing maturation from senescence is essential, as it denotes the eventual deterioration of fruits post-maturity (Thomas, 2013).

Physical and visual attributes for watery fruits like grapes and tomatoes serve as markers of fruit quality and readiness for harvest. Parameters such as skin colour and fruit firmness determine the optimal harvest time (Xu et al., 2022). However, the complexity of watery fruit characteristics is further compounded by the interrelationship among different fruit tissues, including the skin, pulp, and seeds (Huang et al., 2018b).

Watery fruits like grapes and tomatoes are primarily assessed based on their physical and visual appearance (Matthews et al., 2009). Skin colour is a key parameter in these fruits, allowing for estimating maturation levels. Similarly, fruit firmness is commonly used to evaluate texture and determine harvest readiness (Maury et al., 2009; Thomas et al., 2006).

The different morphological tissues of fruits play roles in fruit maturation. For instance, the skin accounts for about 25% of the fruit mass in grapes, while seeds represent only

3.8% (Tian et al., 2015; Sparrow et al., 2015). Consequently, the skin, pulp, and seeds exhibit varying levels of pigmentation that affect different physiological processes. The skin contributes to metabolism processes in some grape varieties, such as Sauvignon Blanc, whereas the pulp and seeds play different roles (Tian et al., 2015).

Mechanical properties, such as penetration force and firmness, are crucial for assessing fruit maturation and storage effects (Leiva-Valenzuela et al., 2013; Rolle et al., 2011; Maury et al., 2009). For example, the decrease in grape berry firmness during maturation is attributed to the loss of turgor pressure in cell walls, affecting mechanical properties (Maury et al., 2009; Thomas et al., 2006). Similarly, changes in elasticity and turgor pressure are observed in tomatoes during maturation, impacting their mechanical properties (Saladié et al., 2007).

Upon reaching the market, consumers often assess fruit quality by touch. However, tactile assessments may cause injuries, compromising fruit quality and increasing disease susceptibility (Porat et al., 2018; Ahmad & Siddiqui, 2015). Thus, minimising physical damage during handling and storage is essential for maintaining fruit integrity and extending shelf life, ultimately enhancing consumer satisfaction and reducing food waste.

2.3.3. Tomato maturation: biochemical and biophysical transformations

The dynamics of tomato maturation show a range of biochemical and biophysical transformations influenced by varietal characteristics and environmental factors. Mature tomatoes exhibit a distinctive red hue primarily attributed to carotenes, with lycopene influencing their colouration (Gawad Saad et al., 2017). However, the modulation of these pigments' concentration and expression by environmental factors and storage conditions can impact matured tomatoes' visual appearance and nutritional attributes (Konagaya et al., 2020).

Beyond observable biophysical and visual attributes, exploring fruit tissues shapes tomato quality and nutritional content. The active involvement of the skin, pulp, and seeds in these processes influences pigmentation and triggers specific physiological events (Yuan et al., 2016). Pigmentation, driven by compounds like lycopene and chlorophylls, not only determines visual appearance but also imparts antioxidant properties and contributes to flavour development (Kumar et al., 2021a). Enzymatic activities, gene expression patterns, and metabolic pathways collectively constitute watery fruits' quality, nutritional composition, and sensory attributes, providing insights into tomato maturation dynamics relevant to agricultural practices (Yuan et al., 2016; Chandra & Ramalingam, 2011).

Understanding the biochemical and biophysical details of tomato maturation is essential for optimising cultivation practices, ensuring fruit quality, and meeting consumer preferences. By clarifying the causal processes shaping tomato characteristics, farmers can make informed decisions regarding harvesting, storage, and post-harvest handling, ultimately enhancing tomatoes' marketability and nutritional value.

2.3.4. Grape maturation: biochemical and biophysical transformations

Grape maturation involves a complex interaction of biochemical and biophysical transformations influenced by varietal characteristics and environmental factors. As grapes mature, they undergo metabolic changes that contribute to developing their distinct flavours, aromas, and textures, governed by anthocyanins, flavonoids, and phenolic compounds that shape grapes' colour, taste, and health attributes (Teixeira et al., 2013). The transition from green to red or purple hues in grape berries signifies the accumulation of anthocyanins, serving as key indicators of maturation (Palliotti & Poni, 2015).

External factors such as temperature, light exposure, and water availability also affect grape maturation (Duan et al., 2019). Environmental conditions modulate grape physiology, influencing pigment dynamics and flavour development. Elevated temperatures during maturation expedite biochemical reactions within grape berries, intensifying sugar accumulation and flavour synthesis (Medda et al., 2022). Conversely, water scarcity from drought conditions can prompt stress responses in grapevines, leading to alterations in phenolic composition and flavour profiles (Ribalta-Pizarro et al., 2021). Understanding the relationship between environmental stressors and grape physiology highlights the adaptive mechanisms grapevines employ under varying conditions.

Beyond changes in colour and taste, grape maturation comprehends a range of physiological processes. The skin, pulp, and seeds of grape berries actively participate in these processes, influencing pigment accumulation, aroma synthesis, and flavour maturation (Serrano et al., 2017). Enzymatic activities and gene expression patterns within grape tissues coordinate the synthesis of flavour precursors, shaping the sensory attributes of mature grapes (Silva & Queiroz, 2016). Additionally, macronutrients and micronutrients such as nitrogen (N), potassium (K), and magnesium (Mg) contribute to metabolic pathways during grape development, further shaping mature grapes' quality and nutritional composition (Tewari et al., 2021).

To complement this understanding, the assessment of grape maturation involves a combination of physiological, physical, and mechanical measurements. Studies by

Doumouya et al. (2014) have explored correlations between these parameters for maturation assessment, revealing links between tissue morphological analysis, berry strength, and elasticity during grape ripening (Segade et al., 2011; Gibeaut et al., 2005). Common indicators such as SSC, acidity, pH, and phenolic content are used to estimate grape maturity, with anthocyanin levels increasing as SSC levels rise (Ristic & Iland, 2005). The evolution of seed coat colour also indicates seed and berry maturation stages (Rousserie et al., 2019).

Mechanical and physical properties provide reliable indices of maturation, with grape seeds undergoing three distinct phases of development characterised by changes in weight, appearance, and tannin composition (Ristic & Iland, 2005). The relationship between seed development and grape berry development underscores the importance of assessing mechanical, physical, and physiological parameters to provide insights into the soil-climate-plant relationship (Doumouya et al., 2014).

Unravelling the mechanisms governing grape maturation and integrating advanced sensing technologies into PA systems, strategies can be devised to optimise viticultural practices and enhance grape quality. Real-time monitoring of grape maturation dynamics, enabled by these technologies, facilitates informed decision-making and promotes sustainable grape production practices, aligning with the overarching goals of PA.

2.3.5. Nutrient and environmental influence on grape and tomato maturation

Macronutrients and micronutrients influence the physiological processes and quality parameters associated with fruit maturation. Among the essential macronutrients, N, P, K, Ca, and Mg influence metabolic pathways during fruit development and ripening (Broadley et al., 2012). For instance, N contributes to protein synthesis and chlorophyll production (Ulissi et al., 2011), while P is vital for energy transfer processes and nucleic acid synthesis (Li et al., 2021). K regulates water uptake, sugar transport, and enzyme activation, impacting fruit size, firmness, and sweetness (Zhang et al., 2018b), whereas Ca strengthens cell walls and enhances fruit firmness (Hocking, 2015). Conversely, Mg is crucial for chlorophyll synthesis, photosynthetic carbon fixation, and enzyme activation (Tewari et al., 2021).

Additionally, micronutrients such as iron (Fe), manganese (Mn), zinc (Zn), and copper (Cu) are essential for antioxidant defence, hormone regulation, and secondary metabolite synthesis, further influencing fruit maturation and quality (Dhaliwal et al., 2022). These nutrients are cofactors for enzymes involved in biochemical and

physiological pathways, ensuring proper fruit development, colouration, and nutritional content.

Environmental factors also impact the maturation process of grapes and tomatoes, affecting pigment dynamics, physiological parameters, and responses to environmental inducements. Grapes, for example, experience increased anthocyanin levels in the skin during ripening, enhancing colour development, while chlorophyll content decreases (Palliotti & Poni, 2015). Similarly, tomatoes undergo internal changes such as acidity and sugar content alterations, directly influencing their flavour profile during maturation (Garcia & Barrett, 2006). The relationship of pigments, including chlorophyll, carotenoids, and anthocyanins, reflects on plant health and responses to environmental conditions, with changes in pigment concentrations serving as critical determinants of fruit quality and consumer preference (Mariz-Ponte et al., 2019; Ciaccheri et al., 2018; Adams & Winsor, 2015).

Moreover, climacteric and non-climacteric behaviours, influenced by temperature and other environmental stressors such as salinity and water scarcity, further impact fruit maturation dynamics (Ribalta-Pizarro et al., 2021; Bertin et al., 2015). Agronomic practices, including irrigation and fertilisation, also determine crop quality and yield, offering perceptions of PA strategies (Brunori et al., 2019).

Integrating advanced sensing technologies into PA systems represents an advancement in understanding fruit maturation processes, enhancing the efficiency and applicability of monitoring techniques in agriculture (Sohaib Ali Shah et al., 2020; Hahn & Omenetto, 2012). By unravelling the complex relationships between nutrients, environmental factors, and fruit maturation dynamics, this thesis aims to provide perceptions for improving agricultural practices and fruit quality assessment in the context of PA.

2.3.6. Tomography-like method: assessing fruit maturation

Traditional methods of assessing fruit maturation involve manual sampling techniques tailored to specific crops, such as grapes and tomatoes. For grapes, standard tools include the refractometer, which measures SSC in grape juice to indicate maturation and manual sampling for visual and tactile assessments (Kalopesa et al., 2023; Piazzolla et al., 2016). Similarly, tomato cultivation often relies on visual inspection and manual sampling, assessing attributes like colour, firmness, and aroma (Bertin & Génard, 2018). However, recent advancements in PA have introduced sophisticated sensors and monitoring approaches to enhance fruit maturation assessment, providing detailed perceptions of internal and external fruit tissues for informed decision-making throughout the fruit supply chain.

While traditional methods have been worthy, recent innovations have addressed limitations in obtaining detailed information about internal tissues and predicting biochemical and biophysical characteristics. One such innovative approach is the tomography-like method, which integrates advanced sensors and Vis-NIR spectroscopy for non-destructive monitoring of fruit tissues (Martins et al., 2023). This method enables the reconstruction of internal tissue structures with high accuracy and detail, leveraging spectral data to capture changes in pigmentation and physiological events during fruit maturation.

The operationalisation of the tomography-like method involves techniques such as principal component analysis (PCA), which utilises latent variables (LV) for orthogonal transformations, offering a reduced subspace defined by principal components (Dahlstrand et al., 2019). This method has practical implications in monitoring the maturation of grapes and tomatoes, as these watery fruits facilitate light transmittance through a point of measurement (POM), a critical factor in non-destructive assessment (Sohaib Ali Shah et al., 2020; Martens et al., 2003). Figure 2-2 illustrates the basic principle of the tomography-like method. This principle allows a bi-directional reconstruction through the creation of a feature space that englobes data of the entire fruit and the different inner tissues, offering real-time insights into maturation levels to enable optimal harvest times and support PA practices, such as irrigation and fertilisation while selecting appropriate storage methods and conditions in storage facilities to minimise post-harvest losses and ensure fruit quality.

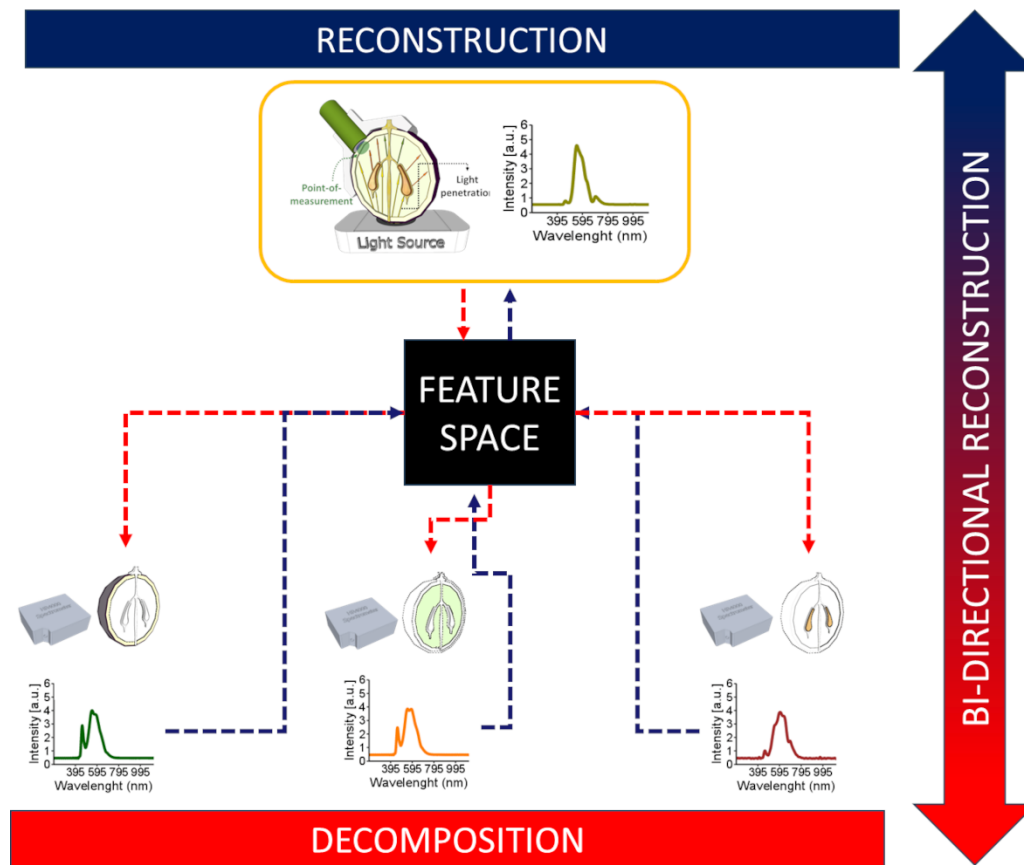


Figure 2-2. Basic principle of the tomography-like method.

The integration of advanced sensing techniques for fruit maturation assessment aligns with the broader objective of PA, contributing to improved agricultural practices, optimised resource utilisation, and enhanced crop quality and productivity (Verma et al., 2021). Accurate monitoring of fruit maturation enables growers to make informed decisions regarding resource allocation and harvest timing, leading to higher yields and increased efficiency. Additionally, precise monitoring ensures that only the highest-quality produce reaches the market, enhancing consumer satisfaction and reducing waste.

The tomography-like method extends beyond academic interest, offering potential applications across the fruit supply chain. From production to harvest monitoring and storage, this method provides real-time insights into maturation levels, supporting informed decisions and addressing consumer concerns about fruit quality (Stiletto & Trestini, 2021). By offering a touchless and reliable method for assessing ripeness, the tomography-like method contributes to dynamic shelf-life determination, minimising post-harvest losses and maximising the value of fruit produce.

2.4. System biology applied in precision agriculture

Crop growth models, such as WOFOST (van Diepen et al., 1989), APSIM (Holzworth et al., 2014), STICS (Brisson et al., 1998), and others, have long been instrumental in predicting the interaction between environmental variables, crop physiology, and agricultural management practices. These models, founded on factors like temperature, soil type, and location, simulate crop growth under varying conditions, aiming to forecast crop performance over time. However, these crop growth models often oversimplify the complex physiological responses of plants to their environment, limiting their ability to capture the full range of plant responses.

In contrast, systems biology takes a broader approach, considering the interactions between genes, proteins, and metabolites within biological systems (Trewavas, 2006). By integrating omics data with computational modelling, systems biology elucidates the molecular mechanisms governing plant growth, development, and responses to environmental inducements (Pazhamala et al., 2021). While traditional models focus on predicting crop performance based on external factors and management practices, systems biology delves into the underlying biological mechanisms, offering a more detailed and mechanistic understanding of plant physiology (Pazhamala et al., 2021; Trewavas, 2006).

Current physiological diagnosis in phenotyping relies on advanced sensing technologies, including remote sensing platforms with multispectral and hyperspectral sensors (Fuentes et al., 2023), thermal (Petrie et al., 2019) and chlorophyll fluorescence imaging (Agati et al., 2008), contributing to a data-rich environment for physiological diagnosis and enabling the assessment of plant responses to environmental conditions. Challenges exist in data integration, model refinement, and validation (Guo et al., 2017), with future directions involving refining sensing technologies (Reis Pereira et al., 2023), advancing computational models (Ahmad et al., 2022), and establishing standardised protocols (Tardieu et al., 2017) to create a robust framework for effortlessly integrating physiological insights into the phenotyping pipeline.

To overcome the constraints of traditional models, the emergence of systems biology offers a holistic perspective on understanding crop physiology and environmental interactions (Lo Presti et al., 2023), integrating various biological components to unravel the complex relationships between genes, proteins, metabolites, and physiological processes (Li et al., 2010). This combination bridges traditional agricultural practices with cutting-edge scientific methodologies, signifying a paradigm shift in advancing the precision and sustainability of modern agriculture (Lo Presti et al., 2023), acknowledging the interconnection of crops and their environment and seeking to leverage this

understanding to optimise agricultural outcomes. Through the lens of systems biology, spectral data becomes not just a tool for observation but a means to delve deeper into the underlying biological mechanisms governing crop growth, development, and responses to environmental conditions (van Wyngaard et al., 2021).

One of the critical applications of systems biology in agriculture is the development of virtual phenotyping, involving creating a computational model to predict plant characteristics based on metabolite profiles, enzymatic activities, and gene expressions (Pazhamala et al., 2021). Integrating bioinformatics tools, such as Flux Balance Analysis (FBA) (Orth et al., 2010), enhances the understanding by providing perceptions of cellular components and connectivity within metabolic pathways. With virtual phenotyping established as a powerful tool, its applications across various crops have been extensively studied, enabling the examination of metabolic shifts, enzymatic responses, and gene activation in plant physiology, thereby optimising agricultural practices for enhanced productivity and sustainability. Incorporating physiological diagnosis into virtual phenotyping assesses plant responses to diverse environmental conditions (Gerlin et al., 2022; Rai et al., 2019), contributing to informed decision-making in crop management practices.

Figure 2-3 illustrates the integration of systems biology within the framework of PA. This approach begins with acquiring plant phenotype measurements through Vis-NIR data, followed by omics analyses encompassing laboratory metabolomic, proteomic, and genomic analysis. Leveraging the hypothesis that physicochemical parameters can be inferred from Vis-NIR data, the process reconstructs metabolic pathways, offering perceptions of physiological metabolism. Ultimately, this methodology generates a digital crop phenotype, facilitating a deeper understanding of its biological processes and responses to environmental conditions that can support agricultural practices (e.g., precision harvest).

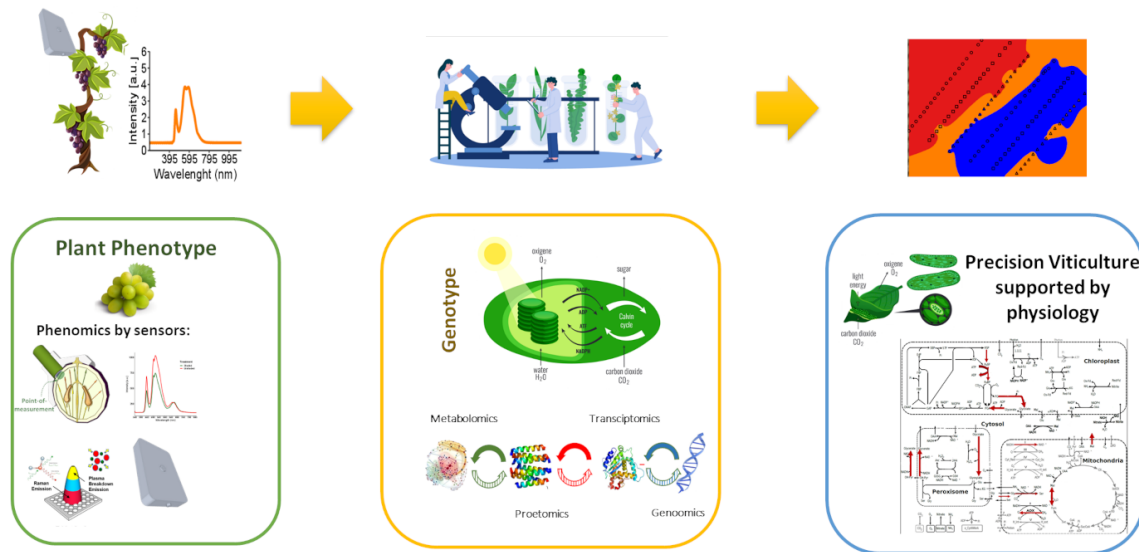


Figure 2-3. Schematic representation of how system biology can be incorporated in the context of precision agriculture.

Metabolic mapping, within the context of PA, represents an approach to understanding the biochemical pathways underlying plant metabolism and physiology. This concept entails visualising and analysing plant metabolic processes, incorporating omics data such as genomics, transcriptomics, and metabolomics (Pazhamala et al., 2021). By integrating these multi-omics datasets and employing computational modelling techniques, metabolic maps provide detailed insights into the molecular mechanisms governing plant growth, development, and responses to environmental conditions. This holistic understanding enables unravelling complex interactions between genes, proteins, and metabolites, offering a mechanistic perspective on plant physiology (Li et al., 2010). Moreover, metabolic mapping facilitates the identification of key metabolic pathways and regulatory networks associated with agronomically important traits, allowing for targeted interventions to optimise crop performance and yield in diverse environmental conditions (Zakhartsev et al., 2016). Metabolic mapping embodies systems biology principles, emphasising the interconnectedness of biological components and processes within living organisms, providing a robust framework for advancing PA practices and enhancing agricultural sustainability (Lo Presti et al., 2023). Systems biology has been integrated into the scope of agriculture to enhance crop physiological understanding. Investigations have focused on model organisms like *Arabidopsis thaliana*, which presents simulated photosynthesis and photorespiration dynamics (de Oliveira Dal'Molin et al., 2010), developed metabolic models elucidating central carbon and energy pathways (Zakhartsev et al., 2016), and explored seed metabolism, including the synthesis of vitamin E (Mintz-Oron et al., 2012). Similarly, in maize studies, efforts have been made to reconstruct cellular structures (Saha et al.,

2011), and investigations into maize yield have provided comprehension into crop productivity (Canas et al., 2017). These examples underscore the practical impact of systems biology in PA, demonstrating its potential to tailor solutions for crop improvement and sustainable agricultural practices.

Through the utilisation of Vis-NIR data, real-time monitoring of various biological tissues can be achieved, providing insights into the prediction of metabolites present in biological materials (van Wyngaard et al., 2021). When incorporated into systems biology models, these metabolites facilitate the reconstruction of phenotypes influenced by the interaction among genetic factors, environmental conditions, and cultural practices. Moreover, it is hypothesised that Vis-NIR data can effectively model metabolites (Falcioni et al., 2023), enhancing the capabilities of system biology models. This study thus represents a novel approach by integrating systems biology with Vis-NIR imaging to explore the relationship between phenotype and genotype, environment, and cultural practices.

Incorporating systems biology into PA aligns effortlessly with the search for a more dynamic and adaptive approach. By providing detailed insights into the physiological and metabolic pathways of crops, virtual phenotyping allows the study of how specific cultural traits (e.g. shad) impact plant processes in the field (Fuentes et al., 2023; Dersch et al., 2016), addressing challenges related to transferability across different years and locations. This integration addresses current limitations and paves the way for innovative solutions to enhance agricultural sustainability. Systems biology fills this gap by reaching into crops' cellular functions, genetic information, and biochemical compositions to address these limitations (Li et al., 2010). In the ongoing OMICBOTS project (<https://www.fc.up.pt/omicbots/>), systems biology is applied to study omics processes, aiming to generate virtual crop phenotypes that intricately capture metabolic pathways, enzymatic activities, and gene activation in different grape metabolic pathways.

2.4.1. Leveraging molecular biology and bioinformatics for enhanced crop phenotyping

Molecular biology approaches offer a powerful means to understand the intricate relationship between physiological processes and cultural practices (Degu et al., 2014). When integrated with systems biology techniques and bioinformatics tools, these approaches provide perceptions into the molecular foundations of plant physiology (Gouot et al., 2019; Hochberg et al., 2013; Dai et al., 2013).

Bioinformatics tools enable the explanation of metabolic engineering in plants by representing metabolic pathways through stoichiometric matrices and employing network-based or constraint-based modelling techniques (Tomar & De, 2013). These

models, often available in the Systems Biology Markup Language (SBML) format, facilitate understanding metabolic pathways despite occasional limitations in accessing genetic and proteomic information (Heirendt et al., 2019).

Creating a virtual phenotype relies on quantifying and identifying metabolites within target pathways and defining reaction constraints through Flux Balance Analysis (FBA) (Aurich et al., 2016). Utilising open databases (e.g. EMBL-EBI, <https://www.ebi.ac.uk/biomodels/>) compiling in-silico crop information, such as those available through EMBL-EBI, enables the exploration of plant physiology when metabolites are quantified (Orth et al., 2010).

Software tools like the COBRA toolbox for Matlab (Heirendt et al., 2019) streamline the creation of virtual phenotypes by facilitating SBML compatibility and providing essential FBA analysis functionalities (Heirendt et al., 2019; Aurich et al., 2016; Tomar & De, 2013). Furthermore, integrating bioinformatics tools with on-the-go sensors capable of quantifying and mapping in-situ metabolites in plants holds promise for real-time crop physiological monitoring (Martins, 2019).

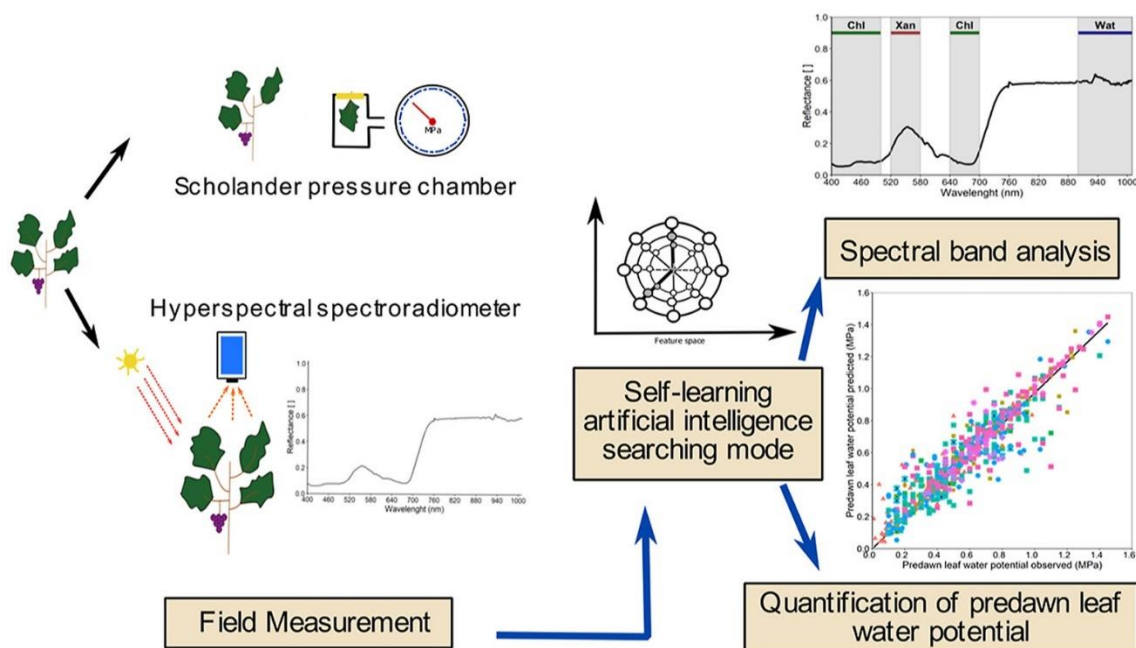
SECTION II

CHAPTER 3

CASE STUDY 1: STUDY ON PREDAWN LEAF WATER POTENTIAL AND ITS IMPACT ON POST-VERAISON PIGMENTS

3. Case study I: study on predawn leaf water potential and its impact on post-veraison pigments.

Tosin, R., Martins, R., Pôças, I., and Cunha, M. (2022). Canopy VIS-NIR spectroscopy and self-learning artificial intelligence for a generalised model of predawn leaf water potential in *Vitis vinifera*. *Biosystems Engineering* 219, 235-258. doi: 10.1016/j.biosystemseng.2022.05.007



Graphical abstract. 3-1. Grafical abstract case study I (Tosin et al., 2022).

Canopy VIS-NIR spectroscopy and self-learning artificial intelligence for a generalised model of predawn leaf water potential in *Vitis vinifera*

Renan Tosin^{1,2}, Rui Martins^{2,*}, Isabel Pôças³, Mario Cunha^{1,2}

¹ Faculdade de Ciências da Universidade do Porto (FCUP), Rua do Campo Alegre, Porto 4169-007, Portugal. RT: renan.tosin@fc.up.pt – MC: mccunha@fc.up.pt

² Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, Porto 4200-465, Portugal.

³ Forestwise, Quinta de Prados, Campus da Universidade de Trás-os-Montes e Alto Douro (UTAD), Vila Real 5001-801, Portugal. IP: isabel.pocas@forestwise.pt

* Corresponding author e-mail: rmcm@inesctec.pt

Highlights

- VIS-NIR spectroscopy was used to predict predawn leaf water potential (Ψ_{pd}).
- Benchmark of SL-AI against state-of-the-art chemometrics methods in modelling Ψ_{pd} .
- Chlorophyll, xanthophyll, and water absorbance bands related to the Ψ_{pd} .
- Cause-effect of plant response to abiotic stresses explained through VIS-NIR bands.

Abstract

This paper focuses on predicting predawn leaf water potential through a self-learning artificial intelligence (SL-AI) algorithm, a novel spectral processing algorithm that is based on the search for covariance modes, providing a direct relationship between spectral information and plant constituents. The SL-AI algorithm was applied in a dataset containing 847 observations obtained with a handheld hyperspectral spectroradiometer (400-1010 nm), structured as: three grapevine cultivars (Touriga Nacional, Touriga Franca and Tinta Barroca), collected in three years (2014, 2015 and 2017), in two test sites in the renowned Douro Wine Region, northeast of Portugal. The Ψ_{pd} SL-AI quantification was tested both in regressive ($R^2 = 0.97$, MAPE = 18.30%) and classification (three classes; overall accuracy = 86.27%) approaches, where the radiation absorption spectrum zones of the chlorophylls, xanthophyll and water were identified along the vegetative growth cycle. The dataset was also tested with Artificial Neural Networks with Principal Component Analysis (ANN-PCA) and Partial Least Square (PLS), which presented worse performance when compared to SL-AI in the regressive (ANN-PCA - $R^2 = 0.85$, MAPE = 43.64%;

PLS - $R^2 = 0.94$, MAPE = 28.76%) and classification (ANN-PCA - overall accuracy: 72.37%; PLS - overall accuracy: 73.79%) approaches. The Ψ_{pd} modelled with SL-AI demonstrated, through hyperspectral reflectance, a cause-effect of the grapevine's hydric status with the absorbance of bands related to chlorophyll, xanthophylls and water. This cause-effect interaction could be explored to identify cultivars and cultural practices, hydric, heating and lighting stresses.

Keywords

Hydric potential; self-learning; spectroscopy; plant stress diagnosis

3.1. Introduction

Water potential (Ψ) is a direct measurement of the canopy water status, being used to diagnose plant water uptake and irrigation. The most used methodology to measure Ψ is based on a pressure chamber (Scholander et al., 1965), a reliable, high-labour, destructive and low-throughput method that does not enable the mapping of vineyards water status. The Ψ can be measured using the stem and/or the leaf, both recognised as good indicators of the vine water status (Netzer et al., 2019; Mirás-Avalos et al., 2014). The choice of the plant organ is mainly based on the feasibility of the measurement (Rodríguez-Pérez et al., 2007), although the stem measurement is more laborious and time-consuming than the leaf (Begg & Turner, 1970). The measurement of Ψ is performed using the leaf at predawn in the Mediterranean climate due to the quick temperature change from predawn to midday (Santesteban et al., 2019).

Visible and near-infrared (VIS-NIR) spectral optical sensors have become non-destructive tools to estimate Ψ in the context of Precision Viticulture (PV) due to the possibility of mapping Ψ . This paper presents a literature review in Table 3-1 and Table 3-2, where Ψ was modelled through spectroscopy techniques performed in several wine regions worldwide. These studies show that the VIS-NIR zone is highly correlated with leaf pigments, which can be used as a proxy of crop water status due to the physiological dependence of leaf pigments on water. The short-wave infrared (SWIR) zone has bands characteristic of water absorption (Kandylakis et al., 2020; Rodríguez-Pérez et al., 2007; Peñuelas & Filella, 1998). Spectroscopy devices, both handheld and on-the-go (mounted on different platforms), are frequently used in the estimation of Ψ due to the detailed spectral information (e.g., Martins et al., 2022; Mariotto et al., 2013). Likewise, multispectral imagery obtained through drones and satellites imagery showed great potential in estimating Ψ (e.g., Romero et al., 2018; Poblite et al., 2017). Nevertheless, the use of spectral optical sensors for estimate Ψ is still under research (Laroche-Pinel et al., 2021; Loggenberg et al., 2018), and the results need both statistical and causal validation for providing reliable estimation (Pôças et al., 2020; Loggenberg et al., 2018; Giovenzana et al., 2018; Poblite et al., 2017; Baluja et al., 2012a).

The current techniques applied in modelling Ψ are mainly based on machine learning algorithms and spectral vegetation indices (VIs) - a data-driven approach - with limitations regarding a robust validation in the field. The data-driven approach presents a few gaps such as: i) lack of cause-effect with the plant physiology, ii) lack of explanation of the interrelationship of different types of abiotic stresses (e.g., hydric, thermal, and lighting), and iii) limitation of the transferability to distinct spatiotemporal conditions (or even between years on the same plot) where the original data were collected and processed. Pointing out these gaps of the data-driven approaches for modelling Ψ , innovative techniques (e.g., sensors, modelling and agronomy) are paramount to allow detailed agronomic information *in-situ* and in real-time about the grapevine's physiology related to the hydric conditions.

Abiotic stresses (e.g., lighting and heating), expressed through the pigment's content, are frequently associated with the canopy water status. Xanthophylls (carotenes group) oxidation is an indicator of heat and lighting stress (Féret et al., 2017; Peñuelas & Filella, 1998) as well as the chlorophylls, which decreases under heating stress and water scarcity (Zarco-Tejada et al., 2013; Sims & Gamon, 2003). However, the different abiotic stresses that affect the canopy are related to the canopy status and measurement time (Palliotti & Poni, 2015). Lighting and heating stresses may not indicate water stress in the canopy (Palliotti & Poni, 2015); on the other hand, if the canopy presents water stress, it is more likely to show light and/or heat stress (Palliotti & Poni, 2015; Chaves et al., 2010; Williams & Ayars, 2005).

Pigments are a proxy of grapevine water status and other abiotic stress (Tosin et al., 2021; Rapaport et al., 2015; Chaves et al., 2010). Pigment's quantification, frequently measured through destructive and very time-laboured methods, is considered a hindrance and non-viable in most cases. VIS-NIR spectroscopy has excellent sensitivity for the quantification of plant pigments and water since these are dominant absorbance constituents in the 300-1200 nm wavelengths (Martins et al., 2022; Tosin et al., 2021; Rapaport et al., 2015; Peñuelas & Filella, 1998). Such makes VIS-NIR spectroscopy ideal for abiotic stress physiological causal diagnosis. Self-learning artificial intelligence (SL-AI) (Martins, 2018) was developed to take advantage of the multi-scaled interferences and matrix effects present in complex spectral signals. The matrix effect is the interaction between all the sample constituents' physical and chemical characteristics, reflected in the spectral information. Spectral information is the super-position of the absorbance of the different constituents, which have their information scattered along the wavelengths according to their absorbance coefficients and proportionality to their concentration (Pasquini, 2018). This multi-scale interference also records physical properties, such as scattering plant tissue/cell structures and molecules. All this information is both convoluted and super-imposed in the spectra (Martins et al., 2022; Pasquini, 2018). In the case of Ψ , water is a major absorbent in the near-

infrared (NIR) region as well as plant stress response given by pigmentation. These are the major interferences analysed in this manuscript that provide a physiological basis for plant water stress. At the core of the SL-AI's methodology lies the concept of covariance-mode (CovM), with the isolation of spectral interference capturing the relationship between the parameter to be quantified and spectral information (Martins et al., 2022; Barroso et al., 2021c; Martins, 2018). The CovM searches for a group of samples that hold the same interference characteristics concerning constituent concentration. Inside this group, interference with other constituents and matrix effects are proportional to the constituent concentration. In this reasoning, CovM does not 'filter' interference, but uses the quantitative information of a particular interference, isolating the different modes of interference present in the data (Martins et al., 2022; Barroso et al., 2021c; Martins, 2018). An optimal CovM is a group of samples that hold the same constituent's composition and spectrum information directly expressed in their latent variables (LV) (Martins et al., 2022). This means that, although spectra and constituents are expressed differently, they hold the same eigen structure, consequently, the same information. SL-AI has been successfully applied to medical point-of-care applications, such as for hemogram quantification (Barroso et al., 2021c), as well as, in precision agriculture for fertilisation management in hydroponics (Silva et al., 2021; Monteiro-Silva et al., 2019) and quantifying sugars and acid in grapes (Martins et al., 2022). In this research, SL-AI is explored as a tool for quantification of predawn leaf water potential (Ψ_{pd}) and diagnosis of *Vitis vinifera* hydric stress.

Thereby, we hypothesise the SL-AI algorithm can estimate the Ψ_{pd} by isolating the CovMs. The pigments and water absorbance contains information that can be proportional to the plant hydric stress stage, allowing a deep analysis of the VIS-NIR data due to the minimisation of the matrix effect, caused by the signal intensity of the different compounds in the hyperspectral data. This plausible hypothesis is based on previously reported observations linking leaf pigment change as a physiological response to hydric stress. This physiological approach is supported by the VIS-NIR zone of the spectrum (400-1010nm), where pigments content is frequently reported (Tosin et al., 2021; Rapaport et al., 2015). The dataset structure of this paper is based on a single plant as an observation unit (instead of a mean of plant group of previous paper) with a strong increase of spectral and Ψ_{pd} data variability. Using a single plant is very relevant to infer the cause-effect explanation, which is critical to transfer technology to decision making in the field. Therefore, we explore the entire spectra because this variability is not compatible with the vegetation indices approach used in previous papers.

Thus, this paper aims: i) to estimate Ψ_{pd} using hyperspectral data from the VIS-NIR zones coupled with SL-AI, able to correct the spectral signal intensity and to minimise the matrix effect, ii) to present a detailed analysis of the different years, cultivars and test sites, iii) to identify through SL-AI and VIS-NIR data (400-1010 nm) how foliar pigments work as a proxy in the grapevine

water status, aiming a cause-effect of the hydric status of the plant and iv) to present a benchmark of SL-AI in modelling Ψ using spectroscopy aiming good statistical goodness of fit for precise information with physiological adherence and to demonstrate the generalisation of the model.

Table 3-1. Bibliography review of the prediction of Ψ in a regressive approach.

Field test conditions	Spectral sensor		Modelling					Reference
	Platform	Spectral Range (nm)	Nobs	Bands (nm) or VI	Range (Ψ) MPa	Method	Statistics	
Ψ: Midday (stem) Location: Chile Cultivar: Carménère	Drone	Multispectral (530-800)	32	550, 570, 670, 700, 800	$-0.6 \geq \Psi \geq -1.8$	ANN	R²: 0.87 RMSE (MPa): 0.12 MAPE (%): 13.3	(Poblete et al., 2017)
Ψ: Midday (stem) Location: Spain Cultivar: Tempranillo	Spectrometer (on-the-go)	Hyperspectral (1100-2100)	540	1450 – 1940	$-0.55 \geq \Psi \geq -2.25$	PLS	R²: 0.69 RMSE (MPa): 0.19 MAPE (%): 21.2	(Diago et al., 2018)
Ψ: Midday (stem) Location: Israel Cultivar: Merlot, Cabernet Sauvignon, Petit Verdot, Syrah, Malbec	Satellite (Planet Satellite)	Multispectral (455-860)	60	GNDVI, NDVI, EVI, SAVI	$-0.63 \geq \Psi \geq -1.41$	SLR	R²: 0.71 RMSE (MPa): 0.11 MAPE (%): 12.2	(Helman et al., 2018)
Ψ: Midday (stem) Location: USA Cultivar: Pinot noir	Spectroradiometer (handheld)	Hyperspectral (350-2500)	125	MBD	$-0.10 \geq \Psi \geq -0.66$	SLR	R²: 0.51 RMSE (MPa): 0.11 MAPE (%): 22.2	(Rodríguez-Pérez et al., 2007)
Ψ: Midday (stem) Location: USA Cultivar: Pinot noir	Spectroradiometer (handheld)	Hyperspectral (350-2500)	125	BA	$-0.10 \geq \Psi \geq -0.66$	SLR	R²: 0.48 RMSE (MPa): 0.12 MAPE (%): 22.8	(Rodríguez-Pérez et al., 2007)
Ψ: Midday (stem) Location: USA Cultivar: Pinot noir	Spectroradiometer (handheld)	Hyperspectral (350-2500)	125	RGI	$-0.10 \geq \Psi \geq -0.66$	SLR	R²: 0.62 RMSE (MPa): 0.1 MAPE (%): 19.4	(Rodríguez-Pérez et al., 2007)
Ψ: Midday (stem) Location: USA Cultivar: Pinot noir	Spectroradiometer (handheld)	Hyperspectral (350-2500)	125	SIPI	$-0.10 \geq \Psi \geq -0.66$	SLR	R²: 0.54 RMSE (MPa): 0.11 MAPE (%): 21.4	(Rodríguez-Pérez et al., 2007)
Ψ: Midday (stem) Location: Spain	Drone	Multispectral (530-800)	10	NDVI	$-1.01 \geq \Psi \geq -1.5$	SLR	R²: 0.68	(Baluja et al., 2012a)

Cultivar: Tempranillo							RMSE (MPa): 0.1 MAPE (%) : 11.1	
Ψ: Midday (stem) Location: Italy Cultivar: Moscato Reale	Satellite (Landsat-8 OLI)	Multispectral (630–1660)	120	NDVI	$-0.40 \geq \Psi \geq -1.68$	SLR	R²: 0.00 MAE (MPa): 0.13 MAPE (%) : 11.55	(Borgogno-Mondino et al., 2018)
Ψ: Midday (stem) Location: Spain Cultivar: Carmenere, Marselan, Godello, Grenache, Pedro Ximénez, Tempranillo	Spectrometer	Hyperspectral (1600-2400)	118	1940	$-0.2 \geq \Psi \geq -1.8$	RF	R²: 0.80 RMSE (MPa): 0.18 MAPE (%) : 20.3	(Gutierrez et al., 2016)
Ψ: Midday (stem) Location: Spain Cultivar: Carmenere, Marselan, Godello, Grenache, Pedro Ximénez, Tempranillo	Spectrometer	Hyperspectral (1600-2400)	118	1940	$-0.2 \geq \Psi \geq -1.8$	M5	R²: 0.71 RMSE (MPa): 0.21 MAPE (%) : 23.7	(Gutierrez et al., 2016)
Ψ: Midday (stem) Location: Spain Cultivar: Albariño, Pedro Ximenez, Verdejo, White Grenache	Spectrometer (handheld)	Hyperspectral (1600-2400)	160	Na	$0.42 \geq \Psi \geq -1.85$	MPLS	R²: 0.67 SE (MPa): 0.23 MAPE (%) : 25.6	(Tardaguila et al., 2017)
Ψ: Midday (stem) Location: Spain Cultivar: Tempranillo	Spectrometer (on-the-go)	Hyperspectral (1100-2100)	180	1420-1450, 1690-1740, 1880-1970	$-0.72 \geq \Psi \geq -2.20$	PLS	R²: 0.86 RMSE (MPa): 0.18 MAPE (%) : 20	(Fernández-Navales et al., 2018)
Ψ: Midday (stem) Location: China Cultivar: Cabernet Sauvignon, Chardonnay, Merlot	Drone	Multispectral (530-810)	270	DVI, GI, MSAVI, NDVI, NDGI, NDRE, OSAVI, RGRI, RDVI, SR	$-0.5 \geq \Psi \geq -1.5$	ANN	R²: 0.53 RMSE (MPa): 0.15 MAPE (%) : 16.7	(Romero et al., 2018)
Ψ: Midday (stem) Location: France Cultivar: Syrah, Chardonnay, Grenache	Spectroradiometer (handheld)	Hyperspectral (350-2500)	118	NRI	$-0.51 \geq \Psi \geq -1.64$	SLR	R²: 0.45 RMSE (MPa): na MAPE (%) : na	(Laroche-Pinel et al., 2021)
Ψ: Midday (stem and leaf) Location: Australia Cultivar: Shiraz	Spectrophotometers (handheld)	Hyperspectral (300-1100) Hyperspectral (350 – 1850)	386	970 1445	$-0.48 \geq \Psi \geq -1.15$	PLS	R²: 0.85 SE (MPa): 0.09 MAPE (%) : 8.1	(De Bei et al., 2011)
Ψ: Midday (leaf) Location: Spain	Spectroradiometer (handheld)	Hyperspectral (350-2500)	36	826, 1110, 1450, 2060	$-0.86 \geq \Psi \geq -1.8$	PLS	R²: 0.54	(González-Fernández et al., 2019)

Cultivar: Tempranillo							RMSE (MPa): 0.18 MAPE (%) : 16.4	
Ψ: Midday (leaf) Location: Israel Cultivar: Cabernet Sauvignon	Spectroradiometer (handheld) Hyperspectral cameras	Hyperspectral (400-2000) Hyperspectral cameras: (400-1000) (900-1700)	32	538, 710, 1400, 1480–1510	$-0.69 \geq \Psi \geq -1.63$	PLS	R²: 0.88 RMSE (MPa): 0.15 MAPE (%) : 13.6	(Rapaport et al., 2015)
Ψ: Midday (leaf) Location: South Africa Cultivar: Pinotage	Spectrometer (handheld)	Hyperspectral (200-900)	29	Na	$-0.68 \geq \Psi \geq -1.49$	PLS	R²: 0.79 RMSE (MPa): 0.09 MAPE (%) : 8.1	(Barnard et al., 2019)
Ψ: Midday (leaf) Location: USA Cultivar: Syrah, Merlot, Cabernet Sauvignon	Spectrometer (handheld)	Hyperspectral (1100-2300)	904	1450 1930	$-0.35 \geq \Psi \geq -1.55$	PLS	R²: 0.76 SE (MPa): 0.15 MAPE (%) : 13.5	(Santos & Kaye, 2009)
Ψ: Midday (leaf) Location: Israel Cultivar: Cabernet Sauvignon	Spectroradiometer (handheld)	Hyperspectral (400-1600)	18	WABI	$-0.88 \geq \Psi \geq -1.11$	SLR	R²: 0.85 RMSE (MPa): na MAPE (%) : na	(Rapaport et al., 2017)
Ψ: Predawn and midday (leaf) Location: USA Cultivar: Thompson Seedless	Drone	Multispectral (530-800)	18	PRInorm	$-0.72 \geq \Psi \geq -1.20$	SLR	R²: 0.77 RMSE (MPa): na MAPE (%) : na	(Zarco-Tejada et al., 2013)
Ψ: Predawn and midmorning (leaf) Location: Italy Cultivar: Biancolella	Spectrometer (handheld)	Hyperspectral (400-1000)	440	704-1893	$-0.10 \geq \Psi \geq -2.25$	PLS	R²: 0.67 RMSE (MPa): 0.31 MAPE (%) : 41.6	(Giovenzana et al., 2018)
Ψ: Predawn and midmorning (leaf) Location: Italy Cultivar: Biancolella	Spectrometer (handheld)	Hyperspectral (1100-2000)	440	704-1893	$-0.10 \geq \Psi \geq -2.25$	PLS	R²: 0.84 RMSE (MPa): 0.23 MAPE (%) : 30.9	(Giovenzana et al., 2018)
Ψ: Predawn and midmorning (leaf) Location: Italy Cultivar: Biancolella	Spectrometer (handheld)	Hyperspectral (400-1000)	440	704-1893	$-0.10 \geq \Psi \geq -2.25$	MLR	R²: 0.68 RMSE (MPa): 0.32 MAPE (%) : 43	(Giovenzana et al., 2018)
Ψ: Predawn and midmorning (leaf) Location: Italy Cultivar: Biancolella	Spectrometer (handheld)	Hyperspectral (1100-2000)	440	704-1893	$-0.10 \geq \Psi \geq -2.25$	MLR	R²: 0.63 RMSE (MPa): 0.33 MAPE (%) : 44.3	(Giovenzana et al., 2018)

Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	127	NRI, PWI, D1	$-0.05 \geq \Psi \geq -0.95$	GBM	R²: 0.80 RMSE (MPa): 0.11 MAPE (%): 28.2	(Pôças et al., 2017)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	127	NRI, PWI, D1	$-0.05 \geq \Psi \geq -0.95$	BRNN	R²: 0.78 RMSE (MPa): 0.12 MAPE (%): 30.8	(Pôças et al., 2017)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	127	NRI, PWI, D1	$-0.05 \geq \Psi \geq -0.95$	B-MARS	R²: 0.79 RMSE (MPa): 0.11 MAPE (%): 28.2	(Pôças et al., 2017)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	127	NRI, PWI, D1	$-0.05 \geq \Psi \geq -0.95$	GAM	R²: 0.78 RMSE (MPa): 0.12 MAPE (%): 30.8	(Pôças et al., 2017)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	127	NRI, PWI, D1	$-0.05 \geq \Psi \geq -0.95$	GBM	R²: 0.85* RMSE (MPa): 0.09* MAPE (%): 23.1*	(Pôças et al., 2017)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI, PWI	$-0.05 \geq \Psi \geq -1.05$	GPR	R²: 0.68 RMSE (MPa): 0.14 MAPE (%): 36.4	(Pôças et al., 2020)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI, PWI	$-0.05 \geq \Psi \geq -1.05$	VH-GPR	R²: 0.67 RMSE (MPa): 0.14 MAPE (%): 36.4	(Pôças et al., 2020)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI, PWI	$-0.05 \geq \Psi \geq -1.05$	BT	R²: 0.68 RMSE (MPa): 0.15 MAPE (%): 37.4	(Pôças et al., 2020)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI, PWI	$-0.05 \geq \Psi \geq -1.05$	RF	R²: 0.64 RMSE (MPa): 0.15 MAPE (%): 38.7	(Pôças et al., 2020)
Ψ: Predawn (leaf) Location: Portugal	Spectroradiometer (handheld)	Hyperspectral (400-1010)	105	SPVlopt1, SPVlopt2, PRI_Cl2opt, NPClopt	$-0.02 \geq \Psi \geq -0.58$	B-MARS	R²: na RMSE (MPa): 0.05	(Tosin et al., 2021)

Cultivar: Tinta Roriz, Touriga Franca, Touriga Nacional							MAPE (%) : 11.6	
Ψ: Predawn (leaf) Location: Portugal Cultivar: Tinta Roriz, Touriga Franca, Touriga Nacional	Spectroradiometer (handheld)	Hyperspectral (400-1010)	105	NRlgreen_LUT, NRlgreen_LWC	$-0.02 \geq \Psi \geq -0.58$	B-MARS	R²: na RMSE (MPa): 0.05 MAPE (%) : 12	(Tosin et al., 2021)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional	Spectroradiometer (handheld)	Hyperspectral (400-1010)	30	VARlopt	$-0.03 \geq \Psi \geq -0.91$	SLR	R²: 0.75 RMSE (MPa): 0.12 MAPE (%) : 30.8	(Pôças et al., 2015)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional	Spectroradiometer (handheld)	Hyperspectral (400-1010)	30	NDGlopt	$-0.03 \geq \Psi \geq -0.91$	SLR	R²: 0.75 RMSE (MPa): 0.12 MAPE (%) : 30.8	(Pôças et al., 2015)
Ψ: Predawn (leaf) Location: Spain Cultivar: Chardonnay	Spectroradiometer (Handheld)	Hyperspectral (400-1100)	30	NDVI	$-0.4 \geq \Psi \geq -1.36$	SLR	R²: 0.87 RMSE (MPa): na MAPE (%) : na	(Serrano et al., 2010)
Ψ: Predawn (leaf) Location: Italy Cultivar: Pinot Blanc	Spectroradiometer (Handheld)	Hyperspectral (480-2200)	96	510-590, 680-720, 1380-1420	$-0.15 \geq \Psi \geq -0.55$	PLS	R²: 0.87 RMSE (MPa): 0.6 MAPE (%) : 7.52	(Pampuri et al., 2021a)
Ψ: Midmorning and midday (leaf) Location: Portugal Cultivar: Touriga Nacional	Robot	Multispectral (532-810)	36	NDVI, PRI	$-0.98 \geq \Psi \geq -1.68$	SLR	R²: 0.69 RMSE (MPa): 0.94 MAPE (%) : na	(Saiz-Rubio et al., 2021)
Ψ: Midmorning (leaf) Location: Portugal Cultivar: Touriga Nacional	Robot	Multispectral (470-970)	84	NDVI, GNDVI, 560, 840	$-0.96 \geq \Psi \geq -1.45$	PLS	R²: 0.57 RMSE (MPa): 0.19 MAPE (%) : 22.47	(Fernández-Novales et al., 2021)
Ψ: Midday (leaf) Location: Portugal Cultivar: Touriga Nacional	Robot	Multispectral (470-970)	84	NDVI, GNDVI, 560, 840	$-1.26 \geq \Psi \geq -1.52$	PLS	R²: 0.42 RMSE (MPa): 0.14 MAPE (%) : 12.64	(Fernández-Novales et al., 2021)
Ψ: Midday (steam) Location: Italy Cultivar: Sauvignese R10	Drone	Multispectral and thermal (530-810)	14	CWSI	$-0.72 \geq \Psi \geq -1.0$	SLR	R²: 0.59 RMSE (MPa): na MAPE (%) : na	(Puig-Sirera et al., 2021)

Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	400, 520-630, 700, 1890, 2400	$0.31 \geq \Psi \geq -1.34$	PLS	R²: 0.81 RMSE (MPa): 0.12 MAPE (%): 13.33	(Wei, Grafton, Bretherton, Irwin, & Sandoval, 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	400-430, 720, 1049, 1400, 1565-1595, 1890, 2250, 2370	$0.31 \geq \Psi \geq -1.34$	RF	R²: 0.70 RMSE (MPa): 0.15 MAPE (%): 16.89	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	400, 430	$0.31 \geq \Psi \geq -1.34$	SVC	R²: 0.74 RMSE (MPa): 0.14 MAPE (%): 15.56	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	400, 520-630, 700, 1890, 2400	$0.31 \geq \Psi \geq -1.34$	PLS	R²: 0.79 RMSE (MPa): 0.13 MAPE (%): 14.11	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	740, 1220, 1700	$0.31 \geq \Psi \geq -1.34$	RF	R²: 0.70 RMSE (MPa): 0.15 MAPE (%): 17.11	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	400, 715-760, 800-1870, 2250-2390	$0.31 \geq \Psi \geq -1.34$	SVC	R²: 0.67 RMSE (MPa): 0.16 MAPE (%): 17.89	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	650-750, 1155, 1370-1400, 1720, 1870	$0.31 \geq \Psi \geq -1.34$	PLS	R²: 0.65 RMSE (MPa): 0.17 MAPE (%): 18.44	(Wei et al., 2021)**

Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	700-1410	$0.31 \geq \Psi \geq -1.34$	RF	R²: 0.71 RMSE (MPa): 0.15 MAPE (%): 16.67	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	650-750, 1155, 1370-1400, 1720, 1870	$0.31 \geq \Psi \geq -1.34$	SVC	R²: 0.68 RMSE (MPa): 0.16 MAPE (%): 17.56	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	670, 970, 1175, 1440, 1929	$0.31 \geq \Psi \geq -1.34$	PLS	R²: 0.70 RMSE (MPa): 0.15 MAPE (%): 16.89	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	670, 970, 1175, 1440, 1929	$0.31 \geq \Psi \geq -1.34$	RF	R²: 0.66 RMSE (MPa): 0.16 MAPE (%): 18	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	Na	$0.31 \geq \Psi \geq -1.34$	SVC	R²: 0.63 RMSE (MPa): 0.17 MAPE (%): 18.89	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	500-650, 700-730, 1400, 1700, 1900, 200-2400	$0.31 \geq \Psi \geq -1.34$	PLS	R²: 0.85 RMSE (MPa): 0.11 MAPE (%): 12.22	(Wei et al., 2021)**
Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	1050-1150, 1200-1300, 1800	$0.31 \geq \Psi \geq -1.34$	RF	R²: 0.67 RMSE (MPa): 0.16 MAPE (%): 17.78	(Wei et al., 2021)**

Ψ: Midday (steam) Location: New Zealand Cultivar: Chardonnay, Pinot Noir, Sauvignon Blanc	Spectroradiometer (Handheld)	Hyperspectral (400-2400)	85	400-500, 750-1300, 1500-1850, 1900-2400	$0.31 \geq \Psi \geq -1.34$	SVC	R²: 0.78 RMSE (MPa): 0.13 MAPE (%): 14.56	(Wei et al., 2021)**

Nobs: dataset size, Na: data non-available, megapascal (MPa), Coefficient of determination (R²), Root mean square error (RMSE), Mean absolute predicted error (MAPE), Artificial Neural Networks (ANN), Partial Least Square Regression (PLS), Green Normalised Difference Vegetation Index (GNDVI), Normalised difference vegetation index (NDVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), Single Linear Regression Model (SLR), Maximum Band Depth (MBD), Band Area (BA), Relative Greenness Index (RGI), Red Green Ratio Index (RGRI), Structure Intensive Pigment Index (SIPI), Random Forest (RF), M5 tree (M5), Modified Partial Least Square Regression (MPLS), Difference Vegetation Index (DVI), Green Index (GI), Modified Soil Adjusted Vegetation Index (MSAVI), Normalised Difference Greenness Vegetation Index (NDGI), Normalised Difference Red Edge Index (NDRE), Optimal Soil Adjusted Vegetation Index (OSAVI), Renormalised Difference Vegetation Index (RDVI), Simple ratio Index (SR), Normalised Reflectance Index (NRI), Water Balance Index (WABI), Photochemical Reflectance Index normalised (PRI_{norm}), Multiple Linear Regression (MLR), Plant Water Index (PWI), Generalised Boosted Model (GBM), Bayesian regularisation-feed-forward neural network (BRNN), Bagged Multivariate Adaptive Regression Splines (B-MARS), Generalised Additive Model (GAM), Gaussian Process Regression (GPR), Variational Heteroscedastic Gaussian Process Regression (VH-GPR), Bagging Trees (BT), Spectral Polygon Vegetation Index optimised 1 (SPV_{lopt1}), Spectral Polygon Vegetation Index optimised 2 (SPV_{lopt2}), Photochemical Reflectance Index with Chlorophyll Index Optimised (PRI_{CI2opt}), Normalised Pigment Chlorophyll Index Optimised (NP_{CIopt}), Normalised Ratio Index Optimised for the Lutein content in the green zone of the spectrum (NRI_{green_LUT}), Normalised Ratio Index Optimised for the leaf water content in the green zone of the spectrum (NRI_{green_LWC}), Visible Atmospherically Resistant Index Optimised (VAR_{lopt}), Normalised Difference Greenness Vegetation Index Optimised (NDGI_{lopt}), Photochemical Reflectance Index (PRI), Crop Water Stress Index (CWSI), Support Vector Regression (SVC). *The dataset used for training considered the cultivar Touriga Nacional and the model's validation considered the cultivars Touriga Franca and Tinta Barroca. **The PLS, RF and SVC were applied in five methods for variable selection (Row reflectance, first derivate, second derivate, continuum removal variable and simple ration indices.

Table 3-2. Bibliography review of the prediction of Ψ in a classification approach.

Field crop	Spectral sensor		Modelling						Reference
	Platform	Spectral Range (nm)	Nobs	Bands (nm) or VI	Range (Ψ) – MPa	Method	Overall accuracy (%)	Number of classes	
Ψ: Midday (stem) Location: South Africa Cultivar: Shiraz	Spectrometer (handheld)	Hyperspectral (450-1000)	60	520.31-585.12	$-1 \geq \Psi \geq -1.8$	XGBoost	80	2	(Loggenberg et al., 2018)
Ψ: Midday (stem) Location: South Africa Cultivar: Shiraz	Spectrometer (handheld)	Hyperspectral (450-1000)	60	520.31-585.12	$-1 \geq \Psi \geq -1.8$	RF	83	2	(Loggenberg et al., 2018)
Ψ: Midday (stem) Location: China Cultivar: Cabernet Sauvignon, Chardonnay, Merlot	Drone	Multispectral (530-810)	270	DVI, GI, MSAVI, NDVI, NDGI, NDRE, OSAVI, RGRI, RDVI, SR	$-0.5 \geq \Psi \geq -1.5$	ANN	83.3	3	(Romero et al., 2018)
Ψ: Midday (stem) Location: France	Spectroradiometer (handheld)	Hyperspectral (350-2500)	118	NRI	$-0.51 \geq \Psi \geq -1.64$	SVM	85	4	(Laroche-Pinel et al., 2021)

Cultivar: Syrah, Chardonnay, Grenache									
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	325	ARlopt NRI	$-0.1 \geq \Psi \geq -1.45$	OLR	73.2	3	(Tosin et al., 2020)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI PWI	$-0.05 \geq \Psi \geq -1.05$	GPR	82	3	(Pôças et al., 2020)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI PWI	$-0.05 \geq \Psi \geq -1.05$	VH-GPR	83	3	(Pôças et al., 2020)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI PWI	$-0.05 \geq \Psi \geq -1.05$	BT	83	3	(Pôças et al., 2020)
Ψ: Predawn (leaf) Location: Portugal Cultivar: Touriga Nacional, Touriga Franca, Tinta Barroca	Spectroradiometer (handheld)	Hyperspectral (400-1010)	190	NRI PWI	$-0.05 \geq \Psi \geq -1.05$	RF	82	3	(Pôças et al., 2020)

Nobs: dataset size, megapascal (MPa), Extreme Gradient Boosting (XGBoost), Random Forest (RF), Difference Vegetation Index (DVI), Green Index (GI), Modified Soil Adjusted Vegetation Index (MSAVI), Normalised Difference Vegetation Index (NDVI), Normalised Difference Greenness Vegetation Index (NDGI), Normalised Difference Red Edge Index (NDRE), Optimal Soil Adjusted Vegetation Index (OSAVI), Red Green Ratio Index (RGRI), Renormalised Difference Vegetation Index (RDVI), Simple ratio Index (SR), Artificial Neural Networks (ANN), Normalised Reflectance Index (NRI), Support Vector Machines (SVM), Anthocyanin Reflectance Index Optimised (ARlopt), Ordinal Logistic Regression (OLR), Plant Water Index (PWI), Gaussian Process Regression (GPR), Variational Heteroscedastic Gaussian Process Regression (VH-GPR), Bagging Trees (BT).

3.2. Material and Methods

3.2.1. Test site

The research was carried out in the Douro Wine Region, northeast of Portugal (Figure S 3-1, in Supplementary material), which has gained notoriety from the quality of its main product, Port wine. The viticulture in this region takes place under particularly rigorous climatic conditions, where vineyards are mainly built over terraces and slopes with soils mostly derived from shale. The vineyards are in some of the most arid regions of Europe. During the ripening period (around 20 July to 20 September), the rainfall is ≤ 28 mm in 80% of the years, and the available water reserve at the end of the ripening period is always $\leq 20\%$, causing lower berry weight and, consequently, lower wine yield, when watering is not possible (Cunha & Richter, 2016). Two commercial vineyards were considered for the study: i) Test Site 1 (TS 1) – Quinta dos Aciprestes (wine company Real Companhia Velha) – Latitude 41.21° N; Longitude 7.43° W; 145 m a.s.l., and ii) Test Site 2 (TS 2) – Quinta do Ataíde (wine company Symington), - Latitude 41.25° N; Longitude 7.11° W; 161 m a.s.l.). A detailed description of the TS, cultivars and irrigation amount is presented in the Supplementary material (Table S 3-1, Table S 3-2 and Table S 3-3).

3.2.2. Predawn leaf water potential

The predawn leaf water potential (Ψ_{pd}) is chosen over the midday leaf water potential due to the quick change of temperature after sunrise in the Douro region, reaching very high temperatures at midday during the summer period, thus making Ψ_{pd} more adequate (Santesteban et al., 2019). Ψ_{pd} was measured using a Scholander chamber (Scholander et al., 1965) (PMS600, Albany, OR, USA) with an associated error of measurement of 0.5%. The Ψ_{pd} were measured in three years (2014, 2015 and 2017) between post-flowering and harvest, within June and September, collecting a random leaf per vine on each date in two test sites (TS1 and TS2). The years 2014 (113 observations, 5 dates in TS 1), 2015 (369 observations, 4 dates in TS 1 and 4 dates in TS 2), and 2017 (365 observations, 4 dates in TS 1 and 4 dates in TS 2) totalled 847 observations for both test sites (Table S 3-1 and Table S 3-2). In 2014, no data were recorded in the TS 2, while in the TS 1 only data of cultivar TN were sampled. The dataset structure, including measurements dates of each year in the test sites, cultivars, and irrigation mode, is presented in the supplementary material (Table S3 and S4). A one-way analysis of variance (ANOVA) with a p-value associated with the Fischer and Duncan post-hoc test ($\alpha = 5\%$) were performed to compare the Ψ_{pd} values between years according to the test site location, irrigation treatment and the cultivars.

3.2.3. Spectroscopy methodology

A portable passive spectroradiometer (Handheld 2, ASD Instruments, Boulder, CO, USA) was used to collect the hyperspectral reflectance from the canopy. The measurements were performed on the same days when Ψ_{pd} was assessed, on cloud-free days, between 11 and 13 h (local time), to minimise changes in solar zenith angle. Reflectance was recorded in the range between the wavelength 325 nm and 1075 nm, with an interval of 1 nm. However, due to the intrinsic noise of data in the spectroradiometer spectral limits, the wavelengths below 400 nm and above 1010 nm were not considered. A more detailed description of ground measurements of Ψ_{pd} and spectral data can be found in previous work (Tosin et al., 2020; Pôças et al., 2020; Pôças et al., 2017; Pôças et al., 2015).

Multiplicative scattering correction logarithm was used to correct baseline and scattering artifacts present in the recorded spectra prior to quantitative algorithms processing (Martins et al., 2022).

3.2.4. Self-learning artificial intelligence

The SL-AI method foundation searches for stable and systematic relationships between the Ψ_{pd} and spectral features (Martins, 2018). Stable covariance is related to the Beer-Lambert law, which is observed at each CovM (Martins et al., 2022). In this sub-group of samples, the covariance is described by a single eigenvector, allowing to unscramble the complex multi-scale interferences between the leaf metabolic composition present in the spectra that provide the direct relationship to Ψ_{pd} . The SL-AI method is performed into two main steps: i) feature space optimisation to maximise covariance between Ψ_{pd} and spectral features; and ii) covariance search for a group of samples that hold consistent covariance is performed to allow an accurate and causal relationship with Ψ_{pd} . Due to the SL-AI architecture, some groups of observations cannot be optimised due to the lack of data to represent the clusters generated through SL-AI. Thus, clusters with insufficient observations are not counted in the model. This paper expects SL-AI to identify zones of the spectrum relevant to the Ψ_{pd} estimation and a cause-effect insight about the grapevine water status. A detailed description of the full methodology is presented by Martins (2018).

The benchmark of SL-AI in quantifying Ψ is presented in both regressive and classification approaches with two sub-structures: i) a comparison with works that used diverse spectroscopy techniques in modelling Ψ , and ii) a comparison of the SL-AI performance against a few methodologies referenced in the literature to assess the dataset presented in this paper and the algorithm's performance.

3.2.4.1. Benchmark on the remote sensing-based previsions of Ψ

The literature review compared diverse works where Ψ was modelled through spectroscopy techniques. Table 3-1 and Table 3-2 present 31 studies performed in several wine regions

worldwide based on spectral optical sensors for estimation of Ψ . The methodologies applied in modelling Ψ followed: i) regressive approach where the dependent variable (Ψ) is assumed as a continuum (studies in Table 3-1) and ii) classification way, where there were established classes for the Ψ according to the stress level (studies in Table 3-2).

The benchmark of the 31 studies referred above was organised and described according to: i) characterisation of the field test, where it is presented the type of Ψ , location and cultivar assessed, ii) sensor's characterisation, describing the type of sensor and spectral range considered, and iii) modelling approaches (regressive and classification), where is detailed the number of observations, bands considered, range of Ψ measured, methodology applied and the statistical results.

In the 31 studies, the unit of Ψ measured was uniformed to megapascal (MPa), and statistical results were standardised in the regressive approach (Table 3-1) by comparison effects. Most of the works presented the coefficient of determination (R^2). However, for those who gave the correlation coefficient (R), this value was squared. The main statistical error indicators were presented through the root mean square error (RMSE), mean absolute error (MAE) and/or standard error (SE). The mean absolute prediction percentage error (MAPE), a bias indicator, was considered. For MAPE calculation (Equation 1), the mean values of Ψ_{pd} were extracted from Santesteban et al. (2019), which presented a detailed metadata assessment for the different types of Ψ (leaf and stem) at different times (predawn and midday) in grapevines.

$$MAPE(\%) = \frac{100 * (\text{statistical error indicator})}{(\text{mean value of } \Psi_{pd})} \quad (1)$$

In the classification approach (Table 3-2), the overall accuracy (%) and the number of observations and classes were considered for comparison.

3.2.4.2. Performance of SL-AI against other modelling techniques

In the same dataset (hyperspectral vs Ψ_{pd}), the performances of SL-AI in predicting the Ψ_{pd} was compared against Artificial Neural Networks with Principal Component Analysis (ANN-PCA) and Partial Least Square (PLS) for benchmarking purposes. ANN-PCA and PLS were chosen due to the good performance in modelling Ψ as shown in Table 3-1 and Table 3-2 (González-Fernández et al., 2019; Romero et al., 2018; Fernández-Novales et al., 2018; Poblete et al., 2017; Rapaport et al., 2015; De Bei et al., 2011).

All models were optimised and validated using cross-validation. ANN-PCA and PLS cross-validation using a split of 70% training set ($n = 595$ for ANN-PCA, and $n = 594$ for PLS) and 30% test set ($n = 252$ for ANN-PCA and $n = 253$ for PLS) were used, iterated for all the dataset at each cross-validation sub-sets. For SL-AI, leave-one-out cross-validation (LOOCV) was used because local clusters have a much smaller number of samples, to which LOOCV is more appropriate for representative sampling and testing.

In n-k cross-validation, the dataset structure and information restrict the number of k samples. If the dataset is significantly large and contains samples holding the same relationship, k can be a large value; on the contrary, if the dataset is less representative, low k is used (Krstajic et al., 2014). In either case, one must never use repeated measures (e.g. replicates) in the dataset, otherwise, the cross-validation null hypothesis cannot be tested because one knows a priori that the information contained in the training set is the same from the test set (Krstajic et al., 2014). LOOCV has a big advantage in complex datasets, such as the one that involves biological variance. The high biological variance in agricultural studies, such as the one presented makes it difficult for researchers to know the exact representativeness of the datasets (Martins et al., 2022). Leave-one-out ensures the researchers that at least we are using the maximum information available in the dataset to predict an unknown sample. Among the works listed in Table 3-1 and Table 3-2, only three used leave-one-out (González-Fernández et al., 2019; Borgogno-Mondino et al., 2018; Baluja et al., 2012a). Of course, studies of prediction error with the increase of k samples are important to demonstrate representativeness of sampled data and perhaps in the future should be standard in agricultural studies. This manuscript presents the feasibility of measuring Ψ_{pd} and the consequences of hydric stress in leaf pigmentation as a tool for water stress management.

The results of the training dataset were applied in the full dataset, aiming to predict Ψ_{pd} in the different years, cultivars and test sites. The ANN-PCA architecture was preceded by a applied in the hyperspectral data (611 bands). The PCA is a non-supervised technique that recognises the dataset geometry through maximisation of the variance in orthogonal components (PCs). Next, the eigenvalues of each PC were assessed, aiming to achieve the maximum variance with fewer PCs. The first five PCs were chosen to perform the ANN-PCA due to the high variance (99.4%). The ANN-PCA model was built following the parameters: i) input layer: the five first principal components of the hyperspectral dataset; ii) hidden layers: there were set three hidden layers, with twenty, ten and five nodes, respectively, and iii) one output layer, where the Ψ_{pd} was predicted. The tangent function was used in the hidden layers and the linear response for Ψ_{pd} . The PLS analysis was performed according to Geladi and Kowalski (1986), which determines the best model when the optimal number of the principal component as input achieves the minimum cross-validation predicted error sum of squares (PRESS).

For the classification approach, SL-AI, ANN-PCA and PLS were based on the three classes of Ψ_{pd} used by Pôças et al. (2020): Low ($\Psi_{pd} \geq -0.25$), Medium ($-0.25 > \Psi_{pd} \leq -0.5$) and High ($\Psi_{pd} < -0.5$).

All the computational processes were performed in R (R Core Team, 2021) with the following packages: “caret” (Kuhn, 2020), “neuralnet” (Fritsch et al., 2019), “pls” (Kristian Hovde Liland et al., 2021), “ggplot2” (Wickham, 2016), “ggpubr” (Kassambara, 2020), “tidyverse” (Wickham et al.,

2019), “ggrepel” (Slowikowski, 2021), “factoextra” (Kassambara & Mundt, 2020), “FactoMineR” (Le et al., 2008) and “hyperSpec” (Beleites & Sergo, 2020). SL-AI was also implemented in R (R Core Team, 2021).

3.2.5. VIS-NIR regions of interest in modelling Ψ_{pd}

PCA was applied in the reflectance data. The first two principal components (PC 1 and PC 2) were considered to explain the total data variance. The loadings of the first principal component (PC 1), which explains the dataset variability, were extracted to identify the main wavelengths (nm) according to the irrigation clusters.

Afterwards, considering the loadings of PC 1, three spectral regions of interest (ROI) in the VIS-NIR related to foliar pigments content and water absorption were established: i) chlorophyll (400-500 nm and 640-700 nm), due to its relation with the water status (Zarco-Tejada et al., 2013), ii) xanthophylls (520-580 nm), a pigment of the carotene group related to the heating and water stress in the canopy (Middleton et al., 2012) and iii) water (900-1010 nm) absorption (Peñuelas & Filella, 1998). For all the ROIs established in this paper, we adopt a macro qualitative (large band width) interpretation of the spectra (Pasquini, 2018). In the case of the water absorption, the SWIR zone is more relevant for water absorption (Kandylakis et al., 2020; Rodríguez-Pérez et al., 2007; Peñuelas & Filella, 1998), while the NIR zone of the spectrum relies more on the biomass (González-Fernández et al., 2015). Thus, for this ROI, we can assume that the leaf thickness varies less than percentage of water (González-Fernández et al., 2015; Thenkabail et al., 2012; Thenkabail et al., 2000).

3.3. Results

3.3.1. Data characterisation

The climatic conditions during the ground measurements period presented large variability among years. In 2017, the weather was characterised by high temperatures and very low precipitation during the summer, while 2014 was the colder and wetter year (Pôças et al., 2020). Such climatic conditions can explain the significant differences in the non-irrigated grapevines that presented higher Ψ_{pd} from 2014 to 2017 (Table 3-3), affecting the irrigation amounts for the same period (Table S 3-3). The irrigation amounts/dates per test site, plot and irrigation treatment were managed by each wine company following the Ψ_{pd} regular measurements (Table S 3-3).

Table 3-3. Statistical results of predawn leaf water potential (Ψ_{pd} , MPa) for the different test sites, irrigation regime and cultivars among the three years of study.

Structural Parameters	N	Year (Ψ_{pd} , MPa)			Mean Ψ_{pd} (MPa)	p-value
		2014	2015	2017		
<u>Test site</u>						
Test site 1 (TS 1)	352	-0.40 ± 0.26	-0.66 ± 0.28	-0.52 ± 0.31	-0.53 ± 0.30	0.070
Test site 2 (TS 2)	495	Na	-0.51 ^a ± 0.26	-0.69 ^b ± 0.30	-0.59 ± 0.30	< 1*10⁻³
p-value	-	-	< 1*10⁻³	< 1*10⁻³	< 1*10⁻³	-
<u>Irrigation regime</u>						
No irrigation (NI)	322	-0.40 ^a ± 0.27	-0.63 ^b ± 0.30	-0.88 ^c ± 0.26	-0.71 ± 0.32	< 1*10⁻³
Irrigation (I)	525	-0.40 ± 0.27	-0.51 ± 0.25	-0.47 ± 0.23	-0.48 ± 0.25	0.690
p-value	-	0.990	< 1*10⁻³	< 1*10⁻³	< 1*10⁻³	-
<u>Cultivars</u>						
Touriga Nacional (TN)	651	-0.40 ^a ± 0.26	-0.57 ^b ± 0.28	-0.65 ^c ± 0.32	-0.57 ± 0.31	< 1*10⁻³
Touriga Franca (TF)	154	Na	-0.51 ± 0.22	-0.57 ± 0.30	-0.54 ± 0.26	0.210
Tinta Barroca (TB)	42	Na	-0.51 ± 0.30	-0.54 ± 0.14	-0.52 ± 0.24	0.700
p-value	-	-	0.190	0.070	0.300	-
Overall mean	847	-0.40 ^a ± 0.26	-0.56 ^b ± 0.27	-0.63 ^c ± 0.31	-0.57 ± 0.30	< 1*10⁻³

Na: data non-available. Means with a p-value < 0.05 is considered statistically different. For the cases that present p-value < 0.05, within lines, different letters signify the means difference according to Duncan's post-hoc test (α = 5%).

The values observed in the measured Ψ_{pd} show significant differences among the test sites and irrigation regime in the three years of data (Table 3-3). TS 1 did not present significant differences between the Ψ_{pd} measured in the three years regarding the test sites. Similarly, the irrigated grapevines did not show significant differences among the years. Due to the colder weather observed in 2014, the Ψ_{pd} in TS 1 for irrigated and non-irrigated grapevines was the same. This atypical condition is supported by Table S 3-3, which shows lower irrigation events and lower irrigation amount in 2014 for TS 1 compared with 2015 and 2017. Despite a non-significative difference in the Ψ_{pd} among the cultivars, Touriga Nacional (TN) showed significant differences among the years, probably due to this cultivar's higher number of observations (Table 3-3). Such variability allowed to test the model's robustness by checking if the performance achieved with the full dataset is maintained when the different years, test sites and cultivars are considered.

3.3.2. Performance of SL-AI against other modelling techniques

In the regressive approach, considering the whole dataset, SL-AI presents better R^2 (0.97 - Figure 3-1i) when compared to ANN-PCA (R^2 = 0.85 - Figure 3-2i) and PLS (R^2 = 0.94 - Figure 3-3i) and lower bias (MAPE = 18.3 % - Figure 3-1i). Although ANN-PCA presented a very good R^2 , the model's error is high (MAPE = 43.64% - Figure 3-2i). PLS presented a good R^2 and a lower bias (MAPE = 28.76% - Figure 3-3i) in the same validation dataset compared to ANN-PCA. For all methodologies PRESS was computed, and SL-AI presented the lowest value, supporting the good predictions (Table S 3-4). Also, all the methodologies were tested according to the years, cultivars and test sites to provide possible data segregation in the Ψ_{pd} modelling. A detailed description of the statistical goodness of fit is presented in Table S 3-4.

Although ANN-PCA and PLS are considered suitable methods for Ψ estimation, a higher bias is expressed through the root mean squared error (RMSE - MPa) and MAPE (%). Also, according to the PRESS calculated, the predictions for ANN-PCA and PLS are worse when compared to SL-AI (Table S 3-4). The statistical goodness of fit was better for this latter, presenting a low model bias.

Figure 3-1 reports the different conditions applied in the SL-AI. When the years are considered, 2014 (Figure 3-1a) reports a higher bias expressed through MAPE (>25%) when compared to the other two years of study (2015 and 2017). The high variability of the data collected in 2014 could be related to an early field data collection (May) in a phase of the vine cycle characterised by very low water stress. Also, the regression methods performed with SL-AI demonstrated higher values of Ψ_{pd} in 2017 compared to 2014 and 2015. Similarly, when the two test sites are compared (Figure 3-1g, 1h) TS 2 presents higher values of Ψ_{pd} , suggesting more drought conditions when compared to TS 1. Regarding the cultivars assessed, all the cultivars present similar statistical goodness of fit (Figure 3-1d, 1e, 1f) except the cultivar Tinta Barroca (TB), which shows lower bias (MAPE < 15%). When performed with SL-AI, the lower number of observations for this cultivar may justify the better results. Regarding the Ψ_{pd} estimated, for TN, especially in 2017, there is a higher level of water stress, suggesting that 2017 was a very drought year and TN is more likely to present hydric stress when compared to Touriga Franca (TF) and TB.

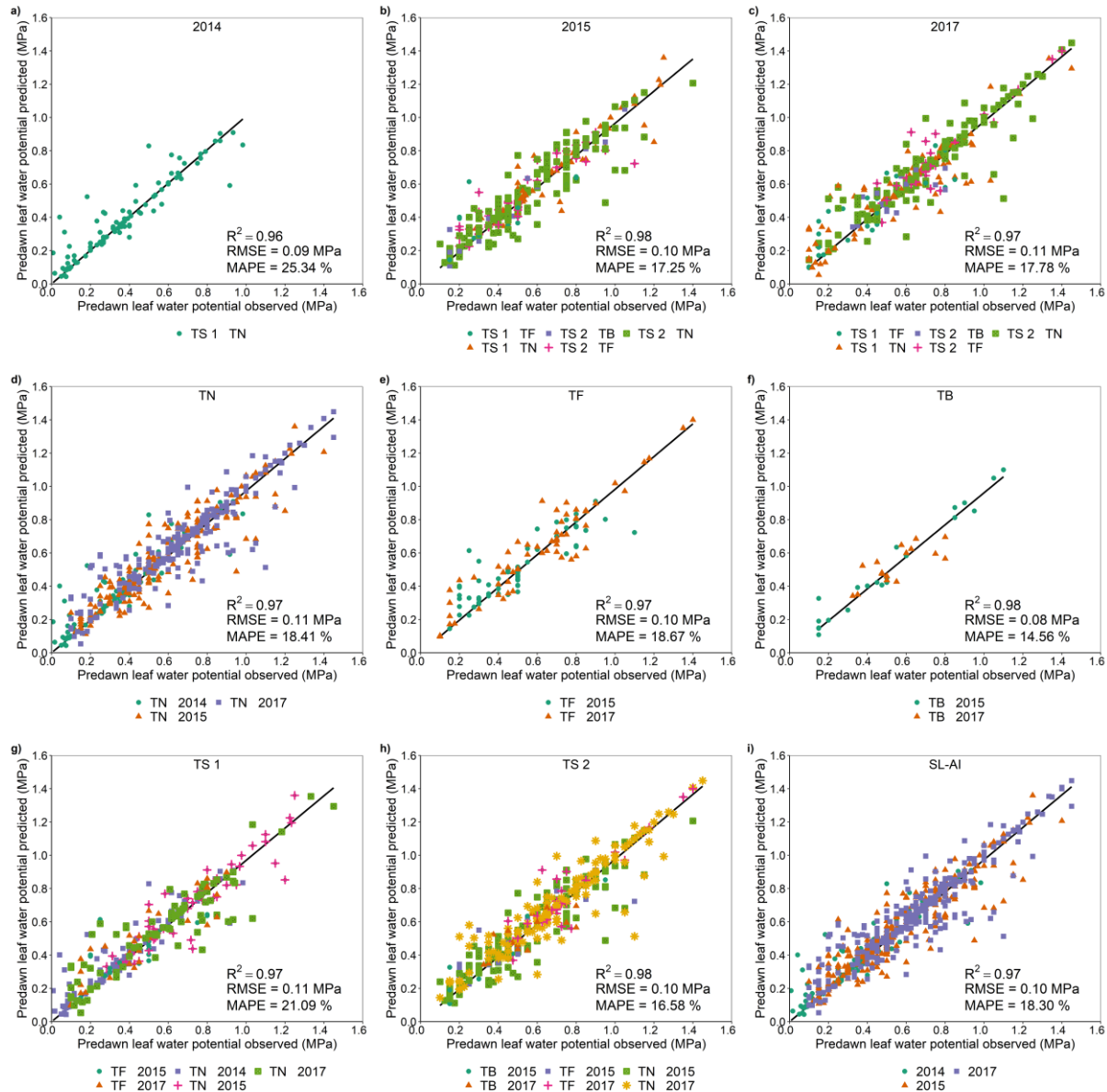


Figure 3-1. Regressions with Self-learning Artificial Intelligence (SL-AI) in the three years of study (i). Data was performed separately according to the years (a: 2014, b: 2015 and c: 2017), cultivars (d: TN: Touriga Nacional, e: TF: Touriga Franca and f: TB: Tinta Barroca) and test site (g: TS 1: Quinta dos Aciprestes, h: TS 2: Quinta do Ataíde).

Considering the tests with ANN-PCA (Figure 3-2), although all cases presented higher R^2 , the bias shown is too high. The higher bias in the ANN-PCA is due to the lower variance error (Geman et al., 1992). The ANN-PCA structure is very good to discriminate non-linear groups. In linear variables like Ψ_{pd} , it is challenging to optimise the spectral data due to the higher gradients observed, which does not allow to assume the classical boundary classification techniques, promoting a higher bias as observed (Martins et al., 2022). Also, in spectral data, ANN is very good to adapt the network according to the structure and neurons parameters, but it is an inefficient method for unscrambling spectral interference and matrix effects (Barroso et al.,

2021c). The better statistical goodness of fit presented for this methodology is when TS 2 was considered (Figure 3-2h), possibly due to the lower data variation and better clustering data.

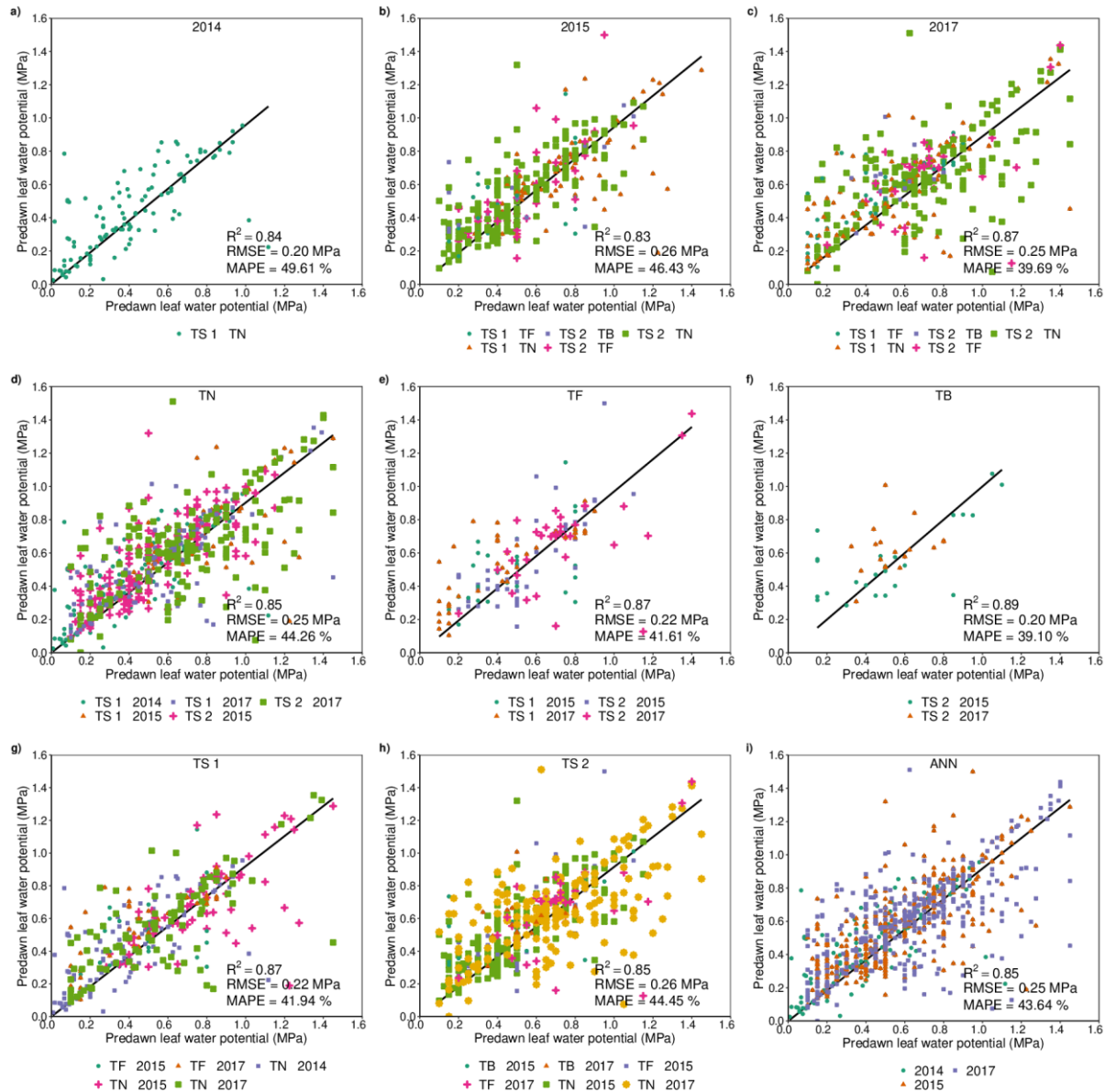


Figure 3-2. Regressions with Artificial Neural Network with Principal component analysis (ANN-PCA) in the three years of study (i). Data was regressed separately according to the years (a: 2014, b: 2015 and c: 2017), cultivars (d: TN: Touriga Nacional, e: TF: Touriga Franca and f: TB: Tinta Barroca) and test site (g: TS 1: Quinta dos Aciprestes, h: TS 2: Quinta do Ataíde).

Regarding the PLS methodology (Figure 3-3), the R^2 and RMSE presented are robust in all cases, but MAPE shows some inconsistencies in the results. PLS is robust if the covariance is stable (Barroso et al., 2021c), which is not a characteristic of the dataset. Under complex co-variance, each PLS latent variable converges towards the maximum global co-variance, which is an average of the information present in the dataset. Under these circumstances, PLS is able to represent the main relationship, but lacks detail for accurately representing interference details that enable high-quality quantifications in complex datasets. As observed in SL-AI, the year 2014

(Figure 3-3a) presents a higher bias when compared to other years (2015 and 2017), perhaps due to the higher data variability in 2014. Possibly due to a lower number of observations, the cultivar TB presented a higher bias (Figure 3-3f). Likewise, TS 1 presents a higher bias (Figure 3-3g), suggesting a high data variability in this TS.

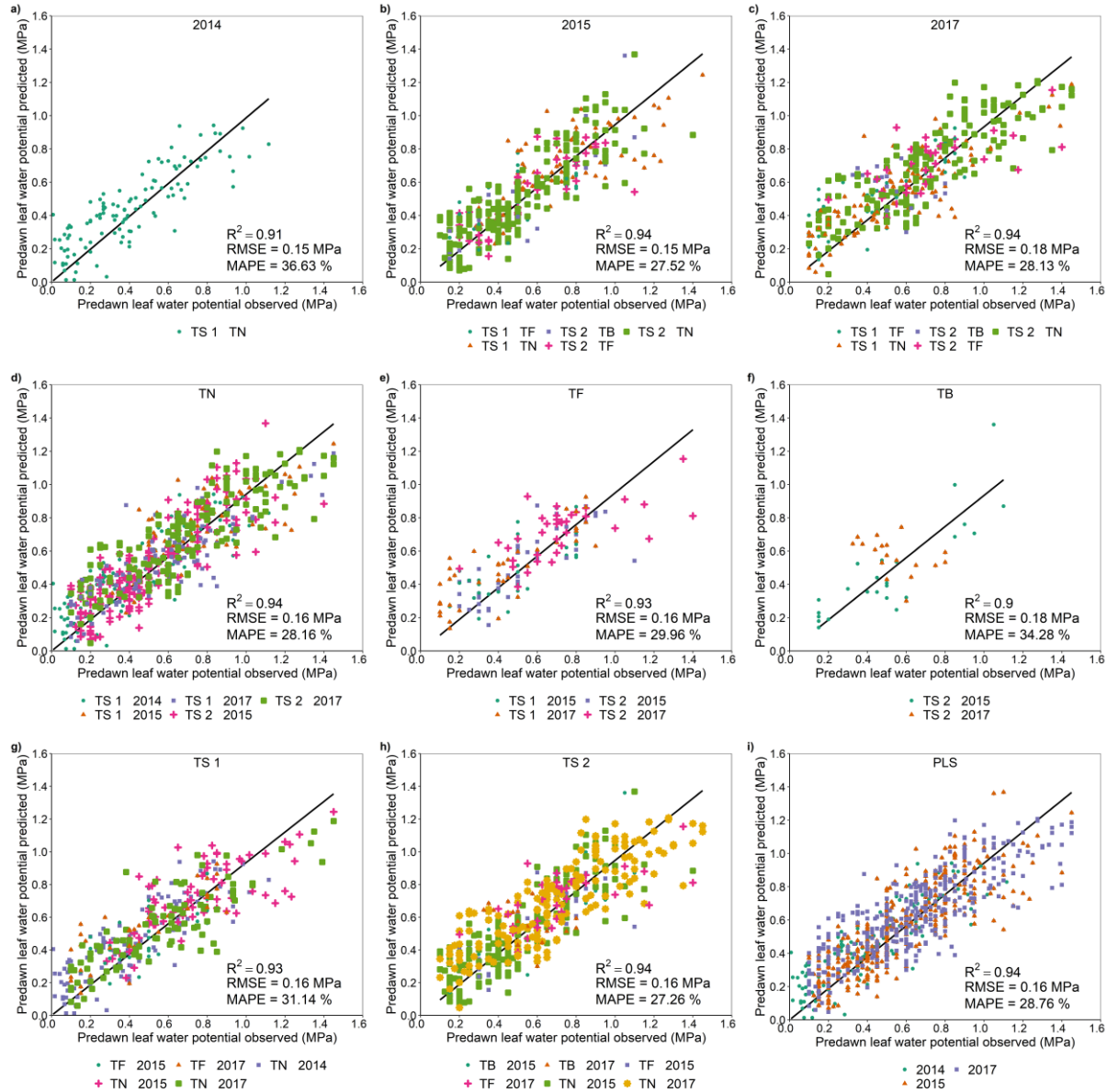


Figure 3-3. Regressions with Partial Least Square (PLS) in the three years of study (i). Data was performed separately according to the years (a: 2014, b: 2015 and c: 2017), cultivars (d: TN: Touriga Nacional, e: TF: Touriga Franca and f: TB: Tinta Barroca) and test site (g: TS 1: Quinta dos Aciprestes, h: TS 2: Quinta do Ataíde).

Table 3-4 presents a confusion matrix where the entire dataset, performed with SL-AI, ANN-PCA and PLS, were classified. Although the overall accuracy is considered high for SL-AI (>86%), in a very heterogeneous and imbalanced dataset, the balanced accuracy (%), a metric that considers the average corrected proportion of each class individually, should be considered (Brodersen et al., 2010). When each class is assessed separately, the balanced accuracy

penalises the class with lower observations (Low); still, it improves the classification of the classes with a higher number of observations (Medium and High). The same trends were observed in the performance of ANN-PCA and PLS, supporting the use of balanced accuracy. PLS has an overall accuracy of 73.79% and ANN-PCA of 43.24 %, despite an R^2 of 0.865. The high bias reflects the lack of capacity of ANN-PCA to resolve local interferences and matrix effects, resulting in lower classification accuracy.

Table 3-4. Confusion matrix of the classification of predawn leaf water potential (Ψ_{pd}) in the three classes established (Low: $\Psi_{pd} \geq -0.25$ MPa, Medium: -0.25 MPa $> \Psi_{pd} \geq -0.5$ MPa and High: $\Psi_{pd} < -0.5$ MPa) using the Self-Learning Artificial Intelligence, Artificial Neural Networks with Principal Component Analysis and Partial Least Square.

Ψ_{pd} Predicted	Ψ_{pd} Observed			Prediction class (%)	by	Balanced accuracy (%)	Overall prediction (%)
	Low	Medium	High				
<i>Self-Learning Artificial Intelligence (SL-AI)</i>							
Low	85	5	0	94.44		85.28	86.27
Medium	28	169	17	78.97		84.85	
High	6	40	349	88.35		90.77	
<i>Principal Component Analysis (ANN-PCA)</i>							
Low	72	17	13	68.82		72.40	72.37
Medium	44	171	48	61.90		72.23	
High	15	95	372	79.78		81.15	
<i>Partial Least Square (PLS)</i>							
Low	63	19	1	75.90		69.29	73.79
Medium	82	166	36	58.45		71.59	
High	7	77	396	82.5		85.58	

The overall classification according to the year, cultivar, and test site is presented in Figure S 3-2. SL-AI showed higher accuracy (> 80 %) for the years sampled, cultivars, and test sites. TB presents slight better overall accuracy (Figure S 3-2) when compared to the other cultivars, possibly due to the lower number of observations. Also, TS 2 presents a slightly higher accuracy (Figure S 3-2), possibly due to lower variability in the data. Regarding the ANN-PCA, a very low overall accuracy is presented in all cases. However, a slightly better overall accuracy is shown in 2015 and TS 2, suggesting the impact of the higher number of observations. Regarding the PLS, there are differences according to the cultivars. TN presents better overall accuracy, related to the lower bias obtained in the regression approach for this cultivar (Figure 3-3d). Therefore, a higher MAPE (%) will contribute to the misclassification of Ψ according to the established classes. This study shows the feasibility of Ψ_{pd} quantification using the VIS-NIR spectra, where SL-AI has improved accuracy when compared to widely used methods such as PLS and ANN-PCA. Cross-validation null hypothesis by random sampling ensures that the information contained in the 3 years of the study can predict Ψ_{pd} within the high variability of the dataset. In future studies, with an increased database, more complex validations should be performed to assess if i) there is time sequence validation - previous years' data can predict current year data; ii) different varieties can predict Ψ_{pd} of an unknown variety; and iii) the quality of Ψ_{pd} predictions under interpolation or extrapolation using data augmentation.

3.3.3. Physiological insights in modelling Ψ_{pd}

The irrigation regime affects the Ψ_{pd} as observed in the field (Table 3-3), and it is expected to influence the spectral reflectance in the grapevines sampled. We hypothesised that water stress is related to the foliar pigment's content along with vegetative growth. While Table 3-3 shows statistical differences according to the irrigation regime and test site, Figure 3-4 identifies differences between the irrigation treatments according to the loading extracted from PC 1 to find the representative bands for the three ROIs established.

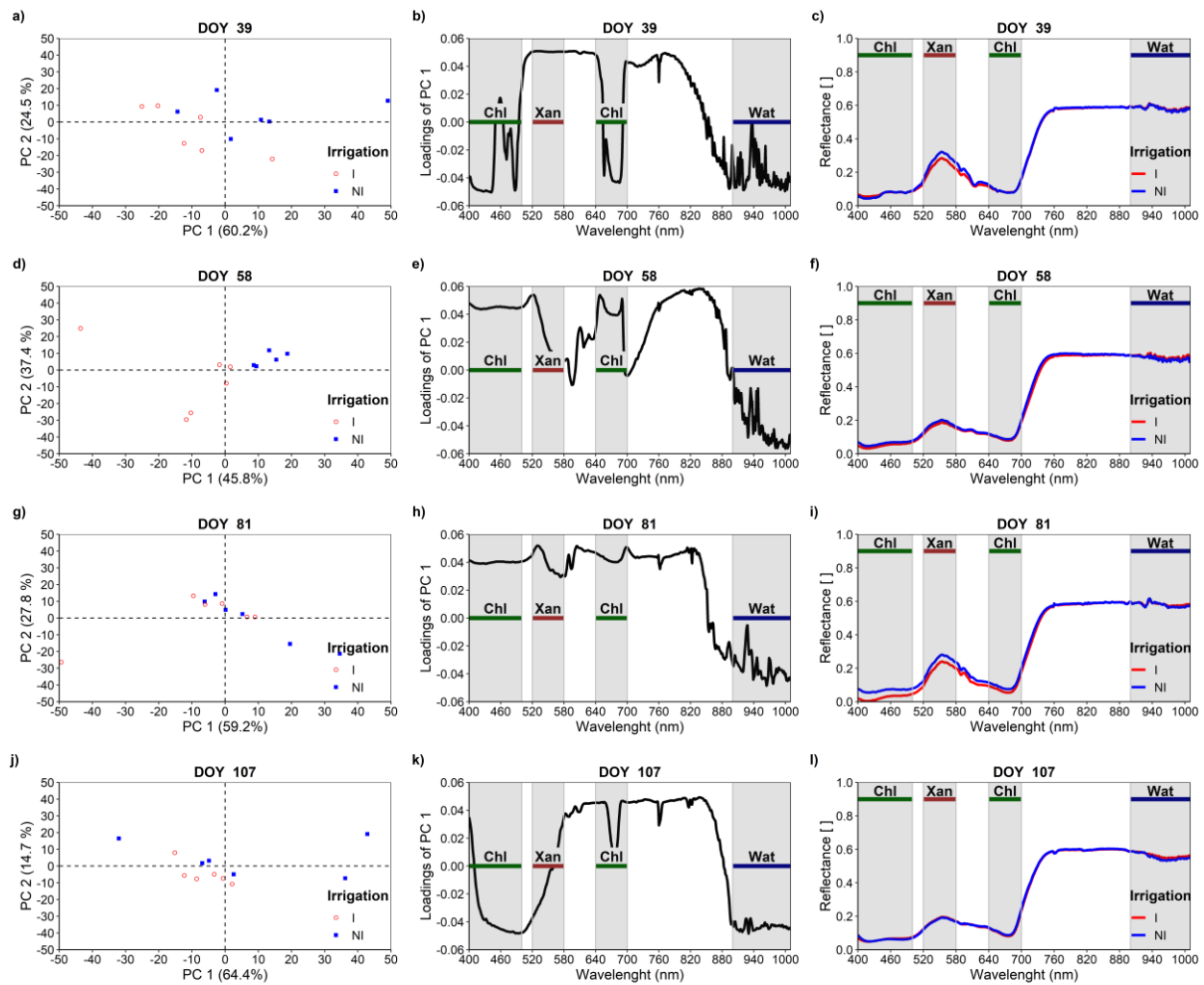


Figure 3-4. Principal Component Analysis (PCA) with the loadings for the test site 2 (Quinta do Ataíde) in the cultivar Touriga Franca in 2015. 'a', 'd', 'g' and 'j' represent the PCA with the two first Principal Components (PC 1 and PC 2) with the variance explained in each component. No filled circles mean irrigation treatment (I) and filled squares the non-irrigated treatment (NI). 'b', 'e', 'h' and 'k' represent the loadings extracted from PC 1 (Y-axis) per wavelengths (X-axis) corresponding to the group separation presented in 'a', 'd', 'g' and 'j'. 'c', 'f', 'i' and 'l' represent the mean reflectance according to the irrigation regime. The zones of the spectrum were identified according to chlorophyll (Chl), Xanthophyll (Xan) and Water (Wat) zones of the spectrum.

