UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E METEOROLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO

DINIZ CARVALHO DE ARRUDA

IMPLEMENTAÇÃO DE MODELOS DE ESPECTROSCOPIA HIPERESPECTRAL E NANOSSATÉLITE NA IDENTIFICAÇÃO DE CULTIVARES DE VITIS VINIFERA E SUAS VARIAÇÕES REGIONAIS.

PORTO ALEGRE

2022

DINIZ CARVALHO DE ARRUDA

IMPLEMENTAÇÃO DE MODELOS DE ESPECTROSCOPIA HIPERESPECTRAL E NANOSSATÉLITE NA IDENTIFICAÇÃO DE CULTIVARES DE VITIS VINIFERA E SUAS VARIAÇÕES REGIONAIS.

Tese de doutorado apresentada ao Programa de Pós-Graduação em Sensoriamento Remoto como requisito parcial para a obtenção do título de doutor em Sensoriamento Remoto e Geoprocessamento.

Orientador: Prof. Dr. Jorge Ricardo Ducati

Coorientador: Dra. Rosemary Hoff

PORTO ALEGRE

2022

CIP - Catalogação na Publicação

Carvalho de Arruda, Diniz
IMPLEMENTAÇÃO DE MODELOS DE ESPECTROSCOPIA HIPERESPECTRAL E NANOSSATÉLITE NA IDENTIFICAÇÃO DE CULTIVARES DE VITIS VINIFERA E SUAS VARIAÇÕES REGIONAIS. / Diniz Carvalho de Arruda. -- 2022.
103 f.
Orientador: Jorge Ricardo Ducati.

Coorientadora: Rosemary Hoff.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre, BR-RS, 2022.

1. Vinhedos . 2. Radiometria. 3. Hiperespectral. 4. Aprendizagem de máquina. I. Ducati, Jorge Ricardo, orient. II. Hoff, Rosemary, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

INSERIR FOLHA DE APROVAÇÃO

RESUMO

O Brasil é destaque na produção mundial de uvas e demonstra uma constante evolução ao longo de sua histórica, desde 1980, com o Estado do Rio Grande do Sul, no topo da lista de produtores. Diversas regiões produtoras de uvas e vinhos no Brasil tem organizado suas atividades no sentido de se tornarem reconhecidas como "Indicações de Procedência" (IP), dando tipicidade e caráter regional aos seus produtos. Esta caracterização requer descrições dos impactos das condições ambientais e do trabalho humano. A utilização de dados adquiridos por sensoriamento remoto, incluindo dados proximais hiperespectrais e de satélites, permitem classificar e caracterizar as variedades de uvas e suas respectivas unidades produtoras de diversas localidades, sob condições climáticas e antrópicas diferenciadas. Esta tese tem como principal objetivo desenvolver uma metodologia para aquisição de dados, treinamento de modelos de hiperespectrais por sensor proximal e imagens via nanossatélite. A área de estudo é composta por oito vinhedos comerciais localizados no Rio Grande do Sul, Brasil. Na primeira fase deste estudo, a unidade de análise foi a folha isolada da videira em diferentes regiões. Posteriormente foi realizado o levantamento dos parâmetros de clorofila, Teor de Sólidos Totais (TST) ou °Brix da uva, espectros de reflectância hiperespectral e imagens de nanossatélite em parcelas de Cabernet Sauvignon em uma vinícola da Serra Gaúcha. Algoritmos de aprendizado de máquina foram aplicados na discriminação de vinhedos por região e por variedade, e na estimativa dos parâmetros clorofila e ºBrix da uva. Os modelos *Light Gradient Booster Machine* (LGBM) e *Random Forest* (RF) obtiveram as melhores acurácias na discriminação espectral em regiões do ultravioleta (UV) e visível (VIS). As estimativas apresentaram elevados R^2 com o modelo de regressão RF. O índice de Gini teve maiores valores para comprimentos de onda no UV/VIS/NIR e o índice de vegetação *Plant Senescence Reflectance Index* (PSRI) teve melhor desempenho para predição dos parâmetros de clorofila, e o *Triangular Greenness Index* (TGI)/*Normalized Difference Vegetation Inde*x (NDVI) para o ºBrix da uva, utilizando como dados a reflectância hiperespectral e a reflectância de superfície. Desenvolvimentos futuros incluem o levantamento de dados com maior número de planta e variedades, auxiliando a compreender as assinaturas espectrais de cada variedade como subsídio para um melhor manejo da produção.

Palavras-chave: Vinhedos, Radiometria, Hiperespectral, Aprendizagem de Máquina

ABSTRACT

Brazil has had an increasing prominence in the production of grapes in the world and the country's production history since the 80's demonstrates this constant evolution. At the top of the list of producers is the State of Rio Grande do Sul. Several grape and wine producing regions in Brazil have organized their activities in order to become recognized as “Indications of Origin” (IO), giving their products typicality and regional character. This characterization requires descriptions of environmental conditions and the impacts of these conditions and human work. The use of remote sensing data, including proximal hyperspectral and satellite data, allow us to classify and characterize grape varieties and their respective producing units from various locations, under different climatic and anthropic conditions. The main objective of this thesis is to develop a methodology for data acquisition, training of plant spectroscopy models with a hyperspectral proximal sensor and for nanosatellite imaging. The study area consists of eight commercial vineyards found in Rio Grande do Sul, Brazil. In the first phase of this study, the unit of analysis was the leaf isolated from the vine in different regions. Subsequently, a survey of chlorophyll parameters, Total Solids Content (°Bx) of the grape, hyperspectral reflectance spectra and nanosatellite images were conducted in Cabernet Sauvignon plots in a Serra Gaúcha winery. Machine learning algorithms were applied in the discrimination of vineyards by region and by variety, and in the estimation of the chlorophyll and Brix parameters of the grape. The Light Gradient Booster Machine (LGBM) and Random Forest (RF) models obtained the best accuracies in spectral discrimination in the ultraviolet (UV) and visible (VIS) regions. The estimates showed high R^2 with the RF regression model. The Gini index had higher values for UV/VIS/NIR wavelengths, and the Plant Senescence Reflectance Index (PSRI) had better performance for predicting chlorophyll parameters, and the Triangular Greenness Index (TGI)/Normalized Difference Vegetation Index (NDVI) for the degree Brix, using as data the hyperspectral reflectance and the surface reflectance. Future developments include collecting data with a greater number of plants and varieties, helping to understand the spectral signatures of each variety as a subsidy for better production management.

Keywords: Vineyards, Radiometric data, Hyperspectral, Machine Learning

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS8

ABNT	Associação Brasileira de Normas Técnicas
CDA	*Canonical Discriminant Analysis*
CEPSRM	Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
IV	Índice de Vegetação
HVI	*Hyperspectral Vegetation Index*
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Brasileiro de Pesquisas Espaciais
LGBM	*Light Gradient Booster Machine*
NASA	*National Aeronautics and Space Administration*
NDVI	*Normalization Difference Vegetation Index*
PPGSR	Programa de Pós-Graduação em Sensoriamento Remote
RBF	*Radial Basis Function*
RFR	*Random Forest Regression*
SVM	*Support Vector Machine*
TGI	*Triangular Greenness Index or RGB index for chlorophyll sensitivity*
VARI	*Visible Atmospherically Resistant Index*
UFRGS	Universidade Federal do Rio Grande do Sul
KPCA	*Kernel by Principal Component Analysis*

SUMÁRIO

1 CONSIDERAÇÕES INICIAIS

O Brasil assume uma posição de destaque na produção mundial de uvas em constante histórica desde 1980. No topo dessa lista se encontra o estado do Rio Grande do Sul, que mantém esta posição até o último levantamento, mesmo com o avanço da viticultura no nordeste do país. O reconhecimento do mercado interno e externo ocorre graças às condições ambientais e climáticas benéficas ao cultivo da vinha, principalmente às variedades de *Vitis vinifera,* castas de origem européia, que compõem 7% da produção de uvas do país, exclusivas para a produção de vinhos finos (EMBRAPA, 2022).

A importância dessa cultura para o país é apresentada, como exemplo, na delimitação geográfica de regiões vitivinícolas, as chamadas Indicações Geográficas – IGs, regulamentadas e definidas pelo marco legal das IGs no Brasil, sendo separadas em Indicação de Procedência – IP e Denominação de Origem – DO (BRASIL, 1996). A DO tem como pressuposto que o produto, nesse caso o vinho, expressa uma tipicidade atribuída das características da produção da uva influenciada pelos dos fatores edafoclimáticos e antrópicos. Alguns autores atribuem essa relação ao termo *terroir*, que estabelece uma delimitação geográfica entre regionalidade e tipicidade (VAUDOUR; CAREY; GILLIOT, 2010; VAUDOUR, 2003).