Due to the high number of parcels, cultivars, and years, the cultivar TF in TS 2 in 2015 (four measurement dates) is selected to represent the spectral difference, according to the irrigation regime (Figure 3-4). The variance of the first two principal components is presented, and ROIs

are identified through the loading from PC 1. Likewise, the mean reflectance according to the irrigation regime.

The patterns observed in Figure 3-4b, 4e, 4h and 4k for the xanthophyll bands, which is reflective in the near-infrared heat dissipation (Middleton et al., 2012), suggest the oxidation of this pigment along the time due to the leaf senescence and the higher absorption in the xanthophyll zone (Figure 3-4c, 4f, 4i and 4l). Also, xanthophylls are more likely to oxidate in non-irrigated grapevines to cope with the lighting, heating, and water stress (Figure 3-4). This evidence is supported in Figure 3-4c, 4f, 4i and 4l, where the oxidation of this pigment suggests higher concentrations, expressed through lower reflectance, in the non-irrigated vines due to the higher water stress submitted.

The chlorophyll content, associated with water content and vegetative growth (Zarco-Tejada et al., 2013), indicates a difference in the reflectance data according to the irrigation treatment along the time. When higher are the chlorophylls concentration, and higher is water content, lower is the reflectance in the established chlorophyll ROIs. In TS 2, DOYs 39 and 107 (Figure 3-4) are the most representative to report this dynamic. According to the irrigation regime illustrated through the PCA analysis, there is a good group separation in these cases. The loadings of PC 1 report the relevance of the established ROIs according to the irrigation regime and support that the chlorophyll zones are directly related to the zones of water in the irrigated grapevines. However, this evidence is not fully noticed in the reflectance data.

3.3.4. Benchmark on the remote sensing-based previsions of Ψ

A literature review of modelling Ψ (based on Table 3-1 and Table 3-2) is represented in Figure 3-4, where the regressive and classification approaches are presented. In the regressive approach, works that present MAPE > 20% allows a qualitative assessment of the results, for those which present a MAPE < 5% allow an analytic assessment. A semi-quantitative assessment is characterised in works with MAPE between 5% to 20%. Similarly, in the classification approach, studies that present an overall accuracy > 80% allows a qualitative assessment.

In the benchmark, SL-AI performed the best statistical goodness of fit in both modelling approaches. SL-AI corrected the different intensities of signal and identified the spectral zones relevant to predict Ψ_{pd} , minimising the bias effect in the data and performing robust statistical goodness of fit (Figure 3-1). According to the literature review presented in Figure 3-4, there are three studies (Wei et al., 2021; Fernández-Novales et al., 2018; De Bei et al., 2011) with a higher number of observations ($n > 60$) that is in the zone of semi-quantitative assessment. However, these studies do not provide higher data variability as presented in our dataset, because: i) two (Fernández-Novales et al., 2018; De Bei et al., 2011) estimated Ψ with only one cultivar, and ii) one (Wei et al., 2021) is more than one cultivar but only one year is studied. Additionally, the Ψ_{pd}

predictions with SL-AI are stable according to multiple conditions presented in our dataset (Figure 3-1, Table S 3-4).

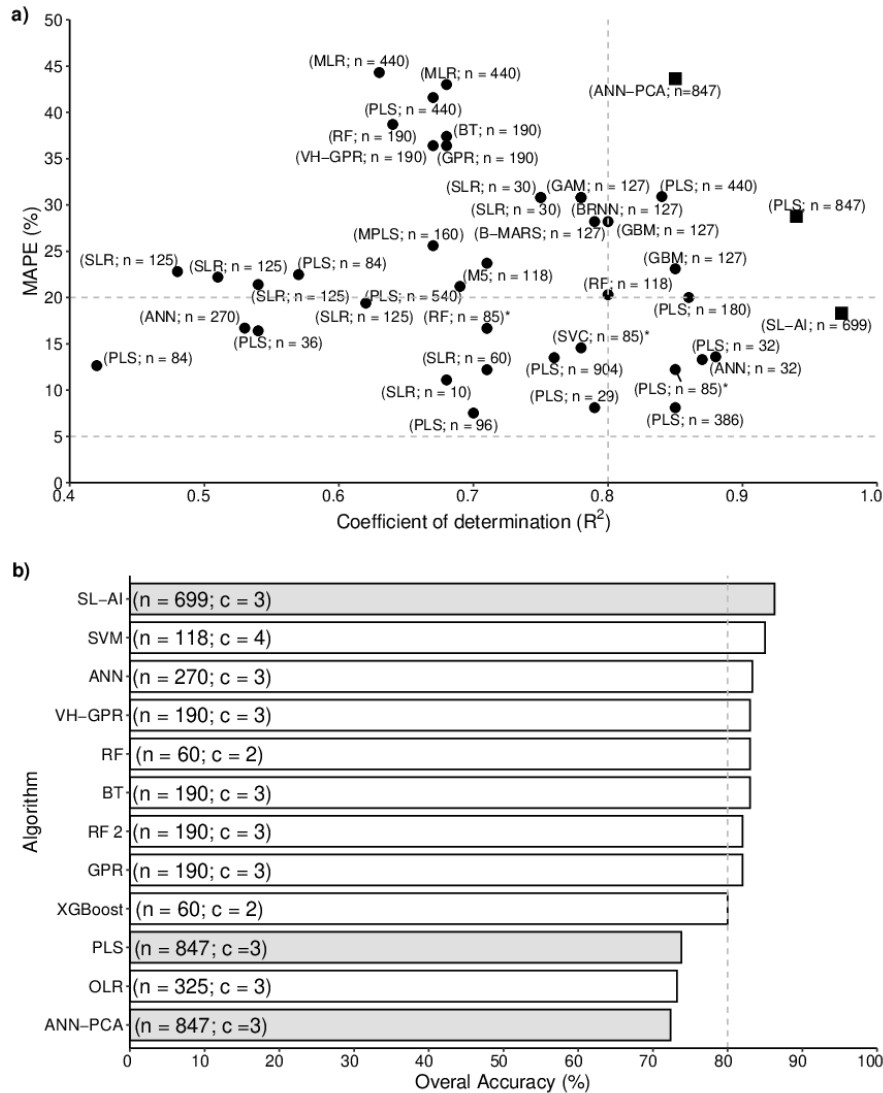


Figure 3-5. Benchmark of estimation of water potential (Ψ) based on optical devices and in a regression ('a') and classification ('b') approach, n: dataset size, c: number of classes of Ψ , PLS: Partial Least Square Regression, SLR: Single Linear Regression Model, MPLS: Modified Partial Least Square Regression, MLR: Multiple Linear Regression, GBM: Generalised Boosted Model, BRNN: Bayesian regularisation-feed-forward neural network, B-MARS: Bagged Multivariate Adaptive Regression Splines, GAM: Generalised Additive Model, ANN: Artificial Neural Network Regression, RF and RF 2: Random Forest, GPR: Gaussian Process Regression, VH-GPR: Variational Heteroscedastic Gaussian Process Regression, BT: Bagging Trees, XGBOOST: Extreme Gradient Boosting, OLR: Ordinal Logistic Regression, SVM: Support Vector Machines, SL-AI: Self-learning artificial intelligence, ANN-PCA: Artificial Neural Network with Principal Component Analysis and SVC: Support Vector Regression. In the regressive approach, squared dots, and in the classification approach, grey bars represent the results performed with the dataset presented in this work. *Due to the similarity performance and numerous variable selection methods, the best result of each methodology tested by Wei et al. (2021) is represented in this figure.

Similar reasons to other studies (Martins et al., 2022; Barroso et al., 2021c) explain lower performances of ANN-PCA and PLS. ANN-PCA tries to map spectral patterns to Ψ_{pd} . However, as spectral interference is constructive or destructive, infinite combinations are possible for the same Ψ_{pd} value. Such information scrambling does not allow ANN to efficiently use the principal component scores information to be recombined inside the hidden layers to connect a spectral

fingerprint to a Ψ_{pd} value. PLS optimises the number of LVs to cope with interference and matrix effects, which introduces non-linearities. PLS is generally thought a linear method, but the combination of information using LVs is non-linear (Cook & Forzani, 2021). However, as PLS uses all data to extract LVs, these represent the major interferences, not allowing to go deeper into unscrambling the existing covariance modes (Martins et al., 2022).

3.4. Discussion

The studies reported in the state-of-the-art (Table 3-1 and Table 3-2) present the following insufficiencies in modelling Ψ : i) lack of genetic variability, presenting only one cultivar in modelling Ψ , and a low number of samples ($n < 60$), reducing the data variability; ii) low range of the spectrometer used, not enabling to register changes in both key pigments and water absorption bands; iii) the data-driven techniques applied in modelling Ψ , in some cases, cannot provide if there is a physiological cause-effect in the results or if there is a statistical coincidence. Although most works have reported coherent zones related to the water absorption in the NIR zone, many did not present a physiological insight in modelling Ψ , nor a temporal assessment of the results. The dataset structure influences the modelling due to the variability promoted and the number of observations sampled. The spectral data is complex and presents vast information in the spectrum (Pasquini, 2018). The VIS-NIR zone of the spectrum reports bands associated with foliar pigments (e.g., chlorophylls, xanthophylls) and water absorption. Regarding the Ψ modelling, a high number of samples should be considered due to high variability among plants. The matrix-effect, caused by an interference of other compounds in the same matrix (Pasquini, 2018; Zhou et al., 2017) is the main challenge in modelling Ψ through VIS-NIR data due to bands related to pigments and water absorption. Regarding the dataset presented in this paper, the high variability and the high number of observations are demonstrated in Table 3, which presents significant differences in the Ψ_{pd} measured among the test sites, irrigation treatments and years of study. Despite our results with these two techniques presenting lower performance, when compared to other studies, ANN-PCA performed good results compared to works with very low observations and only one cultivar (e.g., Poblete et al., 2017). ANN has been used as a black-box approach to model spectroscopy due to its capacity to adapt to non-linear responses, mapping inputs to outputs by non-linear functions. ANN class of models are inherently unstable under optimisation of the architecture with complex spectroscopy data. As we show with CovM, quantitative information must be unscrambled by local models to try to find stable covariance between composition and spectral (Martins et al., 2022; Barroso et al., 2021c). ANN struggles with many local minimums to cope with a general architecture that can fit the global minimum, which is not to the performance of the several sub-sets local minimum. Similarly, when the PLS is compared with works containing only one cultivar (e.g., González-Fernández et al., 2019;

Giovenzana et al., 2018), our results performed better statistical goodness of fit. However, neither ANN-PCA nor PLS could not solve the matrix-effect in the spectral data in modelling Ψ_{pd} due to the high gradient level presented in the hyperspectral data, which means that unknown variables are selected to estimate Ψ_{pd} . On the other hand, SL-AI mitigated the multi-scale interferences (e.g., light conditions, footprint size) and minimised the matrix-effect (e.g., absorption bands related to chlorophyll, xanthophyll, water) for Ψ_{pd} , presenting a cause-effect in the results. The matrix-effect correction and positive cause-effect interpretation of the results through SL-AI were observed in other areas of science (Martins et al., 2022; Silva et al., 2021; Barroso et al., 2021c, 2021d, 2021b, 2021a; Monteiro-Silva et al., 2019). Nevertheless, SL-AI did not count the gradient clusters with insufficient observation to be representative in modelling Ψ_{pd} .

The VIS-NIR spectral data presents detailed information able to detect several biophysical parameters (Pasquini, 2018). However, the literature review showed some studies had used devices with a limited band range, missing important zones of the spectrum related to other abiotic stresses associated with Ψ . Leaf pigments (e.g., chlorophyll, xanthophyll), mainly detectable in the visible zone of the spectrum, work as a proxy of abiotic stresses (e.g., water, lighting, heating). Simultaneously with water stress, lighting and heating stresses may occur at the moment of measurement (Palliotti & Poni, 2015). For example, Rapaport et al. (2015) reported spectral zones associated with the xanthophylls (531-538 nm) and water absorption (1400-1500 nm) in the assessment of vineyard water status indicators, including Ψ .

Similarly, Tosin et al. (2021) estimated Ψ_{pd} in grapevines based on optimised VIs integrating bands in the range 520-551 nm. Nevertheless, none of the works provided an analysis of the abiotic stresses along the grapevine vegetative growth cycle. This paper could provide an analysis and a more detailed insight into grapevine physiology through the robust results of SL-AI performance. Figure 3-4 shows the analysis of the hyperspectral information, concerning the characteristic spectral domains for chlorophyll, xanthophyll and water absorption.

The data-driven techniques have been highly reported as a main technique in modelling Ψ . Although some works presented very good statistical goodness of fit, the lack of interaction with other abiotic stress, promoting a physiological adherence to the results obtained, is one of the main limitations of this technique (Rudin, 2019). Spectral vegetation indices (VIs) (with two or more electromagnetic spectrum bands) have been highly reported as good indicators for estimating Ψ in 50% of the studies (Table 3-1 and Table 3-2). Although the use of VIs are good indicators of crop biophysical characteristics, it is not possible to distinguish if the same bands maintain along the crop vegetative growth cycle due to the different types of abiotic stresses that the plant may present (Gold et al., 2019; Palliotti & Poni, 2015). Yet, using the same dataset and different modelling techniques in estimating Ψ , some authors reported differences in the results (Pôças et al., 2020; Giovenzana et al., 2018; Pôças et al., 2017; Gutierrez et al., 2016; Pôças et

al., 2015; Rodríguez-Pérez et al., 2007). SL-AI demonstrated that for the specific ROIs established, the bands change according to the irrigation treatment along the grapevine vegetative growth cycle (Figure 3-4). Non-irrigated grapevines tend to present more oxidated xanthophylls and lower concentrations of chlorophylls, suggesting beyond water stress, heating, and lighting stresses (Figure 3-4). Also, xanthophyll oxidation along the dates suggests grapevine leaf senescence (Figure 3-4).

The good results performed with SL-AI allow complete information about the relationship between the spectral response and cause-effect in grapevines, distinguishing positively from the other data-driven techniques presented in the benchmarking (Figure 3-4). Compared to other works presented in the literature review, our dataset is very heterogenous from a genetic and environmental point of view. Even though SL-AI minimised the multi-scales interferences and corrected the matrix effect in the hyperspectral data and predicted with high precision Ψ_{pd} . The information processed with SL-AI allowed a precise estimation of Ψ_{pd} due to the discrimination of the different bio compounds present in the leaf obtained through the VIS-NIR zone of the spectrum, allowing an interpretation of the abiotic stresses that affect the grapevines. More important to quantify Ψ_{pd} , SL-AI allows identifying other abiotic stress expressed in the VIS-NIR data (Figure 3-4), promoting an interpretation of cause-effect in the plant.

The results presented in this work with SL-AI show a time-dynamic interaction along the grapevine vegetative growth cycle, linking the loadings of PC 1 for the established ROIs (chlorophylls, xanthophylls and water absorption) to the identification of lighting, heating, and water stress during the grapevine growth cycle. Based on the information presented in Figure 3-4, it is possible to notice the leaf senescence through the zone of the xanthophyll absorption and the differences according to the irrigation regime (from DOY 37 to DOY 107). Non-irrigated leaves tend to present more concentration of oxidated xanthophylls to deal with the high heating and less available water (Middleton et al., 2012; Chaves et al., 2010). The findings of this work are supported by Tosin et al. (2021) when they reported the increase of the lutein content (carotene group) along the vegetative cycle. Parallely, the loadings of PC 1 in the water absorption ROI is always related to the irrigated grapevines due to the higher concentration of water (Figure 3-4) and the high correlation with the NIR zone of the spectrum (Peñuelas & Filella, 1998). The chlorophyll content, in various studies, was associated with increased water concentration (Zarco-Tejada et al., 2013; Lichtenthaler & Babani, 2004; Sims & Gamon, 2003), showed a positive association with irrigated grapevines in some of the cases (Figure 3-4). Chlorophylls are also an indicator of leaf senescence, decreasing the concentration in older leaves (Tosin et al., 2021; Lichtenthaler & Babani, 2004). Despite this work has not quantified pigments, this relationship present in the literature was not observed in our results.

The results presented in this paper, supported by works that performed pigment quantification, showed that the bands in the VIS-NIR domains associated with chlorophyll, xanthophyll, and water change along the grapevine cycle, suggesting lighting, heating, and water stress in the grapevines (Pallioti & Poni, 2015; Chaves et al., 2010). Thus, SL-AI estimated Ψ_{pd} with causality be used in different parts of the globe under diverse environmental conditions.

3.5. Conclusion

This study shows the feasibility of Ψ_{pd} quantification using VIS-NIR spectroscopy. SL-AI could predict Ψ_{pd} with very high precision in a heterogeneous dataset composed of different cultivars, test sites, years of sampling and cultural practices. Despite this study being limited to only three years, the robust results show the potential generalisation of this approach and capacity in providing a physiological cause-effect diagnosis, linking absorbance bands and plant hydric stress.

Ψ_{pd} estimation through SL-AI is directly related to the chlorophyll, xanthophyll, and water absorbance bands, which support cause-effect relationships extraction from the dataset. We further demonstrated that it is possible to associate lighting, heating, and water stresses in grapevines to the corresponding absorbance bands, leading to leaf senescence under different environmental conditions.

Further applications of SL-AI in viticulture can be: i) quantification of pigments through ground truth methods to re-validate the results in a qualitative approach and ii) quantification of pigments and measurements of Ψ_{pd} data to provide an *in-vivo* causality.

Acknowledgments

Renan Tosin acknowledges the “Fundação para a Ciência e a Tecnologia (FCT)”, Portugal for the PhD grant (SFRH/BD/145182/2019). Rui Martins acknowledges “Fundação para a Ciência e Tecnologia (FCT)” for the research contract grant (CEEIND/017801/2018). The authors also thank the wine company Real Companhia Velha (and its Coordinator for Viticulture Rui Soares) and the wine company Symington Family Estates (and its R&D Viticulture Manager Fernando Alves) for the facilities provided for fieldwork, as well as the Associação para o Desenvolvimento da Viticultura Duriense (ADVID) for funding part of the field-work missions. This study was implemented under a cooperation protocol integrating Faculty of Sciences, University of Porto, ADVID, Real Companhia Velha, and Symington Family Estates.

CRedit author statement

Renan Tosin: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Writing – original draft preparation, Writing – review and editing. **Rui Martins:** Conceptualisation,

Methodology, Software, Formal analysis, Investigation, Resources, Writing – original draft preparation, Writing – review and editing, Supervision. **Isabel Pôças:** Investigation, Writing – review and editing. **Mario Cunha:** Conceptualisation, Investigation, Writing – original draft preparation, Writing – review and editing, Supervision.

Supplementary material

Figures

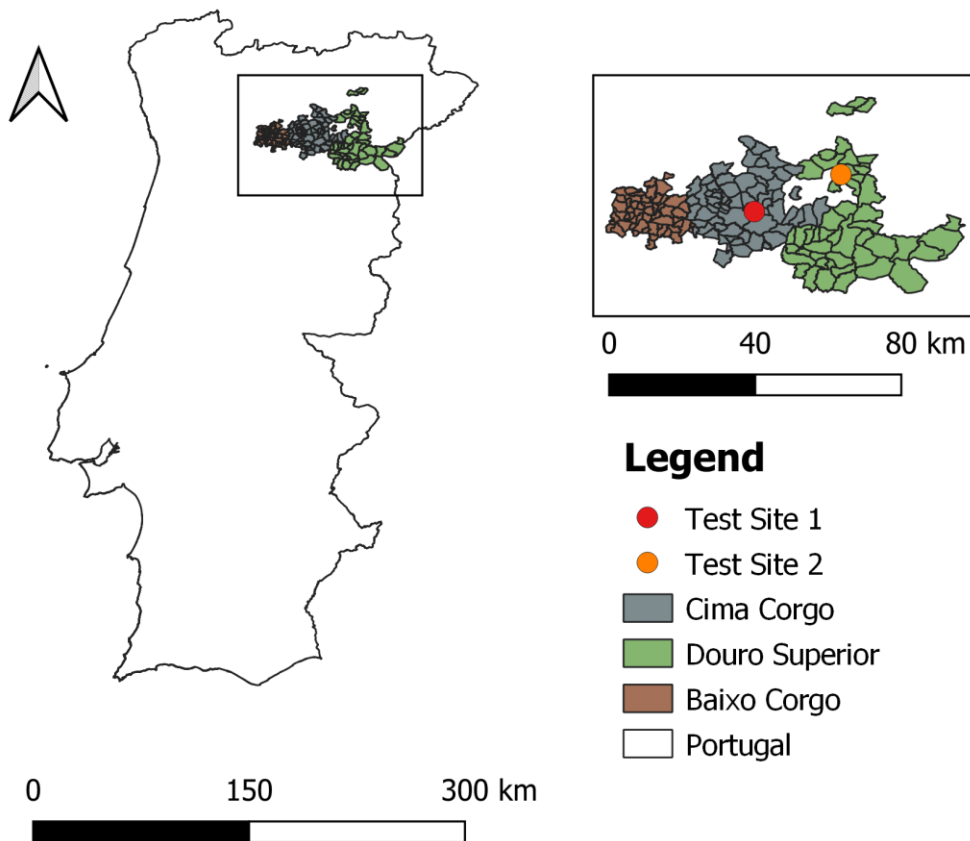


Figure S 3-1. The study area's location identifies the Test Sites 1 and 2 in the Douro Wine Region, Northeast Portugal, and respective the sub-regions: Baixo Corgo, Cima Corgo and Douro Superior.

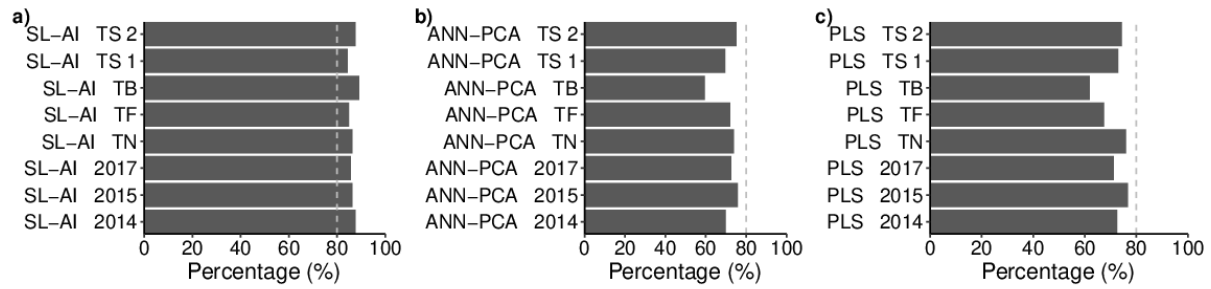


Figure S 3-2. The overall accuracy of the predicted predawn leaf water potential (Ψ_{pd}) based on the three classes of Ψ_{pd} used by Pôças et al. (2020). "a", "b" and "c" are the results performed with Self-learning artificial intelligence (SL-AI), Artificial Neural Network with Principal Component Analysis (ANN-PCA) and Partial Least Square (PLS) respectively, according to the year (2014, 2015, 2017), cultivar (TB: Tinta Barroca, TF: Touriga Franca and TN: Touriga Nacional) and test site (TS 1- Quinta dos Aciprestes, TS 2 – Quinta do Ataíde).

Tables

Table S 3-1. The number of grapevines sampled per test site, cultivar, and irrigation conditions.

	2014		2015		2017		Total	
	I	NR	I	NR	I	NR	I	NR
Teste site S1								
TN	74	39	50	23	64	31	188	93
TF	-	-	31	0	40	-	71	-
Test site S2								
TB	-	-	24	-	18	-	42	-
TF	-	-	24	23	18	18	42	41
TN	-	-	94	100	88	88	182	188
Total	74	39	223	146	228	137	525	322

TN: Touriga Nacional, TF: Touriga Franca, TS2: TB: Tinta Barroca, I: irrigation, NI: no irrigation.

Table S 3-2. Dates of predawn leaf water potential measurement and number of each cultivar assessed in each test site in the sampling dates.

Test site 1			Test site 2			
Measurement date	TN	TF	Measurement date	TN	TF	TB
2014						
June 16	24	-	-	-	-	-
July 10	19	-	-	-	-	-
July 26	22	-	-	-	-	-
August 19	24	-	-	-	-	-
September 9	24	-	-	-	-	-
2015						
June 25	11	9	June 26	48	12	6
July 16	20	7	July 15	50	11	6
August 6	20	6	August 7	51	12	6
September 3	22	9	September 2	45	12	6
2017						
July 5	24	10	July 4	40	-	-
July 20	24	10	July 21	52	12	6
August 3	23	10	August 4	52	12	6
August 31	24	10	September 1	32	12	6

TN: Touriga Nacional, TF: Touriga Franca, TB: Tinta Barroca.

Table S 3-3. Irrigation events and irrigation amounts (L/Plant/day) per test site and irrigation treatment (adapted from Pôças et al. (2020)).

	Irrigation events			Irrigation amount (L/plant/Day)		
	2014	2015	2017	2014	2015	2017
TS 1						
TN_IT1	3	7	6	32	62	92
TN_IT2	4	9	8	36	130	124
TF_IT	-	7	6	-	62	92
TS 2						
TN1_IT	-	5	10	-	95	157
TN2_IT	-	5	10	-	95	157
TF_IT	-	5	8	-	95	125
TB_IT	-	5	9	-	95	154

Test Site 1 (TS 1): TN_IT1– Touriga Nacional – irrigation treatment 1; TN_IT2– Touriga Nacional – irrigation treatment 2; TF_IT – Touriga Franca – irrigation treatment. Test Site 2 (TS 2): Test site 2: TN1_IT –Touriga Nacional, plot 1 – irrigation treatment; TN2_IT – Touriga Nacional, plot 2 – irrigation treatment; TF_IT – Touriga Franca – irrigation treatment; TB_IT – Tinta Barroca– irrigation treatment.

Table S 3-4. Models tested with Self-learning Artificial Intelligence (SL-AI), Artificial Neural Networks with Principal Component Analysis (ANN-PCA) and Partial Least Square Regression (PLS) according to the training, validation, full dataset, year (2014, 2015 and 2017), Cultivar (TN: Touriga Nacional; TF: Touriga Franca; TB: Tinta Barroca) and Test Site (TS 1: Test Site 1 - Quinta dos Aciprestes); TS 2: Test Site 2 - Quinta do Ataíde).

	SL-AI					ANN-PCA					PLS				
	N	R ²	RMSE (MPa)	MAPE (%)	PRESS	N	R ²	RMSE (MPa)	MAPE (%)	PRESS	N	R ²	RMSE (MPa)	MAPE (%)	PRESS
Training	Na	Na	Na	Na	Na	595	0.96	0.13	23.28	10.29	594	0.93	0.15	30.28	14.11
Validation	Na	Na	Na	Na	Na	252	0.85	0.25	43.97	38.32	253	0.92	0.13	33.89	4.44
Full data	699*	0.97	0.10	18.30	7.53	847	0.85	0.25	43.64	50.70	847	0.94	0.16	28.76	22.41
Year															
2014	97*	0.96	0.09	25.34	0.81	113	0.84	0.20	49.61	4.26	113	0.91	0.15	36.63	2.35
2015	309*	0.98	0.10	17.25	2.87	369	0.83	0.26	46.43	23.75	369	0.94	0.15	27.52	8.66
2017	293*	0.97	0.11	17.78	3.84	365	0.87	0.25	39.69	22.57	365	0.94	0.18	28.13	11.36
Cultivar															
TN	536*	0.97	0.11	18.41	5.98	651	0.85	0.25	44.26	41.78	651	0.94	0.16	28.16	17.05
TF	126*	0.97	0.10	18.67	1.35	154	0.87	0.22	41.61	7.23	154	0.93	0.16	29.96	4.05
TB	37*	0.98	0.08	14.56	0.22	42	0.89	0.20	39.10	1.66	42	0.90	0.18	34.28	1.43
Test Site															
TS 1	295*	0.97	0.11	21.09	3.42	352	0.87	0.22	41.94	16.96	352	0.93	0.16	31.14	9.49
TS 2	404*	0.98	0.10	16.58	4.13	495	0.85	0.26	44.45	33.85	495	0.94	0.16	27.26	12.98

N: Number of observations; Na: Data non-available, *number of observations according to the clusters generated through SL-AI.

CHAPTER 4

- CASE STUDY II: LIBS FOR NUTRIENT IDENTIFICATION IN SMALL BIOLOGICAL TISSUES
- CASE STUDY III: QUANTITATIVE VIS-NIR FOR NUTRIENT IDENTIFICATION IN APPLE AND GRAPE LEAVES

4.1. Case study II: LIBS for nutrient identification in small biological tissues.

Tosin, R., Monteiro-Silva, F., Martins, R., and Cunha, M. (2024). A New Approach for Element Characterization of Grapevine Tissue with Laser-Induced Breakdown Spectroscopy. *Horticulturae* 10, 82. doi: doi.org/10.3390/horticulturae10010082

A New Approach for Element Characterisation of Grapevine Tissue with Laser-Induced Breakdown Spectroscopy

Renan Tosin^{1,2}, Filipe Monteiro-Silva², Rui Martins^{2,*} and Mario Cunha^{1,2,*}

¹ Department of Geosciences, Environment and Spatial Planning, Faculty of Sciences of the University of Porto, Rua do Campo Alegre, S/N, 4169-007 Porto, Portugal; renan.tosin@fc.up.pt

² INESC TEC—Institute for Systems and Computer Engineering, Technology and Science, Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, S/N, 4200-465 Porto, Portugal; filipe.m.silva@inesctec.pt

*Correspondence: rui.c.martins@inesctec.pt (R.M.); mccunha@fc.up.pt (M.C.)

Abstract

The determination of grape quality parameters is intricately linked to the mineral composition of the fruit; this relationship is increasingly affected by the impacts of climate change. The conventional chemical methodologies employed for the mineral quantification of grape tissues are expensive and impracticable for widespread commercial applications. This paper utilised Laser-Induced Breakdown Spectroscopy (LIBS) to analyse the mineral constituents within the skin, pulp, and seeds of two distinct *Vitis vinifera* cultivars: a white cultivar (Loureiro) and a red cultivar (Vinhão). The primary objective was to discriminate the potential variations in the calcium (Ca), magnesium (Mg), and nitrogen (N) concentrations and water content among different grape tissues, explaining their consequential impact on the metabolic constitution of the grapes and, by extension, their influence on various quality parameters. Additionally, the study compared the mineral contents of the white and red grape cultivars across three distinct time points post veraison. Significant differences ($p < 0.05$) were observed between the Loureiro and Vinhão cultivars in Ca concentrations across all the dates and tissues and for Mg in the skin and pulp, N in the pulp and seeds, and water content in the skin and pulp. In the Vinhão cultivar, Ca differences were found in the pulp across the dates, N in the seeds, and water content in the skin, pulp, and seeds. Comparing the cultivars within tissues, Ca exhibited differences in the pulp, Mg in the skin and pulp, N in the pulp and seeds, and water content in the skin, pulp, and seeds. These findings provide insights into the relationship between the grape mineral and water content, climatic factors, and viticulture practices within a changing climate.

Keywords

precision viticulture; maturation; nutrients; plant sensing

4.1.1. Introduction

The taste and overall quality of grapes are influenced by their chemical composition (Rogiers et al., 2006). The essential macronutrients, namely calcium (Ca), magnesium (Mg), and nitrogen (N), contribute to the growth, development, and metabolic processes of grapevines (Tewari et al., 2021). Ca contributes to membrane stability, enhances disease resistance, and supports vital metabolic pathways, such as Ascorbate-Glutathione, contributing to the overall health of grapevines (Tewari et al., 2021; Özcan et al., 2017). Mg, which is essential for photosynthesis and enzyme activation, promotes the general health of grapevines (Rogiers et al., 2006). Concurrently, the N profiles proteins that optimise the grapevine canopy's growth, development, and vitality (Tewari et al., 2021). The allocation of these components within grape tissues impacts the final yield and the overall quality of the harvested fruit (Bertoldi et al., 2011).

Environmental, genetic, and agricultural factors collectively influence the plant nutrient transport mechanisms, impacting ion and water uptake efficiency and mineral assimilation (Verdenal et al., 2021; Guo et al., 2016). Beyond the factors shaping grape composition and taste, the dynamic relationship of these external elements is consequential in the context of climate change and viticulture practices. Understanding the elemental composition of grapes is essential for optimising cultivation practices, encompassing factors like irrigation and fertilization, and ensuring the long-term sustainability of vineyards. Climate change-induced water availability shifts directly affect the grape composition (Candar et al., 2021; Buesa et al., 2017). Moreover, pollutants present in the environment (Bora et al., 2020) and soil degradation attributed to heavy metals (Singh, 2021) can manifest in various grape organs, potentially interfering with the overall composition.

While the existing research has underscored the importance of Ca (Potortiota et al., 2017), Mg (Guo et al., 2016), and N (Verdenal et al., 2021), there is a lack of studies exploring the variations in their distribution across different grape tissues like the skin, pulp, and seeds, during maturation. This paper explores the impact of elemental dynamics on grape maturation and quality parameters, which can benefit grape growers and winemakers.

The elemental quantification of grape tissues has relied on traditional methods such as acid digestion coupled with atomic absorption spectrophotometry (AAS) or inductively coupled plasma optical emission spectrometry (ICP-OES) (Senesi et al., 2019).

However, these conventional techniques present inherent constraints that limit their applicability for routine analyses in the viticultural sector. The sample preparation requirements, time-intensive procedures, and dependence on specialised equipment, reagents, and skilled operators pose challenges, particularly in large-scale vineyards (Silva et al., 2023b; Singh, 2021). By recognizing these limitations, alternative non-destructive methods have been developed to address these challenges. For example, X-ray fluorescence has demonstrated efficacy in characterising the element composition of biological samples (Andrade & Pereira-Filho, 2016). However, the limitations in quantifying the elements with atomic numbers (Z) higher than 12 and the need to consider radiation effects during analysis necessitate the careful evaluation of these methods. Also, Vis-NIR has been used to quantify the nutrients (Silva et al., 2023a; Monteiro et al., 2019).

Laser-Induced Breakdown Spectroscopy (LIBS) has emerged as a promising alternative method for elemental analysis in agriculture and food science (Andrews et al., 2023b). This rapid and destructive analytical approach involves using a neodymium-doped yttrium aluminium garnet laser to generate pulses, ablating and exciting the targeted object, such as the grape tissue, and producing a plasma plume. The optical radiation emitted from this plasma is then captured and examined to interpret the sample's elemental composition. Despite the weakness of producing limited molecular structure information, LIBS presents distinct advantages over the traditional techniques. In contrast to the costly conventional methods, LIBS has minimal sample preparation requirements, a high spatial resolution, and reduced operational costs (Senesi et al., 2019), making it an attractive choice for the elemental tissue analysis of small samples of grapes (Tosin et al., 2023b).

The literature reports portable LIBS that can be used for element quantification (Arantes de Carvalho et al., 2018; Rakovský et al., 2014). This study, building upon the groundwork laid by Tosin et al. (2023c), seeks to leverage the advantages of LIBS in characterizing the elemental composition of grapes, explicitly focusing on calcium (Ca), magnesium (Mg), and nitrogen (N) in the skin, pulp, and seeds.

Integrating the relationships systematically, quantitatively, and causally between the soil, climate, and plant physiology is imperative. The utilisation of LIBS-based nutrition diagnosis has emerged as a tool, providing reinforced precision in discerning the grapevine nutritional requirements (Andrade & Pereira-Filho, 2016), which helps to improve the grape quality and facilitates wise and precise management practices, thereby advocating for more sustainable viticultural methods in response to dynamic climate change scenarios (Miras-Avalos & Intrigliolo, 2017). While the focal point of this

research centres on precision viticulture, these discernments and technological developments hold the potential for broader applications encompassing diverse fruit crops.

This study aims to address and overcome the limitations identified in the existing research concerning the application of LIBS in the analysis of plant tissues, with a specific focus on vineyards. Constraints, such as the influence of water content on measurements, the necessity to handle sample volumes, and the requirement for sample preparation pellets, have been noted in prior studies, as identified by Senesi et al. (2019). Furthermore, the challenges have been acknowledged in quantifying the elements within dried tissues, along with considerations regarding the influence of nitrogen in the atmosphere (Ribeiro et al., 2021; Moros & Laserna, 2019). These challenges posed barriers to the implementation of LIBS in viticultural backgrounds.

This work tests and evaluates the possible resolution of these challenges, contributing to the advancement and refinement of LIBS application in plant tissue analysis, particularly within viticulture. Thus, this paper has two main objectives: (i) to extract elemental samples (Ca, Mg, and N) from the grape tissues, specifically the skin, pulp, and seeds; (ii) to conduct comparative analyses between the two grape varieties and across different dates after veraison. The specific goals include utilising the same volume (0.098 cm³) for the skin, pulp, and seeds to identify the elemental constituents and water content.

4.1.2. Materials and Methods

4.1.2.1. Test Site

This research was conducted within the Região dos Vinhos Verdes located in the northwestern region of Portugal, Vila do Conde, specifically at the Campus Agrário de Vairão (41°19'31.91" N; 8°40'27.45" W). The area is defined by a Mediterranean climate, presenting an average annual temperature of 14.9 °C and an average yearly precipitation of 1149 mm (Climate-Data.org). The soil composition is predominantly from shales of the shale–greywacke complex.

The grapes were randomly sampled in the year 2020 on three dates after the veraison (14/08/2020) of two grape cultivars: Loureiro (Vitis International Variety Catalogue (VIVC) 25085), characterised as a white grape cultivar; and Vinhão (VIVC 13100), classified as a red grape cultivar (Table 4.1-1). On each sampling date, the following considerations were made, aiming to minimise the potential sources of variability: (i) prioritising grape bunches that were representative and uniformly sized and matured within the same cultivar; (ii) considering the position of grape bunches on the vine to

mitigate variations stemming from differing sunlight exposure, nutrient availability, and microclimatic conditions; and (iii) incorporating the accumulation of biomass and grape bunch size in the selection process to ensure that the chosen samples were at similar stages of development.

Table 4.1-1. Sampling dates of the Loureiro and Vinhão cultivars.

Date	Days After Veraison (DAV)	Samples (n)	
		Loureiro	Vinhão
S1: 30/09/2020	47	20	16
S2: 12/10/2020	59	9	8
S3: 14/10/2020	61	10	5

4.1.2.2. Characterisation of the Grapes

Using a scalpel, the separation of grape tissues into skin, pulp, and seeds was meticulously executed. After weighing, the tissues dried in an oven set at 45 °C for four days. Following this drying period, re-weighing the dried tissues allowed the water content (%) calculation based on the observed weight differential.

The dried tissues were individually macerated using a grinder, and from these processed tissues, a pellet with an approximate volume of 0.098 cm³ was assembled (Figure 4.1-1). This prepared palette served as the sample for subsequent analysis utilising LIBS.

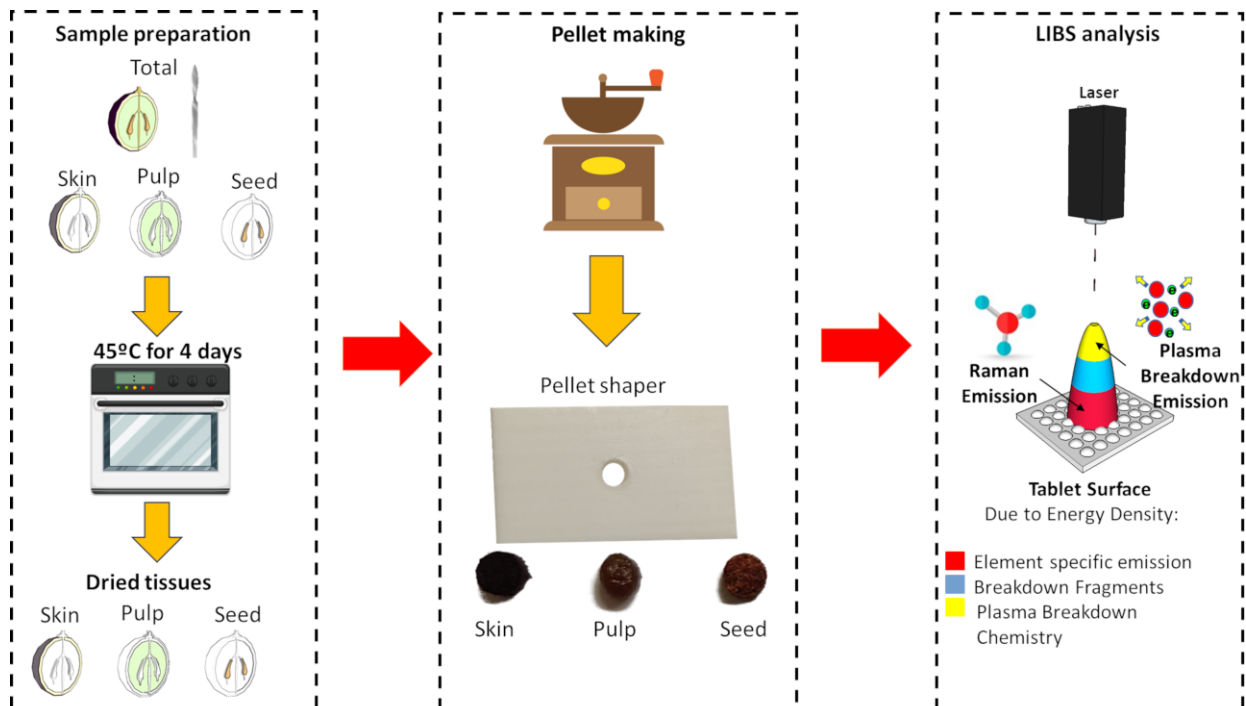


Figure 4.1-1. Sample preparation, pellet making, and LIBS analysis of the skin, pulp, and seeds of the Loureiro and Vinhão cultivars.

2.3. LIBS Instrumentation

The in-house-assembled system utilized a Q-Switched Nd:YAG (neodymium-doped yttrium aluminium garnet) laser operating at a fundamental wavelength of 1064 nm, featuring an 8 ns pulse duration and a maximum adjustable energy of 200 mJ (Martins et al., 2018). The arrangement included lenses and mirrors for the laser pulse focusing on the sample surface, complemented by two crossed visible diode laser pointers aligned with the primary laser's focal point. This alignment facilitated positioning the sample's region of interest at the optimal height and X–Y coordinates (Figure 4.1-2).

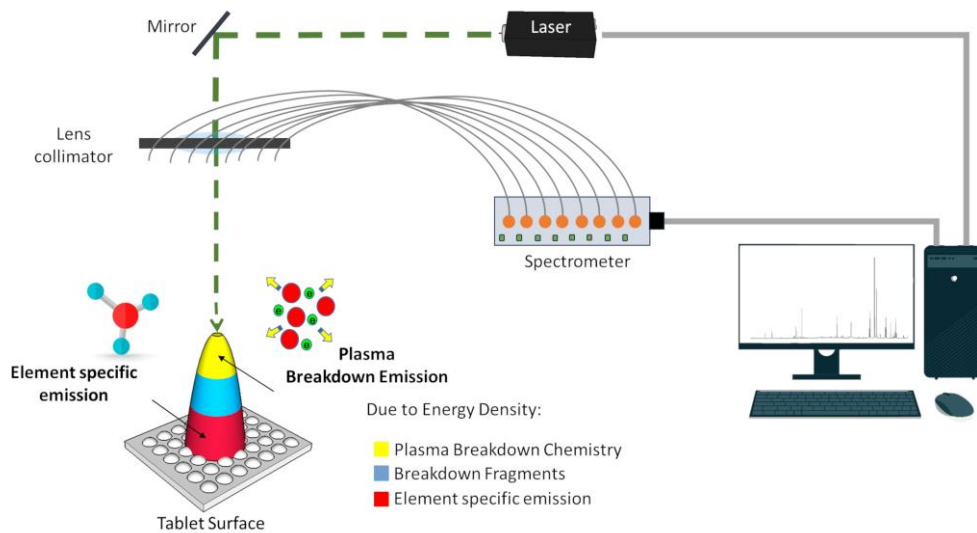


Figure 4.1-2. Experimental setup for Laser-Induced Breakdown Spectroscopy (LIBS), showcasing laser incidence, spectral data acquisition, and manual adjustments for focus. Adapted from Ribeiro et al. (2021).

The sample holder relied on three stages, with control over X, Y, and Z via manual adjustment directions using precision knobs. The system comprises eight optical fibres and collimators aligned with the system's focal point, collecting and directing plasma light into an eight-channel CCD spectrometer (Avantes, Apeldoorn, The Netherlands). This spectrometer enabled spectral detection within the wavelength range of approximately 180 to 920 nm, boasting a resolution between 0.06 nm and 0.18 nm. Three spectra were acquired for each sample, and the resulting averaged spectrum was considered for analysis.

All the tests were conducted in ambient air, and the distance between the sample and the collecting optics varied randomly, aiming to cover the entire longitudinal extent of the sample during scanning. A detailed overview of this specific LIBS setup can be found in the description provided by Ribeiro et al. (2021).

2.4. Element Identification

This paper adopted a qualitative methodology to discern the presence of calcium (Ca), magnesium (Mg), and nitrogen (N) in grape skin, pulp, and seeds. The primary goal of LIBS lies in deducing the chemical composition of samples through the analysis of their spectra. Each chemical element exhibits distinct wavelengths of light, giving rise to well-defined lines within the spectrum. However, instances may arise where the emission wavelengths of different elements overlap. The resulting LIBS spectra had a high resolution of approximately 0.1 nm, and the number of variables for this configuration was 16361. As a result of the large amount of data, principal component analysis (PCA) was applied to study the main differences among the elements assessed between the tissues (skin, pulp, and seeds) and the two grape cultivars.

Two reference databases, the OSCAR (OSCAR, 2020) and NIST databases (NIST, 2020), were evaluated regarding their suitability for interpreting the samples' spectra and building a database. Both databases were employed for elemental computation from the periodic table, with certain elements exclusively identified in one of the databases. In the whole analysis, a total of 37 elements from the periodic table were computed. However, only Ca, Mg, and N were identified in the assessed samples.

The spectral data obtained from LIBS were compared with the spectra available in the OSCAR (OSCAR, 2020) and NIST (NIST, 2020) databases for matches or similarities between the peaks or patterns. The elements (Ca, Mg, and N) present in each sample were identified by matching the peaks or patterns from the LIBS spectra with the corresponding entries in the built database.

2.5. Statistical Analysis

Repeated measures analysis of variance (RM-ANOVA), employed to account for non-identical samples, was complemented by a post hoc test using Duncan's method to assess the variations in water content between the two grape cultivars across the skin, pulp, and seeds on the three evaluation dates. Furthermore, RM-ANOVA, followed by the post hoc test, was conducted to discern significant statistical differences in Ca, Mg, and N characterisation within the skin, pulp, and seeds of both grape cultivars across the three assessment dates. This study considered both the non-significant ($p > 0.05$) and significant ($p < 0.05$) outcomes for comprehensive examination and interpretation.

4.1.3. Results

Figure 4.1-3 presents the average signal intensity relating to the skin, pulp, and seeds of the Loureiro and Vinhão cultivars in S1 (47 days after veraison). The diagram

incorporates the respective wavelengths corresponding to each element identified according to the generated database. Ca is characterized by five peaks at 317.936 nm, 617.128 nm, 646.273 nm, 649.378 nm, and 866.097 nm. Mg exhibits three peaks at 279.535 nm, 280.24 nm, and 285.18 nm; N shows six peaks at 744.2 nm, 746.753 nm, 821.533 nm, 822.248 nm, 824.184 nm, and 867.938 nm.

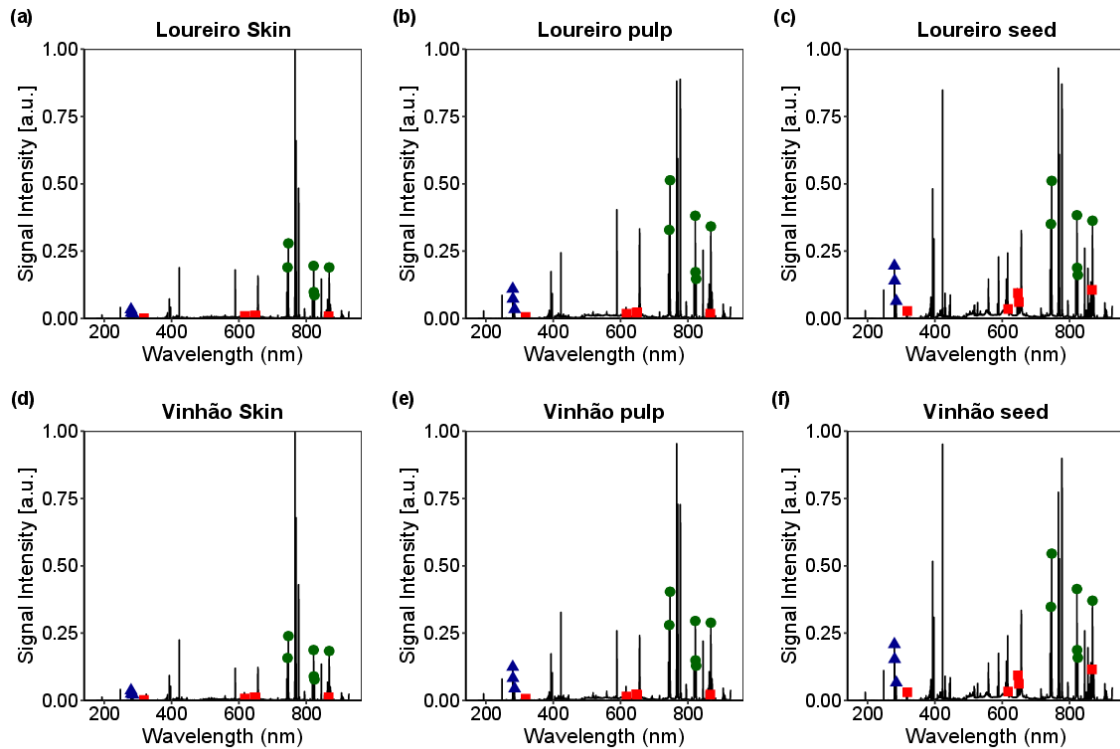


Figure 4.1-3. The average signal intensities of the Ca (■), Mg (▲), and N (●) in the skin, pulp, and seed tissues of the Loureiro cultivar (a–c) and the Vinhão cultivar (d–f) 47 days after veraison (S1). The sample size for each group is specified as $n = 20$ for Loureiro and $n = 16$ for Vinhão.

Table 4.1-2 presents statistical differences in the Ca, Mg, and N concentrations across the skin, pulp, and seed tissues of the two grape cultivars during the three assessment dates. Regarding the Ca concentrations within these tissues, the Loureiro cultivar displays significant variations across all three assessment dates for all the tissues. The higher Ca levels in Loureiro compared to that of Vinhão at S3 may be attributed to overripening, which is known to elevate the Ca concentration in the tissues (Kurt et al., 2017). In the white grapes, as they ripen and overripen, cell wall breakdown occurs, releasing or redistributing Ca ions from the cell walls, potentially influencing the measured Ca content (Hocking, 2015).

Table 4.1-2. The average LIBS spectra intensity peaks of Ca, Mg, and N and water content (% w/w) for each tissue on the three assessment dates (S1, S2, and S3) for the Loureiro and Vinhão cultivars.

Loureiro		S1 (n = 20)	S2 (n = 9)	S3 (n = 10)	p-Value
Element					
Ca	Skin	3258.04 b ± 901.34	3713.58 ab ± 938.70	4189.18 a ± 844.82	0.03
	Pulp	635.30 b ± 120.46 *	588.21 b ± 94.34 *	2003.24 a ± 490.12 *	<0.001
	Seed	354.30 b ± 76.48 *	299.02 b ± 23.71 *	521.00 a ± 128.50	<0.001
Mg	Skin	6699.93 b ± 2180.83	7061.50 b ± 2172.96	10581.13 a ± 1819.63 *	<0.001
	Pulp	2786.03 b ± 933.15 *	2630.43 b ± 627.41 *	5052.57 a ± 701.98	<0.001
	Seed	1035.31 ab ± 297.23 *	976.59 b ± 151.11 *	1223.95 a ± 287.39	0.11
N	Skin	16368.47 a ± 2994.20	17277.07 a ± 1290.88	18,034.47 a ± 2490.27	0.247
	Pulp	11790.15 b ± 2080.38	12896.65 b ± 1892.37 *	14984.23 a ± 1672.66 *	0.001
	Seed	7560.83 a ± 1690.68	6251.21 b ± 1448.79	8009.86 a ± 919.62 *	0.04
Water					
	Skin	77.01 b ± 0.94	73.45 c ± 1.03	79.06 a ± 1.74 *	<0.001
	Pulp	83.41 a ± 2.86 *	79.11 b ± 3.65	81.25 ab ± 2.03 *	0.002
	Seed	56.28 a ± 4.04 *	53.13 a ± 3.54	54.41 a ± 2.80 *	2.53
Vinhão		S1 (n = 16)	S2 (n = 8)	S3 (n = 5)	p-Value
Element					
Ca	Skin	3481.31 a ± 1143.97	3415.04 a ± 942.08	3715.16 a ± 750.24	0.873
	Pulp	1069.07 a ± 321.37 *	703.63 b ± 65.40 *	1298.10 a ± 340.52 *	0.002
	Seed	403.60 a ± 51.57 *	380.63 a ± 91.76 *	402.44 a ± 34.14	0.691
Mg	Skin	7211.87 a ± 2161.640	6926.83 a ± 2940.079	8197.33 a ± 2226.33 *	0.639
	Pulp	4781.77 a ± 961.56 *	4107.27 a ± 556.89 *	4433.30 a ± 414.37	0.165
	Seed	1256.39 ab ± 224.45 *	1191.98 b ± 213.88 *	1464.43 a ± 107.59	0.08
N	Skin	17384.43 a ± 1967.02	16,611.39 a ± 1582.07	17,820.58 a ± 1495.51	0.461
	Pulp	14196.74 ab ± 1965.16 *	14769.43 a ± 1488.09 *	12529.07 b ± 1476.33 *	0.1
	Seed	7305.67 a ± 831.55	6276.87 b ± 695.47	6375.35 b ± 1241.16 *	0.02
Water					
	Skin	76.27 a ± 1.72	73.34 b ± 2.62	71.48 b ± 5.18 *	0.004
	Pulp	81.19 a ± 2.47 *	80.99 a ± 4.87	71.86 b ± 8.57 *	0.002
	Seed	48.27 b ± 3.45 *	52.70 a ± 2.42	51.09 ab ± 2.41 *	0.007

Being on the same line (among the dates) and labelled with different letters means the results are significantly different according to Duncan's post hoc test ($\alpha < 5\%$). The comparison between the cultivars for the same variables and on the same days after veraison was considered ($p < 0.05$). An asterisk identifies the statistical differences between the cultivars.

On the contrary, the Vinhão cultivar only diverges in the pulp tissue. The comparative analysis of the two cultivars reveals uniform Ca concentrations in the skin and consistent concentrations in the seeds during S3.

Regarding the Mg content, the Loureiro cultivar displays fluctuations in the skin and pulp tissues across the three assessment dates. Conversely, the Vinhão cultivar maintains consistent Mg concentrations in internal tissues throughout the three assessment dates. Cross-cultivar analysis reveals variations in the Mg content in the skin tissue of the Loureiro cultivar during S3, as well as in the pulp and seed tissues during S1 and S2. Concerning N content analysis, the Loureiro cultivar lacks significant differences in the skin tissue across the three assessment dates. In contrast, the Vinhão cultivar demonstrates statistical distinctions only in the seed tissue. The assessment of N concentrations between the two cultivars indicates disparities in the pulp tissue for both the grape cultivars across the three assessment dates and in the seed tissue during S3. Regarding the water content, there are alterations in the skin and pulp tissues of the Loureiro cultivar, while the seed tissue remains unaffected. On the contrary, the Vinhão

cultivar experiences fluctuations in water content across all the three tissues. The variations in water content for both the cultivars were observed in the skin tissue during S3 and in the pulp and seed tissues during S1 and S3. Additionally, the impact of water content on the studied macro elements appears to be more pronounced in the Loureiro cultivar than that in the Vinhão cultivar.

Figure 4.1-4 shows the Ca, Mg, and N alterations across the Loureiro cultivar's skin, pulp, and seeds during the three post-veraison dates. These dynamic changes were observed throughout the assessment period. The representations of principal component 1 (PC1) and PC2 in Figure 4.1-4 capture more than 90% of the dataset's variance.

The rationale behind employing PCA in this study is to demonstrate multivariate information from the elemental concentrations in the grape tissues and across the dates into a concise set of principal components (PC1 and PC2). These two components explain most of the variance within the data, offering a more streamlined representation. By transforming the data into a lower-dimensional space, PCA aids in identifying the variations and similarities among the samples. The integration of PCA with the results presented in Table 4.1-2 enriches the understanding of elemental dynamics during grape maturation. This approach allows to visualise how different grape tissues evolve regarding elemental composition across the assessed sampling dates and improves our insights into the clustering or dispersion of samples in a reduced dimensional space, facilitating the identification of underlying patterns or trends that are not immediately apparent in the tabulated data.

Moreover, the alignment of the PCA results with Table 4.1-2 demonstrates the differences in Ca concentrations across the skin, pulp, and seeds for the three assessment dates, the variations in Mg content within the skin and pulp, and the distinctions in the pulp and seeds concerning N. PCA validates the results presented in Table 4.1-2, providing a robust analytical framework for interpreting the results.

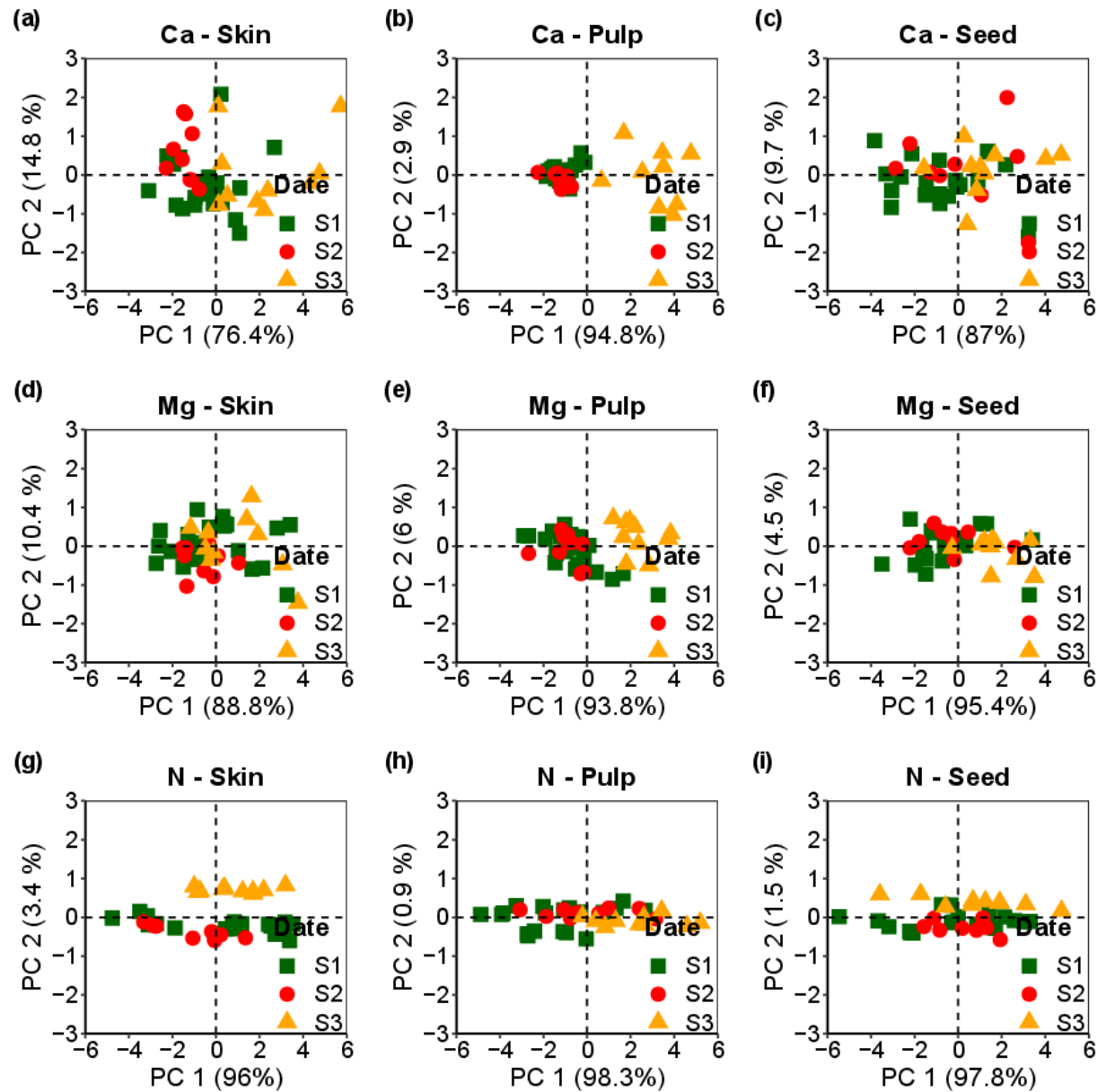


Figure 4.1-4. Principal component analysis of Loureiro cultivar across three assessment dates (S1(■), S2 (●), S3(▲)); a-c represent calcium (Ca) concentrations in skin, pulp, and seeds, while d-f show magnesium (Mg) concentrations, and g-i illustrate nitrogen (N) concentrations in the corresponding tissues.

Figure 4.1-5 shows the variations in Ca, Mg, and N concentrations within the Vinhão cultivar's skin, pulp, and seeds over the three post-veraison dates. Compared to the Loureiro cultivar (Figure 4.1-4), these results highlight the differences in macronutrient dynamics during grape maturation between the two cultivars.

As indicated in Table 4.1-2, the Loureiro cultivar exhibits statistically significant differences in the Ca, Mg, and N concentrations across the assessed tissues. In contrast, the Vinhão cultivar shows significant deviations only in Ca within the pulp and N within

the seeds. These distinctions suggest that the macronutrient dynamics vary during the maturation process, further accentuating the unique responses of the two grape cultivars.

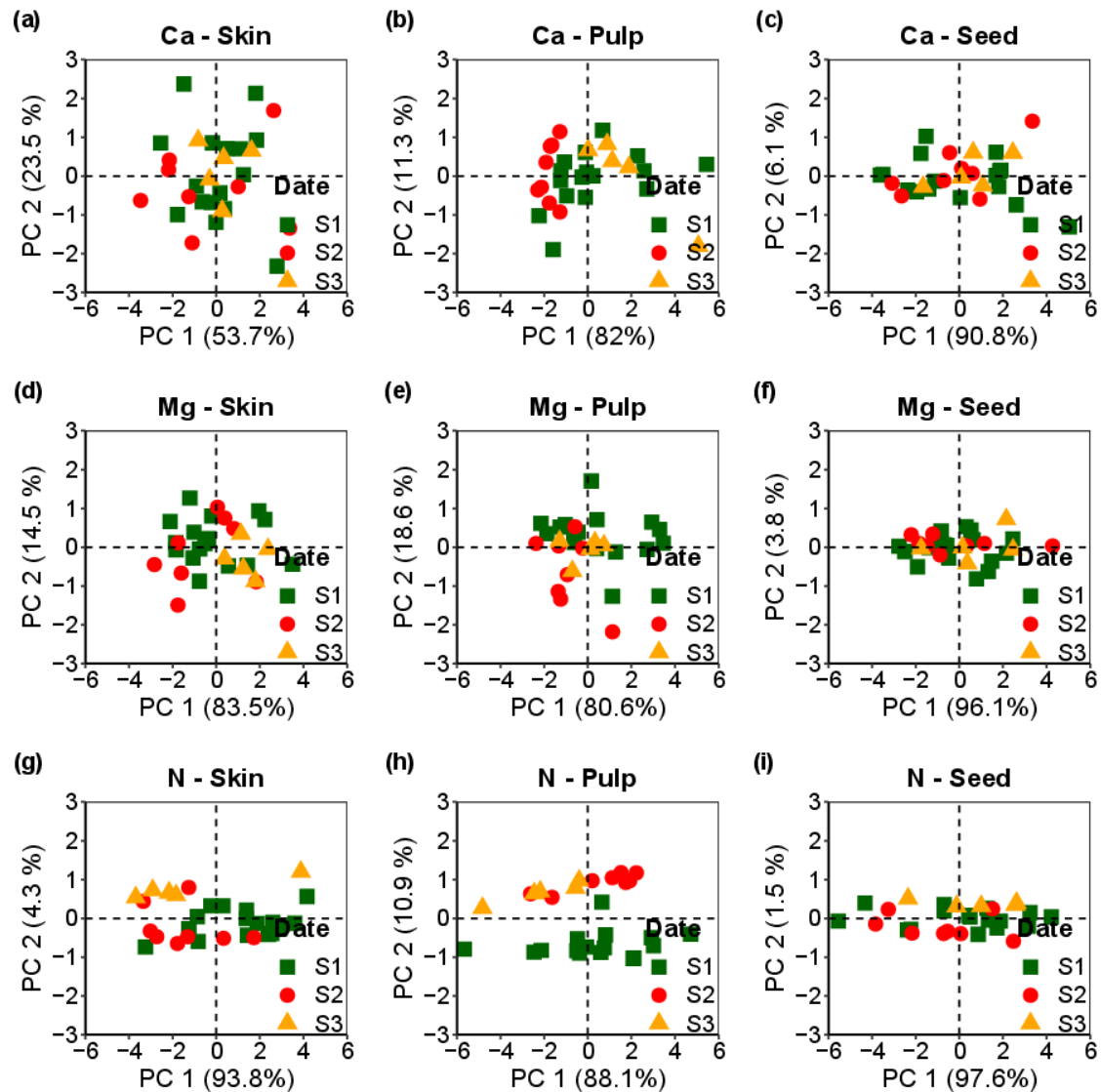


Figure 4.1-5. Principal component analysis of Vinhão cultivar across three assessment dates (S1(■), S2(●), S3(▲)); a-c represent calcium (Ca) concentrations in skin, pulp, and seeds, while d-f show magnesium (Mg) concentrations, and g-i illustrate nitrogen (N) concentrations in the corresponding tissues.

4.1.4. Discussion

Incorporating Laser-Induced Breakdown Spectroscopy (LIBS) into viticulture marks a departure from the traditional, resource-intensive wet laboratory techniques such as ICP-OES for the elemental quantification of grape tissues. In this study, the potential utilisation of LIBS across various grape tissues and cultivars signifies a progressive step toward exploring its applicability and efficacy, particularly in the context of shifting climate patterns. By employing a qualitative framework, this paper enhances the discernment of essential elements, such as calcium (Ca), magnesium (Mg), and nitrogen (N), within the

internal tissues of grapes, specifically the skin, pulp, and seeds, across two distinct grape cultivars.

The relationship between water availability and macronutrient dynamics, particularly in the white cultivars, is a complex aspect of plant physiology and nutrient management (Yuan et al., 2018; Rogiers et al., 2006). Water, serving as a conduit for nutrient uptake, facilitates the absorption of essential macronutrients, like Ca, Mg, and N, by the plant roots (Yuan et al., 2018). The variations in water content influence the elemental constitution of the tissues of grape cultivars, especially in the white cultivars (Martins et al., 2021), as demonstrated in this study (Table 4.1-2). The physical and chemical characteristics of the soil, influenced by the water content, directly impact the solubility and mobility of these macronutrients, influencing their availability to the plant (Ozcan et al., 2017). For example, the Ca levels increase in the grapes under well-hydrated conditions (Martins et al., 2021). The dynamic nature of the rhizosphere, where the root and microbial activities intersect, is susceptible to changes in the water content, further influencing nutrient cycling (Lynch, 2019).

Additionally, the physiological responses of plants to water availability, particularly stress-induced reactions, can affect nutrient uptake and assimilation (Martins et al., 2021; Yuan et al., 2018). The white cultivars may exhibit specific sensitivities and responses to the variations in water availability, emphasizing the importance of adapted irrigation strategies (Buesa et al., 2017). Therefore, the relationship between the water and macronutrients necessitates a better understanding of the optimised agricultural practices (e.g., irrigation), enhanced nutrient management (e.g., fertilisation), and the robust growth and quality of grape cultivars (e.g., pruning and harvest time).

The differences in Ca, Mg, and N concentrations between the Loureiro and Vinhão cultivars (Table 4.1-2) reflect the complex interaction of the genetic, physiological, and environmental factors governing grapevine nutrient dynamics (Martins et al., 2021; Yuan et al., 2018). The differences across the assessed grape cultivars at the three distinct assessment dates emphasise the unique nutrient profiles inherent to each variety. While the Loureiro cultivar is more susceptible to macronutrient changes (Martins et al., 2021) during the maturation of all the tissues, the Vinhão cultivar tends to equilibrate the concentration (Table 4.1-2). During grape maturation, Ca tends to accumulate in the grapes, especially contributing to the flavonoid pathways (Deluc et al., 2007), and Mg is present in the enzyme activator (Guo et al., 2016). The differences in N content observed in both the cultivars from 47 days after veraison (DAV) until 61 DAV, especially in the seeds (Table 4.1-2 and Figures 4 and 5), explain the need for this element during seed development (Walker et al., 1999). Also, N is fundamental in metabolic processes,

emphasising the dynamic relationship between grape maturation, nutrient accumulation, and the metabolic intricacies specific to each grape cultivar (Tewari et al., 2021). However, the Loureiro cultivar in S3 exhibits signs of overripeness due to the higher concentration of Ca (Kurt et al., 2017). This observation aligns with the phenomenon observed in the white grapes, where the process of ripening and overripening triggers cell wall breakdown (Hocking, 2015). During this process, Ca ions are released or redistributed from the cell walls, introducing a potential influence on the measured Ca content (Martins et al., 2021; Hocking, 2015).

The application of LIBS in the element analysis of grapes presents distinct advantages. Firstly, LIBS offers a novel and exploratory approach, providing a potential opportunity for refining and validating the technique for element analysis (Andrews et al., 2023b; Singh, 2021). The promise lies in its ability to offer cost-effective and expeditious alternatives to the traditional methods employed in evaluating the elemental contents of grape tissues (Arantes de Carvalho et al., 2018). Also, it acknowledges the challenges associated with wet lab procedures, particularly the limitations imposed by the availability of biological material and emphasises the potential of LIBS to overcome such constraints. In this study, characterising the elements (Ca, Mg, and N) in the skin, pulp, and seeds with the same low sample volume (0.098 cm^3) was possible. This may be a solution to the need for a low-cost method to assess a small sample of biological tissues (Tosin et al., 2023b; Da Costa Martins et al., 2023). The adaptability of LIBS technology can mitigate the cost and time commitments typically associated with wet lab procedures, positioning it as a pragmatic solution for on-field elemental composition characterisation (Senesi et al., 2019). However, LIBS still needs a sample preparation for analysis.

In this study, the Ca, Mg, and N analyses of grape tissues relied on the data inferred from the online databases OSCAR (OSCAR, 2020) and NIST (NIST, 2020). This approach was chosen because the standardised methods (e.g., ICP-OES) are non-optimised for small tissue aliquots (e.g., 0.098 cm^3), non-expedited, expensive, and time-consuming (Andrade & Pereira-Filho, 2016; Mitić et al., 2011). These standardised methods were deemed impractical for this study's low technology-readiness level (TRL) due to the associated costs of the standardised methods for small tissue samples, highlighting the need for an early-stage, cost-effective technology. This paper's results serve as proof of the use of LIBS in demonstrating the feasibility of internal tissue quantification.

The application of LIBS faces more limitations in comparison to those of the standardised methodologies like ICP-OES. The qualitative nature of LIBS, influenced by the uncertainties in plasma formation and chemistry (Hahn & Omenetto, 2012), coupled with

dynamic complexities introduced by factors, like the humidity, carbon content, and other elements (Peng et al., 2016), poses challenges to accurate quantitative measurements. Despite its potential for studying a broader spectrum of elements (Senesi et al., 2019), it is fundamental to recognise and acknowledge the current limitations of LIBS in achieving precise quantitative analyses.

Despite its advantages, LIBS is not without challenges related to atomization and ionization, particularly after the material undergoes the drying process (Ferreira et al., 2022). The emission peaks generated by LIBS often exhibit a low intensity, creating difficulties in determining the elements at low concentrations without potential recourse to pyrolysis (Moros & Laserna, 2019), which emphasises the necessity of studying the LIBS technique's efficacy in analyzing dried plant materials, recognising the inherent limitations in emission intensity.

In an agricultural context, a consideration for LIBS is its capability to detect light elements (Ribeiro et al., 2021), such as Boron, which are important for plant nutrition (Rogiers et al., 2006). However, the carbon present in plant tissues absorbs a part of the laser energy, restricting the energy for atomizing the elements in lower concentrations (Moros & Laserna, 2019). This interference raises doubts about LIBS's accuracy in detecting and quantifying trace elements and may generate uncertainty about its use for agricultural applications.

In addition, atmospheric N introduces complexities in LIBS analysis. Although analysis did not occur under vacuum conditions, the detected N during analysis is likely to be atmospheric N, rather than an accurate representation of the N content in the plant material (Moros & Laserna, 2019). This atmospheric interference poses potential challenges in result interpretation, emphasizing the need for consideration when analysing the nitrogen-related data acquired through LIBS from plant samples.

Despite the limitations, using LIBS in on-field elemental composition (Rakovský et al., 2014) characterisation introduces new perspectives in metabolomics research. Nevertheless, validating LIBS could pave the way for broader applications (Jull et al., 2018; Nunes et al., 2010), potentially inspiring and facilitating future metabolomics research endeavours. The advantages of LIBS (in situ), including its non-destructive nature and rapid analysis capabilities, make it an attractive tool for exploring the intricate elemental composition of grapes in situ. This technology addresses the current challenges and stimulates a forward-looking perspective by encouraging a more extensive and nuanced exploration of the metabolomic landscape. It raises an understanding of the complex interplay between the elements in grapes and their

potential implications for viticulture and oenology, which are particularly relevant for climate change adaptation and sustainable agricultural practices.

4.1.5. Conclusions

Incorporating Laser-Induced Breakdown Spectroscopy (LIBS) into viticulture represents a departure from the traditional, resource-intensive wet laboratory techniques, such as ICP-OES, for the elemental quantification of grape tissues. This study explores the potential utility of LIBS across various grape tissues and cultivars, marking a progressive step in understanding its applicability.

The complex relationship between water availability and the macronutrient dynamics, especially in the white cultivars, emerges as a critical aspect. The differences in Ca, Mg, and N concentrations between the Loureiro and Vinhão cultivars reflect genetic, physiological, and environmental factors governing the grapevine nutrient dynamics. Despite its advantages, LIBS has more constraints compared to those of the standardised methodologies like ICP-OES, focusing on a limited set of elements due to the small sample size limitations. The challenges related to atomization, ionization, and interference with atmospheric N require careful consideration. This study's limitations underscore the need for future research to enhance LIBS's capabilities and explore its potential in broader agricultural applications.

Further wet lab validation of LIBS results suggests broader applicability, extending beyond the studied elements (Ca, Mg, and N), thereby facilitating its potential integration for determining the various macro and micro elements within grapes and, by extension, other fruits. This multifaceted utility positions LIBS to introduce new avenues for precision viticulture and metabolomic investigations in the context of climate change and sustainable agriculture.

Author Contributions

Conceptualization, R.T., R.M. and M.C.; methodology, F.M.-S. and R.M.; validation, R.M. and M.C.; formal analysis, R.T.; investigation, R.T., R.M. and M.C.; resources, R.M.; data curation, R.T.; writing—original draft preparation, R.T.; writing—review and editing, R.T., F.M.-S., R.M. and M.C.; supervision, R.M. and M.C.; project administration, M.C.; funding acquisition, R.M. and M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: Not applicable

Acknowledgments

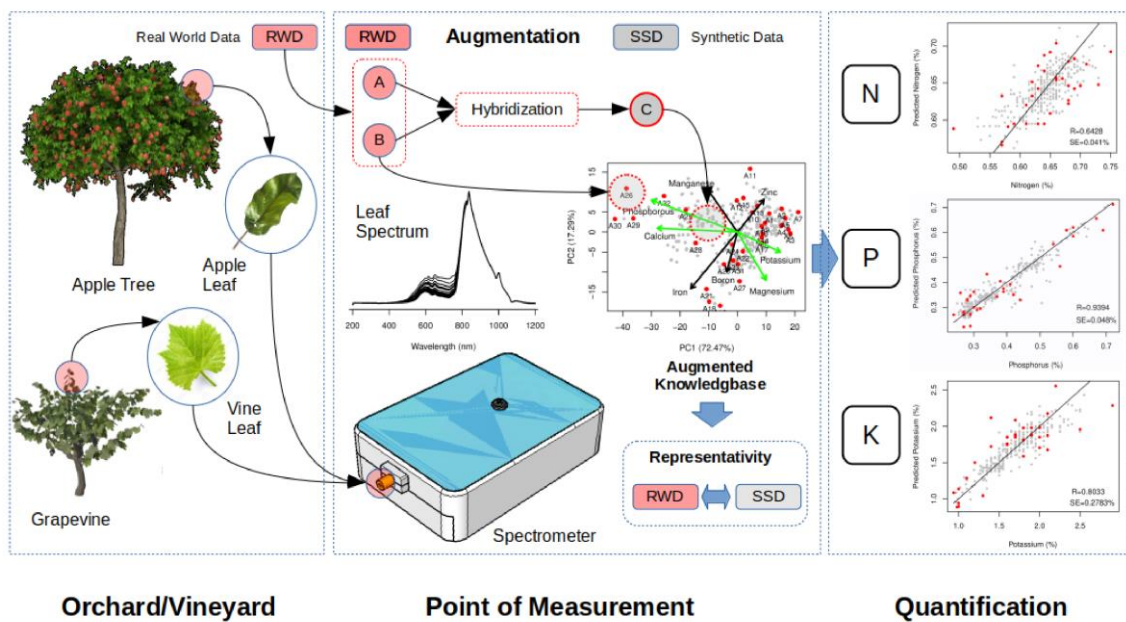
Renan Tosin and Filipe Monteiro-Silva acknowledge Fundação para a Ciência e Tecnologia (FCT) PhD research grant refs. SFRH/BD/145182/2019 and SFRD/BD/09136/2020. Rui Martins acknowledges the Fundação para a Ciência e Tecnologia (FCT) research contract grant (CEEIND/017801/2018). This work is financed by National Funds through the FCT, Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project OmicBots, High-Throughput Integrative Omic-Robots Platform for a Next-Generation Physiology-based Precision Viticulture, with the reference PTDC/ASP-HOR/1338/2021.

Conflicts of Interest: The authors declare no conflicts of interest.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

4.2.Case study III: quantitative Vis-NIR for nutrient identification in apple and grape leaves.

Martins, R.C., Queirós, C., Silva, F.M., Santos, F., Barroso, T.G., **Tosin, R.**, Cunha, M., Leão, M., Damásio, M., Martins, P., and Silvestre, J. (Revised version under review). Spectral Data Augmentation for Dealing with Data Scarcity in Leaf Nutrient Uptake Quantification.



Graphical abstract. 4.2-1. . Graphical abstract case study III (Martins et al., Under review).

Spectral Data Augmentation for Dealing with Data Scarcity in Leaf Nutrient Uptake Quantification

Martins RC¹, Queirós C², Silva FM^{1,3}, Santos F¹, Barroso TG⁴, **Tosin R**^{1,3}, Cunha M^{1,3}, Leão M⁵, Damásio M⁵, Martins P⁵, Silvestre J⁵

¹INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, R. Dr. Roberto Frias, 4200-465 Porto – Portugal

²LAQV/REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences of the University of Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal

³Faculty of Sciences of the University of Porto, Rua do Campo Alegre, 4169-007, Porto – Portugal

⁴TOXRUN – Toxicology Research Unit, University Institute of Health Sciences, CESPU, CRL, 4585-116 Gandra, Portugal

⁵INIAV - National Institute of Agricultural and Veterinary Research, I.P., Quinta da Almoíña EN 374, 2565-191 Dois Portos, Portugal

Highlights

- Hybridization expands the information of the original dataset by inter- and extrapolation;
- Synthetic data reproduces spectral and compositional information of the real world samples;
- Hybridization creates samples that could have been sampled in real-world;
- Representativity of the real-world data is essential for information expansion;
- Data augmentation as a diagnostic tool of knowledgebase representativity.

Abstract

Data scarcity is a major hurdle for physiology-based precision agriculture. Measuring nutrient uptake by Visible-Near Infrared spectroscopy implies collecting spectral and compositional data from low-throughput, such as inductively coupled plasma optical emission spectroscopy. This paper introduces data augmentation in spectroscopy by hybridisation for expanding real-world data into synthetic datasets statistically representative of the real data, allowing the quantification of macro- (N, P, K, Ca, Mg, and S) and micro-nutrients (Fe, Mn, Zn, Cu, and B). Partial least squares (PLS), local partial least squares (LocPLS), and self-learning artificial intelligence (SLAI) were used to determine the capacity to expand the knowledge base. PLS using only real-world data

cannot quantify some nutrients (N and Cu in grapevines and K, Ca, Mg, S, and Cu in apples). In the synthetic dataset of the study it allowed predicting real-world leaf composition of macro- (N, P, K, Ca, Mg, and S) ($R \sim 0.6073 - 0.9394$ and $SE \sim 0.0425-0.0484\%$) and micro-nutrients (Fe, Mn, Zn, Cu, and B) ($R \sim 0.6577 - 0.9098$ and $SE \sim 0.8753-3.978$ ppm) in grapevine leaves using LocPLS and SLAI. The synthetic dataset loses significance if the real-world counterpart has low representativity, resulting in poor quantifications of macro ($R \sim 0.5139-0.7207$ and $SE \sim 0.0194-0.1269\%$) and micronutrients ($R \sim 0.5250-0.7599$ and $SE \sim 8.892-37.89$ ppm), and not allowing S quantification ($R=0.3737$, $SE=0.0084$) in apple tree leaves. Representative real-world sampling makes data augmentation in spectroscopy very efficient in expanding the knowledge base and nutrient quantifications, highlighting the importance of carefully designing experiments in field experiments for a representative knowledge base.

Keywords

Nutrient uptake; Leaf spectroscopy, Precision Fertilisation, Chemometrics

4.2.1. Introduction

Precision fertilisation is a new emergent technology for adapting fertilisation to plant nutritional needs based on soil and plant nutrient uptake analysis. Nutrient uptake evaluation comprises the analysis of nutrients in sap and leaves or other plant parts (Mota et al., 2022). Leaf analysis directly measures nutrients used for growth related to photosynthesis and biomass accumulation rate. Recent studies show that leaf blades better predict nitrogen (N) and potassium (K) grapevine status (Schreiner & Scagel, 2017; Benito et al., 2013). Nutrient uptake in perennial crops has been interpreted mainly by the sufficiency range method (Partelli et al., 2007) due to the straightforward interpretation of results to manage fertiliser application properly (Ge et al., 2018).

Today, laboratory-based macro-and micronutrient leaf quantification is performed using inductively coupled plasma-optical emission spectroscopy (ICP-OES). This ground-truth technique is destructive, laborious, expensive, and has low throughput. Comprehensive sampling in vineyards or orchards becomes unpractical as metabolism dynamics push plants into a new physiological state, resulting in the correction of deficiencies based on late data, which may not provide the necessary expected results in productivity and quality (Kumar et al., 2021b; Wawrzynska & Sirko, 2014). Leaf tissue composition is inherently variable, influenced by factors such as leaf age, maturation stage, and nutrient interactions during uptake and translocation (Kumar et al., 2021b; Wawrzynska & Sirko, 2014).

Nutrient sensing with spectroscopy

Nutrient uptake leaf monitoring using visible near-infrared (Vis-NIR) spectroscopy is a rapidly developing field currently at the forefront of precision fertilisation, comprising the measurement of spectral reflectance, signal processing and chemometrics/artificial intelligence (Tosin et al., 2022; Martins et al., 2022).

Point of Measurement (POM) nutrient quantification using ultraviolet-visible-near-infrared (UV-Vis-NIR) is reported for N, phosphorus (P), K, calcium (Ca), magnesium (Mg), iron (Fe), manganese (Mn), zinc (Zn), and copper (Cu) using partial least squares (PLS), being extensively reviewed (Prananto et al., 2021; van Maarschalkerweerd & Husted, 2015). Quantification has been shown in leaves for tomato (Ulissi et al., 2011), olive tree (Rotbart et al., 2013), cotton (Tarpley et al., 2000), pear (Jie et al., 2014), wheat and barley (Zerner & Parker, 2019), oil palm (Santoso et al., 2018), orange (Menesatti et al., 2010), persimmon (Acosta et al., 2023), eucalyptus (Oliveira & Santana, 2020), citrus (Galvez-Sola et al., 2015) and sugarcane (Yarce & Rojas, 2012) (Supplementary material - Tables A4 and A5). Most nutrient quantifications are performed using dried and ground leaves. Prananto et al. (2021) also measured 'in-vivo' cotton leaves.

Spectroscopy uses the relationship between spectral characteristics and leaf composition, registering each sample's multi-scaled interferences and matrix effects. It provides an information-rich signal, which requires pre-processing for the removal of non-absorbance effects (e.g., scattering) and advanced chemometrics/artificial intelligence methods for unscrambling the information contained in the spectra (Martins et al., 2022; Silva et al., 2021; Barroso et al., 2021c; Monteiro-Silva et al., 2019; Martins, 2018). Accurate predictions depend on dataset representativity and model efficiency in capturing the data information.

Data scarcity

Data abundance in agriculture is a key factor for state-of-the-art artificial intelligence models such as large-scale deep learning (Rahim et al., 2022; van Klompenburg et al., 2020; Kamilaris & Prenafeta-Boldú, 2018), artificial neural networks (Martins et al., 2022), and strategies such as reinforcement learning (Gautron et al., 2022) or adversarial generative models (Lu et al., 2022; Nazki et al., 2020; Barth et al., 2020). Data abundance originated from satellite and drone imaging, robot cameras, and light detection and ranging (LIDAR) or Internet of Things (IoT) climate instrumentation. In contrast, data scarcity is common in highly valuable and expensive information about a plant's physiological state, such as in 'omics' and nutritional uptake, because the information is not yet available as a high-throughput sensing system in the field. The

corresponding laboratory alternative is low-throughput and expensive, such as the case of nutrient quantification using ICP-OES.

Data scarcity also predominates in ‘omics’ exploratory studies. The uncertainty in the experiment’s outcome makes researchers use a broad design of experiment (DoE) ranges to determine the feasibility, contrasts, and effects before scaling up experiments to increase the knowledge base and detailed experimental representativity. Data scarcity constrains the use of widespread artificial intelligence (AI) methods (e.g. deep-learning (DL) or artificial neural networks (ANN)) because these require significant amounts of data to be successfully applied. Methods such as partial least square (PLS) or local methods, such as local-PLS or self-learning artificial intelligence (SLAI), are more appropriate in these circumstances because of their need for less data and interpretability due to direct relationships with theoretical relationships, such as the Beer-Lambert law (Martins et al., 2022; Barroso et al., 2021c).

Data augmentation (DA) can enhance the knowledge base’s diversity and improve the accuracy of model predictions (Barroso et al., 2022; Shorten & Khoshgoftaar, 2019; Nalepa et al., 2019). The biological variance of plants in agricultural ecosystems is essential, and ‘omics’ state-of-the-art technologies still provide limited representativeness of real-world data (RWD), which imposes challenges for developing robust constituent quantifications using non-invasive spectroscopy-based technologies, as there is no capacity for establishing an extensive data knowledge base from RWD records. This problem is partially solved by developing self-learning models, recognising whether the measured spectra are known or unknown before making a prediction and suggesting a knowledge base update (Martins et al., 2022).

It is proposed DA for enhancing leaf nutrients spectral quantification by introducing an ‘in silico’ synthetic spectroscopy dataset (SSD) (Figure. 1c). SSD is created by random hybridisation of two samples, building new spectra and corresponding nutrient composition with 1:1 information mixture; being comparable to a hybrid leaf with the average metabolic state of the two RWD samples.

This research introduces and evaluates the effectiveness of hybridisation as a DA technique in spectroscopy for dealing with data scarcity in field data, analysing the capacity of knowledge base expansion in spectroscopy for leaf nutrient quantification in grapevine and apple leaves, with the following specific objectives: i. determine if SSD is representative of RWD information; ii. assess if SSD can maintain the information equivalence principle between the spectra and nutrient composition; iii. understand how SSD expands information in RWD, improving model accuracy; and iv. how can this new method provide insights towards the efficient upscaling design of experiments (DoE) to

enable spectral nutrient uptake quantification and physiology-based precision agriculture (PA)?

4.2.2. Materials and Methods

4.2.2.1. Experimental Design

A 20-year-old commercial rain-fed vineyard located in Dois Portos (Torres Vedras: 39.043 N, 9.182 W) is planted on deep sandy loam soil with three varieties: Touriga Nacional (Vitis International Variety Catalogue (VIVC) - 12594), Tempranillo (VIVC - 12350), and Marselan (VIVC - 16383). The number of plants is 2958 per hectare (2.6 m x 1.30 m). The experiment involved applying different amounts of N and P to the soil at the winter's end, followed by superficial tillage in eight blocks of 60 plants per block, as shown in Table S 4.2-1 (Supplementary Material). Climate Csb class (Koppen-Geiser) with an average annual temperature of 15.8°C and 638 mm of rainfall, of which 169 mm occur during the growing season.

The experimental apple orchard is located in Alcobaça (39,549 N; 8,957 W) planted with *Malus domestica* Borkh 'Gala Brookfield', grafted on MM106 rootstock, with 1852 trees per hectare (4.50 x 1.50) and a drip irrigation system consisting in a single pipe of drippers per row, 1.2 meters apart and a water flow of 4.0 l h⁻¹ per dripper. The experimental trial consists of a randomised complete block design, with three different fertilisation levels, according to Table S 4.2-2, with twelve blocks per treatment and ten trees per block. Climate Csb class (Koppen-Geiger) with an average annual temperature of 15.1 °C and rainfall of 819 mm, of which 329 mm occur during the growing season.

4.2.2.2. Sampling

In the vineyard case study, healthy leaves (free from visible pests or diseases) were collected at veraison (August 11th) from the 4th to 6th node of vines from the centre of each block to minimise interblock influence. As mineral nutrient concentrations vary throughout the growing season, sampling leaf blades or petioles at bloom or veraison is recommended for routinely monitoring vine nutrient status. In addition, veraison sampling may offer a more accurate indication of grapevine macronutrient levels like N, P, and K due to their greater fluctuations during bloom and their relative mobility within the plant. Two leaves were collected from 15 plants in each of the eight blocks, performing 960 leaves. Leaves were excised with the entire petiole, placed in plastic bags and immediately transported to the lab. In the apple case study, eight healthy leaves per tree were detached by the petiole in the morning of August 8th from the middle third of the

extension shoots with a length between 15 and 30 cm, two from each of the four quadrants, half located outside and the other half located inside the canopy. Eight healthy leaves per tree were collected in the apple trees' case. Since there are 12 blocks of ten plants and three fertilisation treatments, 2880 leaves were collected. Each block was separated, placed in plastic bags, and immediately transported to the lab. Leaves were firstly subjected to spectral recording, and afterwards, samples were washed with distilled water, dried at 65 ± 5 °C and ground to a particle size of 1 ± 0.1 mm for laboratory analysis.

4.2.2.3. Spectroscopy

Leaf spectra were recorded with a USB4000 Ocean Optics spectrometer (200-1200nm), with reflectance probe 600um core (Ocean Insight), and a tungsten-halogen light source (360-2400 nm, Ocean Insight). The integration time was optimised to allow optimal spectral recording within the linear range of the spectrometer detector. Spectra were obtained at room temperature of 18 ± 2 °C, and the tungsten lamp was left to stabilise for 15 min. The dark spectra were recorded, and linear and electric dark correction measurements were taken. Both light spectra were monitored by statistically assessing the reproducibility of the light source, using measurements of light during the several days of the experiment. Ten spectra of each leaf were recorded at different positions in the dark to avoid external light influence, and the average spectra of each leaf were considered for analysis (see Figure 4.2-1).

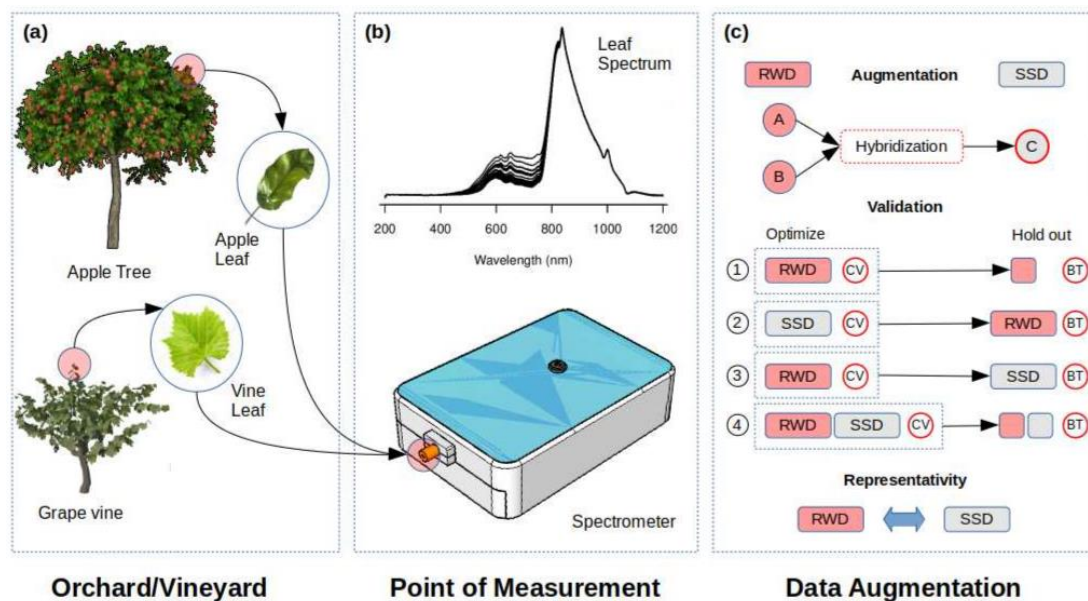


Figure 4.2-1. Nutrient quantification using visible near-infrared (Vis-NIR) spectroscopy: (a) Apple and Grapevine leaf sampling for spectral and inductively coupled plasma-optical emission spectroscopy (ICP-OES) analysis; (b) leaf spectral measurement; and (c) data augmentation for exploratory analysis including synthetic spectroscopy dataset (SSD)

generation from SSD using hybridisation and validation schemes (1 to 4) to determine SSD representativity of real-world data (RWD).

The robust multiplicative scatter correction algorithm (RMSC) was applied to the spectral records to perform scattering correction. The method is a modified version of (Gallagher et al., 2005; Martens et al., 2003) but uses re-weighted least-squares and Huber functions. A high breakdown point of 50% is implemented. Therefore, outliers will not affect scatter correction parameters, which are determined using only the consistently relevant spectral bands (Martins et al., 2022; Barroso et al., 2021c; Monteiro-Silva et al., 2019).

4.2.2.4. Element quantification

Total N and S (sulphur) determination were performed by catalytic pyrolysis followed by gas-chromatographic (GC) separation: 6 mg of vegetal material, weighted with an analytical scale, was placed in a Thermo Flashsmart Total Nitrogen, Carbon, and Sulfur (NCS) analyser and then submitted to catalytic pyrolysis at 950 °C in a reactor.. The resultant oxidised gases are then separated in a GC column. N, C and S are detected and quantified by thermoconductivity. Results are referred to dry material at (100-105) °C. Total P, K, Ca, Mg, Fe, Mn, Zn Cu and boron (B) determination by inductively coupled plasma optical emission spectrometry: the vegetal material is rigorously weighted (500 mg) and reduced to ashes in a muffle at 500 ± 10 °C over 4th. The ashes were solubilised in hydrochloric acid 3M. P, K, Ca, Mg, Fe, Mn, Zn Cu and B were determined in the resulting solution by ICP-OES (Thompson, 2012; Chaves et al., 2010; Duarte et al., 1998).

4.2.2.5. Spectral Data Augmentation

The synthetic dataset is generated by the hybridisation of two random RWD samples. Hybrid samples do not occur naturally in the RWD and may or may not reflect existing leaves. However, by artificially combining information, knowledge base expansion is achieved by spawning spectral gradients proportional to the nutrient concentration. A total of 300 SSD samples were obtained by non-repeated random mixing of pairwise RWD spectral samples (X) and ICP-OES leaf nutrient composition (Y), where $X_{SSD} = \frac{1}{2}(X_i + X_j)$ and $Y_{SSD} = \frac{1}{2}(Y_i + Y_j)$, with i and j random leaf samples.

Figure 4.2-1c describes how SSD was tested for RWD representativeness (null hypothesis) by splitting model optimisation by cross-validation (CV) and blind testing (BT): (1) RWD benchmark leave-one-out optimisation and leave-one-out hold-out blind

testing in PLS models (RWD); (2) SSD Optimised - leave-one-out CV optimisation and blind testing with RWD (SSD→RWD); (3) Optimised leave-one-out CV optimisation and BT testing with SSD (RWD→SSD); and (4) Joint RWD and SSD and random splitting into optimisation using CV and BT datasets (90:10) (Figure 1c, RWD+SSD). Rotation to RWD and SSD datasets in schemes 1 (RWD) and 4 (RWD+SSD) was performed during optimisation to evaluate all samples as BT.

The increase in the information in two-dimensional feature space by data augmentation is graphically represented in Figure 4.2-2. RWD restricts the knowledge base to a polygon area within the feature space (green area of Figure 2a). Inside it, SSDs are generated by interpolating the RWD edges of the polygon, being expected to preserve the natural information present in the RWD and increase the representativeness of local covariances - allowing unscrambling covariance modes (CovMs).

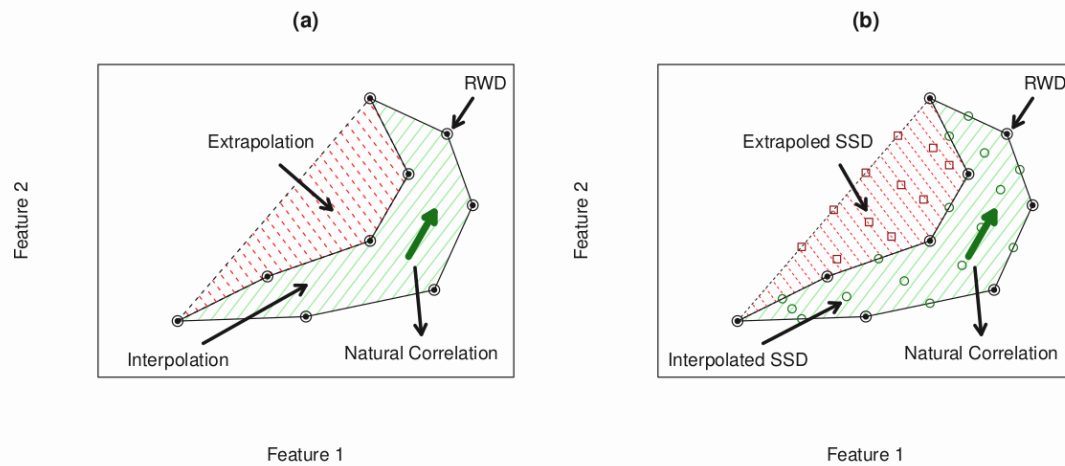


Figure 4.2-2. Real-world data (RWD) and synthetic sample generation by hybridisation: (a) RWD polygon defining the existing scarce dataset characteristics: edges represent the multi-dimensional boundaries of RWD in the feature space where data can be interpolated, and natural correlations are observed (green area), and non-observable characteristics by extrapolation of convex face of the RWD (red area); (b) synthetic data generated by hybridisation of RWD sample edges preserving the information of the RWD in interpolation (◊) and extrapolation (◻) areas.

The convex area in red (Figure 4.2-2a) is obtained by extrapolating SSD through the hybridisation of the convex boundary samples of the RWD polygon (Figure 4.2-2b). Extrapolated SSD does not hold the natural correlation of the original dataset but is expected to retain the information on spectral gradients and interferences to be proportional to nutrient composition, extrapolating unknown characteristics.

4.2.2.6. Benchmarking

CovM search methodology was benchmarked against chemometrics methods designed to handle limited proof-of-concept datasets:

i. Partial least squares regression (PLS). PLS maximises the covariance between the spectra matrix X and leaf composition matrix Y by determining the eigenvectors of X^tY (Ergon, 2009; Geladi & Kowalski, 1986). The method forces the latent structures of spectra and composition (t and u) to be equal (nonlinear iterative partial least squares (NIPALS) or singular value decomposition (SVD) algorithm) to determine each correspondent basis P^t and Q^t (Geladi & Kowalski, 1986) where $X=TP^t$ and $Y=UQ^t$. It proceeds with deflation and sequential orthogonal eigenvectors of the remaining information in X^tY . The number of latent variables (LV) is optimised by cross-validation minimal predicted sum of squares. PLS uses an oblique projection to determine b_{pls} coefficients in $Y=X b_{pls}$ (Ergon, 2009; Phatak & De Jong, 1997).

ii. LocPLS: uses k-nearest neighbour (KNN) clustering to create local sub-groups, where local PLS models are optimised. The KNN clusters are obtained in the principal component analysis (PCA) scores space. The number of clusters and principal components (PC) are optimised by cross-validation (CV)/hold-out (HO) samples (Shen et al., 2019).

iii. Self-learning Artificial Intelligence: uses a two-step approach: a) features space optimisation for providing equivalence between the spectral information (T) and leaf nutrient composition (U) latent spaces; and b) performing local CovM search for unscrambling groups of samples that have the same type of interferences, providing a direct quantification between spectra and nutrient with a single dimension (or eigenvector) exhibiting stable covariance (X^tY) (Silva et al., 2023b; Martins et al., 2022; Barroso et al., 2021c; Martins, 2019).

4.2.2.7. Validation

A two-step approach validated models: (i) CV to optimise model parameters within the training set and (ii) prediction for HO samples to estimate the error. CV assumes training set representativity by performing a null hypothesis test - the null effect is expected if a particular sample is present or not in the training dataset, attaining statistically similar results. By leaving several samples out of model estimation, CV provides the error estimated for each sample in the training set if this sample is unknown, allowing one to decide which are the optimal parameters of each model that best depict the representative features of the dataset and not the particularities of each sample. If the dataset and model are representative, the null effect is expected when using the model to predict hold-out datasets, holding similar error results to the training dataset. CV is

used to avoid over-optimising the dataset (overfitting) and hold-out samples (HO) for null hypothesis testing, evaluating representativity and possible generalisation of the chosen model. Non-optimal models are more robust to generalisation at the cost of accuracy, being a trade-off necessary under data scarcity.

Models are chosen for optimal generalisation at minimum prediction error of CV (e.g., LV in PLS, Loc-PLS and 1LV in CovM). Method performance was evaluated by computing the standard error (SE) estimate obtained by the root mean squared error of the BT data, the Pearson correlation (R) as a metric of linearity between predicted and measured values, and the coefficient of determination (R^2) as a goodness of fit metric, and the p-value was computed among the validation combinations to assess the statistical similarity between RWD and SSD in PLS, Local-PLS and SLAI as presented in section 2.5 (Figure 4.2-1c).

A two-step approach validated models: (i) CV to optimise model parameters within the training set and (ii) prediction for HO to estimate the error. CV assumes training set representativity by performing a null hypothesis test - the null effect is expected if a particular sample is present or not in the training dataset, attaining statistically similar results. By leaving several samples out of model estimation, CV provides the error estimated for each sample in the training set if this sample is unknown, allowing one to decide which are the optimal parameters of each model that best depict the representative features of the dataset and not the particularities of each sample. If the dataset and model are representative, the null effect is expected when using the model to predict HO datasets, holding similar error results to the training dataset. CV is used to avoid over-optimisation of the dataset (overfitting), and hold-out samples (HO) are used for null hypothesis testing, an evaluation of representativity, and possible generalisation of the chosen model. Non-optimal models are more robust to generalisation at the cost of accuracy, being a trade-off necessary under data scarcity.

All models were constructed with the mean spectra of 10 leaf replicate measures in different leaf spots (Figure 4.2-1.). CRAN-R was used for all computations (PLSR packages, LocPLS and SLAI utilising the author's code) (R Core Team, 2020).

4.2.3. Results and Discussion

4.2.3.1. Spectral characterisation

This study reveals that DA enhances the performance of spectral quantification methods, particularly LocPLS and SLAI. When confronted with RWD alone, traditional methods like PLS struggle to quantify certain nutrients in grapevines and apple trees accurately. However, with the integration of SSD, LocPLS and SLAI demonstrate accuracy in

predicting the leaf composition of both macro and micro-nutrients. This improvement underscores the importance of DA in expanding the capabilities of predictive models, enabling more precise quantification of complex biological samples.

Grapevine and apple tree leaves main spectral variance occurs in the visible part of the spectrum, which is mainly associated with the leaf pigments (e.g. chlorophylls, pheophytins, carotenes) in the region of 500 nm to 850 nm (Figure 4.2-3a and 3b). The 550 nm to 750 nm region exhibits a higher variance, which is more significant in apple trees (Figure 4.2-3c) than in grapevine (Figure 4.2-3a).

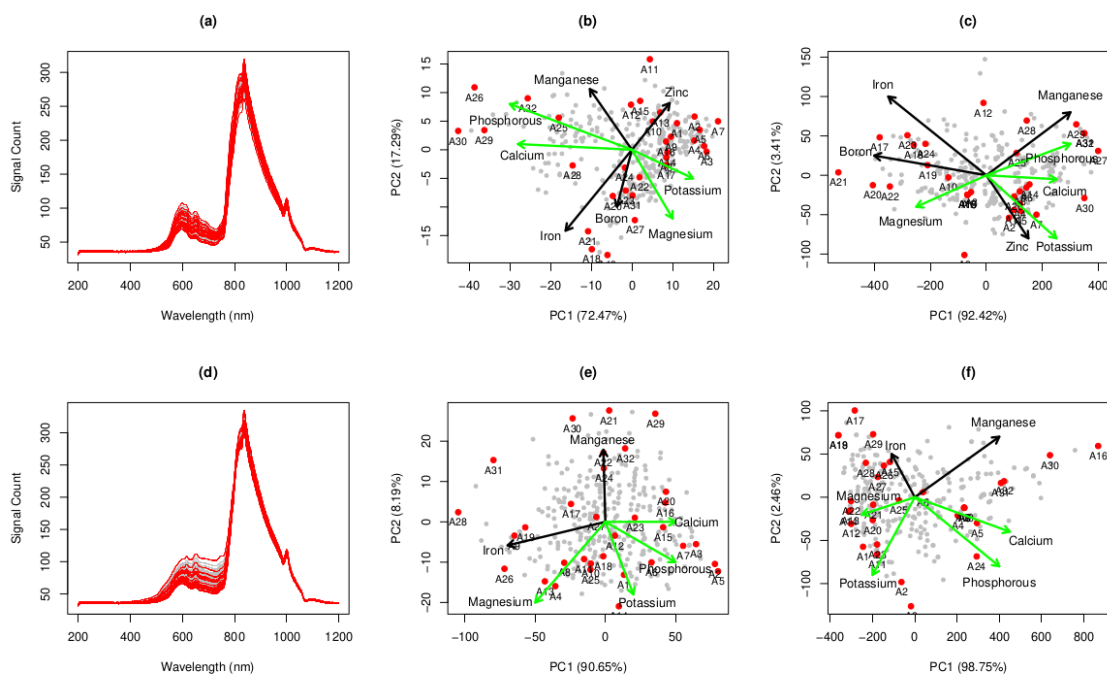


Figure 4.2-3. Spectra information and nutrient composition relationships: (a) grapevine leaf spectra: (■) real-world data (RWD) and (□) synthetic spectroscopy dataset (SSD); (b) Principal component analysis (PCA) scores and loadings of grape leaf nutrient composition; (c) grapevine leaf spectral PCA scores and nutrient gradient vector; (d) apple tree leaf spectra; (e) PCA scores and loadings apple tree leaf nutrient composition; and (f) grapevine leaf spectral PCA scores and nutrient gradient vector; where (■) are RWD samples and (□) are SSD samples for grapevine and apple tree leads.

Spectral differences are associated with the grapevine leaf composition. 89.76% of the compositional variance is explained by two orthogonal PCs, holding PC1 (72.47%) and PC2 (17.29%) of variance (Figures 3b and 3d). Most variance is associated with macronutrients P, K, Ca, Mg, Fe, B, and Zn micronutrients. Increased levels of P and Ca are related to lower levels of K and Mg, and higher levels of Mn, Fe, and B are associated with lower amounts of Zn. A lower degree of variance is given by samples with increased levels of Mn and Zn with lower levels of Fe and B (e.g., samples A26, A11, A32 in Figure 4.2-3b) and vice-versa (e.g. samples A16, A19 and A21 in Figure 4.2-3b).

Macro-nutrients variance is dominated by K (19600 ppm) and Ca (12000 ppm) (Table S 4.2-3 – Supplementary material). Micro-nutrients variance is dominated by Fe

(39.1 ppm), Mn (47 ppm), Zn (34 ppm), and B (15 ppm). The following nutrients are auto-correlated: i. P, Ca and S ($R \sim 0.7047-0.8213$); ii. Mg, B ($R=0.6271$); and iii. Fe, Mn ($R=0.6184$).

Fe dominates compositional variance in apple tree leaves in PC1 (90.65%) and Mn in PC2 (8.19%) (Figure 4.2-3d). Apple tree leaf macro-nutrient composition is mainly attributed to K (6000 ppm) and Ca (5000 ppm). Micro-nutrient variance is attributed to Fe (183 ppm), Mn (78 ppm), Zn (13 ppm), and B (7 ppm) (Table S 4.2-3. The following nutrients are auto-correlated: i. Fe, Zn ($R=0.4516$); ii. P, Ca ($R=0.6394$); and iii, K, B ($R=0.6282$).

Information Similarity

Spectral quantification occurs when the information in the spectra (X) and the nutrient composition (Y) are equivalent. A direct way to qualitatively determine whether the information is equivalent is to independently compare PCA of the composition and spectra (Figure 4.2-3b and 3c for grapevine and Figure 4.2-3f and 3g for an apple tree, respectively). PCA maximises the variance of X or Y; therefore, if the eigenstructure is similar in both datasets, it forms the basis for consistent global covariance X^tY (Martins et al., 2022; Barroso et al., 2022; Barroso et al., 2021c; Martins et al., 2018). Therefore, U and T are equivalent, carrying the same information but on a different basis, where $X=TP^t$ and $Y=UQ^t$, P and Q are PC.

In the grapevine leaves, T and U have similarities. Sample groups are generally maintained within the two latent spaces, with their relative nutrient composition maintained in the T scores (Figure 4.2-3b and 3c). For example, i. macro-nutrients: high levels of Ca and P (e.g. A26, A29, A30, A32); high levels of K (e.g. A2, A4, A5, A6, A7); and high levels of Mg (A18, A20-A22); and ii. micro-nutrients: high levels of Zn in U are preserved in T (e.g. group A3-A7, A11 and A15); high levels of Fe (e.g., groups A18, A20, A22, and A23); high Mn (e.g. A12, A25, A28, A29, A32). A few samples are inconsistent in T; for example, A27 was expected to be within a group of high B and A30 in a high Fe group.

Fe and Mn dominate the spectral and compositional variance in the apple tree leaves. High/low Mn and Fe groups are mostly consistent between T and U in Figure 4.2-3e and 3f. For example, groups of high Fe (A17-A19, A22, A26, A27), high Mn (A24, A30, A31, A32), low Fe (A1-A7, A10-A15, A23), and low Mn (A1-A14, A20, A25) are examples of variance consistency among spectral and compositional information for these dominant nutrients. A few samples show inconsistencies, such as A22/A29, which were expected to be within a high-Mn group, and A16 was expected to be near a group of average Mn content in T.

SSD representativity is observed in the T and U spaces of the grapevine and apple tree leaves (Figure 4.2-3b and 3c, and 3e and 3f), which can be evaluated in terms of the stability of (T, U) and (P, Q) calculated solely using RWD or SSD, which prove to be statistically similar ($p < 10^{-6}$). RWD and SWD carry the same nutrient gradient information. Still, SSD expands the RWD space into sample characteristics previously under-represented, such as regions between low and high Ca and P concentrations in grapevine (Figure 4.2-3b and 3c) or high/low Mn, Ca and P concentrations in apple tree leaves (Figure 4.2-3e and 3f).

The qualitative analysis of X and Y PCA scores and loadings demonstrate the high level of information similarity between T and U latent spaces (Figure 4.2-3b and 3c; Figure 4.2-3c and 3f). Both plant leaves show the association of P and Ca and how increased levels of these nutrients correspond to lower levels of Mg and K. The association between Fe and Mn is also similar in both species, maintaining a high degree of orthogonality. Both plants' T and U spaces maintain stable associations between nutrients and sample groups of characteristic composition. A similar eigenstructure (T~U) indicates global covariance (X^tY) stability, allowing the extraction of consistent covariance eigenvectors proportional to nutrient quantification.

The association of Ca and P gradients show that PC1 spectral characteristics are related to these nutrients, whereas K and Mg gradients are related to combinations of PC1 and PC2 characteristics. The same is observable for B in PC1 and Fe, Mn, and Zn in combinations of PC1 and PC2.

4.2.3.2. Significance and Representativity

The impact of DA on nutrient quantification reveals advancements in predictive accuracy, particularly with the utilisation of LocPLS and SLAI. By incorporating augmented datasets, these methods can capture a more granular representation of nutrient gradients within grapevine and apple tree leaves, leading to more robust and accurate predictive models. This enhancement in representativity underscores the critical role of DA in refining the precision of nutrient quantification methods, thereby advancing the understanding of plant physiology and nutritional dynamics.

Table 4.2-1 shows the p-value of the F-test of RWD predictions, where p-values > 0.005 are considered non-significant for quantitative regression. PLS-RWD show non-significant predictions for N and Cu in grapevines and K, Ca, Mg, S, and Cu in apple leaves ($p < 0.005$, Table 4.2-1). Macronutrients that were dominant in spectral and compositional variance (Figure 4.2-3b and 3c for grapevine, and Figure 4.2-3e and 3f for apple leaves), such as P, Ca, K, Mg, and micronutrients Fe, B, Mn, and Zn are

significantly predicted by PLS-RWD in grapevine, which is not the case for K, Ca, and Mg in apple leaves, where RWD has insufficient information to provide stable covariance and significant prediction.

Table 4.2-1. The F-test p-value of RWD prediction: PLS cross-validated with RWD (PLS-RWD), PLS optimised with SSD (PLS-SSD→RWD), Loc-PLS optimised with SSD (LocPLS-SSD→RWD), and SLAI optimised with SSD (SLAI-SSD→RWD). Non-significant quantitative predictions are marked in underlined bold format (p-values > 0.005).

Nutrient	PLS-RWD		PLS-SSD→RWD		LocPLS-SSD→RWD		SLAI-SSD→RWD	
	G	A	G	A	G	A	G	A
N	<u>0.5142</u>	1.90E-05	1.10E-03	3.80E-06	2.30E-04	1.70E-04	7.30E-05	3.30E-03
P	1.70E-09	3.60E-05	2.30E-12	8.90E-06	6.60E-14	5.60E-06	1.70E-15	7.22E-07
K	1.30E-08	<u>0.392</u>	1.50E-09	0.005	7.30E-10	0.002	3.80E-08	<u>0.0193</u>
Ca	0.002	<u>0.1504</u>	6.80E-04	2.00E-05	6.00E-10	5.80E-05	4.80E-06	4.00E-04
Mg	2.40E-11	<u>0.2435</u>	1.50E-09	<u>0.0169</u>	1.10E-08	7.80E-04	3.50E-09	2.20E-04
S	1.40E-08	<u>0.3619</u>	4.10E-07	4.10E-04	5.70E-10	<u>0.0351</u>	4.60E-08	<u>0.339</u>
Fe	4.10E-05	6.00E-06	8.20E-06	1.37E-07	5.50E-10	0.002	2.40E-07	0.001
Mn	5.50E-06	7.69E-06	2.80E-06	2.60E-07	8.50E-10	4.50E-07	3.50E-09	1.20E-04
Zn	1.50E-08	4.30E-06	3.20E-08	1.40E-07	5.40E-09	2.40E-06	5.50E-13	7.60E-06
Cu	<u>0.422</u>	<u>0.031</u>	<u>0.022</u>	1.00E-04	0.001	4.00E-04	4.30E-05	2.00E-03
B	1.70E-05	7.00E-03	9.80E-06	2.37E-05	1.60E-06	1.50E-04	1.90E-05	<u>0.017</u>

G - Grapevine; A - Apple

The SSD can create a stable global covariance matrix ($X_{SSD}^t Y_{SSD}^t$), which is representative of the RWD—except for Cu ($p=0.022$) in grapevines and Mg ($p=0.0169$) in apple leaves, all other nutrients attained significant predictions when using SSD as a training set of PLS (PLS-SSD→RWD). Most quantifications increase their statistical significance from PLS-RWD to PLS-SSD→RWD. For example, K, Ca and S quantification in apple leaves becomes significant (Table 4.2-1).

The increase in the significance of PLS-SSD→RWD shows that SSD represents RWD, further complementing missing spectral and nutrient gradients in the knowledge base. Hybridisation completes nutrient gradient concentrations corresponding to spectral variances not present in the RWD, such as N, K, Ca, S, and Cu (Table 4.2-1).

Biological samples' covariance is unstable, containing many local covariances due to multi-scaled interference, matrix effects, and a high diversity of physiological states, biodiversity, and tissue age. LocPLS and SLAI attempt to isolate and unscramble the different types of interference in the dataset. While LocPLS uses local neighbours in the PCA scores space grouped by the Euclidean distance, SLAI isolates CovMs, samples corresponding to a single-dimension eigenvector, with a direct relationship between composition and spectral characteristics isolating an interference mode. Both methods break down the complexity of the original dataset into local covariance.

Only S quantification in apple leaves with LocPLS models was not statistically significant ($p=0.035$). S variance in the apple dataset was minimal (0.05%, Table S 4.2-3 - Supplementary material), near the ground truth detection limit. Most nutrient quantification increases their significance when using LocPLS-SSD→RWD: i. grapevine:

N, P, K, Ca, S, Fe, Mn, Zn, Cu, and B, and ii. Apple: P, K, Mg, S, Mn, Zn, Cu, and B. Such occurs at the expense of the high dimension of locPLS models (Tables A7 and A8 – Supplementary Material) combining different interference modes. LocalPLS models using SSD (LocPLS-SSD→RWD) can represent RWD using complete SSD local information.

SLAI-SSD→RWD nutrient quantification models are statistically significant in grapevines but fail for K, S, and B in apple leaves. K and S in apple have a very low variance (Table S 4.2-3 - Supplementary Material), and significant local CovMs require higher degrees of variance (Table S 4.2-3 - Supplementary Material). SSD augmentation allows the extraction of CovMs for other nutrients, but their p-values are higher than those of the LocPLS-SSD→RWD models. These are non-optimal compared to LocPLS using more than one LV; this effect is more pronounced in the apple leaf dataset in K, S, and B (Table S 4.2-3 - Supplementary Material).

Results show that SSD expands the existing information in the original RWD dataset when RWD is representative. Increasing the RWD representativeness by scaling upsampling cannot provide gains in spectral nutrient quantification.

4.2.3.3. Nutrient quantification

The comparative analysis of prediction methods with augmented data underscores differences in their efficacy for nutrient quantification. While traditional methods like PLS exhibit limitations in capturing the detailed covariance structures inherent in biological samples, LocPLS and SLAI leverage the increased granularity provided by augmented datasets to achieve more accurate predictions. This observation highlights the importance of DA in refining predictive models, enabling more precise quantification of nutrient concentrations in grapevine and apple tree leaves. Moreover, augmented datasets' enhanced robustness and generalisability further validate the utility of LocPLS and SLAI in advancing the understanding of plant nutrition and metabolic processes.

Results in literature

Nutrient quantification using Vis-NIR spectroscopy state-of-the-art is resumed in Table S 4.2-4 (Supplementary Material) for macronutrients N, P, K, Ca, Mg, and S, and Table S 4.2-5 (Supplementary Material) for micronutrients Fe, Mn, Zn, Cu, and B. Most PLS models in the literature exhibit abnormally high numbers of LVs, indicating potential biases in the sampling process. These biases can lead to stable covariance structures favouring the performance of a significant number of deflations of X and Y, thereby biasing the PLS model towards the dataset. This observation aligns with the findings of

previous studies (Zerner & Parker, 2019; Santoso et al., 2018; Menesatti et al., 2010), highlighting the importance of ensuring dataset representativeness and addressing potential biases in model development.

Results from different plants show quantitative models for N, K, Ca, and Fe, while others are quantified semi-quantitatively, such as P, Mg, S, Mn, Zn, Cu, and B. All quantifications were performed using PLS models. Figures 4a and 4b illustrate the performance of existing models for macro and micronutrients. Most results are quantitative or semi-quantitative PLS models ($R \sim 0.65 - 0.95$, Figure 4.2-4). N, K, Ca, and Mg exhibit higher dispersion on the R and SE metrics (Figure 4.2-4a). Micronutrient quantification performance is highly dependent on the dataset, where half of the results R are below 0.65, indicating that nutrient quantification under data scarcity leads to inconsistent results.

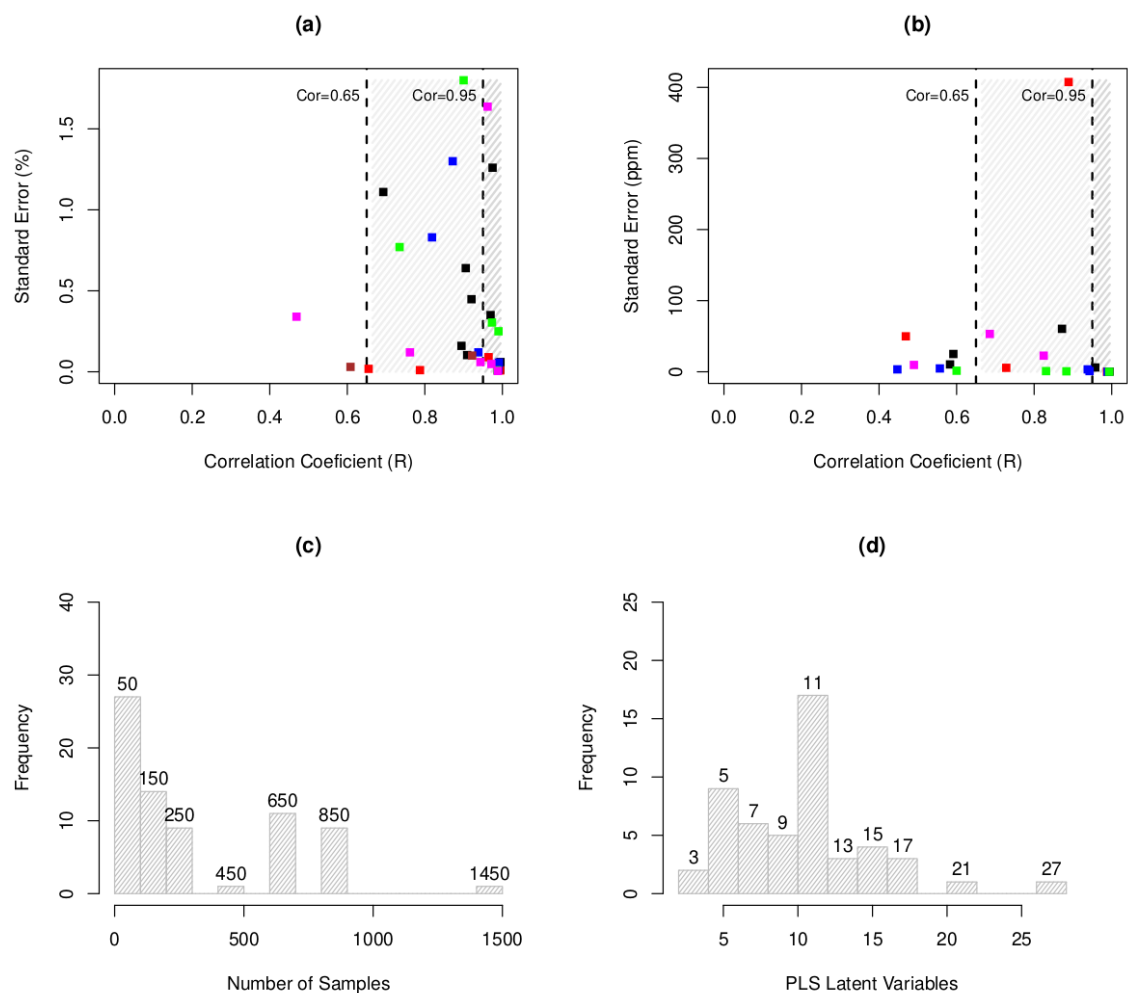


Figure 4.2-4. State-of-the-art figures of merit (Tables A4 and A5): leaf nutrient quantification figures of merit: (a) Correlation vs Standard error for macronutrients – nitrogen (N) (■), phosphorus (P) (■), potassium (K) (■), calcium (Ca) (■), magnesium (Mg) (■), sulfur (S) (■); (b) Correlation vs Standard error for macronutrients – iron (Fe) (■), manganese (Mn) (■), zinc (Zn) (■), copper (Cu) (■), boron (B) (■); (c) histogram of number of samples used in partial least square (PLS) model; and (d) histogram of number of latent variable (LV) used in PLS model.

In the case of Prananto et al. (2021), cubist regression tree modelling was used. This approach can also be regarded as an ensemble of local linear models that perform better than the global PLS, according to Ng et al. (2019) and Tang et al. (2020). The nutrient composition of leaves was evaluated under different conditions: dried and ground, fresh and removed, and fresh and intact. Validation results demonstrate that only dried and ground samples consistently achieve high R^2 values within the interval of $0.88 < R^2 < 0.92$ for macronutrients and $0.21 < R^2 < 0.78$ for micronutrients. However, performance degrades for fresh and intact samples, with R^2 values ranging from 0.21 to 0.82 for macronutrients and 0.07 to 0.74 for micronutrients. The results are consistently superior for dried and ground leaves, emphasizing the importance of sample homogeneity and controlled laboratory conditions. The difference in performance is attributed to the low water content effect on NIR and the concentration of constituents, with ground samples exhibiting higher homogeneity.

Data scarcity is common in the literature, where RWD has less than 250-300 samples, and the median sample size is 50 (e.g., Santoso et al., 2018; Menesatti et al., 2010). The low representativeness of biodiversity, physiological state, multi-scaled interferences, and matrix effects of RWD can provide a misleading optimistic PLS quantification because of inherent colinearities within each dataset, which the high number of PLS LVs (e.g., 10 to 27 LVs, Table S 4.2-4 (Supplementary Material) (Oliveira & Santana, 2020; Galvez-Sola et al., 2015; Menesatti et al., 2010) show that even in small datasets, the number of existing types of inferences, and not by their direct spectral information. Nutrients are mostly quantified by their effect on pigment metabolism, and colinearity between pigments and nutrients is an advantage of Vis-NIR spectroscopy, as it can unscramble interference to exploit nutrient signal amplification by effects in leaf pigments, allowing, for example, the quantification of micro-nutrients. Many LVs are also present in larger datasets (e.g., 12-14, (Acosta et al., 2023; Zerner & Parker, 2019). The median number of LVs in the PLS models is 11, with a wide range of LVs (3-27, Figure 4.2-4).

Figures 5 and 6 show the experimental vs predicted macro and micronutrients using SLAI-SSD→RWD (Tables A6 to A8 - Supplementary Material). The majority of macronutrients in grapevines have consistent quantitative results. N has a low R (0.643) and very low SE (0.041%) because of the low range (0.26%) of the dataset (Table S 4.2-3 - Supplementary Material). All other macronutrients (P, K, Ca, Mg, and S) exhibit significant correlations ($R \sim 0.712$ - 0.939) with low errors ($SE \sim 0.025\%$ to 0.279%) (Figure 4.2-5 and Table S 4.2-7 - Supplementary Material). Micronutrients also show

significant correlations ($R \sim 0.6577$ - 0.9098) and low errors (SE_{pl}^b 0.8759 to 7.062 ppm), except for one outlier sample in Ca, Fe, and Mn (Figure 4.2-5d, 5g and 5h).

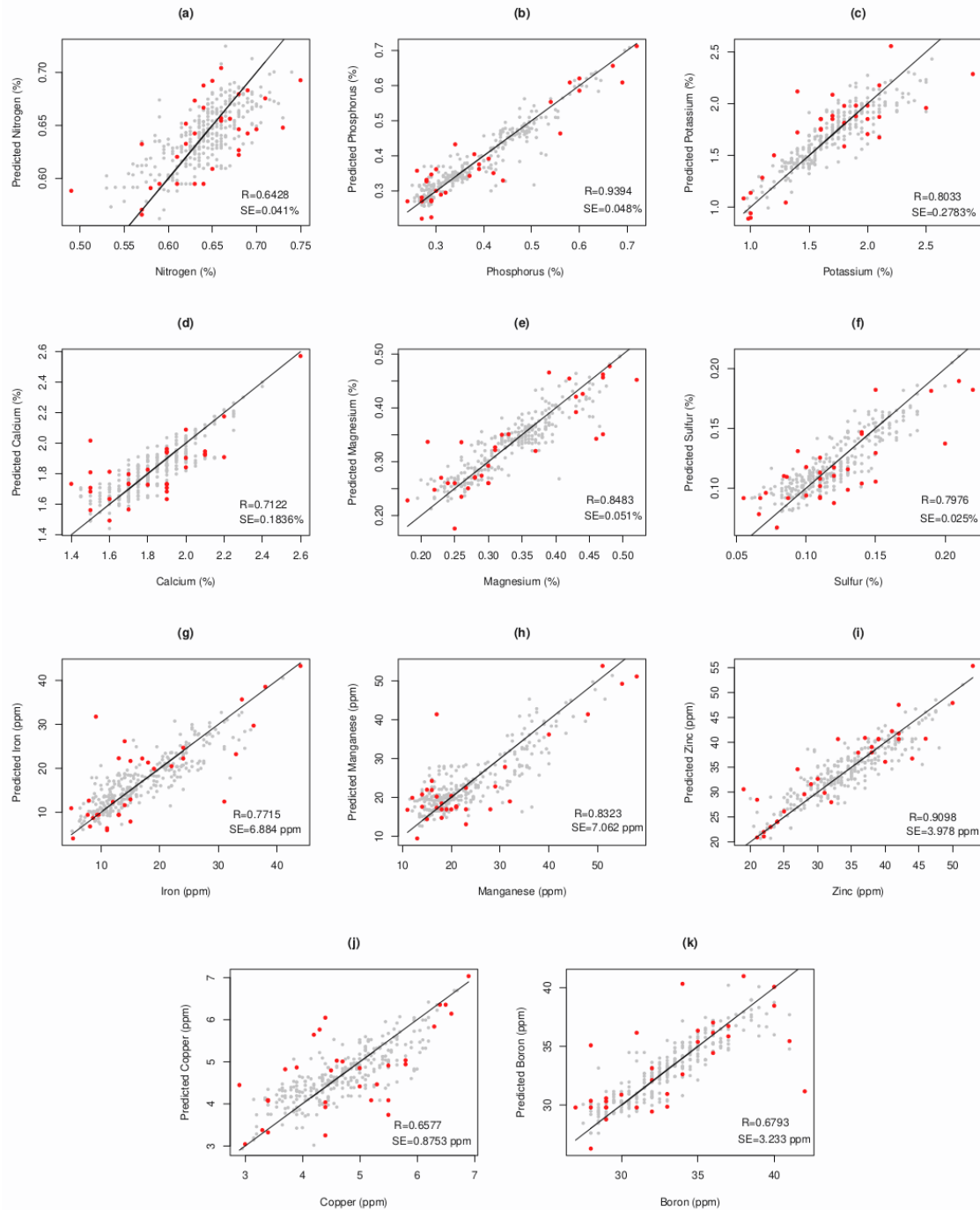


Figure 4.2-5. Nutrient quantification in grapevine: (a) nitrogen, (b) phosphorus, (c) potassium, (d) calcium, (e) magnesium, (f) sulfur, (g) iron, (h) manganese, (i) zinc, (j) copper and (k) boron - where (•) are RWD samples and (•) are SSD samples for grapevine leaves.

SLAI-SSD→RWD in apple is not as efficient - N ($R=0.504$), K ($R=0.414$), S ($R=0.037$), Ca ($R=0.491$), Fe ($R=0.544$), Cu ($R=0.522$) and B ($R=0.416$). The information in the RWD is insufficient to derive an SSD with local CovM representativity for these nutrients,

which is achieved for P ($R=0.751$), Mg ($R=0.609$), Mn ($R=0.634$), and Zn ($R=0.764$), where SSD CovMs can predict RWD.

A direct comparison between PLS-RWD and SLAI-SSD→RWD is shown in Figure 4.2-7I and 7II, for grapevine and apple leaves: i. macronutrients - major increments in R and SE in P, K and Ca in grapevine, but not in apple leaves. Significant improvement in correlation is observed in N for grapevines and in P, K, Ca, and Mg in apple leaves; and ii. micronutrients - significant SE decreases are obtained for Fe, Mn, and Zn in grapevine leaves, while no significant improvements were observed in apple leaves for these nutrients. A significant increase in the correlation was obtained for Cu (from 0.15 to 0.66) and enhancements in Fe, Mn, and Zn in grapevine leaves. No improvements were observed in micronutrient prediction in the apple leaves.

The analysis of PLS, Loc-PLS and SL-AI models obtained with SSD provides important insights into how SSD information is dependent on RWD diversity and representativity:

- I. RWD may not sustain a valid PLS-RWD model. However, if SSD can spawn RWD into a representative dataset, the missing information is interpolated, leading to quantification improvements, such as observed for N in grapevine leaves, where increases from 0.12 to 0.64;
- II. RWD is representative, but the global covariance is unstable; SSD can provide local representativity and CovMs extraction, leading to substantial improvements in quantification (e.g. P, K, Ca, Mg in grapevine and apple leaves);
- III. RWD only has the major variance but lacks the details necessary to extract local information from SSD. There are no significant improvements in using local models (e.g. Fe, Mn, Zn).

Further analysis is made when the knowledge base is increased by joining RWD and SSD (Table S 4.2-6 - Supplementary Material). All models for both micro and macronutrients in grapevine and apple leaves benefit from the knowledgebase expansion, being especially significant for local models (LocPLS-RWD+SSD and SLAI-RWD+SSD), increasing local representativity where localPLS models have low dimensionality (2 to 4 LVs) and CovMs (1LV), and corresponding correlations for all macro and micronutrients in grapevines ($R \sim 0.579$ to 0.854) and apple leaves ($R \sim 0.751$ to 0.951) (Table S-4.2-6 - Supplementary Material). Improvements are attributed to including RWD polygon edges to better represent extreme values of nutrients.

4.2.3.4. Bias-Variance Analysis

Correlation-prediction error plots are presented in Figure 4.2-8. Grapevine and apple tree leaves show the following quantitative analysis for macronutrients:

- i. Nitrogen: in grapevine leaves, all PLS models are below the quantitative R limit ($R=0.65$). Only PLS-RWD+SSD and SL-AI-RWD+SSD can quantify inside the R interval 0.65 to 0.95 (Figure 4.2-8a). In apple, local models SSD optimised (LocPLS-SSD→RWD and SLAI-SSD→RWD) cannot hold a quantitative relationship PLS models and Local models when RWD is combined with SSD, can hold N quantifications ($R \sim 0.699 - 0.821$, $SE \sim 0.068 - 0.081$) (Figure 4.2-8c). Results from the literature show N with significant correlations but very high SE (Figure 4.2-4) because of their increased range.
- ii. Phosphate: all results are inside the quantitative interval ($R \sim 0.789 - 0.939$) with very low error ($SE \sim 0.048\% - 0.046\%$) for grapevine leaves. All results are within the quantitative interval for apple leaves ($R \sim 0.663 - 0.753$, $SE \sim 0.014 - 0.021$). Showing that P quantification benefits from SSD knowledgebase expansion, allowing the determination of CovMs. Quantitative PLS models in literature attain similar results ($R \sim 0.62 - 0.988$, $SE \sim 0.008\% - 0.09\%$).
- iii. Potassium: all results are inside the quantitative interval ($R \sim 0.796 - 0.152$) with very low error ($SE \sim 0.152\% - 0.152\%$) for grapevine leaves (Figure 4.2-8a). Only the SLAI-RWD+SSD model in apple leaves can provide quantitative results ($R=0.773$, $SE=0.071\%$) (Figure 4.2-8c). All local models outperform PLS models, benefiting from SSD. The obtained R values are consistent with published data, but SE is significantly lower ($R \sim 0.43 - 0.98$, $SE \sim 0.013\% - 1.3\%$) (Figure 4.2-4a).
- iv. Calcium: All PLS models fail Ca quantification in grapevine leaves. Only local models can provide quantitative results ($R \sim 0.712 - 0.861$; and $SE \sim 0.098\% - 0.184\%$) (Figure 4.2-8a). In apple leaves, only RWD+SSD models provide quantitative results ($R \sim 0.662 - 0.884$; and $SE \sim 0.056\% - 0.065\%$) (Figure 4.2-8c). PLS results in the literature are quantitative, and this discrepancy can be attributed to the higher range of the datasets ($R \sim 0.81-0.98$; $SE \sim 0.009\% - 1.8\%$).
- v. Magnesium: Models in grapevine leaves are inside the quantification interval for grapevine leaves ($R \sim 0.648 - 0.910$; and $SE \sim 0.028\% - 0.038\%$) (Figure 4.2-8a). In apple leaves, only SLAI-RWD+SSD mode is quantitative ($R = 0.696$, $SE = 0.015\%$) (Figure 4.2-8c). Most literature results are inside the quantitative interval (Figure 4.2-4a, Table S 4.2-4 - Supplementary Material).
- vi. Sulfur: All models but PLS-RWD→SSD are quantitative ($R \sim 0.662 - 0.871$; and $SE \sim 0.015\% - 0.031\%$; Figure 4.2-8a). The opposite occurs in apple

leaves, where the best result is attained by SLAI-RWD+SSD ($R=0.579$, $SE=0.006\%$), which can be attributed to the very low range of variation in the dataset (Figure 4.2-8c). Literature results on S are not abundant, but quantifications seem highly dependent on S concentration, being only quantitative at higher ranges (Oliveira & Santana, 2020).

Grapevine and apple tree leaves show the following quantitative analysis for micronutrients:

- i. Iron: All models but PLS-RWD→SSD are inside the quantification interval ($R \sim 0.659 - 0.846$; and $SE \sim 0.311 \text{ ppm} - 7.818 \text{ ppm}$), consistently increasing the quantification performance in local models in grapevine (Figure 4.2-8b). Consistent quantification results in apple leaves are obtained when the RWD+SSD dataset is used ($R \sim 0.772 - 0.781$, $SE \sim 20.15 \text{ ppm} - 21.91 \text{ ppm}$) (Figure 4.2-8d). Results are consistent with PLS literature models ($R \sim 0.583 - 0.995$; and $SE \sim 0.15 \text{ ppm} - 76.29 \text{ ppm}$).
- ii. Manganese: All models but PLS-RWD→SSD are inside the quantification interval ($R \sim 0.710 - 0.862$; and $SE \sim 4.858 \text{ ppm} - 9.028 \text{ ppm}$), consistently increasing the quantification performance in local models in grapevine (Figure 4.2-8b). In apple leaves, most models achieved quantification ($R \sim 0.662 - 0.792$; and $SE \sim 5.280 \text{ ppm} - 6.310 \text{ ppm}$) (Figure 4.2-8d). Results are consistent with PLS literature models ($R \sim 0.469 - 0.988$; and $SE \sim 0.034 \text{ ppm} - 407.5 \text{ ppm}$).
- iii. Zinc: All models are inside the quantification interval ($R \sim 0.710 - 0.862$; and $SE \sim 4.858 \text{ ppm} - 9.028 \text{ ppm}$), consistently increasing the quantification performance in local models in grapevine (Figure 4.2-8b). In apple leaves, all models are quantitative ($R \sim 0.764 - 0.854$; and $SE \sim 1.941 \text{ ppm} - 3.641 \text{ ppm}$) (Figure 4.2-8d). Results are consistent with PLS literature models ($R \sim 0.469 - 0.988$; and $SE \sim 0.034 \text{ ppm} - 407.5 \text{ ppm}$).
- iv. Copper: only SL-AI models are inside the quantification interval ($R \sim 0.658 - 0.820$; and $SE \sim 0.491 \text{ ppm} - 0.875 \text{ ppm}$) (Figure 4.2-8b). Only local models using RWD+SSD can quantify Cu in apple leaves ($R \sim 0.622 - 0.772$; and $SE \sim 0.122 \text{ ppm} - 1.426 \text{ ppm}$), resulting from the low range of Cu (Figure 4.2-8d). Most PLS models in the literature are inside the quantification interval (Figure 4.2-4), which may be attributed to Cu's higher range and concentration.
- v. Boron: All models, except PLS-RWD→SSD, are inside the quantification interval ($R \sim 0.673 - 0.896$; and $SE \sim 0.896 \text{ ppm} - 1.422 \text{ ppm}$), consistently increasing the quantification performance in local models in grapevine (Figure

4.2-8b). In apple leaves, quantification is achieved by PLS-SSD→RWD ($R=0.674$, $SE=0.721$ ppm) and SLAI-RWD+SSS ($R=0.722$, $SE=0.646$ ppm) models (Figure 4.2-8d). Only one PLS model can quantify B ($R=0.68$, $SE=22.66$ ppm).

4.2.3.5. Data Augmentation Analysis

Through hybridisation, DA has shown two key advantages: i. knowledge base expansion and ii. performance of the analysis on the limits of RWD information. Results show that once RWD is representative of biodiversity and spectral interferences, the samples that compose the boundary polygon in the feature space (Figure 4.2-2) can spawn SSD with information gain at the CovMs, generating a more complex eigenstructure than the original RWD. Such is the case with P, K, Ca, Fe, Mn, Zn, and Cu in grapevines. On the contrary, when RWD has no range representativeness in diversity and interference (biased and class-imbalanced dataset), such as in the apple leaves dataset, SSD does not provide an efficient information gain (Figure 4.2-7II a-d). Fe, Mn, and Zn quantification in apple leaves is an example of how nutrients with higher ranges better perform the apple dataset (Figure 4.2-6g-i).

DA is useful to determine the level of detail in RWD spectral information. The representativeness of RWD in complex biological samples, such as leaves, can theoretically be given by a minimum set of RWD samples that contain all the information necessary for SSD expansion of the feature space. RWDs with maximum diversity and variance ranges imposed by experimental designs planned for representativity are paramount for an optimal knowledge base. RWD vs SSD and RWD+SSD prediction analysis enables understanding which nutrients have a good representation of the

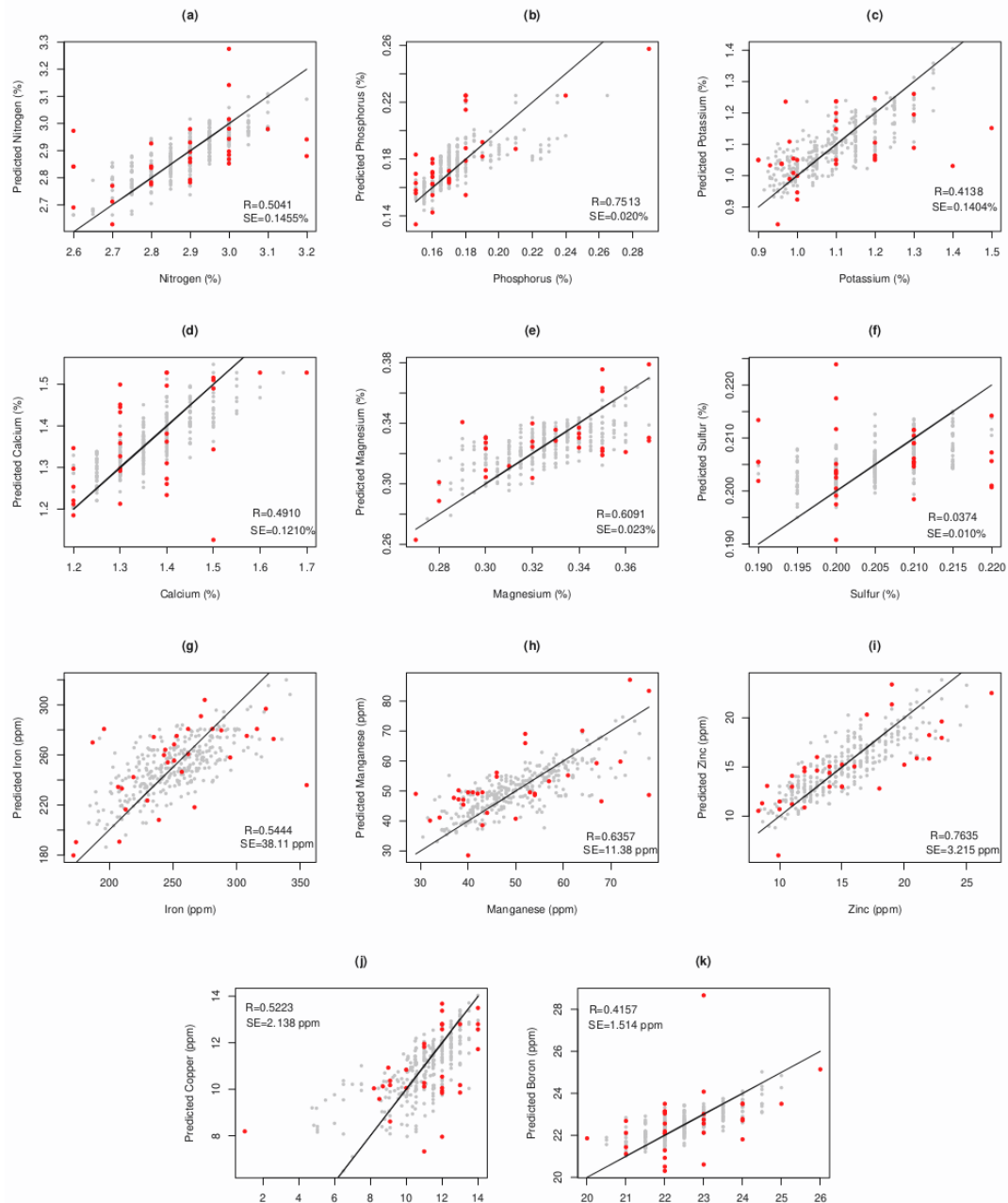
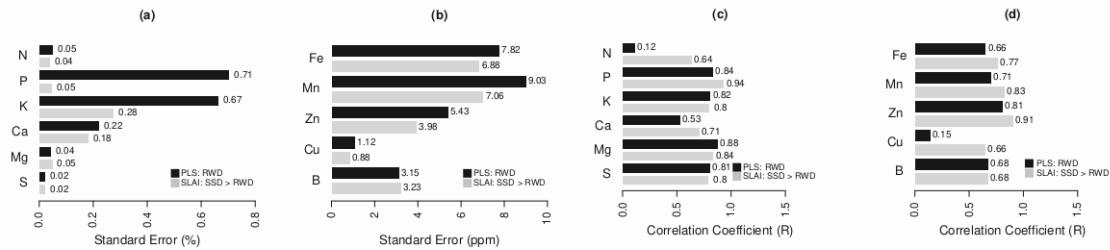


Figure 4.2-6. Nutrient quantification in apple tree leaves: (a) nitrogen, (b) phosphorus, (c) potassium, (d) calcium, (e) magnesium, (f) sulfur, (g) iron, (h) manganese, (i) zinc, (j) copper and (k) boron - where (•) are RWD samples and (•) are SSD samples for apple tree leaves.

spectral gradient equivalence to concentrations, providing information on which ranges and diversity should be expanded in a more complete study.

I.



II.

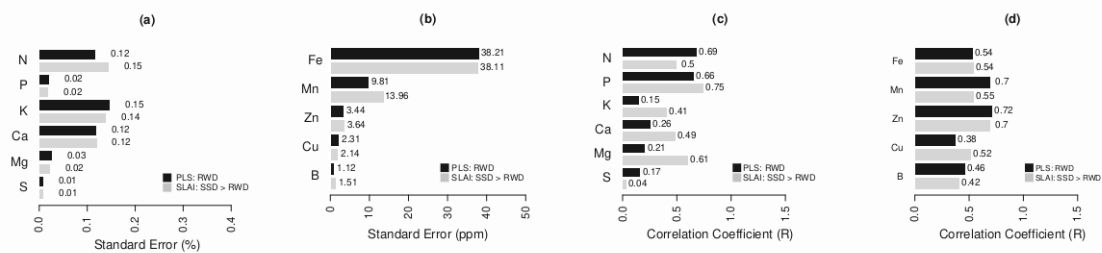


Figure 4.2-7. The standard error (SE) and correlation coefficient (R) benchmarks between PLS-RWD, and SLAI:SSD→RWD are for I. grapevine leaves and II—apple leaves.

DA is also relevant for non-dominant spectral information. Information detail in SSD leads to PLS models with many LVs, which try to account for the different information within the global covariance (X^tY). It also results in high-dimensional LocPLS models, meaning that even at a more restricted sub-set, information in the RWD is representative. Such information is also expected to allow the extraction of CovMs with a single LV, providing a direct and interpretable relationship between spectra and composition. One can, therefore, expect that if RWD is representative of the spectra-composition information in detail, SSD hybridisation provides the necessary information for creating quantitative models with a minimal number of samples in RWD, making it possible to extract CovMs, such as for P, K, Ca, Fe, Mn, Zn and Ca in grapevine leaves (Table S 4.2-7 - Supplementary Material).

Optimal RWD can be seen as the minimum number of samples representing the plant's bio-diversity and physiological states, represented by spectral and compositional spaces with equivalent information. SSD generated using an optimal RWD can unscramble local details as CovMs and sustain the quantification of macro- and micro-nutrients. The limitations on the range representativeness of the apple leaf dataset highlight the importance of efficient DoE in creating knowledge bases.

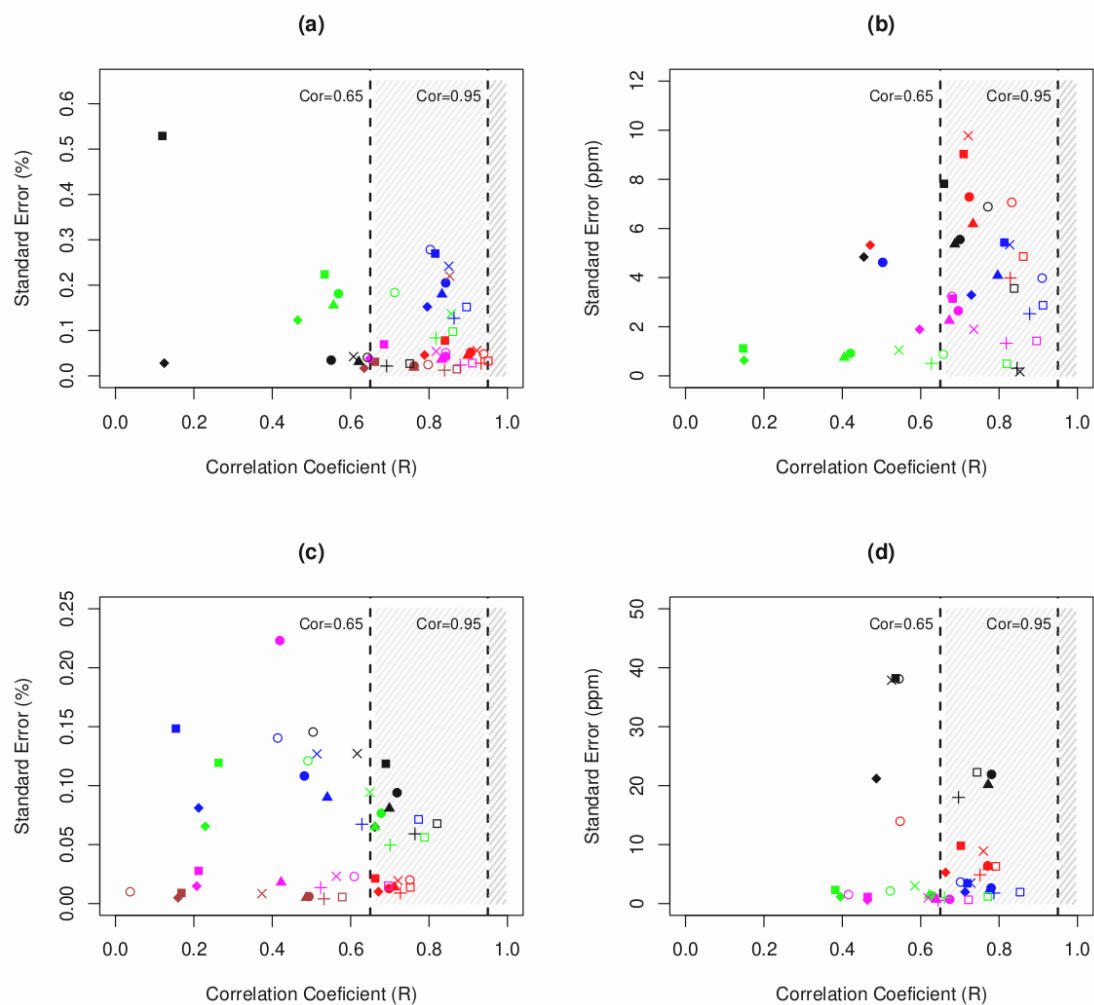


Figure 4.2-8. Correlation vs Standard errors Benchmarks: (a) grapevine leaf macronutrients; (b) grapevine micro-nutrients; (c) apple macro-nutrients; and (d) apple micro-nutrients; where: ■ PLS-RWD, • PLS-SSD→RWD, ♦ PLS-RWD→SSD, ▲ PLS-RWD+SSD, × LocPLS-SSD→RWD, + LocPLS-RWD+SSD, ° SLAI-SSD→ RWD, □ SLAI-RWD+SSD. Macronutrients: nitrogen (N) (■), phosphorus (P) (●), potassium (K) (♦), calcium (Ca) (▲), magnesium (Mg) (×), sulfur (S) (°). Micronutrients: iron (Fe) (■), manganese (Mn) (●), zinc (Zn) (♦), copper (Cu) (▲), boron (B) (×).

Field DoE in agriculture has low control over external variables (e.g. sunlight, temperature, rain, pests), and despite imposing variance over controllable variables (e.g. fertiliser composition, watering), effects can have different amplitudes. A representative sampling of RWD must become efficient to cover the range of variation and overcome the limitations of low-throughput laboratory methods (e.g., ICP-OES). New experimental designs for data scarcity must be developed, such as combining a high-throughput technique to identify ‘in-vivo’ extreme samples defining the RWD boundaries and interpolating samples to create a representative RWD knowledge-based expandable to local characteristics by SSD, which is especially relevant for non-dominant or lower concentration.

Vis-NIR spectroscopy can be used as a high-throughput technique for selecting samples that define the edge of the polygon in Figure 4.2-2 for quantification by ICP-OES. Samples inside the polygon can be used to characterise the natural correlation (for example, Figure 4.2-2) to provide higher local information representativeness, leading to an efficient use of SSD under data scarcity.

4.2.4. Conclusions

DA through hybridisation allows: i. knowledge base expansion and increasing model accuracy by reconstructing missing spectral and composition gradients in RWD, but also as ii—analytical tool for diagnosing RWD representativity and implementing new DoEs and sampling methods.

The results presented in this study highlight the importance of the RWD representativity. Agricultural datasets are obtained directly in the field, and even with a DoE for fertilisation differences, the effects on plants may depend on other uncontrollable factors, not only on physiology. Collecting scarce data for ICP-OES analysis can lead to the non-representativity of biodiversity - as presented in the apple dataset—an efficient data sampling design for achieving RWD representativity for effective SSD reconstruction. New experimental designs should take advantage of high-throughput ‘in-vivo’ technologies to determine representative RWD that can be expanded in the SSD, optimising the low-throughput laboratory resources and enabling physiology-based agriculture.

CRedit authorship contribution statement

Martins RC: Conceptualization, Investigation, Methodology, Software, Funding, writing - original draft; **Queirós C, Silva FM, Santos F, Barroso TG, Tosin R, Cunha M:** Investigation, methodology, validation, writing - review and editing; **Leão M, Damásio M, Martins P, Silvestre J:** Resources, Validation, Writing - Review & Editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that the research was conducted without any commercial or financial relationships that could be considered a potential conflict of interest.

Acknowledgements

This research was funded by national funds from Compete 2020 and PT2020 through POCI/FEDER project Smart Farm 4.0 - Soluções Inteligentes para uma agricultura sustentável, preditiva e autónoma (Ref. POCI-01-0247-FEDER-046078). Silva FM and Tosin R acknowledge Fundação para a Ciência e Tecnologia (FCT) PhD research grants Ref. DFA/BD/9136/2020 and SFRH/BD/145182/2019. Martins RC acknowledges Fundação para a Ciência e Tecnologia (FCT) research contract grant (CEEIND/017801/2018).

Supplementary Material

Tables

Table S 4.2-1. Fertilisation experimental design N-P levels (kg/ha) in Vines.

Block	Fertilisation (kg/ha)															
	1		2		3		4		5		6		7		8	
Variety	N	P	N	P	N	P	N	P	N	P	N	P	N	P	N	P
Touriga Nacional	0	0	15	30	30	60	60	0	60	60	30	30	60	30	30	0
Tempranillo I	0	0	15	30	30	60	60	0	60	60	30	30	60	30	30	0
Tempranillo II	0	0	15	30	30	60	60	0	60	60	30	30	60	30	30	0
Marselan	0	0	15	30	30	60	60	0	60	60	30	30	60	30	30	0

Table S 4.2-2. Apple tree fertilisation experimental design.

Fertiliser		Date of application	kg/ha	N	P	K
Farmer´s fertilisation	Entec 14-7-17	April 29th	340	47,6	23,8	58
	Orgevit 4-3-2,5+65%OM	May, 2nd	250	10	7,5	6,3
Prodl fertilisation	Labin 14-7-14	May, 3rd	200	28	14	28
	Nutricomplex 15-5-30	June, 17th	150	21	7,5	45
2xProdl fertilisation	Labin 14-7-14	May, 3rd	400	56	28	56
	Nutricomplex 15-5-30	June, 17th	300	42	15	90

Table S 4.2-3. Nutrient concentrations dataset characteristics of grapevine and apple tree.

	N (%)	P (%)	K (%)	Ca (%)	Mg (%)	S (%)	Fe (ppm)	Mn (ppm)	Zn (ppm)	Cu (ppm)	B (ppm)
Grapevine											
Minimum	0.49	0.24	0.94	1.4	0.18	0.06	4.9	11	19	2.9	27
Maximum	0.75	0.72	2.9	2.6	0.52	0.22	44	58	53	6.9	42
Variance	0.26	0.48	1.96	1.2	0.34	0.16	39.1	47	34	4	15
Apple Tree											
Minimum	0.25	0.14	0.9	1.2	0.27	0.17	172	29	8.3	1	19
Maximum	0.32	0.29	1.5	1.7	0.39	0.22	355	107	27	14	26
Variance	0.07	0.15	0.6	0.5	0.12	0.05	183	78	18.7	13	7

Table S 4.2-4. Leaf nutrient macronutrient quantification benchmarks in the bibliography.

Nutrient	Wavelength	Plant	Range	No Samples	R	R ²	SE	LV	Method	Source
N (%)	Vis	Tomato	2.5-5.5	117	0.9695	0.94	0.35	4 – 11	PLS	Ulissi et al. (2011)
	Vis-NIR	Olive	0.5-2.5	200	0.6928	0.48	1.11	7	PLS	Rotbart et al. (2013)
	Vis-NIR	Cotton	18.2-42.7dw	190	0.8426	0.71	NA	NA	PLS	Tarpley et al. (2000)
	Vis-NIR	Pear	15-43dw	450	0.9203	0.847	0.448	4 – 10	PLS	Jie et al. (2014)
	Vis-NIR	Wheat, Barley	1.0-7.0	1500	0.9055	0.82	0.64	12 – 14	PLS	Zerner and Parker (2019)
	Vis-NIR	Oil palm	1.8-3.2	42	0.9219	0.85	NA	27	PLS	Santoso et al. (2018)
	Vis-NIR	Orange	2.24 – 2.8	50	0.9091	0.8265	0.103	15	PLS	Menesatti et al. (2010)
	Vis-NIR	Persimmon	1.06-2.82	687	0.8944	0.8	0.16	6	PLS	Acosta et al. (2023)
	NIR	Sugarcane	NA	900	0.9945	0.989	0.012	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	1.53-5.49	115	0.7338	0.5385	0.29	NA	PLS	Prananto et al. (2021)
	Vis-NIR	Eucalyptus	2.36 – 18.95	45	0.9746	0.95	1.26	6	PLS	Oliveira and Santana (2020)
	NIR	Citrus	1.07 – 3.86	217	0.9949	0.99	0.06	11	PLS	Galvez-Sola et al. (2015)
P (%)	Vis-NIR	Orange	0.091 – 0.224	50	0.6549	0.429	0.018	9	PLS	Menesatti et al. (2010)
	Vis-NIR	Oil palm	0.16 – 0.20	42	0.8246	0.68	NA	12	PLS	Santoso et al. (2018)
	Vis-NIR	Persimmon	0.04 – 0.30	687	0.7874	0.62	0.01	5	PLS	Acosta et al. (2023)
	NIR	Sugarcane	NA	900	0.9940	0.988	0.008	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	0.14-0.87	115	NA	NA	0.09	NA	PLS	Prananto et al. (2021)
	Vis-NIR	Eucalyptus	0.18 – 1.24	45	0.9643	0.93	0.09	5	PLS	Oliveira and Santana (2020)
K (%)	Vis-NIR	Orange	0.346 – 1.730	50	0.9910	0.9821	0.057	15	PLS	Menesatti et al. (2010)
	Vis-NIR	Oil palm	0.86 – 1.49	42	0.6557	0.43	NA	5	PLS	Santoso et al. (2018)
	Vis-NIR	Persimmon	0.83 – 2.96	687	0.8185	0.67	0.83	10	PLS	Acosta et al. (2023)
	NIR	Sugarcane	NA	900	0.9894	0.979	0.013	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	0.69-4.23	115	NA	NA	0.37	NA	PLS	Prananto et al. (2021); Prananto et al. (2020)
	Vis-NIR	Eucalyptus	1.95 – 12.46	45	0.8718	0.76	1.3	11	PLS	Oliveira and Santana (2020)
Ca (%)	NIR	Citrus	0.42 – 1.92	217	0.9380	0.88	0.12	17	PLS	Galvez-Sola et al. (2015)
	Vis-NIR	Orange	3.180 – 10.30	50	0.9731	0.947	0.304	7	PLS	Menesatti et al. (2010)
	Vis-NIR	Oil palm	0.42 – 0.91	42	0.6856	0.47	NA	22	PLS	Santoso et al. (2018)
	Vis-NIR	Persimmon	0.30 – 6.52	687	0.7348	0.54	0.77	12	PLS	Acosta et al. (2023)
	NIR	Sugarcane	NA	900	0.9854	0.974	0.009	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	2.04-7.26	115	0.7280	NA	0.53	NA	PLS	Prananto et al. (2021); Prananto et al. (2020)
Mg (%)	Vis-NIR	Eucalyptus	3.45 – 19.71	45	0.9000	0.81	1.8	12	PLS	Oliveira and Santana (2020)
	NIR	Citrus	0.89 – 6.77	217	0.9899	0.98	0.25	13	PLS	Galvez-Sola et al. (2015)
	Vis-NIR	Orange	0.218 – 0.927	50	0.9716	0.944	0.048	13	PLS	Menesatti et al. (2010)
	Vis-NIR	Oil palm	0.16 – 0.39	42	0.80	0.64	NA	12	PLS	Santoso et al. (2018)
	Vis-NIR	Orange	6.69 – 42.30	50	0.9617	0.925	1.637	5	PLS	Menesatti et al. (2010)
	Vis-NIR	Persimmon	0.17 – 0.95	687	0.7616	0.58	0.12	11	PLS	Acosta et al. (2023)
S (%)	NIR	Sugarcane	NA	900	0.9879	0.976	0.005	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	0.37-1.38	115	NA	NA	0.09	NA	PLS	
	Vis-NIR	Eucalyptus	1.03 – 3.00	45	0.4690	0.22	0.34	11	PLS	Oliveira and Santana (2020)
	NIR	Citrus	0.14 – 0.80	217	0.9434	0.89	0.06	6	PLS	Galvez-Sola et al. (2015)
	Vis-NIR	Persimmon	0.11 – 0.29	687	0.6082	0.37	0.03	12	PLS	Acosta et al. (2023)
	NIR	Cotton	0.57-3.09	115	NA	NA	0.32	NA	PLS	Prananto et al. (2021); Prananto et al. (2020)
	Vis-NIR	Eucalyptus	0.93 – 2.02	45	0.9219	0.85	0.1	16	PLS	Oliveira and Santana (2020)

NA - Not Available

Table S 4.2-5. Leaf nutrient micronutrient quantification benchmarks in bibliography.

Nutrient	Wavelength	Plant	Range	No Samples	R	R ²	SE	LV	Method	Source
Fe (ppm)	Vis-NIR	Orange	66.9 – 198.00	50	0.9576	0.917	6.054	10	PLS	Menesatti et al. (2010)
	Vis-NIR	Persimmon	13.22 – 81.77	687	0.5831	0.34	10.32	12	PLS	Acosta et al. (2023)
	NIR	Sugarcane	NA	900	0.9945	0.989	0.15	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	73.65-900.8	115	NA	NA	76.29	NA	PLS	Prananto et al. (2021); Prananto et al. (2020)
	Vis-NIR	Eucalyptus	24.28 – 182.52	45	0.5196	0.35	25	11	PLS	Oliveira and Santana (2020)
Mn (ppm)	NIR	Citrus	39.54 – 636.75	217	0.8718	0.76	60.4	15	PLS	Galvez-Sola et al. (2015)
	Vis-NIR	Persimmon	32.81 – 295.02	687	0.4690	0.22	49.85	9	PLS	Acosta et al. (2023)
	NIR	Sugarcane	NA	900	0.9879	0.976	0.034	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	27.5-367.0	115	NA	NA	45.59	NA	PLS	Prananto et al. (2021)
	Vis-NIR	Eucalyptus	288.97 – 4066.0	45	0.8888	0.79	407.5	7	PLS	Oliveira and Santana (2020)
Zn (ppm)	NIR	Citrus	18.69 – 70.48	217	0.7280	0.53	5.55	12	PLS	Galvez-Sola et al. (2015)
	Vis-NIR	Orange	6.390 – 24.80	50	0.9428	0.889	0.972	8	PLS	Menesatti et al. (2010)
	Vis-NIR	Oil palm	7.98 – 26.97	42	0.4898	0.24	NA	8	PLS	Santoso et al. (2018)
	Vis-NIR	Persimmon	2.02 – 15.68	687	0.4472	0.2	3.37	12	PLS	Acosta et al. (2023)
	NIR	Sugarcane	NA	900	0.9894	0.979	0.002	NA	PLS	Yarce and Rojas (2012)
Cu (ppm)	NIR	Cotton	8.14-64.62	115	NA	NA	7.02	NA	PLS	Prananto et al. (2021); Prananto et al. (2020)
	Vis-NIR	Eucalyptus	7.83 – 35.83	45	0.5568	0.31	4.67	2	PLS	Oliveira and Santana (2020)
	NIR	Citrus	9.01 – 43.92	217	0.9380	0.88	3.34	17	PLS	Galvez-Sola et al. (2015)
	Vis-NIR	Oil palm	1.52 – 9.43	42	0.7870	0.62	NA	17	PLS	Santoso et al. (2018)
	Vis-NIR	Persimmon	1.33-8.97	687	0.8306	0.69	0.93	6	PLS	Acosta et al. (2023)
B (ppm)	NIR	Sugarcane	NA	900	0.9935	0.987	0.015	NA	PLS	Yarce and Rojas (2012)
	NIR	Cotton	1.92-12.5	115	NA	NA	1.35	NA	PLS	Prananto et al. (2021); Prananto et al. (2020)
	Vis-NIR	Eucalyptus	7.60 – 13.10	45	0.8832	0.78	0.63	12	PLS	Oliveira and Santana (2020)
	NIR	Citrus	0.83 – 11.24	217	0.60	0.36	1.47	4	PLS	Galvez-Sola et al. (2015)
	Vis-NIR	Oil palm	7.52 – 25.82	42	0.7550	0.57	NA	11	PLS	Santoso et al. (2018)
B (ppm)	Vis-NIR	Persimmon	12.18 – 102.10	687	0.4899	0.24	9.59	11	PLS	Acosta et al. (2023)
	NIR	Cotton	3.15-185.88	115	NA	NA	17.44	NA	PLS	Prananto et al. (2021); Prananto et al. (2020)
	Vis-NIR	Eucalyptus	25.30 – 171.40	45	0.8246	0.68	22.66	7	PLS	Oliveira and Santana (2020)
	NIR	Citrus	24.22 – 509.93	217	0.6556	0.47	53.2	6	PLS	Galvez-Sola et al. (2015)

NA - not available

Table S 4.2-6. Results benchmarks for PLS, LocPLS and SLAI for real-world data and a combination of real-world data and synthetic samples data in grapevine and apple leaves.

Model	Element	LV	Grapevine leaves				LV	Apple leaves			
			R	R ²	SE(ppm)	DL(ppm)		R	R ²	SE(ppm)	DL(ppm)
PLS RWD ⁽¹⁾	N (%)	1	0.1196	0.014	0.5290	5.29	3	0.6897	0.4757	0.1186	1.186
	P (%)	3	0.8405	0.7065	0.0780	0.78	4	0.6626	0.4390	0.0213	0.2130
	K (%)	2	0.8158	0.6658	0.2697	2.697	1	0.1541	0.0240	0.1484	1.484
	Ca (%)	2	0.5335	0.2876	0.2237	2.237	1	0.2629	0.0690	0.1194	1.194
	Mg (%)	2	0.6851	0.4694	0.0695	0.6950	1	0.2121	0.04497	0.0277	0.2770
	S (%)	2	0.6620	0.4382	0.031	0.3100	2	0.1683	0.02832	0.0090	0.090
	Fe (ppm)	3	0.6593	0.4346	7.818	78.18	4	0.5358	0.2871	38.21	382.1
	Mn (ppm)	3	0.7095	0.5033	9.028	90.28	3	0.7023	0.4932	9.810	98.10
	Zn (ppm)	3	0.8134	0.6624	5.428	54.28	3	0.7190	0.5170	3.440	34.40
	Cu (ppm)	1	0.147	0.022	1.117	11.17	1	0.3820	0.1459	2.310	23.10
	B (ppm)	3	0.6817	0.4646	3.145	31.45	3	0.4646	0.2158	1.120	11.20
	N (%)	6	0.6207	0.3853	0.0311	0.3110	4	0.6989	0.4886	0.0807	0.8070
	P (%)	6	0.8991	0.8084	0.0449	0.4490	6	0.7117	0.566	0.0139	0.1390
	K (%)	5	0.8329	0.6938	0.1799	1.799	5	0.5404	0.292	0.09	0.9000
PLS RWD+SSD ⁽²⁾	Ca (%)	3	0.5562	0.3094	0.1556	1.556	5	0.662	0.4382	0.065	0.6501
	Mg (%)	6	0.8321	0.6924	0.0364	0.3640	4	0.4225	0.1785	0.018	0.1800
	S (%)	5	0.7614	0.5797	0.0186	0.1860	6	0.4865	0.2367	0.005	0.0500
	Fe (ppm)	6	0.6869	0.4719	5.361	53.6	8	0.7721	0.5961	20.15	201.50
	Mn (ppm)	4	0.7334	0.538	6.183	61.83	5	0.7717	0.5955	6.252	62.52
	Zn (ppm)	4	0.7964	0.6343	4.083	40.83	5	0.7756	0.6015	2.24	22.40
	Cu (ppm)	3	0.4048	0.1639	0.7456	7.456	5	0.6221	0.3871	1.426	14.26
	B (ppm)	5	0.6729	0.4528	2.252	22.52	5	0.6394	0.4089	0.683	6.830
	N (%)	4	0.6918	0.4787	0.0217	0.2174	4	0.6989	0.4886	0.0807	0.8070
	P (%)	4	0.9326	0.8699	0.0278	0.2780	6	0.7117	0.566	0.0139	0.1390
	K (%)	3	0.8639	0.7455	0.1269	1.269	5	0.5404	0.292	0.090	0.9000
	Ca (%)	3	0.8178	0.6688	0.0837	0.8370	5	0.662	0.4382	0.0650	0.6500
	Mg (%)	4	0.8794	0.7734	0.024	0.2401	4	0.4225	0.1785	0.0180	0.1800
	S (%)	3	0.8395	0.7046	0.0125	0.1250	6	0.4865	0.2367	0.005	0.0500
LocPLS RWD+SSD ⁽²⁾	Fe (ppm)	4	0.8458	0.7154	0.3114	3.114	8	0.7721	0.5961	20.15	201.5
	Mn (ppm)	2	0.8284	0.6731	3.985	39.85	5	0.7717	0.5955	6.252	62.52
	Zn (ppm)	3	0.8781	0.7711	2.527	25.27	5	0.7756	0.6015	2.240	22.4
	Cu (ppm)	2	0.6275	0.3938	0.5072	5.072	5	0.6221	0.3871	1.426	14.26
	B (ppm)	3	0.8183	0.6496	1.325	13.25	5	0.6394	0.4089	0.6830	6.830
	N (%)	1	0.7507	0.5635	0.0271	0.2710	1	0.8206	0.6734	0.0678	0.6780
	P (%)	1	0.9511	0.9046	0.033	0.33	1	0.7523	0.5654	0.0137	0.1370
	K (%)	1	0.8956	0.8022	0.1519	1.519	1	0.7733	0.5975	0.07136	0.7136
	Ca (%)	1	0.8605	0.7544	0.0975	0.975	1	0.7884	0.6217	0.0562	0.5620
	Mg (%)	1	0.9103	0.8287	0.0282	0.282	1	0.6959	0.4843	0.0151	0.1510
	S (%)	1	0.8712	0.759	0.0148	0.1480	1	0.5786	0.3347	0.0055	0.0550
	Fe (ppm)	1	0.8383	0.7891	3.557	35.57	1	0.7438	0.5533	22.26	222.6
	Mn (ppm)	1	0.8618	0.7415	4.858	48.58	1	0.7915	0.6265	6.310	63.10
	Zn (ppm)	1	0.912	0.8318	2.876	28.76	1	0.8535	0.7286	1.941	19.41
SLAI RWD+SSD ⁽²⁾	Cu (ppm)	1	0.8197	0.6715	0.4909	4.909	1	0.7716	0.5954	1.218	12.18
	B (ppm)	1	0.8957	0.8023	1.422	14.22	1	0.7219	0.5213	0.6456	6.456

⁽¹⁾ Cross-validation optimisation and hold-out metrics calculation using RWD.

⁽²⁾ Cross-validation optimisation and hold-out metrics calculation using RWD and SSD.

Table S 4.2-7. Results benchmark for real-world data vs synthetic samples data for PLS, Loc-PLS and SLAI models in grapevine leaves.

Model	Element	LV	Training Set				Test Set			
			R	R ²	SE	DL	R	R ²	SE	DL
PLS SSD→RWD ⁽¹⁾	N (%)	6	0.6377	0.4067	0.0290	0.2901	0.5503	0.3029	0.0346	0.3460
	P (%)	6	0.8983	0.8070	0.0428	0.428	0.9062	0.8105	0.0520	0.5200
	K (%)	5	0.8300	0.6891	0.2050	2.05	0.8424	0.7096	0.2050	2.050
	Ca (%)	3	0.5577	0.3111	0.1475	1.475	0.5687	0.3235	0.1812	1.812
	Mg (%)	6	0.8274	0.6847	0.0346	0.3460	0.8425	0.7097	0.0426	0.4260
	S (%)	5	0.7566	0.5725	0.0177	0.1770	0.7618	0.5804	0.0219	0.2190
	Fe (ppm)	6	0.6836	0.4679	5.109	51.09	0.7001	0.4902	5.550	55.50
	Mn (ppm)	4	0.7327	0.5368	5.876	58.76	0.7239	0.5241	7.284	72.84
	Zn (ppm)	4	0.7939	0.6304	3.856	38.56	0.5030	0.6449	4.619	46.19
	Cu (ppm)	3	0.4063	0.1651	0.7100	7.100	0.4209	0.1623	0.9155	9.155
	B (ppm)	5	0.6659	0.4435	2.154	21.54	0.6959	0.4843	2.648	26.48
	N (%)	1	0.1196	0.014	0.5290	5.29	0.1243	0.015	0.0280	0.2800
	P (%)	3	0.8405	0.7065	0.0780	0.78	0.7886	0.6219	0.0460	0.4600
	K (%)	2	0.8158	0.6658	0.2697	2.697	0.7955	0.6328	0.1523	1.523
PLS RWD→SSD ⁽²⁾	Ca (%)	2	0.5335	0.2876	0.2237	2.237	0.4649	0.2161	0.1230	1.230
	Mg (%)	2	0.6851	0.4694	0.0695	0.6950	0.6484	0.4204	0.0382	0.3820
	S (%)	2	0.6620	0.4382	0.031	0.3100	0.6345	0.4026	0.0169	0.1690
	Fe (ppm)	3	0.6593	0.4346	7.818	78.18	0.4549	0.2069	4.838	48.38
	Mn (ppm)	3	0.7095	0.5033	9.028	90.28	0.4712	0.2221	5.322	53.22
	Zn (ppm)	3	0.8134	0.6624	5.428	54.28	0.7293	0.5320	3.290	32.90
	Cu (ppm)	1	0.147	0.0220	1.117	11.17	0.1492	0.0220	0.6335	6.335
	B (ppm)	3	0.6817	0.4646	3.145	31.45	0.5972	0.3567	1.893	18.93
	N (%)	3	0.5414	0.2994	0.0314	0.3140	0.6073	0.3669	0.0425	0.4250
	P (%)	4	0.9187	0.8440	0.0384	0.384	0.9221	0.8504	0.0560	0.5600
	K (%)	4	0.8630	0.7430	0.1555	1.555	0.8503	0.7231	0.2417	2.417
	Ca (%)	5	0.8234	0.6863	0.0990	0.9900	0.8565	0.7336	0.1371	1.371
	Mg (%)	3	0.8299	0.6889	0.0343	0.3430	0.8180	0.6691	0.054	0.5400
	S (%)	3	0.8201	0.6726	0.0155	0.1550	0.8528	0.7273	0.2200	2.200
LocPLS SSD→RWD ⁽¹⁾	Fe (ppm)	3	0.7975	0.6359	4.216	42.16	0.8532	0.7297	0.1671	1.671
	Mn (ppm)	3	0.8215	0.6748	4.915	49.15	0.7213	0.5203	9.778	97.78
	Zn (ppm)	3	0.8683	0.7541	3.142	31.42	0.8269	0.5838	5.338	53.38
	Cu (ppm)	3	0.6954	0.4836	0.5575	5.575	0.5445	0.2965	1.046	10.46
	B (ppm)	3	0.7642	0.5840	1.859	18.59	0.7357	0.5413	1.893	18.93
	N (%)	1	0.7394	0.5468	0.0253	0.2530	0.6428	0.4133	0.0406	0.406
	P (%)	1	0.9488	0.9002	0.0307	0.307	0.9394	0.8825	0.0484	0.484
	K (%)	1	0.8914	0.7996	0.1397	1.397	0.8033	0.6452	0.2783	2.783
	Ca (%)	1	0.8766	0.7684	0.0854	0.8540	0.7122	0.5073	0.1836	1.836
	Mg (%)	1	0.9127	0.833	0.0251	0.2510	0.8423	0.7095	0.0510	0.5100
	S (%)	1	0.8818	0.7775	0.0128	0.1280	0.7976	0.6362	0.0248	0.248
	Fe (ppm)	1	0.8738	0.7633	3.400	34.00	0.7715	0.5952	6.884	68.84
	Mn (ppm)	1	0.5500	0.7322	4.469	44.69	0.8323	0.6927	7.062	70.62
	Zn (ppm)	1	0.9054	0.8197	2.690	26.90	0.9098	0.8278	3.978	39.78
SLAI SSD→RWD ⁽²⁾	Cu (ppm)	1	0.8257	0.6818	0.4376	4.376	0.6577	0.4326	0.8753	8.753
	B (ppm)	1	0.9028	0.8150	0.2400	2.400	0.6793	0.4612	3.233	32.33

⁽¹⁾ Training set: Synthetic samples data; Test set: Real-world data.

⁽²⁾ Training set: Real-world data; Test set: Synthetic samples data.

Table S 4.2-8. Results benchmark for real-world data vs synthetic samples data for PLS, Loc-PLS and SLAI models in apple leaves.

Model	Element	Training Set				Test Set				
		LV	R	R ²	SE	DL	R	R ²	SE	DL
PLS SSD→RWD ⁽¹⁾	N (%)	4	0.6949	0.4830	0.0765	0.765	0.7184	0.5162	0.0940	0.9400
	P (%)	6	0.7129	0.5082	0.01314	0.1314	0.6979	0.487	0.0127	0.1270
	K (%)	5	0.5445	0.2965	0.0853	0.853	0.4819	0.2322	0.1082	1.082
	Ca (%)	5	0.6473	0.4190	0.0628	0.628	0.6781	0.4598	0.0766	0.7660
	Mg (%)	4	0.4220	0.1787	0.01721	0.1721	0.4192	0.1758	0.2230	2.230
	S (%)	6	0.4826	0.2306	0.00533	0.0533	0.4934	0.2434	0.0059	0.0594
	Fe (ppm)	8	0.7696	0.5922	19.16	191.6	0.7806	0.6094	21.91	219.1
	Mn (ppm)	5	0.7675	0.5891	5.989	59.89	0.7708	0.5941	6.395	63.95
	Zn (ppm)	5	0.7723	0.5965	2.144	21.44	0.7796	0.6078	2.643	26.43
	Cu (ppm)	5	0.6182	0.3823	1.367	13.67	0.6301	0.3971	1.299	12.99
	B (ppm)	5	0.6418	0.4119	0.6445	6.445	0.6737	0.4539	0.7209	7.209
	N (%)	3	0.6897	0.4757	0.1186	1.186	0.6615	0.4375	0.0650	0.6500
	P (%)	4	0.6626	0.4390	0.0213	0.2130	0.6708	0.4499	0.0101	0.1010
	K (%)	1	0.1541	0.0240	0.1484	1.484	0.2122	0.0450	0.0811	0.8110
PLS RWD→SSD ⁽²⁾	Ca (%)	1	0.2629	0.0690	0.1194	1.194	0.2289	0.0520	0.0655	0.6550
	Mg (%)	1	0.2121	0.04497	0.0277	0.2770	0.2072	0.0429	0.0148	0.1480
	S (%)	2	0.1683	0.02832	0.0090	0.090	0.1596	0.0255	0.0049	0.0490
	Fe (ppm)	4	0.5358	0.2871	38.21	382.1	0.4868	0.2369	21.20	212.0
	Mn (ppm)	3	0.7023	0.4932	9.810	98.10	0.6634	0.4401	5.280	52.80
	Zn (ppm)	3	0.7190	0.5170	3.440	34.40	0.7133	0.5088	1.960	19.60
	Cu (ppm)	1	0.3820	0.1459	2.310	23.10	0.3952	0.1561	1.151	11.51
	B (ppm)	3	0.4646	0.2158	1.120	11.20	0.464	0.2153	0.5931	5.931
	N (%)	4	0.7202	0.5187	0.07373	0.7373	0.6166	0.3808	0.1272	1.272
	P (%)	4	0.7166	0.5136	0.13	1.3	0.7207	0.5196	0.0194	0.194
	K (%)	3	0.6313	0.3986	0.0787	0.787	0.5139	0.2641	0.1269	1.269
	Ca (%)	3	0.7293	0.5318	0.0563	0.563	0.6489	0.421	0.0943	0.943
	Mg (%)	3	0.5353	0.2865	0.16	1.6	0.5636	0.3176	0.0231	0.2310
	S (%)	4	0.4593	0.2109	0.0053	0.053	0.3737	0.1397	0.0084	0.084
LocPLS SSD→RWD ⁽¹⁾	Fe (ppm)	4	0.5749	0.3305	24.5	245	0.5258	0.2765	37.89	378.9
	Mn (ppm)	3	0.7387	0.5458	6.286	62.86	0.7599	0.5775	8.892	88.92
	Zn (ppm)	3	0.7956	0.633	2.041	20.41	0.7275	0.5293	3.493	34.93
	Cu (ppm)	4	0.6555	0.4296	1.311	13.11	0.5851	0.3429	3.001	30.01
	B (ppm)	4	0.6052	0.3663	0.6679	6.679	0.6191	0.3833	0.9808	9.808
	N (%)	1	0.8293	0.6877	0.0590	0.59	0.5041	0.2542	0.1455	1.455
	P (%)	1	0.7528	0.5667	0.0123	0.1230	0.7513	0.5645	0.0200	0.2000
	K (%)	1	0.8040	0.6465	0.0604	0.6040	0.4138	0.1692	0.1404	1.404
	Ca (%)	1	0.7981	0.6368	0.0495	0.4950	0.4910	0.2410	0.1210	1.210
	Mg (%)	1	0.7002	0.4902	0.0135	0.1350	0.6091	0.3710	0.0229	0.229
	S (%)	1	0.6273	0.3935	0.004	0.040	0.0374	0.0010	0.0100	0.100
	Fe (ppm)	1	0.7081	0.5015	21.15	211.5	0.5444	0.2963	38.11	381.1
	Mn (ppm)	1	0.8031	0.6445	5.558	55.58	0.6357	0.4042	11.38	113.8
	Zn (ppm)	1	0.8475	0.7184	1.788	17.88	0.7635	0.5829	3.641	36.41
SLAI SSD→RWD ⁽²⁾	Cu (ppm)	1	0.7153	0.5117	1.214	12.14	0.5223	0.2728	2.138	21.38
	B (ppm)	1	0.7574	0.5737	0.547	5.47	0.4157	0.1728	1.514	15.14

⁽¹⁾ Training set: Synthetic samples data; Test set: Real-world data.

⁽²⁾ Training set: Real-world data; Test set: Synthetic samples data.

CHAPTER 5

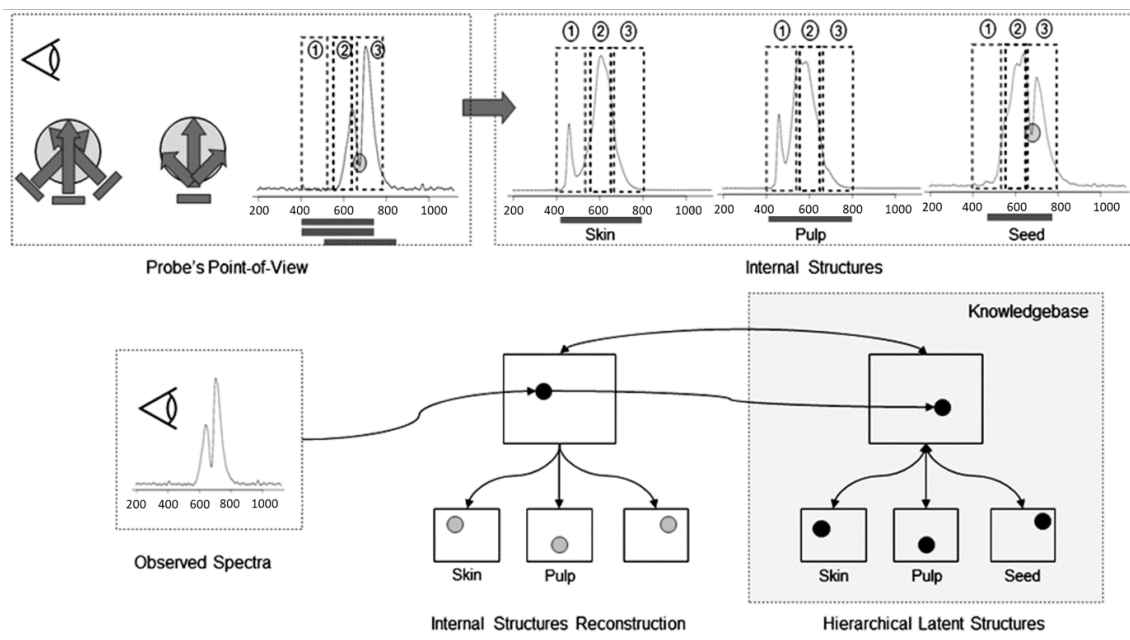
- PATENT: TOMOGRAPHY-LIKE METHOD FOR INTERNAL FRUIT TISSUE ASSESSMENT.
- CASE STUDY IV: TOMATO CASE STUDY: IDENTIFYING TISSUE CHANGES DURING MATURATION.
- CASE STUDY V: GRAPE CASE STUDY: COMPARING CULTIVARS, TRACKING QUALITY CHANGES DURING MATURATION.

5.1. Patent: Tomography-like method for internal fruit tissue assessment.

Martins, R.C., Santos, F., Cunha, M., Monteiro-Silva, F., **Tosin, R.**, Magalhães, S., and Pereira, M.R. (2023). Method and device for non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues. PCT/EP2022/088102. Available:

https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2023126532&_cid=P12-LK2EKH-19952-1.

1



Graphical abstract. 5.1-1. Graphical abstract of the patent (Martins et al., 2023).

¹ NOTE: The patent's content available in this thesis regards only for the specific bi-directional reconstruction of internal tissues. However, in the full document a part related to early detection in disease is also presented.

Method and device for non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues

Inventors

Martins, R. C., Santos, F., Cunha, M., Monteiro-Silva, F., **Tosin, R.**, Magalhães, S. and Pereira, M. R.

Applicants

INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia E Ciencia; Campus da Feup, Rua Dr Roberto Frias, 378, 4200-465 Porto (PT).
Universidade Do Porto; Praça Gomes Teixeira, S/N, 4099-002 Porto (PT).

Abstract

The present document discloses a method and device for a non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues. In particular, a computer-implemented method for non-invasive tomographic metabolite characterisation of a vegetable or animal sample to be characterised, the sample comprising a plurality of differentiated tissues from a set of previously-characterised differentiated tissues, by using a multidimensional latent structure model of differentiated tissues.

5.1.1. Description

5.1.1.1. Technical field

The present disclosure relates to a method and device for non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues.

5.1.1.2. Background

Mainstream analytical chemistry and biosensor approaches to spectroscopy information complexity are to use physical separation unit operations (Arshavsky-Graham & Segal, 2022; Nishat et al., 2021), reaction- specific labelling (Zhou et al., 2021; Jiang et al., 2021), or more recently label-free reaction specificity (Prakash et al., 2021). State-of-the-art chemometrics makes use of macroscopic hyperspectral or multispectral imaging to relate surface characteristics of plant tissues (e.g., fruits or leaves) to pigmentation and

disease symptoms (Liu & Wang, 2021). The previous art methods are cumbersome to be used *in vivo*, particularly in an open field context:

- i. Biosensors need destructive sampling, consumables, provide limited sampling or even need sample preparation, all of which make them impractical to be implemented on robotic platforms for high-throughput monitoring of vast agricultural areas;
- ii. Hyperspectral cameras capture a picture in a wavelength interval of light reflected from the plant surface. The reflected light is non-absorbed light; therefore, it carries little information about the sample composition, with insufficient information relating to omics quantitative information.

These facts are disclosed in order to illustrate the technical problem addressed by the present disclosure.

5.1.2. General description

The present document discloses a computer-implemented method for non-invasive tomographic metabolite characterisation of a vegetable or animal sample to be characterised, the sample comprising a plurality of differentiated tissues from a set of previously-characterised differentiated tissues, by using a multidimensional latent structure model of differentiated tissues; acquiring hyperspectral spectrum or spectra for each differentiated tissue sample; analysing the same tissue samples for the metabolites to be characterised; calculating the latent structure model from the acquired spectra and analysed metabolites; the method comprising the steps of: spatially scanning the sample to be characterised with a hyperspectral point-of-measurement, POM, optical device and a spectrometer; acquiring hyperspectral spectrum or spectra from each of a plurality of spatial positions and/or orientations of the POM optical device and spectrometer in respect of the sample to be characterised; using the previously obtained model to deconvolute the metabolites to be characterised for each of the differentiated tissues from the acquired spectra.

In an embodiment, the scanning of tissue samples belonging to each tissue of the set with a hyperspectral microscope and spectrometer is microscopic, and the spatially scanning the sample to be characterised with a hyperspectral POM optical device and spectrometer is macroscopic.

In an embodiment, the spatially scanning comprises varying one or more of the following to obtain a spatial diversity of the POM spatial scanning: displacement of the POM optical device; point-of-view angle of the POM optical device; different scales of magnification

of the POM optical device; moving the POM optical device lines of sight to cause a parallax effect.

In an embodiment, the multidimensional latent structure model is a hierarchical latent structure model.

In an embodiment, the hierarchical latent structure model comprises a latent variable model comprising a latent variable super-space and corresponding latent variable sub-spaces.

In an embodiment, the hierarchical relationship of the hierarchical latent structure model comprises super-space and corresponding sub-spaces: a latent sub- space for spectral data of each of the differentiated tissues; a super-space for each of the metabolites to be characterised, translating the combination of hyperspectral data to the metabolites to be characterised.

In an embodiment, the sample to be characterised is a three-dimensional tissue structure or organ.

In an embodiment, the metabolites to be characterised comprise pigments, tannins, photosynthesis, energy storage, nutrient uptake, necrosis, infection probability, or combinations thereof.

In an embodiment, the sample to be characterised is vegetable sample and the differentiated tissues comprise skin, pulp, and seed tissue.

In an embodiment, the scanning of tissue samples belonging to each differentiated tissue is carried out along a sampling grid.

In an embodiment, the scanning of tissue samples belonging to each differentiated tissue is further carried out along discrete sampling depths.

In an embodiment, the spectra are UV-VIS-NIR.

It is further disclosed a device for non-invasive tomographic metabolite characterisation of a vegetable or animal sample to be characterised, the sample comprising a plurality of differentiated tissues from a set of previously-characterised differentiated tissues, including a computer processor and a non-transitory computer- readable medium comprising a multidimensional latent structure model of differentiated tissues, which was previously obtained by: scanning tissue samples belonging to each differentiated tissue of the set with a hyperspectral spectrometer; acquiring hyperspectral spectrum or spectra for differentiated each tissue sample from the spectrometer; analysing the same tissue samples for the metabolites to be characterised; calculating the latent structure model from the acquired spectra and analysed metabolites; wherein the computer processor is configured for: spatially scanning the sample to be characterised with a hyperspectral point-of-measurement, POM, optical device and spectrometer; acquiring hyperspectral

spectra from a plurality of spatial positions and/or orientations of the POM optical device and spectrometer in respect of the sample to be characterised; using the previously obtained model to deconvolute the metabolites to be characterised for each of the differentiated tissues from the acquired spectra.

It is also presented a non-transitory computer-readable medium comprising computer program instructions for implementing a device for non-invasive tomographic metabolite characterisation of a vegetable or animal sample to be characterised, which when executed by a processor, cause the processor to carry out the disclosed method.

One of the present disclosure aims is to develop a non-invasive spectroscopy tomography system for compositional imaging of plant and animal tissues. This technology is demonstrated in application to precision agriculture and plant physiology. The present disclosure is based on image reconstruction using latent hierarchical information fusion methods to unscramble the observed spectroscopy signal into its different components of the observed tridimensional structures, particularly the different plant tissues being recorded, such as in grapes the skin, pulp, and seeds. Hierarchical relations exploit the parallax effect, which changes the spectral fingerprint depending upon the point-of-view angle of the spectroscopy probe. Reconstruction is performed by the hierarchical latent relationships between spectral patterns of the observed tissues. This principle is used to develop:

- i. macroscopic tomography - for measuring macroscopic properties of plant internal tissues with case studies of grape and tomato fruits (e.g., skin, pulp, and seeds);
- ii. Multi-scaled tomography results - exploiting the hierarchical parallax relationship between latent spaces of different scales of magnification to provide microscopic reconstructions based on macroscopic measurements.

Results show the feasibility of relationship reconstruction from latent structures, as well as, providing relationship between macroscopic and microscopic spectral data.

Spectral data carries both physical and chemical information. The observed spectrum of plant tissue is the superposition of all structures in the field of view, scattering, matrix effects, and multi-scaled interference. Being able to use these complex interactions is paramount to unscrambling the information present in each recorded spectrum. The complexity of spectral data has been the foremost hurdle to the existence of *in vivo*, reagent-less, and non-invasive chemical or metabolic diagnosis.

The macroscopic spectra are recorded and related to the hyperspectral image by their corresponding latent space.

The capacity of internal structures reconstruction, compositional, and disease diagnosis at very early stages centres molecular biology and plant physiology as the main driver of information in precision agriculture and plant biotechnology. The present disclosure bridges the information between *in vitro* laboratory and *in vivo* field studies by allowing high-throughput and multi-scaled monitoring of plants in their ecosystem without disrupting the interaction of plants with soil, climate, and ecosystem. Allowing information transfer between the field and laboratory tests is a critical enabling transformation in precision agriculture that allows evaluating to be much faster, thus promoting scientific advances in both phytopharmaceuticals and agricultural practices for early-stage and low-impact practices.

One of the main objectives of the disclosure is to provide a macroscopic, microscopic, and multi-scaled spectral reconstruction, compositional, and infection diagnosis through spectroscopy POM devices and advanced signal/information processing, demonstrating the feasibility of tomographic reconstruction by hierarchical latent structures methods.

5.1.3.Detailed description

Figure 5.1-1 presents a device, robotic platform and metabolic quantification according to the disclosure. The device comprises a spectroscopy Point-of- Measurement (POM) to record the spectral fingerprint of plant tissues. The POM can record the spectra of the internal structures of plant tissues, allowing the measurement of key metabolites representative of biochemical cycles and regulatory mechanisms, such as pigments, tannins, photosynthesis, energy storage, and nutrient uptake.

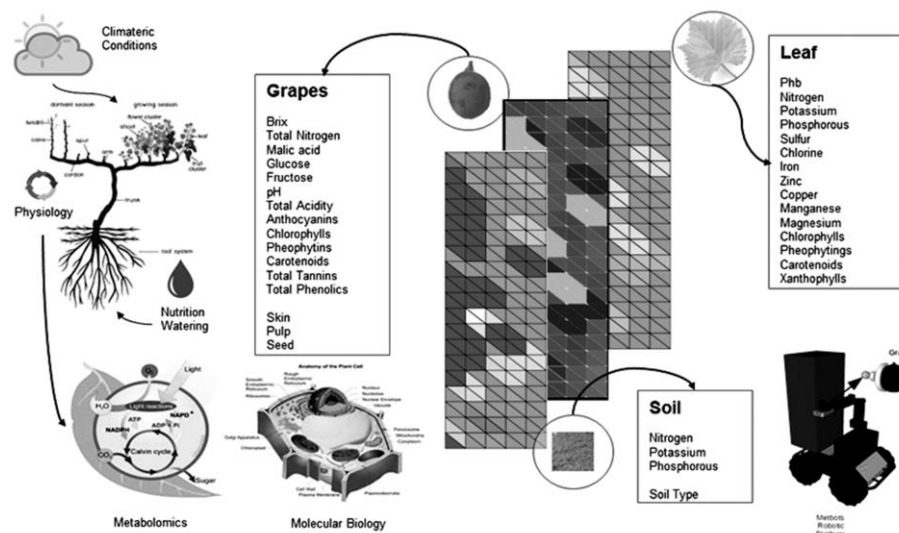


Figure 5.1-1. Schematic representation of an embodiment of a device, robotic platform and metabolic quantification according to the disclosure.

In an embodiment, the present disclosure is deployed/integrated on existing systems and devices, thus extending their technical capabilities. It enables the existing system to explore in detail plant internal structures and composition for a new *omics* approach to precision agriculture, where molecular biology and plant physiology are key enablers of new diagnosis and agricultural practices. The disclosure includes, Figure 5.1-2. the basic principles for macroscopic tomography reconstruction using latent variables. The method is based in two parts: i. spectral recording of samples at different point-of-view; and ii. establishment of internal structures spectral database.

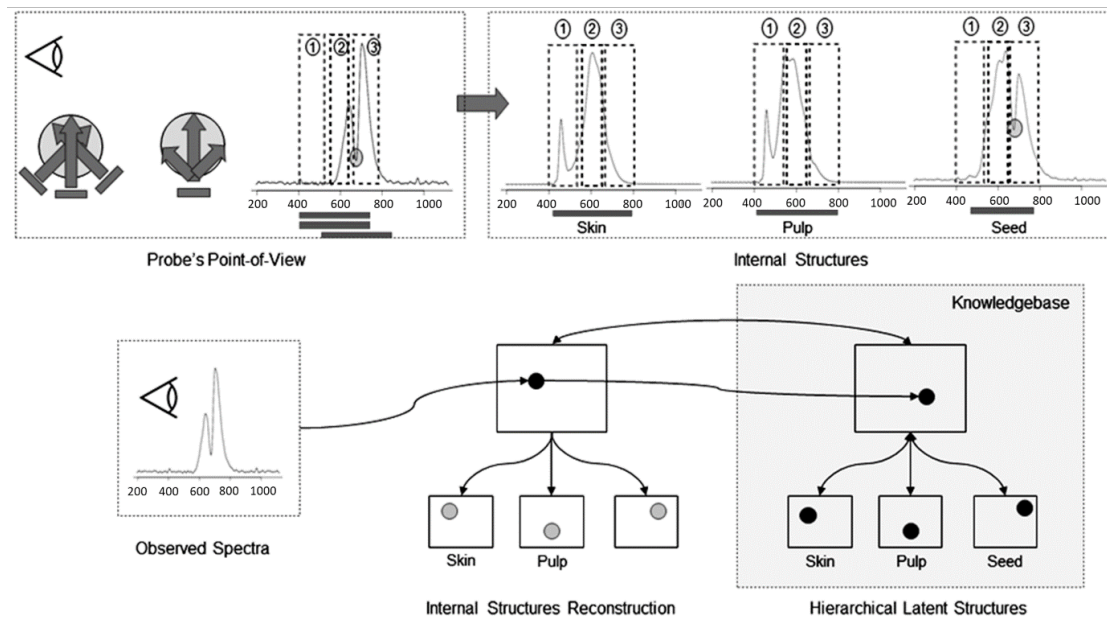


Figure 5.1-2. Schematic representation of an embodiment of 3D internal spectroscopy reconstruction method using hierarchical latent spaces.

In an embodiment, the two are related by hierarchical latent structures models in Figure 5.1-2, whereby the information of the object's different perspectives are matched against the knowledge base of tissues to provide a reconstruction of internal sample structures from the observed spectra.

The basis of information relationship within different scales of spectral measurements is performed by hierarchical relationship of super-space and corresponding sub-spaces. The sub-spaces conserve the latent space variables of the different layers of a hyperspectral microscope image, that is, the features of these images are linear combinations of the vectors spawning the sub-space at a given space coordinate (de Noord & Theobald, 2006; Westerhuis et al., 1998). These characteristics are used to derive a super-space of characteristics, translating the combination of a particular hyperspectral image or tomography. The super-space characteristics are afterwards

possible to be related to the macroscopic feature space to establish the relationship between macroscopic and microscopic measurements.

Results of the reconstruction of the macroscopic tomography using hierarchical latent variables are presented in Figure 5.1-3. In this example, we demonstrate that although the observed spectra in different types of equipment, e.g., robotic device and HR4000 spectrometer, are not similar in terms of their shape, these carry the same information. The two signals' variance space is equivalent. Such means that both are observing the same spectral data but with a different type of spectral fingerprint due to the optical configuration. This fingerprint carries the same physical and chemical information about the observed grapes.

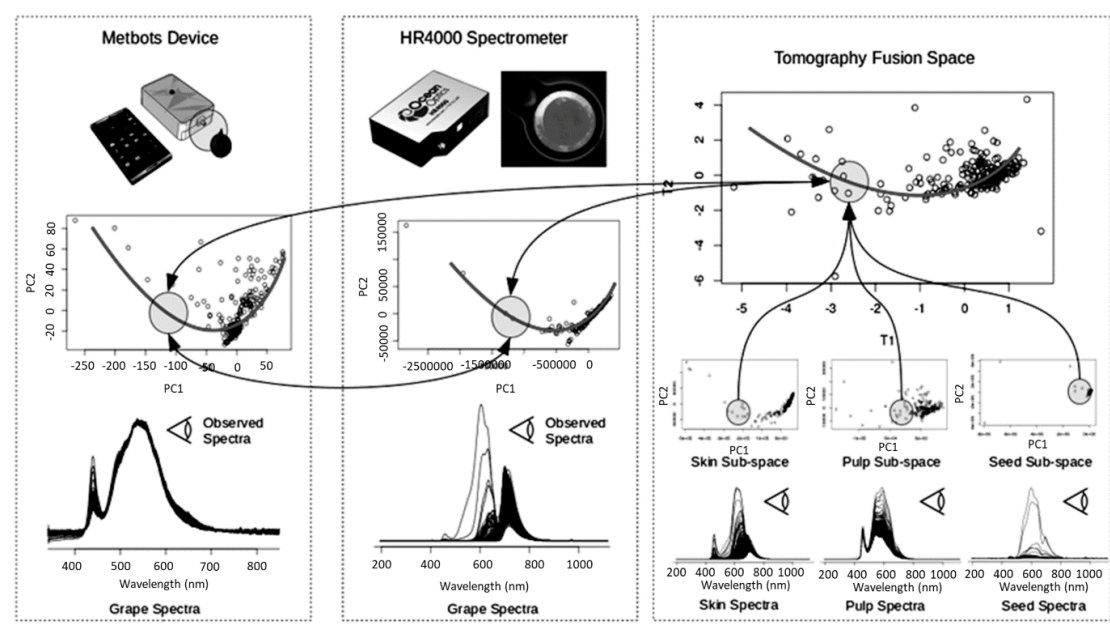


Figure 5.1-3. Schematic representation of an embodiment of macroscopic tomography in grapes: measurements made with a robotic device and HR4000 spectrometer to the hole grape; correspondence between spectral variance space of the robotic device and HR4000 spectrometer and its relationship to the tomography fusion space, obtained by using the information from skin, pulp and seed variance sub-spaces.