No auxílio da valorização e proteção da produção de uvas, os dados adquiridos de forma remota entram com apoio no embasamento de decisões técnicas, permitindo uma visão espaço-temporal do desenvolvimento do vinhedo ao longo do ciclo por meio o vigor vegetativo da folha e geometria do dossel. O sensoriamento remoto na viticultura vem sendo aplicado em diferentes plataformas de aquisição: Orbitais (DUCATI; SARATE; FACHEL, 2014; HELMAN *et al.*, 2018; JOHNSON *et al.*, 2003; KARAKIZI; OIKONOMOU; KARANTZALOS, 2016; KNIPPER *et al.*, 2019; SILVA; DUCATI, 2009); Aéreos (ACEVEDO-OPAZO *et al.*, 2008; HALL; LOUIS; LAMB, 2003; TARANTINO; FIGORITO,

2012; VANEGAS *et al.*, 2018); e proximais(JUNGES *et al.*, 2020; MARTÍN *et al.*, 2007; REYNOLDS, ANDREW G. ; BROWN, 2015; RODRÍGUEZ-PÉREZ *et al.*, 2007).

1.1 Sensoriamento Remoto aplicado à viticultura

Em sincronia com as definições amplamente discutidas na literatura sobre a definição de sensoriamento remoto, podemos citar que na viticultura a detecção remota é um conjunto de técnicas e ferramentas para aquisição de dados espectrais, em diversas plataformas. Por intermédio da interação do fluxo radiante com a superfície da folha e/ou dossel da videira, os sensores remotos captam a energia para quantificar e caracterizar as propriedades biofísicas. As tecnologias remotas compõem uma linha de pesquisa e aplicação na viticultura de precisão.

A padronização das assinaturas espectrais apresentadas na literatura, por imageamento via Satélite ou leituras espectrorradiométricas de campo, é desenvolvida com objetivo de discriminar/separar espectralmente variedades de videiras. Em alguns trabalhos imagens de satélite de média resolução foram utilizados para confeccionar padrões espectrais de parcelas de vinhas, identificando na composição do pixel assinatura espectral misturada com outros usos da parcela, como sombreamento e cobertura entre fileiras(JOHNSON et al., 2003; SILVA; DUCATI, 2009). Em nível de campo, assinaturas hiperespectrais de dosséis vegetativos ou folhas são coletadas para identificar variações sutis em comprimentos de onda ou bandas específicas (GONZATTI BOMBASSARO, 2016; PITHAN *et al.*, 2021; SCHOEDL *et al.*, 2012; THUM *et al.*, 2020). A discriminação de assinaturas espectrais entre videiras também ocorreu com a análise de imagens de alta resolução de espacial, identificou separabilidade por meio das feições espaciais e texturais (KARAKIZI; OIKONOMOU; KARANTZALOS, 2016).

A aplicabilidade monitoramento de vinhedos por imagem é interligada ao interesse de reconhecer áreas susceptíveis a receber novos cultivos e ampliar o rendimento de uma parcela. A relação da área foliar por índices de vegetação e condições hídricas é comprovada com o uso

de imagens de UAV, numa escala de maior detalhamento (POBLETE-ECHEVERRÍA *et al.*, 2017; RODRÍGUEZ-PÉREZ *et al.*, 2007; TIAN *et al.*, 2017).

As características espectrais são usadas para criar zonas homogêneas de produção em parcelas, potencializando as características do vinho a sua tipicidade local. A heterogeneidade espacial de uma parcela é um fator determinante nas alternâncias da qualidade da uva. A variabilidade do vigor do dossel e folhas são parâmetros monitorados remotamente para identificar características do desenvolvimento da vinha intra-parcela(GONZÁLEZ-FERNÁNDEZ; CATANZARITE TORRES; RODRIGUEZ-PÉREZ, 2010). As alterações na superfície da folha são utilizadas como indicadores de zoneamentos vitícolas, para mapear blocos com diferentes potenciais de qualidade no vinhedo (BONNARDOT *et al.*, 2012; BRAMLEY; HAMILTON, 2004; STREVER, 2003).

1.2 Aplicações Sensoriamento Remoto Proximal Hiperespectral

Há muitos anos fala-se sobre as potencialidades dos sensores hiperespectrais e sua aplicabilidade no monitoramento de cultivos agrícolas. O diferencial dos dados hiperespectrais está na sua característica principal, um extenso número de bandas ou *wavebands* (HENNESSY; CLARKE; LEWIS, 2020), com bandas estreitas quando comparado ao sensor multiespectral. A superioridade na resolução espectral apresenta para o setor agrícola uma expansão no quesito qualidade de informação, e aqui os dados hiperespectrais apresentam um perfil detalhado da resposta do alvo agrícola, folha isolada ou dossel (FORMAGGIO; SANCHES, 2017), uma sensibilidade espectral sobre as características da parte estrutural/morfológica e alcança fases complexas do metabolismo da planta, como reações físico-químicas e composição biológica (CABELLO-PASINI; MACÍAS-CARRANZA, 2011; GITELSON *et al.*, 2002; ORDÓÑEZ *et al.*, 2018; PEÑUELAS; FILELLA; GAMON, 1995; SLATON; HUNT; SMITH, 2001).