Spectra of the grapes along the maturation follow the displayed parabola of Figure 5.1-3, in both variance spaces of the robotic device and HR4000 spectrometer. This parabolic shape is due to the evolution of grape maturation and is also observed in the tomography fusion space, which is obtained by the hierarchical relationship between the latent sub-spaces of the skin, pulp, and seeds tissues spectral data.

The tomographic fusion space is obtained by the information fusion of the variance sub-spaces of spectral characteristics of the skin, pulp, and seeds, obtained by their observed spectra, as seen on Figure 5.1-3. The relationship between the robotic device, HR4000, tomography fusion space and the information of the corresponding tissue is bi-

directional, *i.e.*, it possible to infer the tissues spectra from the grape spectra - and vice-versa - to estimate the observed grape spectra from the tissue spectra.

This case-study shows how spectral data is correctly represented by its latent structures, spaces, and sub-spaces. The spectral data in tomography, as well as, between different physical scales, is well represented by a hierarchical relationship of the latent structures from each scale. The use of latent structures is highly computationally efficient, as these methods perform all computations in a compressed space, and therefore, as the computational cost is low, the proposed method is ideal for running in low-cost IoT (Internet of Things) hardware, allowing the development of very cost-effective hardware/software relation for metabolic tomography.

Figure 5.1-4 exemplifies the reconstruction of the skin, pulp, and seed spectra from the grape spectra recorded by the robotic device from two different grapes collected during veraison and near harvest. At the human eye, these grapes look very similar, and the recorded spectra from the robotic device to the untrained eye show low variation. The grapes are however in very distinct development stages, and their tissues have very significant compositional differences. By using the relationship between the robotic variance space and tomography fusion space, the computed spectra from the skin, pulp, and seed spectra exhibit very significant differences typical of veraison and near harvest development stages. These significant differences correspond to compositional differences in skin pigmentation (chlorophylls vs anthocyanins), sugars and acids accumulated in the pulp tissues, and seed ripeness status (proportional to lignification and tannins) with estimates presented in Figure 5.1-4.

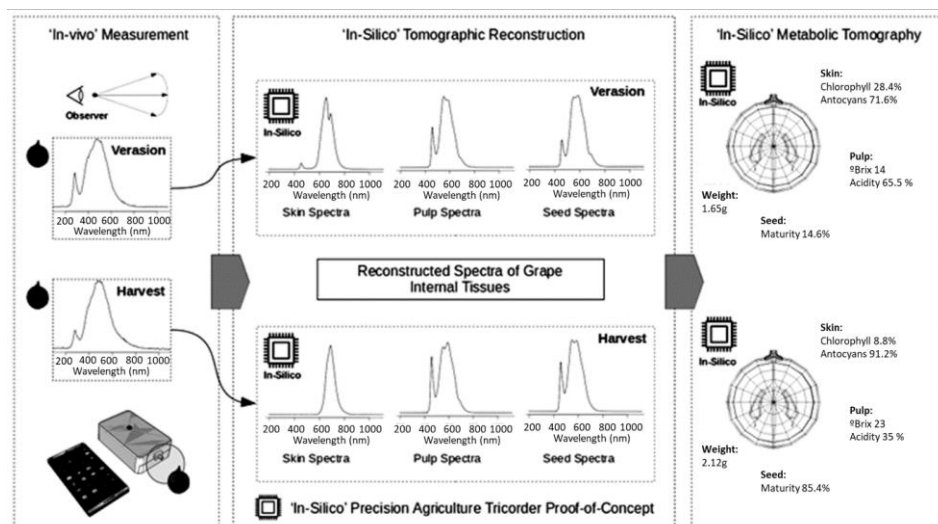


Figure 5.1-4. Schematic representation of an embodiment of macroscopic tomography reconstruction: 'in- vivo' measurement of two different grapes at veraison and near harvest time, corresponding 'in- silico' reconstruction of the skin, pulp and seed tissues spectra and estimated relevant properties of each tissue.

To our best knowledge, the state-of-the-art does not provide a tool similar to the present disclosure. Results show that the disclosed technology has high potential to revolutionise precision agriculture in its dimensions by being able to monitor the omics of each plant in vast plantations, providing information for the study of the complex relationship of plant-soil-limate-ecosystem and agricultural practices.

The benchmarks presented in Table 1 show how this type of technology can improve precision agriculture's state of the art. By focusing precision agriculture on plant physiology and its molecular biology, new horizons can be achieved by: i. developing new agricultural practices with omics scientific evidence; ii. introducing a modern way of developing agricultural practices by the use of *in silico* models from bioinformatics and systems biology; iii. linking plant biotechnology molecular biology *in vitro* results towards *in vivo* field applications; and iv. understanding how the physiology of plants inserted into a natural ecosystem responds when compared to laboratory observations.

Flow diagrams of particular embodiments of the presently disclosed methods are depicted in figures. The flow diagrams illustrate the functional information one of ordinary skill in the art requires to perform said methods required in accordance with the present disclosure.

It will be appreciated by those of ordinary skill in the art that unless otherwise indicated herein, the particular sequence of steps described is illustrative only and can be varied without departing from the disclosure. Thus, unless otherwise stated, the steps described are so unordered meaning that, when possible, the steps can be performed in any convenient or desirable order.

It is to be appreciated that certain embodiments of the disclosure as described herein may be incorporated as code (e.g., a software algorithm or program) residing in firmware and/or on computer useable medium having control logic for enabling execution on a computer system having a computer processor, such as any of the servers described herein. Such a computer system typically includes memory storage configured to provide output from execution of the code which configures a processor in accordance with the execution. The code can be arranged as firmware or software and can be organised as a set of modules, including the various modules and algorithms described herein, such as discrete code modules, function calls, procedure calls or objects in an object-oriented programming environment. If implemented using modules, the code can comprise a single module or a plurality of modules that operate in cooperation with one another to configure the machine in which it is executed to perform the associated functions, as described herein.

The disclosure should not be seen in any way restricted to the embodiments described and a person with ordinary skill in the art will foresee many possibilities to modifications thereof. The above-described embodiments are combinable. The following claims further set out particular embodiments of the disclosure.

5.1.4.Claims

1. Computer-implemented method for non-invasive tomographic metabolite characterisation of a vegetable or animal sample to be characterised, the sample comprising a plurality of differentiated tissues from a set of previously- characterised differentiated tissues, by using a multidimensional latent structure model of differentiated tissues, which was previously obtained by:

- Scanning tissue samples belonging to each differentiated tissue of the set with a hyperspectral microscope and spectrometer;
- Acquiring hyperspectral spectrum or spectra for each differentiated tissue sample from the hyperspectral microscope and spectrometer;
- Analysing the same tissue samples for the metabolites to be characterised; calculating the latent structure model from the acquired spectra and analysed metabolites;

Wherein the method comprises the steps of:

- Spatially scanning the sample to be characterised with a hyperspectral point-of- measurement, POM, optical device and a spectrometer;
 - Acquiring hyperspectral spectrum or spectra from each of a plurality of spatial positions and/or orientations of the POM optical device and spectrometer in respect of the sample to be characterised;
 - Using the previously obtained model to deconvolute the metabolites to be characterised for each of the differentiated tissues from the acquired spectra.
2. Method according to the previous claim wherein the scanning tissue samples belonging to each tissue of the set with a hyperspectral microscope and spectrometer is microscopic, and the spatially scanning the sample to be characterised with a hyperspectral POM optical device and spectrometer is macroscopic.
 3. Method according to the previous claim wherein spatially scanning comprises varying one or more of the following to obtain a spatial diversity of the POM spatial scanning:

- Displacement of the POM optical device;
 - Point-of-view angle of the POM optical device;
 - Different scales of magnification of the POM optical device;
 - Moving the POM optical device lines of sight to cause a parallax effect.
4. Method according to any of the previous claims wherein the multidimensional latent structure model is a hierarchical latent structure model.
 5. Method according to the previous claim wherein the hierarchical latent structure model is comprised of a latent variable model comprising a latent variable super- space and corresponding latent variable sub-spaces.
 6. Method according to the previous claim wherein the latent variable super-space is a super-space for each of the metabolites to be characterised, translating the combination of hyperspectral data to the metabolites to be characterised; each of the latent variable sub-spaces is a latent sub-space for spectral data of each of the differentiated tissues.
 7. Method according to any of the previous claims wherein the sample to be characterised is a three-dimensional tissue structure or organ.
 8. Method according to any of the previous claims wherein the metabolites to be characterised comprise pigments, tannins, photosynthesis, energy storage, nutrient uptake, necrosis, infection probability, or combinations thereof.
 9. Method according to any of the previous claims wherein the sample to be characterised is a vegetable sample and the differentiated tissues comprise skin, pulp and seed tissue.
 10. Method according to the previous claim wherein the sample to be characterised is a vegetable sample, in particular the sample to be characterised is a tomato or grape tissue sample.
 11. Method according to any of the previous claims wherein the scanning of tissue samples belonging to each differentiated tissue is carried out along a sampling grid.
 12. Method according to any of the previous claims wherein the scanning of tissue samples belonging to each differentiated tissue is carried out along discrete sampling depths.
 13. Method according to any of the previous claims wherein the spectra are UV-VIS-NIR.

14. Device for non-invasive tomographic metabolite characterisation of a vegetable or animal sample to be characterised, the sample comprising a plurality of differentiated tissues from a set of previously-characterised differentiated tissues, comprising a computer processor and a non-transitory computer-readable medium comprising a multidimensional latent structure model of differentiated tissues, which was previously obtained by:

- Scanning tissue samples belonging to each differentiated tissue of the set with a hyperspectral microscope and spectrometer;
- Acquiring hyperspectral spectrum or spectra for each differentiated tissue sample from the hyperspectral microscope and spectrometer;
- Analysing the same tissue samples for the metabolites to be characterised; calculating the latent structure model from the acquired spectra and analysed metabolites;

Wherein the computer processor is configured for:

- Spatially scanning the sample to be characterised with a hyperspectral point-of-measurement, POM, optical device and a spectrometer;
- Acquiring hyperspectral spectrum or spectra from each of a plurality of spatial positions and/or orientations of the POM optical device and spectrometer in respect of the sample to be characterised;
- Using the previously obtained model to deconvolute the metabolites to be characterised for each of the differentiated tissues from the acquired spectra.

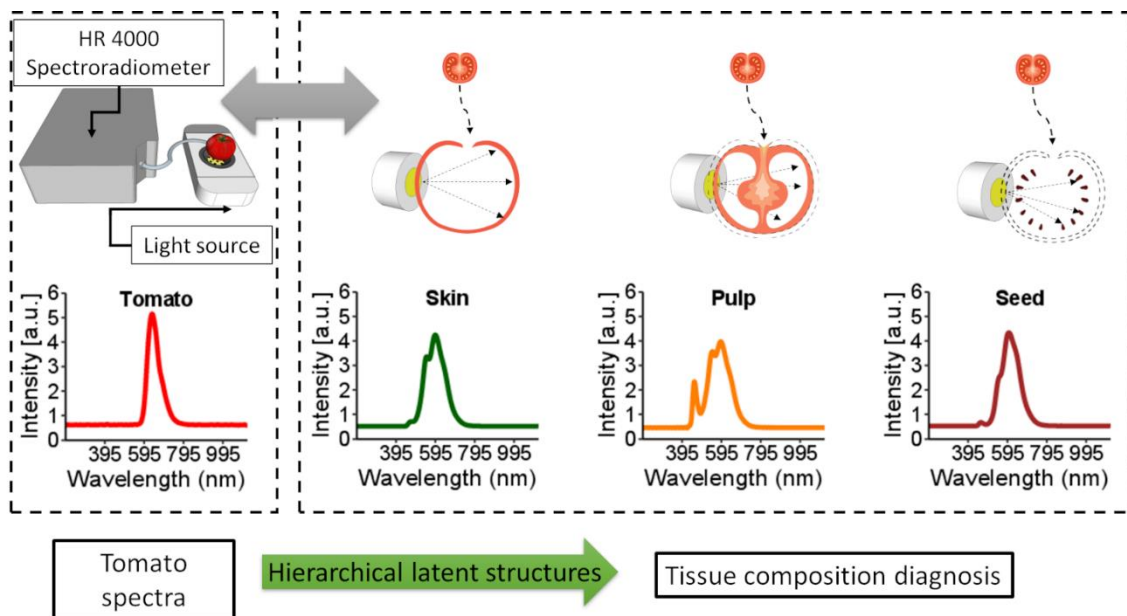
15. Non-transitory computer-readable medium comprising computer program instructions for implementing a device for non-invasive tomographic metabolite characterisation of a vegetable or animal sample to be characterised, which when executed by a processor, cause the processor to carry out the method of any of the claims 1-13.

Acknowledgments

It is acknowledged the contribution of Fundação Amadeus Dias, Santander Universidades, and the University of Porto in respect of the BIP PROOF 2021 AWARD. This work was supported by Project "OmicBots: High-Throughput Integrative Omic-Robots Platform for a Next Generation Physiology-based Precision Viticulture/PTDC/ASP-HOR/1338/2021", financed by the National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology).

5.2.Case study IV: Tomato case study: identifying tissue changes during maturation.

Tosin, R., Cunha, M., Monteiro-Silva, F., Santos, F., Barroso, T., and Martins, R. (2024). Bi-directional hyperspectral reconstruction of cherry tomato: diagnosis of internal tissues maturation stage and composition. *Frontiers in Plant Science*. doi: 10.3389/fpls.2024.1351958



Graphical abstract. 5.2-1. Graphical abstract case study IV (Tosin et al., 2024a).

Bi-directional hyperspectral reconstruction of cherry tomato: diagnosis of internal tissues maturation stage and composition

Renan Tosin^{1,2}, Mario Cunha^{1,2*}, Filipe Monteiro-Silva², Filipe Santos², Teresa Barroso², Rui Martins²

¹ Department of Geosciences, Environment and Spatial Planning, Faculty of Sciences of the University of Porto, Rua do Campo Alegre, S/N, 4169-465, Porto, Portugal.

² INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, S/N, 4200-465, Porto, Portugal.

* Correspondence:

Corresponding Author

mccunha@fc.up.pt

Abstract

Introduction: Precision monitoring maturity in climacteric fruits like tomato is crucial for minimising losses within the food supply chain and enhancing pre- and post-harvest production and utilisation.

Objectives: This paper introduces an approach to analyse the precision maturation of tomato using hyperspectral tomography-like.

Methods: A novel bi-directional spectral reconstruction method is presented, leveraging visible to near-infrared (Vis-NIR) information gathered from tomato spectra and their internal tissues (skin, pulp, and seeds). The study, encompassing 118 tomatoes at various maturation stages, employs a multi-block hierarchical principal component analysis combined with partial least squares for bi-directional reconstruction. The approach involves predicting internal tissue spectra by decomposing the overall tomato spectral information, creating a superset with eight latent variables for each tissue. The reverse process also utilises eight latent variables for reconstructing skin, pulp, and seed spectral data.

Results: The reconstruction of the tomato spectra presents a mean absolute percentage error of 30.44 % and 5.37 %, 5.25 % and 6.42 % and Pearson's correlation coefficient of 0.85, 0.98, 0.99 and 0.99 for the skin, pulp and seed, respectively. Quality parameters, including soluble solid content (%), chlorophyll (a.u.), lycopene (a.u.), and puncture force (N), were assessed and modelled with PLS with the original and reconstructed datasets, presenting a range of R² higher than 0.84 in the reconstructed dataset. An empirical

demonstration of the tomato maturation in the internal tissues revealed the dynamic of the chlorophyll and lycopene in the different tissues during the maturation process.

Conclusion: The proposed approach for inner tomato tissue spectral inference is highly reliable, provides early indications and is easy to operate. This study highlights the potential of Vis-NIR devices in precision fruit maturation assessment, surpassing conventional labour-intensive techniques in cost-effectiveness and efficiency. The implications of this advancement extend to various agronomic and food chain applications, promising substantial improvements in monitoring and enhancing fruit quality.

Keywords

Fruit Maturation; Latent Structures; Precision Agriculture; Spectral Reconstruction; Spectroscopy.

5.2.1. Introduction

Tomato is a climacteric fresh fruit composed of multiple tissues with diverse physical and biochemical compositions relevant to defining its quality through the food chain. The constitution of these tissues undergoes significant dynamic changes throughout the maturation process and in the post-harvest phase until the consumer such as the levels of antioxidants, lycopene, ascorbic acid, phenols and free radicals (Chandra & Ramalingam, 2011) and bioactive compound (e.g. flavonoids (Tamasi et al., 2019)).

The tomato fruit maturation process is marked by tissue specialisation, which promotes biochemical and physical changes in all tissues (Moco et al., 2007). During the ripening process, the tomato changes colour from green to red, resulting in morphological and biochemical modifications. The cultivar, environmental conditions (e.g., soil, light, temperature) and agronomic practices (e.g., irrigation, fertilisation) are essential factors that contribute to the tomato maturation process. For instance, in diverse regions, variations in antioxidants and phenols have been observed (Chandra et al., 2012). Under salinity conditions, morphological aspects such as size, water content, and colour undergo changes, impacting sugar content and acidity (Pascale et al., 2015). Additionally, genetic factors exhibit divergence even under identical conditions (Toor & Savage, 2005).

Several non-destructive techniques are available in the literature for characterising fruit maturation. However, none of these techniques can provide detailed information about the internal tissues of the fruit while also predicting its biophysical and biochemical characteristics efficiently. Although similar works approached the reconstruction of fruits

using other techniques, such as electrical impedance to detect the tomato level of maturation (Verma et al., 2021), computed tomography to detect bitter pit in apples (Si & Sankaran, 2016) and X-ray in predicting the sugar content in kiwi fruit (Kanno & Kuroyama, 2020) and phenotyping characteristics of seed in predicting the seed length, width, thickness, and radius of soybean and wheat with an accuracy ranging 80-96% (Liu et al., 2020b), some of these techniques are expensive and not expedite for in situ measurements. Additionally, these methods are mainly used to identify physical damage, and none of them analysed each tissue individually or presented a qualitative approach (Donis-González et al., 2014). In biomedical science, non-destructive tissue characterisation (human and animal) has been developed through Vis-NIR spectroscopy, particularly for detecting tissue anomalies (Dahlstrand et al., 2019; Malone et al., 2014) and guiding the right incision during surgery (Stelzle et al., 2012; Vo-Dinh et al., 2010), which indicate that similar techniques can be applied to vegetation tissues. Monitoring tomato precision maturation throughout the food chain is crucial for minimising losses within the food supply chain and improving pre- and post-harvest production and utilisation (Garcia & Barrett, 2006). From the high-tech horticulture point of view, monitoring maturation is essential to improve cultural practices, such as irrigation, fertilisation and canopy management, which are directly related to the pigments, organic acids and sugars in the tomato fruit (Bertin & Génard, 2018). In addition, to ensure that the seeds are fully developed for tomato seed production (Shrestha et al., 2016). They can guarantee that all tissues are well developed and that the gustative parameters favour the final consumer.

Traditional methods like chemical assays and chromatography used to characterise tomato biochemical parameters (e.g., lycopene) in different tissues are, in most cases, destructive, very expensive, non-adapted for small amount of tissues and time-consuming, which hinders the assessment of tomato quality parameters based on the composition of each tissue. Therefore, despite the high cost and time consumption of traditional analysis, alternative non-destructive techniques have been developed to assess tomato maturation and quality parameters, including colourimetric (Gómez et al., 2001), fluorescence (Konagaya et al., 2020; Wu & Wang, 2014), and Vis-NIR techniques (Zhu et al., 2015; Torres et al., 2015).

Vis-NIR devices have demonstrated outstanding potential for estimating fruit's biochemical and biophysical parameters like lycopene and β -carotene (Tilahun et al., 2018), water potential (Tosin et al., 2021) and sugars and acids (Martins et al., 2022). Numerous studies on tomatoes have utilised Vis-NIR techniques to estimate various biochemical and biophysical parameters. These include soluble solid content (SSC) with

reported R^2 of 0.87 (Ecartot et al., 2013), 0.60-0.77 (Torres et al., 2015), and 0.88-0.98 (Ding et al., 2016). The pH levels were determined with a R^2 of 0.80 (Huang et al., 2018a), while color attributes showed R^2 values ranging from 0.91-0.99 (Ecartot et al., 2013). Vitamin C content was estimated with a R^2 of 0.67 (Azadshahraki et al., 2018), total acidity with R^2 of 0.66-0.94 (Ding et al., 2016) and R^2 of 0.91-0.98 (Najjar & Abu-Khalaf, 2021), and firmness exhibited R^2 of 0.70-0.72 (Najjar & Abu-Khalaf, 2021). The analysis of β -carotene revealed R^2 values of 0.77-0.88 (Tilahun et al., 2018), phenols showed R^2 values of 0.76-0.98 (Ding et al., 2016), malic acid with R^2 of 0.27-0.42 (Torres et al., 2015), citric acid with R^2 of 0.66-0.94 (Ding et al., 2016) and lycopene with R^2 values ranging from 0.45-0.75 (Clément et al., 2015), 0.73-0.83 (Ciaccheri et al., 2018) and 0.85-0.89 (Tilahun et al., 2018). However, tomato exhibits differences in the structural and biochemical characteristics of different tissues, which leads to significant ramifications in the absorption and scattering of light inside the tomato fruit (Skolik et al., 2019). These complex structures of tomato make it challenging to investigate the inner tissues through Vis-NIR spectroscopy and how they behave during maturation.

Spectral information about the inner tissues of fruits can be acquired using Vis-NIR data (Martins et al., 2023). Martins et al. (2022) demonstrated empirically that different grape tissues influence the whole fruit during the maturation process and that the concentration of pigments changes during maturation in various tissues. Obtaining spectral information of the internal tissues of fruits through Vis-NIR requires appropriate modelling techniques. Principal component analysis (PCA) is one such technique that uses latent variables (LV) to perform an orthogonal transformation of the original dataset onto a reduced subspace that is spanned by the principal components (Dahlstrand et al., 2019). Conversely, the combination of LV creates a superset, presenting a direct relationship with the original information. LV models can deal with significant and correlated variables (Trygg & Wold, 2003). Multi-block hierarchical PCA (HPCA) and hierarchical partial least squares (HPLS) are frequently used in chemometrics to deal with spectral information from different sensors and a batch of data (Martins et al., 2023; Mishra et al., 2021). Therefore, multi-block analysis can create a bi-directional reconstruction of whole tomato fruit from the skin, pulp, and seed spectral data.

Tissue reconstruction is based on hierarchical latent relationships between the spectral patterns of the observed tissues, providing details of the internal plant structures (Martins et al., 2023). This manuscript uses the term “tomography-like” to partly define the capacities of data-driven class reconstruction using hierarchical relationships. The practicality of bidirectionality involves connecting the principal latent space derived from tomato tissues to the entire tomato spectrum and executing the reverse process, which

consists of breaking down the tomato fruit's spectra into the spectra of its tissues, namely the skin, pulp, and seed. This methodology was recently applied to grapes, as demonstrated by Tosin et al. (2023b), and it is expected to work in tomato, in which the tissue composition is different from grapes. The term tomography-like is debatable in that it only implies the resolution of the tissue image; here, it is aimed to provide a median spectrum of each tissue, given the fruit spectra in a non-destructive way. Once several spectra are taken from different positions on the fruit, a 3D resolved image can be reconstructed by relating the positions $[x, y, z]$ and spectral gradients within internal tissues (Tosin et al., 2023b; Martins et al., 2022). In this sense, this work can be considered 'tomography-like'. Data-driven reconstruction does not use the same numerical approaches and solutions as the classical approaches but is entering several research areas due to their computational efficiency (Martins et al., 2022; Bar-Sinai et al., 2019a).

Furthermore, spatially resolved tissue fruit composition is yet very complex to be obtained experimentally (Tosin et al., 2023b). There are still very significant constraints at the level of analytical chemistry state-of-the-art on the quantity of sample that can be used to quantify parameters considered in today's routine analysis, such as SSC and pigments. Therefore, metabolic or compositional imaging validation is still limited to the laboratory ground truth methods.

This research used a point-of-measurement (POM), where the light enters the fruit and has internal reflections, being only able to return to the spectrometer through a centre fibre optics pinhole, meaning that all light reaching the spectrometer interacts with the inner fruit's tissues, maximising the spectral information on all internal tissues.

In multispectral or hyperspectral imaging, light is generally illuminated outside the fruit. This demonstrates that most light reaching the imaging sensor is reflected, non-absorbed, and carries little information about internal tissue composition.

Therefore, relationships with the recorded spectra using this method are generally limited to covariant information with pigmentation, which has the danger of quantifying through correlation and not due to the causal characteristic features present in the spectra. At present, it is believed that POM devices can be of more practical application to field studies than hyperspectral cameras.

Through multi-block analysis, this research facilitates the reconstruction of hyperspectral data by utilising information from individual spectra of tomato tissues, namely skin, pulp, and seed. This method enables the decomposition of the overall tomato spectrum into its constituent tissues, offering a bi-directional relationship. This study further provides a qualitative analysis of the tomato ripening process, elucidating the maturation levels in

the skin, pulp, and seed at various developmental stages. By employing a POM sensing approach that maximises spectral information from internal tissues, the research addresses the limitations of traditional destructive methods, providing a non-destructive alternative for characterising tomato biochemical parameters. This contribution to the high-tech horticulture food supply chain can be used to ensure superior quality produce, reduce waste, enhance market value, and advance agricultural practices.

Therefore, three main objectives have been established in this work: i) to reconstruct the tomato hyperspectral data using information from the skin, pulp, and seed spectra through multi-block analysis; ii) to demonstrate that the spectral information of the entire tomato can be decomposed into the skin, pulp, and seed; and iii) to provide a qualitative analysis of the dynamics of the tomato ripening process, demonstrating the maturation levels in the skin, pulp, and seed, and how these tissues behave at different stages of the maturation process.

5.2.2. Material and methods

5.2.2.1. Sampling and tomato properties

A total of 118 cherry tomatoes, freshly picked at several maturation stages, were promptly taken to the laboratory for analysis. Puncture force (N) and SSC (%) were performed after measuring the tomato spectra using a digital penetrometer (model PCE-PTR 200, PCE Group, D-59872 Meschede, Germany), registering the resistance force and maximal force until puncture and a hand refractometer Milwaukee model MR32ATC, with a scale range of SSC from 0 to 32.0 %, respectively.

The tomato skin, pulp, and seeds were methodically extracted from each tomato ($n = 118$) and subjected to individual analysis to obtain their respective spectral records. Then, aliquots measuring approximately 0.5 cm² were taken using a lancet and deposited onto a glass microscope slide. The procedure involved peeling the tomato skin, slicing the pulp to a thickness of approximately 3 mm, and directly placing a single well-developed seed, selected from the various seeds present in the tomato, onto the microscope slide. This meticulous process ensured the separation of tomato tissues and allowed the acquisition of specific spectral data for the skin, pulp, and seeds.

The tomato process of maturation progresses from green to red and can be classified into six different colours: i) green, ii) breakers, iii) turning, iv) pink, v) light-red and vi) red (USDA, 1991).

Tomato is a complex fruit regarding internal tissues (Figure 5.2-1, Figure S 5.2-1). At the green stage of maturation, the tissues are not well developed, which makes tissue separation hard. Therefore, this work considered the epidermis as the skin, columella,

placenta and pericarp as pulp and seeds as seeds (Figure 5.2-1). The jelly parenchyma and sepal were not considered in the analysis.

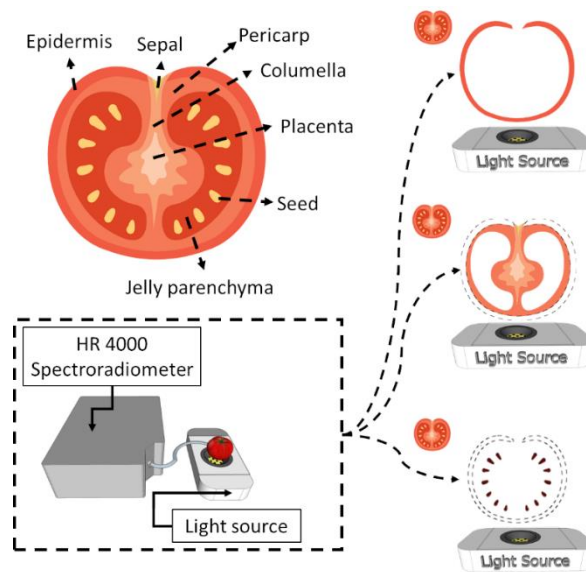


Figure 5.2-1. Internal tomato tissues and the spectroscopy system used to obtain spectral information from the entire tomato and the respective tissues (skin, pulp and seed).

5.2.2.2. Spectroscopy

Tomato spectra were recorded with a white LED platform (Supplementary Figure 1). The platform comprises a reflection disk with a power LED (6500K, Philips SpotOn Ultra 69141/31/PH) at the bottom. The spectral range of the LED emits light from 380 nm to 780 nm. Therefore, LED spectra were used as a reference to check measurement and light emission stability. The tomato is placed above the LED, and the measurement is performed by collecting reflectance with a fibre optic probe (Ocean Insite). Skin, pulp, and seeds are placed on the microscope slide centred with the LED, and the reflectance probe also collects light. Spectra were recorded by a high-resolution spectroradiometer (Ocean Insite HR4000), which obtains information from 195.34 nm to 1118.33 nm; the integration time was optimised for each sample to maintain most of the spectra within the linear response.

After collecting all the spectral data, a logarithm multiplicative scattering correction (Martins et al., 2022) was applied to normalise and reduce the noise in the spectral information. The correction is a widely used method that addresses the issue of light scattering, which can distort the spectral signal and lead to inaccuracies in the measurements.

The logarithmic transformation helps to remove this scattering effect and improve the accuracy of the spectral data.

5.2.2.3. Hierarchical latent structures reconstruction

Latent structures are spaces obtained by matrix decomposition into their eigenvectors, a new basis where the contained information is projected. Latent structures provide a geometrical interpretation of the dataset and its samples by understanding their position on the eigenvector basis. Eigenvectors can be extracted with different properties, but one of the most common decompositions is PCA, where orthogonal eigenvectors are obtained by maximising the dataset variance, allowing to provide the interpretation of relevant variation. PCA can be obtained from the dataset $\mathbf{X}$ by singular value decomposition (SVD) after subtracting the mean of $\mathbf{X}$: $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^t$; where $\mathbf{U}$ is the left singular, $\mathbf{S}$ the singular values and $\mathbf{V}^t$ the left singular. In PCA, the scores (aka latent structures) $\mathbf{t}$ are given by $\mathbf{U}\mathbf{S}$, and the loadings (aka basis/eigenvectors) by $\mathbf{p}^t = \mathbf{V}^t$.

The geometry of information contained in $\mathbf{X}$ can be studied to determine what eigenvectors represent non-random information by performing randomisation tests (Martins et al., 2022), where the spectra dataset reconstruction can be decomposed into $\mathbf{X} = \mathbf{t}\mathbf{p}^t + \mathbf{e}$, where $\mathbf{e}$ is random information, irrelevant for spectral reconstruction (Martins et al., 2023).

Let's consider the corresponding database of tissue spectra: skin $\mathbf{X}_1$, pulp $\mathbf{X}_2$ and seeds $\mathbf{X}_3$ and their corresponding relevant PCA decomposition:

$$\mathbf{X}_1 = \mathbf{t}_1\mathbf{p}_1^t + \mathbf{e}_1 \quad (1)$$

$$\mathbf{X}_2 = \mathbf{t}_2\mathbf{p}_2^t + \mathbf{e}_2 \quad (2)$$

$$\mathbf{X}_3 = \mathbf{t}_3\mathbf{p}_3^t + \mathbf{e}_3 \quad (3)$$

Where $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_3$ are the relevant latent features that reconstruct the original tissue spectra $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$, and $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ discarded random spectral information. The latent information associated with each tissue can now be fused by determining the relevant common dimensions of their latent variance in geometry along each eigenvector.

Let's take $\mathbf{t}_f = [\mathbf{t}_1^i, \mathbf{t}_2^i, \mathbf{t}_3^i]$ as the concatenation of the i dimension of $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_3$, to be fused into a superset latent space $\mathbf{T}$ by finding the relevant information of each sub-level. The superset latent space can be determined by:

$$\mathbf{t}_f = \mathbf{T}_i\mathbf{P}_i^t \quad (4)$$

Being $\mathbf{T}_i$ the superset latent information of the i dimension of the subsets (tissue spectra), and $\mathbf{P}_i^t$ provide the contribution of each subset to the fused information $\mathbf{T}_i$.

If the information of $[\mathbf{t}_1^i, \mathbf{t}_2^i, \mathbf{t}_3^i]$ has the same direction, $\mathbf{T}_i$ will be described by a single eigenvector $\mathbf{P}_i^t$ or a single dimension; otherwise, further relevant dimensions are added to $\mathbf{T}_i$.

The superset latent structure is constructed for each dimension of $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$, representing the relevant information of the tissue dataset, that is, the relevant characteristics from the skin, pulp, and seeds that relate to the observed tomato spectra.

5.2.2.4. Association and bi-directionality

The relevant features extracted from the sub-levels represented in the superset $\mathbf{T}$ have a similar latent structure to the direct PCA decomposition of the tomato spectra ($\mathbf{Y}$). By performing a PCA decomposition to $\mathbf{Y}$, it is gotten $\mathbf{Y} = \mathbf{U}\mathbf{C}^t$; where a direct association between the latent spaces $\mathbf{T}$ and $\mathbf{U}$ is expected ($\mathbf{T} \approx \mathbf{U}$). One can expect that samples with similar composition and morphological characteristics will generate cluster aggregations in $\mathbf{T}$ and $\mathbf{U}$, reflecting the different skin, pulp, and seed maturation state combinations (Martins et al., 2023).

Therefore, one can establish a direct relationship between neighbouring samples in $\mathbf{T}$ or $\mathbf{U}$, ensuring bi-directionality between the subsets $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{Y}$ (Figure 5.2-2).

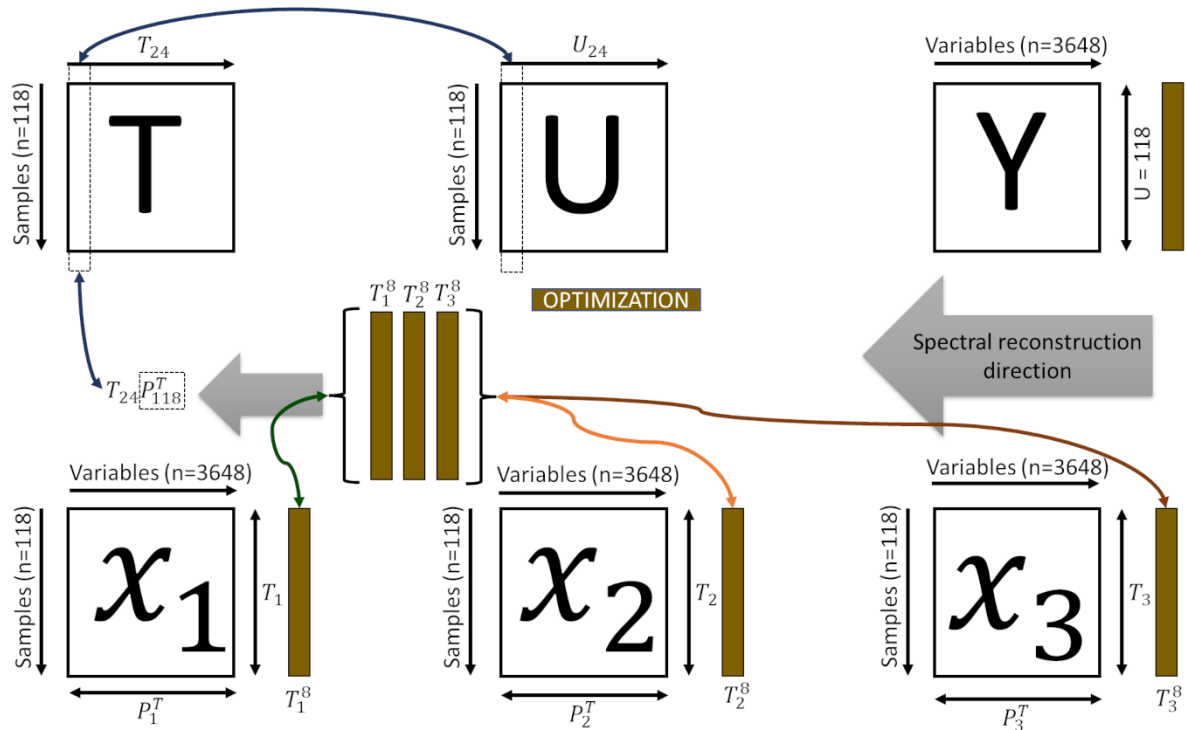


Figure 5.2-2. Representation of the bi-directional reconstruction process and decomposition of the spectral information. $\mathbf{T}$ is the superset latent space; $\mathbf{T}_{24}$ the superset latent information of the dimension of the subsets ($\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$); $\mathbf{U}$ is the feature space of each subset; $\mathbf{U}_8$ the superset latent information of the dimension of the subsets $\mathbf{Y}$ is the tomato spectra; $\mathbf{P}_{118}^t$ is the contribution of each subset ($\mathbf{P}_1^t$, $\mathbf{P}_2^t$, $\mathbf{P}_3^t$); $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ spectra of the skin, pulp and seed, respectively.

Inferring the internal tissues for a given unknown sample is performed by projecting the spectra $\mathbf{Y}$ into the feature space $\mathbf{U}$, by $\mathbf{U} = \mathbf{Y}\mathbf{C}$, and finding the neighbouring samples ($\mathbf{k}$) in this feature space. The $\mathbf{k}$ can be used to verify its propagation from $\mathbf{T}$ to the sub-level spaces $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$, by reconstructing $\mathbf{t}_f$ (eq. 4).

5.2.2.5. Validation

The hierarchical latent structure model was optimised and validated in a two-step approach: i. cross-validation (CV) to optimise the number of principal components (PC) of the sub-spaces $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_3$ and superspace $\mathbf{T}$; and ii. hold-out samples (HO) are used to test predictions and provide quantitative metrics.

CV is a test to the null hypothesis, and HO samples are double-check confirmations of the CV metrics. If the knowledge base is representative, any unknown HO or removed from the dataset, CV should provide statistically similar prediction metrics, proving the null hypothesis. By leaving samples out, CV provides the determination of the optimal error of each tissue reconstruction, $\mathbf{e} = [\mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3]$ (eqs. 1-3). For each sample, the training set, the CV algorithm removes one sample (leave-one-out) for determining the error ($\mathbf{e}$) for increasing the number of PCs of sub-space and superset. The optimal number of PCs is considered the one that provides minimal CV errors, preventing over-characterisation of random features at the sub-level passing into the superset. Suppose the training set is representative, CV and HO errors are expected to be similar. In that case, the model can efficiently reproduce the spectral information, and the null hypothesis is verified.

After reconstructing and decomposing the spectral data, a standardisation of the original, reconstructed, and decomposed spectral data was applied to mitigate the effects of signal intensity. Standardisation is a common practice used in data analysis to rescale variables with a mean of zero and a standard deviation of one. In this study, the formula $(\mathbf{x} - \mathbf{mean})/\mathbf{sd}$ was used to standardise the spectral data. This method allows us to compare and analyse the spectral data more accurately, as it removes any differences in signal intensity that could impact the interpretation of the results.

The following metrics were used to benchmark spectral reconstruction: i. mean standard error (MSE) in counts/wavelength (nm), representing the average reconstruction error per wavelength (nm); ii. mean absolute percentage error (MAPE) in % wavelength (nm), representing the average bias per wavelength (nm); iii. Pearson's correlation coefficient and the p-value were extracted to check for significant differences between the original and reconstructed datasets; and iv. Euclidean distances in $\mathbf{T}$ and $\mathbf{U}$ of CV and HO

samples: This metric measures the knowledge-base representativeness and stability, which is important for evaluating the performance of the spectral reconstruction model.

5.2.2.6. Prediction of tomato quality

As a proof of concept, the organoleptic characteristics of tomatoes were predicted by comparing quantification results obtained from real spectral datasets with those obtained from reconstructed spectra using latent hierarchical structures. The visible range of the spectral data, which falls between 400-700 nm, contains valuable information related to pigments, specifically chlorophyll and lycopene content. Based on the findings of Ciaccheri et al. (2018) and Moco et al. (2007), the inferred chlorophyll content used the ratio of green (520-570 nm) to red (571-700 nm) spectral bands, and lycopene content used the ratio of red to green spectral bands. The results demonstrate that the reconstructed tissue spectra provide a good relationship to the estimates obtained with the fruit spectra, thus serving as proof of the principle of internal tissue quantification. Also, it presents other parameters obtained with the fruit, such as puncture force and SSC, which correlate with the reconstructed tissue spectra, further demonstrating the feasibility of the approach. Although there is a state-of-the-art, optimised ground truth method for measuring fruit composition at the fruit level for small fruits such as grapes (Tosin et al., 2023b; Martins et al., 2022), recording that data would not provide significant advantages as it does not allow better tissue resolution quantification than the one presented in this work.

This investigation employed a computational approach to assess tomato pigment content in tissue reconstruction, driven by the need for a time-efficient evaluation of lycopene and chlorophyll. Wet lab analyses commonly used for larger tissue samples were unsuitable due to the small tissue quantities (around 0.5 cm²) involved in the spectral analysis (Tosin et al., 2023b). Routine methods for these analytes require larger tissue quantities, hindering direct comparison with results obtained through computational methods (Tilahun et al., 2018; Clément et al., 2008). Sophisticated analytical methods for small tissue quantities are costly and impractical for evaluating a system at a low technology readiness level (TRL). Therefore, validating the results using expedited and cost-effective methods suitable for assessing the mentioned pigments and potentially other analytes is advisable. Adopting expedited approaches and avoiding expensive wet lab methods can validate the proof of concept without incurring high costs, facilitating a smoother transition for further system development and refinement.

The study employed a partial least squares (PLS) approach to predict the quality parameters of tomatoes. The dataset comprised 118 samples, divided into two sets: 70%

($n=82$) for training and 30% ($n=36$) for validation. This division of the dataset into training and validation sets ensures the utilisation of a significant portion of the data for model development while still allowing for robust evaluation and assessment of the model's performance.

A robust validation technique, leave-one-out cross-validation (LOOCV), was employed to evaluate the model's performance. This approach involved systematically excluding one sample at a time during the evaluation process, allowing for an accurate estimation of the model's predictive ability and mitigating the risk of overfitting.

The determination of the optimal number of LV in PLS model was carried out through an assessment of root mean square error (RMSE) values. This integral step in PLS modelling aimed to minimise the RMSE, underlining its fundamental role in refining the model for superior precision and effectiveness in predicting outcomes. The selection of the ideal number of LV was strategically driven by the overarching goal of achieving the most accurate and reliable results, a chase evident in the search of minimised RMSE values.

Within the data-driven analysis, representativeness and hypothesis-testing principles serve as foundational pillars. Representativeness ensures that the dataset employed for training and validation accurately represents the entire population of interest. Hypothesis testing facilitates the formulation and evaluation of statistical hypotheses, guaranteeing the results' reliability and significance.

By adhering to these principles, it is possible to construct robust models and generate reliable predictions in data-driven analysis.

Benchmarks were performed using the following modelling approaches: i. Similarity(Sim)-Euclidean distance as a metric of the spectral and compositional similarity between neighbouring samples in the feature space (e.g., Fachada et al., 2014); ii. Principal component regression (PCR) - where the latent structures of the sub-level spectra $[t_1 \ t_2 \ t_3]$, superset $\mathbf{T}$ and tomato spectra $\mathbf{U}$; PLS maximises the covariance between the spectra $\mathbf{X}$ and tomato composition $\mathbf{Y}$ by determining the eigenvectors of $\mathbf{X}^t \mathbf{Y}$ (Martins et al., 2023). This method forces the latent structures of spectra and composition (PLS scores - $\mathbf{U}$) to be equal (NIPALS algorithm) (Ergon, 2009) for the determination of each correspondent basis $\mathbf{U}^t$ and $\mathbf{Q}^t$ (Geladi & Kowalski, 1986). It proceeds with deflation and sequential orthogonal eigenvectors of the remaining information in $\mathbf{X}^t \mathbf{Y}$ (Phatak & De Jong, 1997). The number of deflation or LV is optimised by cross-validation/hold-out samples with minimal predicted sum of squares (PRESS) (Krstajic et al., 2014). PLS uses an oblique projection to determine the $\mathbf{b}_{pls}$ coefficients in $\mathbf{Y} = \mathbf{X}\mathbf{b}_{pls}$ (Ergon, 2009; Phatak & De Jong, 1997).

Require: X_1, X_2, X_3
Ensure: $j = \operatorname{argmin}(X_i - \mathbf{t}_i \mathbf{p}_i^t)$
 $X_1 = \mathbf{t}_1 \mathbf{p}_1^t$
 $X_2 = \mathbf{t}_2 \mathbf{p}_2^t$
 $X_3 = \mathbf{t}_3 \mathbf{p}_3^t$
While $j > j_{min}$ **do**
 $\tau_i \leftarrow [\mathbf{t}_1^i | \mathbf{t}_2^i | \mathbf{t}_3^i]$
 $\tau_i = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^t$
 $\mathbf{T}_i = \mathbf{U}_i \mathbf{S}_i$
 $\mathbf{P}_i^t = \mathbf{V}_i^t$
end while
Output: $\mathbf{T} = [\mathbf{T}_1 \dots \mathbf{T}_n]; \mathbf{P} = [\mathbf{P}_1 \dots \mathbf{P}_n]$

Algorithm 5.2-1. Hierarchical latent structures algorithm.

Require: $\mathbf{Y}, \mathbf{T}$
Ensure: $\mathbf{Y} = \mathbf{U} \mathbf{Q}^t$
 $\mathbf{U}^k = \mathbf{T} \mathbf{b}$
 $\mathbf{T}^k = \mathbf{U} \mathbf{c}$
 $\mathbf{Y}^k = (\mathbf{T} \mathbf{b}) \mathbf{Q}^t$
Output: $\mathbf{Y}^k$

Require: $\mathbf{Y}, \mathbf{T}$
Ensure: $\mathbf{Y} = \mathbf{U} \mathbf{Q}^t$
 $\mathbf{U}^k = \mathbf{T} \mathbf{b}$
 $\mathbf{T}^k = \mathbf{U} \mathbf{c}$
 $\mathbf{Y}^k = (\mathbf{T} \mathbf{b}) \mathbf{Q}^t$
Output: $\mathbf{Y}^k$

Algorithm 5.2-2. Outer relationships for reconstruction.

Require: $Y, T, P, [t_1, t_2, t_3]$

Ensure: $j = \operatorname{argmin}(X_i - t_i p_i^t)$

$$Y = UQ^t$$

$$T = Ub$$

$$[t_1, t_2, t_3] = T_p P^t$$

$$X_1 = t_1 p_1^t$$

$$X_2 = t_2 p_2^t$$

$$X_3 = t_3 p_3^t$$

Output: X_1, X_2, X_3

Require: $Y, T, P, [t_1, t_2, t_3]$

Ensure: $j = \operatorname{argmin}(X_i - t_i p_i^t)$

$$Y = UQ^t$$

$$T = Ub$$

$$[t_1, t_2, t_3] = T_p P^t$$

$$X_1 = t_1 p_1^t$$

$$X_2 = t_2 p_2^t$$

$$X_3 = t_3 p_3^t$$

Output: X_1, X_2, X_3

Algorithm 5.2-3. Latent structures reconstruction algorithm.

5.2.3. Results

5.2.3.1. Tomato tissue reconstruction

Figure 5.2-3 presents an application of PCA to investigate the spectral data of tomatoes, including the entire tomato and its internal tissues (skin, pulp, and seeds). Each data point represents a unique spectral measurement obtained from an individual tomato. The primary objective of PCA is to identify a lower-dimensional representation of the data that captures the most important information.

In this study, the analysis classified the tomatoes into three maturation stages: i) green, ii) turning, and iii) red. A random sample from each maturation stage was selected to represent the feature space of these samples.

Figure 5.2-3a illustrates the comprehensive space occupied by the assessed tomatoes. Each tissue of the tomato, namely the skin, pulp, and seeds, possesses its distinct feature space, as illustrated in Figures 3c, 3d, and 3e, respectively. PCA enables the decomposition of the space, resulting in individual spaces corresponding to each tomato fraction.

In Figure 5.2-3b, the plot demonstrates the reconstructed spectral data of the tomato obtained by combining the information from the skin, pulp, and seeds. The reconstructed data provides a condensed version of the original spectral data while preserving essential characteristics. It is also possible to reverse the process, enabling the data decomposition into the individual feature spaces of the tomato's skin, pulp, and seeds. The application of PCA in this study enhances the understanding of the tomato's spectral data and internal structures. In addition, it provides insights into the relationships between different tomato fractions and their corresponding spectral properties, shedding light on the distinct characteristics of each tissue within the tomato.

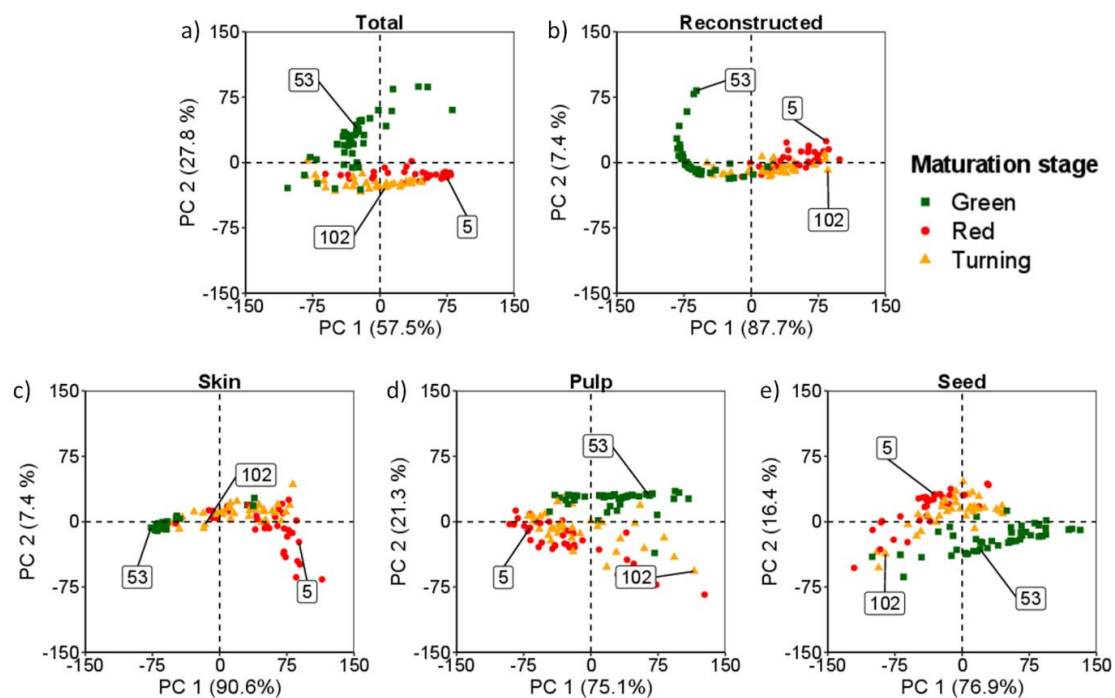


Figure 5.2-3. Spectral bi-directionality in reconstructing tomato components: skin, pulp, and seeds, at different maturation levels. The numbers in the box indicate the observations from the dataset for each maturation stage: ■ (53) for green, ▲ (102) for turning, and ● (5) for red. a) and b) represent the feature space for the total and total reconstructed spectra, respectively. c), d) and e) define the feature space of the skin, pulp and seed, respectively.

Figure 5.2-4 shows the spectral signature of tomato tissues at different stages of maturation: green, middle stage (turning), and ripened (red). The bi-directional spectral reconstruction (Figure 5.2-2) works better for internal tomato tissues than for the entire tomato (Figure 5.2-4, Table 5.2-1). In addition, during the experiments, it was observed that the light had more difficulty passing through the green tomatoes (Figure 5.2-4), which is a limitation in obtaining complete information about the internal tissues. It is suggested that at this stage of maturation, the tissues are not fully developed and contain a high concentration of pectin, which interferes with the optical properties of the tomato.

Table 5.2-1. Mean square error (MSE), mean absolute percentage error (MAPE), p-value and Pearson's correlation coefficient (r) for the reconstruction of tomato spectra and decomposition of the entire tomato spectra into skin, pulp and seeds spectra.

Reconstruction	MSE(Count/wavelength (nm))	MAPE (%)	p-value	r
Tomato Spectra	0.30	30.44	< 0.001	0.85
Skin Spectra	0.04	5.37	< 0.001	0.98
Pulp Spectra	0.02	5.25	< 0.001	0.99
Seeds Spectra	0.03	6.42	< 0.001	0.99

Mean square error (MSE), mean absolute error (MAPE), and p-value < 0.001 indicate that there is no significant difference between the original and reconstructed spectral matrices.

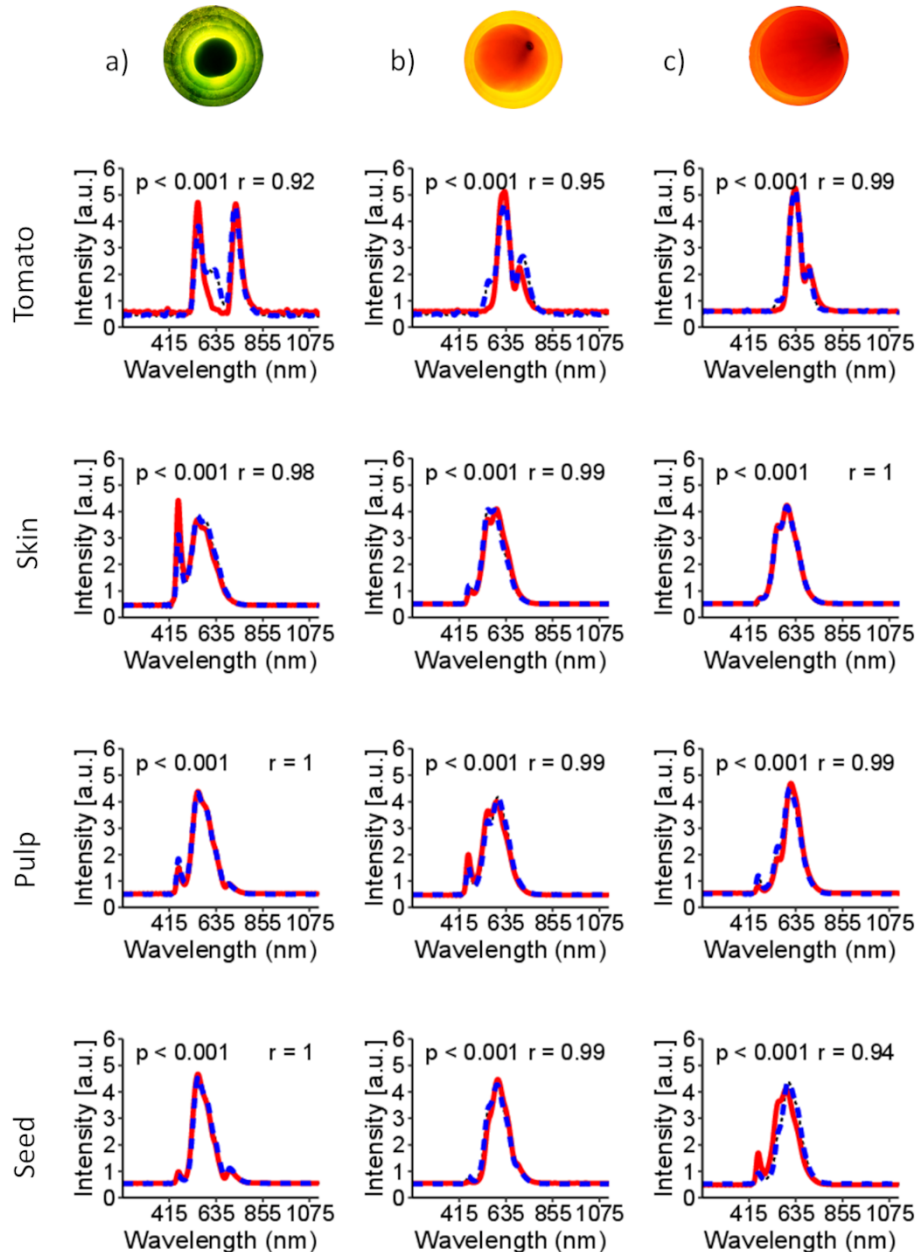


Figure 5.2-4. Spectral signatures of the tomato and the internal tissues. In the red line (—) is the original spectral signature, and in the blue dashed line (---) is the reconstructed spectral. Figures a) to c) are tomatoes subjected to LED light. Figure a) is a tomato in the green stages of maturation; b) tomato in the turning stages of maturation; c) tomato in the red stage of maturation. The p-value < 0.001 indicates no significant difference between the original and reconstructed spectra. r is Pearson's correlation.

The dynamics of maturation in tomato is shown in Figure 5.2-4. Based on these spectral signatures, different band peaks in the spectral signature are suggested to be related to tissue pigments (e.g., chlorophyll and lycopene). For example, green tomato presents higher signal intensity in the bands 460-500 nm and 660-700 nm, suggesting a correlation with chlorophyll content. Likewise, the 500-550 nm range of bands is more related to the carotene group, probably related to lycopene content in the more advanced stages of maturation.

Table 5.2-1 presents a benchmark for spectral reconstruction and shows that the spectral reconstruction did not present significative differences (p -value < 0.001) over the original spectral data. However, compared with the respective tissues, the total spectral of tomato shows a higher MSE (0.30) and MAPE (30.4%) and a lower Pearson's correlation coefficient ($r=0.85$; p -value < 0.001). Three leading causes can explain the low accuracy of the whole tomato: i) the green stage of maturation seems to be a hindrance for the light going through the internal tissues; ii) during the green stage of maturation, the tissues are not fully developed; and iii) the acquisition of the spectral information with an optical probe aimed at finding the best position to obtain the light signal through the internal tissues.

On the other hand, the decomposition of the whole tomato to predict the internal tissues worked better (Figure 5.2-4, Table 5.2-1). These results suggest that the spectral data of the tomato presented sufficient information on the internal tissues studied in this work. The lower accuracy in the reconstruction of tomato is discussed in the next section.

5.2.3.2. Quality parameters evaluation

The original and reconstructed spectral information were used to predict the quality parameters of the tomato and their respective tissues. Overall analysis showed that the original and reconstructed spectral data were consistent and robust for predicting the quality parameters analysed (Table 5.2-2). Additionally, the different tissues of tomato could be used to predict the quality parameters assessed in this experiment. It is important to highlight that for SSC and puncture force, the entire tomato was considered for the measurement, and the different tissues of the tomato could predict these values for the whole tomato fruit. The results presented for chlorophyll and lycopene used an empirical dry lab method based on the spectral information of each tissue and the whole tomato, which helped to infer the pigment concentration in each tissue individually and in the entire tomato. Regression plots of the SSC (%), chlorophyll (a.u.), lycopene (a.u.) and puncture force for the skin, pulp and seed are presented in the Supplementary Materials (Figure S 5.2-2, 3, 4 and 5).

Table 5.2-2. Reconstruction benchmark of the tomato, skin, pulp and seed in predicting the soluble solid content (SSC), chlorophyll, lycopene and puncture force. The original dataset exclusively predicted SSC and puncture force for the entire tomato. Chlorophyll and lycopene content were predicted solely in the reconstructed dataset, as these pigments were estimated using real data, rendering their prediction in the original dataset nonsensical.

Property	Metric	Real dataset Tomato	Skin	Pulp	Seeds
SSC (%)	r	0.99	-	-	-
	R ²	0.98	-	-	-
	MSE (%)	0.85	-	-	-
	MAPE (%)	10.78	-	-	-
	LV	3	-	-	-
Puncture force (N)	r	0.98	-	-	-
	R ²	0.97	-	-	-
	MSE (N)	2.94	-	-	-
	MAPE (%)	15.87	-	-	-
	LV	3	-	-	-
Reconstructed Dataset					
SSC (%)	r	0.99	-	-	-
	R ²	0.98	-	-	-
	MSE (%)	0.86	-	-	-
	MAPE (%)	11.12	-	-	-
	LV	3	-	-	-
Chlorophylls (a.u.)*	r	0.95	0.99	0.99	0.99
	R ²	0.90	0.99	0.99	0.99
	MSE (a.u.)	0.49	0.64	0.53	0.51
	MAPE (%)	51.32	6.01	10.06	13.51
	LV	2	2	3	2
Lycopene (a.u.)*	r	0.92	0.99	0.98	0.99
	R ²	0.84	0.99	0.96	0.98
	MSE (a.u.)	0.62	0.64	0.73	0.57
	MAPE (%)	37.68	6.44	11.09	12.91
	LV	2	2	2	2
Puncture force (N)	r	0.90	-	-	-
	R ²	0.95	-	-	-
	MSE (N)	3.84	-	-	-
	MAPE (%)	21.83	-	-	-
	LV	2	-	-	-

*Values computed with the spectral information (nm); the number of latent variables (LV) used in the partial least square (PLS). SSC and puncture force measurements were exclusively conducted for the entire tomato. Prediction based on individual tissues such as skin, pulp, and seeds is deemed impractical, given the integrated nature of these tissues. As a result, a dash (–) is denoted to signify the exclusion of these components in the predictive analysis.

Figure 5.2-5 presents the changes in chlorophyll and lycopene concentrations in different tomato tissues during ripening. As the tomato ripens, chlorophyll concentration decreases while lycopene increases, as demonstrated by the spectral information and dynamic concentration data in Figure 5.2-5.

During the early maturation stage (Figure 5.2-5a), the tomato skin has a higher concentration of chlorophyll and a lower concentration of lycopene. Likewise, the pulp has a higher concentration of chlorophyll and a lower concentration of lycopene than other tissues of the tomato. This distribution of chlorophyll and lycopene is also reflected

in the peaks observed in the tomato spectra. Specifically, peaks in the 460-500 nm and 670-700 nm range are associated with chlorophyll, while those in the 530-560 nm range are associated with carotenes, including lycopene. These wavelength assignments are drawn from the findings of Ciaccheri et al. (2018) and Moco et al. (2007).

During the ripening process, the peaks associated with chlorophyll decrease while those associated with lycopene increase, as shown in Figure 5.2-5 for the tomato skin, pulp, and seed. These changes in pigment concentrations are responsible for the observed colour changes in the tomato from green to red as it ripens.

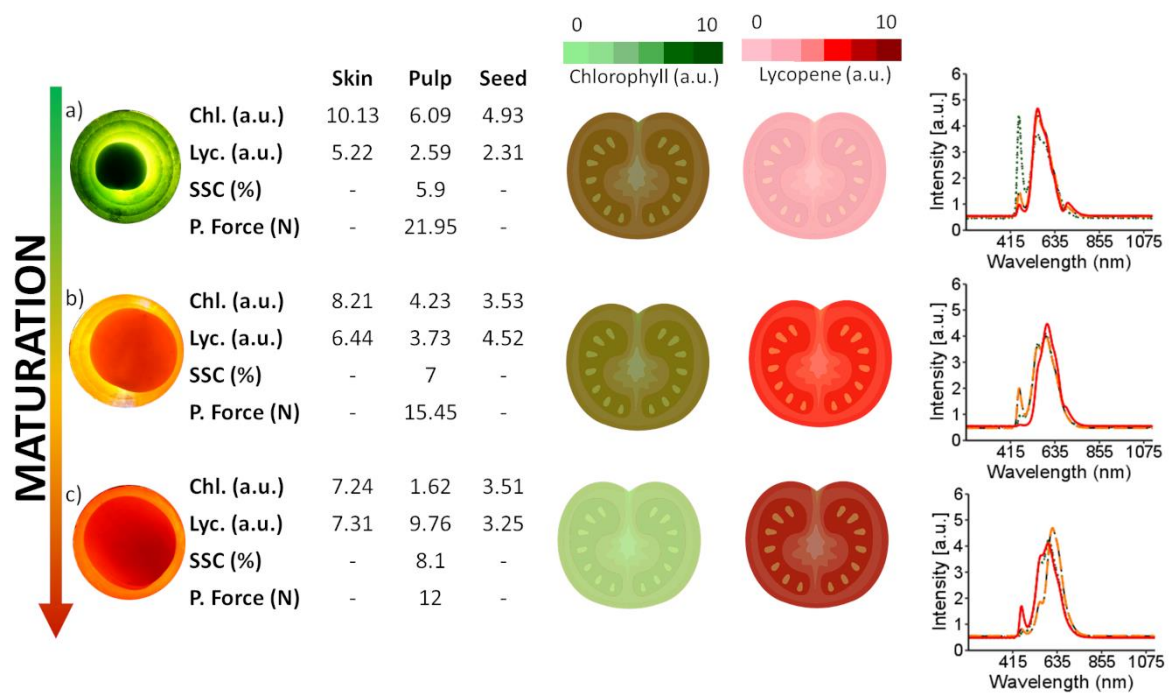


Figure 5.2-5. Spectral information obtained from tomatoes at different maturation stages and internal tissues and the dynamics of lycopene and chlorophyll during maturation. Chlorophyll content was empirically quantified, with light green representing lower levels and dark green representing higher levels. Lycopene levels were also quantified, with dark red representing lower levels and light red representing higher levels. a), b) and c) are red, turning and green tomatoes over the light source, respectively. The green spectra signature (□) represents the spectral signature of the skin, the orange spectra signature (□) represents the spectral signature of the pulp, and the red spectra signature (□) represents the spectral signature of the seed.

5.2.4. Discussion

This paper presents a disruptive methodology for the bi-directional reconstruction of whole tomato hyperspectral data and internal tissues (skin, pulp, and seeds). This non-destructive method can explain how different tomato tissues behave at various stages of the ripening process. As described in Figures 1, 2 and 3, this work aims to bi-directionally reconstruct the spectral information of the tomato from the data of the skin, pulp, and seed and to decompose the information of the tomato spectra into the internal tissues. Although each tomato tissue presented a particular space (Figure 5.2-3), creating a superset with the scores of each tomato fraction made it possible to reproduce the entire

spectral tomato (Figure 5.2-4). The most important LV of each fraction formed the superset used to reconstruct the tomato spectra. Through hierarchical PLS, the tissues could predict the entire tomato spectral data. The decomposition of the whole tomato data and the same number of LV combined in the hierarchical PLS can predict the tomato tissues. The literature reports that the tomography-like approach (Martins et al., 2023) successfully worked in grapes (Tosin et al., 2023b), and the results of this paper support that it can be applied to aqueous fruits like tomato. Due to the complex nature of tomato maturation and the diverse biochemical compositions of its internal tissues (Skolik et al., 2019), encompassing skin, pulp, and seed, this work presents a technique for the bi-directional spectral reconstruction of tomatoes using Vis-NIR data.

The literature offers several methodologies (Mishra et al., 2021) that could be used for spectral reconstruction. For instance, O2-PLS (Trygg & Wold, 2003) and OnPLS (Lofstedt et al., 2013a) utilise spectral data's local and global joints, bioheat models (Marin et al., 2021; Alzahrani & Abbas, 2019) could be adapted to predict the internal tomato tissues and adaptive neuro-fuzzy inference system (Abdullahi et al., 2022; Abdullahi et al., 2021), a hybrid computational model that combines the adaptive capabilities of neural networks with the interpretability of a mathematical framework that deals with uncertainty and imprecision in decision-making. However, these methods are not hierarchical and do not allow for convolution and fusion of information in a superset or deconvolution in a reverse way. Similarly, advanced approaches such as deep learning (DL) can deal with complex data (Mishra et al., 2022) and reconstruct the whole tomato with spectral information from the different tissues. Nevertheless, the decomposition of the tomato spectra into the spectral data of its tissues becomes even more complex and challenging. Nonetheless, the results presented in Table 5.2-1 show that hierarchical PCA combined with PLS can effectively perform bi-directional modelling with the directions $\mathbf{X} - \mathbf{Y}$ and $\mathbf{Y} - \mathbf{X}$, and remove orthogonal data between $\mathbf{Y}$ and $\mathbf{X}$, facilitating the reconstruction of the tomato and its internal tissues.

Owing to the multiple internal tissues that compose the tomato fruit, it faced a challenge in reconstructing the complete spectral information of the tomato (Figure 5.2-4). Furthermore, separating and identifying different tissues in green tomato are difficult because green tomato presents more fibre concentration (Chandra & Ramalingam, 2011) and pectin (Huang et al., 2018b; Moco et al., 2007). As a solution, all internal tissues (except the jelly parenchyma) were considered the pulp. Nevertheless, green tomatoes have less interest when compared with more advanced stages of maturation. Spectral data acquisition encountered a few challenges related to the position of the tomato in the system used to obtain spectral data. First, green tomato is opaque, limiting

light's ability to pass through (Figure 5.2-4). This limitation requires the optical fibre probe to be well-positioned to obtain more light signals. However, searching for light using the probe may not obtain a signal from all internal structures, which limits tissue reconstruction. A similar effect was observed in matured tomato. Depending on the probe position, some internal tissues may not be assessed, or less information may be obtained. Second, the tissues considered as the pulp will affect the reconstruction of the entire tomato.

During the data acquisition of the entire tomato, almost all the internal tissues are expected to be evaluated using a fibre probe. However, to provide detailed information in the superset utilised to reconstruct the entire tomato, it may be necessary to individually assess each tissue considered as pulp and examine their specific details. Therefore, the errors observed in Table 5.2-1 and Figure 5.2-4 indicate less accuracy in the reconstruction of the entire green tomato and less in the other tomato fractions of the matured tomato. Nevertheless, this method can reconstruct the Vis-NIR spectra of tomato and internal tissues and be used for different purposes.

This work assessed the SSC, chlorophyll, lycopene, and puncture force of the tomato using the original and reconstructed spectra for each fraction of the tomato (Table 5.2-2). The SSC and puncture force were assessed in the entire tomato, and the spectral data of each fraction were considered in the modelling. Chlorophyll and lycopene provided spectral information for each tissue and empirically demonstrated the concentration of each pigment in the tissues. The original and reconstructed spectra results were very similar for reconstructing the entire tomato and the reconstructed spectra (Table 5.2-2).

Among the quality parameters assessed, the SSC results were the most stable and robust for the original and reconstructed spectra. Nevertheless, it is essential to highlight that the different tomato tissues present distinct SSC concentrations, and the individual tissues could have been assessed to predict the concentration of SSC in each fraction. For chlorophyll and lycopene, the ratio of the zones of the spectrum empirically demonstrated the different concentrations of pigments in the distinct tissues (Figure S 5.2-2, 3, 4 and 5). Considering all the tomato tissues, the skin is the part that is indicated to have more concentration of chlorophylls. This study reveals that the skin has the highest concentrations of chlorophyll and lycopene, except during the red maturation stage. These results align with existing literature, particularly Chandra et al. (2012), highlighting the skin's elevated lycopene levels compared to pulp and seeds. However, in terms of chlorophyll content, the skin ranks as the second lowest tissue during maturation, as observed by Moco et al. (2007). The qualitative approach to classifying

internal tomato tissues throughout maturation may contribute to these variations compared to the quantitative methods in the cited literature.

Consistent with prior research (Chandra et al., 2012; Moco et al., 2007), this paper reveals that the skin exhibits a higher concentration of lycopene, as shown in Figure S 3. Puncture force analysis indicates elevated values in green tomatoes, probably attributable to heightened fibre and pectin content. This coincides with lower levels of SSC and lycopene, alongside increased chlorophyll concentrations (Huang et al., 2018b; Moco et al., 2007).

The temporal dynamics of maturation across distinct tomato tissues are demonstrated in Figure 5.2-5, where chlorophyll concentrations, particularly in the skin, are higher in green tomatoes (Figure 5.2-5a). Conversely, matured tomatoes (Figure 5.2-5c) tend to exhibit increased lycopene concentrations, mainly in the skin. Considering the spectral signatures of the tomato fractions (Figure 5.2-5), further studies can be conducted to determine the type of information that can be extracted. Empirically assessing the full tomato spectrum (Figure 5.2-5), two bands peak near 500 nm and 690 nm in green tomato, probably related to chlorophyll (Huang et al., 2018b; Ecartot et al., 2013). In the spectral skin signature, a similar peak (near 490 nm) in the green tomato may be related to chlorophyll a. When the pulp and seed were analysed, the same peak (near 490 nm) was present but with less intensity, suggesting a lower chlorophyll concentration in those tissues. The peaks near 550 nm are related to the carotene group (Ciaccheri et al., 2018), especially the lycopene concentration. For matured tomato, these peaks present more intensity in the full tomato spectra, skin, and pulp, suggesting a higher concentration of lycopene when compared with less mature tomatoes.

Vis-NIR data can enhance the efficiency and quality of crop production by providing valuable information for optimising various agricultural practices, such as irrigation, fertilisation, and pruning (Xia et al., 2021). Furthermore, by leveraging this data, crop growers can gain precise insights into the maturation process of fruits, which is influenced by a range of biotic and abiotic factors, including diseases, water availability, temperature, and light intensity.

The methodology presented in this paper has the main advantage of providing more accurate and detailed information about the internal structure of the fruit as a tomography-like system when compared to the whole-fruit measurement by using hyperspectral or multispectral data (e.g., Mishra et al., 2023; Mishra & Woltering, 2023) that obtain majority external information. The tomography-like presented in this paper allows for the assessment of individual tissues, facilitating the acquisition of more accurate and detailed information about the internal structure of the fruit (Martins et al.,

2023). Figure 5.2-3 represents the feature space of the entire tomato and the skin, pulp and seed, where the different maturation stages occupy distinct feature spaces. In contrast, traditional whole-fruit measurements often lack precision in providing insights into the inner tissue of the fruit. Also, the methodology presented in this paper can lead to more accurate predictions of the internal tissue properties and better quality control. It can also enable a more specific and targeted analysis of internal tissue properties by offering comprehensive and high-dimensional data. These data provide more detailed information about the fruit's internal structure and can also be utilised to determine additional components beyond those presented, particularly in supporting metabolomic studies. This knowledge can help to fine-tune agricultural practices and mitigate potential risks, ultimately leading to improved crop yields and higher-quality produce.

This paper presents a novel technique for determining the quality parameters SSC (%), chlorophyll (a.u.), lycopene (a.u.) and puncture force (N) of fruits using visible and near-infrared (Vis-NIR) spectroscopy of their skin, pulp, and seed. The approach builds upon a growing body of research demonstrating the ability of hyperspectral sensors to measure a wide range of quality parameters non-destructively and accurately in crops (Tosin et al., 2023b; Tosin et al., 2022; Martins et al., 2022). Furthermore, by leveraging the high-dimensional data obtained for each tissue, the method showcased in this study has the potential to unlock a multitude of additional quality parameters during fruit maturation. This capacity for rapid and precise determination could enhance fruit production's efficiency and effectiveness while elevating the final product's overall quality. Finally, it is worth noting that the extensive dimensionality of the data obtained for each tissue opens possibilities for identifying and characterising other components beyond those currently presented, particularly in supporting metabolomic studies.

5.2.5. Conclusion

This paper proposes a tomography-like system that can predict the Vis-NIR information of the internal tissue. Applying multi-block hierarchical component analysis in conjunction with PLS enables the bi-directional reconstruction of spectral information, facilitating the prediction of internal tissue spectra (skin, pulp, and seed) and the decomposition of the overall tomato spectral information into its constituent tissues.

This novel approach allows assessing tomato maturation dynamics by analysing internal tissue characteristics, offering pertinent information for precision agricultural practices. Moreover, the method can identify physiological issues related to abiotic (e.g., water stress, high temperature) and biotic (e.g., bacterial infection). These identified stressors

can be integrated into multifaceted omics techniques to understand the plant's physiological responses.

Building on successful testing in grapes this technique, demonstrates its efficacy in the complex tissue structure of tomato. Thus, the same approach could be applied to other aqueous fruits, such as blueberries. However, further work is necessary to test the applicability of this technique in other fruits, to study the dynamic of the Vis-NIR information with the internal tissues during the maturation process, and to incorporate additional analytical data for validation.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Renan Tosin: Investigation, Methodology, Writing- Original draft preparation, Writing- Reviewing and Editing. **Mario Cunha:** Supervision, Investigation, Methodology, Writing - Original draft preparation, Writing – Reviewing and Editing. **Filipe Monteiro-Silva:** Methodology. **Filipe Santos:** Investigation. **Teresa Barroso:** Methodology. **Rui Martins:** Supervision, Investigation, Methodology, Writing- Original draft preparation, Writing- Reviewing and Editing.

Acknowledgements

Renan Tosin and Filipe Monteiro-Silva acknowledge Fundação para a Ciência e Tecnologia (FCT) PhD research grants Ref. SFRH/BD/145182/2019 and SFRD/BD/09136/2020. Martins RC acknowledges Fundação para a Ciência e Tecnologia (FCT) research contract grant (CEEIND/017801/2018). This work is financed by National Funds through the FCT – Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project OmicBots – OmicBots: High-Throughput Integrative Omic-Robots Platform for a Next Generation Physiology-based Precision Viticulture, with reference PTDC/ASP-HOR/1338/2021.

Supplementary Material

Figures

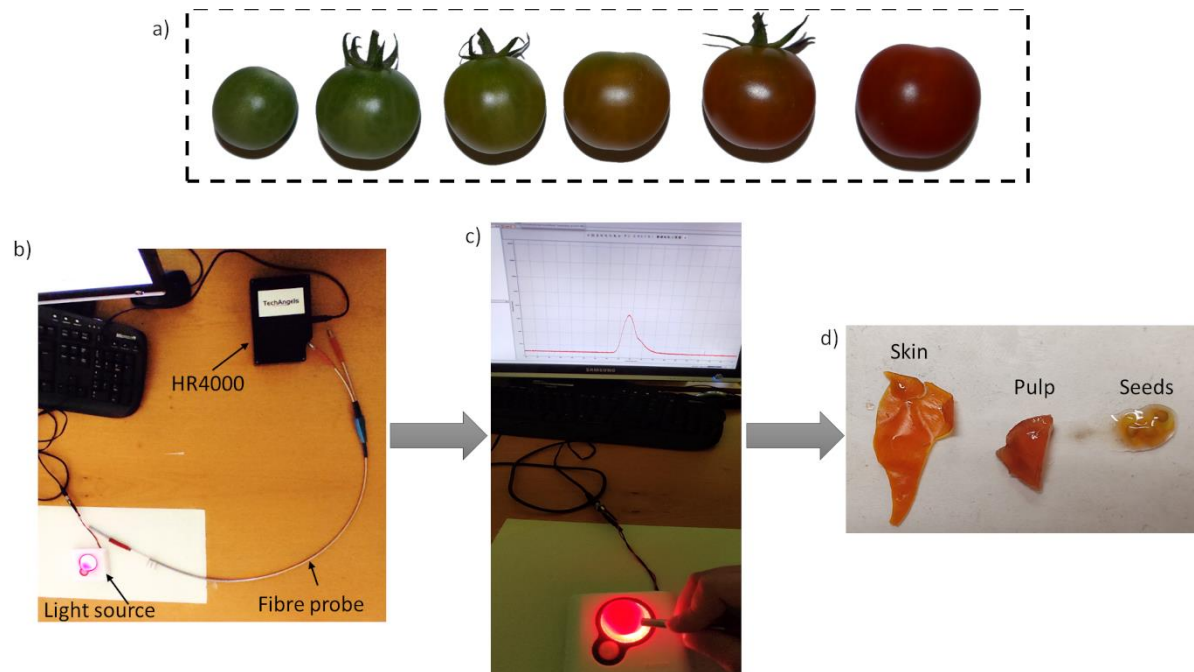


Figure S 5.2-1. Gradient of tomatoes (a) in different stages of maturation used in the experiment and the tomography-like system used to obtain the spectral information of the tomato and the internal tissues (skin, pulp and seed). In (b), the spectrometer (HR4000), the fibre probe, and the light source with the tomato are described; (c) it demonstrates how the spectral information is obtained; (d) shows the aliquots of the skin, pulp, and seeds considered for measurements.

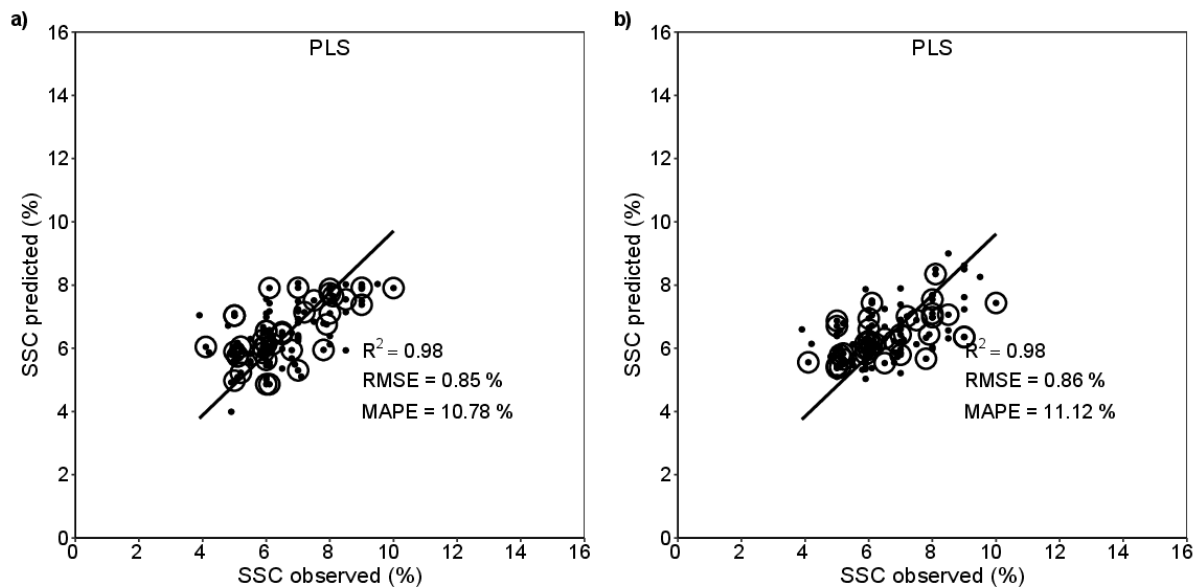


Figure S 5.2-2. Regression plots for the soluble solid content (SSC) in the original (a) and reconstructed (b) spectra considering the whole tomato. Circled observations represent the validation dataset.

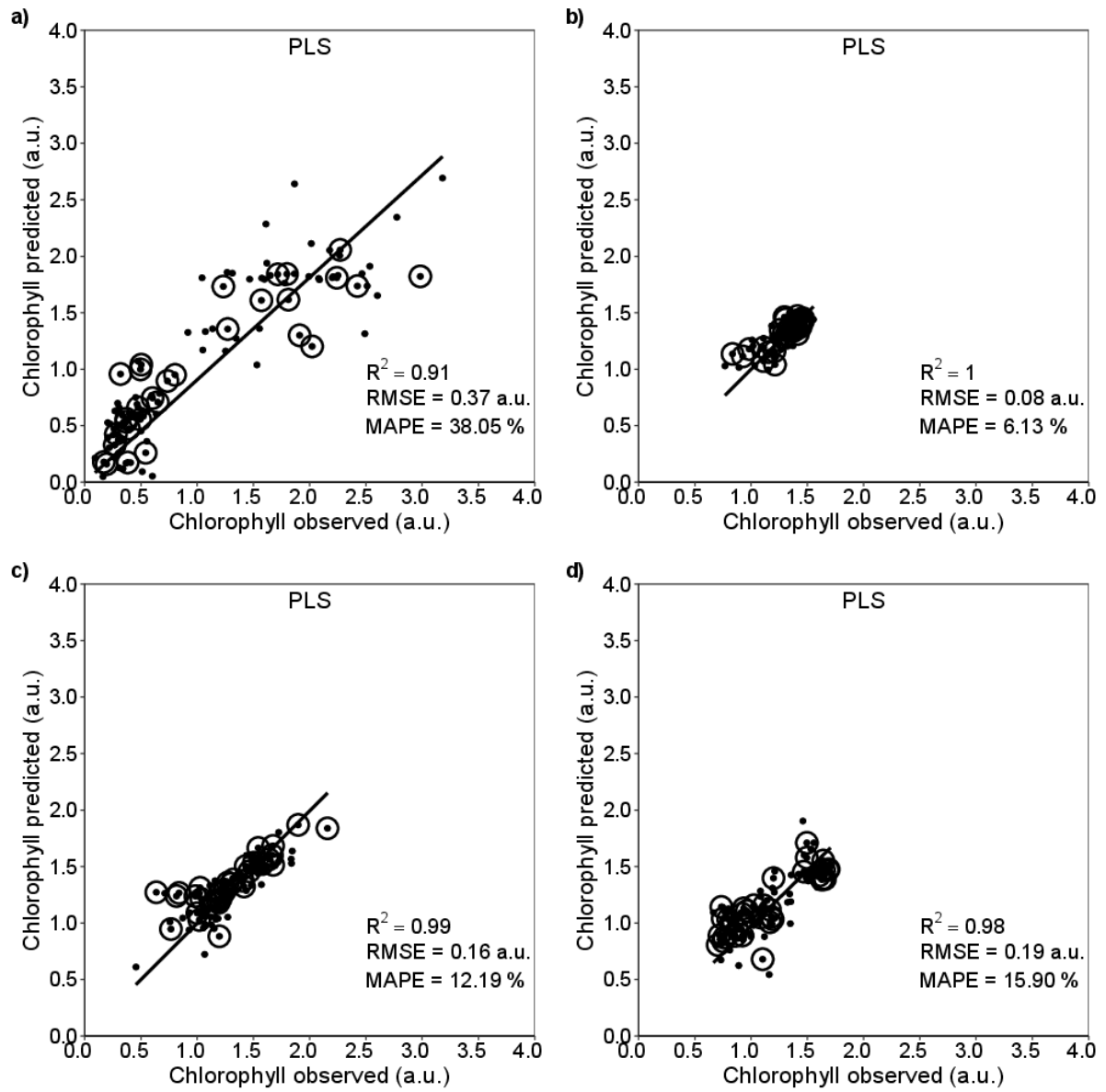


Figure S 5.2-3. Regression plots for the chlorophyll (a.u.) in the reconstructed spectra considering the whole tomato (a) and the skin (b), pulp (c) and seed (d). Circled observations represent the validation dataset.

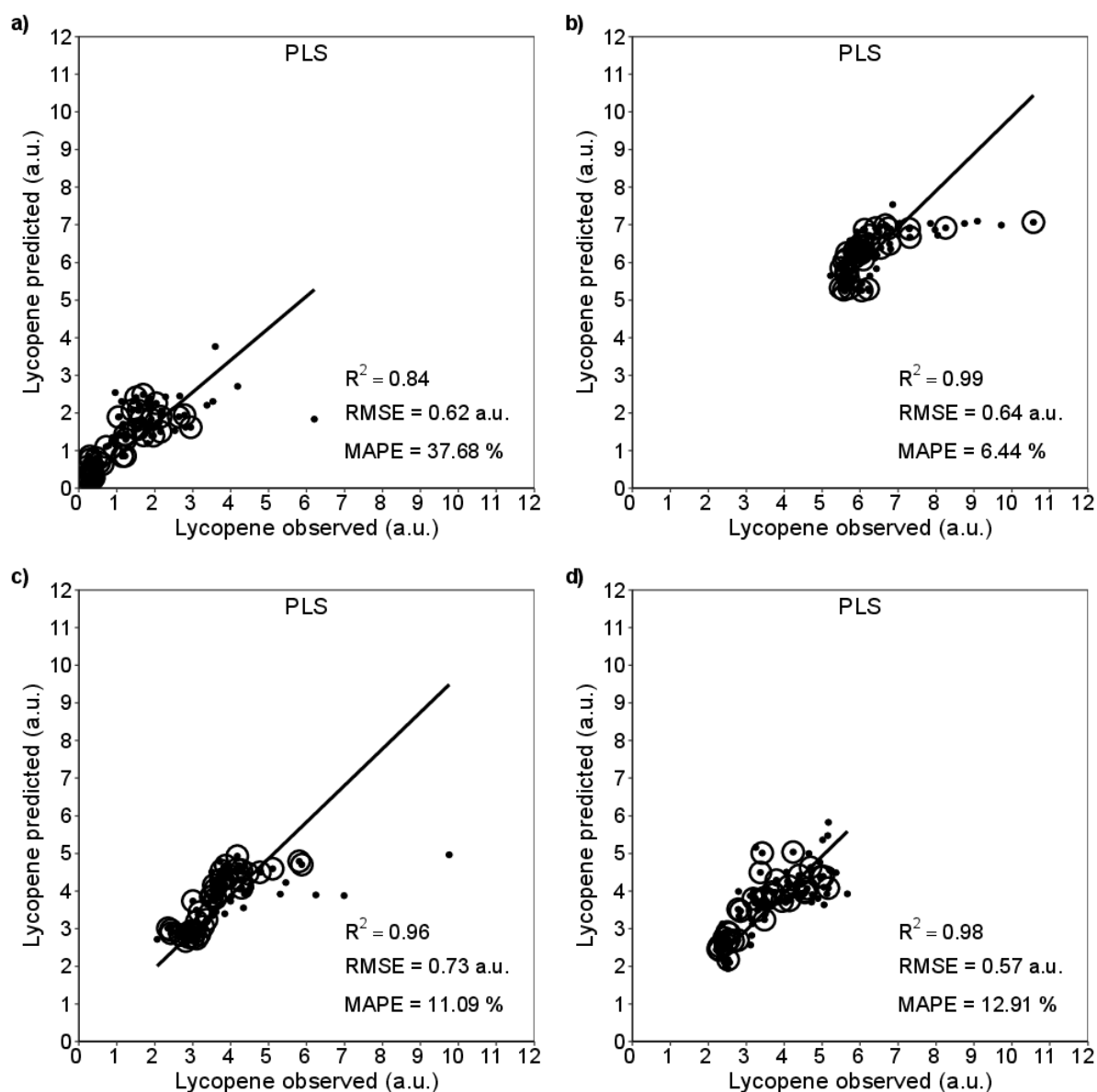


Figure S 5.2-4. Regression plots for the lycopene (a.u.) in the reconstructed spectra considering whole tomato (a) and the skin (b), pulp (c) and seed (d). Circled observations represent the validation dataset.

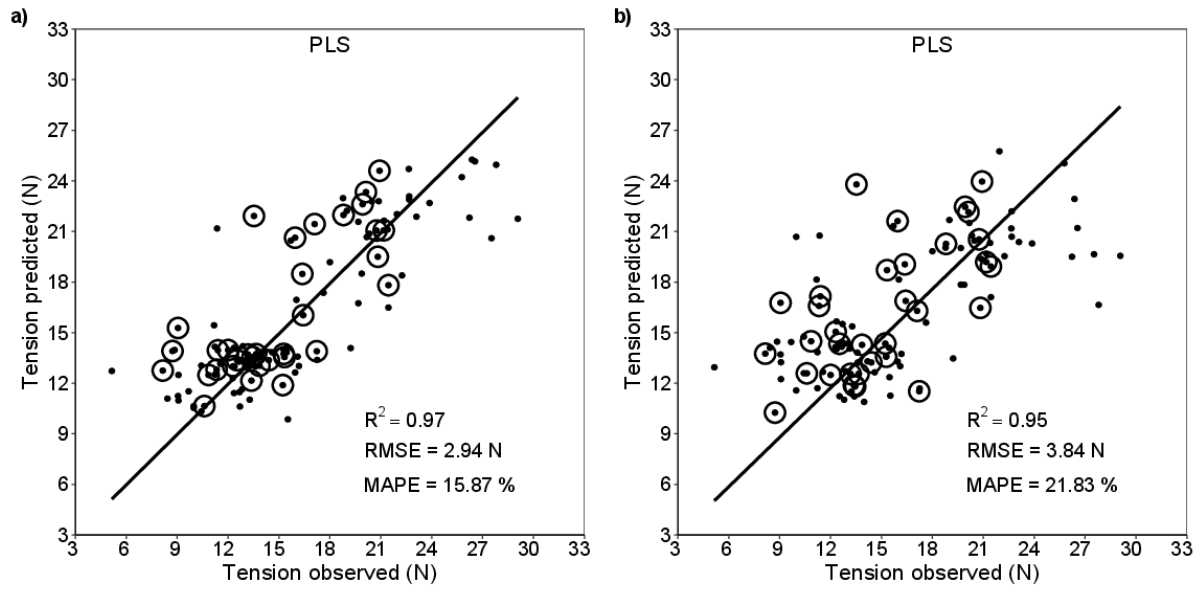
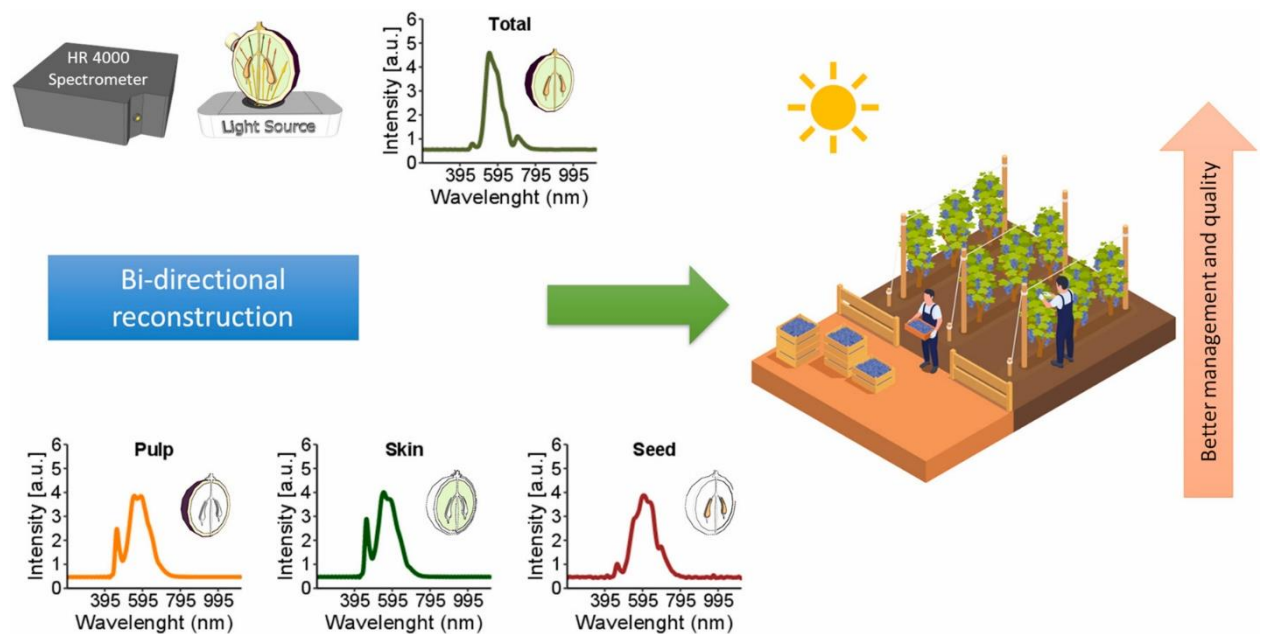


Figure S 5.2-5. Regression plots for the puncture force (N) in the original (a) and reconstructed (b) spectra considering the whole tomato. Circled observations represent the validation dataset.

5.3. Case study V: Grape case study: comparing cultivars, tracking quality changes during maturation.

Tosin, R., Monteiro-Silva, F., Martins, R., and Cunha, M. (2023). Precision maturation assessment of grape tissues: Hyperspectral bi-directional reconstruction using tomography-like based on multi-block hierarchical principal component analysis. *Biosystems Engineering* 236, 147-159. doi: 10.1016/j.biosystemseng.2023.10.011



Graphical abstract. 5.3-1. Graphical abstract case study V (Tosin et al., 2023b).

Precision maturation assessment of grape tissues: hyperspectral bi-directional reconstruction using tomography-like based on multi-block hierarchical principal component analysis

Renan Tosin^{1,2}, Filipe Monteiro-Silva², Rui Martins^{2,*} and Mario Cunha^{1,2}

¹ Department of Geosciences, Environment and Spatial Planning, Faculty of Sciences of the University of Porto, Rua do Campo Alegre, S/N, 4169-007, Porto-Portugal

² INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, S/N, 4200-465, Porto-Portugal

* rui.c.martins@inesctec.pt

Highlights

- Bi-directional spectral reconstruction predicts the skin, pulp, seed, and total grape.
- Spectral reconstruction works for a white and red grape cultivar.
- Reconstructed spectra can estimate SSC, chlorophyll, anthocyanin and puncture force.
- Tomography-like can support viticultural decisions based on grape physiology.

Abstract

This paper introduces a tomography-style method for assessing grape maturation. It analyses inner tissue spectra through point-of-measurement (POM) sensing. A multi-block hierarchical principal component analysis (MHPCA) algorithm was used for the spectral reconstruction of total grapes (skin, pulp, and seed). Two grape cultivars, Loureiro (white; $n = 216$) and Vinhão (red; $n = 205$) were measured at 12 dates after veraison. The reconstructed spectra showed no significant differences ($p < 0.001$) from the originals for both grapes. Loureiro had better statistical metrics (total grape: $r = 0.99$, skin: $r = 1$; pulp: $r = 1$, seed: $r = 0.94$) than Vinhão (total grape: $r = 0.92$, skin: $r = 0.92$; pulp: $r = 0.95$, seed: $r = 0.95$). Using self-learning artificial intelligence (SL-AI), the following parameters were predicted for both grapes: soluble solids content (%; MAPE < 13%), puncture force (N; MAPE < 29%), chlorophyll content (a.u.; MAPE < 29%), and anthocyanin content (a.u.; MAPE < 17%, Vinhão only). When comparing observed values with predicted skin, pulp, and seed spectra, Vinhão showed no statistical differences for most parameters, except pulp chlorophyll on one DAV in the final maturation stage. The same was done with the Loureiro cultivar. Although Loureiro

mostly showed no statistical differences in assessed parameters across tissues and dates, variations were found in pulp and skin chlorophyll content and puncture force. This tomography-like approach based on tissue maturation can help viticulturists to access instant data on grape maturation, supporting informed decision-making and promoting more sustainable agricultural practices.

Keywords

Fruit quality; precision maturation; SCC; pigments; spectroscopy; proximal sensing

5.3.1. Introduction

Grape maturation involves complex dynamic transformations in various tissues responsible for its quality. During the maturation, the skin (SK) undergoes colour changes and phenolic compounds accumulation (Farhadi et al., 2016), the pulp (PU) changes the pigmentation, increases the phenolic compound and sugar content (Carvalho et al., 2015), and the seeds (SE) become darker in colour and harder in texture, turning from a greenish hue to brown or black, also changing their chemical composition by increasing tannin levels while the seed's water content decreases (Ribalta-Pizarro et al., 2021; Ristic & Iland, 2005). This maturation dynamic process is influenced by environmental conditions and agricultural practices, which, in turn, affect the composition of the inner grape tissues (Yuan et al., 2018), possibly impacting the maturation time lag of these tissues (Ribalta-Pizarro et al., 2021). Despite the effectiveness of traditional grape assessment methods in predicting quality parameters, they often lack to account for the maturation variability across individual grape tissues. Traditional wet lab assessments of grape maturation involve harvesting grapes at different stages and locations to monitor the process. These destructive methods use chemistry-based techniques to measure sugar content, acidity, and phenolic compounds (e.g., Fernandez-Novales et al., 2019a). However, they are costly, time-consuming, and labour-intensive. Additionally, these methods cannot map important parameters like sugar content and phenolic composition throughout the entire field, including grape SK, PU, and SE, due to the high debit of data needed.

To address these limitations, non-destructive assessment techniques using spectral devices offer alternatives for evaluating grape maturation. Spectroscopy devices equipped with Vis-near-infrared spectroscopy (Martins et al., 2022; Tosin et al., 2021), Raman spectroscopy (Ding et al., 2021; Nogales-Bueno et al., 2017), fluorescence spectroscopy (Agati et al., 2013; Baluja et al., 2012b), thermal imaging (Gouot et al., 2019), and magnetic resonance imaging (Andaur et al., 2004) provides non-destructive

Spectral devices used for grape maturation assessment can utilise different methods for light interaction. Image and no-image sensors generally measure the reflectance of solar incident light (Figure 5.3-1a) with a field of view angle (e.g., Tosin et al., 2021). Still, they can also capture spectral information when coupled with a light source (Fernandez-Novales et al., 2019a). On the other hand, point-of-measurement devices (POM) can operate with light transmittance (Martins et al., 2022), determining the amount of light that passes through the sample by employing a light source that traverses the fruit (Figure 5.3-1b). When used with a LED to obtain light transmittance in the fruit, POM devices appear more capable of getting spectral information on the internal tissues than reflectance-based sensors (Figure 5.3-1a). Thus, it is hypothesised that POM can provide detailed information on grape SK, PU, and SE to assess the maturation of each tissue individually.

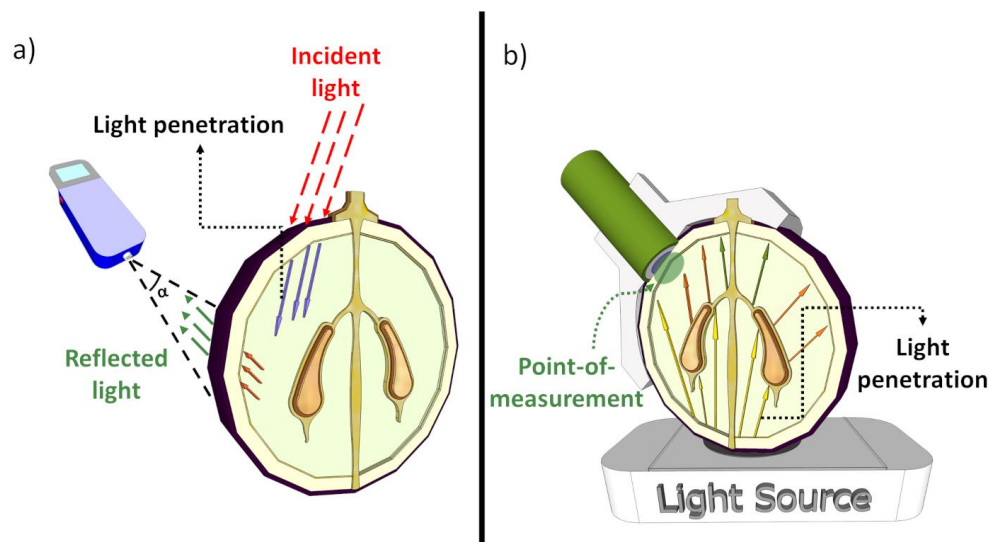


Figure 5.3-1. Comparison of spectral information acquisition methods for grapes. (a) Proximal sensing device measurement obtains spectral information until a particular depth within the grape, α is the conical field-of-view. This method allows for the assessment of the grape's spectral characteristics from its surface to a certain internal depth. (b) Point-of-measurement enables the acquisition of spectral data from main internal grape tissues. This method captures spectral information from specific points within the grape's internal structure, providing detailed information into its inner composition and characteristics.

As the total grape spectra obtained from POM devices can contain detailed information on the SK, PU and SE spectral data, specific modelling techniques are required to process these data. For instance, principal component analysis (PCA) can effectively handle and correlate variables, making it suitable for addressing this challenge (Trygg & Wold, 2003). Other methods, such as multi-block hierarchical PCA (MHPCA) and hierarchical partial least squares (HPLS), can handle spectral information from various sensors and datasets (Mishra et al., 2021). The application of MHPCA analysis is hypothesised to fitting for the bi-directional spectral reconstruction of the total grape (TO) using spectral data from the SK, PU, and SE, creating a technique called “tomography-like”.

The term “tomography-like” used in this work for grape tissue reconstruction is based on the hierarchical latent relationships between the observed spectral patterns of grape tissues. The method provides detailed information about the internal structures of grapes. MHPCA establishes bi-directional connections between the overall grape spectra and the individual grape SK, PU, and SE spectra. This non-destructive approach enables the assessment of the composition of grape tissues. The process involves capturing multiple spectra from different grape positions to create a 3D image with resolved information on tissue distribution. Although the term “tomography-like” is debated as it primarily implies tissue image resolution, this approach extends to provide median spectra for each tissue. Data-driven reconstruction, such as MHPCA, is an approach that differs from transitional imaging methods based on inverse problem numerical methods (e.g. magnetic resonance imaging), known for its efficiency and versatility, making it applicable in various research areas (Martins et al., 2022; Bar-Sinai et al., 2019b).

Therefore, by conducting spectral analysis of grape tissues from white and red grapes throughout their maturity process, this study aims i) to compare the spectral characteristics of each tissue obtained from intact grapes with those obtained from destructive samples, ii) to assess the accuracy of the reconstructed spectra for each tissue in determining the biophysical characteristics of that tissues: soluble solids content (SSC), chlorophyll, anthocyanin content, and puncture force (PF) and iii) to explore the sensitivity of the developed non-destructive tomography-like to discriminate between different states of tissues during maturation, as well as their analytical composition in white and red grape varieties. The specifics goals include: i) developing the main algorithms to establish the MHPCA analysis for the bi-directional reconstruction and decomposition of the grape tissues, ii) to infer on hyperspectral sensing technologies for bi-directional grape tissue reconstruction, iii) developing a proof-of-concept for the causal

relationships between specific spectra characteristics and the physiological changes that occur during the grape tissues maturation and iv) to provide a foundation for further research on translation between wet lab and dry lab techniques to support advanced viticulture and wine practices.

5.3.2. Material and Methods

5.3.2.1. Test site

The experiment was conducted in the Região dos Vinhos Verdes, located in the northwest region of Portugal, specifically at the Campus Agrário de Vairão (41°19'31.91" N; 8°40'27.45" W) in Vila do Conde. Grape sampling took place in 2020 after veraison (August 14th) for two grape cultivars: Loureiro (Vitis international variety catalogue (VIVC) 25085), a white grape cultivar, and Vinhão (VIVC 13100), a red grape cultivar. Sampling for both cultivars was performed on August 19th (05 days after veraison, 05 DAV), 21st (07 DAV), 25th (11 DAV), September 3rd (20 DAV), 7th (24 DAV), 8th (25 DAV), 24th (41 DAV), 28th (45 DAV), 29th (46 DAV), 30th (47 DAV), October 12th (59 DAV – only Vinhão), October 13th (60 DAV – only Loureiro), and 14th (61 DAV). A total of 20 grapes per sampling date were randomly selected, considering different grapevines and bunches for each cultivar with a concerted effort to minimise potential sources of variability by: i) prioritising grape bunches for their representativeness and uniformity in size and maturity within the same cultivar; ii) taking into account the position of grape bunches on the vine to mitigate variations arising from differing sunlight exposure, nutrient availability, and microclimatic conditions; and iii) factoring in the accumulation of biomass and grape bunch size in the selection process to ensure that the chosen samples were at similar stages of development. Nevertheless, some grape samples were excluded from the study due to unexpected damage suffered during transportation to the laboratory. Therefore, this study analysed 216 grapes of Loureiro and 205 grapes of the Vinhão cultivar. Figure S 5.3-1 describes the meteorological conditions during the sampling dates.

5.3.2.2. Grape measurements

Grape width (mm) and high (mm) were measured with a digital calliper (LiMiT Caliper Digital 150mm 190140103) to calculate the grape volume (cm³) through the sphere volume formula:

$$Volume(cm^3) = (4/3) * \pi * (a^2) * b \quad (1)$$

Where a is half of the width (cm) of the grape and b is half of the length (cm) of the grape.

The maximum PF (N) was measured using a digital penetrometer (model PCE-PTR 200, PCE Group, D-59872 Meschede, Germany). Subsequently, SSC (%) was immediately measured using a hand refractometer (Milwaukee model MR32ATC) with a scale range from 0 to 32.0%.

Skin (SK), pulp (PU), and seeds (SE) were meticulously isolated using a lancet for each grape in both cultivars. About 0.5 cm² and 3 mm thick portions were placed on separate slides. This method entailed precise SK peeling, PU slicing, and putting a single developed SE on each slide. This method entailed precise SK peeling, PU slicing, and putting a single developed SE on each slide. After veraison, the size and shape of the seeds do not present substantial changes (Ristic & Iland, 2005), but for this study, measuring one seed is enough to make the proof-of-concept.

This rigorous process enabled individual tissue acquisition for distinct SK, PU, and SE spectral data.

The content of chlorophyll and anthocyanins in different tissues was empirically determined using spectral information collected from the small aliquots of each tissue. The quantification of chlorophyll (a.u.) can consider the green (520-570 nm) and the red domain of the spectrum (571-700 nm), as referred by Zhu et al. (2020). Therefore, this work considered the mean band ratio between the green spectral zone (520-570 nm) and the red spectral zone (571-700 nm) in estimating the chlorophyll content. Similarly, anthocyanin content is highly associated with the red-edge zone of the spectra, reporting the wavelengths 680, 675, and 700 as positively correlated with this pigment in grapes (Rustioni et al., 2013). Therefore, this work estimated anthocyanin (a.u.) by calculating the band ratio 705 nm by 680 nm.

The pigments were inferred through spectroscopy due to the available methods presented in the literature (e.g., Farhadi et al., 2016) for small tissue aliquots (0.5 cm² and 3 mm thick) being non-expedite, expensive, time-consuming and non-optimised. Additionally, these conventional methods were ill-suited for the low technology readiness level (TRL) by the substantial expenses associated with sophisticated wet lab methods for small tissue samples, making them impractical for early-stage technology development. The resulting data demonstrated the efficacy of reconstructed tissue spectra in establishing a relationship with estimates derived from grape spectra (Martins et al., 2022), which, in essence, serves as a proof-of-concept for the underlying principle of internal tissue quantification.

5.3.2.3. Spectral measurements

The spectral measurement of the total grape and the SK, PU and SE involved the following steps: i) positioning the grape above a platform equipped with a deep-power LED (6500k, Philips SpotOn Ultra 69141/31/PH, with light emission from 380 nm to 780 nm) at the bottom. ii) obtain spectral data a high-resolution spectrometer (Ocean Insite HR4000, 195.34-1118.33 nm) equipped with a reflectance fibre optic probe (Ocean Insite). The fibre probe was positioned to obtain the maximum transmitted light from the tissues assessed (skin, pulp and seed). The incident light comes from the skin's inner side, and the local with maximum light transmittance was chosen to position the fibre probe. For the seed's spectra recording, the pulp involving the seed was completely removed, and the seeds were positioned on the dorsal side. The positioning of the fibre probe had to be adjusted in the seed due to the characteristics of the LED. The desirable position would be where structures such as nucellus, endosperm, chalaza, vascular bundles, inner integument and outer tegument could be obtained in the light transmittance. However, due to the characteristics of the LED these tissues could not be obtained in most cases due to the lack of transmitted light (Figure 5.3-2). Therefore, we had to choose the local with maximum transmitted light in these cases, probably skipping the chalaza, vascular bundles and part of the endosperm and nucellus and iii) optimising the integration time for each sample to adjust data acquisition durations to ensure that most spectra remained within the linear response range of the spectroradiometer. This customisation balanced spectral information capture and detector saturation avoidance. Longer integration times were used for samples with low light transmittance to improve signal-to-noise ratios and detail in spectral data. Conversely, shorter integration times were applied to samples with higher optical properties to prevent detector saturation. This approach optimised precision in the spectroscopic analysis while accommodating the varying optical characteristics of grape tissues. The key consideration was that these varied integration times did not significantly impact the signal-to-noise ratio, ensuring that the overall measurement results remained robust, remaining within the sensor linear response. Despite the limited spectral range of the LED illumination, all the obtained spectra data from the spectrometer were included in the subsequent analysis.

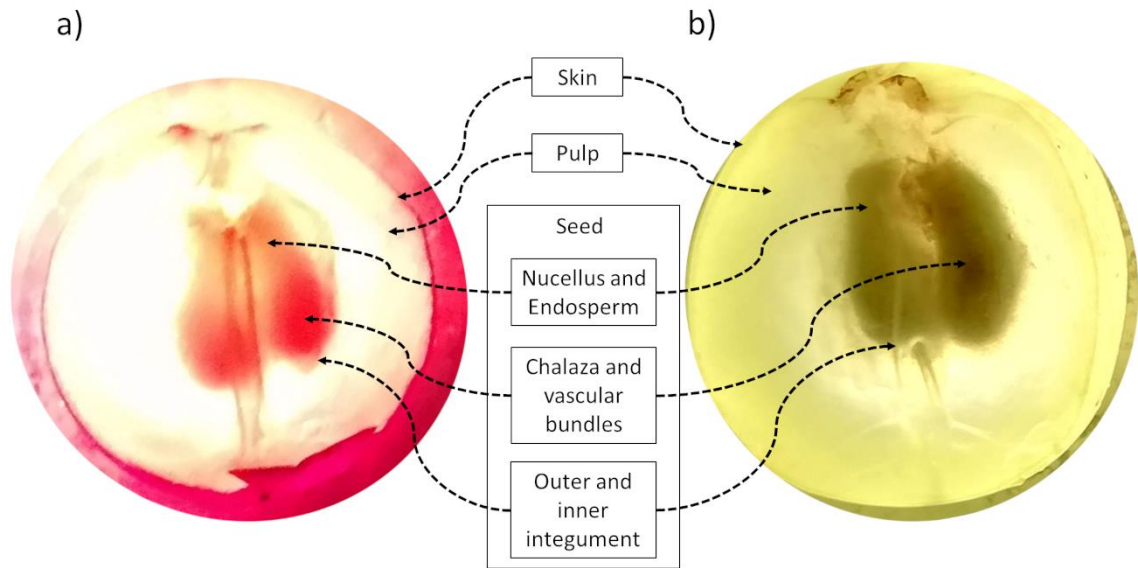


Figure 5.3-2. A transversal cut of the grape of Vinhão (a) and Loureiro (b) cultivar to show the light transmittance and morphological description of the tissues and details of the seeds. The figure shows the distinct light transmittance patterns within the grapes and provides the structural characteristics of the seeds, showing their optical characteristics.

5.3.2.4. Bi-directional spectral reconstruction

Grapes' spectral components are interconnected, allowing bi-directional information flow and enabling identifying unknown samples by comparing them with similar ones. Analysing these connections reveals unique spectral patterns in different grapes, even when specific information is absent. Multi-block hierarchical principal component analysis (MHPCA) dissects data into latent structures for distinct pattern comprehension (Westerhuis et al., 1998). MHPCA hierarchically breaks down data into eigenvectors, creating a new basis for projecting dataset information. PCA derives orthogonal eigenvectors to maximise variance within the dataset. Singular value decomposition (SVD) is used after subtracting the mean of dataset X to perform PCA. Scores (latent structures) originate from the left singular matrix U , while loadings (eigenvectors) come from suitable matrix V^t .

Examining eigenvectors using randomisation tests reveals non-random information. Reconstructing spectra separates random and relevant data. For grape tissues, PCA decomposition is applied to skin X_1 , pulp X_2 , and seeds X_3 , producing latent features $([t_1, t_2, t_3])$ reconstructing tissue spectra (Figure 5.3-3) and discarding random data (e_1, e_2, e_3) . Latent information for each tissue and variance along eigenvectors' geometry is determined. Concatenating dimensions t_1 , t_2 , and t_3 forms a superset latent space T . This can identify relevant superset latent information (T_i) and each subset's contribution to fused information (P_i^t) (Figure 5.3-3).

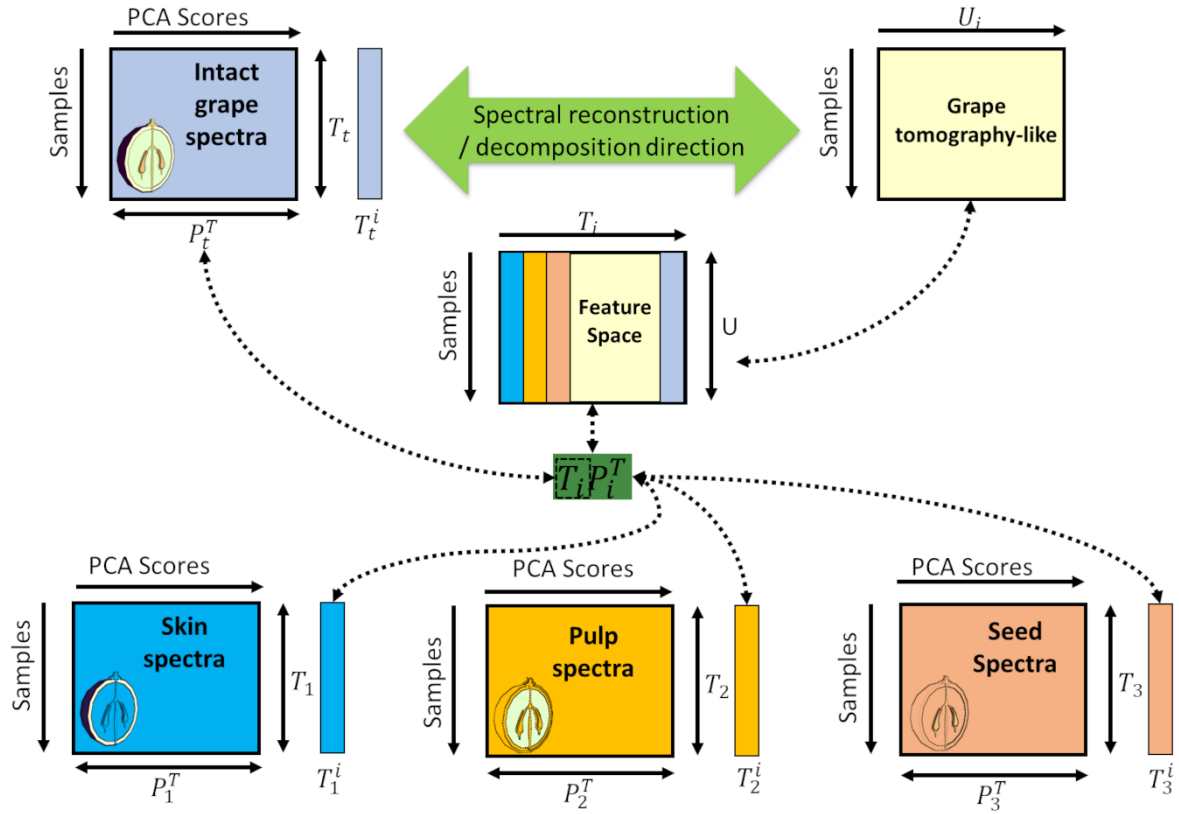


Figure 5.3-3. Representation of the spectral reconstruction and decomposition of the grape and the internal tissues. T is the superset latent space; T_i the superset latent information of the dimension of the subsets (T_1, T_2, T_3); U is the feature space of each subset; U_i the superset latent information of the dimension of the subsets of the total grape spectra; P_i^T is the contribution of the SK, PU and SE of each subset (P_1^T, P_2^T, P_3^T).

By constructing the superset latent structure, relevant SK, PU, and SE characteristics within the tissue dataset are represented and connected to grape spectra (Algorithm S 5.3-1 - Supplementary Material). Similar spectral patterns among samples lead to consistent superset and feature space patterns, aiding the identification of grape tissues in unknown samples. Integrating information from various grape parts captures the unique grape spectral traits comprehensively. The whole grape decomposition yields specific tissue info, enhancing individual spectral property understanding.

Relevant SK, PU, and SE characteristics in superset T are similar to direct PCA grape spectra decomposition (Y). This implies a direct link between latent spaces T and U , suggesting similar clustering for compositionally alike samples in both (Algorithm S 5.3-1 - Supplementary Material). This parallels grape SK, PU, and SE maturation stages.

Direct T and U linkage ensures bi-directionality between subsets X_1 (skin), X_2 (pulp), X_3 (seed), and Y . To determine unknown sample's tissues, Y is projected into U 's feature space ($U = YC$, being projection matrix). Neighbouring samples (k) of the unknown can be identified in this space. Verifying k propagation from T to t_1, t_2 , and t_3 is by reconstructing t_f , $t_f = [t_1^i, t_2^i, t_3^i]$, concatenating i^{th} dimension of t_1, t_2, t_3 . This fused

information constructs superset latent space T using Algorithm S 5.3-2 (Supplementary Material). Superset space is determined as:

$$t_f = T_i P_i^t \quad (2)$$

Here, T_i is relevant sub-level info and P_i^T is each subset's fused info contribution.

5.3.2.5. Validation of spectral reconstruction

Grape spectral components interconnect, enabling bi-directional info flow for identifying traits in unknown samples. This approach aids pattern discovery and understanding, improving grape study data analysis and interpretation. Spectral reconstruction validation occurred in two phases: i) hierarchical latent structure development – optimisation and validation, and ii) spectral data preprocessing – scattering correction and standardisation.

Hierarchical latent structure

This model involves optimisation and validation. Firstly, cross-validation determines optimal principal component (PC) numbers for t_1 , t_2 , and t_3 , sub-spaces and superspace T . Prediction tests and quantitative metrics are applied to hold-out samples. For tissue spectra database: skin X_1 , pulp X_2 and seed X_3 and relevant PCA decomposition:

$$X_1 = t_1 p_1^t + e_1 \quad (3)$$

$$X_2 = t_2 p_2^t + e_2 \quad (4)$$

$$X_3 = t_3 p_3^t + e_3 \quad (5)$$

Cross-validation (CV) determines the optimal error for each tissue reconstruction, $e = [e_1 e_2 e_3]$ (equations 3-5), for increasing the number of sub-space and superset principal components (PCs). The optimal number of PCs is considered the one that provides minimal CV errors and prevents over-fitting of random features at the sub-level passing into the superset. Hold-out (HO) samples double-check the accuracy of the CV metrics (Algorithm S 5.3-3 – Supplementary Material).

Bi-direction spectra reconstruction considered the following metrics: i) Euclidean distance to measure the similarity between neighbouring samples in the feature space (Fachada et al., 2014; Ramirez-Lopez et al., 2013), and ii) Principal component regression (PCR) to analyse the latent structures of the sub-level spectra, superset, and grape spectra. PLS was used to maximise the covariance between the spectra and grape composition, utilising eigenvectors of $X^t Y$ to ensure similar latent structures (Ergon, 2009; Phatak & De Jong, 1997). The number of deflation or latent variables was optimised through cross-validation/hold-out samples, minimising the predicted sum of

squares (PRESS) (Krstajic et al., 2014). PLS utilised an oblique projection to determine the coefficients $bpls$ in $Y = Xbpls$ (Ergon, 2009; Phatak & De Jong, 1997).

The evaluation of the reconstructed spectra of skin X_1 , pulp X_2 , seed X_3 , and Y considered the following metrics: i.) mean standard error (MSE) in counts/wavelength (nm); ii.) mean absolute percentage error (MAPE) in %/wavelength (nm), iii.) Euclidean distances in T and U of CV and HO samples and iv) Pearson's correlation associated with a p-value. These metrics determine the average reconstruction error per wavelength (nm) and the average bias per wavelength (nm) and measure the representativeness and stability of the knowledge base. If the training set is representative, CV and HO errors expect to be similar, and the model can efficiently reproduce the spectral information, thus verifying the null hypothesis.

Prepossessing spectral data

A logarithm multiplicative scattering correction (MSC-log) method was applied to both the original and reconstructed spectra, which performs baseline and scattering artefact correction (Martins et al., 2022). This correction mitigates unwanted variations and ensures more accurate and reliable data analysis. Additionally, a standardisation was applied to the original and reconstructed data to account for differences in signal intensity.

The standardisation process involved calculating the z-scores by subtracting the mean and dividing by the standard deviation of each data point. By standardising the data, the effects of signal intensity were minimised, allowing for more precise comparisons and analyses. These preprocessing steps played a crucial role in mitigating the spectral data insufficiencies, ensuring accurate and meaningful interpretation of the results.

5.3.2.6. Predictive modelling of SSC, puncture force and pigments

The SL-AI algorithm was used for the prediction of the following biophysical characteristics: SSC (%), PF (N) chlorophyll (a.u.) and anthocyanin (a.u.) in both original and reconstructed datasets. The SL-AI establishes the connection between spectral features and the biophysical variables studied by exploring covariance mode (CovM). The use of the SL-AI is justified based on its ability to handle datasets with complex spatiotemporal variability and its good performance in predicting grape biophysical proprieties, as demonstrated in previous works (Tosin et al., 2022; Martins et al., 2022). The following statistical metrics: coefficient of determination (R^2), root means square error (RMSE) and mean absolute percentage error (MAPE - %) assessed the SL-AI robustness in predicting SSC, PF, chlorophyll and anthocyanin content.

5.3.2.7. Data characterisation

Among the 12 DAVs, three specific DAVs representing distinct stages of the maturation process, namely the beginning (DAV 05), middle (DAV 24), and end (DAV 47) stages, were selected to assess the differences between the total grape (TO), SK, PU and SE. These three DAVs were considered in the following analysis: i) characterisation of both grape cultivars regarding the biophysical parameters measured, iii) spectral signature reconstruction, iv) analysis of the first two principal components of PCA to evaluate the maturation differences based on the spectral data of the total grape as well as its SK, PU, and SE components and iv) to evaluate the prediction of the biophysical parameters with SL-AI for each grape cultivar in the SK, PU and SE.

A one-way analysis of variance (ANOVA) with a p-value associated with the Fischer followed (when needed) by Duncan post hoc test ($\alpha = 5\%$) was performed to compare the volume, SSC, PF, chlorophyll and anthocyanin values among the 12 DAV, for both cultivars.

The comparison of the volume, SSC, PF, chlorophyll and anthocyanin for the measured and predicted values among the 12 DAV through SL-AI for each cultivar considered a t-test.

ANOVA and t-test considered the following levels of significance: i) non-significant (ns); ii) significant ($p < 0.05$, *); iii) highly significant ($p < 0.01$, **) and iv) very highly significant ($p < 0.001$, ***).

5.3.3. Results

5.3.3.1. Grape characterisation

Table 5.3-1 presents for the three DAVs (05, 24 and 47) the evolution throughout the maturation for both grape varieties regarding the grape volume, SSC and PF for the total grape and the spectra-derived pigments (anthocyanin and chlorophyll) in TO, SK, PU, and SE. A detailed comparison of the 12 dates following the veraison is presented in Table S 5.3-1 of the Supplementary Material.

Table 5.3-1. Comparison of volume (cm³), soluble solids content (SSC), puncture force (PF), anthocyanin (Ant), and chlorophyll (Chl) in total grape (TO), skin (SK), pulp (PU), and seed (SE) of Loureiro and Vinhão cultivars across three assessment dates from days after veraison.

Variables	Days after veraison			p-value
	05	24	47	
<i>Volume (cm³)</i>	***	ns	ns	
Loureiro	1.32 ^a ± 0.36	1.37 ^a ± 0.37	1.26 ^a ± 0.27	0.688
Vinhão	0.93 ^b ± 0.16	1.39 ^a ± 0.28	1.42 ^a ± 0.30	p < 0.001
<i>SSC (%)</i>	***	***	***	
Loureiro	16.31 ^b ± 0.76	17.06 ^a ± 1.15	16.20 ^b ± 1.10	0.030
Vinhão	9.23 ^c ± 1.29	11.85 ^b ± 3.41	17.24 ^a ± 1.65	p < 0.001
<i>PF (N)</i>	***	***	ns	
Loureiro	2.97 ^c ± 0.33	4.47 ^a ± 1.01	3.76 ^b ± 0.53	p < 0.001
Vinhão	7.08 ^a ± 2.62	7.71 ^a ± 3.15	4.16 ^b ± 0.58	p < 0.001
<i>Chl TO (a.u.)</i>	***	***	***	
Loureiro	7.59 ^a ± 0.51	6.18 ^b ± 0.55	4.88 ^c ± 0.39	p < 0.001
Vinhão	2.45 ^a ± 1.87	1.81 ^a ± 1.19	2.32 ^a ± 0.11	0.268
<i>Chl SK (a.u.)</i>	***	***	***	
Loureiro	22.28 ^a ± 0.79	22.41 ^a ± 0.76	19.87 ^b ± 0.94	p < 0.001
Vinhão	1.76 ^a ± 1.72	1.32 ^a ± 1.23	1.75 ^a ± 0.54	0.384
<i>Chl PU (a.u.)</i>	***	***	***	
Loureiro	20.29 ^a ± 0.59	19.56 ^b ± 0.66	18.19 ^c ± 0.82	p < 0.001
Vinhão	3.50 ^a ± 0.14	3.22 ^b ± 0.37	2.18 ^c ± 0.54	p < 0.001
<i>Chl SE (a.u.)</i>	**	*	ns	
Loureiro	5.63 ^a ± 1.19	4.93 ^b ± 0.75	5.13 ^{ab} ± 0.85	0.079
Vinhão	4.53 ^a ± 0.73	4.26 ^a ± 0.91	4.77 ^a ± 0.95	0.163
<i>Ant (a.u.)</i>				
Ant TO (a.u.)	4.25 ^a ± 0.71	3.37 ^b ± 0.34	2.61 ^c ± 0.44	p < 0.001
Ant SK (a.u.)	1.51 ^c ± 0.14	1.84 ^b ± 0.52	3.31 ^a ± 0.62	p < 0.001
Ant PU (a.u.)	11.77 ^a ± 0.29	12.00 ^a ± 0.75	10.93 ^b ± 0.62	p < 0.001
Ant SE (a.u.)	9.56 ^a ± 0.89	9.68 ^a ± 1.25	8.54 ^b ± 1.05	0.009

In the same line (among dates), means labelled with different letters were significantly different according to Duncan's post hoc test ($\alpha = 5\%$). The comparison between cultivars for the same variables and same days after veraison considered the following levels of significance: non-significant (ns), significant ($p < 0.05$, *), highly Significant ($p < 0.01$, **), and very highly significant ($p < 0.001$, ***).

Grape SE from both Loureiro and Vinhão cultivars were quantified from DAV 45. This analysis revealed an average of two SE per grape, with the mode also consistently at two, indicating that SE count was steady and uniform for both grape cultivars, underscoring the stability of SE maturation processes across these cultivars.

Results showed significant DAV and cultivar differences for SSC and chlorophyll in PU, implying varied maturation lags. Volume comparison of both grape cultivars indicated significance only at DAV 05, suggesting Vinhão enlarges faster but maintains size like Loureiro. This Vinhão delay implies Loureiro's quicker growth. PF varied across DAVs for both cultivars but not at DAV 47. This links with grape sugar accumulation; Loureiro had higher SSC at the maturation start (DAV 05).

Chlorophyll in TO and SK differed only for Loureiro, with consistent differences across dates. DAV differences were Loureiro-specific, not Vinhão. Comparing cultivars, SE differed in DAV 05 and 24 but not 47.

Anthocyanin increased significantly in TO, SK, PU, and SE across DAVs, typical for red grapes such as Vinhão.

The Loureiro variety demonstrates higher SSC concentrations than Vinhão (Figure 5.3-4). However, as shown in Figure 5.3-4, Loureiro is more susceptible to the effects of rainfall on the SSC levels, likely attributed to a more developed protective of the protective wax layer on the red grape cultivar (Rustioni et al., 2016).

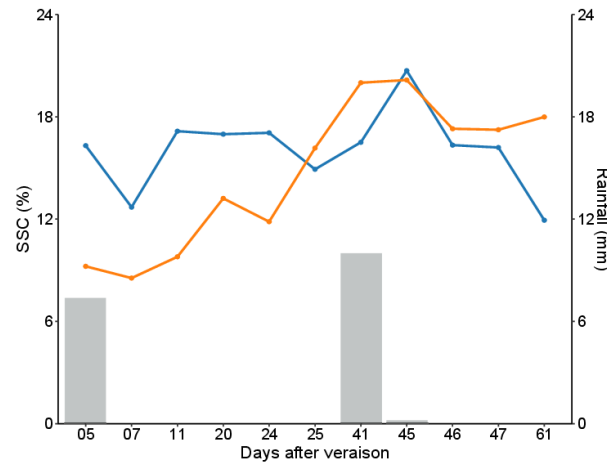


Figure 5.3-4. Comparative analysis of soluble solids content (SSC - %) levels between Loureiro (blue line) and Vinhão (orange line) cultivars, based on the total rainfall (grey bars, mm) on the sampling dates.

5.3.3.2. Bi-directional spectral reconstruction

The algorithms based on MHPCA resulted in a bi-directional reconstruction of the TO and SK, PU and SE spectra. Table 5.3-2 presents the statistical metrics for the spectral reconstruction of TO and its internal tissues (SK, PU, and SE) for the Loureiro and Vinhão cultivars across the 12 assessment dates. Vinhão cultivar generally presented higher values of MSE (≥ 0.09) and MAPE ($> 11\%$) when compared to Loureiro (MSE ≤ 0.02 , MAPE $< 7\%$), except Loureiro SE, which presented higher MSE (> 0.20) and higher MAPE ($> 40\%$). The high Pearson's correlation coefficient ($r \geq 0.90$) and non-significant p-values ($p < 0.001$) also support the robustness of the spectral reconstruction.

Table 5.3-2. Statistics for the spectral reconstruction compared with the original spectra of the total grape and the skin, pulp and seed spectra for the Loureiro and Vinhão cultivar in the 12 assessment dates.

Statistics	Total	Skin	Pulp	Seed
<i>Loureiro (n = obs)</i>				
MSE (count/wavelength (nm))	0.02	0.00	0.01	0.23
MAPE (%)	6.08	1.78	0.23	42.20
r	0.99	1	1	0.94
p-value	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
<i>Vinhão (n = obs)</i>				
MSE (count/wavelength (nm))	0.16	0.15	0.09	0.09
MAPE (%)	33.32	23.25	11.15	14.67
r	0.92	0.92	0.95	0.95
p-value	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), the p-value < 0.001 , obtained through the Person's correlation coefficient (r), indicates that there is no significant difference between the original and reconstructed spectral matrices.

A comparison of the metrics reveals that the MHPCA algorithm yielded better results for the Loureiro cultivar than Vinhão. This difference can be attributed to the enhanced light

transmittance observed in the Loureiro grapes' PU and SK, facilitating signal acquisition by the fibre probe (Figure 5.3-1). The low precision observed in SE reconstruction for both cultivars can be attributed to the small size of this tissue and the inherent challenges associated with identifying a location that yields a strong signal intensity.

Consequently, the measurement of SE necessitated a longer integration time to improve the results' accuracy. This extended integration time increased the probe's sensitivity, making the spectral data more vulnerable to noise caused by even the slightest hand movements.

Despite the difficulties in accurately measuring specific tissues, such as SE, and the challenges posed by dye cultivars like Vinhão, the overall outcomes of the spectral reconstruction were considered satisfactory.

Figure 5.3-5 represents the mean original spectral of the TO, SK, PU and SE compared with the reconstructed spectra of the total grape and internal tissues for both grape cultivars at the three DAVs (05, 24,47). The reconstructed spectra for both grape cultivars exhibit similarity to the original spectra, consistent with the statistical parameters presented in Table 5.3-2, which considers all assessment dates.

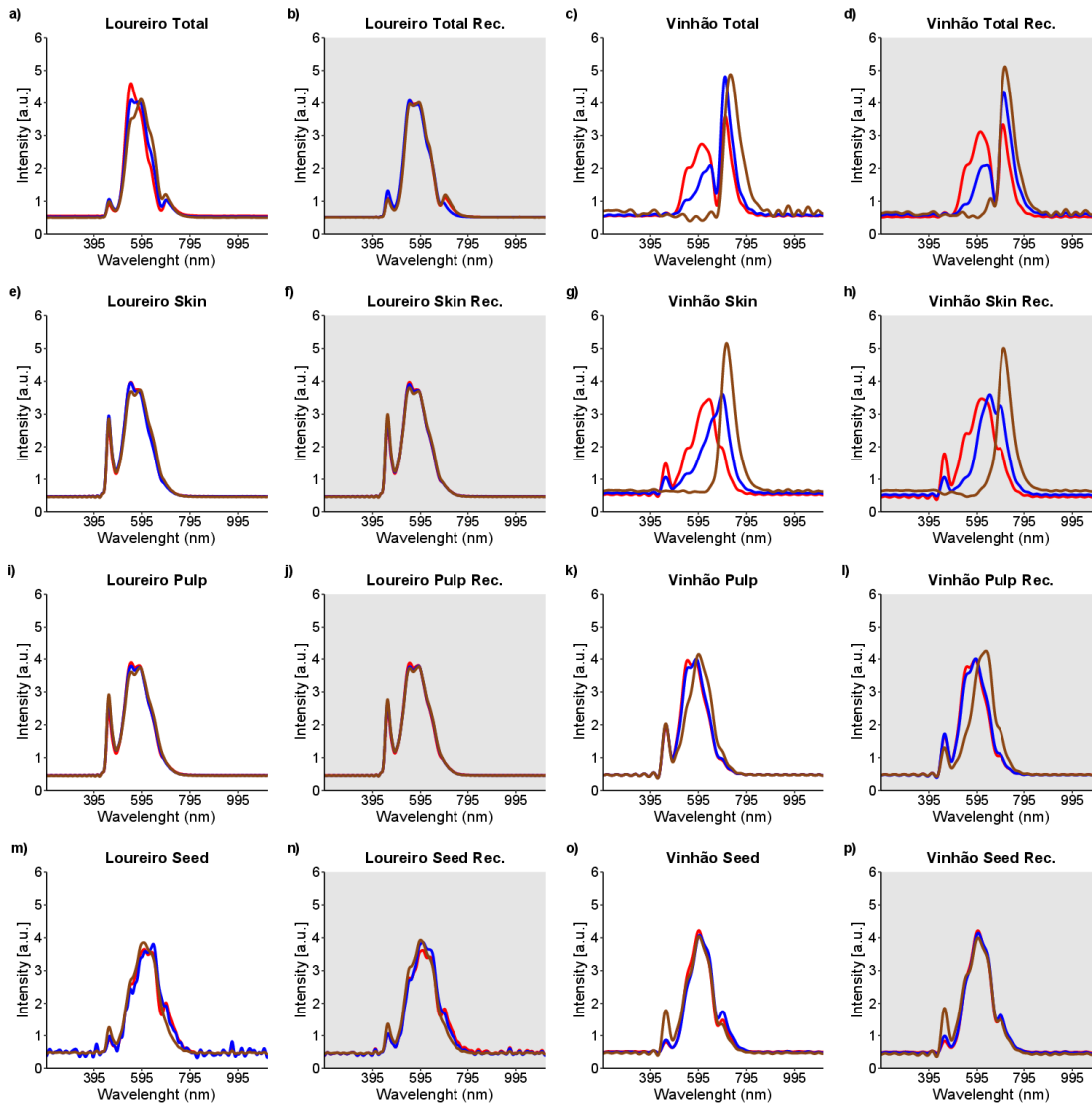


Figure 5.3-5. Comparison of original and reconstructed (Rec.) (grey background) spectral signatures for grape cultivars (Loureiro e Vinhão) for total grape and its tissues at different stages of maturation: 5 (red line), 24 (blue line), and 47 (brown line) days after veraison (DAV). For Loureiro, it includes the original (a, e, i, m) and reconstructed (b, f, j, n) spectral signatures of the total grape, skin, pulp, and seed, respectively. Likewise, Vinhão provides the same grape tissues' original (c, g, k, o) and reconstructed (d, h, l, p) spectral signatures.

The similarity observed between the original and reconstructed spectra renders the reconstructed spectra suitable for analysing grape quality parameters, including the biophysical variables in different tissues.

During maturation, spectral differences emerge between the Loureiro and Vinhão grape cultivars, mainly in the visible spectrum. These differences can be influenced by factors like pigment concentration, SSC and texture inferred by PF, as demonstrated in Table 5.3-1. In the case of the Loureiro cultivar, spectral signature differences are challenging to discern in TO, SK, PU and SE, as the maturation process of white grapes changes from green to yellow-green. Conversely, the Vinhão cultivar, in which the grape's colour

changes from green to purple, exhibits more prominent spectral changes in TO, SK, PU and SE due to its increased anthocyanin content and SSC.

5.3.3.3. Spectral patterns of grapes composition during maturation

Figure 5.3-6 demonstrates the application of PCA to analyse the spectral data of grapes involving their internal tissues (SK, PU and SE). Each data point corresponds to a unique spectral measurement from an individual grape, enabling PCA to identify a lower-dimensional representation that captures essential information.

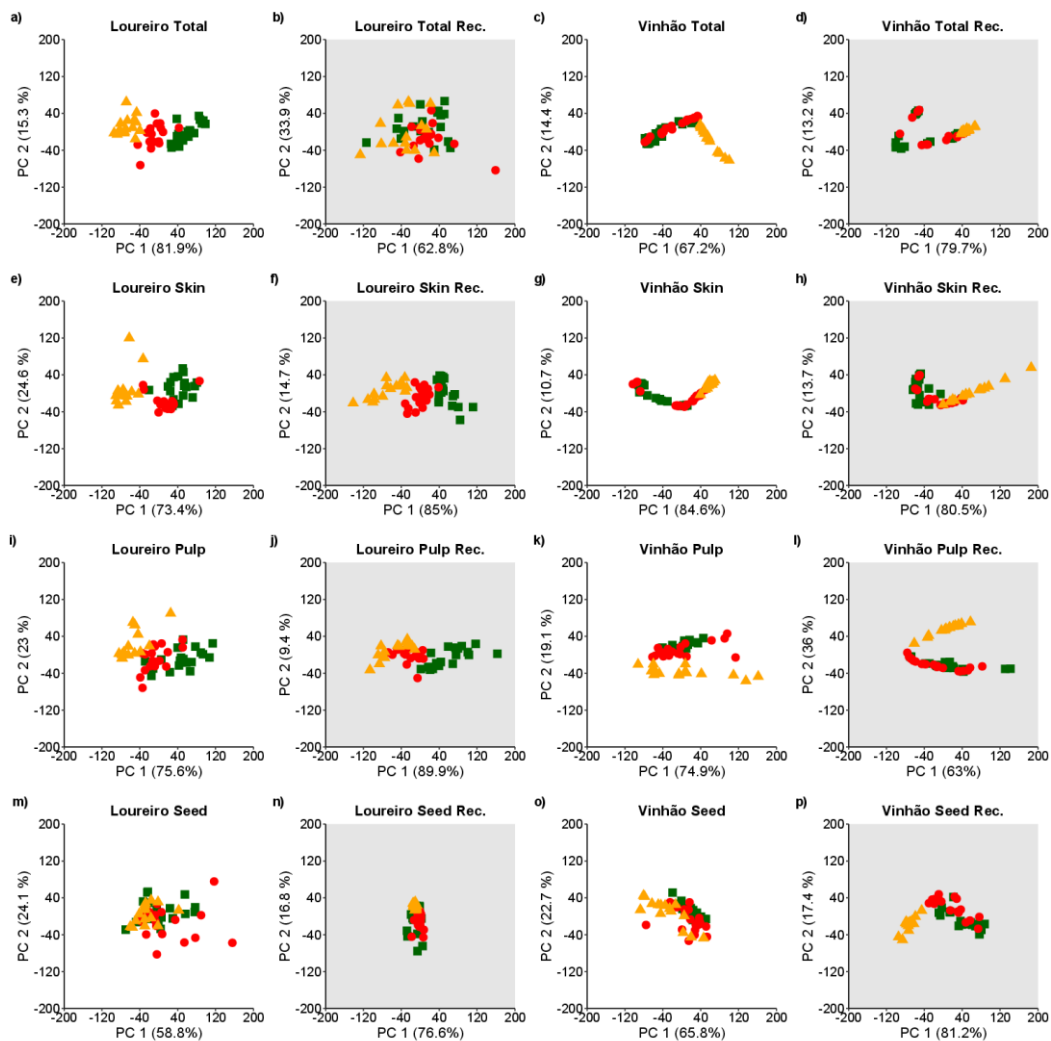


Figure 5.3-6. Principal component analysis (PCA) illustrating variations of original and reconstructed (Rec.) (light grey background pattern) spectral data for Loureiro and Vinhão cultivars over time (5 (green square), 24 (red circle), and 47 (yellow triangle) days after veraison) and differentiation of grape tissues. For Loureiro, it includes the original (a, e, i, m) and reconstructed (b, f, j, n) spectral signatures of the total grape, skin, pulp, and seed, respectively. Likewise, Vinhão provides the same grape tissues' original (c, g, k, o) and reconstructed (d, h, l, p) spectral signatures.

Figures 5a and c demonstrate the overall space occupied by the grapes of Loureiro and Vinhão, respectively. Furthermore, Figures 5e, 5g, 5i, and 5k display grape cultivars'

distinct SK, PU and SE feature spaces. PCA effectively decomposes the space, creating separate spaces for each grape tissue.

Figures 5b and d present the reconstructed spectral data obtained by combining information from the SK, PU and SE. This reconstructed data provides a condensed version of the original spectral data, retaining crucial characteristics. The process can be reversed to decompose grape data into individual feature spaces for the grape's SK, PU and SE, as illustrated in Figures 5f, 5h, 5j, and 5l, respectively.

PCA's application enhances understanding of the grape's spectral data and internal structures. It reveals relationships between different grape fractions and their spectral properties, shedding light on each tissue's distinct characteristics within the grape.

In Figure 5.3-6, PC1 and PC2 represent over 90% of the data for both cultivars' original and reconstructed TO, SK and PU spectra. This suggests that these two principal components capture the relevant spectral information for these specific grape varieties. When considering SE and the entire Vinhão grape cultivar, the combination of PC1 and PC2 falls below the 90% threshold. The decrease in accuracy observed can be attributed to several challenges encountered during spectral acquisition from the tissues under investigation. Firstly, the samples' small size and darker nature (SE) presented difficulties obtaining consistent and reliable spectral data. Secondly, obtaining signal intensity from these samples posed additional challenges, affecting the quality of the acquired spectral information. As a result of these combined factors, the precision of the results was negatively impacted in this specific context.

When analysing the original spectra, differences in the maturation process of the Loureiro and Vinhão grape cultivars are identified. DAV 05, 24, and 47 variations in the Loureiro cultivar are observed in TO and SK spectra. However, the separation of seed spectra is less distinct, likely due to challenges associated with accurately capturing SE spectra. Consequently, the reconstructed TO spectra of the Loureiro cultivar lack clear differentiation among the assessed dates. On the other hand, the Vinhão cultivar exhibits a distinct difference at 47 DAV in TO and PU spectra.

Likewise, SK and SE spectra of the Vinhão cultivar exhibit differences across all assessed dates. This dynamic pattern of separation observed through PCA suggests a delayed maturation among the tissues in the Vinhão cultivar, where the SK develops rapidly. Still, PU may not mature within the same temporal period. Interestingly, differences in SE spectra are observed across the three dates for the Vinhão cultivar, providing hints into seed maturation processes.

5.3.3.4. Modelling grape SSC, puncture force and pigments using spectral data

SL-AI assessed the efficacy of reconstructed spectral information in predicting the biophysical parameters analysed in this paper in the different grape tissues. The predictions of these parameters were based on the original and reconstructed spectral data, considering the 12 assessment dates. The analysis demonstrated that the original and reconstructed data provided reliable and consistent predictions for the examined biophysical parameters (Table 5.3-3). Detailed results and regression plots for estimating the biophysical parameters of both cultivars are provided in the Supplementary Materials (Figures S 5.3-2 to 8).

Table 5.3-3. Anthocyanin (Ant), chlorophyll (Chl), SSC (%) and puncture force (PF -N) were predicted with the self-learning artificial intelligence (SL-AI) algorithm for the original and reconstructed dataset for the grape tissues of Loureiro and Vinhão cultivar for 12 assessment dates.

	Original				Reconstructed			
	R ²	RMSE	MAPE (%)	Nobs.	R ²	RMSE	MAPE (%)	Nobs.
Loureiro								
<i>Chl (a.u.)</i>								
Total	-	-	-	-	0.98	0.88	11.41	210
Skin	-	-	-	-	1	0.73	2.69	211
Pulp	-	-	-	-	1	0.82	3.29	212
Seed	-	-	-	-	0.97	0.93	13.8	211
<i>SSC (%)</i>								
Total	0.99	1.9	9.55	216	0.97	2.74	12.89	216
<i>PF (N)</i>								
Total	0.92	1.28	26.34	215	0.88	1.76	25.13	209
Vinhão								
<i>Ant (a.u.)</i>								
Total	-	-	-	-	0.93	0.87	16.94	205
Skin	-	-	-	-	0.95	0.58	14.11	205
Pulp	-	-	-	-	1	0.72	4.5	205
Seed	-	-	-	-	0.99	0.89	7.73	204
<i>Chl (a.u.)</i>								
Total	-	-	-	-	0.94	0.59	15.24	205
Skin	-	-	-	-	0.91	0.54	28.78	205
Pulp	-	-	-	-	0.95	0.61	23.24	205
Seed	-	-	-	-	0.96	0.89	15.65	205
<i>SSC (%)</i>								
Total	0.99	1.75	8.99	202	0.99	1.87	9.57	202
<i>PF (N)</i>								
Total	0.93	1.49	18.79	202	0.91	1.69	20.51	202

Mean squared error (MSE); mean absolute percentage error (MAPE), number of observations (Nobs.)

By comparing the original and reconstructed spectra, the Loureiro cultivar exhibited higher statistical performance for the pigment's quantification than the Vinhão cultivar. For chlorophyll, Loureiro obtained $R^2 \geq 0.97$ and $MAPE < 14\%$, and Vinhão reached R^2

≥ 0.91 and $\text{MAPE} < 29\%$. Anthocyanin in Vinhão shows $R^2 \geq 0.93$ and $\text{MAPE} < 17\%$. On the other hand, Vinhão presented better statistical metrics for SSC and PF when compared to Loureiro. Regarding PF, Loureiro scored $R^2 \geq 0.88$ and $\text{MAPE} < 26\%$, and Vinhão achieved $R^2 \geq 0.91$ and $\text{MAPE} < 21\%$. For SSC, Loureiro achieved $R^2 \geq 0.97$ and $\text{MAPE} < 13\%$, while Vinhão got $R^2 \geq 0.99$ and $\text{MAPE} < 10\%$.

The application of SL-AI considered a single latent variable (LV) for all parameters listed in Table 5.3-3. However, due to the structure of SL-AI with covariance mode (CovM), specific parameters could not cover the entire dataset leading to the exclusion of observation in the estimation process: i) Loureiro - chlorophyll (6 observations) and PF estimations excluded a maximum of seven samples, respectively ii) Vinhão - the anthocyanin estimation did not consider one observation, and the SSC and PF estimations excluded up to three samples in the reconstructed dataset, respectively.

5.3.3.5. Biophysical parameters along the maturation

Figure 5.3-7 presents the biophysical parameters studied in this paper estimated with SL-AI using the TO, SK, PU, and SE reconstructed dataset for DAVs 5, 24, and 47 of the Vinhão cultivar. The values in Figure 5.3-7 are also compared with the observed values in Table 5.3-1. For most parameters assessed in the Vinhão cultivar, non-significant values were observed compared to the observed values. However, a significance ($p < 0.05$) difference (1.74 vs 2.18) was found for chlorophyll in PU at the 47 DAV. This difference may be related to the reconstruction of PU for this date. As this work modelled chlorophyll and anthocyanin using spectral data, the errors associated with the reconstruction (Table 5.3-2) may affect the estimation of chlorophyll in PU for DAV 47. These findings reinforce the success of spectral reconstructions (Table 5.3-2) and demonstrate the capability of SL-AI in predicting the values (Table 5.3-3) for each biophysical parameter considered in this study.

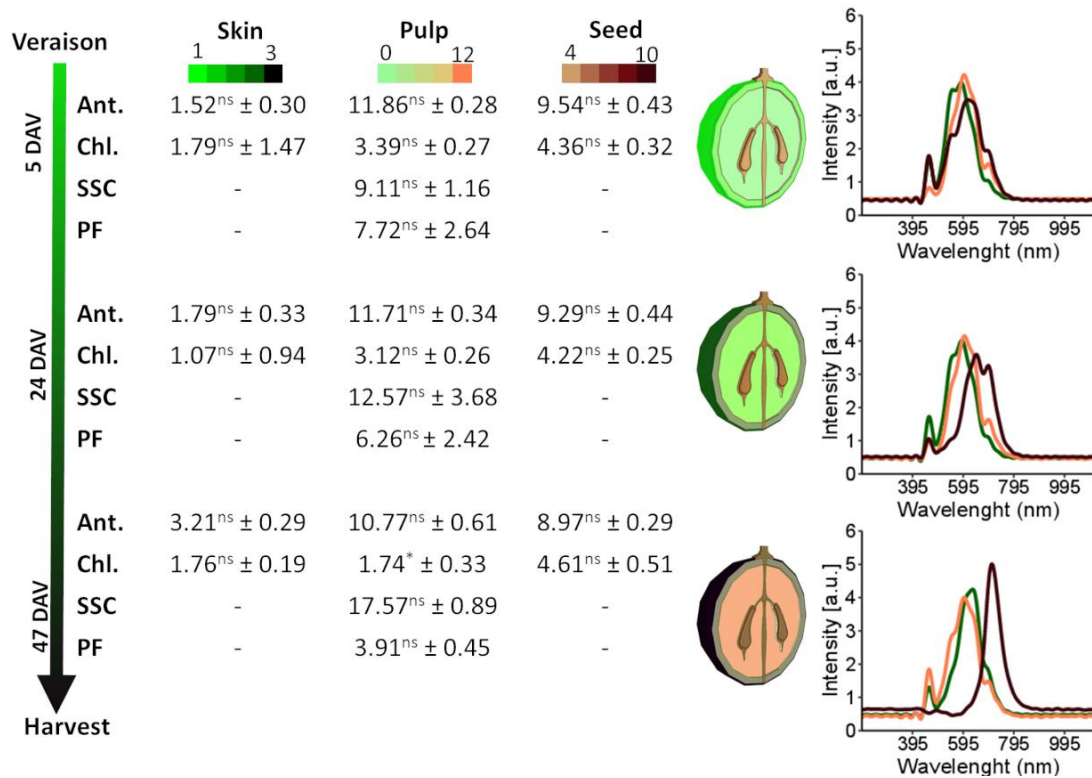


Figure 5.3-7. Temporal progression of maturation in the Vinhão cultivar, presented through the reconstructed spectral dataset of the total grape, skin, pulp and seed at 5, 24, and 47 days after veraison (DAV). Anthocyanin (Ant.) and chlorophyll (Chl.) in the skin, pulp, and seed (expressed in arbitrary units, a.u.), and predicted values of soluble solids content (SSC - %) and puncture force (PF - N) for the total grape. Reconstructed spectra of the skin (brown line), pulp (green line), and seed (orange line) for the respective DAVs. A t-test compared the observed (Table 5.3-1) with the predicted values of Ant., Chl., SSC, and PF using the reconstructed datasets through self-learning artificial intelligence (SL-AI). * indicates statistical significance at $p < 0.05$, and "ns" denotes non-significance.

Table 5.3-4, likewise Figure 5.3-7, compares the observed biophysical parameters with the estimated values using the reconstructed dataset through SL-AI for DAVs 05, 24, and 47 for the Loureiro cultivar. Significant differences were found in the three DAVs assessed. For DAV 05, PF presented significant differences (3.76 vs 2.97). For DAV 24, chlorophyll in PU (19.18 vs 19.56) and PF (3.91 vs 4.47), and for DAV 47, chlorophyll in SK (20.44 vs 19.87) and PU (18.82 vs 18.19) presented significant differences.

Table 5.3-4. The statistical description of the reconstructed spectral dataset of the total grape, skin, pulp and seed using self-learning artificial intelligence (SL-AI) in predicting chlorophyll (Chl – a.u.) for the skin, pulp and seed of Loureiro cultivar and the predicted values for the soluble solid content (SSC - %) and puncture force (PF – N) for the total grape in the 05, 24 and 47 days after veraison (DAV).

DAV		Skin	Pulp	Seed
05	Chl (a.u.)	22.19 ^{ns} ± 0.11	20.09 ^{ns} ± 0.35	5.32 ^{ns} ± 0.55
	SSC (%)	-	16.02 ^{ns} ± 1.14	-
	PF (N)	-	3.76* ± 0.66	-
24	Chl (a.u.)	22.05 ^{ns} ± 0.25	19.18* ± 0.31	4.97 ^{ns} ± 0.32
	SSC (%)	-	16.63 ^{ns} ± 0.73	-
	P. Force (N)	-	3.91* ± 0.31	-
47	Chl (a.u.)	20.44* ± 0.36	18.82* ± 0.22	5.34 ^{ns} ± 0.14
	SSC (%)	-	16.74 ^{ns} ± 1.18	-
	P. Force (N)	-	4.53 ^{ns} ± 1.45	-

A t-test compared the observed (Table 5.3-1) and reconstructed datasets through SL-AI to compare the predicted Chl, SSC, and PF values. * indicates statistical significance at $p < 0.05$, and “ns” denotes non-significance differences.

As this work used spectral data to infer chlorophyll in the grapes, the errors associated with the reconstruction may have affected the estimation of chlorophyll, mainly in the PU of Loureiro. Although Table 5.3-2 indicates robust metrics for the spectral reconstruction of the SK and PU ($r = 1$), and SL-AI demonstrated strong results ($\text{MAPE} < 4\%$), the error estimation may be concentrated more in this specific group of samples considered.

Despite the favourable metrics of spectral reconstruction for the Loureiro TO ($r = 0.99$), SL-AI faced challenges predicting PF, resulting in a mean absolute percentage error (MAPE) $< 26\%$. The errors associated with the spectra reconstruction and the lower performance in modelling PF for Loureiro corroborate by the significant differences found.

5.3.4. Discussion

This paper proposes a non-destructive tomography-like for analysing internal tissue spectra through point-of-measurement (POM) sensing. By applying the algorithm based on multi-block hierarchical principal component analysis (MHPCA), it is possible to establish a bi-directional spectral reconstruction considering the total grape (TO) and the skin (SK), pulp (PU) and seed (SE). The reconstructed spectra were subsequently used to estimate the biophysical properties of grape tissues through SL-AI.

The tomography-like presented in this study offers potential support for viticultural management practices and decisions based on the assessment of the internal tissues. Unlike previous studies that utilised X-ray (Liu et al., 2020a; Kanno & Kuroyama, 2020) and ultrasound-based methods (Li et al., 2019; Goławski et al., 2017) to assess fruits at a cellular level, the tomography-like approach in this study examines grape berries using Vis-NIR information, providing hints into their physiological characteristics. The

Vis-NIR non-destructive information obtained from different grape tissues facilitates the modelling of essential parameters such as sugar content, acidity levels, and phenolic compounds, which are crucial for advanced monitoring of grape maturation (Fernandez-Navales et al., 2019a; Giovenzana et al., 2015).

This study, as a proof of concept and low TRL, demonstrated that through the total spectra of the grape, it is possible to predict chlorophyll and anthocyanin in the SK, PU and SE. The results of this paper are an improvement when compared to other research studies that empirically demonstrated that it is possible to predict biophysical parameters in the internal tissues (Martins et al., 2022) and employed spectral information to predict pigments (Tosin et al., 2021).

The bi-directional spectral reconstruction using MHPCA combined with PLS yielded a strong correlation ($r \geq 0.92$; $p < 0.001$) between the original and reconstructed spectra (Table 5.3-2). Alternative methods, such as deep learning (Mishra et al., 2022), O2-PLS (Trygg & Wold, 2003), and On-PLS (Lofstedt et al., 2013b) may present a possible solution for spectral reconstruction of TO using the spectral information from the SK, PU, and SE. However, these methods pose complexities due to the need for extensive data, limited interpretability in deep learning, and challenges in data preprocessing and model selection. Moreover, over-fitting impacts the generalisation of results to new data, and the availability of suitable software and tools might be limited, adding complexity. Further research is needed to refine these methods, improve interpretability, and adapt them to grape spectra characteristics.

POM devices present a promising non-destructive approach for grape maturation assessment. These instruments offer detailed spectral information from the grape tissues, including SK, PU and SE (Martins et al., 2022). The tomography-like approach can enable timely decision-making during grape maturation, optimising harvest timing and grape quality through on-site, accurate measurements for precise management of grape clusters, including irrigation, nutrient, and canopy management to improve resource utilisation. Moreover, the efficiency of POM devices in acquiring data without incurring a substantial data load makes them particularly suitable for large-scale measurements (Nicolai et al., 2007), facilitating more extensive implementation and enhanced grape production management practices.

The study validates the spectral bi-directional reconstruction approach by closely resembling the original and reconstructed spectral signatures (Figure 5.3-5). Applying PCA analysis (Figure 5.3-6) supports the similarity between the two sets of spectra, demonstrating comparable geometric properties in space for the two components. Although the spectra data contains information on several biophysical parameters, as

shown in this work, appropriate analysis techniques (e.g., PCA) provides a better comprehension of the dynamic of the maturation. For example, the decrease in chlorophyll content in PU for both cultivars is due to PU being related to the sugar accumulation and contributes to the aromatic and flavour profiles of the grapes (Gouot et al., 2019; Agati et al., 2008). This evidence is shown in the different geometric spaces of PCA in PU when other dates have separate spaces for this tissue (Figure 5.3-6). Contrarily, the changes in SE related to size, structure, and colour (Yang et al., 2011; Kennedy et al., 2000) were unclear in our study due to the geometrical data points for the three DAVs being very close and presenting less than 90% of representability of the data when considered PC1 and PC2 (Figure 5.3-6).

The literature encompasses various modelling techniques using Vis-NIR for predicting biophysical variables. Among these techniques, PLS is widely employed (Fernandez-Novales et al., 2019a). However, PLS models exhibit limitations in their static nature, necessitating retraining if substantial dataset changes occur (Ergon, 2009). In contrast, as Martins et al. (2022) argued, SL-AI offers distinct advantages, such as finding covariance mode (CovM) of non-dominant interference constituents and reducing prediction errors. As shown in Table 5.3-3, SL-AI demonstrates the potential of utilising spectral information to biophysical parameters in the original and reconstructed dataset for the total grape and internal tissues, producing overall MAPE < 29 % and $R^2 \geq 0.88$. Figure 5.3-7 and Table 5.3-3 also support the use of SL-AI in predicting parameters for different grape cultivars. The robust results provided in this paper support the use of SL-AI in modelling biophysical parameters.

Although the reconstruction process yielded robust outcomes (Table 5.3-2), limitations were encountered during the reconstruction of the total grape for both cultivars. The considerable variations in grape size, especially seeds, posed challenges in consistently positioning them within the tomography system. Consequently, the placement of the fibre probe relied on identifying an optimal location that facilitated maximum LED light transmission. The variability in positioning may have influenced the signal intensity obtained, potentially resulting in the omission of spectral information from the seeds. This lack of precision interfered with reconstructing the total grape and internal tissues (Table 5.3-2).

As this technology is currently at a low technology readiness level (TRL), further research is necessary to address these limitations and refine the reconstruction techniques for more precise and reliable results. First, future studies should focus on optimising the positioning of the grapes within the tomography system, improving light transmission through dyed grapes, especially for SE spectra due to the lack of a technology that

measure consistently the entire SE spectra, and exploring advanced algorithms or machine learning approaches to enhance spectral reconstruction accuracy and parameter prediction. Second, as this work demonstrated, it is possible to infer the pigment composition in SK, PU and SE through the TO. A wet lab method, like HPLC (e.g., Farhadi et al., 2016), should be adapted for the small aliquots of the tissues to validate qualitatively the pigment content in the inner tissues of the grapes.

The tomography-like system has the potential for high-throughput spatiotemporal sampling, allowing rapid and accurate measurements of grape maturation at various locations within the vineyard. This capability can transform vineyard management practices by providing real-time information on grape quality parameters. The non-destructive method can be integrated into existing grape sorting and grading systems, offering the grape industry an efficient and accurate way to monitor grape maturation and ensure harvest quality. The system can also be applied to other grape cultivars and support metabolomics studies. Implementing the tomography-like system into a robotic platform may present challenges related to POM setup and light transmittance due to the challenge in developing a system that allows the obtention of the transmittance information in-situ of the target area. Still, successful implementation holds great promise for advancing grape cultivation and management techniques. Still, successful implementation holds great promise for advancing grape cultivation and management techniques.

5.3.5. Conclusion

The tomography-like system presented in this paper has demonstrated its effectiveness in establishing bi-directional spectral reconstruction for the TO, encompassing the SK, PU, and SE proving the TO spectra contain spectra information of the inner tissues.

Despite the efficiency of the reconstruction process, challenges were encountered in accurately capturing light dynamics passing through the internal tissues of the Vinhão cultivar due to the high concentration of red pigments that promoted lower statistical metrics compared to the Loureiro cultivar. Addressing these challenges is crucial for the continued advancement of this technology. In future studies, efforts like using a more powerful LED and developing huck support for the fibre probe can enhance the measurements' accuracy and improve the reconstruction output.

The tomography-like system can offer high-throughput capabilities for rapid and precise grape quality parameter measurements at multiple locations within vineyards, where viticulturists can access real-time information on grape maturation, facilitating data-driven decision-making and more sustainable viticultural practices.

CRediT authorship contribution statement

Renan Tosin: Investigation, Methodology, Writing- Original draft preparation, Writing- Reviewing and Editing. **Filipe Monteiro-Silva:** Methodology, Writing - Reviewing and Editing. **Rui Martins:** Supervision, Investigation, Methodology, Writing- Original draft preparation, Writing- Reviewing and Editing. **Mario Cunha:** Supervision, Investigation, Methodology, Writing - Original draft preparation, Writing – Reviewing and Editing.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Renan Tosin and Filipe Monteiro-Silva acknowledge Fundação para a Ciência e Tecnologia (FCT) PhD research grants Ref. SFRH/BD/145182/2019 and SFRD/BD/09136/2020. Rui Martins acknowledges Fundação para a Ciência e Tecnologia (FCT) research contract grant (CEEIND/017801/2018). This work is financed by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project OmicBots - OmicBots: High-Throughput Integrative Omic-Robots Platform for a Next Generation Physiology-based Precision Viticulture, with reference PTDC/ASP-HOR/1338/2021.

Supplementary material

Supplementary materials are provided to improve the comprehension of the research findings presented in the main paper. This document includes the algorithms developed for this paper and supporting table and figures, which provide additional context and detailed data that contribute to understanding the methodologies and results discussed in the main manuscript.

Algorithms

Require: X_1, X_2, X_3 Ensure: $j = \operatorname{argmin}(X_i - t_i p_i^t)$	$\begin{aligned} X_1 &= t_1 p_1^t \\ X_2 &= t_2 p_2^t \\ X_3 &= t_3 p_3^t \end{aligned}$
While $j > j_{\min}$ do	$\begin{aligned} \tau_i &\leftarrow [t_1^i t_2^i t_3^i] \\ \tau_i &= U_i S_i V_i^t \\ T_i &= U_i S_i \\ P_i^t &= V_i^t \end{aligned}$
end while Output: $T = [T_1 \dots T_n]; P = [P_1 \dots P_n]$	
Require: X_1, X_2, X_3 Ensure: $j = \operatorname{argmin}(X_i - t_i p_i^t)$	$\begin{aligned} X_1 &= t_1 p_1^t \\ X_2 &= t_2 p_2^t \\ X_3 &= t_3 p_3^t \end{aligned}$
While $j > j_{\min}$ do	$\begin{aligned} \tau_i &\leftarrow [t_1^i t_2^i t_3^i] \\ \tau_i &= U_i S_i V_i^t \\ T_i &= U_i S_i \\ P_i^t &= V_i^t \end{aligned}$
end while Output: $T = [T_1 \dots T_n]; P = [P_1 \dots P_n]$	

Algorithm S 5.3-1. Hierarchical latent structures algorithm

Require: Y, T

Ensure: $Y = UQ^t$

$U^k = Tb$

$T^k = Uc$

$$Y^k = (Tb)Q^t$$

Output: Y^k

Require: Y, T

Ensure: $Y = UQ^t$

$U^k = Tb$

$T^k = Uc$

$$Y^k = (Tb)Q^t$$

Output: Y^k

Algorithm S 5.3-3. Outer relationships for reconstruction

Require: $Y, T, P, [t_1, t_2, t_3]$

Ensure: $j = \operatorname{argmin}(X_i - t_i p_i^t)$

$$Y = UQ^t$$

$$T = Ub$$

$$[t_1, t_2, t_3] = T_p P^t$$

$$X_1 = t_1 p_1^t$$

$$X_2 = t_2 p_2^t$$

$$X_3 = t_3 p_3^t$$

Output: X_1, X_2, X_3

Require: $Y, T, P, [t_1, t_2, t_3]$

Ensure: $j = \operatorname{argmin}(X_i - t_i p_i^t)$

$$Y = UQ^t$$

$$T = Ub$$

$$[t_1, t_2, t_3] = T_p P^t$$

$$X_1 = t_1 p_1^t$$

$$X_2 = t_2 p_2^t$$

$$X_3 = t_3 p_3^t$$

Output: X_1, X_2, X_3

Algorithm S 5.3-2. Latent structures reconstruction algorithm

Figures

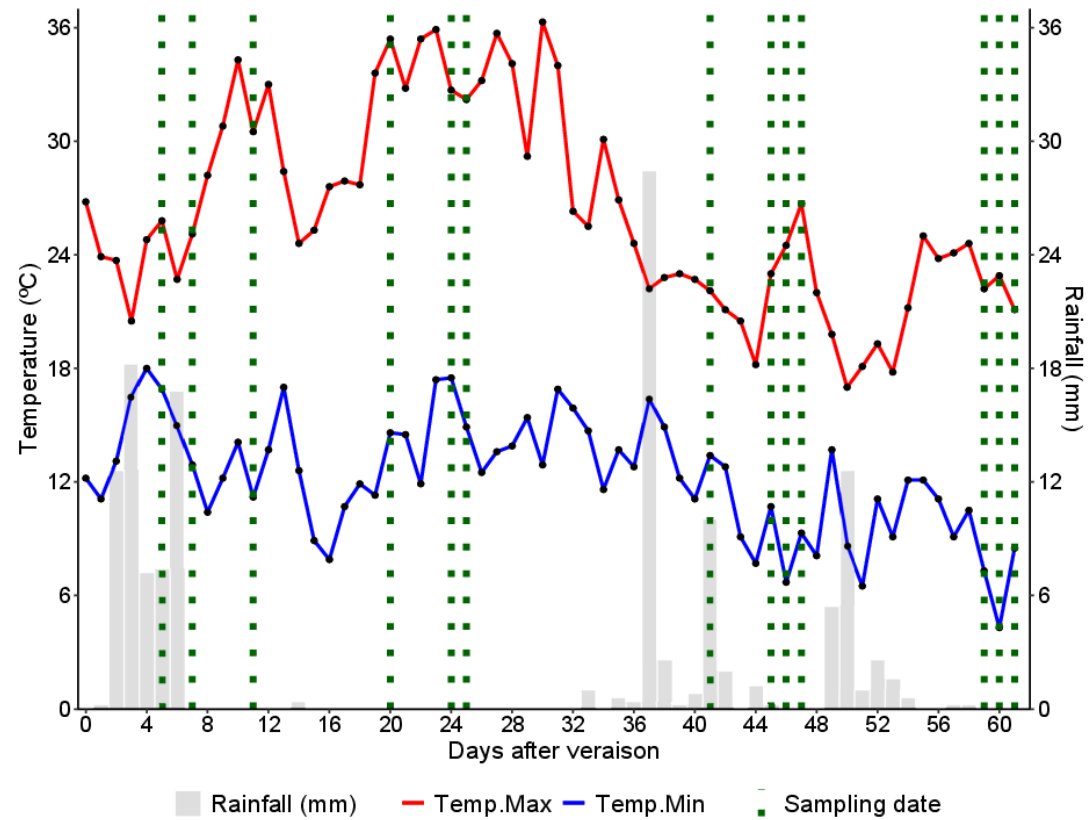


Figure S 5.3-1. Climatic data from the veraison (14/08/2020) until the last sampling date (14/10/2020). Green dotted point represents the sampling dates.

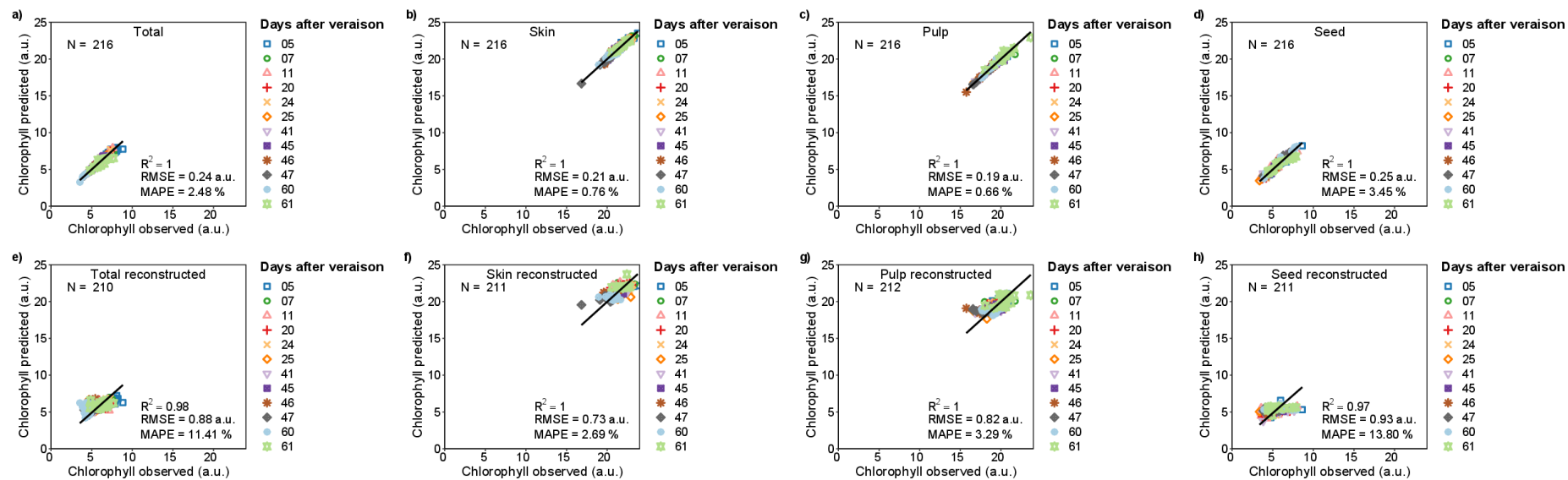


Figure S 5.3-2. Comparison of chlorophyll (a.u.) levels predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Loureiro grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a), skin (b), pulp (c) and seed (d). Reconstructed spectra for the total grape (e), skin (f), pulp (g) and seed (h).

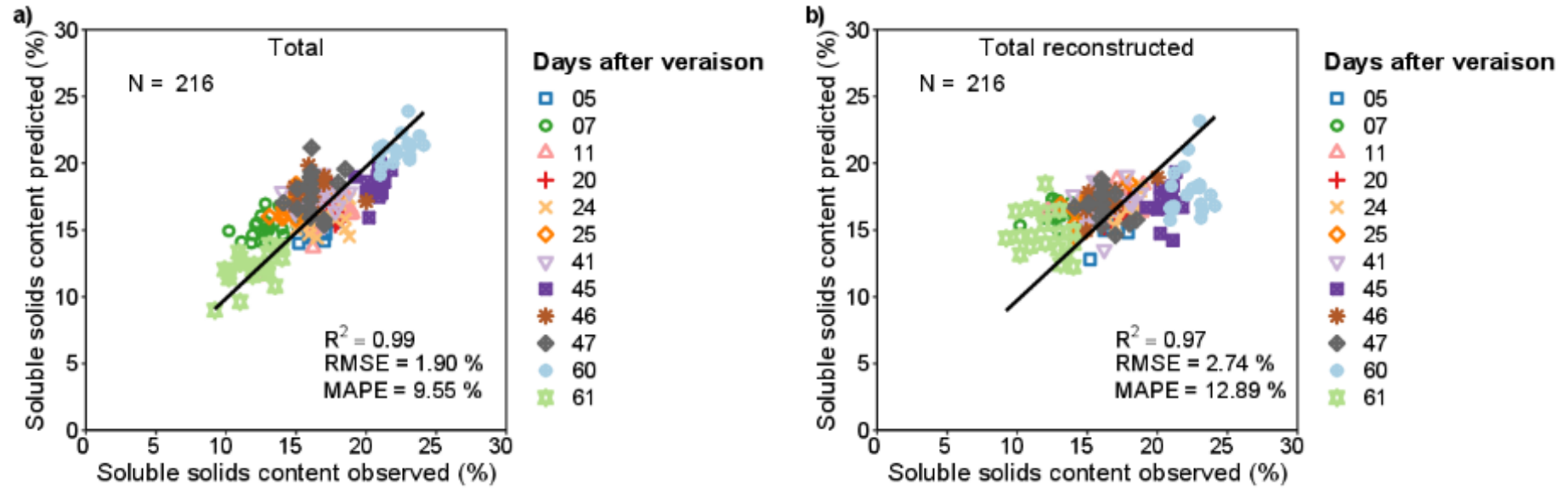


Figure S 5.3-3. Comparison of soluble solids content predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Loureiro grape. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).

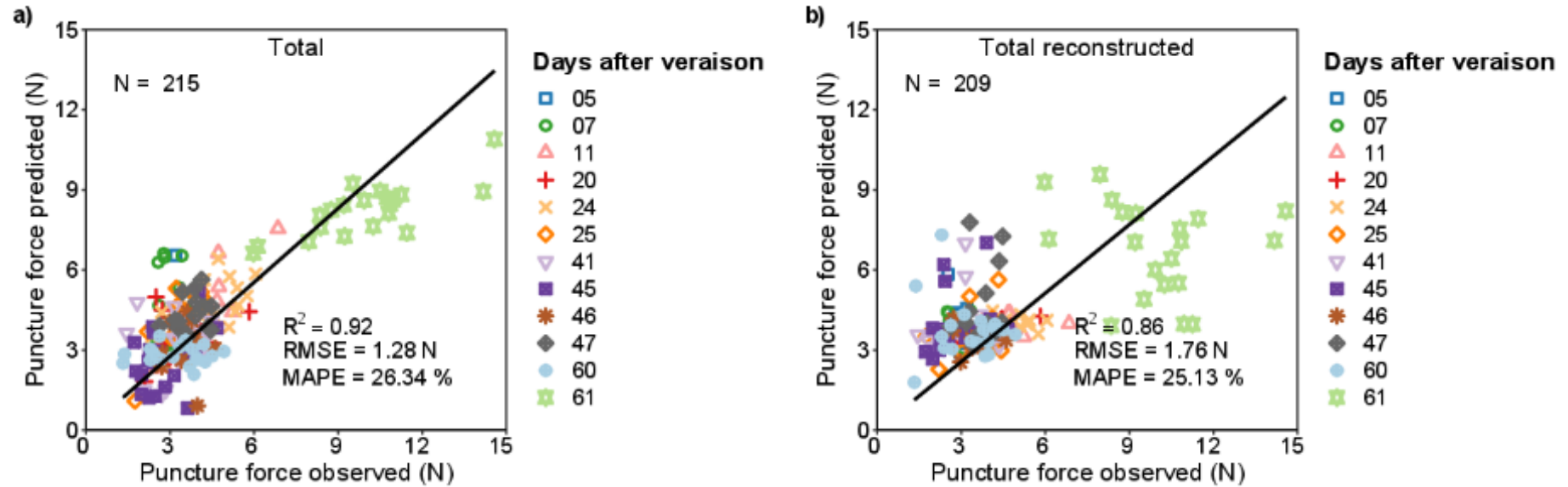


Figure S 5.3-4. Comparison of puncture force (N) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Loureiro grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).

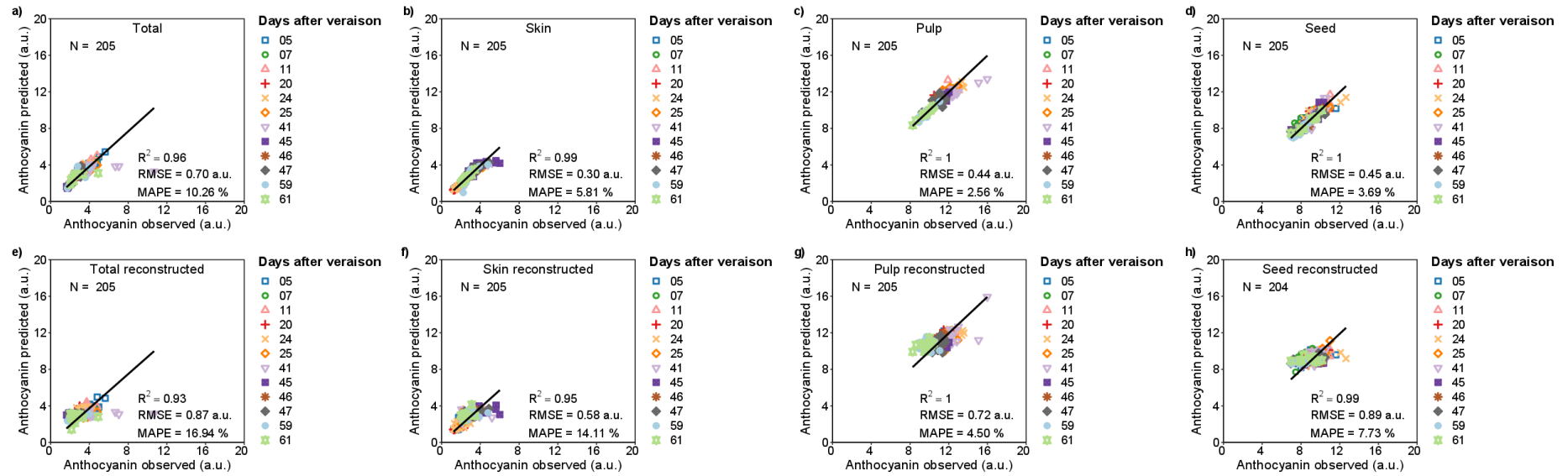


Figure S 5.3-5. Comparison of anthocyanin levels (a.u.) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a), skin (b), pulp (c) and seed (d). Reconstructed spectra for the total grape (e), skin (f), pulp (g) and seed (h).

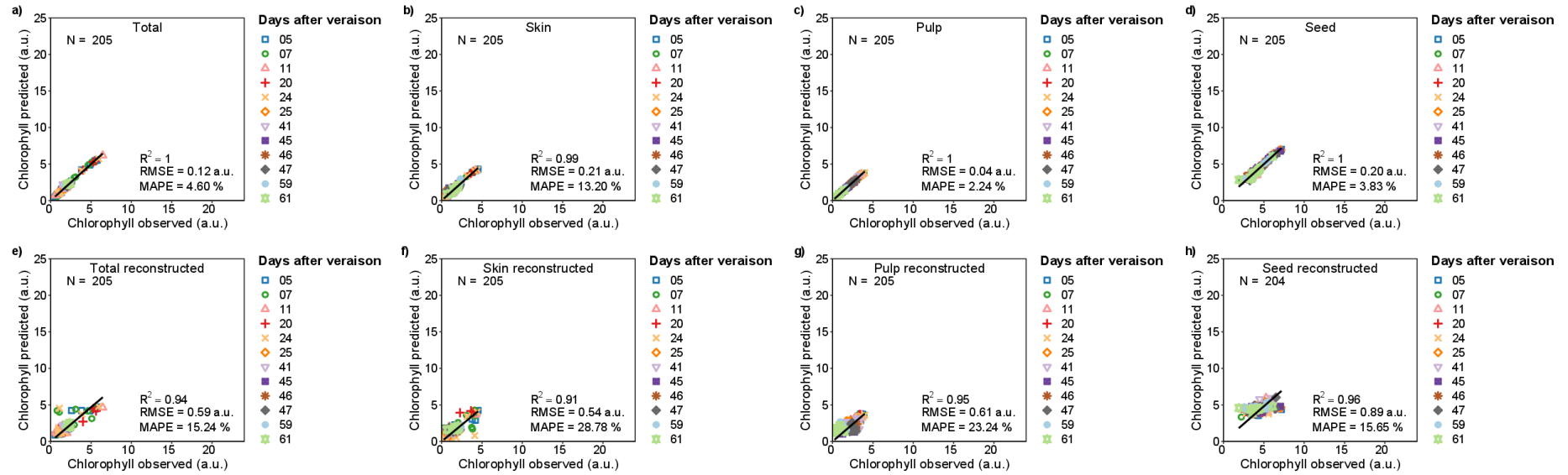


Figure S 5.3-6. Comparison of chlorophyll levels (a.u.) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a), skin (b), pulp (c) and seed (d). Reconstructed spectra for the total grape (e), skin (f), pulp (g) and seed (h).

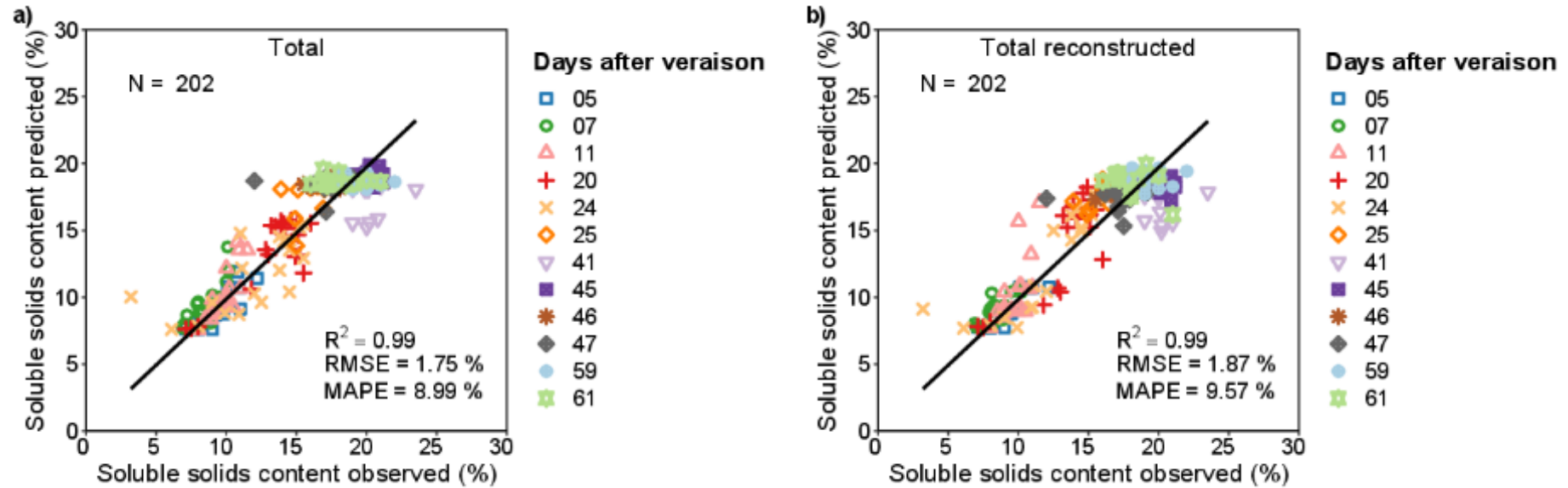


Figure S 5.3-7. Comparison of soluble solids content predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).

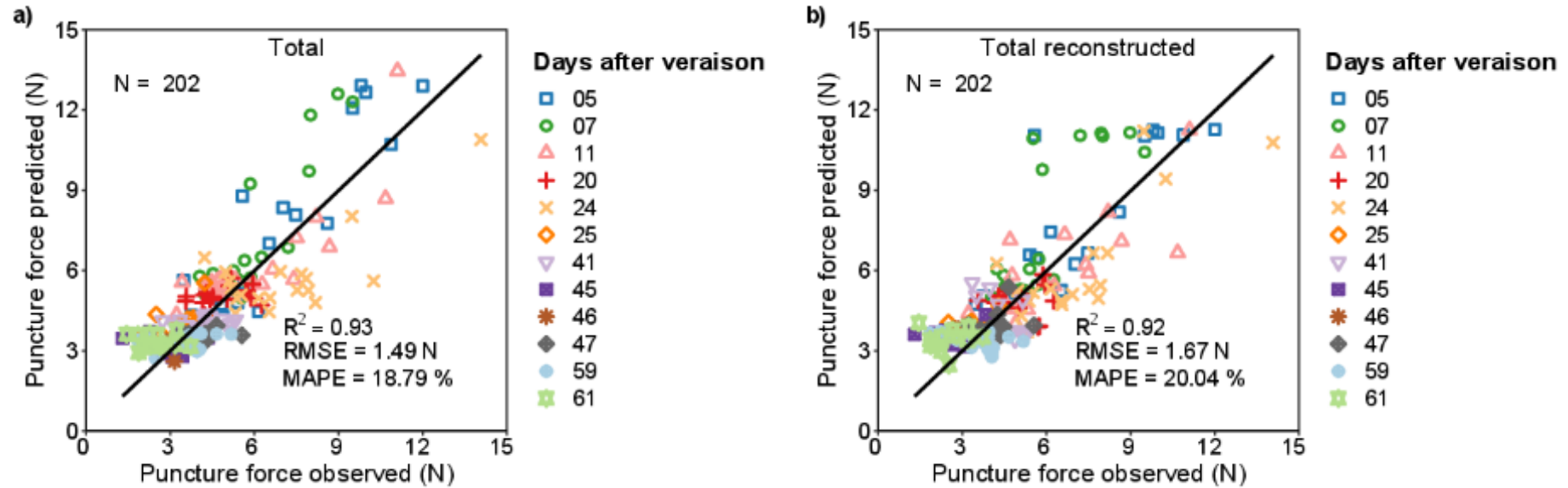


Figure S 5.3-8. Comparison of puncture force (N) predicted using a self-learning artificial intelligence algorithm (SL-AI) for the original and predicted datasets of total Vinhão grape, including the skin, pulp, and seed. N represents the number of observations considered by SL-AI. Original spectral data for the total grape (a) and reconstructed spectra for the total grape (b).

Table

Table S 5.3-1. Comparison of volume (cm³), soluble solids content (SSC - %), puncture force (PF - N), anthocyanin (Ant – a.u.), and chlorophyll (Chl – a.u.) in total grape (TO), skin (SK), pulp (PU), and seed (SE) of Loureiro and Vinhão cultivars across three assessment dates from days after veraison.

Variable	Days after veraison													p-value
	05	07	11	20	24	25	41	45	46	47	59	60	61	
Vol (cm³)	***	***	ns	ns	ns	ns	ns	ns	***	ns			**	
Loureiro	1.32 ^{abc} ± 0.36	1.50 ^a ± 0.46	1.07 ^{cd} ± 0.25	1.39 ^{ab} ± 0.40	1.37 ^{ab} ± 0.37	1.34 ^{ab} ± 0.38	1.36 ^{ab} ± 0.38	1.20 ^{bcd} ± 0.40	1.29 ^{abcd} ± 0.28	1.26 ^{abcd} ± 0.27	-	1.03 ^d ± 0.30	0.55 ^e ± 0.08	p < 0.001
Vinhão	0.93 ^e ± 0.16	0.97 ^e ± 0.20	1.20 ^{cd} ± 0.26	1.22 ^{cd} ± 0.27	1.39 ^{bc} ± 0.28	1.31 ^{bcd} ± 0.17	1.69 ^a ± 0.31	1.18 ^d ± 0.13	0.71 ^f ± 0.23	1.42 ^b ± 0.30	0.86 ^{ef} ± 0.16	-	0.73 ^f ± 0.21	p < 0.001
SSC (%)	***	***	***	***	***	**	***		*	ns			***	
Loureiro	16.31 ^{cd} ± 0.76	12.70 ^f ± 1.06	17.16 ^c ± 1.59	16.98 ^{cd} ± 0.67	17.06 ^{cd} ± 1.15	14.92 ^e ± 1.10	16.51 ^{cd} ± 1.27	20.71 ^b ± 0.60	16.34 ^{cd} ± 1.50	16.20 ^d ± 1.10	-	22.11 ^a ± 1.07	11.94 ^g ± 1.47	p < 0.001
Vinhão	9.23 ^f ± 1.29	8.54 ^f ± 0.99	9.79 ^f ± 0.95	13.22 ^d ± 2.65	11.85 ^e ± 3.41	16.17 ^c ± 1.15	20.00 ^a ± 1.05	20.16 ^a ± 0.82	17.30 ^{bc} ± 1.10	17.24 ^{bc} ± 1.65	19.48 ^a ± 1.13	-	17.99 ^b ± 1.44	p < 0.001
PF (N)	***	***	**	**	***	ns	**	ns	ns	ns			***	
Loureiro	2.97 ^{de} ± 0.33	3.11 ^{de} ± 0.52	4.00 ^{bc} ± 0.97	3.54 ^{cde} ± 0.87	4.47 ^b ± 1.01	3.54 ^{cde} ± 0.76	3.35 ^{cde} ± 0.87	2.77 ^e ± 0.82	3.50 ^{cde} ± 0.60	3.76 ^{bcd} ± 0.53	-	3.27 ^{cde} ± 1.02	9.95 ^a ± 2.16	p < 0.001
Vinhão	7.08 ^a ± 2.62	5.95 ^a ± 1.58	6.14 ^a ± 2.27	6.99 ^a ± 5.20	7.71 ^a ± 3.15	3.46 ^b ± 0.68	4.22 ^b ± 0.74	3.14 ^b ± 0.79	3.43 ^b ± 0.51	4.16 ^b ± 0.58	3.58 ^b ± 0.85	-	2.49 ^b ± 0.66	p < 0.001
C_{TO} (a.u.)	***	***	***	***	***	***	***	***	***	***			***	
Loureiro	7.59 ^a ± 0.51	7.02 ^b ± 0.45	6.65 ^c ± 0.64	6.49 ^{cd} ± 0.58	6.18 ^{de} ± 0.55	5.47 ^g ± 0.62	5.86 ^{ef} ± 0.49	5.78 ^{fg} ± 0.42	5.63 ^{fg} ± 0.34	4.88 ^h ± 0.39	-	4.41 ⁱ ± 0.38	6.20 ^{de} ± 0.74	p < 0.001
Vinhão	2.45 ^a ± 1.87	1.86 ^a ± 1.52	1.64 ^a ± 1.23	2.04 ^a ± 1.30	1.81 ^a ± 1.19	2.13 ^a ± 0.21	2.06 ^a ± 0.28	2.41 ^a ± 0.05	2.27 ^a ± 0.11	2.32 ^a ± 0.11	2.30 ^a ± 0.16	-	2.32 ^a ± 0.11	0.25
C_{SK} (a.u.)	***	***	***	***	***	***	***	***	***	***			***	
Loureiro	22.28 ^a ± 0.79	22.04 ^{ab} ± 0.77	22.02 ^{ab} ± 0.63	22.00 ^{ab} ± 0.65	22.41 ^a ± 0.76	21.54 ^{bc} ± 0.69	21.48 ^c ± 0.70	21.06 ^{cd} ± 0.58	20.56 ^d ± 0.68	19.87 ^e ± 0.94	-	20.62 ^d ± 0.61	21.98 ^{ab} ± 0.61	p < 0.001
Vinhão	1.76 ^a ± 1.72	1.49 ^{ab} ± 1.61	0.85 ^b ± 0.90	1.28 ^{ab} ± 0.97	1.32 ^{ab} ± 1.23	1.35 ^{ab} ± 0.50	1.51 ^{ab} ± 0.49	2.02 ^a ± 0.35	1.41 ^{ab} ± 0.28	1.75 ^a ± 0.54	1.95 ^a ± 0.26	-	1.59 ^{ab} ± 0.41	0.013
C_{PU} (a.u.)	***	***	***	***	***	***	***	***	***	***			***	
Loureiro	20.29 ^a ± 0.59	19.64 ^b ± 0.75	19.61 ^b ± 0.44	19.12 ^{bcd} ± 0.59	19.56 ^{bc} ± 0.66	18.59 ^{ef} ± 0.97	18.97 ^{de} ± 0.86	19.49 ^{bcd} ± 0.72	18.19 ^f ± 1.15	18.19 ^f ± 0.82	-	19.07 ^{cde} ± 0.68	20.40 ^a ± 1.08	p < 0.001
Vinhão	3.50 ^a ± 0.14	3.38 ^a ± 0.28	3.31 ^{ab} ± 0.21	3.08 ^b ± 0.26	3.22 ^{ab} ± 0.37	2.44 ^{cd} ± 0.53	2.61 ^c ± 0.58	1.87 ^e ± 0.52	2.49 ^c ± 0.28	2.18 ^d ± 0.54	0.75 ^f ± 0.38	-	0.65 ^f ± 0.40	p < 0.001
C_{SE} (a.u.)	**	*	ns	***	*	ns	ns	ns	ns	**			***	

Loureiro	5.63 ^a	±	4.87 ^{ab}	±	4.89 ^{ab}	±	4.97 ^{ab}	±	4.93 ^{ab}	±	4.98 ^{ab}	±	5.30 ^{ab}	±	5.17 ^{ab}	±	4.67 ^b	±	5.13 ^{ab}	±	-	5.39 ^{ab}	±	5.44 ^{ab}	±	0.195
	1.19		0.86		1.20		0.84		0.75		0.91		1.13		0.74		0.55		0.85		-	1.35		1.02		
Vinhão	4.53 ^{bcd}	±	4.21 ^{bcd}	±	4.41 ^{bcd}	±	3.95 ^d	±	4.26 ^{cd}	±	4.94 ^{ab}	±	5.40 ^a	±	4.87 ^{abc}	±	4.36 ^{bcd}	±	4.77 ^{abc}	±	4.14 ^{cd}	±	-	3.91 ^d	±	p <
	0.73		1.10		0.85		0.71		0.91		0.98		0.75		1.07		0.83		0.95		0.98		-	0.98		0.001
A TO	4.25 ^a	±	3.75 ^{ab}	±	3.73 ^{ab}	±	3.18 ^{bc}	±	3.37 ^b	±	3.42 ^b	±	4.09 ^a	±	2.11 ^e	±	2.75 ^{cd}	±	2.61 ^{cde}	±	2.54 ^{de}	±	-	2.67 ^{cde}	±	p <
(a.u.)	0.71		0.41		0.47		0.43		0.34		0.57		1.96		0.35		0.43		0.44		0.52		-	0.79		0.001
A SK	1.51 ^e	±	1.46 ^e	±	1.55 ^e	±	1.89 ^e	±	1.84 ^c	±	2.95 ^c	±	3.41 ^b	±	3.83 ^a	±	3.10 ^{bc}	±	3.31 ^{bc}	±	2.97 ^c	±	-	2.43 ^d	±	p <
(a.u.)	0.14		0.12		0.19		0.39		0.52		0.78		0.84		1.03		0.40		0.62		0.64		-	0.43		0.001
A PU	11.77 ^{bc}	±	11.84 ^{bc}	±	11.94 ^{ab}	±	11.63 ^{bcd}	±	12.00 ^{ab}	±	11.56 ^{bcd}	±	12.36 ^a	±	11.41 ^{cd}	±	11.23 ^{de}	±	10.93 ^e	±	9.79 ^f	±	-	9.61 ⁱ	±	p <
(a.u.)	0.29		0.44		0.47		0.36		0.75		0.52		1.22		0.38		0.24		0.62		0.54		-	0.65		0.001
A SE	9.56 ^{ab}	±	8.42 ^d	±	9.98 ^a	±	9.45 ^{abc}	±	9.68 ^{ab}	±	9.73 ^{ab}	±	9.42 ^{abc}	±	9.23 ^{bc}	±	9.46 ^{abc}	±	8.54 ^d	±	8.38 ^d	±	-	8.81 ^{cd}	±	p <
(a.u.)	0.89		0.60		0.55		0.64		1.25		0.63		0.79		1.11		0.73		1.05		0.94		-	0.95		0.001

Means with a p-value < 0.05 is considered statistically different. For cases with p-value < 0.05, within lines, different letters signify the mean difference according to Duncan's post hoc test ($\alpha = 5\%$). The comparison between cultivars for the same variables and same days after veraison considered the following levels of significance: non-significant (ns), significant ($p < 0.05$, *), highly Significant ($p < 0.01$, **) and very highly significant ($p < 0.001$, ***).

CHAPTER 6

CASE STUDY VI: DOURO WINE REGION CASE
STUDY: METABOLIC MAPS FOR TISSUE EVOLUTION
AND MATURATION

6. Case study VI: Douro wine region case study: metabolic maps for tissue evolution and maturation.

In preparation

Tosin, R. and supervision team.

TRIBE Team led by Doctor Filipe Santos.

Activities within the project OmicBots - OmicBots: High-Throughput Integrative Omic-Robots Platform for a Next Generation Physiology-based Precision Viticulture, with reference PTDC/ASP-HOR/1338/2021.

Metabolic maps for precision grape maturation: integrating physicochemical parameters using sensing high-throughput tomography-like method to define site-specific management

Abstract

This study investigates grape maturation dynamics and terrain variations in a vineyard within the Douro Wine Region under different irrigation levels (no irrigation, 30% Etc, and 60% Etc). Metabolic maps were generated to analyse grape phenotype using a tomography-like method based on Vis-NIR data collected from a point-of-measurement sensor over two consecutive years (2021 and 2022). Non-destructive measurements assessed physicochemical parameters, including soluble solids content (SSC), puncture force (PF), volume, weight/volume ratio, chlorophyll, and anthocyanin in the skin, pulp, and seed across six post-veraison dates. Comparative analysis between random forest (RF) and self-learning artificial intelligence (SL-AI) algorithms showed SL-AI's superior robustness within the destructive dataset (SL-AI – $R^2 \geq 0.87$, $MAPE \leq 37.55\%$; RF – $R^2 \geq 0.86$, $MAPE \leq 46.20\%$). At the same time, RF exhibited greater stability in predicting the non-destructive dataset, offering more logical results in physicochemical parameter evolution (e.g., SSC) during grape maturation. Altitude and NDVI maps revealed differences within the vineyard parcel, sometimes unrelated to irrigation treatments. Significant variations ($p < 0.05$) in SSC, chlorophyll, and anthocyanin levels, particularly in non-irrigated locations, suggested environmental factors' influence on grape maturation. Simulations for winemaking goals based on biophysical variables advanced viticulture practices, providing insights into grapevine responses to environmental and terrain characteristics. By embracing site-specific management (SSM) principles, tailored resource allocation and agronomic strategies addressed specific needs across the vineyard, enhancing wine production efficiency through precision viticulture (PV) practices. This approach fosters differential actions, ensuring interventions are precisely calibrated, offering practical applications for decision-making in PV and enhancing wine production efficiency.

Keywords

Precision viticulture, terrain mapping, spatialisation, SSC, chlorophyll, anthocyanin

6.1. Introduction

Grape maturation is influenced by various environmental factors, including temperature, sunlight, and water availability (Webb et al., 2012), which limits informed decisions for

vineyard management in the scope of precision viticulture (PV). Understanding this variability is crucial for making informed decisions regarding PV.

The maturation process of grapes determines their quality and suitability for winemaking. Grape maturation encompasses several physicochemical transformations, including alterations in soluble solids content (SSC), puncture force (PF), volume, weight-to-volume ratio, and pigment content within the internal tissues (Rogiers et al., 2022).

Among the several physicochemical parameters evaluated during grape maturation, SSC, which represents the concentration of sugars in the grape berries, is one of the most used. However, relying solely on SSC can be challenging as it may not provide accurate and complete information due to its dependence on other physicochemical parameters such as PF, pigmentation, and environmental conditions. PF holds implications, particularly in mechanical harvesting or table grape production, where berry firmness affects handling and transport (Piazzolla et al., 2016).

Besides SSC, PF and pigment content could also be considered when assessing grape maturation. PF measures the firmness of grape berries, indicating their structural integrity and readiness for harvesting (Piazzolla et al., 2016). Pigment content, including chlorophylls (CHL) and anthocyanins (ANT), contributes to grape colouration and can vary throughout maturation (Massonnet et al., 2017).

The nonlinear behaviour of CHL and ANT during grape maturation means that their concentration tends to fluctuate unpredictably over time (Zhao et al., 2021). This variability minimises due to changes in environmental conditions such as temperature, sunlight exposure, and soil composition, which can directly influence the synthesis and degradation of ANT within the grape tissues (Garrido et al., 2023). Genetic factors within grapevine cultivars may also contribute to the nonlinear patterns observed, as different varieties may exhibit varying responses to environmental conditions during maturation (Rogiers et al., 2022; Zhao et al., 2021). Furthermore, interactions between ANT and other compounds within the grape, such as tannins and sugars, could further affect their behaviour, leading to the observed unpredictability in their concentration levels throughout the maturation process (Rousserie et al., 2019). Winemakers typically seek a balance among various parameters to achieve desired wine characteristics, often aiming for high SSC, elevated ANT levels, and low CHL content (Pampuri et al., 2021b). Additionally, grape assessment is related to volume measurements, which can estimate productivity (Anastasiou et al., 2022). These processes can be monitored throughout the vegetative cycle through cultural practices such as irrigation management or harvest decisions.

While various physicochemical parameters are evaluated during grape maturation to measure readiness for harvest and winemaking, conventional techniques such as high-performance liquid chromatography (HPLC) are restrictive due to their destructiveness, expense, and time-consuming nature (Roitsch et al., 2019). Additionally, the lack of optimised protocols to assess small amounts of tissues individually results in a limited understanding of grape development and hinders the transferability of PV approaches across diverse environmental conditions (Tosin et al., 2024b). Consequently, there is a pressing need for improved methods that can comprehensively and non-destructively evaluate grape maturation parameters, enabling more informed decision-making in vineyard management practices. Explicitly acknowledging this gap and advocating for the development of innovative techniques would provide a more explicit rationale for conducting further research in this area.

Therefore, this study is built upon three main hypotheses. Firstly, considering the successful validation of the tomography-like method in cherry tomatoes (Tosin et al., 2024a) and grapes from the Região dos Vinhos Verdes with cultivars Loureiro and Vinhão (Tosin et al., 2023b), where it effectively characterised physicochemical parameters of the SK, pulp (PU), and SE, it is hypothesised that this method will also be effective in a different grape cultivar within a drought-prone region, such as the Douro Wine Region. Secondly, the tomography-like method will consistently produce results that align across two years of data. Thirdly, the tomography-like method will perform effectively under varied irrigation treatments, altitudes, and vineyard vigour zones, from post-veraison to harvest.

A thorough understanding and continuous monitoring of grape maturation are imperative for grape growers and winemakers to harvest grapes at the optimal maturity level, ensuring the production of high-quality wines aligned with predetermined goals. However, these parameters exhibit temporal and spatial variability influenced by various factors, including soil characteristics, vineyard vigour, climate conditions, and tissue types. Therefore, precise and periodic analyses of these physicochemical parameters are necessary to monitor grape maturation and make well-informed harvesting decisions effectively. By combining these phenotype characteristics, a more accurate assessment of grape maturation can be achieved, facilitating the identification of the appropriate harvest time tailored to specific winemaking objectives (Sohaib Ali Shah et al., 2020) once changes in weather conditions, such as rainfall, can impact readings of SSC between measurements, resulting in inaccurate assessments of grape maturity (Xu et al., 2022; Piazzolla et al., 2016). Therefore, implementing site-specific management (SSM) strategies enables vineyard managers to optimise grape maturation processes,

considering multiple physicochemical parameters and making informed harvesting decisions aligning with specific winemaking goals.

One indicator of vineyard vigour and changes in physicochemical parameters (e.g., SSC, hydric status) is the normalised difference vegetation index (NDVI), making it a tool for decision-making, such as irrigation scheduling and fertilisation practices (Meyers et al., 2020; Webb et al., 2012), but it lacks in provides precise information at the grape level. The necessity to map spatial and temporal variations of grape physicochemical parameters arises due to the better explanation of grape maturation and SSM. Bramley et al. (2017) examined the spatial variation of rotundone, a compound responsible for “pepperiness” in grape berries, and its stability from year to year, shedding light on its influence on wine style. Similarly, Yu et al. (2020) explored the relationship between soil, plant water status, and grapevine physiology, providing insights into how these factors affect grape quality. Despite their contributions, these studies did not incorporate metabolic mapping techniques. However, Yu et al. (2021) and Shmulevitz et al. (2024) addressed this gap by mapping the spatial variability of grape berry maturation at the molecular level and employing sensors and gene expression analysis to understand differences in vineyard vigour and fruit quality. Nevertheless, a spatial variation of the inner tissues is still needed. This absence highlights the need to integrate metabolic mapping into research methodologies to advance the understanding grape physiology and facilitate informed decision-making in vineyard management practices through SSM. By understanding how grape maturation evolves across the internal tissues, the tomography-like approach can offer the capability to map grape maturation *in situ*, contributing to the comprehension of the terrain characteristics influencing grape development, allowing viticulturists to tailor cultural practices, such as irrigation and pruning, according to the specificities of the vineyard location. Additionally, incorporating the concept of metabolic maps, representing the interactions within an organism, can further enhance the understanding of grape maturation. Metabolic maps illustrate the physicochemical parameters occurring within grapes during maturation, providing information on nutrient uptake, sugar synthesis, and phenolic compound production (Gago et al., 2017). By integrating metabolic mapping into this tomography-like method, viticulturists can gain deeper perceptions of the metabolic processes driving grape maturation. This allows precise SSM for optimising grape quality according to specific goals and vineyard productivity.

Hence, this study seeks to generate metabolic maps showing the temporal progression of grape maturation processes while integrating biophysical parameters through a tomography-like method. With this aim, the research addresses the following questions:

i) does the tomography-like method work in different conditions where it has been previously tested?, ii) what is the influence of environmental factors on the temporal dynamics of grape maturation? iii) how do terrain characteristics contribute to shaping the developmental dynamics of grapes within the surveyed parcel? and iv) how can metabolic mapping advance the comprehension of grape physiology and facilitate informed decisions in vineyard management through the SSM?

6.2. Material and Methods

6.2.1. Test site

The Douro Wine Region is famous for its Port wine production. Viticulture in this area faces tough environmental conditions, including steep slopes, shale-derived soils, and one of Europe's driest climates. During the maturation period from around 20 July to 20 September, rainfall often stays below 28 mm in approximately 80% of years. This lack of water leads to a limited water reserve, typically not exceeding 20% by the end of the maturation. Consequently, berry weight is affected, reducing wine production yield and quality, mainly when irrigation is impossible (Cunha & Richter, 2016).

The research was conducted at Quinta dos Aciprestes in the Douro Wine Region, which Real Companhia Velha owns. The vineyard is situated at Latitude 41.21° N and Longitude 7.43° W, and the studied parcel has an elevation of approximately 187 meters above sea level. The study encompasses the period from post-veraison to harvest in 2021 and 2022. The parcel includes the Touriga Nacional grape variety (Vitis International Variety Catalogue VIVC - 12594), established on Bilateral Royat pruning systems and planted with a spacing of 2.2 m between rows and 1 m between individual plants.

The company implemented and supervised the irrigation system, utilising drip irrigation for accurate and effective water distribution to the vines. Emitters are strategically positioned 1 meter apart, each delivering water at a rate of 2 L h⁻¹ per emitter, ensuring optimal water supply to meet the vine's needs during the entire growth cycle. The vineyard was divided into two plots, each subjected to three distinct irrigation treatments: non-irrigated (NI), irrigation to replace 30% of evapotranspiration (Etc 30 %) water volume, and 60% Etc (Etc 60%). The wine company managed these treatments to monitor and regulate water usage.

6.2.2. Vineyard's spatiotemporal variability based on satellite imagery

Altitude mapping characterises slopes and water stress, as it helps detect variability across the terrain (Sassu et al., 2021). The altitude data for the studied parcel was sourced from the Advanced Land Observing Satellite Digital Elevation Model (ALOS DEM) with a 12.5 m resolution, accessed through the Alaska Satellite Facility (ASF) Distributed Active Archive Center (DAAC). This dataset is a product of the Shuttle Radar Topography Mission (SRTM) utilising the ALOS Phased Array type L-band Synthetic Aperture Radar (ALOS PALSAR) (©JAXA/METI ALOS PALSAR, 2007).

NDVI can describe vegetation vigour by quantifying the density of vegetation based on the differential reflectance of near-infrared and red light (Kandylakis et al., 2020). Sentinel-2 Level 2A (L2A) data featuring atmospheric correction was used to compute NDVI. Sentinel-2 has a spectral resolution of 10 m in the Vis-NIR spectral range and a temporal frequency of five days.

The NDVI computation employed the formula $(\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED})$, with the satellite's bands 4 (665 nm) and 8 (842 nm) representing the red and near-infrared (NIR) spectral regions. Cloud-free images were selected from the most proximal sampling dates. Table 6-1 presents the data for measurement and satellite imagery.

Table 6-1. Field sampling dates and corresponding proximal satellite imagery were selected to represent each sampling date in 2021.

	Field Sampling	Satellite imagery
S1	29/07/2021	28/07/2021
S2	05/08/2021	07/08/2021
S3	12/08/2021	12/08/2021
S4	19/08/2021	17/08/2021
S5	26/08/2021	27/08/2021
S6	02/09/2021	11/09/2021

The predawn leaf water potential (PLWP) was used to characterise zones of different altitudes and vigour of the studied parcel. A pressure chamber (Scholander et al., 1965) (PMS600, Albany, OR, USA) was used to quantify PLWP in the parcel.

6.2.3. Experimental design of grape sampling

The temporal grape sampling underwent a study across six post-veraison dates in 2021 and 2022. In 2021, the veraison occurred on 24/07/2021, and the sampling dates were designated as S1 (29/07/2021 – 05 days after veraison, DAV)), S2 (05/08/2021 – 12 DAV), S3 (12/08/2021 – 19 DAV), S4 (19/08/2021 – 26 DAV), S5 (26/08/2021 – 33 DAV), and S6 (02/09/2021 – 40 DAV). Similarly, in 2022, veraison took place on 25/07/2022, and the six sampling dates ranged from 03 to 38 DAVs, labelled as S1 (28/07/2022 – 03

DAV), S2 (04/08/2022 – 17 DAV), S3 (11/08/2022 – 24 DAV), S4 (18/08/2022 – 31 DAV), S5 (25/08/2022 – 31 DAV), and S6 (01/09/2022 – 38 DAV).

Sampling involved collecting one random grape per grapevine in the studied grapevine lines, considering well-developed bunches, similar solar exposure, and grapes at the bunch's top. In 2021, 20 grapes per treatment line were collected, totalling 120 grapes per date, with 32 selected for destructive analyses. In 2022, six grapes per treatment line were collected, totalling 36 grapes per date that underwent non-destructive and destructive analysis (Figure 6-1).

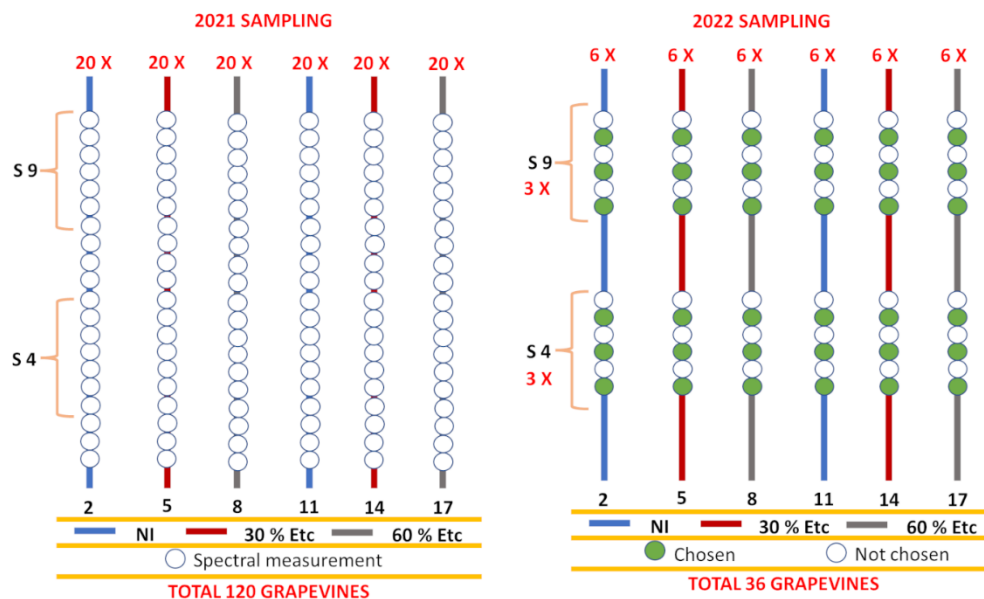


Figure 6-1. Description of the sampling approach employed in 2021 and 2022. In 2021, 20 grapes per irrigation line were selected for spectral analysis, with 32 samples randomly chosen from the 120 collected per date for subsequent assessment of internal tissues. S4 is the fourth set of grapevines in the row with a lower altitude, whereas S9 is the ninth sector with a higher altitude.

The destructive measurements were employed to construct the foundational knowledge base for the tomography-like method. Subsequently, the non-destructive measurements were utilised to predict the internal tissues of the grapes, leveraging the knowledge base established through the destructive analysis (see subsection 2.5). However, some grapes suffered damage during transportation and were excluded from destructive measurements (Table 6-2).

Table 6-2. The number of grapes used in destructive and non-destructive measurements during the evaluation dates in 2021 and 2022.

	2021			2022		
	DAV	Non-destructive	Destructive	DAV	Non-destructive	Destructive
S1	05	120	32	03	36	36
S2	12	120	32	10	36	36
S3	19	120	31	17	36	8

S4	26	120	32	24	36	17
S5	33	120	32	31	36	33
S6	40	120	32	38	36	36
Total	-	720	191	-	216	166

Days after veraison (DAV). In 2021, S1 (29/07/2021), S2 (05/08/2021), S3 (12/08/), S4 (19/08/2021), S5 (26/08/2021), and S6 (02/09/2021). In 2022, S1 (28/07/2022), S2 (04/08/2022), S3 (11/08/2022), S4 (18/08/2022), S5 (25/08/2022), and S6 (01/09/2022).

6.2.4. Grape measurements

6.2.4.1. Spectra measurement

All the grapes sampled (destructive or not) were analysed spectrally using a high-resolution spectrometer (Ocean Insite HR4000, 195.34-1118.33 nm) equipped with a reflectance fibre optic probe. The LED (6500k, Philips SpotOn Ultra 69141/31/PH) illuminated the grape positioned above a platform, and the fibre probe was adjusted to capture maximum transmitted light from the tissues. Integration times were optimised for each sample to balance data acquisition durations, ensuring linear response range adherence and preventing detector saturation. Longer integration times were applied to samples with low light transmittance, while shorter ones were used for more accessible light transmittance samples (Tosin et al., 2023b). Despite the limited spectral range of LED illumination, all acquired spectral data were included in subsequent analyses (195 to 1118 nm). However, the LED lighting setup restricted the utilisation of the full spectrum of wavelengths available in the equipment.

6.2.4.2. Physicochemical parameters measurements

Before the destructive analysis, the selected grapes were weighed with a precision scale. The grape volume (VOL - cm³) was calculated using the sphere volume formula based on their width and length measured using a digital calliper (LiMiT Caliper Digital 150mm). Next, the relation of weight/volume (W/VOL – g/ cm³) was established.

Regarding the destructive analysis, the grapes were first submitted to the maximum PF, expressed in Newtons (N), determined using a digital penetrometer (model PCE-PTR 200, PCE Group, D-59872 Meschede, Germany). Subsequently, SSC in percentage (%) was immediately measured with a hand refractometer (Milwaukee model MR32ATC) with a scale range from 0 to 32.0%.

Then, SK, PU, and SE were separated for each grape by using a scalpel. Portions measuring approximately 0.5 cm² and 3 mm thick were carefully arranged on distinct slides. Due to the minimal alterations in SE size and shape post-veraison, one single seed can represent the entire grape (Ristic & Iland, 2005). Immediately, the spectra data of each tissue was obtained through the same spectroradiometer previously mentioned.

The content of CHL and ANT in the entire grape (TO), SK, PU and SE was determined indirectly using spectral information (a.u.) collected from small aliquots of each tissue. CHL quantification considered the green (520-570 nm) and red (571-700 nm) spectral zones, while ANT content was estimated through the band ratio 705 nm by 680 nm (Tosin et al., 2023b). Spectroscopy was chosen for pigment inference due to the limitations of conventional methods, such as their non-expedient, expensive, and time-consuming nature for small tissue samples. The resulting data showcased the effectiveness of reconstructed tissue spectra in establishing a relationship with estimates derived from grape spectra, validating the proof-of-concept for internal tissue quantification (Tosin et al., 2023b).

6.2.4.3. Spectra processing

The obtained spectra from TO and internal tissues underwent a logarithm multiplicative scattering correction (MSC-log) method, a validated technique for baseline and scattering artefact correction (Martins et al., 2022). This correction effectively reduces unwanted variations, enhancing the accuracy and reliability of data analysis. Standardisation was also applied to original and reconstructed data to address signal intensity differences. Standardisation involves calculating z-scores by subtracting the mean and dividing by the standard deviation for each data point. Standardising the data minimised the impact of signal intensity, enabling more precise comparisons and analyses (Tosin et al., 2023a). These preprocessing measures played a vital role in addressing spectral data insufficiencies, ensuring the accuracy and meaningful interpretation of the results.

Martins et al. (2023) and Tosin et al. (2023b) describe detailed information about the algorithm structure.

6.2.5. Tomography-like method

The tomography-like method (Martins et al., 2023) employed in this study operates in two directions. Initially, the reconstruction of TO involves integrating the spectral data of SK, PU, and SE. Subsequently, TO spectra are decomposed to reconstruct individual internal tissue spectra. Grape tissue spectra demonstrate interconnected patterns, facilitating bidirectional information flow and the identification of unknown samples by comparing them with similar ones (Figure 6-2).

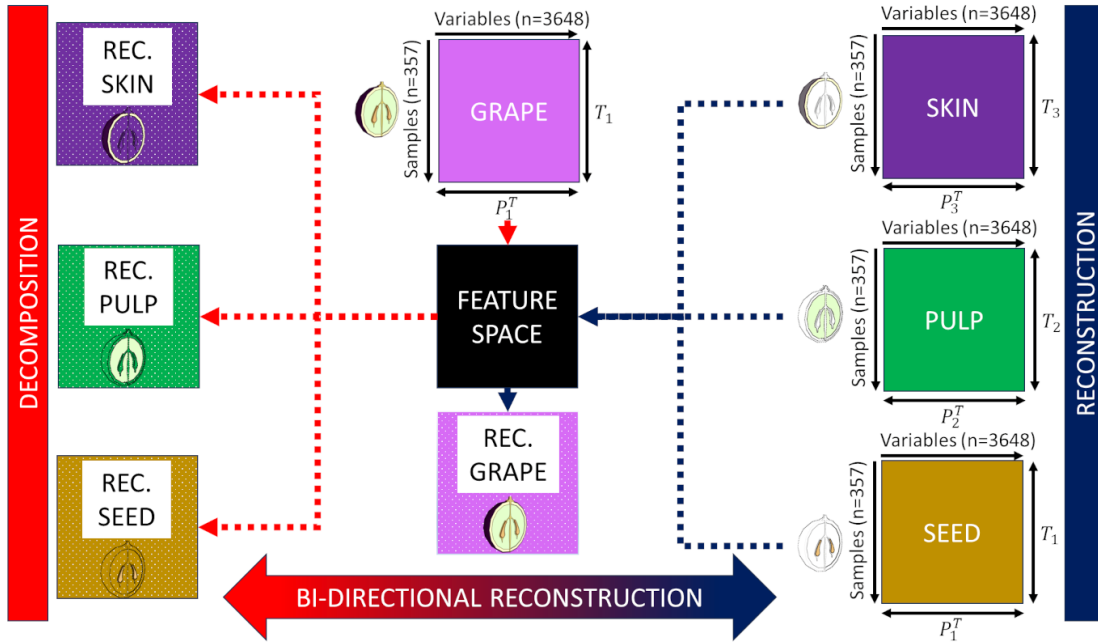


Figure 6-2. Exemplification of the tomography-like method. P_i^t is the contribution of the SK, PU and SE of each subset (P_1^t , P_2^t and P_3^t). The variables (n=3648) refer to each principal component (PC) of the spectral data. Sample (n=357) refers to the destructive measurements of grapes that were used to construct the knowledge base.

This method segregates grape tissue data and reveals unique spectral patterns. Depositing the data into eigenvectors creates a new basis for projecting dataset information. Examining eigenvectors using randomisation tests enables the separation of random and relevant data. For grape tissues, PCA decomposition is applied to SK X_1 , PU X_2 , and SE X_3 , producing latent features ($[t_1, t_2, t_3]$) that reconstruct tissue spectra while discarding random data. Concatenating dimensions t_1 , t_2 , and t_3 forms a feature space T , identifying relevant superset latent information (T_i) and each subset's contribution to fused information (P_i^t). The direct linkage between superset T and PCA grape spectra decomposition (Y) suggests similar clustering for compositionally alike samples, reflecting grape SK, PU, and SE maturation stages. This bi-directionality between subsets X_1 (SK), X_2 (PU), X_3 (SE), and Y allows for identifying unknown sample tissues. The validation process involves hierarchical latent structure development (optimisation and validation) and cross-validation to determine the optimal number of principal components (PCs) for sub-spaces (t_1, t_2, t_3) and T . Validation of tissue spectra reconstruction is performed through cross-validation, considering metrics such as Euclidean distance and principal component regression (PCR) to minimise errors and prevent overfitting. This methodology has been detailed by Tosin et al. (2024a) and Tosin et al. (2023b).

6.2.5.1. Identifying internal tissues of non-destructive grapes

To identify the characteristics of the non-destructive grapes' internal tissues measured in 2021 ($n=720$), the k-nearest neighbours (KNN) algorithm, a non-parametric and instance-based supervised learning method, was utilised. This algorithm predicts the value of a new data point based on the values of its k nearest neighbours in the feature space (Zhang et al., 2018a). The term "nearest neighbours" denotes the data points most similar to the query point, determined by the Euclidean distance.

The KNN algorithm was used to identify similar grape clusters from the non-destructive samples, thereby facilitating the identification of the SK, PU, and SE spectra from the non-destructive grape measurements.

6.2.6. Modelling and mapping physicochemical parameters

After implementing the KNN algorithm, the SK, PU and SE were predicted on grapes not subjected to destructive analysis. This prediction was based on a dataset collected over two years, totalling 357 grapes observed.

By discerning similarities and cluster patterns among the grapes, the KNN method facilitated the prediction of non-destructive samples by leveraging the shared characteristics of their destructively assessed equivalences. Following this, the physicochemical parameters measured in this study were estimated using the random forest (RF) algorithm (Breiman, 2001) and the self-learning artificial intelligence (SL-AI) algorithm (Martins, 2018). RF is widely recognised and employed in the estimation of physicochemical parameters in various crops (e.g., Kalopesa et al., 2023; Zahir et al., 2022) as well as the new SL-AI (e.g., Tosin et al., 2023b; Martins et al., 2022).

For the RF modelling approach, the first 10 PCs of the total SK, PU, and SE spectra were considered for modelling, while SL-AI considered the entire spectral range.

To assess the robustness of RF and SL-AI in predicting SSC, PF, CHL, and ANT content, the statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE - %), were employed.

For the RF model, the dataset ($n=357$) was divided into training (90%, $n=325$) and validation (10%, $n=32$) sets.

The non-destructive dataset, comprising 720 samples, was utilised for modelling the physicochemical parameters of each georeferenced grapevine. Through inverse distance weighted (IDW) interpolation of a point vector layer, the predicted parameters were spatialised for each grapevine. Subsequently, metabolic maps were generated by discretely interpolating the quartile values of the variables by using only three classes.

QGIS software (version 3.16.0-Hannover, 2020) was employed to conduct geographic information system (GIS) analysis.

6.2.6.1. Decision maps criteria based on site-specific management

This study aimed to utilise metabolic maps analysis, supporting decision-making in vineyard management and winemaking processes in the context of SSM. These diagrams were constructed through the implementation of three distinct models:

- Model 1: This model does not assign specific weightings to the variables. It is suitable for scenarios where higher SSC values, elevated ANT levels in grape skin SK, and lower CHL content in PU are desired.
- Model 2: In this model, 60% of the criteria are attributed to SSC values, followed by 30% for ANT levels in SK and 10% for CHL content in PU. This model prioritises SSC values but also considers ANT levels and CHL content.
- Model 3: ANT levels in SK are considered the most crucial variable, commanding 60% importance. SSC values account for 30%, while CHL content in PU holds a 10% weight. This model places a greater emphasis on ANT levels than the other variables.

The variables were normalised for comparison and analysis using the maximums method for these hypothetical simulations. These models showcase how various grape characteristics obtained from metabolic maps can be grouped to inform decision-making in viticulture and winemaking. While these specific weightings are presented as hypotheses, exploring other combinations to tailor the models according to specific objectives and preferences in vineyard management and winemaking practices is possible.

6.3. Results

6.3.1. Meteorology during the maturation period

Figure 6-3 presents the average temperature and precipitation from 1 July to 30 September, illustrating the period of grape maturation and assessment dates from both years.

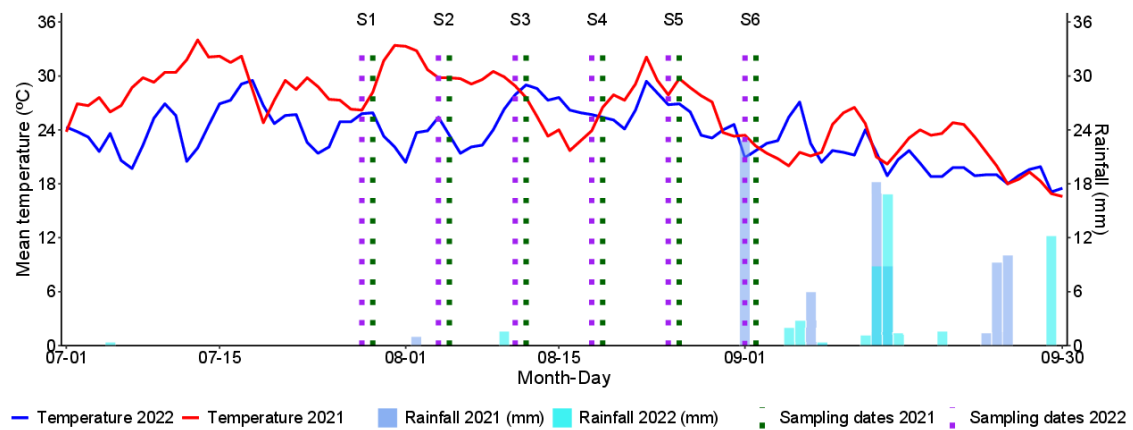


Figure 6-3. Daily mean temperature and total precipitation between 1 July and 30 September 2021 and 2022. In 2021 S1 (29/07/2021), S2 (05/08/2021), S3 (12/08/2021), S4 (19/08/2021), S5 (26/08/2021), and S6 (02/09/2021). In 2022, S1 (28/07/2022), S2 (04/08/2022), S3 (11/08/2022), S4 (18/08/2022), S5 (25/08/2022), and S6 (01/09/2022).

A consistent climatic pattern characterised by minimal precipitation and elevated temperatures was observed throughout the observation periods covering from S1 to S6 in 2021 and 2022. Precipitation was almost absent during these intervals in both years, except in S6 (02/09/2021) in 2021, when precipitation was observed after the measurement in S6. The mean yearly temperature was approximately 26 °C, and 2022 was hotter than 2021. Although 2022 was warmer and drier compared to 2021.

6.3.2. Vineyard spatiotemporal variability

The variation in altitude in the assessed lines presents a slope of approximately 15%, with a higher slope observed in the lines from right to left of the parcel studied (Figure 6-4).

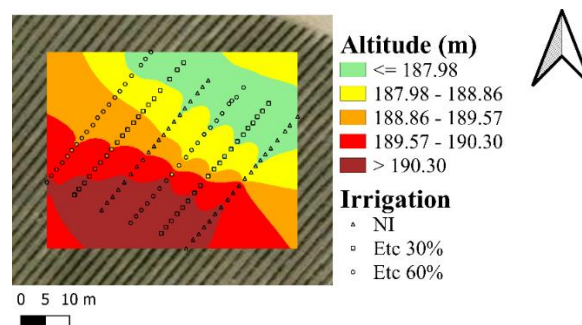


Figure 6-4. Test site characterisation regarding the irrigation treatments (NI: non-irrigated, Etc 30%, Etc 60%) and altitude.

The examination of altitude (Figure 6-4) and NDVI (Figure 6-5) revealed two distinct irrigation zones. The first zone, characterised by higher vigour, encompasses the three

lines from right to left, while the second zone, with lower vigour, comprises the remaining lines.

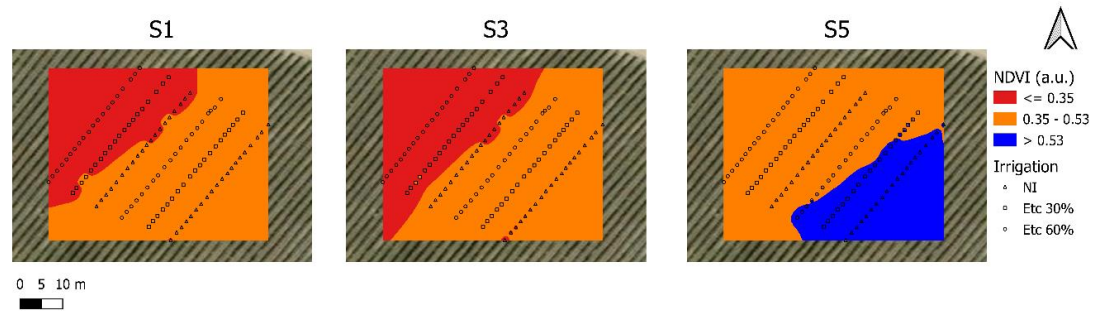


Figure 6-5. The normalised difference index (NDVI) based on Sentinel imagery mapped across the three assessment dates in 2021. S1 (05 days after veraison – DAV), S3 (19 DAV) and S5 (33 DAV).

In the 2021 growing season, vine vigour was assessed using NDVI (Figure 6-5). The focus on 2021 data was due to the availability of grapes not destructively assessed during this period, which was used to validate the tomography-like method in predicting internal grape tissues (SK, PU, and SE). This estimation allows for determining physicochemical parameters established in this study. Figure 6-5 also shows variations across specific assessment dates (S1, S3, and S5), representing distinct grape maturation stages. These dates were strategically selected to capture differences in vine vigour at different maturation stages, showing minimal change from S1 to S3 and increased vigour at S5 (Figure 6-5). The observed NDVI fluctuations indicate a dynamic response of grapevines to changing environmental conditions, underscoring the complex interaction between irrigation practices, terrain characteristics, and grapevine physiology. These maps contribute to a better understanding of vineyard dynamics and grape quality, which is essential for informed SSM in viticulture and winemaking practices. Detailed statistical analysis revealed significant differences in NDVI values among sampling dates, reflecting temporal variations in vine growth and development (Table S 6-1).

In 2022, PLWP exhibited elevated water stress levels compared to 2021 (Figure 6-6), indicating variations in environmental conditions between the two years. The observed pattern of PLWP concerning the vigour zones of the irrigation treatment remained consistent, with higher stress levels observed in the less vigorous zone within the irrigated lines. Conversely, in the non-irrigated lines, the most vigorous line exhibited higher water stress than the less vigorous zone, demonstrating the influence of both irrigation and vigour zone factors on the hydric status of the parcel. Detailed PLWP data for each assessment date for each year of the study is provided in the Supplementary Material (Table S 6-2).

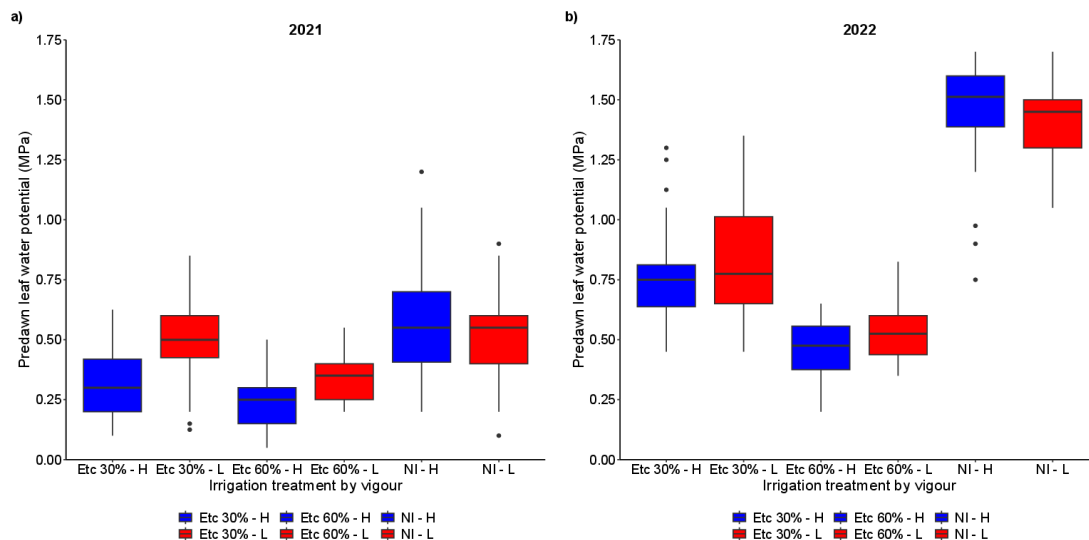


Figure 6-6. Mean of the predawn leaf water potential of the irrigation treatment (NI: non-irrigation, Etc 60% and Etc 30%), regarding the vigour zones of the parcel (H: high vigour, L: low vigour) in 2021 (a) and 2022 (b).

6.3.3. Tomography-like of grapes

The tomography-like method consists of spectral reconstruction of TO, which combines the spectra of SK, PU, and SE. On the other hand, the spectral decomposition involves reconstructing the SK, PU, and SE from the TO spectra (Figure 6-2).

Table 6-3 shows the statistical metrics of the spectra reconstructions of the grape's internal tissues (SK, PU, and SE) from TO and the decomposition of TO into SK, PU and SE in assessing the tomography-like method's performance over the two study years. The results demonstrate the efficacy of the tomography-like method for Touriga Nacional grapes in the Douro Wine Region across both years. In 2021, a high Pearson's correlation coefficient ($r \geq 0.97$) indicated a robust linear relationship between the predicted and observed values. The mean square error (MSE) for grape reconstruction and decomposition remained low (≤ 0.13), signifying accurate predictions. Similarly, the mean absolute percentage error (MAPE) was relatively minimal ($\leq 15.79\%$), indicating a close alignment between predicted and actual values. These results were statistically significant ($p < 0.001$), validating the reliability of the predictions. Conversely, in 2022, although the MSE (≤ 0.08) suggested continued accuracy in the predictions compared to 2021, the MAPE increased to 27.53%, indicating a higher error level. Nevertheless, Pearson's correlation coefficient remained high ($r \geq 0.95$), demonstrating a robust linear relationship between predicted and observed values. As in 2021, these results were statistically significant ($p < 0.001$), reinforcing the reliability of the predictions.

Table 6-3. Statistical metrics of the bi-directional spectral reconstruction of the grape's tissues in 2021 and 2022.

Grape/ Tissue	2021 (n= 191)				2022 (n=166)			
	MSE	MAPE (%)	r	p-value	MSE	MAPE (%)	r	p-value
Total	0.05	10.58	0.97	< 0.001	0.04	27.53	0.97	< 0.001
Skin	0.13	15.79	0.94	< 0.001	0.05	8.19	0.95	< 0.001
Pulp	0.01	2.28	1	< 0.001	0.06	7.46	0.97	< 0.001
Seed	0.05	8.56	0.97	< 0.001	0.08	12.33	0.95	< 0.001

MSE: Mean square error, MAPE: Mean absolute percentage error, r: Pearson's correlation coefficient.

6.3.4. Estimating the physicochemical variables based on the destructive dataset

Table 6-4 presents a comparative analysis of the performance of SL-AI and RF algorithms in predicting various physicochemical variables in the destructive dataset (n=357). The supplementary material presents the regression plot (Figure S 6-1 and S 2).

Table 6-4. Benchmark of self-learning artificial intelligence (SL-AI) and random forest (RF) in predicting the physicochemical variables of the destructive dataset (n=357) encompassing the data of 2021 (n=191) and 2022 (n=166).

Variable	SL-AI			RF		
	R ²	RMSE	MAPE	R ²	RMSE	MAPE
SSC (%)	0.99	1.93	7.51	0.99	1.89	7.12
PF (N)	0.97	0.57	14.42	0.94	0.79	27.34
VOL (cm ³)	0.95	0.34	19.76	0.86	0.50	41.20
W/VOL (g/cm ³)	0.99	0.11	6	0.98	0.18	8.45
CHL TO (a.u.)	0.87*	0.24	37.55	0.94*	0.14	22.34
CHL SK (a.u.)	0.98*	0.08	13.75	0.96*	0.12	14.55
CHL PU (a.u.)	0.99*	0.05	7.06	0.96*	0.14	20.39
CHL SE (a.u.)	1*	0.02	2.91	0.93*	0.17	24.97
ANT TO (a.u.)	0.94*	0.12	18.40	0.98*	0.09	8.98
ANT SK (a.u.)	1*	0.02	5.24	0.86*	0.14	33.83
ANT PU (a.u.)	1*	0.06	2.14	1*	0.07	4.50
ANT SE (a.u.)	0.99*	0.09	6.39	1*	0.04	2.61

R²: coefficient of determination, RMSE: root mean square error, MAPE: mean absolute percentage error. *means measured through the spectra.

Table 6-4 presents the performance of the SL-AI and RF algorithms in predicting the physicochemical parameters of grapes and tissues. Both algorithms exhibit high predictive performance, supported by high R² values, low RMSE and MAPE for most physicochemical variables. However, SL-AI outperformed RF, demonstrating superior statistical metrics with R² ≥ 0.94 and MAPE ≤ 37.55%, whereas RF yielded R² ≥ 0.86 and MAPE ≤ 41.20%.

6.3.5. Predicting and mapping the physicochemical variables on the non-destructive dataset

The knowledge base established through the destructive measurement of grapes (n=357) during the tomography-like methodology served as a foundation for predicting the spectral characteristics of the grape SK, PU, and SE for grapes that were not destructively measured (2021). These predictions were used to estimate the physicochemical parameters measured in this study using both the SL-AI and RF algorithms. However, despite SL-AI demonstrating superior statistical metrics compared to RF in estimating reconstructed spectra (Table 6-4), it exhibited inconsistent values for the physicochemical variables across grape maturation stages (Table S 6-3). Conversely, RF (Table S 6-4) produced more realistic values reflecting the behaviour of grape maturation.

The grape maturation was showcased by selecting the following DAVs: 19 DAV (S3), 26 DAV (S4) and 40 DAV (S6). The 19 and 26 DAVs were selected to exemplify the changes that occur during the mid-season of grape maturation. In contrast, the S6 was selected

to demonstrate the physicochemical parameters of the grapes at the harvest date defined according to the vinification objectives of the company responsible for the vineyard under study.

This study chooses CHL and ANT as representative indicators of temporal evolution during grape maturation. This decision stems from the spectral measurements obtained from SK, PU and SE, which provide perceptions of physiological and biochemical changes occurring during maturation. CHL content is a proxy for photosynthetic activity, while ANT levels reflect colour development, both alterations throughout grape maturation. By focusing on these two parameters, the study aims to capture the dynamic changes in grape physiology and biochemistry across different stages of maturation. This approach allows for a concise yet informative representation of grape maturation dynamics, facilitating a better understanding of the processes involved and their implications for SSM and winemaking practices.

6.3.5.1. The dynamic of chlorophyll content in the skin, pulp and seed

The dynamics of CHL content in SK, PU, and SE indicate the maturation process of grapes, displaying a consistent downward trend, as shown in Figure 6-7 and complemented by Table S 6-4 (Supplementary Material). Specifically, CHL content in TO gradually decreases as grapes mature, reflecting CHL degradation. CHL content in SK continually decreases throughout maturation, particularly between assessment dates S3 and S4. Variations in CHL content are observed among different zones of vineyard vigour and irrigation lines (Figure 6-7), with areas of lower vigour and reduced irrigation typically associated with lower CHL levels. However, by assessment date S6, a different pattern emerges, where higher altitude zones tend to display elevated CHL content in SK, although with exceptions in specific zones subject to irrigation treatments.

In the case of CHL content in PU, a relatively uniform distribution is observed during S3, with exceptions in higher altitude areas of less vigorous zones under specific irrigation treatments (Figure 6-7). However, by S4, CHL levels in PU remain generally high, with decreases observed in specific areas, including less vigorous zones and certain altitudes within different irrigation treatments. Subsequently, in S6, CHL levels in PU decreased across the parcel, with higher concentrations observed in more vigorous zones than less vigorous ones (Figure 6-7).

While CHL content in SE remains relatively stable across assessment dates, no distinct separation based on vineyard vigour or irrigation treatments is observed at S3 (Figure 6-7). However, certain zones exhibit lower CHL accumulation in SE. Nonetheless, by S4,

a separation based on terrain topography becomes apparent, with higher altitudes displaying increased CHL levels in the seed (Figure 6-7). Specific irrigation treatments in less vigorous zones at S4 show lower CHL in the seed, indicating a more advanced physiological grape maturity. By S6, higher vigour zones tend to exhibit lower CHL levels in the seed, particularly in irrigated lines at higher altitudes. In contrast, less vigorous zones tend to display higher CHL concentrations, especially in lower altitudes of irrigated lines.

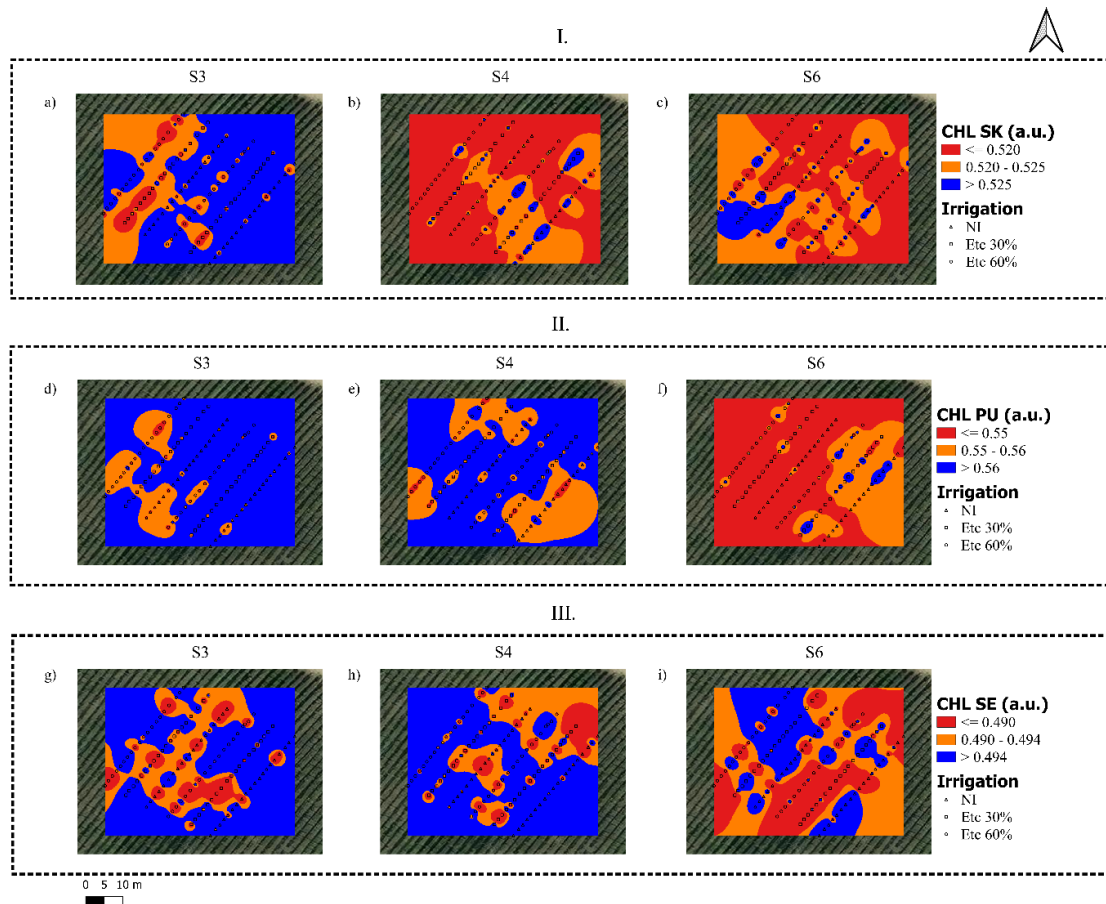


Figure 6-7. Chlorophyll content in the skin (CHL SK – I. a), b), and c)), in the pulp (CHL PU – II. d), e) and f)), and the seed (CHL SE – III. G), h), i)) maps for Touriga Nacional cultivar across three post-veraison sampling dates (S3, S4 and S6) in 2021 to represent the prediction of the grapes that were measured using only spectral data. S3 (19 days after veraison (DAV), a), d) and g)), S4 (26 DAV, b), e) and h)) and S6 (40 DAV, c), f) and i)). NI: non-irrigated, 30% Etc: 30% of evapotranspiration and 60% Etc: 60%.

6.3.5.2. The dynamic of anthocyanin content in the skin, pulp and seed

The dynamics of ANT content across grape tissues present a complex pattern, with variations observed between TO and SK content assessment dates, as indicated in Figure 6-8 and Table S 6-3, suggesting dynamic shifts in pigmentation during maturation. Specifically, ANT in SK exhibits higher concentrations in the less vigorous zones of the

parcel during S3, pronounced in higher altitude areas subject to irrigation treatments (Figure 6-8). This trend persists into S4, with elevated ANT content observed in the less vigorous zone, particularly in higher altitude regions within irrigated lines. However, by S6, an increase in ANT in SK is noted throughout the parcel, although with lower values observed in higher altitude zones, while remaining high in more vigorous zones and irrigated lines.

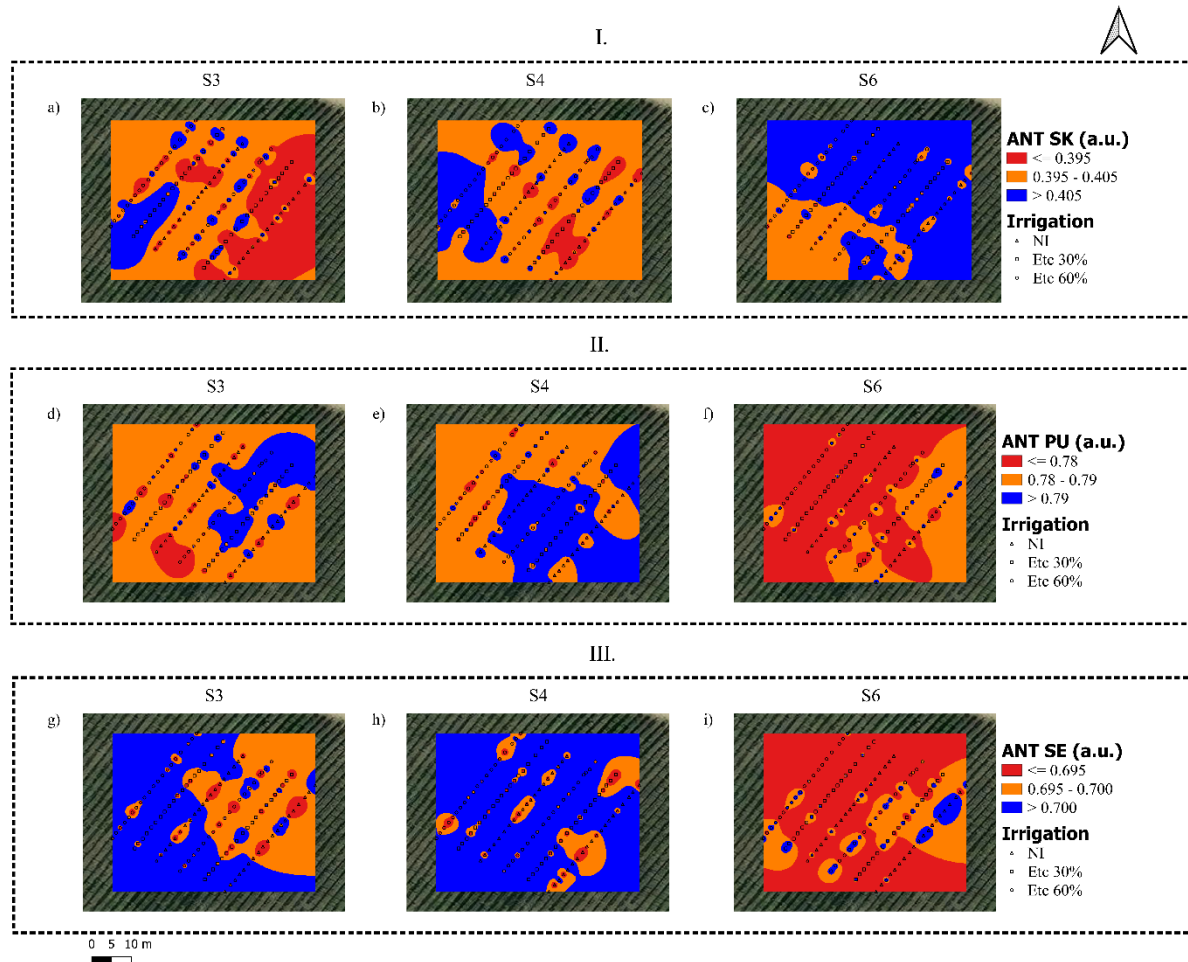


Figure 6-8. Anthocyanin content in the skin (ANT SK – I. a), b), and c)), in the pulp (ANT PU – II. d), e) and f)), and the seed (ANT SE – III. G), h), i)) maps for Touriga Nacional cultivar across three post-veraison sampling dates (S3, S4 and S6) in 2021 to represent the prediction of the grapes that were measured using only spectral data. S3 (19 days after veraison (DAV), a), d) and g)), S4 (26 DAV, b), e) and h)) and S6 (40 DAV, c), f) and i)). NI: non-irrigated, 30% Etc: 30% of evapotranspiration and 60% Etc: 60%.

Regarding ANT in PU, higher values are observed in the irrigated lines of the more vigorous zones during S3 (Figure 6-8), maintaining this pattern at S4, where the most vigorous zone exhibits elevated ANT in PU values, particularly within irrigated lines in higher altitude regions. However, ANT in PU content experiences a decrease across the parcel by S6, with higher levels remaining predominantly in the more vigorous zones, especially within lines subjected to specific irrigation treatments. This decrease in ANT

in PU towards the end of maturation may be related to the balance between the synthesis and degradation of organic acids during this phase, which is also highly dependent on meteorological conditions.

ANT in SE, on the other hand, is higher in high altitudes and irrigated lines within the less vigorous part of the parcel during S3. However, by S4, a homogenisation of the parcel is observed, with specific zones displaying lower ANT levels, particularly in non-irrigated lines at higher altitudes. Nevertheless, at S6, like the trend observed in PU, the ANT levels decrease in the entire parcel, with higher levels concentrated in the lower altitude regions of the more vigorous part of the parcel.

6.3.5.3. Assessment of physicochemical parameters to harvest

Figure 6-9 provides an overview of the assessment of various physicochemical parameters for TO and SK, PU, and SE at 40 DAV, representing the harvest date decided for the studied commercial vineyard. SSC demonstrates a consistent time increase, reflecting the gradual accumulation of sugars in the grapes as they ripen, as shown in Table S 6-3. At this stage, the higher SSC values are observed in areas of lower vineyard vigour, particularly at lower altitudes. At this date, irrigated areas coincide with higher SSC values within the same areas of high NDVI values, especially in the parcel's vigorous zone.

Conversely, PF continually decreases throughout the same period, indicating a softening of grape texture as ripening progresses (Table S 6-3). The PF pattern in S6 aligns with SSC values, with elevated levels observed in irrigated lines within less vigorous zones, including those at lower altitudes in non-irrigated lines (Figure 6-9). Similarly, volume measurements display higher values in irrigated lines within less vigorous zones, consistent with PF trends. Volume demonstrates a constant decline as grapes mature, contrasting with the gradual increase in weight per volume (W/VOL) over the same period (Figure 6-9, Table S 6-3). However, W/VOL values are higher in the most vigorous zone, particularly in non-irrigated and Etc 30% lines (Figure 6-9).

At S6, Figure 6-9 shows distinct patterns in pigment distribution across the SK, PU, and SE, considering factors such as altitude, vigour, and irrigation treatments. Specifically, CHL levels in the SK indicate higher values at higher altitudes, particularly in the less vigorous sections of the parcel. Conversely, the PU exhibits a pattern closely correlated with NDVI, with higher CHL concentrations observed in the most vigorous areas, particularly at lower altitudes. Curiously, the SE follows a similar pattern to NDVI but shows an inverse relationship with CHL in the PU. While the lower altitude zones of the

less vigorous sections exhibit high CHL values, the higher altitude zones of the irrigated lines in the most vigorous sections display lower CHL levels.

At S6, ANT levels exhibit distinct patterns across different grape tissues. In SK, high values are consistently observed at lower altitudes, regardless of the parcel's vigour. Conversely, in PU, the pattern mirrors the NDVI values, with the most vigorous areas of the parcel showing higher ANT concentrations, particularly in plots treated with Etc 30% irrigation. Similarly, SE follows this trend, with higher concentrations also found in the most vigorous parts of the parcel, particularly in areas treated with Etc 30% irrigation.

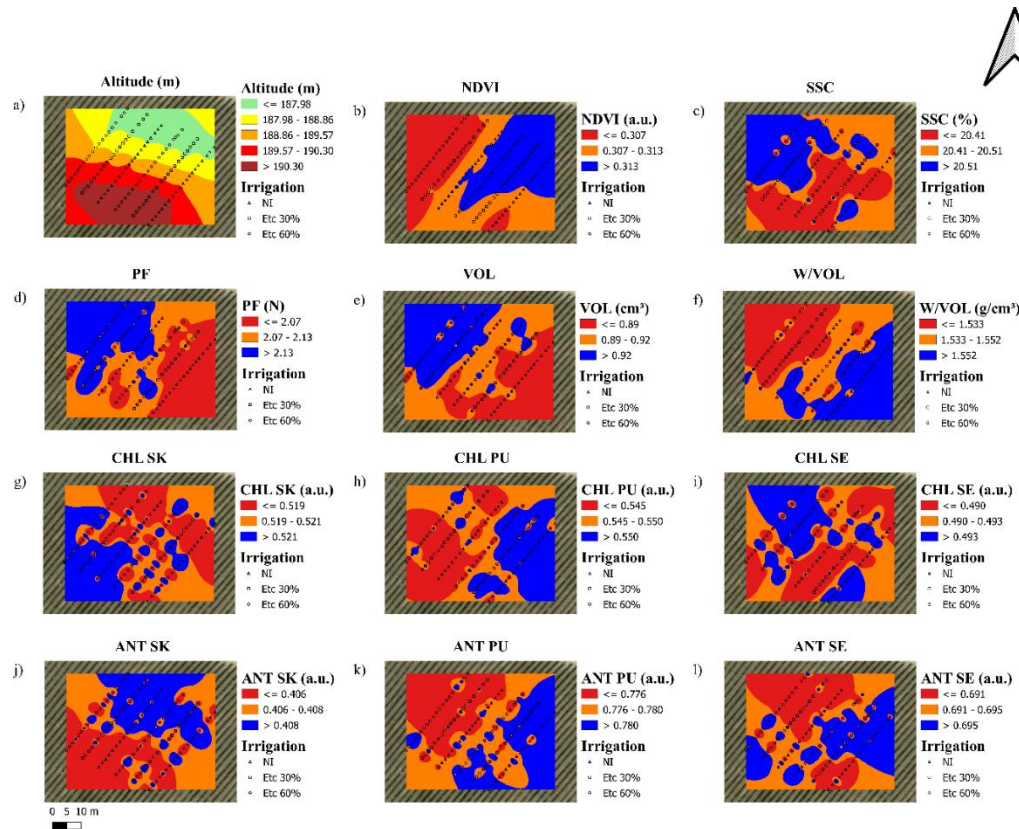


Figure 6-9. Altitude a), normalised difference vegetation index (NDVI - b), solid soluble content (SSC - c), puncture force (PF - d), volume (vol - e), weight per volume (W/VOL - f), chlorophyll in the skin (CHL SK - g), chlorophyll in the pulp (CHL PU - h), chlorophyll in the seed (CHL SE - i), anthocyanin in the skin (ANT SK - j), anthocyanin in the pulp (ANT PU - k), anthocyanin in the seed (ANT SE - l) at 40 days after veraison. NI: non-irrigated, 30% Etc: 30% of evapotranspiration and 60% Etc: 60%.

6.3.5.4. Site-specific management

Advanced grape maturity indicators are crucial in viticulture to attain specific winemaking objectives during grape maturation, which involve integrating various physicochemical grape parameters to define SSM. For instance, winemakers may seek to enhance the ANT and SCC concentration in grapes at harvest while minimising CHL. However, many other combinations of physicochemical parameters of grapes can be established

according to the quality objectives of a particular type of wine or even according to objectives for the commercialisation of table grapes (e.g., PF, volume).

Therefore, this study examines a range of physicochemical parameters to establish distinct goals for grape maturation, which can serve to improve irrigation management and harvest timing according to winemaking quality objectives. These maps can also be utilised with other variables, bridging laboratory analyses and field applications, thus holding agronomic relevance. Leveraging the data presented in Figure 6-9 at 40 DAV, three decision maps were formulated to delineate different harvest objectives, as shown in Figure 6-10.

Model 1 (Figure 6-10 a) assigns equal importance to SSC, ANT in SK, and minimal CHL content in PU. This model promotes higher values, particularly in the lower altitude regions and the irrigated lines of the parcel.

Model 2 (Figure 6-10 b), which prioritises SSC over ANT in SK, presents a more balanced distribution across the vigour zones of the parcel. According to this model, the less vigorous regions suggest higher values, particularly in the irrigated lines from lower to higher altitudes. Conversely, the higher vigour zones of the parcel exhibit elevated values in the lower altitude regions, specifically in NI and Etc 30% lines.

Lastly, model 3 (Figure 6-10 c) prioritises higher values of ANT in SK and SSC. This decision criterion leads to a more pronounced division based on altitude within the parcel. The lower altitude regions across all irrigation treatments and vigour zones exhibit higher values compared to the higher altitude regions.

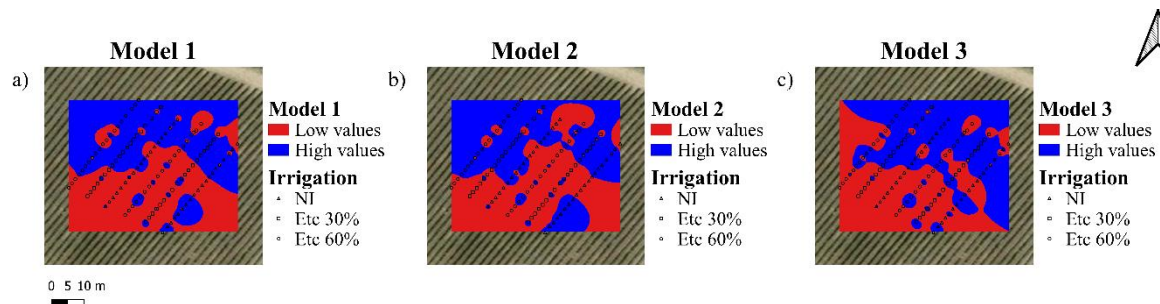


Figure 6-10. Decision maps regarding different goals in the maturation. Model 1 stands for no weight in the soluble solids content (SSC), anthocyanin in the skin (ANT SK) and chlorophyll in the pulp (CHL PU). Model 2 attributes 60% to SSC, 30 % to ANT SK, and 10 % to CHL PU, and Model 3 stands for 30% SSC, 60 % to ANT SK and 10 for CHL PU.

6.4. Discussion

The topographic characteristics of vineyards influence grapevine vigour and grape characteristics. Understanding these dynamics helps to delineate specific zones within vineyards to tailor SSM, such as irrigation management and harvest timing, thereby optimising various parameters according to specific quality objectives. In this study, irrigation was a key factor considered. This information can potentially refine irrigation

schemes, ensuring their optimal utilisation. Moreover, the establishment of decision maps, reflected through the SSM, aids in determining the most suitable irrigation strategies for different vineyard zones, whether it involves adjusting irrigation levels in higher altitude areas or suppressing irrigation in lower regions.

The tomography-like method (Martins et al., 2023) that was previously demonstrated on grapes (Tosin et al., 2023b) and tomatoes (Tosin et al., 2024a) also showed efficacy for reconstructing internal grape tissues in the grape cultivar Touriga Nacional in the Douro Wine Region, considering two data years and different irrigation treatments. This paper successfully demonstrates how environmental factors influence grape maturation within the vineyard parcel using the tomography-like method. It highlights variations in topography and vigour across the parcel and constructs metabolic maps illustrating physicochemical parameters of TO, SK, PU, and SE.

The tomography-like method employs non-invasive imaging to characterise grape physicochemical parameters, facilitating the assessment of grape maturation. In contrast to other techniques, such as electrical impedance (e.g., Verma et al., 2021), which assess internal physical damages of fruits, the tomography-like method offers a detailed prediction of inner tissues without causing potential unknown damages associated with X-ray use (e.g., Liu et al., 2020a). Additionally, by predicting non-destructively the inner tissues of fruits, the tomography-like can improve viticultural practices (e.g., irrigation, fertilisation), allowing for better SSM according to the objectives of winemaking or table grapes.

Analysing metabolic maps (Figure 6-7, 8 and 9) alongside terrain analysis (Figure 6-4 and 5) elucidates variations in grape maturation dynamics across different NDVI-based vigour zones within the vineyard (Al-Saddik et al., 2017). On the other hand, changes in SSC dynamics provide perceptions of maturation trajectories, revealing variations in grapevine firmness (Xu et al., 2022). These fluctuations underscore the relationship between turgor pressure and grape maturation, where higher PF values suggest firmer grapes (Xu et al., 2022; Sohaib Ali Shah et al., 2020).

Assessment of CHL content provides perceptions of the maturation dynamics of grapes, revealing distinct patterns in non-irrigated zones (Figure 6-7). There is a reduction in photosynthetic activity throughout grape maturation, leading to diminished CHL synthesis and stability across grape tissues as energy is redirected towards fruit development and maturation (Li et al., 2022). Enzymatic degradation of CHL, catalysed by chlorophyllase and other enzymes, further contributes to the recycling of essential compounds for photosynthesis into alternate metabolic pathways (Li et al., 2024; Garrido et al., 2023). Fluctuations in CHL content can also signal genetic and environmental

stressors such as nutrient deficiencies, heat stress, or water scarcity, enabling vineyard managers to implement proactive measures to optimise grape production (Medda et al., 2022). Specifically, the decline in CHL content in SK and PU during maturation primarily results from reduced photosynthetic activity and CHL degradation (Figure 6-7). The decrease in CHL in SK and PU is influenced by environmental factors such as higher temperatures and intense sunlight, which accelerate degradation processes (Rousserie et al., 2019; Ristic & Iland, 2005). Moreover, during maturation, CHL content in SE serves as an indicator of the vine's photosynthetic capacity and overall health, offering an understanding of its energy production and growth efficiency and accumulating reserves like sugars and lipids to support subsequent growth, necessitating resource reallocation and potentially limiting CHL synthesis and maintenance (Garrido et al., 2023). CHL increased in SE at S6 during grape maturation, particularly at lower altitudes in the irrigated lines of less vigorous zones within the vineyard parcel (Figure 6-7). This dynamic may be attributed to resource reallocation under stress conditions, such as water deficit (Figure 6-6), prompting compensatory responses in specific tissues (e.g., SE).

During grape maturation, ANT and carotenoids accumulate, enhancing the characteristic colour of ripe grapes. Optimal maturation conditions redirect resources from CHL synthesis towards pigment production, favouring ANT accumulation (Massonnet et al., 2017). However, high temperatures and intense sunlight can disrupt CHL stability, accelerating its degradation (Rogiers et al., 2022). This disruption may lead to decreased ANT levels in specific tissues, such as PU and SE, which is particularly evident near harvest (S6). This decline in ANT content could be attributed to increased CHL breakdown under stressful conditions like high temperatures or water scarcity (Rogiers et al., 2022; Massonnet et al., 2017).

Using metabolic maps to make agronomic decisions is paramount for optimising grape maturation and achieving specific winemaking goals. The results presented in this study underscore the importance of metabolic mapping in delineating distinct objectives and management zones for grape maturation, considering various physicochemical parameters. By integrating data from Figure 6-9 at 40 DAV, three decision maps were formulated to guide harvest goals, as illustrated in Figure 6-10, thus providing SSM. These models prioritise different variables based on the desired wine characteristics. For instance, Model 1 (Figure 6-10 a) is particularly suitable for wines where a harmonious combination of sugar content and pigment concentration is desired.

Conversely, Model 2 (Figure 6-10 b) prioritises higher SSC while placing less emphasis on anthocyanin content, resulting in a more even distribution across different vineyard

vigour zones. This approach benefits wines where sugar content takes precedence over pigment quantification. Finally, Model 3 (Figure 6-10 c) focuses on maximising ANT content in the SK and SSC, which is ideal for wines where pigment intensity is the primary characteristic required, such as those aiming for rich colour profiles. Moreover, if the production goal were table grapes, the emphasis would likely be on factors such as SSC and PF. These decision maps serve as instrumental tools for vineyard management, enabling growers to tailor their cultivation practices to achieve specific quality targets tailored to the desired wine or grape product.

While this study presents a spatiotemporal dynamic of maturation in SK, PU and SE through a dry-lab approach, its limitations and the necessity for further validation and refinement of the tomography-like method across various grape cultivars and environmental conditions are acknowledged. This study did not validate the results concerning pigments, and the discovery of other compounds (e.g., rotundone) may contribute to describing wine quality in wet-lab settings (e.g., Bramley et al., 2017). Therefore, future research endeavours should integrate advanced techniques, such as genomic and metabolomic analyses, to deepen the comprehension of grape physiology and enhance the predictive capabilities of this approach, which will enable the development of more robust SSM for winemakers and grape growers, tailored to specific production objectives and viticultural traits (e.g., irrigation, fertilisation), while also opening new avenues for research in systems biology as the spectral data obtained from this study holds promise for predicting essential metabolites, thereby enriching this research domain (Pazhamala et al., 2021).

6.5. Conclusion

This study shows the efficacy of the tomography-like method in reconstructing internal grape tissues, exemplified in the Touriga Nacional cultivar within the Douro Wine Region. By integrating spectral data and terrain analysis, the study sheds light on the complex dynamics of grape maturation and offers insights into optimising vineyard management practices. In regions like this, as in others worldwide, vineyards exhibit variability in terrain characteristics, influencing vine vigour and maturation parameters. This variability underscores the importance of employing advanced methodologies like the tomography-like approach to understand and address the diverse challenges of grape cultivation.

The results demonstrate the method's ability to delineate physicochemical variations in the total grape and internal tissues across different vineyard zones, providing a basis for targeted interventions to enhance grape quality and wine production efficiency. The study showcases the method's capacity to predict physicochemical parameters non-

destructively, thus facilitating informed decision-making in viticulture. While acknowledging limitations such as the need for further validation across different grape cultivars and environments, the study's original contribution lies in its practical application of advanced techniques to address pressing challenges in vineyard management. The study offers a novel approach to optimising grape maturation and improving wine quality by providing decision maps tailored to specific wine quality objectives.

Future research endeavours should focus on refining the predictive capabilities of the tomography-like method through integrating genomic and metabolomic analyses. Such advancements promise to improve vineyard management practices and elevate wine production standards.

Acknowledgements

The authors also thank the wine company Real Companhia Velha (and its Coordinator for Viticulture Rui Soares) for the facilities provided for fieldwork, as well as the Associação para o Desenvolvimento da Viticultura Duriense (ADVID) for the measurement of the predawn leaf water potential.

Supplementary material

Supplementary figures

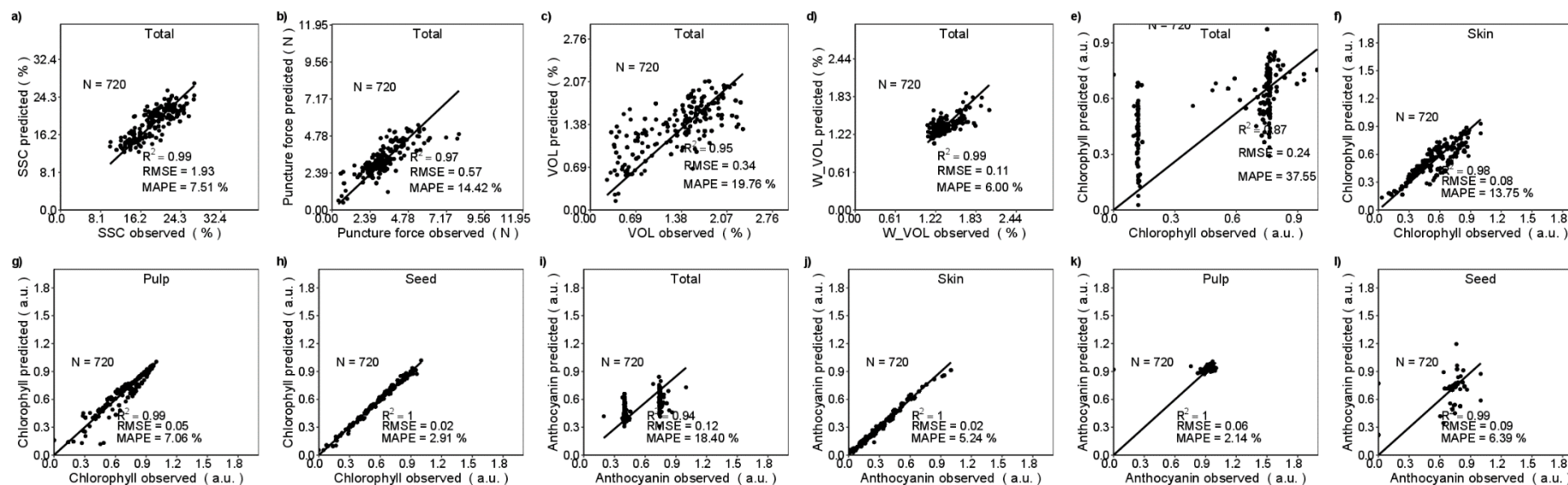


Figure S 6-1. Regression plots by applying self-learning artificial intelligence (SL-AI) in estimating: a) Soluble solid content (SSC - %), b) Puncture force (N), c) Volume (cm³), d) Weight/volume (g/cm³), e) Chlorophyll (a.u.) total in the grape, f) Chlorophyll (a.u.) in the skin, g) Chlorophyll (a.u.) in the pulp, h) Chlorophyll (a.u.) in the seed, i) Anthocyanin (a.u.) total in the grape, j) Anthocyanin (a.u.) in the skin, k) Anthocyanin (a.u.) in the pulp, and l) Anthocyanin (a.u.) in the seed.

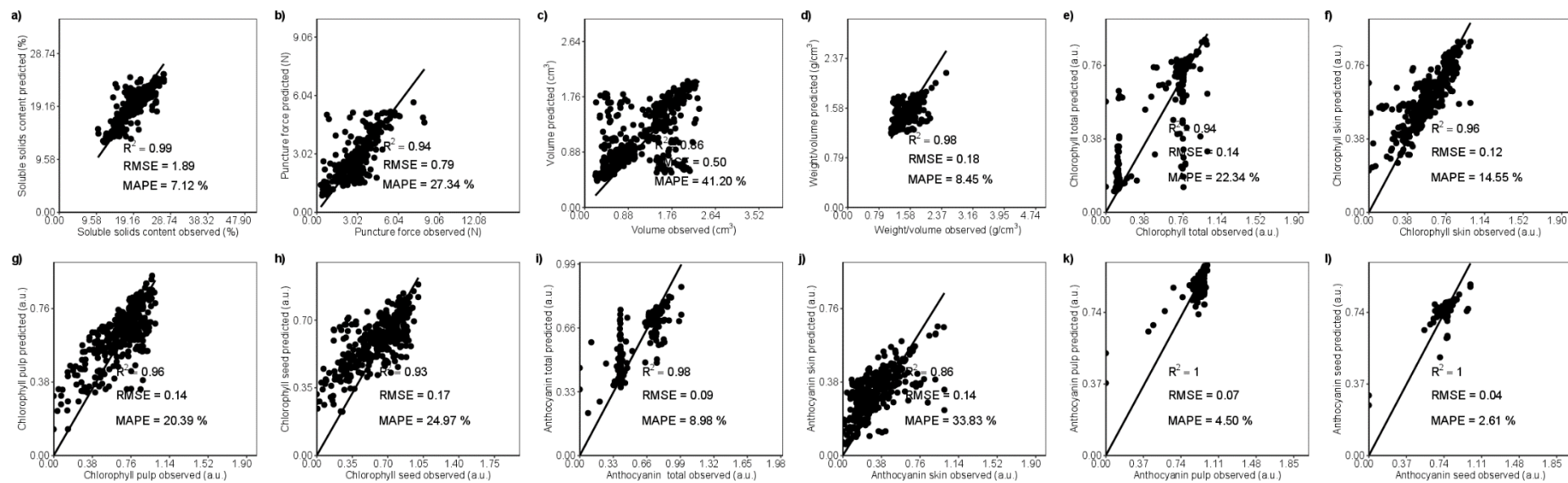


Figure S 6-2. Regression plots by applying random forest (RF) in estimating: a) Soluble solid content (SSC - %), b) Puncture force (N), c) Volume (cm³), d) Weight/volume (g/cm³), e) Chlorophyll (a.u.) total in the grape, f) Chlorophyll (a.u.) in the skin, g) Chlorophyll (a.u.) in the pulp, h) Chlorophyll (a.u.) in the seed, i) Anthocyanin (a.u.) total in the grape, j) Anthocyanin (a.u.) in the skin, k) Anthocyanin (a.u.) in the pulp, and l) Anthocyanin (a.u.) in the seed.

Supplementary tables

Table S 6-1 provides the statistical mean values of the normalised difference index (NDVI) for the year 2021 across six sampling dates: S1 (29/07/2021 – 05 days after veraison (DAV)), S2 (05/08/2021 – 12 DAV), S3 (12/08/2021 – 19 DAV), S4 (19/08/2021 – 26 DAV), S5 (26/08/2021 – 33 DAV), and S6 (02/09/2021 – 40 DAV). The table includes NDVI values for three irrigation treatments: non-irrigated (NI), irrigation to replace 30% of evapotranspiration (Etc 30%), and irrigation to replace 60% of evapotranspiration (Etc 60%). The p-values indicate the level of significance for each treatment and sampling date. Statistical significance, denoted by p-value < 0.05, signifies significant mean differences based on Duncan's post-hoc test ($\alpha = 5\%$). Differences within rows are denoted by lowercase letters, and differences within columns are denoted by uppercase letters, according to Duncan's post-hoc test.

Table S 6-1. The statistical mean value of the normalised difference index (NDVI) for 2021 in the six sampling dates S1 (29/07/2021 – 05 days after veraison (DAV)), S2 (05/08/2021 – 12 DAV), S3 (12/08/2021 – 19 DAV), S4 (19/08/2021 – 26 DAV), S5 (26/08/2021 – 33 DAV), and S6 (02/09/2021 – 40 DAV).

NDVI	S1	S2	S3	S4	S5	S6	p-Value
NI	0.36cA $\pm$ 0.03	0.55aA $\pm$ 0.04	0.35cA $\pm$ 0.02	0.55aA $\pm$ 0.04	0.53bA $\pm$ 0.05	0.31dA $\pm$ 0.02	< 0.001
30% Etc	0.36bA $\pm$ 0.04	0.55aA $\pm$ 0.07	0.35bA $\pm$ 0.04	0.55aA $\pm$ 0.07	0.53aA $\pm$ 0.05	0.31cA $\pm$ 0.02	< 0.001
60 Etc	0.35cA $\pm$ 0.05	0.53aA $\pm$ 0.08	0.35cA $\pm$ 0.05	0.54aA $\pm$ 0.08	0.50bB $\pm$ 0.03	0.31dA $\pm$ 0.03	< 0.001
P-value	0.707	0.449	0.907	0.559	0.014	0.751	
Field	0.36c $\pm$ 0.04	0.54a $\pm$ 0.06	0.35c $\pm$ 0.04	0.55a $\pm$ 0.07	0.52b $\pm$ 0.05	0.31d $\pm$ 0.02	< 0.001

Means with a p-value < 0.05 indicate statistical significance. In cases where the p-value is < 0.05, differing lowercase letters within rows signify significant mean differences based on Duncan's post-hoc test ($\alpha = 5\%$), while differing uppercase letters within columns signify significant mean differences according to Duncan's post-hoc test ($\alpha = 5\%$).

Table S 6-2 presents the predawn leaf water potential (PLWP - MPa) across different irrigation treatments over two consecutive years, 2021 and 2022. The irrigation treatments, including non-irrigated and two irrigation levels (30% and 60% Etc), were evaluated for their impact on PLWP. Statistical analysis, including one-way ANOVA and Fischer's post-hoc Duncan test ($\alpha = 5\%$), was conducted to compare yearly PLWP values among irrigation treatments.

Table S 6-2. Comparison of predawn leaf water potential (PLWP - MPa) across irrigation treatments in 2021 and 2022 in the six post-veraison dates. For 2021, S1 (29/07/2021 – 05 days after veraison (DAV)), S2 (05/08/2021 – 12 DAV), S3 (12/08/2021 – 19 DAV), S4 (19/08/2021 – 26 DAV), S5 (26/08/2021 – 33 DAV), and S6 (02/09/2021 – 40 DAV) and for 2022: S1 (28/07/2022 – 03 DAV), S2 (04/08/2022 – 17 DAV), S3 (11/08/2023 – 24 DAV), S4 (18/08/2022 – 31 DAV), S5 (25/08/2022 – 31 DAV), and S6 (01/09/2022 – 38 DAV).

Evaluation date	Non-irrigated	30% Etc	60% Etc	p-value
2021				
S1	0.47 ± 0.15bA	0.30 ± 0.14aB	0.36 ± 0.17bcB	<0.001
S2	0.51 ± 0.23bA	0.26 ± 0.09aB	0.34 ± 0.15bcB	0.003
S3	0.68 ± 0.19aA	0.29 ± 0.09aC	0.45 ± 0.12bB	<0.001
S4	0.69 ± 0.16aA	0.33 ± 0.10aB	0.59 ± .0.17aA	<0.001
S5	0.67 ± 0.11aA	0.37 ± 0.11aB	0.45 ± 0.12bB	<0.001
S6	0.31 ± 0.11cA	0.34 ± 0.37aA	0.24 ± 0.13cA	0.548
Mean	0.54A*	0.31C*	0.40B*	<0.001
p-value	<0.001	0.67	<0.001	
2022				
S1	1.26 ± 0.24bA	0.47± 0.12bB	0.56 ± 0.07dB	<0.001
S2	1.29 ± 0.19bA	0.58 ± 0.02aC	1.10 ± 0.24aB	<0.001
S3	1.48 ± 0.12aA	0.65 ± 0.13aB	0.73 ± 0.09cB	<0.001
S4	1.49 ± 0.13aA	0.43 ± 0.09bC	0.72 ± 0.10cB	<0.001
S5	1.56 ± 0.10aA	0.45 ± 0.08bC	0.75 ± 0.15cB	<0.001
S6	1.54 ± 0.13aA	0.39 ± 0.12bC	0.96 ± 0.16bB	<0.001
Mean	1.44A*	0.49C*	0.80B*	<0.001
p-value	<0.001	<0.001	<0.001	
Means with a p-value < 0.05 indicate statistical significance. In cases where the p-value is < 0.05, differing lowercase letters within rows signify significant mean differences based on Duncan's post-hoc test ($\alpha = 5\%$), while differing uppercase letters within columns signify significant mean differences according to Duncan's post-hoc test ($\alpha = 5\%$). * indicates there are differences among years ($p<0.001$).				

Table S 6-3 presents the results of the SL-AI algorithm in predicting various physicochemical parameters of grapes and tissues based on the non-destructive dataset (n=720). The table includes data for soluble solids content (SSC), puncture force (PF), volume (VOL), weight per volume (W/VOL), anthocyanin content in the total grape (ANT-TO), skin (ANT-SK), pulp (ANT-PU), and seeds (ANT-SE), as well as chlorophyll content in the total grape (CHL-TO), skin (CHL-SK), pulp (CHL-PU), and seeds (CHL-SE). Each row corresponds to a specific parameter, and each column represents one of the six sampling dates (S1 to S6). The values provided in the table are means with standard deviations, and the significance of differences among the sampling dates is indicated by different letters based on Duncan's post hoc test ($\alpha < 5\%$).

Table S 6-3. Results of SL-AI in predicting soluble solids content (SSC - %), puncture force (PF - N), Volume (VOL - cm³), weight per volume (W/VOL - g/cm³), anthocyanin content in the total grape (ANT-TO - a.u.), skin (ANT-SK - a.u.), pulp (ANT-PU - a.u.) and seed (ANT-SE - a.u.), chlorophyll content in the total grape (CHL-TO - a.u.), skin (CHL-SK - a.u.), pulp (CHL-PU - a.u.) and seed (CHL-SE - a.u.) and the predawn leaf water potential (MPa). Based on the non-destructive dataset (n= 720).

	S1	S2	S3	S4	S5	S6	p-Value
SSC	19.55 ^{cd} ± 2.60	20.69 ^{ab} ± 1.49	20.07 ^{bc} ± 3.90	19.69 ^{cd} ± 2.50	21.23 ^a ± 1.43	19.06 ^d ± 2.97	< 0.001
PF	3.12 ^b ± 0.39	3.04 ^b ± 0.84	2.74 ^c ± 0.33	2.89 ^c ± 0.46	2.85 ^c ± 0.17	3.98 ^a ± 0.85	< 0.001
VOL	1.44 ^b ± 0.48	1.65 ^a ± 0.28	1.43 ^b ± 0.35	1.36 ^{bc} ± 0.35	1.29 ^c ± 0.39	1.59 ^a ± 0.27	< 0.001
W/VOL	1.32 ^{bc} ± 0.08	1.31 ^c ± 0.06	1.35 ^a ± 0.10	1.35 ^{ab} ± 0.09	1.31 ^c ± 0.11	1.35 ^a ± 0.12	< 0.001
ANT-TO	0.54 ^a ± 0.08	0.50 ^b ± 0.07	0.48 ^c ± 0.08	0.47 ^c ± 0.11	0.46 ^c ± 0.07	0.46 ^c ± 0.04	< 0.001
ANT-SK	0.15 ^b ± 0.09	0.34 ^b ± 0.05	0.41 ^{bc} ± 0.14	0.58 ^c ± 0.14	0.32 ^c ± 0.16	0.17 ^{bc} ± 0.05	< 0.001
ANT-PU	0.93 ^b ± 0.03	0.93 ^b ± 0.03	0.95 ^c ± 0.03	0.96 ^a ± 0.01	0.97 ^d ± 0.01	0.97 ^e ± 0.00	< 0.001
ANT-SE	0.77 ^b ± 0.01	0.76 ^d ± 0.01	0.76 ^c ± 0.01	0.77 ^e ± 0.00	0.75 ^a ± 0.02	0.62 ^c ± 0.22	< 0.001
CHL-TO	0.57 ^c ± 0.16	0.59 ^c ± 0.16	0.64 ^{ab} ± 0.09	0.51 ^d ± 0.24	0.60 ^{bc} ± 0.15	0.65 ^a ± 0.08	< 0.001
CHL-SK	0.60 ^a ± 0.19	0.52 ^b ± 0.16	0.48 ^b ± 0.13	0.47 ^b ± 0.14	0.47 ^b ± 0.15	0.48 ^b ± 0.13	< 0.001
CHL-PU	0.80 ^{ab} ± 0.05	0.78 ^a ± 0.05	0.72 ^b ± 0.03	0.84 ^{cd} ± 0.09	0.63 ^{cd} ± 0.28	0.50 ^d ± 0.10	< 0.001
CHL-SE	0.72 ^a ± 0.04	0.51 ^{bc} ± 0.18	0.57 ^b ± 0.13	0.44 ^c ± 0.30	0.83 ^c ± 0.05	0.59 ^c ± 0.15	< 0.001

The same line (among dates) and labelled with different letters mean the results are significantly different according to Duncan's post hoc test ($\alpha < 5\%$). SSC: soluble solids content (%); PF: puncture force (N); VOL: volume (cm³); weight per volume (g/cm³), ANT-TO: anthocyanin in the total grape (a.u.); ANT-SK: anthocyanin in the skin (a.u.); ANT-PU: anthocyanin in the pulp (a.u.); ANT-SE: anthocyanin in the seeds (a.u.); CHL-TO: chlorophyll in the total grape; CHL-SK: chlorophyll in the skin; CHL-PU: chlorophyll in the pulp; CHL-SE: chlorophyll in the seed.

Table S 6-4 presents the results of the RF algorithm in predicting various physicochemical parameters of grapes and tissues based on the non-destructive dataset (n=720). The table includes data for soluble solids content (SSC), puncture force (PF), volume (VOL), weight per volume (W/VOL), anthocyanin content in the total grape (ANT-TO), skin (ANT-SK), pulp (ANT-PU), and seeds (ANT-SE), as well as chlorophyll content in the total grape (CHL-TO), skin (CHL-SK), pulp (CHL-PU), and seeds (CHL-SE). Each row corresponds to a specific parameter, and each column represents one of the six sampling dates (S1 to S6). The values provided in the table are means with standard deviations, and the significance of differences among the sampling dates is indicated by different letters based on Duncan's post hoc test ($\alpha < 5\%$).

Table S 6-4. Results of Random Forest in predicting soluble solids content (SSC - %), puncture force (PF - N), Volume (VOL - cm³), weight per volume (W/VOL - g/cm³), anthocyanin content in the total grape (ANT-TO - a.u.), skin (ANT-SK - a.u.), pulp (ANT-PU - a.u.) and seed (ANT-SE - a.u.), chlorophyll content in the total grape (CHL-TO - a.u.), skin (CHL-SK - a.u.), pulp (CHL-PU - a.u.) and seed (CHL-SE - a.u.) and the predawn leaf water potential (MPa). Based on the non-destructive dataset (n= 720).

Variables	S1	S2	S3	S4	S5	S6	p-Value
SSC	16.64 ^d ± 2.48	19.44 ^c ± 1.91	20.11 ^b ± 0.78	20.40 ^{ab} ± 0.75	20.57 ^a ± 1.05	20.46 ^{ab} ± 1.25	< 0.001
PF	3.76 ^a ± 0.90	2.76 ^b ± 0.76	2.12 ^c ± 0.47	1.92 ^d ± 0.39	2.10 ^c ± 0.48	2.11 ^c ± 0.60	< 0.001
VOL	1.23 ^a ± 0.27	1.16 ^b ± 0.27	0.96 ^c ± 0.23	0.86 ^d ± 0.21	0.91 ^{cd} ± 0.21	0.91 ^{cd} ± 0.19	< 0.001
W/VOL	1.33 ^d ± 0.08	1.41 ^c ± 0.12	1.52 ^b ± 0.12	1.58 ^a ± 0.12	1.55 ^b ± 0.12	1.54 ^b ± 0.11	< 0.001
ANT-TO	0.55 ^c ± 0.08	0.55 ^c ± 0.08	0.56 ^c ± 0.09	0.61 ^b ± 0.08	0.65 ^a ± 0.06	0.65 ^a ± 0.05	< 0.001
ANT-SK	0.36 ^c ± 0.06	0.38 ^b ± 0.05	0.40 ^{ab} ± 0.04	0.40 ^a ± 0.04	0.40 ^{ab} ± 0.04	0.41 ^a ± 0.04	< 0.001
ANT-PU	0.77 ^b ± 0.04	0.77 ^b ± 0.04	0.79 ^a ± 0.04	0.79 ^a ± 0.04	0.78 ^a ± 0.04	0.78 ^{ab} ± 0.04	< 0.001
ANT-SE	0.68 ^b ± 0.04	0.70 ^a ± 0.04	0.70 ^a ± 0.04	0.70 ^a ± 0.04	0.70 ^a ± 0.05	0.69 ^a ± 0.04	< 0.001
CHL-TO	0.50 ^a ± 0.16	0.50 ^a ± 0.16	0.47 ^a ± 0.16	0.39 ^b ± 0.14	0.29 ^c ± 0.11	0.28 ^c ± 0.08	< 0.001
CHL-SK	0.54 ^a ± 0.06	0.53 ^b ± 0.04	0.53 ^{bc} ± 0.04	0.52 ^c ± 0.04	0.52 ^{bc} ± 0.03	0.52 ^{bc} ± 0.04	< 0.001
CHL-PU	0.57 ^a ± 0.04	0.56 ^{bc} ± 0.04	0.57 ^{ab} ± 0.05	0.56 ^{ab} ± 0.05	0.55 ^{bc} ± 0.05	0.55 ^c ± 0.04	< 0.001
CHL-SE	0.50 ^a ± 0.07	0.49 ^a ± 0.07	0.50 ^a ± 0.07	0.50 ^a ± 0.07	0.49 ^a ± 0.07	0.49 ^a ± 0.07	0.967

The same line (among dates) and labelled with different letters mean the results are significantly different according to Duncan's post hoc test ($\alpha < 5\%$). SSC: soluble solids content (%); PF: puncture force (N); VOL: volume (cm³); weight per volume (g/cm³), ANT-TO: anthocyanin in the total grape (a.u.); ANT-SK: anthocyanin in the skin (a.u.); ANT-PU: anthocyanin in the pulp (a.u.); ANT-SE: anthocyanin in the seeds (a.u.); CHL-TO: chlorophyll in the total grape; CHL-SK: chlorophyll in the skin; CHL-PU: chlorophyll in the pulp; CHL-SE: chlorophyll in the seed.

CHAPTER 7

CASE STUDY VII: COMBINING SPECTRAL SENSING
AND SYSTEMS BIOLOGY FOR LEVERAGING
PRECISION VITICULTURE: UNRAVELLING
PHOTOSYNTHETIC PATHWAYS IN THE DOURO WINE
REGION - ASSESSING THE IMPACT OF SHADE
TREATMENT ON GRAPE PHYSIOLOGY

7. Case Study VII: Combining Spectral Sensing and Systems Biology for Leveraging Precision Viticulture: Unravelling Photosynthetic Pathways in the Douro Wine Region - Assessing the Impact of Shade Treatment on Grape Physiology

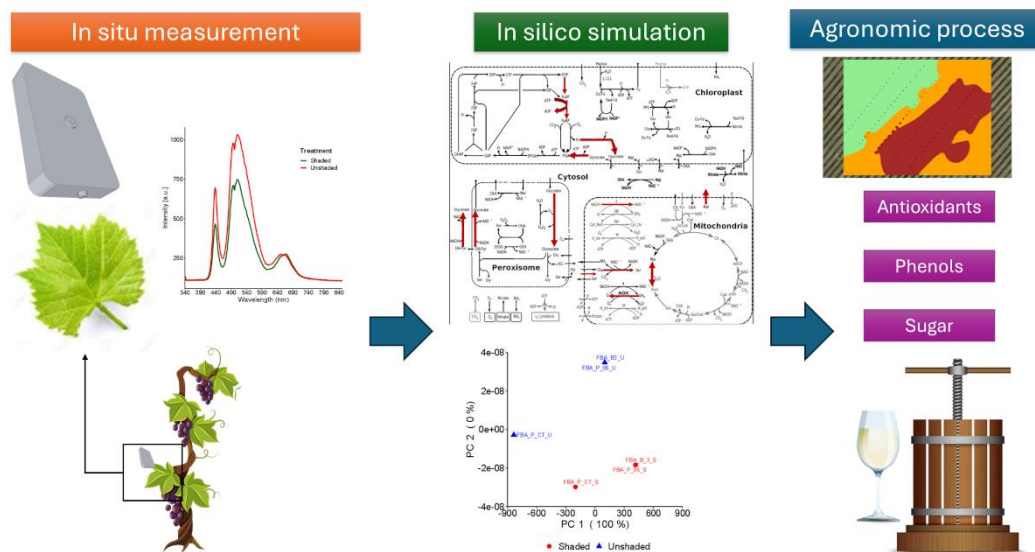
In preparation

Tosin, R. and supervision team.

IB2 Requite team led by Professor Conceição Santos.

TRIBE team led by Doctor Filipe Santos.

Activities within the project OmicBots - OmicBots: High-Throughput Integrative Omic-Robots Platform for a Next Generation Physiology-based Precision Viticulture, with reference PTDC/ASP-HOR/1338/2021.



Graphical abstract. 7-1. Graphical abstract case study VII.

Combining spectral sensing and system biology for leveraging precision viticulture: a case study unveiling the impact of shade curtains on grape physiology

Abstract

This study investigates how grapevines respond physiologically and metabolically to shading induced by artificial curtains. Through a multidisciplinary approach, grapevines' adaptations to shading are presented via biochemical analyses, hyperspectral data, and system biology techniques. Conducted in a Moscatel Galego (VIVIC - 8031) parcel within Portugal's Douro Wine Region, curtain shading was implemented during post-veraison and measured predawn leaf water potential (Ψ_{pd}) to assess water stress. Hyperspectral analysis characterised grapevine leaf spectra under shaded and unshaded conditions, while biochemical analyses examined chlorophyll concentrations and reactive oxygen species (ROS) levels. System biology simulations explored plant responses to varying light conditions, emphasising ROS's role in stress mediation. Results show reduced water stress in shaded conditions compared to unshaded vines.

Hyperspectral analysis revealed higher spectral absorbance in shaded vines, reflecting increased chlorophyll concentrations, with chlorophyll a-b levels 117.39% higher in shaded treatments. Conversely, ROS levels were 52.10% higher in unshaded treatments, indicating greater oxidative stress. The SL-AI model accurately predicted chlorophyll a (R^2 : 0.92, MAPE: 24.39%), chlorophyll b (R^2 : 0.96, MAPE: 17.61%), and ROS levels (R^2 : 0.76, MAPE: 52.17%). *In silico* simulations employing flux balance analysis (FBA) unveiled distinct phenotypes distinguishing shaded and unshaded vines, with reactions of the photorespiration system converging in different conditions and different compartments of the cell, such as in the chloroplast, peroxisomes and mitochondria. Integrating complementary information from biochemical analyses, hyperspectral data, and systems biology provides perception into grapevine responses to environmental stressors. Future research endeavours should prioritise the synergies of integrating these approaches to explain the intricate molecular mechanisms behind these responses, thereby refining viticultural practices for sustainable grape production. Leveraging advanced omics technologies such as genomics, proteomics, and metabolomics, in conjunction with precise metabolic models, holds immense potential to deepen understanding of grapevine metabolism and optimise viticultural practices for enhanced productivity and quality.

Keywords

Photorespiration; ROS; chlorophyll; self-learning; flux balance analysis; phenotype; phenometaboloma.

7.1. Introduction

In grapevine cultivation, shade curtains modulate plant microclimates and mitigate the adverse effects of excessive light exposure and temperature fluctuations driven by changing climatic conditions (Pallas & Christophe, 2015). While this practice contributes to conserving water resources (Rogiers et al., 2022), it also presents a challenge due to the existing gap in knowledge concerning the physiological implications of shade curtains on grapevine irrigation requirements and overall plant protection. Addressing this deficiency necessitates an integrated physiological framework capable of deciphering the interactions among plant physiology, environmental factors, and cultural practices, with support through systems biology.

This paper investigates the effects of shade curtains on grapevines using a systems biology approach supported by spectral data. By focusing on photorespiration, the study aims to demonstrate the potential of combining spectral data with systems biology to describe the central metabolism mechanisms of grapevine responses to shading, specifically regarding reactive oxygen species (ROS) accumulation.

This research hypothesises that integrating spectral data with systems biology can elucidate plant phenotypes and physiological mechanisms influenced by shading, thereby supporting precision viticulture (PV) practices. The concept of “phenometaboloma,” encompassing the entire grapevine’s metabolic phenotypes in response to shading conditions, is presented to explore these effects. This term covers a wide array of metabolic pathways and enzymatic reactions affected by changes in light availability, offering a holistic view of grapevine physiology under varying environmental conditions. By integrating metabolic models and *in silico* data, the study seeks to unravel the metabolic pathways affected by shading conditions, shedding light on the biological interactions driving grapevine responses.

Table 7-1 reviews enzymes for metabolic pathways governing grapevine responses to light-induced stress. This table serves two main proposes: i) identification and analysis, examining enzyme activity under specific light conditions to reveal patterns of activation or repression, thereby providing insights into how grapevines respond to varying light intensities and durations, and ii) qualitative model analysis of critical metabolic processes such as water availability, photorespiration, CO₂ absorption during photosynthesis, and

sugar formation, which are crucial for understanding grapevine responses to light stress within a systems biology approach.

Table 7-1. Enzymes and metabolic pathways in grapevines related by light and shading conditions. The table is categorised into two sections: unshaded and shaded. Each section contains columns indicating metabolic pathways, enzyme expression (whether upregulated or downregulated), enzyme function, and relevant references.

Treatment	Pathway	Expression	Function	Reference
Unshaded				
hydroxycinnamic acids (HCAs)	Phenolic acids	Up	UV protection, pigmentation, cell defence	Teixeira et al. (2013); Adams (2006)
hydroxybenzoic acids (HBAs)	Phenolic acids	Up	UV protection, pigmentation, cell defence	Teixeira et al. (2013); Adams (2006)
Resveratrol	Non-flavonoid polyphenols	Up	UV protection, pigmentation, cell defence	Sternad Lemut et al. (2011); Spayd et al. (2002)
Quercetin	Polyphenol	Up	UV protection	Sternad Lemut et al. (2011); Spayd et al. (2002)
Kaempferol	Polyphenol	Up	UV protection	Sternad Lemut et al. (2011); Spayd et al. (2002)
Myricetin	Polyphenol	Up	UV protection	Sternad Lemut et al. (2011); Spayd et al. (2002)
Lipocalin (s240)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
quinone oxidoreductase-like protein (s472)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
ascorbate peroxidase2 (APX2)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
peroxiredoxin (PRX)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
glutathione-s-transferase(GST)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
catalase (CAT)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
isoflavone-reductase-like protein (IRL)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
nucleoside diphosphate kinase2 (NDPK2)	Oxidative stress defence	Up	Cell defence	Nilo-Poyanco et al. (2013)
Auxin (AUX)	Hormones	UP	Hormone signal	Sun et al. (2017b)
Xylose	Cell membrane	Down	Cell membrane	Ayenew et al. (2015)
Xylobiose	Cell membrane	Down	Cell membrane	Ayenew et al. (2015)
Phenylalanine	Amino acid pathway	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
Light-inducible protein (ELIP1)	Chlorophyll biosynthesis	Up	Regulates the chlorophyll biosynthesis	du Plessis et al. (2017)
photosystem II PsbO protein	Photosynthesis	Down (final stage)	Photosynthesis	du Plessis et al. (2017)
LHB1B1 light-harvesting protein	Photosynthesis	Down (final stage)	Photosynthesis	du Plessis et al. (2017)
polyphenol oxidase chloroplast precursor	Photosynthesis	Down (final stage)	Photosynthesis	du Plessis et al. (2017)

Shade

blue light receptor cryptochrome 2 (CRY2)	Photoreceptors	Down	Light receptors	Sun et al. (2017b)
HY5	Photoreceptor regulator	Down	Light receptors	Sun et al. (2017b)
HY5- homolog (HYH)	Photoreceptor regulator	Down	Light receptors	Sun et al. (2017b)
cytokinin (CTK)	Hormones	Up	Hormone signal	Sun et al. (2017b)
brassinosteroid (BR)	Hormones	Up	Hormone signal	Sun et al. (2017b)
pyrabactin resistance 1/PYR1-like (PYR)	Hormones	Up	Hormone signal	Sun et al. (2017b)
ABA-responsive (element binding factor)	Hormones	Up	Hormone signal	Sun et al. (2017b)
Maleate	Maleic acid	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
beta-alanine	Amino acid pathway	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
Citrate	Amino acid pathway	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
Aspartate	Amino acid pathway	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
procyanidin B1	Polyphenol	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
Epigallocatechin	Polyphenol	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
Catechin	Polyphenol	Up	UV protection, pigmentation, cell defence	Reshef et al. (2017)
Raffinose	Sugars; Carbon metabolism	Up	Carbon metabolism	Reshef et al. (2017)

Up - consists of positive regulation of enzymes involved in metabolic pathways; Down - consists of negative regulation of enzymes involved in metabolic pathways.

Systems biology to unravelling grape physiology under light stress

This article uses systems biology to explain the leaf's physiological effects of shaded conditions on grapevines. Despite its proven efficacy across various disciplines - from microbiology (Dennis et al., 2004) to human health (Joshi et al., 2021) and even food engineering (Liu et al., 2019a) - the application of systems biology in agronomy, particularly for enhancing crop management decisions, remains largely underexplored. Through metabolomics, proteomics, and genomics, this interdisciplinary field offers perceptions of plant physiology that elude traditional data-driven approaches (Kumar et al., 2015). By integrating molecular components into networks, systems biology facilitates modelling dynamic behaviours, paving the way for creating virtual phenotypes. These computational models not only predict observable characteristics but also identify critical genetic and metabolic determinants of plant traits, enabling the development of crops with superior qualities and optimised resource use (de Oliveira Dal'Molin et al., 2010).

The predictive modelling capabilities of systems biology, mainly through techniques such as Flux Balance Analysis (FBA) (Orth et al., 2010), hold the potential to improve the understanding of plant physiology under varying environmental conditions (Lakshmanan et al., 2015; Poolman et al., 2013). Despite these advancements, integrating systems biology into practical agricultural applications has been limited, often hindered by the unidirectional flow of data and the lack of models that summarise the plant-environment interactions.

Integrating plant spectral sensing and system biology

Conventional laboratory-based methodologies for plant biochemical analysis, while indispensable to supporting system biology models, present several limitations, including high costs, laborious procedures, and the consumption of sample volumes, which may not be feasible for rare plants (Tosin et al., 2024b; D'Agostino & Tripodi, 2017). Hence, there is a need for non-destructive, cost-effective, and efficient alternatives that can complement traditional methods and enhance the scope of systems biology research.

Hyperspectral sensors allow for real-time monitoring of biochemical variables like foliar pigments, which are capable of predicting levels of chlorophyll (Tosin et al., 2021), carotenes (Van Wittenberghe et al., 2024), anthocyanin (Liu et al., 2019b), flavonoids (Falcioni et al., 2023), and ROS (Herman et al., 2018). This study hypothesises that hyperspectral data can model foliar pigments and ROS, making it a novel approach by integrating systems biology with hyperspectral data to investigate the impact of shade on photorespiration in grapevines.

Figure 7-1 illustrates an integrated methodology to show the plant responses by merging genomics, metabolomics, and systems biology approaches. This framework enables the establishment of connections between laboratory experiments and field observations, facilitating a more profound knowledge of grapevine physiology under varying environmental conditions. Central to this methodology is integrating sensors for detecting molecular components and monitoring the physiological state of plants in real time. By leveraging high-throughput, high-resolution, and high-dimensionality techniques, this approach offers a holistic view of plants' responses to abiotic stress, paving the way for more precise and tailored agricultural practices.

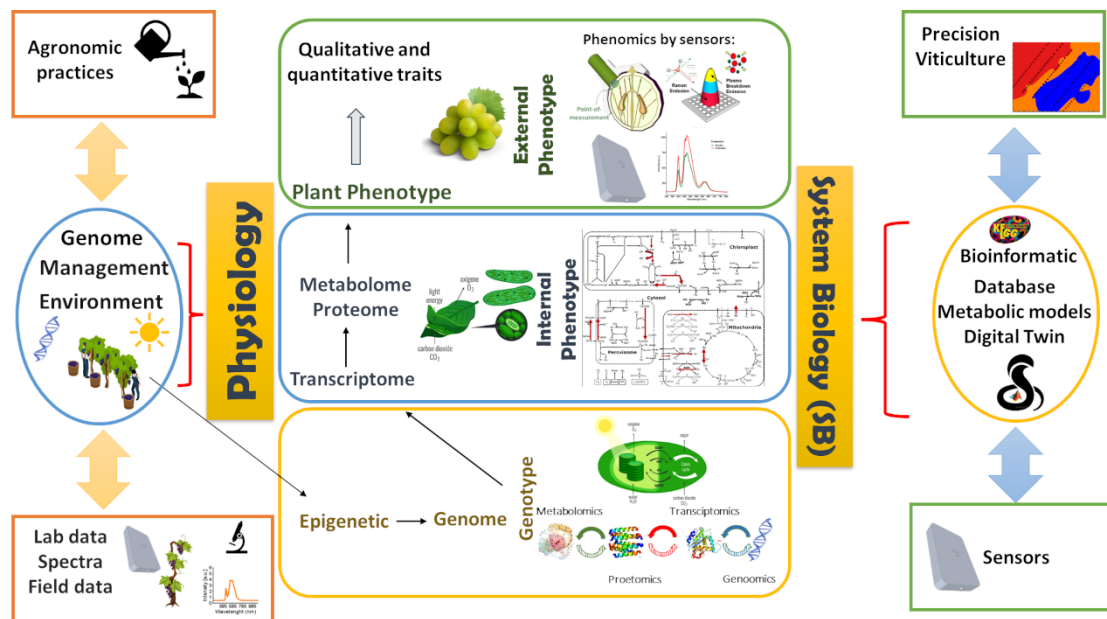


Figure 7-1. Representation of an integrated methodology combining genomics, metabolomics, and systems biology approaches aimed to establish connections between laboratory experiments and field observations. Additionally, the figure demonstrates the integration of sensors for detecting molecular components and monitoring the physiological state of plants to feed the system biology.

By investigating these effects, this paper aims to demonstrate how systems biology can elucidate the phenotype observed in the field, focusing on the specific case of shade curtains. The study utilises a qualitative analysis approach, which involves comparing assumptions with experimental data. This method helps integrate metabolic models and *in silico* data to elucidate how shading conditions affect metabolism. By incorporating spectral data and biochemical variables such as foliar pigments, water status, and ROS levels, the research aims to reveal the complex biological interactions driving grapevine responses to shade. The specific objectives include: i) modelling foliar pigments (chlorophyll) and ROS using hyperspectral data, ii) initiating the development of a virtual phenotype and validating certain potential behaviours within the photorespiration

pathway under both shaded and unshaded conditions, and iii) identifying key enzymes and reactions involved in these processes.

7.2. Material and Methods

7.2.1. Test Site and Sampling

The study was conducted at Quinta Vale dos Cavalos, which belongs to Poças Vinhos, located in Numão in the Douro Wine Region. The Douro region's Mediterranean-like climate entails warm, dry summers with pronounced hydric stress typically observed post-flowering. During the grape ripening period, precipitation generally is lower than 30 mm, accompanied by high solar irradiation values, temperature levels, and vapour pressure deficit prone to combined light, thermic, and hydric plant stresses (Cunha et al., 2016).

The study focused on the grape variety Moscatel Galego Branco (Vitis International Variety Catalogue - VIVC 8031) within a non-irrigated commercial parcel encompassed by sections shaded by a curtain. The experiment was conducted in both shaded and unshaded sections of the parcel. The leaf analysis was carried out at final maturity (S1: 26/08/2022), with ten samples collected from areas shaded by the curtain and ten from unshaded areas.

7.2.2. Phenotype characterisation

Water status

Predawn leaf water potential (ψ_{pd}) of the assessed parcel was measured using a Scholander pressure chamber (PMS600, Albany, OR, USA) (Scholander et al., 1965) on one leaf per vine block sampled, providing a characterisation of the effect of the curtain shades on plant water status, characterising the phenotypes under different treatments.

Hyperspectral data

Hyperspectral data were obtained using a point-of-measurement hyperspectral prototype device (Martins et al., 2019). In this prototype, light is received by a central pinhole fibre and delivered by surrounding fibres, enabling it to penetrate the leaf. This device recorded reflectance signatures across the electromagnetic spectrum (340 nm to 850 nm). Three data points per leaf were collected, and the average was used for spectral representation. The spectral data will be used to develop a predictive model of plant pigment and ROS (see next section).

7.2.3. Biochemical analysis

Pigments analysis

A sample of 100 mg of grapevine fresh leaf was used for pigment quantification. The leaves sampled in the test site were stored at -80°C and processed using liquid nitrogen. Pigments were extracted using an Acetone:Tris-HCl (50 mM) buffer (80:20, pH 7.8) (Sims & Gamon, 2002). Bead beater tubes facilitated maceration, with readings at 663 nm, 537 nm, 647 nm, and 470 nm, using a spectrophotometer (Omega Multimode Microplate Reader from BMG LABTECH) for chlorophyll a and chlorophyll b. Results are presented in mg/g fresh mass of the extract. The unit is presented by milligrams per gram of fresh matter (mg/gFM).

Superoxide (O_2^-) quantification for ROS assessment

A total of 100 mg of fresh leaf was utilised. Bead beater tubes (2 mL) with spheres were used for extraction. The extraction buffer comprised 0.01 M phosphate buffer (pH = 7.8), 0.05% NBT (diluted in 100 µL of DMSO), and 10 mM sodium azide. After shaking, samples were stored in the dark, followed by centrifugation (13 000 xg, 2 minutes, 4°C) and heating (85°C, 10 minutes); after, it was placed on ice for 10 minutes. Superoxide (O_2^-) levels, indicative of ROS activity, were quantified at 580 nm using a spectrophotometer (Omega Multimode Microplate Reader from BMG LABTECH) (Costa-Santos et al., 2021; Gajewska & Skłodowska, 2007). The unit is presented by absorbance per gram of fresh matter (ABS/gFM).

7.2.4. Spectral biochemical modelling approach

The Vis-NIR data predicted the chlorophyll and ROS. The preprocessing steps were undertaken to ensure the reliability and accuracy of the leaf spectral data. Potential outliers were identified and removed to eliminate any abnormal data points that could bias the results. A logarithm multiplicative scattering correction (MSC-log) technique was applied to the spectral data. This correction method helps to enhance the spectral features and mitigate any variations caused by scattering effects, thereby improving the overall quality of the data (Martins et al., 2022).

The modelling of spectra-laboratory analysis (pigments and ROS) was subjected to modelling using a self-learning artificial intelligence algorithm (SL-AI) (Martins, 2018). The utilisation of the SL-AI is validated by its proficiency in establishing correlations between spectral attributes and the investigated biochemical parameters through Covariance Mode (CovM) analysis. The justification for employing the SL-AI stems from

its proficiency in managing datasets characterised by variations and its proven track record in accurately predicting grapevine biochemical characteristics, evidenced in previous studies (Tosin et al., 2023b; Tosin et al., 2022).

The SL-AI base model was evaluated using the statistical metrics: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE - %).

7.2.5. *In silico* simulations

The photorespiration model proposed by Huma et al. (2018) was considered for the *in silico* simulations of the photorespiration process under different shading conditions. The literature review outlined in Table 7-1 shows the potential impacts on metabolites involved in photorespiration synthesis under shaded and unshaded conditions.

The chlorophyll and ROS levels measured in the laboratory and those estimated through SL-AI were incorporated into the photorespiration model to elucidate the differences in each shading condition.

This study hypothesises that under conditions of excessive light exposure, grapevines experience an elevation in ROS levels and a decline in chlorophyll concentrations. Also, grapevines activate a mechanism to increase chlorophyll production under shaded conditions, compensating for reduced photon availability. Two scenarios were considered to represent the impact of plant shade curtains on leaf physiological adaptative mechanisms i) in unshaded scenarios, leaves exhibit a lower photosynthetic efficiency, around 80%, yet remain free from oxidative stress; and ii) in shaded conditions, leaves become more efficient due to increased chlorophyll production, enabling photosynthetic rates to surpass those of sun-exposed leaves, due to the absence of oxidative stress by preventing the accumulation of ROS.

Furthermore, under excessive light conditions, the costs of energy and ATP to mitigate ROS accumulation may compensate for the benefits of increased photon capture. Conversely, shaded plants may achieve greater energy efficiency in the Calvin cycle due to reduced ROS-induced oxidative stress (Adams et al., 2018).

This study employs system biology approaches focusing on key reactions such as the peroxisomal catalase reaction and peroxisomal glycolate oxidase reaction presented in the model used for the simulations (Huma et al., 2018). These reactions are incorporated in the *in silico* model to simulate grapevine photorespiration under different light conditions. Specifically, three photorespiration simulation hypotheses are formulated: i) Photorespiration I (PH I): Modifying the peroxisomal catalase reaction in response to

ROS fluctuations within shaded environments, what changes would occur in fluxes?; ii) Photorespiration II (PH II): How would fluxes be affected by manipulating the peroxisomal glycolate oxidase reaction? and iii) Photorespiration III (PH III): If the peroxisomal catalase reaction and the glycolate oxidase reaction were adjusted, what alterations would be observed in fluxes?

In silico simulations were generated using the Cobratoolbox for Matlab (Heirendt et al., 2019) with the glpk solver, aiming to solve the relationship $S_v = 0$. This equation, fundamental in systems biology, embodies the mass balance principle or stoichiometry. Here, S represents the stoichiometric matrix, delineating the stoichiometric coefficients of metabolites engaged in biochemical reactions within the biological system. Each row of the matrix corresponds to a metabolite, while each column corresponds to a reaction. Conversely, v symbolises the vector of reaction rates or fluxes, indicating the flow or rate of each reaction. The equation $S_v = 0$ stipulates that the product of the stoichiometric matrix S and the vector of reaction rates v equals zero under steady-state conditions. This principle underscores mass conservation, ensuring that the total metabolite production rate equals its consumption within the system.

The *in silico* model utilised for these simulations had specified lower and upper bounds in abstract units. The lower bound was set to zero, while the upper bound was determined as the maximum value for these reactions, also expressed in abstract units. However, for the qualitative analysis in this paper, the lower bound remained at zero. In contrast, the upper bound was set to 100 to represent lower values of ROS observed in laboratory analyses. The percentage variations were applied to the upper bound of these objective reactions in the simulations of varying conditions. Furthermore, Monte Carlo (MC) simulation was performed over ten times from the minimum to the maximum variance to observe changes in the phenotype space, thereby exploring the robustness and variability of the model under different conditions.

Principal Component Analysis (PCA) was used to discriminate the phenotype spaces within photorespiration and the Monte-Carlo simulations.

7.3. Results and Discussion

7.3.1. Phenotype characterisation

This study investigated grapevines' physiological and metabolic responses to varying light conditions, contrasting those subjected to direct sunlight with those shaded by curtains. Results revealed significant differences in biochemical parameters, including water status, hyperspectral signatures, pigment concentrations, and ROS levels, by

supporting research on grapevine physiology (e.g., du Plessis et al., 2017; Nilo-Poyanco et al., 2013) and offering perceptions of the mechanisms driving these responses.

Water status

The differences in Ψ_{pd} between shaded and unshaded grapevines (Figure 7.2) underscore the influence of light availability on plant water status (Zhang et al., 2022; Palliotti & Poni, 2015). Specifically, grapevines in unshaded conditions showed lower Ψ_{pd} values, suggesting elevated water uptake and transpiration rates, possibly due to heightened metabolic activity and photosynthetic capacity in response to abundant sunlight (Adams et al., 2018; Reshef et al., 2017). However, grapevines may close their stomata in stressful conditions, reducing transpiration (Savoi et al., 2016). In such cases, alternative mechanisms are needed to dissipate the excess energy, particularly in the unshaded treatment. These observations align with the concept of phenometaboloma, suggesting that shade curtains modulate grapevine physiology, particularly influencing photorespiration (Ju et al., 2018).

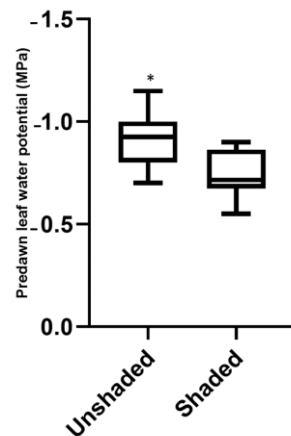


Figure 7-2. Predawn Leaf Water Potential Variation in Grapevine Leaves under unshaded and shaded conditions. * Statistically significant ($p < 0.05$) according to t-test.

Hyperspectral signatures

The hyperspectral analysis reveals a distinct difference in activity between plants subjected to unshaded and those under shaded (Figure 7-3), showing that plants receiving more sunlight have lower values of spectra absorbance when compared to those in shaded conditions.

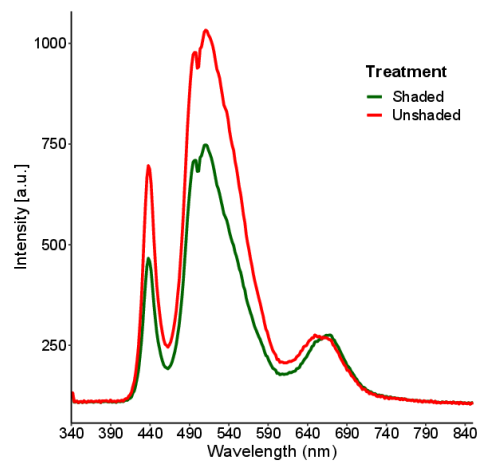


Figure 7-3. Mean spectra from leaves vines exposed to unshaded conditions and shaded conditions.

The observed variations in spectral patterns between grapevines exposed to sunlight and those in shaded conditions indicate distinctions in the vine phenotype (Figure 7-3). The shaded treatment shows a higher absorption across the blue (400-480 nm), green (481-560 nm), and red (561-700 nm) spectral zones, which are commonly associated with foliar pigments like chlorophyll (Lichtenthaler & Babani, 2004).

Pigment and ROS analysis

The pigment analysis in Table 7-2 revealed a significant difference in chlorophyll a and a-b concentrations. Under shaded conditions, plants exhibited heightened production of photosynthetic pigments, enhancing their ability to capture and utilise light energy efficiently. The higher chlorophyll concentrations in shaded grapevines indicate a compensatory mechanism to optimise light capture and energy transduction under reduced light availability (Van Wittenberghe et al., 2024). This aspect can be further clarified by examining the profile of chlorophyll a/b ratios. Specifically, understanding the balance between chlorophyll a (which is typically associated with photosynthetic activity) and chlorophyll b (which plays a role in light harvesting and photoprotection) (Wang et al., 2022; Christ & Hörtensteiner, 2013) could provide perceptions of the metabolic processes. Variations in these ratios may indicate shifts in the plant's balance between catabolic and anabolic pathways, reflecting adaptations to environmental conditions such as light availability (Luo et al., 2019). Therefore, analysing chlorophyll a/b ratios can offer information about the physiological responses of grapevines to varying light conditions, shedding light on their metabolic adjustments for enhanced survival and productivity.

Table 7-2. Comparison of shaded and unshaded conditions for chlorophyll a (mg/gFM), chlorophyll b (mg/gFM), chlorophyll a+b (mg/gFM) and ROS O₂ (ABS/gFM).

Biochemical analytes	Shaded	Unshaded	P-value	Variation %
Chlorophyll a (mg/gFM)	0.19	0.12	0.001*	158.33
Chlorophyll b (mg/gFM)	0.08	0.11	0.006*	71.03
Chlorophyll a+b (mg/gFM)	0.27	0.23	0.037*	117.39
ROS O ₂ (ABS/gFM)	0.99	1.90	0.019*	52.10

* Statistically significant ($p < 0.05$) according to t-test. mg/gFM: milligrams per gram of fresh matter; ABS/gFM: absorbance times grams of fresh matter. Variations were computed considering the initial stage of the unshaded treatments.

Regarding ROS, there was also a decrease in ROS levels (Table 7-2), which may be attributed to a reduction in the plant's photosynthesis and energy activity levels or its energy efficiency despite the greater amount of chlorophylls (Carvalho et al., 2015). The differences in ROS levels between shaded and unshaded plants provide further evidence of the impact of light exposure on grapevine physiology (Sun et al., 2017b; Carvalho et al., 2015). Higher ROS concentrations observed in unshaded conditions suggest a link between light intensity, ROS production, and oxidative stress (Carvalho et al., 2015), which supports the hypothesis that shade curtains mitigate oxidative stress by influencing photorespiration, ultimately affecting ROS levels (Adams et al., 2018). These results show the role of antioxidant defence mechanisms in grapevines under varying light conditions (Sandalio et al., 2013), underscoring the importance of understanding plant-environment interactions for sustainable viticulture practices

ROS at homeostatic levels are essential to the plant under stress conditions, and shade maintains the eustress levels within homeostatic values. On the other hand, in the unshaded condition, chlorophylls were controlled, and ROS increased. When in excess (particularly in organelles that possess electron transport chains, like chloroplasts and mitochondria), ROS can damage genetic material, lipids, and proteins, among other biomolecules, ultimately leading to the degradation of the organelles and cell disturbance. In unshaded leaves, there is a high production of energy and an increase in the transport of electrons from photosynthesis, which may increase the generation of superoxide molecules, hydrogen peroxide and the hydroxyl radical (Iqbal et al., 2021).

7.3.2. Modelling biochemical variables

The relationship between spectral data and metabolites bridges environmental conditions and internal biochemical processes by correlating spectral signatures with metabolites. Vis-NIR has mainly been used to quantify grapevine metabolites (van

Wyngaard et al., 2021). For example, quantifying chlorophyll through the Vis-NIR data (e.g., Tosin et al., 2021) is a proxy for water status and lacks a physiological explanation. Integrating hyperspectral data with systems biology methodologies enables a holistic understanding of how environmental factors influence plant metabolism and phenotype (Pazhamala et al., 2021). It can promote vineyard management practices to enhance grape quality and yield based on a physiological approach.

The implementation of SL-AI in modelling chlorophyll a (mg/gFM) and b (mg/gFM) and ROS (ABS/gFM) yielded robust statistical metrics (Figure 7-4) despite the limitation of a small dataset comprising 19 observations. However, not all parameters were assessed across all observations due to the CovM architecture present in SL-AI. This limitation arises from the predefined number of clusters in the SL-AI, which might not effectively accommodate the diverse variations present in the dataset. Therefore, while the statistical metrics demonstrate the robustness of the model, the interpretation of the results should be cautious, considering the potential impact of data allocation constraints imposed by the SL-AI architecture.

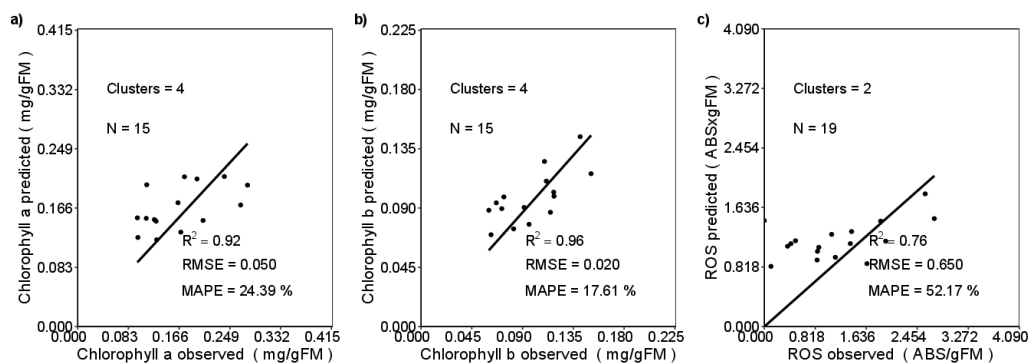


Figure 7-4. Results of the use of foliar spectral data combined with self-learning artificial intelligence (SL-AI) in modelling: a) chlorophyll a (mg/gFM), b) chlorophyll b (mg/gFM) and c) reactive oxygen species (ROS – ABS/gFM). N= number of samples considered. Coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE - %).

7.3.3. *In silico* simulations

This study combined hyperspectral data with system biology, allowing simulations of plant responses under different light conditions and focusing on the role of ROS in mediating plant stress responses, mainly through its correlation with photorespiration mechanisms. The simulations, informed by experimental data and existing literature (Wang et al., 2022; Ju et al., 2018), highlight the central role of enzymes such as peroxisomal catalase and peroxisomal glycolate oxidase in ROS metabolism. Peroxisomal catalase works by transforming hydrogen peroxide into oxygen and water, effectively counteracting the potentially detrimental impacts of ROS (Carmo de Carvalho

e Martins et al., 2022). Similarly, peroxisomal glycolate oxidase converts O_2^- and glycolate into glyoxylate and H_2O_2 . Although H_2O_2 is a reactive oxygen species that can contribute to oxidative stress, it also plays a role in the plant's antioxidant defence system by acting as a signalling molecule and being managed by detoxifying enzymes (Sandalio et al., 2013).

Plants naturally produce ROS when exposed to environmental stresses like high light intensity and temperature. Plants adopt a range of defence mechanisms known as antioxidants to protect themselves from ROS damage. These include small molecules such as ascorbic acid, ascorbate, glutathione, non-protein amino acids, phenolic compounds, α -tocopherol, and certain alkaloids, as well as antioxidant enzymes like superoxide dismutase, catalase, peroxidases, ascorbate peroxidase, monodehydroascorbate reductase, and dehydroascorbate reductase (Berni et al., 2019). In environments with prolonged sunlight exposure, ROS-detoxifying enzymes like catalase, glycolate oxidase, and glutathione reductase may be upregulated to eliminate ROS, thereby safeguarding cellular components from oxidative stress-induced damage. This regulatory response helps maintain cellular stability and function, ensuring plant health and resilience (Wang et al., 2022).

This paper chose PH III hypothesis (see section 7.2.5) to demonstrate the metabolism simulation due to its advanced phase of photorespiration, occurring in peroxisomes—the primary site of ROS detoxification (Geigenberger & Fernie, 2014). This choice aligns with the laboratory data used in this study, particularly the quantification of O_2^- . PH III encompasses a widespread set of reactions involved in photorespiration, including those related to energy metabolism and biosynthesis. PH III accounts for key processes regulating carbon and energy flow within the cell by incorporating both peroxisomal catalase and glycolate oxidase reactions.

Figure 7.5 presents the simulated metabolic map of reactions for PR III under two contrasting conditions: a) shaded and b) unshaded. The map illustrates the hypothetical dynamic regulation of photorespiratory metabolism in response to changes in environmental light conditions theorised in this paper. The reactions highlighted in blue and red show the intensity of the production or consumption of the reactions' products; these reactions include maleate, citrate, and ascorbate, along with their interrelationships derived from literature sources (referenced in Table 7-1).

In the simulations, these reactions have greater fluxes in unshaded conditions compared to shaded conditions. Additionally, PH III facilitates the analysis of metabolic changes in response to environmental factors such as shading, as evidenced by observed alterations in metabolite expression levels.

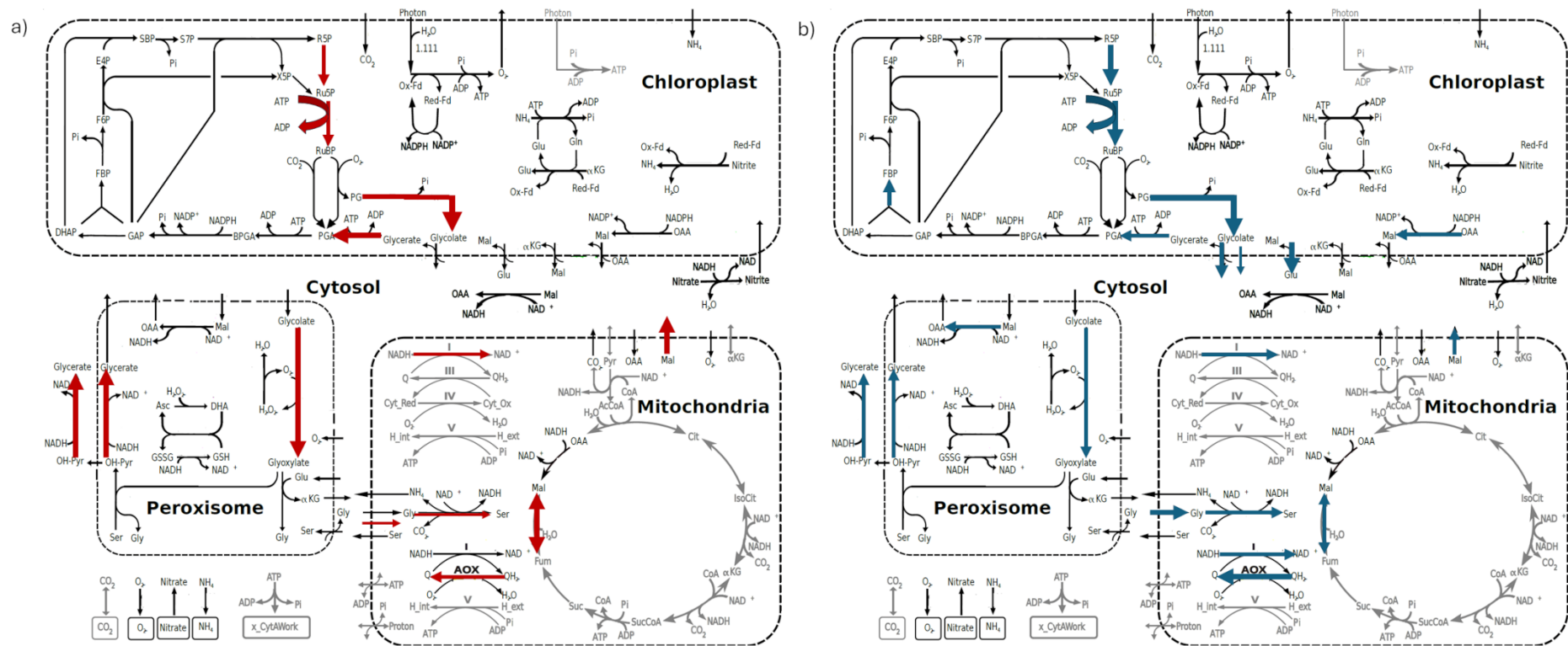


Figure 7-5. Photorespiration cycle under shaded a) and unshaded b) conditions, showcasing the Photorespiration III with the objective reactions in the flux balance analyses of Peroxisomal Catalase and Peroxisomal Glycolate Oxidase. The diagram illustrates the reactions involved, offering a dynamic regulation of photorespiratory metabolism in response to environmental light conditions. The arrows' size indicates each pathway's intensity under the respective condition.

Maleate and citrate induce plant pigmentation in shading conditions (Reshef et al., 2017). Maleate, primarily associated with photorespiration II, is produced from other reactions and transported to the mitochondria (Tomaz et al., 2010). Citrate is converted into oxaloacetate in the cytoplasm and then transferred to the chloroplast, where photorespiration resumes (Zakhartsev et al., 2016). Ascorbate protects against oxidative damage during photorespiration by acting as an antioxidant, neutralising ROS and safeguarding plant cells (Burbidge et al., 2021), being more induced under sun exposure conditions.

Under unshaded conditions, the fluxes indicate increased activity in metabolic pathways associated with the glycolate metabolism and ATP production, as evidenced by the elevated levels of chloroplast glycolate and chloroplast ATP due to increased sun exposure and heightened activity in energy acquisition pathways. The unshaded condition also showed a greater amount of ROS, which alters some mechanisms for ATP synthesis. Conversely, shaded conditions demonstrate a shift in metabolic activity, with a reduction in glycolate metabolism and ATP production, accompanied by changes in the levels of intermediates such as chloroplast glycerate and chloroplast ribulose-1,5-bisphosphate (Figure 7.5)

This paper (Table 7-1 and Figure 7-5) demonstrates the hypothetical metabolic fluxes in prioritise energy conservation and carbon storage in shaded conditions, which leads to decreased levels of maleate, intermediates in energy metabolism and biosynthesis (Reshef et al., 2017). Conversely, unshaded environments with ample sunlight suggest metabolic pathways towards increased energy production and carbon fixation, resulting in higher concentrations of maleate (Adams et al., 2018; Reshef et al., 2017). Additionally, although not shown in the simulations, antioxidants like ascorbate vary with light availability, with lower levels in shaded conditions due to reduced ROS production (Sandalio et al., 2013; Nilo-Poyanco et al., 2013). While Reshef et al. (2017) showcased the light interference through an *in situ* case study, providing a realistic representation of natural conditions, the results of Nilo-Poyanco et al. (2013), conducted under controlled indoor conditions, may introduce variations in the physiological response. This comparison highlights the importance of considering experimental context when interpreting the impact of light on plant metabolism, as controlled conditions might not fully capture the complexity of natural environments.

Figure 7.6 shows the phenotype spaces resulting from the three photorespiration simulation approaches described in section 7.2.5. In panel a) phenotype spaces for five FBA models under unshaded and shaded conditions. Panel b) showcases phenotype spaces derived from MC simulations, elucidating variations in ROS levels observed in laboratory experiments for the peroxisomal catalase reaction (MC_P_CT) and

peroxisomal glycolate oxidase reaction (MC_P_96). Unfortunately, both reactions were set as objective with the same variation in the data, and MC simulations did not present valid results. This non-response for the third hypothesis may be related to the choice of two reactions related to ROS and the same variation used for these reactions, suggesting a static stage for the MC reactions.

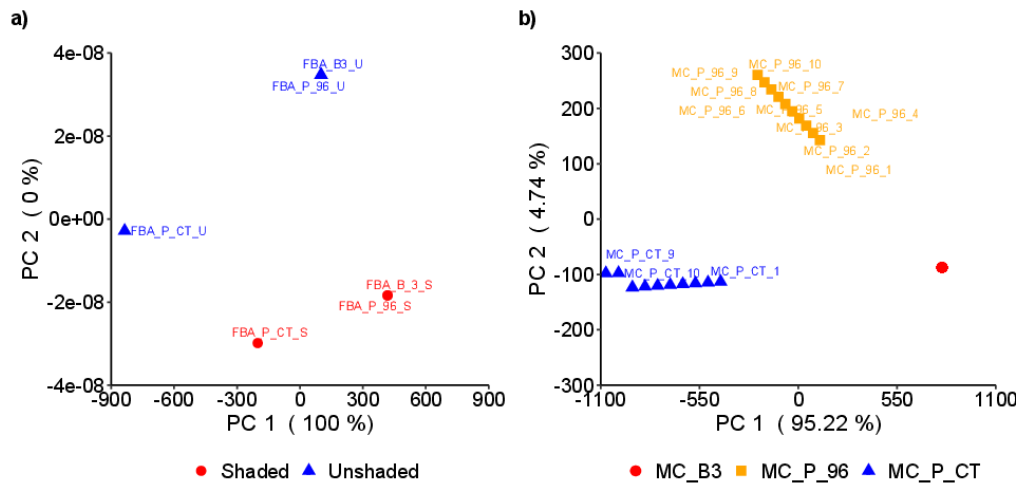


Figure 7-6. Panel a) shows the phenotype spaces for the three Flux Balance Analysis (FBA) models under different photorespiration conditions. For unshaded conditions, FBA models include FBA_P_CT_U (Peroxisomal Catalase Reaction as objective), FBA_P_96_U (Peroxisomal Glycolate Oxidase Reaction), and FBA_B3_U (Peroxisomal Catalase Reaction as objective and Peroxisomal Glycolate Oxidase Reaction as objective). Similarly, shaded conditions are represented by FBA_P_CT_S, FBA_P_96_S and FBA_B3_S, respectively. Panel b) demonstrates the phenotype spaces resulting from Monte Carlo (MC) simulations, illustrating the range of variations in chlorophyll and reactive oxygen species (ROS) levels assessed in laboratory experiments. MC simulations encompass MC_B3 Peroxisomal Catalase Reaction as objective and Peroxisomal Glycolate Oxidase Reaction as objective), MC_P_96 Peroxisomal Glycolate Oxidase Reaction as objective) and MC_P_CT (Peroxisomal Catalase Reaction as objective). In the case of MC_P_96 and MC_P_CT, 1 represents higher ROS (unshaded condition) and 10, lower ROS (shaded condition). MC_B3 did not result in a valid MC simulation.

In shaded conditions, plants can adapt to reduced light availability, leading to changes in metabolic fluxes and enzyme activities. Conversely, in unshaded conditions with ample sunlight exposure, metabolic pathways are geared towards increased energy production and carbon fixation. This differential regulation of metabolic pathways underscores the adaptive strategies employed by grapevines to optimise their performance under varying light conditions. FBA_B3 simulation may better simulate ROS levels than the FBA_P_CT and FBA_P_96 simulations. This inference is drawn from the fact that the FBA_B3 simulation shows the highest loadings for ROS-related reactions among the three simulations. Specifically, the FBA_B3 simulation has the highest loadings for reactions involving peroxisomal catalase and glycolate oxidase, both enzymes in ROS metabolism (Carvalho et al., 2015).

The top 10 reactions identified through PCA loadings in both shaded and unshaded conditions encompass key enzymes involved in various metabolic pathways crucial for

plant physiology, as well as the vector of the stoichiometric matrix for each reaction (Table 7-4). Chloroplast ferredoxin reductase (Chl_FerredoxinReductase), chloroplast glyceraldehyde-3-phosphate dehydrogenase (Chl_G3Pdh), and chloroplast phosphoglycerate kinase (Chl_PGK) represent steps in the Calvin cycle essential for carbon fixation and energy production (Trost et al., 2006). Chloroplast ribulose-5-phosphate kinase (Chl_Ru5Pk) and chloroplast phosphoglycolate phosphatase (Chl_PGlyPase) are associated with the pentose phosphate pathway, facilitating the regeneration of ribulose-1,5-bisphosphate and the production of NADPH, crucial for redox balance and biosynthetic processes (Geigenberger & Fernie, 2014). Additionally, chloroplast ribulose-1,5-bisphosphate oxygenase (Chl_RuBPOxid) participates in photorespiration.

Table 7-3. The top 10 reactions ranked through the PCA loadings contribute to the leave's phenotype in shaded and unshaded conditions and the Flux Balance Analysis (FBA) in a.u. for each simulation.

Top 10 reactions	%	FBA_P_CT_U	FBA_P_CT_S	FBA_P_96_U	FBA_P_96_S	FBA_B3_U	FBA_B_3_S
Chl_FerredoxinReductase	7.39%	-793.91	-525	-396.95	-262.5	-396.95	-262.5
Chl_G3Pdh	7.04%	756.1	500	378.05	250	378.05	250
Chl_PGK	7.04%	756.1	500	378.05	250	378.05	250
Chl_Ru5Pk	4.22%	453.66	300	226.83	150	226.83	150
Chl_PGlyPase	2.81%	302.44	200	151.22	100	151.22	100
Chl_RuBPOxid	2.81%	302.44	200	151.22	100	151.22	100
Chl_TPI	2.81%	302.44	200	151.22	100	151.22	100
Chl_X5Piso	2.81%	302.44	200	151.22	100	151.22	100
Mit_Gly_tx	2.81%	302.44	200	151.22	100	151.22	100
Per_Glycolate_tx	2.81%	302.44	200	151.22	100	151.22	100
Total	42.57%	-	-	-	-	-	-

Chl_FerredoxinReductase: chloroplast ferredoxin reductase, Chl_G3Pdh: chloroplast glyceraldehyde-3-phosphate dehydrogenase, Chl_PGK: chloroplast phosphoglycerate kinase, Chl_Ru5Pk: chloroplast ribulose-5-phosphate kinase, Chl_PGlyPase: chloroplast phosphoglycolate phosphatase, Chl_RuBPOxid: chloroplast ribulose-1,5-bisphosphate oxygenase, Chl_TPI: chloroplast triose phosphate isomerase, Chl_X5Piso: chloroplast xylulose-5-phosphate isomerase, Mit_Gly_tx: mitochondrial glycolate transporter, Per_Glycolate_tx: peroxisomal glycolate transporter. For unshaded conditions, the vector of Flux Balance Analysis (FBA) models include FBA_P_CT_U (Peroxisomal Catalase Reaction as objective), FBA_P_96_U (Peroxisomal Glycolate Oxidase Reaction), and FBA_B3_U (Peroxisomal Catalase Reaction as objective and Peroxisomal Glycolate Oxidase Reaction as objective). Similarly, shaded conditions are represented by FBA_P_CT_S, FBA_P_96_S and FBA_B3_S, respectively.

7.3.4. Innovation for precision viticulture

The integration of systems biology techniques, exemplified in this study, provides an innovative framework to simulate pathways and metabolic processes within plants to practical applications across industries, including wine production, that can be incorporated in PV. Leveraging systems biology methodologies enables the simulation of the effects of environmental variables, such as shaded conditions, on plant metabolism, thereby facilitating investigations into their implications for final products like wine. The synthesis and accumulation of ascorbic acid, an antioxidant compound found

in grapes and wine, can offer a pathway for investigating the relationship among shade treatment, ROS dynamics, and ascorbic acid levels during grape fermentation. Ascorbic acid's role in alleviating oxidative stress and preserving wine quality underscores its importance in winemaking (Burbidge et al., 2021). Unravelling the impact of shaded conditions on ROS dynamics, ascorbic acid metabolism, and oxidant potency during fermentation unveils novel avenues for wine research, offering prospects to refine winemaking techniques and elevate the calibre of the end product.

This study underscores the importance of photorespiration in mitigating oxidative stress, yet further exploration is necessary to clarify whether this mechanism solely protects plants from ROS or serves additional physiological roles. Understanding the specific adaptive strategies employed by C3 plants to stabilise ROS damage and why photorespiration predominates over alternative mechanisms would extend the comprehension of plant stress resilience. Moreover, assessing the limitations and applicability of the metabolic engineering model used in this study is crucial for interpreting field data and capturing grapevine responses to light stress. Given the varied responses of grapevine cultivars to oxidative stress, searching into the genetic and physiological factors contributing to this diversity and its implications for viticultural practices is essential. Additionally, investigating the factors that activate shifts between iso- and anisohydric responses in grapevines, along with each strategy's associated risks and limitations, would advance the understanding of vine physiology and inform adaptive management practices. Isohydric plants maintain a constant leaf water potential by tightly regulating stomatal conductance, prioritising water conservation over photosynthesis. In contrast, anisohydric plants exhibit more variable leaf water potential, with stomatal conductance responding directly to changes in soil moisture, prioritising photosynthesis over water conservation (Pallioti & Poni, 2015; Chaves et al., 2010). Addressing these questions through future research endeavours will refine the perception of grapevine stress responses and enhance the sustainability and resilience of viticultural systems.

Furthermore, integrating hyperspectral sensors within the systems biology framework, particularly in PV, holds promise for enhancing vineyard practices with robust physiological foundations. This approach enables the analysis of hypotheses grounded in plant physiology, thereby improving cultural practices and product quality. Moreover, the combination of spectral analysis and systems biology opens avenues for metabolic mapping, providing knowledge into the plant's phenometaboloma and physiological processes. By leveraging this integrated approach, viticulturists can transcend the limitations of current data-driven PV approaches, often constrained by data

circumstances, and instead adopt information-rich strategies. This cooperative combination of technologies and methodologies enables the generation of high-throughput data, which empowers viticulturists to make data-driven decisions, fostering the development of more sustainable and resilient agricultural systems.

7.4. Conclusions

This study investigated the physiological and metabolic responses through systems biology of grapevines to varying light conditions, focusing specifically on the impact of shading compared to unshaded sunlight exposure. By combining hyperspectral data with biochemical parameters such as pigment concentrations and ROS levels and utilising an *in silico* photorespiration model, the study uncovered the adaptive strategies employed by grapevines in response to changes in light availability.

The observed differences in physiological parameters between shaded and unshaded grapevines underscore the correlation between light availability, plant water status, and metabolic activity. Grapevines exposed to unshaded conditions may demonstrate increased water uptake and transpiration rates. However, defence mechanisms to protect from excess light and heat may be activated, indicative of heightened metabolic activity and photosynthetic capacity.

The integration of hyperspectral data with systems biology techniques facilitated the simulation of plant responses under various light conditions, with a particular focus on the role of ROS in mediating stress responses. The enzymatic reactions associated with ROS metabolism highlighted the critical importance of mitigating oxidative stress under environmental conditions.

While this study effectively addressed its objectives within the systems biology framework, future research should delve deeper into the molecular mechanisms governing grapevine responses to shading and other environmental stressors. Leveraging advanced technologies such as hyperspectral sensors and robotics, in conjunction with physiological modelling, holds promise for refining PV practices and advancing the understanding of plant-environment interactions on a molecular level. However, this emerging field encompasses jargon from diverse areas, such as photonics and smart technologies, resulting in a somewhat insular interaction without a common language to bridge these distinct areas in this field of study.

The integration of systems biology methodologies provides a framework for simulating metabolic processes within plants, offering practical applications across industries, including wine production. Further investigations into the molecular mechanisms

underlying grapevine responses to shading and other stressors are warranted, with the potential to enhance precision viticulture practices and optimise crop productivity.

Acknowledgements

The authors also thank the wine company Poças Junior experimental facilities (and its Coordinator for Viticulture Maria Manuel Maia) for the facilities provided for fieldwork, as well as the Associação para o Desenvolvimento da Viticultura Duriense (ADVID) for the measurement of the predawn leaf water potential.

SECTION III

--- CHAPTER 8

DISCUSSION

8. Discussion

Traditional reliance on satellite or drone-mounted sensors in PA has data transferability and spatial applicability limitations. Low-cost proximal hyperspectral sensors emerge as a technological solution to these challenges in several agronomic processes. These are used in cultivating, managing, and producing crops to optimise crop yields, quality, and sustainability. These sensors, particularly exemplified by the low-cost Metbots prototype (Martins et al., 2019), offer high-dimensional, detailed, and immediate information, addressing cost concerns. The activities developed by this research team, as described in attachment A, various physiological parameters can be determined with this sensor. Preliminary results also showed potential in estimating SSC (Santos-Campos et al., 2023) and predawn leaf water status (David et al., 2023). Integrating these sensors into mobile platforms, such as robots, is highlighted for high-throughput mapping and the adoption of physiological PA.

Advancing PA, this thesis endeavours to reshape crop management and agronomic processes by bridging critical gaps in current models. A key emphasis is placed on introducing a more physiologically adherent modelling approach, recognising the interaction of plant physiological parameters in field settings. Ultimately, the aim is to elevate the precision and effectiveness of agricultural practices. The lack of sufficient physiological adherence within current models prompts the need for innovative methodologies to address the problem of transferability among years and locations of data-driven models (Wagenaar et al., 2021) on which the current PA is mainly based. In response to this challenge, the thesis explores the integration of Vis-NIR spectroscopy (reflectance and transmittance mode), LIBS, and POM techniques. These technologies are strategically employed to bridge the physiological modelling gap and improve the assessment of critical parameters of agronomic processes such as predawn leaf water potential, nutrient concentrations in leaves and fruits, and fruit maturation dynamics. Through this interdisciplinary approach, the thesis aims to contribute to the scientific understanding of agronomic processes and the practical implementation of sustainable and efficient PA strategies. Next, follow the scientific questions presented in the introduction of this thesis and how the case studies answered them.

8.1. Chapter 3: assessing ecophysiological plant parameters in the field

Chapter 3 delves into the assessment of ecophysiological plant parameters in the field. This multidimensional exploration integrates diverse methodologies and techniques

within PA, focusing on Case Study I. This case study examines predawn leaf water potential (PLWP) and its impact on post-veraison foliar pigments. The central aim is to predict predawn leaf water potential in a non-destructive way and decipher its correlation with absorbance bands tied to chlorophyll, xanthophylls, and water. Furthermore, this case study includes a literature review on applying Vis-NIR techniques for measuring PLWP. The comparison of this literature review with the results of this case study facilitated an enhanced approach, leveraging an SL-AI algorithm. This approach was applied to a highly robust dataset encompassing various cultivars, years, test sites, and irrigation regimes, improving statistical metrics and physiological hints.

The traditional measurement of water status in plants, specifically grapevines, is typically done by measuring the PLWP using a pressure chamber. However, this method is labour-intensive, destructive, and does not allow water status mapping (Pôças et al., 2017). This case study aims to solve these limitations using Vis-NIR spectral optical sensors to estimate non-destructively water potential. Using SL-AI based on CovM provides physiological adherence to the results. By analysing the spectral data, the SL-AI model can identify specific absorbance bands related to leaf pigments and water absorption, which serve as water stress indicators. This approach allows for a more detailed and accurate analysis of grapevine water status, considering different cultivars, test sites, and years.

Through the CovM, SL-AI uses spectral data to incorporate factors influencing water potential, including abiotic stressors like temperature, drought, and wind, alongside physiological aspects like foliar pigments and leaf ageing (Santesteban et al., 2019).

The results reveal a correlation between grapevine hydric status and specific spectral characteristics, providing perceptions of stress conditions affecting grapevines. Beyond validating the predictive power of PLWP, the study enhances the understanding of the plant's response to varying water availability, contributing to assessing ecophysiological parameters in vineyard environments.

Spectroscopic techniques and the SL-AI method discussed in Chapters 3 and 5 improve predictive capabilities for precise agronomic practices, including irrigation management. Understanding the dynamic interaction between plant pigments and water status, as revealed by spectroscopic studies (e.g., Rapaport et al., 2015), enriches the perception of crop health (Mazzetto et al., 2010) and stress conditions (Loggenberg et al., 2018). This interrelation of parameters, explored in Chapters 3, 4, and 5, emphasises the value of spectroscopy in monitoring changes and detecting early signs of abiotic (e.g., Janni et al., 2019) and biotic diseases (e.g., Reis Pereira et al., 2023) in crops.

The investigation, grounded in the context of PA, showcases the transformative potential of integrating advanced technologies and methodologies to understand better ecophysiological plant parameters in field conditions, an alternative to traditional, very time-laboured methods that lack physiological explanation. The multidimensional approach, exemplified by Case Study I, contributes to the broader goals of sustainable and efficient use of resources and improved agricultural practices.

8.2. Chapter 4: methodology for conducting a nutritional characterisation of various plant organs

In Chapter 4, the nutritional characterisation of different plant tissues is achieved through advanced techniques that provide insights into the plant/organ elemental composition and nutritional status, employing innovative methods within the framework of PA, as highlighted in the studies presented in Chapters 4 and 5.

Case Study II focuses on utilising LIBS for the elemental characterisation of the small amount of grapevine tissues. The primary objective is to characterise elemental concentrations in grape tissues using LIBS, offering perceptions of the grapes' metabolic constitution and quality parameters. The results suggest the potential for in situ applications (Arantes de Carvalho et al., 2018) and for characterising small samples, a task for which traditional methods like inductively coupled plasma optical emission spectrometry (ICP-OES) may not be optimised or even available by requiring a large amount of biological sample (e.g., Mitić et al., 2011).

While LIBS can provide more qualitative perceptions into elemental composition (Andrews et al., 2023a), Case Study III introduces a complementary approach through Vis-NIR spectroscopy for quantitative nutrient identification in apple and grape leaves. The objective is to quantify nutrients in these leaves using Vis-NIR spectroscopy data augmentation, addressing data scarcity in PA.

In this paper, data augmentation is implemented using a method called hybridisation that consists of creating a synthetic spectroscopy dataset (SSD) by combining two random samples to generate a new spectral profile and its corresponding nutrient composition. This process is done to expand the knowledge base and fill in gaps in the feature space.

The hybridisation operation involves taking two samples' average spectra or composition and creating a new sample with the average nutrient content, which is done to approximate the complex interference along the feature space and capture the multiscale interference nature of the spectral information. The hybridisation process allows for a more granular representation of nutrient gradients within the samples.

Data augmentation through hybridisation differs from deep learning because it does not involve using deep learning models but focuses on expanding the knowledge base and capturing the multiscale interference nature of spectral information. The methods used in this research are partial least squares (PLS), local partial least squares (LocPLS), and self-learning artificial intelligence (SL-AI). These methods are used to model the spectral data and quantify the nutrient composition. The key novelty of the study is demonstrating how the hybridisation dataset (SSD) can improve models and provide more detailed information to predict the real-world dataset (RWD) with higher accuracy. The SSD is statistically representative of the RWD characteristics and feature spaces, and it can enhance the performance of models such as LocPLS and SL-AI in macro and micronutrient quantification from leaf spectra.

Data augmentation through hybridisation is used in this study to overcome data scarcity and improve the accuracy of nutrient quantification in plants. The investigation successfully contributes to precision fertilisation by quantifying nutrient uptake in plant leaves. This approach facilitates on-site sampling with high efficiency and low cost, which is crucial due to the temporal dynamics of nutrients across various plant organs.

These case studies underscore the versatility and applicability of LIBS and Vis-NIR spectroscopy in elemental identification and nutrient quantification in grape and apple tissues. The insights gained from these studies contribute to the development of PA practices, offering tools for optimising crop management strategies according to the spatio-temporal variability of plant and field.

The transition to the quantitative analysis of macro and microelements in plant tissue (van Maarschalkerweerd & Husted, 2015) further emphasises the integration of advanced techniques. Data augmentation techniques can address data scarcity and expand the knowledge base (Lu et al., 2022). The obtained data becomes a practical resource for integrating precise information into systems biology, allowing simulations of the dynamic processes occurring during crop development. The acquired data incorporates precise understandings into systems biology, enabling simulations of the dynamic processes inherent in crop development. Moreover, the techniques developed hold potential for application in other studies where the sampling process is resource-intensive in terms of both human effort and financial resources. The approach outlined will facilitate a more efficient sampling design, thereby optimising the sampling process to yield richer information and mitigate overlapping and information gaps.

The application of LIBS and Vis-NIR spectroscopy discussed in Chapter 4 enhances the understanding of the dynamic relationship between plant nutritional status and maturation processes. These techniques, applied to grape and apple tissues, identify

and quantify macro and micro elements, providing physiological visions beyond traditional methods.

8.3. Chapter 5: assessing the evolution of maturation in fruits

Chapter 5 of the thesis delves into the process of evaluating the evolution of maturation in fruits, employing a novel, non-destructive tomographic analysis method. The primary aim is to monitor and predict maturation in tomatoes and grapes, accentuating the method's reliability, early indication capabilities, and operational simplicity.

Fruit maturation, as explored in this chapter, presents various physiological parameters. Key phenotype traits such as colour and texture indicate maturation stages (Xu et al., 2022). Additionally, the macroscopic insights derived from the size of fruits contribute to a broader understanding of the underlying physiological processes.

The integration of Vis-NIR spectroscopy and SL-AI, as discussed previously in Chapter 3, emerges as an enhancement in PA. This integration effectively utilises non-measurable parameters, predicting physiological states in grapevines and biophysical parameters in internal fruit tissues.

The introduced tomography-like approach responds to challenges in quantifying SSC in tomatoes compared to grapes. While providing a physiological understanding of internal tissue maturation, this method emphasises the dynamics of pigments throughout the maturation process.

The sequential application of LIBS and Vis-NIR spectroscopy, culminating in the tomography-like approach, can solve existing challenges and introduce novel avenues for evaluating and optimising crop maturation processes. The method's efficiency is underscored by its simplicity, featuring a POM device equipped with light, eliminating the need for radiation and enhancing safety, especially when contrasted with X-ray methods (Liu et al., 2020a).

Beyond its technical capabilities, the tomography-like method finds probable applications across various fruits, including kiwi, watermelon, avocado, pears, and apples. This broad application underscores its potential to produce quality and characteristics, impacting multiple segments of the agriculture industry (Truong et al., 2022).

Integrating tomography-like techniques with other technologies, such as SL-AI, forms a multi-modal approach. This holistic strategy facilitates a full fruit quality assessment, considering external and internal aspects. Such an approach aligns with the broader goal of understanding crop health and development. The tomography-like can also prevent the consumer from touching the fruit to assess its quality. The consumer touch may

promote physical damage and accelerate the maturation process in climacteric fruits (e.g. tomato), allowing possible disease development (Ahmad & Siddiqui, 2015).

Challenges in predicting SSC in tomatoes using spectroscopic techniques, as explored in Chapter 5, involve physiological factors and matrix effects (Silva et al., 2023b). Challenges arise in precisely forecasting SSC due to physiological factors, such as the dynamic characteristics of plant tissues and their reactions to environmental conditions (Ribalta-Pizarro et al., 2021). Matrix effects, influenced by factors like skin thickness and internal structures, contribute to variations in spectroscopic responses, demanding a nuanced understanding for improved precision (Martins et al., 2022).

The implications of spectroscopic techniques on the winery industry are connected throughout the thesis. Phenolic compounds, critical for wine flavour and colour, present unique challenges due to their diverse nature. The tomography-like method contributes data for understanding the distribution and composition of these compounds in grape tissues through site-specific management, allowing for determining specific characteristics for the crop quality. For example, acidity, a crucial determinant of wine quality, requires precise measurement during grape maturation (Giovenzana et al., 2015). The comparative analysis of tomography-like methods with other techniques emphasises spectroscopy's potential in providing accurate acidity measurements.

The tomography-like method's contribution extends beyond maturation indicators, supporting irrigation, fertilisation, and pruning practices. The method facilitates more informed decision-making in vineyard management according to a wine-making target by providing detailed physiological information (Tessarín et al., 2018).

8.4. Chapter 6: performing tomography-like assessments in the field to evaluate maturation

Chapter 6 explores the application of tomography-like assessments to evaluate maturation dynamics in grape cultivation, focusing on Case Study VI conducted in the Douro wine region. This study aimed to implement tomography-like assessments in a grapevine parcel to create metabolic maps for tissue maturation and enable site-specific management strategies tailored to grape ripening dynamics.

The tomography-like method involves reconstructing spectral data from internal grape tissues, allowing for predicting physicochemical parameters such as chlorophyll and anthocyanin content. By leveraging spectral information, this non-destructive approach provides insights into grape maturation without causing damage to the grapes themselves (Martins et al., 2022).

Integration with terrain analysis further enhances the utility of tomography-like assessments by facilitating the delineating of specific zones within vineyards. This spatial delineation enables tailored management practices, such as irrigation and fertilisation adjustments, based on predicted physicochemical parameters across different areas of the vineyard terrain (Arantes de Carvalho et al., 2018).

However, limitations associated with tomography-like methods include their reliance on spectral data and the inherent constraints imposed by the available wavelength range of the equipment used for data collection. Additionally, factors such as grape variety and vineyard location may influence the study results, potentially limiting their generalisability to other contexts (Mitić et al., 2011).

Despite these limitations, tomography-like assessments promise to optimise grapevine management practices and improve wine quality. By providing detailed insights into grape maturation dynamics, this approach can inform decisions regarding harvest timing, grape processing techniques, and overall vineyard management strategies (Tessarini et al., 2018).

Metabolic mapping and practical implications for crop management

Metabolic mapping involves the systematic analysis and visualisation of metabolic processes within fruit tissues, facilitated by tomography-like methods. Unlike conventional imaging techniques that primarily capture structural features, tomography-like assessments focus on detailed spectral information, enabling the identification and quantification of specific metabolites essential for creating metabolic maps (Gerlin et al., 2021).

These metabolic maps provide insights into biochemical pathways and metabolic processes underlying grape ripening, offering a comprehensive understanding of changes in metabolites such as sugars, acids, phenolics, and aroma compounds (Bavaresco et al., 2012). By elucidating these metabolic dynamics, farmers can optimise vineyard management practices, grape harvest timing, and wine-making techniques to achieve desired grape and wine characteristics (Gago et al., 2017).

Furthermore, metabolic mapping can contribute to sustainable agriculture by minimising resource inputs and reducing wastage through targeted interventions based on metabolic targets (Lu et al., 2022). By unravelling the complex biochemical pathways within fruits, tomography-like methods promote the establishment of precision agriculture practices that enhance sustainability and resilience in the face of evolving agricultural challenges (Lecourieux et al., 2020).

8.5. Chapter 7: the relationship between precision agriculture and systems biology

Chapter 7 delves into the symbiotic relationship between PA and systems biology, drawing from insights gathered throughout the thesis. Central to this discussion is Case Study VII, which aimed to demonstrate through system biology the complex relationship between agricultural practices, specifically shade treatments, and grapevine metabolism through the integration of PA and systems biology methodologies.

The case study in Portugal's Douro wine region demonstrates the potential of systems biology technology in optimising farming practices by synthesising physiological and quality-related insights from grapevine responses to shade treatments.

The results underscore the impacts of shade treatments on grapevine physiology, unveiling interactions that shape internal leaf structure, elemental composition, and biochemical content, particularly during the critical post-veraison phase. Such insights align with existing literature, such as the work by Palliotti and Poni (2015), which underscores the importance of canopy management techniques in vineyard operations. Integral to this multidimensional approach is predicting and simulating metabolic pathways using computational models rooted in systems biology principles. By integrating information about enzymes and metabolic routes associated with shade utilisation during post-veraison, these virtual phenotypes offer actionable insights for optimising vineyard management practices tailored to specific physiological needs (Lecourieux et al., 2020).

Validation of computational models is paramount for advancing PA methodologies. By leveraging bioinformatics tools and high-throughput sequencing techniques, biological models can be constructed to elucidate complex interactions between genetic factors, environmental conditions, and agronomic practices. Moreover, integrating experimental data from field trials supports the predictive accuracy of these models under real-world conditions (Pazhamala et al., 2021).

This study has implications for sustainable vineyard management, emphasising the importance of canopy manipulation and light exposure regulation during growth phases. Optimising shade utilisation based on grapevine physiological requirements enhances photosynthetic efficiency, balanced nutrient uptake, and improved grape quality.

Moreover, integrating systems biology principles into PA approaches facilitates informed decision-making tailored to the specific needs of individual plants or crop zones (Wurtzel et al., 2019). Detailed insights into metabolic routes, gene expressions, and physiological responses enable targeted interventions such as irrigation scheduling, fertilisation

practices, and pest management, enhancing resource efficiency and promoting sustainable agricultural practices.

At its core, systems biology represents an interdisciplinary approach that endeavours to unravel the complexities of biological systems. Examining components and interactions as an integrated whole offers a holistic understanding of how crops interact with their environment. Integrating systems biology with spectral data promises high-resolution insights into plant-environment interactions, paving the way for a deeper understanding of plant responses and advancing the precision and sustainability of modern agriculture.

8.6. Enhancing consumer awareness and mitigating food waste: a holistic perspective on advanced methodologies for precision agriculture rooted in systems biology

The efficacy of tomography-like methodologies in fruit production is elucidated through preceding research endeavours outlined in earlier sections. These studies have demonstrated the practical utility of such approaches, initially evaluated within grapevine and tomato cultivation domains. However, the applicative scope of these methodologies transcends the boundaries of specific fruit typologies, with broader implications covering the entire fruit and vegetable value chain (Truong et al., 2022).

Beyond the confines of grapevine and tomato cultivation, these methodologies hold promise for deployment across a spectrum of fruits, including but not limited to apples and pears. Moreover, their potential extends across the entire range of activities within the fruit and vegetable value chain, encompassing processing facilities, where their utility in evaluating internal characteristics proximate to seeds can be particularly accentuated. Additionally, the tomography-like approach contributes to shelf-life management. It informs consumers about fruit quality, preventing them from touching the fruits and vegetables and preventing possible physical damage and contamination.

Consumer awareness is pivotal in addressing food waste and enhancing sustainable agricultural practices. The ability of the tomography-like method to assess internal tissues and biochemical properties becomes instrumental in this information paradigm (Truong et al., 2022; Stiletto & Trestini, 2021). Allowing consumers to make informed choices and avoiding touch that can injure the fruit aligns with the broader goals of fostering a discerning consumer base contributing to sustainable and responsible agricultural practices.

Moreover, exploring predictive modelling and artificial intelligence in conjunction with tomography-like methods can open avenues for advancements. Future research

endeavours should delve into refining predictive models, improving the accuracy of physiological parameter predictions, and exploring the integration of emerging technologies for enhanced decision support systems in the context of PA.

The impact of tomography-like on shelf-life extension and logistic challenges can be minimised through advancements in predictive modelling and artificial intelligence. These methods contribute to efficient supply chain management (Dora et al., 2021). Technology is crucial in determining optimal storage times, preventing fungal formation, and restructuring distribution, aligning with contemporary demands for food security and waste reduction.

Implementing the techniques discussed in real-world scenarios requires strategic planning incorporating market studies evaluating consumer expectations and preferences. Aligning these technologies with existing distribution chains is imperative. Partnerships between research institutions, industry stakeholders, and regulatory associations will be instrumental in navigating the complexities of technology integration into existing agricultural practices.

Reducing food waste, as proposed by initiatives like the Farm to Fork Strategy (European Commission, 2020), brings multifaceted economic, environmental, and social benefits. The presented case studies, especially in tomatoes and grapes, serve as proof of concept for tomography-like methods and underscore their potential to reduce food waste.

Recommendations for further research and technology development should focus on refining and expanding the applications of tomography-like techniques. Investigations into additional crop types, varied environmental conditions, and potential limitations will contribute to a better understanding of the technology's capabilities and constraints.

The alignment with the 2030 agenda for sustainable development (UN DESA, 2016) features the commitment to transformative action. This thesis aligns precisely with Goal 12, emphasising responsible consumption and production, and resonates with the core principles of the discussed technologies. By facilitating informed decision-making, reducing food waste, and optimising resource usage, these technologies directly contribute to the sustainable development goals outlined in Agenda 2030.

Similarly, aligning with the European Farm to Fork Strategy (European Commission, 2020) reinforces the commitment to improving food systems. With its focus on environmental impact, biodiversity conservation, and public health promotion, this strategy resonates with the sustainable agricultural practices promoted by the discussed technologies.

The contribution of the discussed technologies to global goals extends beyond the reduction of food waste to the mitigation of climate change. As highlighted in various studies (e.g., Crippa et al., 2021), food loss and waste contribute to greenhouse gas emissions, and reducing these losses is a crucial component in climate change mitigation efforts. Moreover, the role of the discussed technologies in advancing sustainable agriculture is paramount. By providing real-time data, enabling PA, and promoting efficient resource management, these technologies contribute to the overarching goal of creating a more resilient and sustainable food system (European Commission, 2021).

8.7. Perspectives and challenges

In the search for advancing PA and contributing to the understanding of fruit maturation, this thesis encountered several limitations inherent in the applied methods and technologies. It is crucial to critically discuss these limitations to provide a perspective on the scope and potential challenges of the research conducted.

One limitation is applying data-driven methods, particularly in predicting biophysical variables like SSC in fruits. While these methods showcase success in predicting outcomes, the challenge remains in achieving a comprehensive physiological explanation, especially in complex maturation processes. The matrix effect of compounds in fruit tissues, as exemplified in the difficulty of predicting SSC in tomatoes, underscores the limitations of relying solely on data-driven approaches.

Though innovative, the set-up of the tomography-like method faces limitations in obtaining spectral transmittance information from specific tissues, which poses a challenge in achieving a complete characterisation of internal fruit structures, particularly in cases where tissues hinder the transmission of light. It necessitates further development work to enhance the method's ability to capture detailed spectral information from various other types of fruit (e.g. apple, pear, watermelon).

Additionally, combining omics work in crops within field conditions presents challenges. The complexity of plant biology systems and the dynamic nature of field environments introduce uncertainties and variations that may affect the accuracy and reliability of the results (Mishra et al., 2019). Understanding the entire biological system of plants under these conditions remains challenging.

The case studies, encompassing element detection through LIBS analyses on grapes, Vis-NIR on apple leaves, and applying tomography-like techniques, show that this thesis operates at a relatively low technology readiness level (TRL). It predominantly serves as a conceptual proof, spotlighting the potential of these methodologies in PA. While the

results show promise, rigorous validation is essential for a quantitative approach, ensuring their robustness and applicability across diverse agricultural contexts.

Addressing these limitations and challenges requires strategic considerations for future studies. Investments are essential to overcome the tomography-like method's hardware set-up constraints. Improvement in hardware and biochemical validation processes would enhance the method's accuracy and broaden its applicability across various fruits and maturation stages.

Furthermore, financial investment in projects that advance the concept proof presented in this thesis is imperative, as it would elevate the TRL of the methodologies, enabling a quicker transition from conceptual validation to practical implementation in PA. Such investments would catalyse transformative changes in the field, encouraging the development of more reliable and scalable technologies for sustainable and informed agricultural practices and making the transition from academic to corporate business shorter.

The ongoing Omicbots project (<https://www.fc.up.pt/omicbots/>) aims to combine in situ and dry-lab methodologies to enhance agricultural practices, as Figure 8-1 demonstrates.

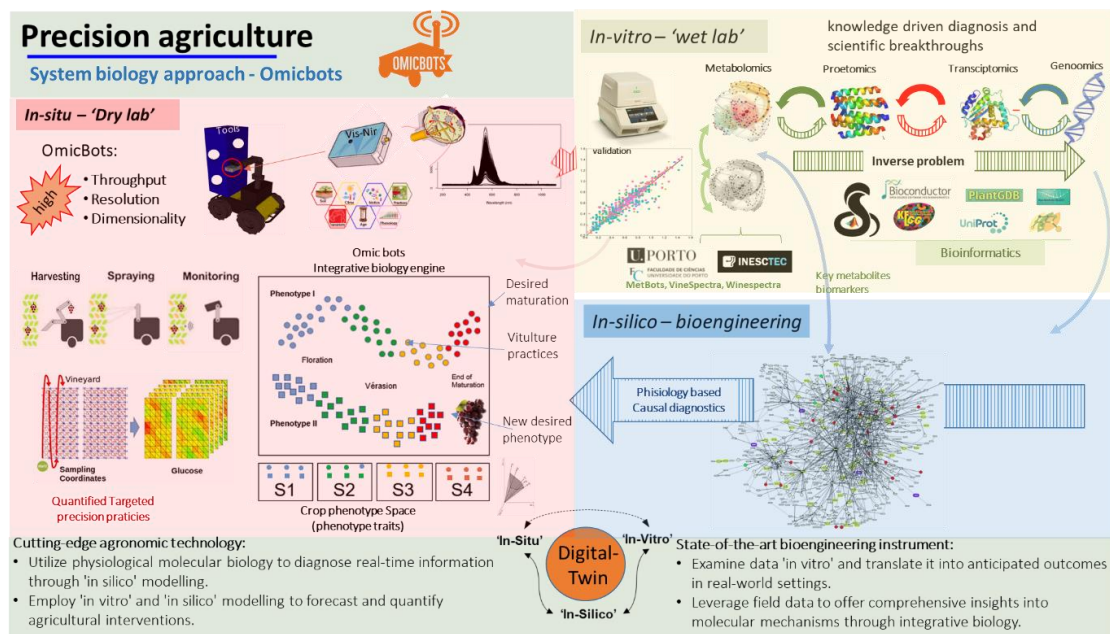


Figure 8-1. Omicbots: integrating in situ, in vitro, and in silico approaches for precision agriculture.

Omicbots employs high-throughput, high-resolution, and high-dimensionality techniques to analyse agricultural traits such as harvesting and spraying monitoring in the dry-lab setting. Omicbots can characterise biophysical parameters like glucose by utilising

vineyard sampling coordinates, providing experimental targets that inform precision practices. The integrative biology engine of Omicbots allows for the characterisation of different phenotypes related to the flowering, veraison, and maturation stages. This information enables approaches tailored to desired maturation outcomes, specific viticultural practices, and the creation of new desired phenotypes.

Omicbots can promote metabolomic, proteomic, transcriptomic, and genomic validations in the wet lab environment. These analyses provide knowledge-driven diagnoses and scientific breakthroughs. Combined with in-silico bioengineering, Omicbots can generate physiological-based causal diagnoses, offering insights into the molecular mechanisms underlying agricultural processes.

Integrating in vitro and in silico approaches allows Omicbots to analyse data obtained in the field and translate it into actionable insights. By leveraging field results, Omicbots can provide integrative biology insights on molecular mechanisms, bridging the gap between in situ, in vitro, and in silico variables.

The Omicbots project represents a high-end agronomic and bioengineering tool that combines in situ, in vitro, and in silico methodologies to advance precision agriculture. Its ability to diagnose in vivo information through physiological molecular biology, predict and quantify agricultural practices, and provide integrative biology insights underscores its potential to revolutionise agricultural management practices. Omicbots pave the way for more sustainable, efficient, and tailored agricultural practices by continuously integrating systems biology approaches.

CHAPTER 9

CONCLUSION

9. Conclusion

Exploring limitations in traditional PA approaches, particularly those relying on satellite and drone-mounted sensors. It provided an understanding of the challenges in capturing detailed physiological information. The introduction and evaluation of point-of-measurement (POM) emerged as a transformative solution, overcoming the shortcomings of traditional methods and delivering real-time and detailed data essential for precise viticultural management.

This thesis has effectively achieved its objectives in assessing ecophysiological plant parameters in the field, developing a methodology for conducting a nutritional characterisation of various plant organs, developing a technology for evaluating the evolution of maturation in fruits, creating metabolic maps to perform to assess the maturation in the field and to demonstrate the how systems biology can be incorporated into PA.

This thesis examines cutting-edge technologies within PA, unravelling multifaceted issues across chapters and papers. Each section contributes to the evolution of PA, with Chapter 3 shedding light on predawn leaf water potential, thus enriching the understanding of plant physiology. Extending the application of laser-induced breakdown spectroscopy (LIBS) to grapes is expounded in Chapter 4, while Paper 3 broadens the scope to Vis-NIR analysis of apple and grape leaves. A pivotal milestone is achieved in Chapter 5 with the introduction of the SpecTom project, resulting in a patent for a tomography-like method, as elaborated in Papers 4 and 5 concerning its application in tomatoes and grapes, respectively. Chapter 6 delves into mapping metabolic parameters in the Douro region, highlighting the technology's potential for offering detailed physiological insights for agricultural management. Chapter 7 further investigates the creation of virtual phenotypes in a Douro region shade case, unravelling the relationship between crops and their environment. These findings promote advancement in PA by addressing critical questions about plant parameters, nutritional characterisations, and the evolution of fruit maturation. Integrating PA with systems biology enriches the understanding of plant physiology mechanisms, effectively bridging the gap between conventional practices and advanced scientific approaches.

The methodologies and technologies presented in this thesis have emerged as potential game-changers in agricultural practices, offering implications for sustainability efforts and food waste reduction. For instance, tomography-like assessments enhance the understanding of fruit maturation dynamics and provide practical insights into optimising vineyard management practices. These assessments enable targeted irrigation, fertilisation, and disease management interventions by precisely visualising the spatial

distribution of critical metabolites within fruit tissues, promoting resource-efficient and sustainable agricultural practices.

Moreover, the findings of this thesis have implications for reducing food waste along the entire fruit and vegetable value chain. The ability of tomography-like methods to assess internal tissues and biochemical properties empowers consumers to make informed choices, thereby minimising wastage and fostering responsible consumption. Additionally, the integration of predictive modelling and artificial intelligence with these methods holds promise for streamlining supply chain management, optimising storage times, and preventing pathological attacks, all of which contribute to reducing post-harvest losses and enhancing food security.

However, it is essential to acknowledge the challenges and limitations associated with implementing these findings in real-world scenarios. Factors such as the complexity of plant biology systems and the dynamic nature of field environments introduce uncertainties that may affect the reliability of results. Furthermore, transitioning from conceptual validation to practical implementation in PA requires strategic planning, financial investment, and collaboration among various stakeholders.

Further research is recommended to refine and expand the applications of these methodologies, particularly in addressing additional crop types and varied environmental conditions, due to limitations inherent in current technologies and the methodologies' low technology readiness level (TRL). The conclusion emphasises the need for continued research, technological refinement, and strategic implementation, focusing on overcoming hardware constraints and increasing the TRL to ensure the robustness and scalability of these technologies in diverse agricultural contexts.

Strategic measures must be taken to address the limitations identified in implementing the findings of this study in real-world scenarios. Firstly, interdisciplinary collaboration among researchers, industry stakeholders, and policymakers is crucial to developing strategies for overcoming the challenges posed by the dynamic nature of field environments and the complexity of plant biology systems. This collaboration can facilitate the integration of cutting-edge technologies with existing agricultural practices, ensuring their practical applicability and scalability. Moreover, financial investment in research and development projects aimed at refining and validating the methodologies presented in this thesis is imperative. By elevating these methodologies' TRL, they can undergo rigorous testing and validation in diverse agricultural contexts, paving the way for their widespread adoption. Furthermore, strategic planning and regulatory support are essential to cross the complexities of technology integration into existing agricultural systems. Addressing these challenges through collaborative efforts and strategic

initiatives can unlock PA’s full potential in achieving sustainable and efficient agricultural management practices.

By addressing these challenges and leveraging the transformative potential of research findings, steps can be made towards achieving sustainable agricultural practices and reducing food waste, aligning with the global goals outlined in the 2030 Agenda for Sustainable Development and the European Farm to Fork Strategy. The thesis contributes to goals such as Zero Hunger, Responsible Consumption and Production, and Climate Action. It also supports the Farm to Fork Strategy by ensuring transparent quality assurance and empowering consumers with detailed information. Additionally, the thesis aids in reducing food waste and mitigating climate change, directly aligning with the objectives of these strategic initiatives.

This thesis advances academic knowledge and lays a solid foundation for the future of PA with practical implications for sustainable and efficient agricultural management.

References

- ©JAXA/METI ALOS PALSAR. (2007). Alpsrp074960810 (Publication no. 10.5067/Z97HFCNKR6VA). Retrieved 10 March 2024 <https://asf.alaska.edu>
- Abdullahi, S. B., Muangchoo, K., Abubakar, A. B., Ibrahim, A. H., & Aremu, K. O. (2021). Data-driven ai-based parameters tuning using grid partition algorithm for predicting climatic effect on epidemic diseases. *IEEE Access*, 9, 55388-55412. doi: 10.1109/access.2021.3068215
- Abdullahi, S. B., Khunpanuk, C., Bature, Z. A., Chiroma, H., Pakkaranang, N., Abubakar, A. B., & Ibrahim, A. H. (2022). Biometric information recognition using artificial intelligence algorithms: A performance comparison. *IEEE Access*, 10, 49167-49183. doi: 10.1109/access.2022.3171850
- Acosta, M., Rodríguez-Carretero, I., Blasco, J., de Paz, J. M., & Quiñones, A. (2023). Non-destructive appraisal of macro- and micronutrients in persimmon leaves using vis/nir hyperspectral imaging. *Agriculture*, 13(4). doi: 10.3390/agriculture13040916
- Adams, D. O. (2006). Phenolics and ripening in grape berries. *American Journal of Enology and Viticulture*, 57(3), 249-256. doi: 10.5344/ajev.2006.57.3.249
- Adams, P., & Winsor, G. W. (2015). Some responses of glasshouse tomatoes to boron. *Journal of Horticultural Science*, 49(4), 355-363. doi: 10.1080/00221589.1974.11514590
- Adams, W. W., Stewart, J. J., & Demmig-Adams, B. (2018). Photosynthetic modulation in response to plant activity and environment. In W. W. Adams Iii & I. Terashima (Eds.), *The leaf: A platform for performing photosynthesis* (pp. 493-563). Cham: Springer International Publishing.
- Agati, G., Traversi, M. L., & Cerovic, Z. G. (2008). Chlorophyll fluorescence imaging for the noninvasive assessment of anthocyanins in whole grape (vitis vinifera L.) bunches. *Photochemistry and Photobiology*, 84(6), 1431-1434. doi: 10.1111/j.1751-1097.2008.00424.x
- Agati, G., D'Onofrio, C., Ducci, E., Cuzzola, A., Remorini, D., Tuccio, L., . . . Mattii, G. (2013). Potential of a multiparametric optical sensor for determining in situ the maturity components of red and white vitis vinifera wine grapes. *Journal of Agricultural and Food Chemistry*, 61(50), 12211-12218. doi: 10.1021/jf405099n
- Ahmad, A., Noor, S. E., Cartujo Cassinello, P., & Martos Núñez, M. V. (2022). Artificial intelligence (ai) as a complementary technology for agricultural remote sensing (rs) in plant physiology teaching.
- Ahmad, M. S., & Siddiqui, M. W. (2015). Factors affecting postharvest quality of fresh fruits *Postharvest quality assurance of fruits* (pp. 7-32).
- Al-Saddik, H., Simon, J. C., & Cointault, F. (2017). Development of spectral disease indices for 'flavescence doree' grapevine disease identification. *Sensors (Basel)*, 17(12). doi: 10.3390/s17122772
- Alzahrani, F. S., & Abbas, I. A. (2019). Analytical estimations of temperature in a living tissue generated by laser irradiation using experimental data. *J Therm Biol*, 85, 102421. doi: 10.1016/j.jtherbio.2019.102421
- Ampatzidis, Y., & Partel, V. (2019). Uav-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing*, 11(4). doi: 10.3390/rs11040410
- Anastasiou, E., Xanthopoulos, G., Templalexis, C., Lentzou, D., Panitsas, F., Mesimeri, A., . . . Fountas, S. (2022). Climatic indices as markers of table-grapes postharvest quality: A prediction exercise. *Smart Agricultural Technology*, 2. doi: 10.1016/j.atech.2022.100059
- Andaur, J. E., Guesalaga, A. R., Agosin, E. E., Guarini, M. W., & Irrarrázaval, P. (2004). Magnetic resonance imaging for nondestructive analysis of wine grapes. *Journal of Agricultural and Food Chemistry*, 52(2), 165-170. doi: 10.1021/jf034886c

- Andrade, D. F., & Pereira-Filho, E. R. (2016). Direct determination of contaminants and major and minor nutrients in solid fertilizers using laser-induced breakdown spectroscopy (libs). *J Agric Food Chem*, 64(41), 7890-7898. doi: 10.1021/acs.jafc.6b04028
- Andrews, H. B., Martin, M. Z., Wymore, A. M., & Kalluri, U. C. (2023a). Rapid in situ nutrient element distribution in plants and soils using laser-induced breakdown spectroscopy (libs). *Plant and Soil*. doi: 10.1007/s11104-023-05988-7
- Andrews, H. B., Martin, M. Z., Wymore, A. M., & Kalluri, U. C. (2023b). Rapid in situ nutrient element distribution in plants and soils using laser-induced breakdown spectroscopy (libs). *Plant and Soil*, 1-10. doi: 10.1007/s11104-023-05988-7
- Arantes de Carvalho, G. G., Bueno Guerra, M. B., Adame, A., Nomura, C. S., Oliveira, P. V., Pereira de Carvalho, H. W., . . . Krug, F. J. (2018). Recent advances in libs and xrf for the analysis of plants. *Journal of Analytical Atomic Spectrometry*, 33(6), 919-944. doi: 10.1039/c7ja00293a
- Arshavsky-Graham, S., & Segal, E. (2022). Lab-on-a-chip devices for point-of-care medical diagnostics. *Adv Biochem Eng Biotechnol*, 179, 247-265. doi: 10.1007/10_2020_127
- Aurich, M. K., Fleming, R. M., & Thiele, I. (2016). Metabotools: A comprehensive toolbox for analysis of genome-scale metabolic models. *Front Physiol*, 7, 327. doi: 10.3389/fphys.2016.00327
- Ayenew, B., Degu, A., Manela, N., Perl, A., Shamir, M. O., & Fait, A. (2015). Metabolite profiling and transcript analysis reveal specificities in the response of a berry derived cell culture to abiotic stresses. *Front Plant Sci*, 6, 728. doi: 10.3389/fpls.2015.00728
- Azadshahraki, F., Jamshidi, B., & Sharabiani, V. R. (2018). Non-destructive determination of vitamin c and lycopene contents of intact cv. Newton tomatoes using nir spectroscopy. *Yuzuncu Yil University Journal of Agricultural Sciences*, 28(4), 389-397.
- Baluja, J., Diago, M. P., Balda, P., Zorer, R., Meggio, F., Morales, F., & Tardaguila, J. (2012a). Assessment of vineyard water status variability by thermal and multispectral imagery using an unmanned aerial vehicle (uav). *Irrigation Science*, 30(6), 511-522. doi: 10.1007/s00271-012-0382-9
- Baluja, J., Diago, M. P., Goovaerts, P., & Tardaguila, J. (2012b). Spatio-temporal dynamics of grape anthocyanin accumulation in a tempranillo vineyard monitored by proximal sensing. *Australian Journal of Grape and Wine Research*, 18(2), 173-182. doi: 10.1111/j.1755-0238.2012.00186.x
- Bar-Sinai, Y., Hoyer, S., Hickey, J., & Brenner, M. P. (2019a). Learning data-driven discretizations for partial differential equations. *Proc Natl Acad Sci U S A*, 116(31), 15344-15349. doi: 10.1073/pnas.1814058116
- Bar-Sinai, Y., Hoyer, S., Hickey, J., & Brenner, M. P. (2019b). Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences of the United States of America*, 116(31), 15344-15349. doi: 10.1073/pnas.1814058116
- Barnard, Y., Strever, A., Bosman, G., & Poblete-Echeverría, C. (2019). Fast and non-destructive method for estimating grapevine water status. *Acta Horticulturae*(1253), 413-420. doi: 10.17660/ActaHortic.2019.1253.54
- Barroso, T., Ribeiro, L., Gregório, H., Santos, F., & Martins, R. (2021a). Quantification of red blood cells and white blood cells using scattering information extraction from vis-swnir spectra: Towards spectroscopy point-of-care cell counts. *CSAC2021: 1st International Electronic Conference on Chemical Sensors and Analytical Chemistry*. doi: <https://doi.org/10.3390/CSAC2021-10436>
- Barroso, T., Ribeiro, L., Gregório, H., Santos, F., & Martins, R. (2021b). Visible-near infrared platelets count: Towards thrombocytosis point-of-care diagnosis.

- CSAC2021: 1st International Electronic Conference on Chemical Sensors and Analytical Chemistry. doi: <https://doi.org/10.3390/CSAC2021-10435>
- Barroso, T., Ribeiro, L., Gregório, H., Santos, F., & Martins, R. (2021c). Point-of-care vis-swnir spectroscopy towards reagent-less hemogram analysis. *Sensors and Actuators B: Chemical*, 343. doi: 10.1016/j.snb.2021.130138
- Barroso, T., Ribeiro, L., Gregório, H., Santos, F., & Martins, R. (2021d). Feasibility of total white blood cell counts by visible-near infrared spectroscopy. *CSAC2021: 1st International Electronic Conference on Chemical Sensors and Analytical Chemistry*. doi: <https://doi.org/10.3390/CSAC2021-10434>
- Barroso, T. G., Ribeiro, L., Gregório, H., Monteiro-Silva, F., Neves dos Santos, F., & Martins, R. C. (2022). Point-of-care using vis-nir spectroscopy for white blood cell count analysis. *Chemosensors*, 10(11). doi: 10.3390/chemosensors10110460
- Barth, R., Hemming, J., & Van Henten, E. J. (2020). Optimising realism of synthetic images using cycle generative adversarial networks for improved part segmentation. *Computers and Electronics in Agriculture*, 173. doi: 10.1016/j.compag.2020.105378
- Barthod, S., Cerovic, Z., & Epron, D. (2007). Can dual chlorophyll fluorescence excitation be used to assess the variation in the content of uv-absorbing phenolic compounds in leaves of temperate tree species along a light gradient? *Journal of Experimental Botany*, 58(7), 1753-1760. doi: 10.1093/jxb/erm030
- Bavaresco, L., Mattivi, F., De Rosso, M., & Flamini, R. (2012). Effects of elicitors, viticultural factors, and enological practices on resveratrol and stilbenes in grapevine and wine. *Mini-Reviews in Medicinal Chemistry*, 12(13), 1366-1381. doi: <http://dx.doi.org/10.2174/13895575112091366>
- Begg, J. E., & Turner, N. C. (1970). Water potential gradients in field tobacco. *Plant Physiology*, 46(2), 343-346. doi: 10.1104/pp.46.2.343
- Beleites, C., & Sergo, V. (2020). Hyperspec: A package to handle hyperspectral data sets in r. Retrieved from <https://github.com/cbeleites/hyperSpec>
- Ben Ghazlen, N., Cerovic, Z. G., Germain, C., Toutain, S., & Latouche, G. (2010). Non-destructive optical monitoring of grape maturation by proximal sensing. *Sensors (Basel)*, 10(11), 10040-10068. doi: 10.3390/s101110040
- Benito, A., Romero, I., Domínguez, N., García-Escudero, E., & Martín, I. (2013). Leaf blade and petiole analysis for nutrient diagnosis in *Vitis vinifera* L. Cv. Garnacha tinta. *Australian Journal of Grape and Wine Research*, 19(2), 285-298. doi: 10.1111/ajgw.12022
- Berni, R., Luyckx, M., Xu, X., Legay, S., Sergeant, K., Hausman, J.-F., . . . Guerriero, G. (2019). Reactive oxygen species and heavy metal stress in plants: Impact on the cell wall and secondary metabolism. *Environmental and Experimental Botany*, 161, 98-106. doi: 10.1016/j.envexpbot.2018.10.017
- Bertin, N., Buret, M., & Gary, C. (2015). Insights into the formation of tomato quality during fruit development. *The Journal of Horticultural Science and Biotechnology*, 76(6), 786-792. doi: 10.1080/14620316.2001.11511446
- Bertin, N., & Génard, M. (2018). Tomato quality as influenced by preharvest factors. *Scientia Horticulturae*, 233, 264-276. doi: 10.1016/j.scienta.2018.01.056
- Bertoldi, D., Larcher, R., Bertamini, M., Otto, S., Concheri, G., & Nicolini, G. (2011). Accumulation and distribution pattern of macro- and microelements and trace elements in *Vitis vinifera* L. Cv. Chardonnay berries. *J Agric Food Chem*, 59(13), 7224-7236. doi: 10.1021/jf2006003
- Bora, F. D., Bunea, C. I., Chira, R., & Bunea, A. (2020). Assessment of the quality of polluted areas in northwest Romania based on the content of elements in different organs of grapevine (*Vitis vinifera* L.). *Molecules*, 25(3). doi: 10.3390/molecules25030750

- Borgogno-Mondino, E., Novello, V., Lessio, A., & de Palma, L. (2018). Describing the spatio-temporal variability of vines and soil by satellite-based spectral indices: A case study in apulia (south italy). *International Journal of Applied Earth Observation and Geoinformation*, 68, 42-50. doi: 10.1016/j.jag.2018.01.013
- Bramley, R. G. V., Siebert, T. E., Herderich, M. J., & Krstic, M. P. (2017). Patterns of within-vineyard spatial variation in the 'pepper' compound rotundone are temporally stable from year to year. *Australian Journal of Grape and Wine Research*, 23(1), 42-47. doi: 10.1111/ajgw.12245
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. doi: 10.1023/A:1010933404324
- Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M. H., Ruget, F., Nicoullaud, B., . . . Delécolle, R. (1998). Stics: A generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie*, 18(5-6), 311-346.
- Broadley, M., Brown, P., Cakmak, I., Rengel, Z., & Zhao, F. (2012). Function of nutrients *Marschner's mineral nutrition of higher plants* (pp. 191-248).
- Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). *The balanced accuracy and its posterior distribution*. Paper presented at the 2010 20th International Conference on Pattern Recognition.
- Brunori, E., Maesano, M., Moresi, F., Scarascia Mugnozza, G., & Biasi, R. (2019). *Towards sustainable viticulture: Key role of vineyard's precision monitoring*. Paper presented at the Proceedings of the International Symposium on Technologies for Smart City, Malaga, Spain.
- Buesa, I., Pérez, D., Castel, J., Intrigliolo, D. S., & Castel, J. R. (2017). Effect of deficit irrigation on vine performance and grape composition of vitis vinifera L. Cv. Muscat of alexandria. *Australian Journal of Grape and Wine Research*, 23(2), 251-259. doi: 10.1111/ajgw.12280
- Burbidge, C. A., Ford, C. M., Melino, V. J., Wong, D. C. J., Jia, Y., Jenkins, C. L. D., . . . Sweetman, C. (2021). Biosynthesis and cellular functions of tartaric acid in grapevines. *Front Plant Sci*, 12, 643024. doi: 10.3389/fpls.2021.643024
- Bwambale, E., Abagale, F. K., & Anornu, G. K. (2023). Data-driven model predictive control for precision irrigation management. *Smart Agricultural Technology*, 3. doi: 10.1016/j.atech.2022.100074
- Canas, R. A., Yesbergenova-Cuny, Z., Simons, M., Chardon, F., Armengaud, P., Quillere, I., . . . Hirel, B. (2017). Exploiting the genetic diversity of maize using a combined metabolomic, enzyme activity profiling, and metabolic modeling approach to link leaf physiology to kernel yield. *Plant Cell*, 29(5), 919-943. doi: 10.1105/tpc.16.00613
- Candar, S., Açıkbaş, B., Ekiz, M., Zobar, D., Korkutal, İ., & Bahar, E. (2021). Influence of water scarcity on macronutrients contents in young leaves of wine grape cultivars. *Ciência e Técnica Vitivinícola*, 36(2), 104-115. doi: 10.1051/ctv/ctv20213602104
- Carmo de Carvalho e Martins, M. d., da Silva Santos Oliveira, A. S., da Silva, L. A. A., Primo, M. G. S., & de Carvalho Lira, V. B. (2022). Biological indicators of oxidative stress [malondialdehyde, catalase, glutathione peroxidase, and superoxide dismutase] and their application in nutrition *Biomarkers in nutrition* (pp. 833-856).
- Carvalho, L. C., Vidigal, P., & Amâncio, S. (2015). Oxidative stress homeostasis in grapevine (vitis vinifera L.). *Frontiers in Environmental Science*, 3. doi: 10.3389/fenvs.2015.00020
- Cerovic, Z. G., Moise, N., Agati, G., Latouche, G., Ben Ghazlen, N., & Meyer, S. (2008). New portable optical sensors for the assessment of winegrape phenolic maturity based on berry fluorescence. *Journal of Food Composition and Analysis*, 21(8), 650-654. doi: 10.1016/j.jfca.2008.03.012

- Cerovic, Z. G., Ghazlen, N. B., Milhade, C., Obert, M., Debuissou, S., & Moigne, M. L. (2015). Nondestructive diagnostic test for nitrogen nutrition of grapevine (vitis vinifera.) based on dual leaf-clip measurements in the field. *Journal of Agricultural and Food Chemistry*, 63(14), 3669-3680. doi: 10.1021/acs.jafc.5b00304
- Chandra, H. M., & Ramalingam, S. (2011). Antioxidant potentials of skin, pulp, and seed fractions of commercially important tomato cultivars. *Food Science and Biotechnology*, 20(1), 15-21. doi: 10.1007/s10068-011-0003-z
- Chandra, H. M., Shanmugaraj, B. M., Srinivasan, B., & Ramalingam, S. (2012). Influence of genotypic variations on antioxidant properties in different fractions of tomato. *Journal of Food Science*, 77(11), C1174-C1178. doi: 10.1111/j.1750-3841.2012.02962.x
- Chaves, M. M., Zarrouk, O., Francisco, R., Costa, J. M., Santos, T., Regalado, A. P., . . . Lopes, C. M. (2010). Grapevine under deficit irrigation: Hints from physiological and molecular data. *Annals of Botany*, 105(5), 661-676. doi: 10.1093/aob/mcq030
- Christ, B., & Hörtensteiner, S. (2013). Mechanism and significance of chlorophyll breakdown. *Journal of Plant Growth Regulation*, 33(1), 4-20. doi: 10.1007/s00344-013-9392-y
- Ciaccheri, L., Tuccio, L., Mencaglia, A. A., Mignani, A. G., Hallmann, E., Sikorska-Zimny, K., . . . Agati, G. (2018). Directional versus total reflectance spectroscopy for the in situ determination of lycopene in tomato fruits. *Journal of Food Composition and Analysis*, 71, 65-71. doi: 10.1016/j.jfca.2018.01.023
- Clément, A., Dorais, M., & Vernon, M. (2008). Nondestructive measurement of fresh tomato lycopene content and other physicochemical characteristics using visible-nir spectroscopy. *Journal of Agricultural and Food Chemistry*, 56(21), 9813-9818. doi: 10.1021/jf801299r
- Clément, A., Bacon, R., Sirois, S., & Dorais, M. (2015). Mature-ripe tomato spectral classification according to lycopene content and fruit type by visible, nir reflectance and intrinsic fluorescence. *Quality Assurance and Safety of Crops & Foods*, 7(5), 747-756. doi: 10.3920/qas2014.0521
- Cogato, A., Pagay, V., Marinello, F., Meggio, F., Grace, P., & De Antoni Migliorati, M. (2019). Assessing the feasibility of using sentinel-2 imagery to quantify the impact of heatwaves on irrigated vineyards. *Remote Sensing*, 11(23). doi: 10.3390/rs11232869
- Cook, R. D., & Forzani, L. (2021). Pls regression algorithms in the presence of nonlinearity. *Chemometrics and Intelligent Laboratory Systems*, 213. doi: 10.1016/j.chemolab.2021.104307
- Costa-Santos, M., Mariz-Ponte, N., Dias, M. C., Moura, L., Marques, G., & Santos, C. (2021). Effect of bacillus spp. And brevibacillus sp. On the photosynthesis and redox status of solanum lycopersicum. *Horticulturae*, 7(2), 24.
- Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F. N., & Leip, A. (2021). Food systems are responsible for a third of global anthropogenic ghg emissions. *Nat Food*, 2(3), 198-209. doi: 10.1038/s43016-021-00225-9
- Cunha, M., Ribeiro, H., & Abreu, I. (2016). Pollen-based predictive modelling of wine production: Application to an arid region. *European Journal of Agronomy*, 73, 42-54. doi: 10.1016/j.eja.2015.10.008
- Cunha, M., & Richter, C. (2016). The impact of climate change on the winegrape vineyards of the portuguese douro region. *Climatic Change*, 138(1), 239-251. doi: 10.1007/s10584-016-1719-9
- Cunha, M., Martins, R., & Neves dos Santos, F. (2022). Sustainable agriculture in the era of field-omics. Let's improve agronomics. *INESC TEC Science&Society*, 1(4).

- D'Agostino, N., & Tripodi, P. (2017). Ngs-based genotyping, high-throughput phenotyping and genome-wide association studies laid the foundations for next-generation breeding in horticultural crops. *Diversity*, 9(3). doi: 10.3390/d9030038
- Da Costa Martins, R. M., Baptista Neves Dos Santos, F., De Miranda Furtado Campos Cunha, M. M., Marques Monteiro Da Silva, A. F., Tosin, R., Costa Magalhães, S. A., & Reis Pereira, M. (2023). Method and device for non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues. WO/2023/126532.
- Dahlstrand, U., Sheikh, R., Dybelius Ansson, C., Memarzadeh, K., Reistad, N., & Malmso, M. (2019). Extended-wavelength diffuse reflectance spectroscopy with a machine-learning method for in vivo tissue classification. *PLoS One*, 14(10), e0223682. doi: 10.1371/journal.pone.0223682
- Dai, Z. W., Leon, C., Feil, R., Lunn, J. E., Delrot, S., & Gomes, E. (2013). Metabolic profiling reveals coordinated switches in primary carbohydrate metabolism in grape berry (*Vitis vinifera* L.), a non-climacteric fleshy fruit. *J Exp Bot*, 64(5), 1345-1355. doi: 10.1093/jxb/ers396
- David, E., Tosin, R., Gonçalves, I., Rodrigues, L., Barbosa, C., Santos, F., . . . Cunha, M. (2023). *In-field hyperspectral proximal sensing for estimating grapevine water status to support smart precision viticulture*. Paper presented at the The 3rd International Electronic Conference on Agronomy session Improving nutrient- and water-use efficiencies.
- De Bei, R., Cozzolino, D., Sullivan, W., Cynkar, W., Fuentes, S., Damberg, R., . . . Tyerman, S. (2011). Non-destructive measurement of grapevine water potential using near infrared spectroscopy. *Australian Journal of Grape and Wine Research*, 17(1), 62-71. doi: 10.1111/j.1755-0238.2010.00117.x
- de Noord, O. E., & Theobald, E. H. (2006). Multilevel component analysis and multilevel pls of chemical process data. *Journal of Chemometrics*, 19(5-7), 301-307. doi: 10.1002/cem.933
- de Oliveira Dal'Molin, C. G., Quek, L. E., Palfreyman, R. W., Brumley, S. M., & Nielsen, L. K. (2010). Aragem, a genome-scale reconstruction of the primary metabolic network in arabidopsis. *Plant Physiol*, 152(2), 579-589. doi: 10.1104/pp.109.148817
- Degu, A., Hochberg, U., Sikron, N., Venturini, L., Buson, G., Ghan, R., . . . Fait, A. (2014). Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between cabernet sauvignon and shiraz cultivars at branching points in the polyphenol pathway. *BMC Plant Biology*, 14(1), 188. doi: 10.1186/s12870-014-0188-4
- Deluc, L. G., Grimplet, J., Wheatley, M. D., Tillett, R. L., Quilici, D. R., Osborne, C., . . . Cramer, G. R. (2007). Transcriptomic and metabolite analyses of cabernet sauvignon grape berry development. *BMC Genomics*, 8, 429. doi: 10.1186/1471-2164-8-429
- Dennis, P. P., Ehrenberg, M., & Bremer, H. (2004). Control of rna synthesis in escherichia coli: A systems biology approach. *Microbiol Mol Biol Rev*, 68(4), 639-668. doi: 10.1128/MMBR.68.4.639-668.2004
- Dersch, L. M., Beckers, V., & Wittmann, C. (2016). Green pathways: Metabolic network analysis of plant systems. *Metab Eng*, 34, 1-24. doi: 10.1016/j.ymben.2015.12.001
- Dhaliwal, S. S., Sharma, V., & Shukla, A. K. (2022). Chapter one - impact of micronutrients in mitigation of abiotic stresses in soils and plants—a progressive step toward crop security and nutritional quality. In D. L. Sparks (Ed.), *Advances in agronomy* (Vol. 173, pp. 1-78): Academic Press.
- Diago, M. P., Fernandez-Novales, J., Gutierrez, S., Maranon, M., & Tardaguila, J. (2018). Development and validation of a new methodology to assess the vineyard water

- status by on-the-go near infrared spectroscopy. *Frontiers in Plant Science*, 9, 59. doi: 10.3389/fpls.2018.00059
- Ding, Q., Rehman Sheikh, A., Pan, W., Gu, X., Sun, N., Su, X., . . . Zhang, T. (2021). In situ monitoring of grape seed protein hydrolysis by raman spectroscopy. *Journal of Food Biochemistry*, 45(4), e13646. doi: 10.1111/jfbc.13646
- Ding, X., Guo, Y., Ni, Y., & Kokot, S. (2016). A novel nir spectroscopic method for rapid analyses of lycopene, total acid, sugar, phenols and antioxidant activity in dehydrated tomato samples. *Vibrational Spectroscopy*, 82, 1-9. doi: 10.1016/j.vibspec.2015.10.004
- Donis-González, I. R., Guyer, D. E., Pease, A., & Barthel, F. (2014). Internal characterisation of fresh agricultural products using traditional and ultrafast electron beam x-ray computed tomography imaging. *Biosystems Engineering*, 117, 104-113. doi: 10.1016/j.biosystemseng.2013.07.002
- Dora, M., Kumar, A., Mangla, S. K., Pant, A., & Kamal, M. M. (2021). Critical success factors influencing artificial intelligence adoption in food supply chains. *International Journal of Production Research*, 60(14), 4621-4640. doi: 10.1080/00207543.2021.1959665
- dos Santos Costa, D., Oliveros Mesa, N. F., Santos Freire, M., Pereira Ramos, R., & Teruel Mederos, B. J. (2019). Development of predictive models for quality and maturation stage attributes of wine grapes using vis-nir reflectance spectroscopy. *Postharvest Biology and Technology*, 150, 166-178. doi: 10.1016/j.postharvbio.2018.12.010
- Doumouya, S., Lahaye, M., Maury, C., & Siret, R. (2014). Physical and physiological heterogeneity within the grape bunch: Impact on mechanical properties during maturation. *American Journal of Enology and Viticulture*, 65(2), 170-178. doi: 10.5344/ajev.2014.13062
- du Plessis, K., Young, P. R., Eyeghe-Bickong, H. A., & Vivier, M. A. (2017). The transcriptional responses and metabolic consequences of acclimation to elevated light exposure in grapevine berries. *Front Plant Sci*, 8, 1261. doi: 10.3389/fpls.2017.01261
- Duan, S., Wu, Y., Fu, R., Wang, L., Chen, Y., Xu, W., . . . Wang, S. (2019). Comparative metabolic profiling of grape skin tissue along grapevine berry developmental stages reveals systematic influences of root restriction on skin metabolome. *Int J Mol Sci*, 20(3). doi: 10.3390/ijms20030534
- Duarte, L., Pacheco, C., & Soveral-Dias, J. (1998). Métodos de análise de material vegetal. *Laboratório Químico Agrícola Rebelo da Silva, INIA, Lisboa*.
- Duncan, E., Glaros, A., Ross, D. Z., & Nost, E. (2021). New but for whom? Discourses of innovation in precision agriculture. *Agric Human Values*, 38(4), 1181-1199. doi: 10.1007/s10460-021-10244-8
- Ecarnot, M., Baczyk, P., Tessarotto, L., & Chervin, C. (2013). Rapid phenotyping of the tomato fruit model, micro-tom, with a portable vis-nir spectrometer. *Plant Physiol Biochem*, 70, 159-163. doi: 10.1016/j.plaphy.2013.05.019
- Ergon, R. (2009). Re-interpretation of nipals results solves plsr inconsistency problem. *Journal of Chemometrics*, 23(2), 72-75. doi: 10.1002/cem.1180
- European Commission. (2020). A farm to fork strategy for a fair, healthy and environmentally-friendly food system Retrieved 19 January 2024, from https://eur-lex.europa.eu/resource.html?uri=cellar:ea0f9f73-9ab2-11ea-9d2d-01aa75ed71a1.0001.02/DOC_1&format=PDF
- European Commission. (2021). European green deal - delivering on our targets Retrieved 19 January 2024, from https://ec.europa.eu/commission/presscorner/api/files/attachment/869807/EGD_brochure_EN.pdf.pdf

- Fachada, N., Figueiredo, M. A. T., Lopes, V. V., Martins, R. C., & Rosa, A. C. (2014). Spectrometric differentiation of yeast strains using minimum volume increase and minimum direction change clustering criteria. *Pattern Recognition Letters*, 45, 55-61. doi: 10.1016/j.patrec.2014.03.008
- Falcioni, R., Goncalves, J. V. F., de Oliveira, K. M., de Oliveira, C. A., Reis, A. S., Crusiol, L. G. T., . . . Nanni, M. R. (2023). Chemometric analysis for the prediction of biochemical compounds in leaves using uv-vis-nir-swir hyperspectroscopy. *Plants (Basel)*, 12(19). doi: 10.3390/plants12193424
- Farhadi, K., Esmailzadeh, F., Hatami, M., Forough, M., & Molaie, R. (2016). Determination of phenolic compounds content and antioxidant activity in skin, pulp, seed, cane and leaf of five native grape cultivars in west azerbaijan province, iran. *Food Chemistry*, 199, 847-855. doi: 10.1016/j.foodchem.2015.12.083
- Féret, J. B., Gitelson, A. A., Noble, S. D., & Jacquemoud, S. (2017). Prospect-d: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment*, 193, 204-215. doi: 10.1016/j.rse.2017.03.004
- Fernandes, A. M., Franco, C., Mendes-Ferreira, A., Mendes-Faia, A., Costa, P. L. d., & Melo-Pinto, P. (2015). Brix, ph and anthocyanin content determination in whole port wine grape berries by hyperspectral imaging and neural networks. *Computers and Electronics in Agriculture*, 115, 88-96. doi: 10.1016/j.compag.2015.05.013
- Fernandez-Navales, J., Garde-Cerdan, T., Tardaguila, J., Gutierrez-Gamboa, G., Perez-Alvarez, E. P., & Diago, M. P. (2019a). Assessment of amino acids and total soluble solids in intact grape berries using contactless vis and nir spectroscopy during ripening. *Talanta*, 199, 244-253. doi: 10.1016/j.talanta.2019.02.037
- Fernandez-Navales, J., Tardaguila, J., Gutierrez, S., & Paz Diago, M. (2019b). On-the-go vis + sw - nir spectroscopy as a reliable monitoring tool for grape composition within the vineyard. *Molecules*, 24(15). doi: 10.3390/molecules24152795
- Fernández-Navales, J., Tardaguila, J., Gutiérrez, S., Marañón, M., & Diago, M. P. (2018). In field quantification and discrimination of different vineyard water regimes by on-the-go nir spectroscopy. *Biosystems Engineering*, 165, 47-58. doi: 10.1016/j.biosystemseng.2017.08.018
- Fernández-Navales, J., Saiz-Rubio, V., Barrio, I., Rovira-Más, F., Cuenca-Cuenca, A., Santos Alves, F., . . . Diago, M. P. (2021). Monitoring and mapping vineyard water status using non-invasive technologies by a ground robot. *Remote Sensing*, 13(14). doi: 10.3390/rs13142830
- Ferreira, M. F. S., Capela, D., Silva, N. A., Gonçalves, F., Lima, A., Guimarães, D., & Jorge, P. A. S. (2022). Comprehensive comparison of linear and non-linear methodologies for lithium quantification in geological samples using libs. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 195. doi: 10.1016/j.sab.2022.106504
- Fritsch, S., Guenther, F., & Wright, M. N. (2019). Neuralnet: Training of neural networks. Retrieved from <https://CRAN.R-project.org/package=neuralnet>
- Fuentes, S., Tongson, E., & Gonzalez Viejo, C. (2023). New developments and opportunities for ai in viticulture, pomology, and soft-fruit research: A mini-review and invitation to contribute articles. *Frontiers in Horticulture*, 2. doi: 10.3389/fhort.2023.1282615
- Gago, J., Fernie, A. R., Nikoloski, Z., Tohge, T., Martorell, S., Escalona, J. M., . . . Medrano, H. (2017). Integrative field scale phenotyping for investigating metabolic components of water stress within a vineyard. *Plant Methods*, 13, 90. doi: 10.1186/s13007-017-0241-z

- Gajewska, E., & Skłodowska, M. (2007). Effect of nickel on ros content and antioxidative enzyme activities in wheat leaves. *BioMetals*, 20(1), 27-36. doi: 10.1007/s10534-006-9011-5
- Gallagher, N. B., Blake, T. A., & Gassman, P. L. (2005). Application of extended inverse scatter correction to mid-infrared reflectance spectra of soil. *Journal of Chemometrics*, 19(5-7), 271-281. doi: 10.1002/cem.929
- Galvez-Sola, L., Garcia-Sanchez, F., Perez-Perez, J. G., Gimeno, V., Navarro, J. M., Moral, R., . . . Nieves, M. (2015). Rapid estimation of nutritional elements on citrus leaves by near infrared reflectance spectroscopy. *Front Plant Sci*, 6, 571. doi: 10.3389/fpls.2015.00571
- Garcia, E., & Barrett, D. M. (2006). Evaluation of processing tomatoes from two consecutive growing seasons: Quality attributes, peelability and yield. *Journal of Food Processing and Preservation*, 30(1), 20-36. doi: 10.1111/j.1745-4549.2005.00044.x
- Garrido, A., Conde, A., Serodio, J., De Vos, R. C. H., & Cunha, A. (2023). Fruit photosynthesis: More to know about where, how and why. *Plants (Basel)*, 12(13). doi: 10.3390/plants12132393
- Gautron, R., Maillard, O.-A., Preux, P., Corbeels, M., & Sabbadin, R. (2022). Reinforcement learning for crop management support: Review, prospects and challenges. *Computers and Electronics in Agriculture*, 200. doi: 10.1016/j.compag.2022.107182
- Gawad Saad, A., Pék, Z. n., Szuvandzsiev, P. t., Hussein Gehad, D., & Helyes, L. (2017). Determination of carotenoids in tomato products using vis/nir spectroscopy. *Journal of microbiology, biotechnology and food sciences*, 7(1), 27-31. doi: 10.15414/jmbfs.2017.7.1.27-31
- Ge, S., Zhu, Z., & Jiang, Y. (2018). Long-term impact of fertilization on soil ph and fertility in an apple production system. *Journal of soil science and plant nutrition*, 18, 282-293.
- Geigenberger, P., & Fernie, A. R. (2014). Metabolic control of redox and redox control of metabolism in plants. *Antioxidants & Redox Signaling*, 21(9), 1389-1421. doi: 10.1089/ars.2014.6018
- Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185, 1-17. doi: [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
- Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1), 1-58. doi: 10.1162/neco.1992.4.1.1
- Gerlin, L., Frainay, C., Jourdan, F., Baroukh, C., & Prigent, S. (2021). Plant genome-scale metabolic networks *Plant metabolomics in full swing* (pp. 237-270).
- Gerlin, L., Cottret, L., Escourrou, A., Genin, S., & Baroukh, C. (2022). A multi-organ metabolic model of tomato predicts plant responses to nutritional and genetic perturbations. *Plant Physiol*, 188(3), 1709-1723. doi: 10.1093/plphys/kiab548
- Gibeaut, D. M., Pauly, M., Bacic, A., & Fincher, G. B. (2005). Changes in cell wall polysaccharides in developing barley (*hordeum vulgare*) coleoptiles. *Planta*, 221(5), 729-738. doi: 10.1007/s00425-005-1481-0
- Giovenzana, V., Civelli, R., Beghi, R., Oberti, R., & Guidetti, R. (2015). Testing of a simplified led based vis/nir system for rapid ripeness evaluation of white grape (*vitis vinifera* l.) for franciacorta wine. *Talanta*, 144, 584-591. doi: 10.1016/j.talanta.2015.06.055
- Giovenzana, V., Beghi, R., Parisi, S., Brancadoro, L., & Guidetti, R. (2018). Potential effectiveness of visible and near infrared spectroscopy coupled with wavelength selection for real time grapevine leaf water status measurement. *Journal of the Science of Food and Agriculture*, 98(5), 1935-1943. doi: 10.1002/jsfa.8676

- Goclawski, J., Sekulska-Nalewajko, J., Korzeniewska, E., & Piekarska, A. (2017). The use of optical coherence tomography for the evaluation of textural changes of grapes exposed to pulsed electric field. *Computers and Electronics in Agriculture*, 142, 29-40. doi: 10.1016/j.compag.2017.08.008
- Gold, K. M., Gevens, A. J., & Townsend, P. A. (2019). System for detection of disease in plants. U.S. Patent No. 11,054,368.
- Gómez, R., Costa, J., Amo, M., Alvarruiz, A., Picazo, M., & Pardo, J. E. (2001). Physicochemical and colorimetric evaluation of local varieties of tomato grown in se Spain. *Journal of the Science of Food and Agriculture*, 81(11), 1101-1105. doi: 10.1002/jsfa.915
- González-Fernández, A. B., Rodríguez-Pérez, J. R., Marabel, M., & Álvarez-Taboada, F. (2015). Spectroscopic estimation of leaf water content in commercial vineyards using continuum removal and partial least squares regression. *Scientia Horticulturae*, 188, 15-22. doi: 10.1016/j.scienta.2015.03.012
- González-Fernández, A. B., Sanz-Ablanedo, E., Gabella, V. M., García-Fernández, M., & Rodríguez-Pérez, J. R. (2019). Field spectroscopy: A non-destructive technique for estimating water status in vineyards. *Agronomy*, 9(8). doi: 10.3390/agronomy9080427
- Gonzalez-Sanchez, A., Frausto-Solis, J., & Ojeda-Bustamante, W. (2014). Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, 12(2). doi: 10.5424/sjar/2014122-4439
- Gouot, J. C., Smith, J. P., Holzapfel, B. P., & Barril, C. (2019). Impact of short temperature exposure of vitis vinifera l. Cv. Shiraz grapevine bunches on berry development, primary metabolism and tannin accumulation. *Environmental and Experimental Botany*, 168. doi: 10.1016/j.envexpbot.2019.103866
- Goussev, V., Zhao, H., Gui, L., & Yin, J. (2019). Nonparametric and parametric methods of spectral analysis. *MATEC Web of Conferences*, 283. doi: 10.1051/mateconf/201928307002
- Grosskinsky, D. K., Syaifullah, S. J., & Roitsch, T. (2018). Integration of multi-omics techniques and physiological phenotyping within a holistic phenomics approach to study senescence in model and crop plants. *J Exp Bot*, 69(4), 825-844. doi: 10.1093/jxb/erx333
- Guo, W., Nazim, H., Liang, Z., & Yang, D. (2016). Magnesium deficiency in plants: An urgent problem. *The Crop Journal*, 4(2), 83-91. doi: 10.1016/j.cj.2015.11.003
- Guo, W., Xu, Y., & Feng, X. (2017). Deepmetabolism: A deep learning system to predict phenotype from genome sequencing. *arXiv preprint arXiv:1705.03094*.
- Gutierrez, S., Tardaguila, J., Fernandez-Navales, J., & Diago, M. P. (2016). Data mining and nir spectroscopy in viticulture: Applications for plant phenotyping under field conditions. *Sensors (Basel)*, 16(2), 236. doi: 10.3390/s16020236
- Hahn, D. W., & Omenetto, N. (2012). Laser-induced breakdown spectroscopy (libs), part ii: Review of instrumental and methodological approaches to material analysis and applications to different fields. *Appl Spectrosc*, 66(4), 347-419. doi: 10.1366/11-06574
- Hazarika, S., Choudhury, R., Montazer, B., Medhi, S., Goswami, M. P., & Sarma, U. (2020). Detection of citrus tristeza virus in mandarin orange using a custom-developed electronic nose system. *IEEE Transactions on Instrumentation and Measurement*, 69(11), 9010-9018. doi: 10.1109/tim.2020.2997064
- Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., . . . Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the cobra toolbox v.3.0. *Nature Protocols*, 14(3), 639-702. doi: 10.1038/s41596-018-0098-2
- Helman, D., Bahat, I., Netzer, Y., Ben-Gal, A., Alchanatis, V., Peeters, A., & Cohen, Y. (2018). Using time series of high-resolution planet satellite images to monitor

- grapevine stem water potential in commercial vineyards. *Remote Sensing*, 10(10). doi: 10.3390/rs10101615
- Herman, J., Zhang, Y., Castranova, V., & Neal, S. L. (2018). Emerging technologies for optical spectral detection of reactive oxygen species. *Anal Bioanal Chem*, 410(24), 6079-6095. doi: 10.1007/s00216-018-1233-1
- Hochberg, U., Degu, A., Toubiana, D., Gendler, T., Nikoloski, Z., Rachmilevitch, S., & Fait, A. (2013). Metabolite profiling and network analysis reveal coordinated changes in grapevine water stress response. *BMC Plant Biology*, 13(1), 184. doi: 10.1186/1471-2229-13-184
- Hocking, B. J. (2015). *The role of calcium in the cell wall of grape berries*. PhD, The University of Adelaide, Adelaide.
- 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. doi: 10.1016/j.envsoft.2014.07.009
- Huang, Y., Lu, R., & Chen, K. (2018a). Assessment of tomato soluble solids content and ph by spatially-resolved and conventional vis/nir spectroscopy. *Journal of Food Engineering*, 236, 19-28. doi: 10.1016/j.jfoodeng.2018.05.008
- Huang, Y., Lu, R., Hu, D., & Chen, K. (2018b). Quality assessment of tomato fruit by optical absorption and scattering properties. *Postharvest Biology and Technology*, 143, 78-85. doi: 10.1016/j.postharvbio.2018.04.016
- Huma, B., Kundu, S., Poolman, M. G., Kruger, N. J., & Fell, D. A. (2018). Stoichiometric analysis of the energetics and metabolic impact of photorespiration in c3 plants. *The Plant Journal*, 96(6), 1228-1241. doi: <https://doi.org/10.1111/tpj.14105>
- Iqbal, Z., Iqbal, M. S., Hashem, A., Abd Allah, E. F., & Ansari, M. I. (2021). Plant defense responses to biotic stress and its interplay with fluctuating dark/light conditions. *Front Plant Sci*, 12, 631810. doi: 10.3389/fpls.2021.631810
- Janni, M., Coppede, N., Bettelli, M., Briglia, N., Petrozza, A., Summerer, S., . . . Zappettini, A. (2019). In vivo phenotyping for the early detection of drought stress in tomato. *Plant Phenomics*, 2019, 6168209. doi: 10.34133/2019/6168209
- Jiang, K., Wu, J., Qiu, Y., Go, Y. Y., Ban, K., Park, H. J., & Lee, J.-H. (2021). Plasmonic colorimetric pcr for rapid molecular diagnostic assays. *Sensors and Actuators B: Chemical*, 337. doi: 10.1016/j.snb.2021.129762
- Jie, W., Hua-bing, Z., Chang-wei, S., qiao-wei, C., Cai-xia, D., & Yang-chun, X. (2014). Determination of nitrogen concentration in fresh pear leaves by visible/near-infrared reflectance spectroscopy. *Agronomy Journal*, 106(5), 1867-1872. doi: <https://doi.org/10.2134/agronj13.0303>
- Jordao, A. M., Correia, A., & Gonçalves, F. (2012). Evolution of antioxidant capacity in seeds and skins during grape maturation and their association with proanthocyanidin and anthocyanin content. *Vitis*, 51(3), 137-139.
- Joshi, A., Rienks, M., Theofilatos, K., & Mayr, M. (2021). Systems biology in cardiovascular disease: A multiomics approach. *Nature Reviews Cardiology*, 18(5), 313-330. doi: 10.1038/s41569-020-00477-1
- Ju, Y. L., Yue, X. F., Zhao, X. F., Zhao, H., & Fang, Y. L. (2018). Physiological, micro-morphological and metabolomic analysis of grapevine (vitis vinifera l.) leaf of plants under water stress. *Plant Physiol Biochem*, 130, 501-510. doi: 10.1016/j.plaphy.2018.07.036
- Jull, H., Künnemeyer, R., & Schaare, P. (2018). Nutrient quantification in fresh and dried mixtures of ryegrass and clover leaves using laser-induced breakdown spectroscopy. *Precision Agriculture*, 19(5), 823-839. doi: 10.1007/s11119-018-9559-4
- Kalopesa, E., Karyotis, K., Tziolas, N., Tsakiridis, N., Samarinas, N., & Zalidis, G. (2023). Estimation of sugar content in wine grapes via in situ vnir-swir point spectroscopy

- using explainable artificial intelligence techniques. *Sensors (Basel)*, 23(3). doi: 10.3390/s23031065
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70-90. doi: 10.1016/j.compag.2018.02.016
- Kandylakis, Z., Falagas, A., Karakizi, C., & Karantzas, K. (2020). Water stress estimation in vineyards from aerial swir and multispectral uav data. *Remote Sensing*, 12(15). doi: 10.3390/rs12152499
- Kanno, I., & Kuroyama, T. (2020). Estimation of the sugar content of fruit by energy-resolved computed tomography using a material decomposition method. *Journal of Nuclear Science and Technology*, 58(5), 533-541. doi: 10.1080/00223131.2020.1845836
- Kassambara, A. (2020). Ggpubr: 'Ggplot2' based publication ready plots. Retrieved from <https://CRAN.R-project.org/package=ggpubr>
- Kassambara, A., & Mundt, F. (2020). Factoextra: Extract and visualize the results of multivariate data analyses. Retrieved from <https://CRAN.R-project.org/package=factoextra>
- Kennedy, J. A., Matthews, M. A., & Waterhouse, A. L. (2000). Changes in grape seed polyphenols during fruit ripening. *Phytochemistry*, 55(1), 77-85. doi: [https://doi.org/10.1016/S0031-9422\(00\)00196-5](https://doi.org/10.1016/S0031-9422(00)00196-5)
- Kok, Z. H., Mohamed Shariff, A. R., Alfatni, M. S. M., & Khairunniza-Bejo, S. (2021). Support vector machine in precision agriculture: A review. *Computers and Electronics in Agriculture*, 191. doi: 10.1016/j.compag.2021.106546
- Konagaya, K., Al Riza, D. F., Nie, S., Yoneda, M., Hirata, T., Takahashi, N., . . . Kondo, N. (2020). Monitoring mature tomato (red stage) quality during storage using ultraviolet-induced visible fluorescence image. *Postharvest Biology and Technology*, 160. doi: 10.1016/j.postharvbio.2019.111031
- Kristian Hovde Liland, Bjørn-Helge Mevik, & Wehrens, R. (2021). Pls: Partial least squares and principal component regression. Retrieved from <https://CRAN.R-project.org/package=pls>
- Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, 6(1), 10. doi: 10.1186/1758-2946-6-10
- Kuhn, M. (2020). Caret: Classification and regression training. Retrieved from <https://CRAN.R-project.org/package=caret>
- Kumar, A., Pathak, R. K., Gupta, S. M., Gaur, V. S., & Pandey, D. (2015). Systems biology for smart crops and agricultural innovation: Filling the gaps between genotype and phenotype for complex traits linked with robust agricultural productivity and sustainability. *OMICS*, 19(10), 581-601. doi: 10.1089/omi.2015.0106
- Kumar, M., Tomar, M., Bhuyan, D. J., Punia, S., Grasso, S., Sa, A. G. A., . . . Mekhemar, M. (2021a). Tomato (solanum lycopersicum l.) seed: A review on bioactives and biomedical activities. *Biomed Pharmacother*, 142, 112018. doi: 10.1016/j.biopha.2021.112018
- Kumar, S., Kumar, S., & Mohapatra, T. (2021b). Interaction between macro- and micro-nutrients in plants. *Front Plant Sci*, 12, 665583. doi: 10.3389/fpls.2021.665583
- Kurt, A., Torun, H., Colak, N., Seiler, G., Hayirlioglu-Ayaz, S., & Ayaz, F. A. (2017). Nutrient profiles of the hybrid grape cultivar 'isabel' during berry maturation and ripening. *J Sci Food Agric*, 97(8), 2468-2479. doi: 10.1002/jsfa.8061
- Lakshmanan, M., Lim, S. H., Mohanty, B., Kim, J. K., Ha, S. H., & Lee, D. Y. (2015). Unraveling the light-specific metabolic and regulatory signatures of rice through combined in silico modeling and multiomics analysis. *Plant Physiol*, 169(4), 3002-3020. doi: 10.1104/pp.15.01379

- Laroche-Pinel, E., Albughdadi, M., Duthoit, S., Chéret, V., Rousseau, J., & Clenet, H. (2021). Understanding vine hyperspectral signature through different irrigation plans: A first step to monitor vineyard water status. *Remote Sensing*, 13(3). doi: 10.3390/rs13030536
- Le, S., Josse, J., & Husson, F. (2008). Factominer: A package for multivariate analysis. *Journal of Statistical Software*, 25(1), 1-18. doi: 10.18637/jss.v025.i01
- Lecourieux, D., Kappel, C., Claverol, S., Pieri, P., Feil, R., Lunn, J. E., . . . Lecourieux, F. (2020). Proteomic and metabolomic profiling underlines the stage- and time-dependent effects of high temperature on grape berry metabolism. *J Integr Plant Biol*, 62(8), 1132-1158. doi: 10.1111/jipb.12894
- Leiva-Valenzuela, G. A., Lu, R., & Aguilera, J. M. (2013). Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. *Journal of Food Engineering*, 115(1), 91-98. doi: 10.1016/j.jfoodeng.2012.10.001
- Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., . . . Laibe, C. (2010). Biomodels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4(1), 92. doi: 10.1186/1752-0509-4-92
- Li, M., Landahl, S., East, A. R., Verboven, P., & Terry, L. A. (2019). Optical coherence tomography—a review of the opportunities and challenges for postharvest quality evaluation. *Postharvest Biology and Technology*, 150, 9-18. doi: 10.1016/j.postharvbio.2018.12.005
- Li, X., Zhang, W., Niu, D., & Liu, X. (2024). Effects of abiotic stress on chlorophyll metabolism. *Plant Sci*, 342, 112030. doi: 10.1016/j.plantsci.2024.112030
- Li, Y. M., You, J. L., Nie, W. F., Sun, M. H., & Xie, Z. S. (2022). Transcription profiles reveal age-dependent variations of photosynthetic properties and sugar metabolism in grape leaves (*vitis vinifera* L.). *Int J Mol Sci*, 23(4). doi: 10.3390/ijms23042243
- Li, Z., Qiu, Q., Chen, Y., Lin, D., Huang, J., & Huang, T. (2021). Metabolite alteration in response to low phosphorus stress in developing tomato fruits. *Plant Physiol Biochem*, 159, 234-243. doi: 10.1016/j.plaphy.2020.12.023
- Lichtenthaler, H. K., & Babani, F. (2004). Light adaptation and senescence of the photosynthetic apparatus. Changes in pigment composition, chlorophyll fluorescence parameters and photosynthetic activity. In G. C. Papageorgiou & Govindjee (Eds.), *Chlorophyll a fluorescence: A signature of photosynthesis* (pp. 713-736). Dordrecht: Springer Netherlands.
- Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Yunsheng, M., . . . Hou, P. (2018). A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. *IEEE Transactions on Services Computing*, 11(2), 249-261. doi: 10.1109/tsc.2017.2662008
- Liu, J., Chan, S. H. J., Chen, J., Solem, C., & Jensen, P. R. (2019a). Systems biology - a guide for understanding and developing improved strains of lactic acid bacteria. *Front Microbiol*, 10, 876. doi: 10.3389/fmicb.2019.00876
- Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: A review. *Plant Methods*, 17(1), 22. doi: 10.1186/s13007-021-00722-9
- Liu, W., Liu, C., Jin, J., Li, D., Fu, Y., & Yuan, X. (2020a). High-throughput phenotyping of morphological seed and fruit characteristics using x-ray computed tomography. *Frontiers in Plant Science*, 11, 601475. doi: 10.3389/fpls.2020.601475
- Liu, W., Liu, C., Jin, J., Li, D., Fu, Y., & Yuan, X. (2020b). High-throughput phenotyping of morphological seed and fruit characteristics using x-ray computed tomography. *Front Plant Sci*, 11, 601475. doi: 10.3389/fpls.2020.601475

- Liu, X., Liu, C., Shi, Z., & Chang, Q. (2019b). Comparison of prediction power of three multivariate calibrations for estimation of leaf anthocyanin content with visible spectroscopy in prunus cerasifera. *PeerJ*, 7, e7997. doi: 10.7717/peerj.7997
- Lo Presti, D., Di Tocco, J., Massaroni, C., Cimini, S., De Gara, L., Singh, S., . . . Cinti, S. (2023). Current understanding, challenges and perspective on portable systems applied to plant monitoring and precision agriculture. *Biosens Bioelectron*, 222, 115005. doi: 10.1016/j.bios.2022.115005
- Lofstedt, T., Hoffman, D., & Trygg, J. (2013a). Global, local and unique decompositions in onpls for multiblock data analysis. *Anal Chim Acta*, 791, 13-24. doi: 10.1016/j.aca.2013.06.026
- Lofstedt, T., Hoffman, D., & Trygg, J. (2013b). Global, local and unique decompositions in onpls for multiblock data analysis. *Analytica Chimica Acta*, 791, 13-24. doi: 10.1016/j.aca.2013.06.026
- Loggenberg, K., Strever, A., Greyling, B., & Poona, N. (2018). Modelling water stress in a shiraz vineyard using hyperspectral imaging and machine learning. *Remote Sensing*, 10(2). doi: 10.3390/rs10020202
- Lozano-Elena, F., Fabregas, N., Coletto-Alcudia, V., & Cano-Delgado, A. I. (2022). Analysis of metabolic dynamics during drought stress in arabidopsis plants. *Sci Data*, 9(1), 90. doi: 10.1038/s41597-022-01161-4
- Lu, Y., Saeys, W., Kim, M., Peng, Y., & Lu, R. (2020). Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress. *Postharvest Biology and Technology*, 170. doi: 10.1016/j.postharvbio.2020.111318
- Lu, Y., Chen, D., Olaniyi, E., & Huang, Y. (2022). Generative adversarial networks (gans) for image augmentation in agriculture: A systematic review. *Computers and Electronics in Agriculture*, 200. doi: 10.1016/j.compag.2022.107208
- Luo, Y. Y., Li, R. X., Jiang, Q. S., Bai, R., & Duan, D. (2019). Changes in the chlorophyll content of grape leaves could provide a physiological index for responses and adaptation to uv-c radiation. *Nordic Journal of Botany*, 37(4). doi: 10.1111/njb.02314
- Lynch, J. P. (2019). Root phenotypes for improved nutrient capture: An underexploited opportunity for global agriculture. *New Phytol*, 223(2), 548-564. doi: 10.1111/nph.15738
- Malone, E., Sato Dos Santos, G., Holder, D., & Arridge, S. (2014). Multifrequency electrical impedance tomography using spectral constraints. *IEEE Trans Med Imaging*, 33(2), 340-350. doi: 10.1109/TMI.2013.2284966
- Marin, M., Hobiny, A., & Abbas, I. (2021). Finite element analysis of nonlinear bioheat model in skin tissue due to external thermal sources. *Mathematics*, 9(13). doi: 10.3390/math9131459
- Mariotto, I., Thenkabail, P. S., Huete, A., Slonecker, E. T., & Platonov, A. (2013). Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the hyspirci mission. *Remote Sensing of Environment*, 139, 291-305. doi: 10.1016/j.rse.2013.08.002
- Mariz-Ponte, N., Martins, S., Gonçalves, A., Correia, C. M., Ribeiro, C., Dias, M. C., & Santos, C. (2019). The potential use of the uv-a and uv-b to improve tomato quality and preference for consumers. *Scientia Horticulturae*, 246, 777-784. doi: 10.1016/j.scienta.2018.11.058
- Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry*, 75(3), 394-404.

- Martins, R. (2018). Big data self-learning methodology for the accurate quantification and classification of spectral information under complex variability and multi-scale interference. WO2018060967. PCT WO2018060967.
- Martins, R. C. (2019). Spectrophotometric method and device for predicting a quantification of a constituent from a sample Retrieved 24/03/2024, from <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2019194693& cid=P20-LTIKEU-66113-1>
- Martins, R. C., Magalhães, S., Jorge, P., Barroso, T., & Santos, F. (2019, 2019//). *Metbots: Metabolomics robots for precision viticulture*. Paper presented at the Progress in Artificial Intelligence, Cham.
- Martins, R. C., Barroso, T. G., Jorge, P., Cunha, M., & Santos, F. (2022). Unscrambling spectral interference and matrix effects in vitis vinifera vis-nir spectroscopy: Towards analytical grade ‘in vivo’ sugars and acids quantification. *Computers and Electronics in Agriculture*, 194. doi: 10.1016/j.compag.2022.106710
- Martins, R. C., Santos, F., Cunha, M., Monteiro-Silva, F., Tosin, R., Magalhães, S., & Pereira, M. R. (2023). Method and device for non-invasive tomographic characterisation of a sample comprising a plurality of differentiated tissues Retrieved 29/01/2024, from <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2023126532& cid=P12-LK2EKH-19952-1>
- Martins, R. C., Queirós, C., Silva, F. M., Santos, F., Barroso, T. G., Tosin, R., . . . Silvestre, J. (Under review). Spectral data augmentation for dealing with data scarcity in leaf nutrient uptake quantification. *Biosystems Engineering*.
- Martins, R. M. D. C., Jorge, P. A. D., Silva, E. A. P., De Almeida, J. M. S., & Martins, A. M. D. O. (2018). A method and apparatus for characterisation of constituents in a physical sample from electromagnetic spectral information. United States Patent No. US20210270744A1.
- Martins, V., Unlubayir, M., Teixeira, A., Lanoue, A., & Geros, H. (2021). Exogenous calcium delays grape berry maturation in the white cv. Loureiro while increasing fruit firmness and flavonol content. *Front Plant Sci*, 12, 742887. doi: 10.3389/fpls.2021.742887
- Massonnet, M., Fasoli, M., Tornielli, G. B., Altieri, M., Sandri, M., Zuccolotto, P., . . . Pezzotti, M. (2017). Ripening transcriptomic program in red and white grapevine varieties correlates with berry skin anthocyanin accumulation. *Plant Physiol*, 174(4), 2376-2396. doi: 10.1104/pp.17.00311
- Matese, A., & Di Gennaro, S. (2018). Practical applications of a multisensor uav platform based on multispectral, thermal and rgb high resolution images in precision viticulture. *Agriculture*, 8(7). doi: 10.3390/agriculture8070116
- Matthews, M. A., Thomas, T. R., & Shackel, K. A. (2009). Fruit ripening in vitis vinifera L.: Possible relation of veraison to turgor and berry softening. *Australian Journal of Grape and Wine Research*, 15(3), 278-283. doi: 10.1111/j.1755-0238.2009.00060.x
- Maury, C., Madieta, E., Le Moigne, M., Mehinagic, E., Siret, R., & Jourjon, F. (2009). Development of a mechanical texture test to evaluate the ripening process of cabernet franc grapes. *Journal of Texture Studies*, 40(5), 511-535. doi: 10.1111/j.1745-4603.2009.00195.x
- Mazzetto, F., Calcante, A., Mena, A., & Vercesi, A. (2010). Integration of optical and analogue sensors for monitoring canopy health and vigour in precision viticulture. *Precision Agriculture*, 11(6), 636-649. doi: 10.1007/s11119-010-9186-1
- Medda, S., Fadda, A., & Mulas, M. (2022). Influence of climate change on metabolism and biological characteristics in perennial woody fruit crops in the mediterranean environment. *Horticulturae*, 8(4). doi: 10.3390/horticulturae8040273

- Menesatti, P., Antonucci, F., Pallottino, F., Rocuzzo, G., Allegra, M., Stagno, F., & Intrigliolo, F. (2010). Estimation of plant nutritional status by vis–nir spectrophotometric analysis on orange leaves [citrus sinensis (l) osbeck cv tarocco]. *Biosystems Engineering*, 105(4), 448-454. doi: 10.1016/j.biosystemseng.2010.01.003
- Meyers, J. M., Dokoozlian, N., Ryan, C., Bioni, C., & Vanden Heuvel, J. E. (2020). A new, satellite ndvi-based sampling protocol for grape maturation monitoring. *Remote Sensing*, 12(7). doi: 10.3390/rs12071159
- Middleton, E. M., Huemmrich, K. F., Cheng, Y.-B., & Margolis, H. A. (2012). Spectral bioindicators of photosynthetic efficiency and vegetation stress. In P. S. Thenkabail, J. G. Lyon & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation* (pp. 265-288). Boca Raton: CRC Press. Taylor & Francis Group.
- Mintz-Oron, S., Meir, S., Malitsky, S., Ruppim, E., Aharoni, A., & Shlomi, T. (2012). Reconstruction of arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc Natl Acad Sci U S A*, 109(1), 339-344. doi: 10.1073/pnas.1100358109
- Miras-Avalos, J. M., & Intrigliolo, D. S. (2017). Grape composition under abiotic constraints: Water stress and salinity. *Front Plant Sci*, 8, 851. doi: 10.3389/fpls.2017.00851
- Mirás-Avalos, J. M., Trigo-Córdoba, E., & Bouzas-Cid, Y. (2014). Does predawn water potential discern between irrigation treatments in galician white grapevine cultivars? *OENO One*, 48(2), 123-127. doi: 10.20870/oeno-one.2014.48.2.1566
- Mishra, B., Kumar, N., & Mukhtar, M. S. (2019). Systems biology and machine learning in plant-pathogen interactions. *Mol Plant Microbe Interact*, 32(1), 45-55. doi: 10.1094/MPMI-08-18-0221-FI
- Mishra, P., Roger, J.-M., Jouan-Rimbaud-Bouveresse, D., Biancolillo, A., Marini, F., Nordon, A., & Rutledge, D. N. (2021). Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC Trends in Analytical Chemistry*, 137. doi: 10.1016/j.trac.2021.116206
- Mishra, P., Passos, D., Marini, F., Xu, J., Amigo, J. M., Gowen, A. A., . . . Nordon, A. (2022). Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry*, 157, 1-8. doi: 10.1016/j.trac.2022.116804
- Mishra, P., Verschoor, J., Vries, M. N.-d., Polder, G., & Boer, M. P. (2023). Portable near-infrared spectral imaging combining deep learning and chemometrics for dry matter and soluble solids prediction in intact kiwifruit. *Infrared Physics & Technology*, 131. doi: 10.1016/j.infrared.2023.104677
- Mishra, P., & Woltering, E. (2023). Semi-supervised robust models for predicting dry matter in mango fruit with near-infrared spectroscopy. *Postharvest Biology and Technology*, 200. doi: 10.1016/j.postharvbio.2023.112335
- Mitić, S. S., Obradović, M. V., Mitić, M. N., Kostić, D. A., Pavlović, A. N., Tošić, S. B., & Stojković, M. D. (2011). Elemental composition of various sour cherry and table grape cultivars using inductively coupled plasma atomic emission spectrometry method (icp-oes). *Food Analytical Methods*, 5(2), 279-286. doi: 10.1007/s12161-011-9232-2
- Moco, S., Capanoglu, E., Tikunov, Y., Bino, R. J., Boyacioglu, D., Hall, R. D., . . . De Vos, R. C. (2007). Tissue specialization at the metabolite level is perceived during the development of tomato fruit. *J Exp Bot*, 58(15-16), 4131-4146. doi: 10.1093/jxb/erm271
- Monteiro-Silva, F., Jorge, P. A. S., & Martins, R. C. (2019). Optical sensing of nitrogen, phosphorus and potassium: A spectrophotometrical approach toward smart nutrient deployment. *Chemosensors*, 7(4), 51. doi: 10.3390/chemosensors7040051

- Monteiro, S., Jorge, & Martins. (2019). Optical sensing of nitrogen, phosphorus and potassium: A spectrophotometrical approach toward smart nutrient deployment. *Chemosensors*, 7(4). doi: 10.3390/chemosensors7040051
- Moros, J., & Laserna, J. (2019). Laser-induced breakdown spectroscopy (libs) of organic compounds: A review. *Applied Spectroscopy*, 73(9), 963-1011. doi: 10.1177/0003702819853252
- Mota, M., Martins, M. J., Policarpo, G., Sprey, L., Pastaneira, M., Almeida, P., . . . Oliveira, C. M. (2022). Nutrient content with different fertilizer management and influence on yield and fruit quality in apple cv. Gala. *Horticulturae*, 8(8). doi: 10.3390/horticulturae8080713
- Mourao, P. R., & Domingues Martinho, V. (2017). Portuguese agriculture and the evolution of greenhouse gas emissions-can vegetables control livestock emissions? *Environ Sci Pollut Res Int*, 24(19), 16107-16119. doi: 10.1007/s11356-017-9257-1
- Najjar, K., & Abu-Khalaf, N. (2021). Non-destructive quality measurement for three varieties of tomato using vis/nir spectroscopy. *Sustainability*, 13(19). doi: 10.3390/su131910747
- Nalepa, J., Marcinkiewicz, M., & Kawulok, M. (2019). Data augmentation for brain-tumor segmentation: A review. *Front Comput Neurosci*, 13, 83. doi: 10.3389/fncom.2019.00083
- Nazki, H., Yoon, S., Fuentes, A., & Park, D. S. (2020). Unsupervised image translation using adversarial networks for improved plant disease recognition. *Computers and Electronics in Agriculture*, 168. doi: 10.1016/j.compag.2019.105117
- Netzer, Y., Munitz, S., Shtein, I., & Schwartz, A. (2019). Structural memory in grapevines: Early season water availability affects late season drought stress severity. *European Journal of Agronomy*, 105, 96-103. doi: 10.1016/j.eja.2019.02.008
- Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., & McBratney, A. B. (2019). Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*, 352, 251-267. doi: 10.1016/j.geoderma.2019.06.016
- Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of nir spectroscopy: A review. *Postharvest Biology and Technology*, 46(2), 99-118. doi: 10.1016/j.postharvbio.2007.06.024
- Nilo-Poyanco, R., Olivares, D., Orellana, A., Hinrichsen, P., & Pinto, M. (2013). Proteomic analysis of grapevine (*vitis vinifera* L.) leaf changes induced by transition to autotrophy and exposure to high light irradiance. *J Proteomics*, 91, 309-330. doi: 10.1016/j.jprot.2013.07.004
- Nishat, S., Jafry, A. T., Martinez, A. W., & Awan, F. R. (2021). Paper-based microfluidics: Simplified fabrication and assay methods. *Sensors and Actuators B: Chemical*, 336. doi: 10.1016/j.snb.2021.129681
- NIST. (2020, 20/05/2020). Nist atomic spectra database lines form Retrieved from 20/05/2020, https://www.physics.nist.gov/PhysRefData/ASD/lines_form.html
- Nogales-Bueno, J., Baca-Bocanegra, B., Rooney, A., Hernandez-Hierro, J. M., Byrne, H. J., & Heredia, F. J. (2017). Study of phenolic extractability in grape seeds by means of atr-ftir and raman spectroscopy. *Food Chemistry*, 232, 602-609. doi: 10.1016/j.foodchem.2017.04.049
- Norberto, M., Sillero, N., Coimbra, J., & Cunha, M. (2023). Filling the maize yield gap based on precision agriculture – a maxent approach. *Computers and Electronics in Agriculture*, 211. doi: 10.1016/j.compag.2023.107970

- Nunes, L. C., Batista Braga, J. W., Trevizan, L. C., Florêncio de Souza, P., Arantes de Carvalho, G. G., Júnior, D. S., . . . Krug, F. J. (2010). Optimization and validation of a libs method for the determination of macro and micronutrients in sugar cane leaves. *Journal of Analytical Atomic Spectrometry*, 25(9). doi: 10.1039/c003620j
- Oduntan, Y. A., Stubbs, C. J., & Robertson, D. J. (2022). High throughput phenotyping of cross-sectional morphology to assess stalk lodging resistance. *Plant Methods*, 18(1), 1. doi: 10.1186/s13007-021-00833-3
- Oliveira, L. F. R. d., & Santana, R. C. (2020). Estimation of leaf nutrient concentration from hyperspectral reflectance in eucalyptus using partial least squares regression. *Scientia Agrícola*, 77(6). doi: 10.1590/1678-992x-2018-0409
- Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nat Biotechnol*, 28(3), 245-248. doi: 10.1038/nbt.1614
- OSCAR. (2020, 20/05/2020). Oscar-optical science center for applied research at delaware state university libs database Retrieved 20/05/2020, from <https://oscar.desu.edu/libs/>
- Ozcan, M. M., Juhaimi, F. A., Gulcu, M., Uslu, N., Gecgel, U., Ghafoor, K., & Dursun, N. (2017). Effect of harvest time on physico-chemical properties and bioactive compounds of pulp and seeds of grape varieties. *Journal of Food Science and Technology*, 54(8), 2230-2240. doi: 10.1007/s13197-017-2658-9
- Özcan, M. M., Al Juhaimi, F., Gülcü, M., Uslu, N., & Geçgel, Ü. (2017). Determination of bioactive compounds and mineral contents of seedless parts and seeds of grapes. *South African Journal of Enology & Viticulture*, 38(2), 212-220. doi: 10.21548/38-2-1605
- Pallas, B., & Christophe, A. (2015). Relationships between biomass allocation, axis organogenesis and organ expansion under shading and water deficit conditions in grapevine. *Funct Plant Biol*, 42(12), 1116-1128. doi: 10.1071/FP15168
- Palliotti, A., & Poni, S. (2015). Grapevine under light and heat stresses *Grapevine in a changing environment* (pp. 148-178).
- Pampuri, A., Tugnolo, A., Bianchi, D., Giovenzana, V., Beghi, R., Fontes, N., . . . Guidetti, R. (2021a). Optical specifications for a proximal sensing approach to monitor the vine water status in a distributed and autonomous fashion. *Biosystems Engineering*. doi: 10.1016/j.biosystemseng.2021.11.007
- Pampuri, A., Tugnolo, A., Giovenzana, V., Casson, A., Guidetti, R., & Beghi, R. (2021b). Design of cost-effective led based prototypes for the evaluation of grape (vitis vinifera L.) ripeness. *Computers and Electronics in Agriculture*, 189. doi: 10.1016/j.compag.2021.106381
- Pandey, P. C., & Pandey, M. (2023). Highlighting the role of agriculture and geospatial technology in food security and sustainable development goals. *Sustainable Development*, 31(5), 3175-3195. doi: 10.1002/sd.2600
- Partelli, F. L., Vieira, H. D., de Carvalho, V. B., & de Assis Alves Mourão Filho, F. (2007). Diagnosis and recommendation integrated system norms, sufficiency range, and nutritional evaluation of arabian coffee in two sampling periods. *Journal of Plant Nutrition*, 30(10), 1651-1667. doi: 10.1080/01904160701615525
- Pascale, S. D., Maggio, A., Fogliano, V., Ambrosino, P., & Ritieni, A. (2015). Irrigation with saline water improves carotenoids content and antioxidant activity of tomato. *The Journal of Horticultural Science and Biotechnology*, 76(4), 447-453. doi: 10.1080/14620316.2001.11511392
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives - a review. *Analytica Chimica Acta*, 1026, 8-36. doi: 10.1016/j.aca.2018.04.004
- Pazhamala, L. T., Kudapa, H., Weckwerth, W., Millar, A. H., & Varshney, R. K. (2021). Systems biology for crop improvement. *Plant Genome*, 14(2), e20098. doi: 10.1002/tpg2.20098

- Peng, J., Liu, F., Zhou, F., Song, K., Zhang, C., Ye, L., & He, Y. (2016). Challenging applications for multi-element analysis by laser-induced breakdown spectroscopy in agriculture: A review. *TrAC Trends in Analytical Chemistry*, 85, 260-272. doi: 10.1016/j.trac.2016.08.015
- Peñuelas, J., & Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, 3(4), 151-156. doi: 10.1016/S1360-1385(98)01213-8
- Pérez-Zavala, R., Torres-Torriti, M., Cheein, F. A., & Troni, G. (2018). A pattern recognition strategy for visual grape bunch detection in vineyards. *Computers and Electronics in Agriculture*, 151, 136-149. doi: 10.1016/j.compag.2018.05.019
- Petrie, P. R., Wang, Y., Liu, S., Lam, S., Whitty, M. A., & Skewes, M. A. (2019). The accuracy and utility of a low cost thermal camera and smartphone-based system to assess grapevine water status. *Biosystems Engineering*, 179, 126-139. doi: 10.1016/j.biosystemseng.2019.01.002
- Phatak, A., & De Jong, S. (1997). The geometry of partial least squares. *Journal of Chemometrics*, 11(4), 311-338. doi: 10.1002/(sici)1099-128x(199707)11:4<311::Aid-cem478>3.0.Co;2-4
- Piazzolla, F., Pati, S., Amodio, M. L., & Colelli, G. (2016). Effect of harvest time on table grape quality during on-vine storage. *J Sci Food Agric*, 96(1), 131-139. doi: 10.1002/jsfa.7072
- Poblete-Echeverría, C., Daniels, A. J., Nieuwoudt, H. H., & Opara, U. L. (2020). *Artificial neural network as alternative method for prediction of sugar and acidity using near-infrared spectroscopy in table grapes*.
- Poblete, T., Ortega-Farias, S., Moreno, M. A., & Bardeen, M. (2017). Artificial neural network to predict vine water status spatial variability using multispectral information obtained from an unmanned aerial vehicle (uav). *Sensors (Basel)*, 17(11). doi: 10.3390/s17112488
- Pôças, I., Rodrigues, A., Gonçalves, S., Costa, P., Gonçalves, I., Pereira, L., & Cunha, M. (2015). Predicting grapevine water status based on hyperspectral reflectance vegetation indices. *Remote Sensing*, 7(12), 16460-16479. doi: 10.3390/rs71215835
- Pôças, I., Gonçalves, J., Costa, P. M., Gonçalves, I., Pereira, L. S., & Cunha, M. (2017). Hyperspectral-based predictive modelling of grapevine water status in the portuguese douro wine region. *International Journal of Applied Earth Observation and Geoinformation*, 58, 177-190. doi: 10.1016/j.jag.2017.02.013
- Pôças, I., Tosin, R., Gonçalves, I., & Cunha, M. (2020). Toward a generalized predictive model of grapevine water status in douro region from hyperspectral data. *Agricultural and Forest Meteorology*, 280. doi: 10.1016/j.agrformet.2019.107793
- Poolman, M. G., Kundu, S., Shaw, R., & Fell, D. A. (2013). Responses to light intensity in a genome-scale model of rice metabolism. *Plant Physiol*, 162(2), 1060-1072. doi: 10.1104/pp.113.216762
- Porat, R., Lichter, A., Terry, L. A., Harker, R., & Buzby, J. (2018). Postharvest losses of fruit and vegetables during retail and in consumers' homes: Quantifications, causes, and means of prevention. *Postharvest Biology and Technology*, 139, 135-149. doi: 10.1016/j.postharvbio.2017.11.019
- Potortiota, A. G., Lo Turco, V., Saitta, M., Bua, G. D., Tropea, A., Dugo, G., & Di Bella, G. (2017). Chemometric analysis of minerals and trace elements in sicilian wines from two different grape cultivars. *Nat Prod Res*, 31(9), 1000-1005. doi: 10.1080/14786419.2016.1261341
- Power, A., Truong, V. K., Chapman, J., & Cozzolino, D. (2019). From the laboratory to the vineyard-evolution of the measurement of grape composition using nir spectroscopy towards high-throughput analysis. *High Throughput*, 8(4). doi: 10.3390/ht8040021

- Prakash, J., Dey, A., Uppal, S., Alexander, R., Kaushal, A., Misra, H. S., & Dasgupta, K. (2021). Label-free rapid electrochemical detection of DNA hybridization using ultrasensitive standalone cnt aerogel biosensor. *Biosens Bioelectron*, 191, 113480. doi: 10.1016/j.bios.2021.113480
- Prananto, J. A., Minasny, B., & Weaver, T. (2020). Chapter one - near infrared (nir) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues. In D. L. Sparks (Ed.), *Advances in agronomy* (Vol. 164, pp. 1-49): Academic Press.
- Prananto, J. A., Minasny, B., & Weaver, T. (2021). Rapid and cost-effective nutrient content analysis of cotton leaves using near-infrared spectroscopy (nirs). *PeerJ*, 9, e11042. doi: 10.7717/peerj.11042
- Puig-Sirera, À., Antichi, D., Warren Raffa, D., & Rallo, G. (2021). Application of remote sensing techniques to discriminate the effect of different soil management treatments over rainfed vineyards in chianti terroir. *Remote Sensing*, 13(4). doi: 10.3390/rs13040716
- R Core Team. (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- R Core Team. (2021). R: A language and environment for statistical computing: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rahim, U. F., Utsumi, T., & Mineno, H. (2022). Deep learning-based accurate grapevine inflorescence and flower quantification in unstructured vineyard images acquired using a mobile sensing platform. *Computers and Electronics in Agriculture*, 198. doi: 10.1016/j.compag.2022.107088
- Rai, A., Yamazaki, M., & Saito, K. (2019). A new era in plant functional genomics. *Current Opinion in Systems Biology*, 15, 58-67. doi: 10.1016/j.coisb.2019.03.005
- Rakovský, J., Čermák, P., Musset, O., & Veis, P. (2014). A review of the development of portable laser induced breakdown spectroscopy and its applications. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 101, 269-287. doi: 10.1016/j.sab.2014.09.015
- Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J. A. M., & Scholten, T. (2013). The spectrum-based learner: A new local approach for modeling soil vis-nir spectra of complex datasets. *Geoderma*, 195-196, 268-279. doi: 10.1016/j.geoderma.2012.12.014
- Rapaport, T., Hochberg, U., Shoshany, M., Karnieli, A., & Rachmilevitch, S. (2015). Combining leaf physiology, hyperspectral imaging and partial least squares-regression (pls-r) for grapevine water status assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 88-97. doi: 10.1016/j.isprsjprs.2015.09.003
- Rapaport, T., Hochberg, U., Cochavi, A., Karnieli, A., & Rachmilevitch, S. (2017). The potential of the spectral 'water balance index' (wabi) for crop irrigation scheduling. *New Phytologist*, 216(3), 741-757. doi: 10.1111/nph.14718
- Reis-Pereira, M., Tosin, R., Martins, R., Neves Dos Santos, F., Tavares, F., & Cunha, M. (2022). Kiwi plant canker diagnosis using hyperspectral signal processing and machine learning: Detecting symptoms caused by pseudomonas syringae pv. Actinidiae. *Plants (Basel)*, 11(16). doi: 10.3390/plants11162154
- Reis Pereira, M., Dos Santos, F. N., Tavares, F., & Cunha, M. (2023). Enhancing host-pathogen phenotyping dynamics: Early detection of tomato bacterial diseases using hyperspectral point measurement and predictive modeling. *Front Plant Sci*, 14, 1242201. doi: 10.3389/fpls.2023.1242201
- Reshef, N., Walbaum, N., Agam, N., & Fait, A. (2017). Sunlight modulates fruit metabolic profile and shapes the spatial pattern of compound accumulation within the grape cluster. *Front Plant Sci*, 8, 70. doi: 10.3389/fpls.2017.00070

- Ribalta-Pizarro, C., Munoz, P., & Munne-Bosch, S. (2021). Tissue-specific hormonal variations in grapes of irrigated and non-irrigated grapevines (*Vitis vinifera* cv. "Merlot") growing under mediterranean field conditions. *Frontiers in Plant Science*, 12, 621587. doi: 10.3389/fpls.2021.621587
- Ribeiro, R., Capela, D., Ferreira, M., Martins, R., Jorge, P., Guimarães, D., & Lima, A. (2021). X-ray fluorescence and laser-induced breakdown spectroscopy analysis of li-rich minerals in veins from Argemela tin mine, central Portugal. *Minerals*, 11(11). doi: 10.3390/min11111169
- Ristic, R., & Iland, P. G. (2005). Relationships between seed and berry development of *Vitis vinifera* L. Cv Shiraz: Developmental changes in seed morphology and phenolic composition. *Australian Journal of Grape and Wine Research*, 11(1), 43-58. doi: 10.1111/j.1755-0238.2005.tb00278.x
- Roberts, D., & Mattoo, A. (2018). Sustainable agriculture—enhancing environmental benefits, food nutritional quality and building crop resilience to abiotic and biotic stresses. *Agriculture*, 8(1). doi: 10.3390/agriculture8010008
- Rodríguez-Pérez, J. R., Riaño, D., Carlisle, E., Ustin, S., & Smart, D. R. (2007). Evaluation of hyperspectral reflectance indexes to detect grapevine water status in vineyards. *American Journal of Enology and Viticulture*, 58(3), 302-317.
- Rodríguez-Pulido, F. J., Barbin, D. F., Sun, D.-W., Gordillo, B., González-Miret, M. L., & Heredia, F. J. (2013). Grape seed characterization by NIR hyperspectral imaging. *Postharvest Biology and Technology*, 76, 74-82. doi: 10.1016/j.postharvbio.2012.09.007
- Rogiers, S. Y., Greer, D. H., Hatfield, J. M., Orchard, B. A., & Keller, M. (2006). Mineral sinks within ripening grape berries (*Vitis vinifera* L.). *Vitis - Journal of Grapevine Research*, 45(3), 115.
- Rogiers, S. Y., Greer, D. H., Liu, Y., Baby, T., & Xiao, Z. (2022). Impact of climate change on grape berry ripening: An assessment of adaptation strategies for the Australian vineyard. *Front Plant Sci*, 13, 1094633. doi: 10.3389/fpls.2022.1094633
- Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jimenez-Berni, J., Pinto, F., & Ober, E. S. (2019). Review: New sensors and data-driven approaches—a path to next generation phenomics. *Plant Sci*, 282, 2-10. doi: 10.1016/j.plantsci.2019.01.011
- Rolle, L., Giacosa, S., Gerbi, V., & Novello, V. (2011). Comparative study of texture properties, color characteristics, and chemical composition of ten white table-grape varieties. *American Journal of Enology and Viticulture*, 62(1), 49-56. doi: 10.5344/ajev.2010.10029
- Romero, M., Luo, Y., Su, B., & Fuentes, S. (2018). Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for irrigation scheduling management. *Computers and Electronics in Agriculture*, 147, 109-117. doi: 10.1016/j.compag.2018.02.013
- Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., . . . Yermiyahu, U. (2013). Estimating olive leaf nitrogen concentration using visible and near-infrared spectral reflectance. *Biosystems Engineering*, 114(4), 426-434. doi: 10.1016/j.biosystemseng.2012.09.005
- Rousserie, P., Rabot, A., & Geny-Denis, L. (2019). From flavanols biosynthesis to wine tannins: What place for grape seeds? *Journal of Agricultural and Food Chemistry*, 67(5), 1325-1343. doi: 10.1021/acs.jafc.8b05768
- Rouxinol, M. I., Martins, M. R., Murta, G. C., Mota Barroso, J., & Rato, A. E. (2022). Quality assessment of red wine grapes through NIR spectroscopy. *Agronomy*, 12(3). doi: 10.3390/agronomy12030637
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *arXiv pre-print server*. doi: None arxiv:1811.10154

- Rustioni, L., Basilico, R., Fiori, S., Leoni, A., Maghradze, D., & Failla, O. (2013). Grape colour phenotyping: Development of a method based on the reflectance spectrum. *Phytochemical Analysis*, 24(5), 453-459. doi: 10.1002/pca.2434
- Rustioni, L., Maghradze, D., & Failla, O. (2016). Optical properties of berry epicuticular waxes in four georgian grape cultivars (vitis vinifera l.). *South African Journal of Enology and Viticulture*, 33(2). doi: 10.21548/33-2-1111
- Sadani, K., Nag, P., & Mukherji, S. (2019). Lspr based optical fiber sensor with chitosan capped gold nanoparticles on bsa for trace detection of hg (ii) in water, soil and food samples. *Biosensors and Bioelectronics*, 134, 90-96. doi: <https://doi.org/10.1016/j.bios.2019.03.046>
- Saha, R., Suthers, P. F., & Maranas, C. D. (2011). Zea mays irs1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS One*, 6(7), e21784. doi: 10.1371/journal.pone.0021784
- Saiz-Rubio, V., Rovira-Más, F., Cuenca-Cuenca, A., & Alves, F. (2021). Robotics-based vineyard water potential monitoring at high resolution. *Computers and Electronics in Agriculture*, 187. doi: 10.1016/j.compag.2021.106311
- Saladié, M., Matas, A. J., Isaacson, T., Jenks, M. A., Goodwin, S. M., Niklas, K. J., . . . Rose, J. K. C. (2007). A reevaluation of the key factors that influence tomato fruit softening and integrity. *Plant Physiology*, 144(2), 1012-1028. doi: 10.1104/pp.107.097477
- Sandalio, L. M., Rodríguez-Serrano, M., Romero-Puertas, M. C., & del Río, L. A. (2013). Role of peroxisomes as a source of reactive oxygen species (ros) signaling molecules. In L. A. del Río (Ed.), *Peroxisomes and their key role in cellular signaling and metabolism* (pp. 231-255). Dordrecht: Springer Netherlands.
- Santesteban, L. G., Miranda, C., Marín, D., Sesma, B., Intrigliolo, D. S., Mirás-Avalos, J. M., . . . Royo, J. B. (2019). Discrimination ability of leaf and stem water potential at different times of the day through a meta-analysis in grapevine (vitis vinifera l.). *Agricultural Water Management*, 221, 202-210. doi: 10.1016/j.agwat.2019.04.020
- Santos-Campos, M., Tosin, R., Rodrigues, L., Gonçalves, I., Barbosa, C., Martins, R., . . . Cunha, M. (2023). *Enhancing grape brix prediction in precision viticulture: A benchmarking study of predictive models using hyperspectral proximal sensors*. Paper presented at the The 3rd International Electronic Conference on Agronomy.
- Santos, A. O., & Kaye, O. (2009). Grapevine leaf water potential based upon near infrared spectroscopy. *Scientia Agricola*, 66, 287-292. doi: <https://doi.org/10.1590/S0103-90162009000300001>
- Santoso, H., Tani, H., Wang, X., & Segah, H. (2018). Predicting oil palm leaf nutrient contents in kalimantan, indonesia by measuring reflectance with a spectroradiometer. *International Journal of Remote Sensing*, 40(19), 7581-7602. doi: 10.1080/01431161.2018.1516323
- Sassu, A., Gambella, F., Ghiani, L., Mercenaro, L., Caria, M., & Pazzona, A. L. (2021). Advances in unmanned aerial system remote sensing for precision viticulture. *Sensors (Basel)*, 21(3). doi: 10.3390/s21030956
- Savoi, S., Wong, D. C., Arapitsas, P., Miculan, M., Bucchetti, B., Peterlunger, E., . . . Castellarin, S. D. (2016). Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (vitis vinifera l.). *BMC Plant Biol*, 16, 67. doi: 10.1186/s12870-016-0760-1
- Scholander, P. F., Bradstreet, E. D., Hemmingsen, E. A., & Hammel, H. T. (1965). Sap pressure in vascular plants. *Negative hydrostatic pressure can be measured in plants*, 148(3668), 339-346. doi: 10.1126/science.148.3668.339

- Schreiber, U., Bilger, W., & Neubauer, C. (1995). Chlorophyll fluorescence as a noninvasive indicator for rapid assessment of in vivo photosynthesis *Ecophysiology of photosynthesis* (pp. 49-70): Springer.
- Schreiner, R. P., & Scagel, C. F. (2017). Leaf blade versus petiole nutrient tests as predictors of nitrogen, phosphorus, and potassium status of 'pinot noir' grapevines. *HortScience*, 52(1), 174-184. doi: 10.21273/hortsci11405-16
- Schwarz, D., Thompson, A. J., & Klaring, H. P. (2014). Guidelines to use tomato in experiments with a controlled environment. *Front Plant Sci*, 5, 625. doi: 10.3389/fpls.2014.00625
- Segade, S. R., Giacosa, S., Gerbi, V., & Rolle, L. (2011). Berry skin thickness as main texture parameter to predict anthocyanin extractability in winegrapes. *LWT - Food Science and Technology*, 44(2), 392-398. doi: 10.1016/j.lwt.2010.09.004
- Senesi, G. S., Cabral, J., Menegatti, C. R., Marangoni, B., & Nicolodelli, G. (2019). Recent advances and future trends in libs applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part ii. Crop plants and their food derivatives. *TrAC Trends in Analytical Chemistry*, 118, 453-469. doi: 10.1016/j.trac.2019.05.052
- Serrano, A., Espinoza, C., Armijo, G., Inostroza-Blancheteau, C., Poblete, E., Meyer-Regueiro, C., . . . Arce-Johnson, P. (2017). Omics approaches for understanding grapevine berry development: Regulatory networks associated with endogenous processes and environmental responses. *Front Plant Sci*, 8, 1486. doi: 10.3389/fpls.2017.01486
- Serrano, L., González-Flor, C., & Gorchs, G. (2010). Assessing vineyard water status using the reflectance based water index. *Agriculture, Ecosystems & Environment*, 139(4), 490-499. doi: 10.1016/j.agee.2010.09.007
- Sharifi, A. (2020). Remotely sensed vegetation indices for crop nutrition mapping. *J Sci Food Agric*, 100(14), 5191-5196. doi: 10.1002/jsfa.10568
- Shen, G., Lesnoff, M., Baeten, V., Dardenne, P., Davrieux, F., Ceballos, H., . . . Fernández Pierna, J. A. (2019). Local partial least squares based on global pls scores. *Journal of Chemometrics*, 33(5). doi: 10.1002/cem.3117
- Shmulevitz, R., Amato, A., Previtali, P., Green, E., Sanchez, L., Alsina, M. M., . . . Fasoli, M. (2024). Spatial variability of grape berry maturation program at the molecular level. *Horticulturae*, 10(3). doi: 10.3390/horticulturae10030238
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1). doi: 10.1186/s40537-019-0197-0
- Shrestha, S., Knapič, M., Žibrat, U., Deleuran, L. C., & Gislum, R. (2016). Single seed near-infrared hyperspectral imaging in determining tomato (*solanum lycopersicum* L.) seed quality in association with multivariate data analysis. *Sensors and Actuators B: Chemical*, 237, 1027-1034. doi: 10.1016/j.snb.2016.08.170
- Si, Y., & Sankaran, S. (2016). Computed tomography imaging-based bitter pit evaluation in apples. *Biosystems Engineering*, 151, 9-16. doi: 10.1016/j.biosystemseng.2016.08.008
- Silva, A. F., Löfkvist, K., Gilbertsson, M., Os, E. V., Franken, G., Balendonck, J., . . . Martins, R. (2021). Hydroponics monitoring through uv-vis spectroscopy and artificial intelligence: Quantification of nitrogen, phosphorous and potassium. *CSAC2021: 1st International Electronic Conference on Chemical Sensors and Analytical Chemistry*. doi: <https://doi.org/10.3390/CSAC2021-10448>
- Silva, C. A. A. C., Fiorio, P. R., Rizzo, R., Rossetto, R., Vitti, A. C., Dias, F. L. F., . . . Bárbara Neto, M. (2023a). Detection of nutritional stress in sugarcane by vis-nir-swir reflectance spectroscopy. *Ciência Rural*, 53(12). doi: 10.1590/0103-8478cr20220543

- Silva, F. M., Queirós, C., Pinho, T., Boaventura, J., Santos, F., Barroso, T. G., . . . Martins, R. C. (2023b). Reagent-less spectroscopy towards npk sensing for hydroponics nutrient solutions. *Sensors and Actuators B: Chemical*, 395. doi: 10.1016/j.snb.2023.134442
- Silva, L. R., & Queiroz, M. (2016). Bioactive compounds of red grapes from dão region (portugal): Evaluation of phenolic and organic profile. *Asian Pacific Journal of Tropical Biomedicine*, 6(4), 315-321. doi: 10.1016/j.apjtb.2015.12.015
- Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2), 337-354. doi: [https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/10.1016/S0034-4257(02)00010-X)
- Sims, D. A., & Gamon, J. A. (2003). Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: A comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sensing of Environment*, 84(4), 526-537. doi: 10.1016/s0034-4257(02)00151-7
- Singh, V. K. (2021). Review: Application of libs to elemental analysis and mapping of plant samples. *Atomic Spectroscopy*, 42(1). doi: 10.46770/as.2020.201
- Skolik, P., Morais, C. L. M., Martin, F. L., & McAinsh, M. R. (2019). Determination of developmental and ripening stages of whole tomato fruit using portable infrared spectroscopy and chemometrics. *BMC Plant Biol*, 19(1), 236. doi: 10.1186/s12870-019-1852-5
- Slowikowski, K. (2021). Ggrepel: Automatically position non-overlapping text labels with 'ggplot2'. Retrieved from <https://CRAN.R-project.org/package=ggrepel>
- Sohaib Ali Shah, S., Zeb, A., Qureshi, W. S., Arslan, M., Ullah Malik, A., Alasmay, W., & Alanazi, E. (2020). Towards fruit maturity estimation using nir spectroscopy. *Infrared Physics & Technology*, 111. doi: 10.1016/j.infrared.2020.103479
- Soubeyrand, E., Colombie, S., Beauvoit, B., Dai, Z., Cluzet, S., Hilbert, G., . . . Gomes, E. (2018). Constraint-based modeling highlights cell energy, redox status and alpha-ketoglutarate availability as metabolic drivers for anthocyanin accumulation in grape cells under nitrogen limitation. *Front Plant Sci*, 9, 421. doi: 10.3389/fpls.2018.00421
- Sousa, N. J. C., Martins, R. M. D. C., & Osorio, R. A. A. (2014). Optical system for parameter characterization of an element of body fluid or tissue.
- Sparrow, A. M., Dambergs, R. G., Bindon, K. A., Smith, P. A., & Close, D. C. (2015). Interactions of grape skin, seed, and pulp on tannin and anthocyanin extraction in pinot noir wines. *American Journal of Enology and Viticulture*, 66(4), 472-481. doi: 10.5344/ajev.2015.15022
- Spayd, S. E., Tarara, J. M., Mee, D. L., & Ferguson, J. C. (2002). Separation of sunlight and temperature effects on the composition of *vitis vinifera* cv. Merlot berries. *American Journal of Enology and Viticulture*, 53(3), 171-182. doi: 10.5344/ajev.2002.53.3.171
- Stelzle, F., Adler, W., Zam, A., Tangermann-Gerk, K., Knipfer, C., Douplik, A., . . . Nkenke, E. (2012). In vivo optical tissue differentiation by diffuse reflectance spectroscopy: Preliminary results for tissue-specific laser surgery. *Surg Innov*, 19(4), 385-393. doi: 10.1177/1553350611429692
- Sternad Lemut, M., Trost, K., Sivilotti, P., & Vrhovsek, U. (2011). Pinot noir grape colour related phenolics as affected by leaf removal treatments in the vipava valley. *Journal of Food Composition and Analysis*, 24(6), 777-784. doi: 10.1016/j.jfca.2011.03.003
- Stiletto, A., & Trestini, S. (2021). Factors behind consumers' choices for healthy fruits: A review of pomegranate and its food derivatives. *Agricultural and Food Economics*, 9(1). doi: 10.1186/s40100-021-00202-7

- Sun, L., Gao, F., Anderson, M., Kustas, W., Alsina, M., Sanchez, L., . . . Post, K. (2017a). Daily mapping of 30 m lai and ndvi for grape yield prediction in california vineyards. *Remote Sensing*, 9(4). doi: 10.3390/rs9040317
- Sun, R. Z., Cheng, G., Li, Q., He, Y. N., Wang, Y., Lan, Y. B., . . . Wang, J. (2017b). Light-induced variation in phenolic compounds in cabernet sauvignon grapes (vitis vinifera l.) involves extensive transcriptome reprogramming of biosynthetic enzymes, transcription factors, and phytohormonal regulators. *Front Plant Sci*, 8, 547. doi: 10.3389/fpls.2017.00547
- Talari, A. C. S., Rehman, S., & Rehman, I. U. (2019). Advancing cancer diagnostics with artificial intelligence and spectroscopy: Identifying chemical changes associated with breast cancer. *Expert Review of Molecular Diagnostics*, 19(10), 929-940. doi: 10.1080/14737159.2019.1659727
- Tamasi, G., Pardini, A., Bonechi, C., Donati, A., Pessina, F., Marcolongo, P., . . . Rossi, C. (2019). Characterization of nutraceutical components in tomato pulp, skin and locular gel. *European Food Research and Technology*, 245(4), 907-918. doi: 10.1007/s00217-019-03235-x
- Tan, H., Hu, Y., Ma, B., Yu, G., & Li, Y. (2024). An improved dcgan model: Data augmentation of hyperspectral image for identification pesticide residues of hami melon. *Food Control*, 157. doi: 10.1016/j.foodcont.2023.110168
- Tang, Y., Jones, E., & Minasny, B. (2020). Evaluating low-cost portable near infrared sensors for rapid analysis of soils from south eastern australia. *Geoderma Regional*, 20. doi: 10.1016/j.geodrs.2019.e00240
- Tantalaki, N., Souravlas, S., & Roumeliotis, M. (2019). Data-driven decision making in precision agriculture: The rise of big data in agricultural systems. *Journal of Agricultural & Food Information*, 20(4), 344-380. doi: 10.1080/10496505.2019.1638264
- Tardaguila, J., Fernandez-Novales, J., Gutierrez, S., & Diago, M. P. (2017). Non-destructive assessment of grapevine water status in the field using a portable nir spectrophotometer. *Journal of the Science of Food and Agriculture*, 97(11), 3772-3780. doi: 10.1002/jsfa.8241
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., & Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr Biol*, 27(15), R770-R783. doi: 10.1016/j.cub.2017.05.055
- Tarpley, L., Reddy, K. R., & Sassenrath-Cole, G. F. (2000). Reflectance indices with precision and accuracy in predicting cotton leaf nitrogen concentration. *Crop Science*, 40(6), 1814-1819. doi: <https://doi.org/10.2135/cropsci2000.4061814x>
- Teixeira, A., Eiras-Dias, J., Castellarin, S. D., & Geros, H. (2013). Berry phenolics of grapevine under challenging environments. *Int J Mol Sci*, 14(9), 18711-18739. doi: 10.3390/ijms140918711
- Tessarini, P., Parpinello, G. P., & Rombolà, A. D. (2018). Physiological and enological implications of postveraison trimming in an organically-managed sangiovese vineyard. *American Journal of Enology and Viticulture*, 69(1), 59-69. doi: 10.5344/ajev.2017.16045
- Tewari, R. K., Yadav, N., Gupta, R., & Kumar, P. (2021). Oxidative stress under macronutrient deficiency in plants. *Journal of Soil Science and Plant Nutrition*, 21(1), 832-859. doi: 10.1007/s42729-020-00405-9
- Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, 71(2), 158-182. doi: [https://doi.org/10.1016/S0034-4257\(99\)00067-X](https://doi.org/10.1016/S0034-4257(99)00067-X)
- Thenkabail, P. S., Lyon, J. G., & Huete, A. (2012). Advances in hyperspectral remote sensing of vegetation and agricultural croplands. In P. S. Thenkabail, J. G. Lyon

- & A. Huete (Eds.), *Hyperspectral remote sensing of vegetation* (pp. 4-35). Boca Raton: CRC Press. Taylor & Francis Group.
- Thomas, H. (2013). Senescence, ageing and death of the whole plant. *New Phytol*, 197(3), 696-711. doi: 10.1111/nph.12047
- Thomas, T. R., Matthews, M. A., & Shackel, K. A. (2006). Direct in situ measurement of cell turgor in grape (*Vitis vinifera* L.) berries during development and in response to plant water deficits. *Plant, Cell & Environment*, 29(5), 993-1001. doi: 10.1111/j.1365-3040.2006.01496.x
- Thompson, M. (2012). *Handbook of inductively coupled plasma spectrometry*: Springer Science & Business Media.
- Tian, B., Harrison, R., Morton, J., & Deb-Choudhury, S. (2015). Proteomic analysis of sauvignon blanc grape skin, pulp and seed and relative quantification of pathogenesis-related proteins. *PLoS One*, 10(6), e0130132. doi: 10.1371/journal.pone.0130132
- Tilahun, S., Park, D. S., Seo, M. H., Hwang, I. G., Kim, S. H., Choi, H. R., & Jeong, C. S. (2018). Prediction of lycopene and β -carotene in tomatoes by portable chromameter and vis/nir spectra. *Postharvest Biology and Technology*, 136, 50-56. doi: 10.1016/j.postharvbio.2017.10.007
- Tilly, N., Hoffmeister, D., Cao, Q., Lenz-Wiedemann, V., Miao, Y., & Bareth, G. (2015). Transferability of models for estimating paddy rice biomass from spatial plant height data. *Agriculture*, 5(3), 538-560. doi: 10.3390/agriculture5030538
- Tomar, N., & De, R. K. (2013). Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene*, 521(1), 1-14. doi: 10.1016/j.gene.2013.03.017
- Tomaz, T., Bagard, M., Pracharoenwattana, I., Lindén, P., Lee, C. P., Carroll, A. J., . . . Millar, A. H. (2010). Mitochondrial malate dehydrogenase lowers leaf respiration and alters photorespiration and plant growth in arabidopsis. *Plant Physiology*, 154(3), 1143-1157. doi: 10.1104/pp.110.161612
- Toor, R. K., & Savage, G. P. (2005). Antioxidant activity in different fractions of tomatoes. *Food Research International*, 38(5), 487-494. doi: 10.1016/j.foodres.2004.10.016
- Torres, I., Pérez-Marín, D., Haba, M.-J. D. I., & Sánchez, M.-T. (2015). Fast and accurate quality assessment of raf tomatoes using nirs technology. *Postharvest Biology and Technology*, 107, 9-15. doi: 10.1016/j.postharvbio.2015.04.004
- Tosin, R., Pôças, I., Gonçalves, I., & Cunha, M. (2020). Estimation of grapevine predawn leaf water potential based on hyperspectral reflectance data in douro wine region. *Vitis: Journal of Grapevine Research*, 59(1), 9-18. doi: <https://doi.org/10.5073/vitis.2020.59.9-18>
- Tosin, R., Pôças, I., Novo, H., Teixeira, J., Fontes, N., Graça, A., & Cunha, M. (2021). Assessing predawn leaf water potential based on hyperspectral data and pigment's concentration of *Vitis vinifera* L. in the douro wine region. *Scientia Horticulturae*, 278. doi: 10.1016/j.scienta.2020.109860
- Tosin, R., Martins, R., Pôças, I., & Cunha, M. (2022). Canopy vis-nir spectroscopy and self-learning artificial intelligence for a generalised model of predawn leaf water potential in *Vitis vinifera*. *Biosystems Engineering*, 219, 235-258. doi: 10.1016/j.biosystemseng.2022.05.007
- Tosin, R., Martins, R., & Cunha, M. (2023a). Tomography-like for hyperspectral bi-directional grape tissue reconstruction based on machine learning: Implications for diagnosis composition and precision maturation monitoring. *BIO Web of Conferences*, 68. doi: 10.1051/bioconf/20236801017
- Tosin, R., Monteiro-Silva, F., Martins, R., & Cunha, M. (2023b). Precision maturation assessment of grape tissues: Hyperspectral bi-directional reconstruction using tomography-like based on multi-block hierarchical principal component analysis.

- Biosystems Engineering*, 236, 147-159. doi: 10.1016/j.biosystemseng.2023.10.011
- Tosin, R., Monteiro-Silva, F., Martins, R., & Cunha, M. (2023c). *Libs-based analysis of elemental composition in skin, pulp, and seeds of white and red grape cultivars*. Paper presented at the Csac 2023.
- Tosin, R., Cunha, M., Monteiro-Silva, F., Santos, F., Barroso, T., & Martins, R. (2024a). Bi-directional hyperspectral reconstruction of cherry tomato: Diagnosis of internal tissues maturation stage and composition. *Frontiers in Plant Science*. doi: 10.3389/fpls.2024.1351958
- Tosin, R., Monteiro-Silva, F., Martins, R., & Cunha, M. (2024b). A new approach for element characterization of grapevine tissue with laser-induced breakdown spectroscopy. *Horticulturae*, 10(1), 82. doi: 10.3390/horticulturae10010082
- Trewavas, A. (2006). A brief history of systems biology. "Every object that biology studies is a system of systems." Francois jacob (1974). *Plant Cell*, 18(10), 2420-2430. doi: 10.1105/tpc.106.042267
- Trost, P., Fermani, S., Marri, L., Zaffagnini, M., Falini, G., Scagliarini, S., . . . Sparla, F. (2006). Thioredoxin-dependent regulation of photosynthetic glyceraldehyde-3-phosphate dehydrogenase: Autonomous vs. Cp12-dependent mechanisms. *Photosynthesis Research*, 89(2-3), 263-275. doi: 10.1007/s11120-006-9099-z
- Truong, V. A., Lang, B., & Conroy, D. M. (2022). When food governance matters to consumer food choice: Consumer perception of and preference for food quality certifications. *Appetite*, 168, 105688. doi: 10.1016/j.appet.2021.105688
- Trygg, J., & Wold, S. (2003). O2-pls, a two-block (x-y) latent variable regression (lvr) method with an integral osc filter. *Journal of Chemometrics*, 17(1), 53-64. doi: 10.1002/cem.775
- Ulissi, V., Antonucci, F., Benincasa, P., Farneselli, M., Tosti, G., Guiducci, M., . . . Menesatti, P. (2011). Nitrogen concentration estimation in tomato leaves by vis-nir non-destructive spectroscopy. *Sensors (Basel)*, 11(6), 6411-6424. doi: 10.3390/s110606411
- UN DESA. (2016). Transforming our world : The 2030 agenda for sustainable development Retrieved 19 January 2024, from <https://wedocs.unep.org/20.500.11822/11125>
- USDA. (1991). United states standards for grades of fresh tomatoes: United States Department of Agriculture. Agricultural Marketing Service.
- van Diepen, C. A., Wolf, J., van Keulen, H., & Rappoldt, C. (1989). Wofost: A simulation model of crop production. *Soil Use and Management*, 5(1), 16-24. doi: <https://doi.org/10.1111/j.1475-2743.1989.tb00755.x>
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177. doi: 10.1016/j.compag.2020.105709
- van Maarschalkerweerd, M., & Husted, S. (2015). Recent developments in fast spectroscopy for plant mineral analysis. *Front Plant Sci*, 6, 169. doi: 10.3389/fpls.2015.00169
- Van Wittenberghe, S., Amin, E., Pascual-Ventoe, A. B., Pérez-Suay, A., Tenjo, C., Sabater, N., . . . Moreno, J. (2024). Retrieval of leaf-level fluorescence quantum efficiency and npq-related xanthophyll absorption through spectral unmixing strategies for future vis-nir imaging spectroscopy. *Remote Sensing of Environment*, 300. doi: 10.1016/j.rse.2023.113879
- van Wyngaard, E., Blancquaert, E., Nieuwoudt, H., & Aleixandre-Tudo, J. L. (2021). Infrared spectroscopy and chemometric applications for the qualitative and quantitative investigation of grapevine organs. *Front Plant Sci*, 12, 723247. doi: 10.3389/fpls.2021.723247

- Verdenal, T., Dienes-Nagy, Á., Spangenberg, J. E., Zufferey, V., Spring, J.-L., Viret, O., . . . Van Leeuwen, C. (2021). Understanding and managing nitrogen nutrition in grapevine: A review. *OENO One*, 55(1), 1-43. doi: 10.20870/oenone.2021.55.1.3866
- Verma, M., Gharpure, D. C., & Wagh, V. G. (2021). Non-destructive testing of fruits using electrical impedance tomography: A preliminary study. *AIP Conference Proceedings*, 2335(1), 100003. doi: 10.1063/5.0043734
- Vo-Dinh, T., Zam, A., Grundfest, W. S., Stelzle, F., Tangermann-Gerk, K., Mahadevan-Jansen, A., . . . Douplik, A. (2010). *Tissue differentiation by diffuse reflectance spectroscopy for automated oral and maxillofacial laser surgery: Ex vivo pilot study*. Paper presented at the Advanced Biomedical and Clinical Diagnostic Systems VIII.
- Wagenaar, D., Hermawan, T., van den Homberg, M. J. C., Aerts, J., Kreibich, H., de Moel, H., & Bouwer, L. M. (2021). Improved transferability of data-driven damage models through sample selection bias correction. *Risk Anal*, 41(1), 37-55. doi: 10.1111/risa.13575
- Walker, R. P., Chen, Z.-H., Técsi, L. I., Famiani, F., Lea, P. J., & Leegood, R. C. (1999). Phosphoenolpyruvate carboxykinase plays a role in interactions of carbon and nitrogen metabolism during grape seed development. *Planta*, 210(1), 9-18. doi: 10.1007/s004250050648
- Wang, G., Zeng, F., Song, P., Sun, B., Wang, Q., & Wang, J. (2022). Effects of reduced chlorophyll content on photosystem functions and photosynthetic electron transport rate in rice leaves. *J Plant Physiol*, 272, 153669. doi: 10.1016/j.jplph.2022.153669
- Wawrzynska, A., & Sirko, A. (2014). To control and to be controlled: Understanding the arabidopsis slim1 function in sulfur deficiency through comprehensive investigation of the eil protein family. *Front Plant Sci*, 5, 575. doi: 10.3389/fpls.2014.00575
- Webb, L. B., Whetton, P. H., Bhend, J., Darbyshire, R., Briggs, P. R., & Barlow, E. W. R. (2012). Earlier wine-grape ripening driven by climatic warming and drying and management practices. *Nature Climate Change*, 2(4), 259-264. doi: 10.1038/nclimate1417
- Wei, H.-E., Grafton, M., Bretherton, M., Irwin, M., & Sandoval, E. (2021). Evaluation of point hyperspectral reflectance and multivariate regression models for grapevine water status estimation. *Remote Sensing*, 13(16). doi: 10.3390/rs13163198
- Wencel, D., Kaworek, A., Abel, T., Efremov, V., Bradford, A., Carthy, D., . . . McDonagh, C. (2018). Optical sensor for real-time ph monitoring in human tissue. *Small*, 14(51), 1803627. doi: 10.1002/sml.201803627
- Westerhuis, J. A., Kourti, T., & MacGregor, J. F. (1998). Analysis of multiblock and hierarchical pca and pls models. *Journal of Chemometrics*, 12(5), 301-321. doi: 10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S
- Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43). doi: 10.21105/joss.01686
- Williams, L. E., & Ayars, J. E. (2005). Grapevine water use and the crop coefficient are linear functions of the shaded area measured beneath the canopy. *Agricultural and Forest Meteorology*, 132(3-4), 201-211. doi: 10.1016/j.agrformet.2005.07.010
- Wu, G., & Wang, C. (2014). Investigating the effects of simulated transport vibration on tomato tissue damage based on vis/nir spectroscopy. *Postharvest Biology and Technology*, 98, 41-47. doi: 10.1016/j.postharvbio.2014.06.016

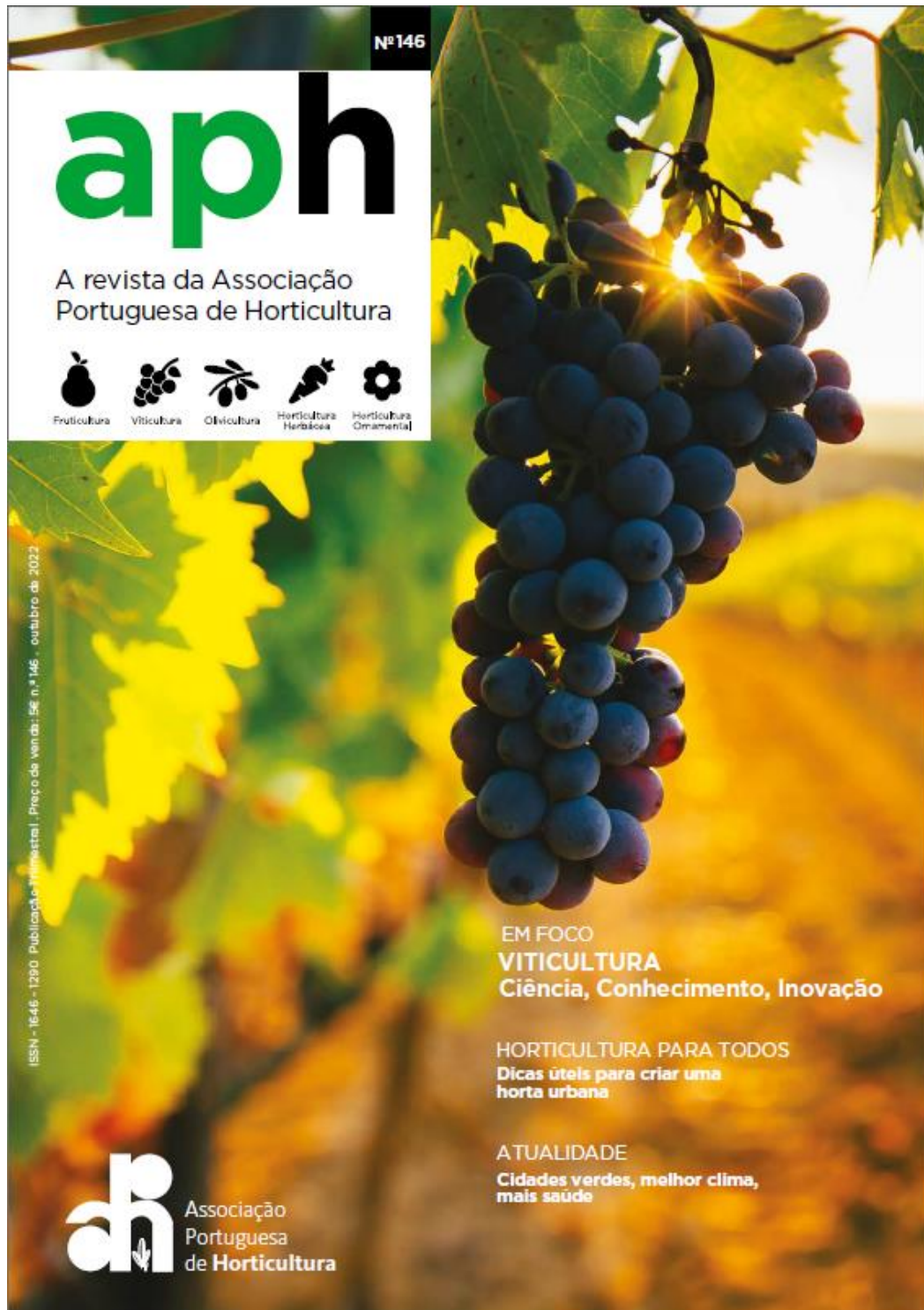
- Wurtzel, E. T., Vickers, C. E., Hanson, A. D., Millar, A. H., Cooper, M., Voss-Fels, K. P., . . . Erb, T. J. (2019). Revolutionizing agriculture with synthetic biology. *Nat Plants*. doi: 10.1038/s41477-019-0539-0
- Xia, J. A., Zhang, W., Zhang, W., Yang, Y., Hu, G., Ge, D., . . . Cao, H. (2021). A cloud computing-based approach using the visible near-infrared spectrum to classify greenhouse tomato plants under water stress. *Computers and Electronics in Agriculture*, 181. doi: 10.1016/j.compag.2020.105966
- Xu, M., Sun, J., Yao, K., Cai, Q., Shen, J., Tian, Y., & Zhou, X. (2022). Developing deep learning based regression approaches for prediction of firmness and ph in kyoho grape using vis/nir hyperspectral imaging. *Infrared Physics & Technology*, 120. doi: 10.1016/j.infrared.2021.104003
- Yandun Narvaez, F., Reina, G., Torres-Torriti, M., Kantor, G., & Cheein, F. A. (2017). A survey of ranging and imaging techniques for precision agriculture phenotyping. *IEEE/ASME Transactions on Mechatronics*, 22(6), 2428-2439. doi: 10.1109/tmech.2017.2760866
- Yang, C., Wang, Y., Wu, B., Fang, J., & Li, S. (2011). Volatile compounds evolution of three table grapes with different flavour during and after maturation. *Food Chemistry*, 128(4), 823-830. doi: 10.1016/j.foodchem.2010.11.029
- Yang, C. (2020). Remote sensing and precision agriculture technologies for crop disease detection and management with a practical application example. *Engineering*, 6(5), 528-532. doi: 10.1016/j.eng.2019.10.015
- Yarce, C. J., & Rojas, G. (2012). Near infrared spectroscopy for the analysis of macro and micro nutrients in sugarcane leaves. *Zuckerindustrie-Sugar Industry*, 137(11), 707.
- Yu, R., Brillante, L., Martinez-Luscher, J., & Kurtural, S. K. (2020). Spatial variability of soil and plant water status and their cascading effects on grapevine physiology are linked to berry and wine chemistry. *Front Plant Sci*, 11, 790. doi: 10.3389/fpls.2020.00790
- Yu, R., Brillante, L., Torres, N., & Kurtural, S. K. (2021). Proximal sensing of vineyard soil and canopy vegetation for determining vineyard spatial variability in plant physiology and berry chemistry. *OENO One*, 55(2), 315-333. doi: 10.20870/oeno-one.2021.55.2.4598
- Yuan, F., Schreiner, R. P., Osborne, J., & Qian, M. C. (2018). Effects of soil npk supply on pinot noir wine phenolics and aroma composition. *American Journal of Enology and Viticulture*, 69(4), 371-385. doi: 10.5344/ajev.2018.17077
- Yuan, H., Cheung, C. Y., Poolman, M. G., Hilbers, P. A., & van Riel, N. A. (2016). A genome-scale metabolic network reconstruction of tomato (*solanum lycopersicum* l.) and its application to photorespiratory metabolism. *Plant J*, 85(2), 289-304. doi: 10.1111/tpj.13075
- Yuan, H., Yang, G., Li, C., Wang, Y., Liu, J., Yu, H., . . . Yang, X. (2017). Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of rf, ann, and svm regression models. *Remote Sensing*, 9(4). doi: 10.3390/rs9040309
- Zahir, S. A. D. M., Omar, A. F., Jamlos, M. F., Azmi, M. A. M., & Muncan, J. (2022). A review of visible and near-infrared (vis-nir) spectroscopy application in plant stress detection. *Sensors and Actuators A: Physical*, 338. doi: 10.1016/j.sna.2022.113468
- Zakhartsev, M., Medvedeva, I., Orlov, Y., Akberdin, I., Krebs, O., & Schulze, W. X. (2016). Metabolic model of central carbon and energy metabolisms of growing arabidopsis thaliana in relation to sucrose translocation. *BMC Plant Biology*, 16(1). doi: 10.1186/s12870-016-0868-3
- Zarco-Tejada, P. J., González-Dugo, V., Williams, L. E., Suárez, L., Berni, J. A. J., Goldhamer, D., & Fereres, E. (2013). A pri-based water stress index combining

- structural and chlorophyll effects: Assessment using diurnal narrow-band airborne imagery and the cwsj thermal index. *Remote Sensing of Environment*, 138, 38-50. doi: 10.1016/j.rse.2013.07.024
- Zerner, M., & Parker, K. (2019). Rapid assessment of crop nitrogen and stress status-in-field assessment of a hand-held near infrared tool Retrieved 4-07-2023, from <https://grdc.com.au/resources-and-publications/grdc-update-papers/tab-content/grdc-update-papers/2019/02/rapid-assessment-of-crop-nitrogen-and-stress-status-in-field-assessment-of-a-hand-held-near-infrared-tool>
- Zhang, P., Dong, T., Jin, H., Pei, D., Pervaiz, T., Ren, Y., . . . Fang, J. (2022). Analysis of photosynthetic ability and related physiological traits in nodal leaves of grape. *Scientia Horticulturae*, 304. doi: 10.1016/j.scienta.2022.111251
- Zhang, S., Cheng, D., Deng, Z., Zong, M., & Deng, X. (2018a). A novel k nn algorithm with data-driven k parameter computation. *Pattern Recognition Letters*, 109, 44-54. doi: 10.1016/j.patrec.2017.09.036
- Zhang, W., Zhang, X., Wang, Y., Zhang, N., Guo, Y., Ren, X., & Zhao, Z. (2018b). Potassium fertilization arrests malate accumulation and alters soluble sugar metabolism in apple fruit. *Biol Open*, 7(12). doi: 10.1242/bio.024745
- Zhao, Y. W., Wang, C. K., Huang, X. Y., & Hu, D. G. (2021). Anthocyanin stability and degradation in plants. *Plant Signal Behav*, 16(12), 1987767. doi: 10.1080/15592324.2021.1987767
- Zhou, T., Lu, D., She, Q., Chen, C., Chen, J., Huang, Z., . . . Lu, Y. (2021). Hypersensitive detection of il-6 on sers substrate calibrated by dual model. *Sensors and Actuators B: Chemical*, 336. doi: 10.1016/j.snb.2021.129597
- Zhou, W., Yang, S., & Wang, P. G. (2017). Matrix effects and application of matrix effect factor. *Bioanalysis*, 9(23), 1839-1844. doi: 10.4155/bio-2017-0214
- Zhu, J., He, W., Yao, J., Yu, Q., Xu, C., Huang, H., & Mhae B. Jandug, C. (2020). Spectral reflectance characteristics and chlorophyll content estimation model of quercus aquifolioides leaves at different altitudes in sejila mountain. *Applied Sciences*, 10(10). doi: 10.3390/app10103636
- Zhu, Q., He, C., Lu, R., Mendoza, F., & Cen, H. (2015). Ripeness evaluation of 'sun bright' tomato using optical absorption and scattering properties. *Postharvest Biology and Technology*, 103, 27-34. doi: 10.1016/j.postharvbio.2015.02.007

_____ ATTACHMENTS

A. Technical paper

Tosin, R., Martins, R.C., Santos, F., and Cunha, M. (2022). Sensores proximais hiperespectrais e algoritmos computacionais para fenotipagem digital de parâmetros fisiológicos da videira. *Rev. Assoc. Port. Hortic.* 2022 146, 24-26.



Sensores proximais hiperespectrais e algoritmos computacionais para fenotipagem digital de parâmetros fisiológicos da videira em viticultura de precisão

Renan Tosin^{1,2}, Rui Martins², Filipe Neves dos Santos² e Mário Cunha^{1,2*}

¹ Faculdade de Ciências da Universidade do Porto (FCUP), Rua do Campo Alegre, Porto, 4169-007, Portugal.

² Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC), Campus da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, Porto, 4200-465, Portugal.

* Autor correspondente: mccunha@fc.up.pt

A Viticultura de precisão (VP) tem privilegiado a utilização de sensores montados em satélites ou *drones*, baseada em abordagens *data-driven* as quais limitadas em conhecimento sobre a fisiologia da videira. Como tal ocorrem insuficiências na transferibilidade útil dos dados entre anos no mesmo local e entre *terroir*. Os sensores proximais hiperespectrais de baixo custo são um contributo tecnológico importante para contornar estas limitações à implementação da VP.

Viticultura de precisão fisiológica

Portugal é um dos produtores de vinho mais importantes a nível mundial em quantidade, qualidade e diversidade que em 2020 atingiu um recorde de vendas de 849M€. Para fazer face à concorrência e aos desafios biofísicos emergentes, os produtores vão depender cada vez mais da Viticultura de Precisão (VP). Por exemplo, a intensidade e frequência de valores estremados de estresse abióticos (ex estresse hídrico) e bióticos verificados nos últimos anos, exigem uma monitorização mais eficiente para suportar práticas culturais inovadoras que permitam aumentar a resiliência do sistema vitícola. Alguns métodos tradicionalmente utilizados para medição de parâmetros fisiológica na vinha, tais como o Potencial Hídrico de Base (PHB) e o controlo de maturação dos bagos (e.x. °Brix) são destrutivos, caros, laboriosos e não permitem a sua espacialização numa ótica de VP o que lhe retira utilidade.

Em VP, a quantidade e a natureza da variação espacial de um sistema vitícola são fatores-chave para atuar de forma diferenciada, na gestão dessa variabilidade, de modo a garantir melhores indicadores agronómicos, económicos e ambientais. Atualmente, a VP baseia-se sobretudo na definição de blocos no interior da vinha cujas características são relativamente uniformes para implementar práticas vitícolas espacialmente

diferenciadas. Esta abordagem designada “*site specific management*” tem privilegiado a aquisição de dados provenientes de sensores remotos montados em satélites ou plataformas aéreas recorrendo à combinação de bandas para definir índices espectrais de vegetação (IV) ou a tecnologias de *data-mining* para processar grandes quantidades de dados.

Ainda que os satélites e *drones* tenham demonstrado grande potencial para estabelecer sistemas de VP, apresentam insuficiências quanto à resolução e rigor das medições, processamento dos dados e compatibilidade com algumas técnicas vitícolas emergentes. A nível espacial, os satélites mais utilizados em viticultura (ex Sentinel 2 e Landsat 8), permitem resoluções de várias bandas até 10 metros, sendo a resolução da ordem de centímetros no caso dos drones. A nível temporal, os satélites, devido a limitações meteorológicas, apenas permitem adquirir algumas imagens por mês, enquanto os drones permitem adquirir imagens com maior frequência, mas os custos são mais elevados. Os sensores com resolução multiespectral ou RGB são os mais utilizados em satélites e *drones* sendo ainda escassa a utilização de sensores hiperespectral.

Em relação ao processamento dos dados dos sensores para suporte à VP, a utilização de IV é, ainda, a abordagem mais utilizada, em que o *Normalized Difference Vegetation Index* (NDVI) assume imperial destaque desde a década de 70. Não obstante a relevância destes IV em estratégias de *Site specific management*, não permitem estabelecer uma relação causal entre as medições do sensor e parâmetros fisiológicos da videira. Também, os cenários climáticos exigem técnicas de mitigação, tais como práticas de sombreamento (ex redes) e/ou reflexão da luz (ex caulino), que impossibilitam a utilização de satélite e *drones*.

Apesar dos profícuos avanços tecnológicos, a abordagem baseada em sensores de baixa resolução não permite aceder ao conhecimento *in-situ* sobre a fisiologia da videira fundamentais para uma gestão diferenciada e eficiente do sistema vitícola. Por outro lado, as abordagens *data-driven* frequentemente associadas a enigmáticos IV não permitem estabelecer relações mecanicistas de causa-efeito para determinar, com o rigor fisiológico, a complexa interação entre solo-clima-planta (Cunha et al., 2022). Como tal, ocorrem limitações na transferibilidade interanual útil dos dados ou mesmo para sua transposição de um *terroir* para outro. Este são alguns dos fatores limitadores do crescimento, implementação e plena adoção de sistemas de VP.

Sensores proximais hiperespectrais em viticultura de precisão

Os sensores de proximidade hiperespectrais solucionam algumas das limitações presentes nos satélites e *drones*, nomeadamente quando montados em plataformas moveis ou mesmo quando utilizados pelo agrónomo numa visita regular de monitorização à vinha. A obtenção das assinaturas espectrais dos diferentes elementos da planta (folha, fruto, caule) permite a obtenção de uma informação muito mais detalhada e precisa que a dos satélites e *drones*, permitindo disponibilizar a informação imediatamente após a aquisição (*on the go*).

Atualmente a sua principal limitação é o custo de aquisição que pode atingir valores em torno de 30 mil euros, o que impede a aquisição por uma grande parte dos viticultores. Todavia, estão em fase de validação sensores hiperspectrais ativos de baixo custo, como por exemplo, o Metbots desenvolvido pelo Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC) no âmbito do projeto Metbots - *Metabolomic robots with self-learning artificial intelligence for precision agriculture* (Martins et al., 2019). O sensor Metbots é um protótipo de baixo custo que permite a aquisição de espectros que abrangem a zona do ultravioleta, visível até o infravermelho próximo. A sua utilização é muito expedita já que tira partido da conectividade, processamento e visualização através do smartphone (Figura 1). Possui uma fonte de luz LED (sensor ativo, Figura 1b) para permitir as medições mesmo durante o período noturno (importante no caso do PHB) e garantir uma fonte de luz estável para a extração de espectros. O sensor utiliza técnicas de fotónica inteligente (*smart photonics*) para permitir a estimação de diversos parâmetros com precisão analítica. O Metbots pode ser incorporado num robô (Figura 1c) para a aquisição e mapeamento dos resultados da fenotipagem digital.

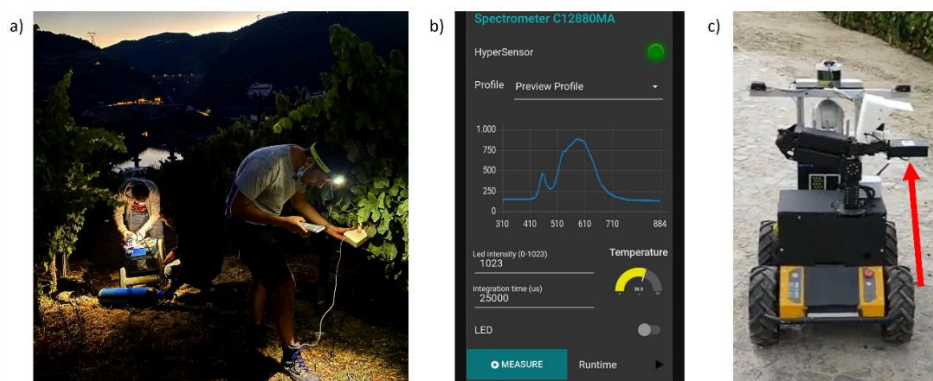


Figure A-1. Utilização do Metbots na medição do potencial hídrico foliar de base nas vinhas do Douro: a) obtenção dos espectros das folhas antes de serem retiradas e medidas na câmara de pressão (Scholander), antes do amanhecer, b) espectro apresentado pelo Metbots e c) O Metbots (indicado com a seta vermelha) acoplado num robô com outros sensores.

Fenotipagem digital

Os sensores de proximidade têm demonstrado grande potencial na fenotipagem digital de parâmetros fisiológicos da vinha. Desse modo, o INESC TEC em parceria com a Faculdade de Ciências da Universidade do Porto (FCUP) tem vindo a desenvolver trabalhos para a validação desses sensores, e assim poder ter uma aderência fisiológica de causa-efeito nos dados espectrais colhidos em vinhas. Os testes desenvolvidos em várias regiões vitivinícolas mostram bons resultados na determinação de diversos parâmetros, tais como: glucose, frutose, ácido málico e tartárico (Martins et al., 2022), carotenos, luteínas, clorofilas e teor de água na folha (Tosin et al., 2021), potencial hídrico foliar de base (Tosin et al., 2022) e doenças do kiwi (Reis-Pereira et al., 2022). A Figura 2 apresenta resultados preliminares obtidos com sensores hiperespectrais de proximidade em diferentes ensaios realizados na região do Douro. A metodologia segue um procedimento genérico de pré-processamento (*scattering correction*), aplicação de técnicas de inteligência artificial (Figura 2a), seleção da técnica de modelação estatística (Figura 2b) ou de *machine learning* (Figura 2c). Os métodos de validação utilizados nesses modelos foram através de *cross-validation* ou por validação externa (*datasets independentes*).

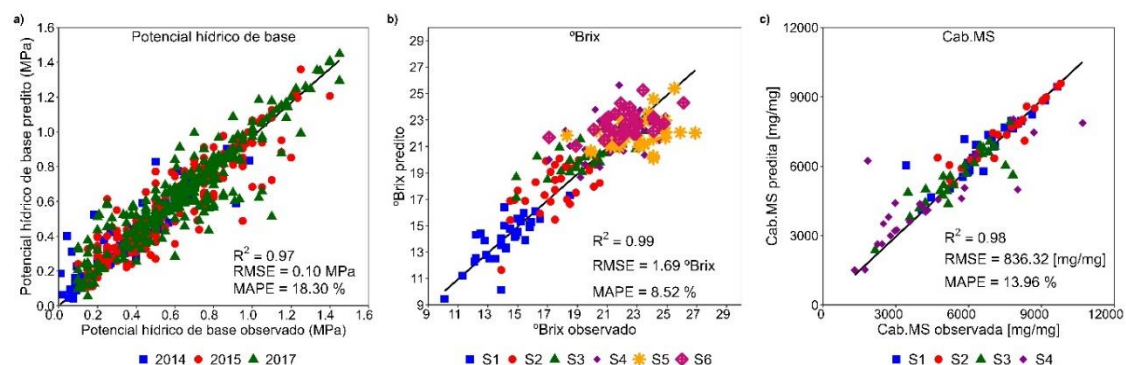


Figure A-2. Relação entre dados observados e estimados com base em sensores hiperespectrais de proximidade e diferentes técnicas de modelação dos seguintes parâmetros biofísicos da vinha na região do Douro. a) PHB utilizando o algoritmo Self-learning artificial intelligence (SL-AI) num dataset com 847 observações recolhidas em 2014, 2015 e 2017; b) modelação com Partial Least Square Regression (PLS) do °Brix em Touriga Nacional (n=192) no ano de 2021 em três diferentes datas (S1: 29/07/2021, S2: 5/08/2021, S3: 12/08/2021, S4: 19/08/2021, S5: 26/08/2021, S6: 02/09/2021); c) modelação das clorofilas ab através do algoritmo Support Vector Machine (SVM) num dataset (n=105) com três cultivares (Touriga Nacional, Touriga Franca e Tinta Roriz) em quatro datas (S1: 03/08/2018, S2: 17/08/2018; S3: 31/08/2018; S4: 12/09/2018). R^2 : coeficiente de determinação, RMSE: Erro quadrático médio, MAPE: Erro médio absoluto previsto.

Alguns destes resultados, ainda que preliminares, têm grande potencial para o desenvolvimento de um modelo de VP fisiológica. Por exemplo, a estimativa e mapeamento do PHB permite suportar técnicas de rega de precisão para aumentar a produtividade da água. A estimativa do teor de pigmentos, como por exemplo das clorofilas, serão um indicador de estresses abióticos da vinha que podem ser utilizados

para a melhor a eficiência das práticas culturais. O mapeamento temporal do °Brix, é importante para monitorizar o processo de maturação e suportar estratégias de vindima localizada de acordo com objetivos enológicos previamente definidos.

A montagem destes sensores em plataformas móveis, tais como um robô, permitirá fazer o mapeamento destes parâmetros com elevado débito para suportar a adoção de VP fisiológica.

Perspetivas dos sensores proximais na vinha

Os sensores proximais tem demonstrado um elevado potencial para a fenotipagem digital das características biofísicas da vinha, como demonstrado na Figura 2. O rigor dos resultados obtidos aliados ao baixo custo e grande dimensionalidade do sensor (elevado número de características fenotípicas analisadas simultaneamente) perspetivam a sua utilização generalizada para melhorar a eficiência das operações vitícolas e enológicas numa ótica de VP fisiológica. Uma próxima etapa será o aperfeiçoamento do sensor hiperespectral num braço robótico para a fenotipagem digital de folhas e bagos de videiras em tempo real. No futuro, parâmetros bioquímicos e metabólicos, relacionados com a fisiologia da videira (ex compostos fenólicos) e com doenças da videira serão testados no âmbito de projetos em curso na FCUP e INESC TEC com particular destaque para o OMICBOTS *High-Throughput Integrative Omic-Robots Platform for a Next Generation Physiology-based Precision Viticulture* (<https://www.fc.up.pt/omicbots/Home.html>).

Agradecimentos

Renan Tosin agradece à “Fundação para a Ciência e a Tecnologia (FCT)”, Portugal pela bolsa de doutoramento (SFRH/BD/145182/2019). Rui Martins agradece à “Fundação para a Ciência e Tecnologia (FCT)” pelo contrato de investigação (CEEIND/017801/2018). Os autores agradecem também às empresas vitivinícolas do Douro: Real Companhia Velha, Symington Family Estates (e ao seu Gestor de Viticultura de I&D Fernando Alves) e Sogrape (e ao seu Gestor de I&D António Graça) e também à Associação para o Desenvolvimento da Viticultura Duriense (ADVID) e estudante Valéria Testa (FCUP) pelas facilidades disponibilizadas para o trabalho de campo. Este trabalho foi parcialmente financiado por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia no âmbito dos projetos POCI-01-0145-FEDER-031124 e PTDC/ASP-HOR/1338/2021