O sensoriamento remoto hiperespectral é uma tecnologia altamente precisa na determinação do comportamento espectral da vegetação, discriminando vegetais de diferentes espécies, sejam agrícolas, invasoras ou nativas de um determinado local (ABBASI *et al.*, 2020; ADAM *et al.*, 2012; BASHEER; EL KAFRAWY; MEKAWY, 2019; ZHANG *et al.*, 2014). A potencialidade dos sensores hiperespectrais na detecção da vegetação é agregar alta resolução espacial com a sensibilidade de centenas de bandas espectrais para determinar características intrínsecas das plantas (AKBARZADEH *et al.*, 2018; BERGSTRÄSSER *et al.*, 2015; ECKERT; KNEUBÜHLER, 2004; FERREIRO-ARMÁN *et al.*, 2007; GALVÃO; FORMAGGIO; TISOT, 2005; LACAR; LEWIS; GRIERSON, 2002; TISOT *et al.*, 2007). Os parâmetros de pigmentos fotossintetizantes, textura e estrutura da folha são analisados por métodos não invasivos, substituindo análises laboratoriais (STEELE; GITELSON; RUNDQUIST, 2008; STEELE *et al.*, 2009).

Autores tem utilizado regiões espectrais na região do visível, especificamente nas regiões do verde, vermelho e infravermelho para mensurar o espectro de reflectância hiperespectral oriundos de espécies de videiras (HUNT JR. *et al.*, 2011; RENZULLO; BLANCHFIELD; POWELL, 2006). O infravermelho próximo é sensível pela área foliar, densidade da folha(DOGAN et al., 2018; QIN et al., 2010). Já a região espectral do infravermelho médio apresenta relações com a estimativo no conteúdo de água, celulose e proteínas das folhas (KOKALY, 2001; ORDÓÑEZ *et al.*, 2018; VILLACRÉS *et al.*, 2021).

1.3 Justificativa e Hipótese do Trabalho

O Rio Grande do Sul é um Estado com significativo número de implementações de Indicações Geográficas (IGs) no Brasil, o que no caso da produção de vinhos representa uma notoriedade e reconhecimento mais amplo para suas fronteiras vitícolas. O desenvolvimento de técnicas e ferramentas para auxiliar na identificação e implementação de IGs é necessário, e é grande o potencial da detecção remota para este objetivo, incluindo a implementação de

modelos espectrais, que possa fornecer informações precisas sobre estado de saúde dos vinhedos. A demanda de informações de campo precisa e precoce, não somente na questão espacial, mas também espectral, auxilia o produtor na organização de planos de controle e ações efetivo, com a finalidade de manter e elevar a qualidade de suas uvas.

A delimitação do local de produção da uva é uma forma para reconhecer e identificar a tipicidade dos vinhos dali derivados, e para isto pode-se utilizar dados proximais ou remotos como fontes de informações na busca da identidade de regiões produtoras e suas variedades. O sensoriamento remoto hiperespectral é uma ferramenta importante para discriminação de variedades de uvas das espécies de *Vitis vinifera* no local onde são cultivadas, a partir do espectro de reflectância hiperespectral da folha isolada. Para isto, pode-se vincular dados de campo com sensores orbitais para monitoramento diário da evolução do cultivo, visando uma calibração das condições da vinha frente as variações espectrais, advindas das características espaciais e temporais do vinhedo ao longa da safra.

Nesta perspectiva os modelos de aprendizagem de máquina surgem como apoio às análises dos espectros de reflectância hiperespectral e para as imagens multiespectrais, na criação de modelos de redução dimensional, classificação de variedades/região e a predição dos parâmetros de clorofila e grau ºBrix da uva. Este trabalho tem como hipótese de que a espacialização de vinhedos e a discriminação de variedades, como pré-condição à caracterização de IGs, são possíveis, baseadas na coleta e análise de dados hiper/multiespectrais, permitindo classificar e caracterizar de forma não-invasiva a interação das condições ambientais e características das variedades, em um estudo aplicado em vinícolas localizadas no Rio Grande do Sul, sob condições climáticas e antrópicas diferenciadas. A unidade de análise em nível de folha representa um fator inovador no campo, já que as características das vinhas não são tão variantes em parcelas pequenas, como é o caso de parcelas localizadas no RS.

1.4 Objetivo Geral

O trabalho tem como seu principal objetivo desenvolver metodologias para, a partir de dados orbitais e proximais, classificar vinhedos, derivar pigmentos fotossintéticos e o ºBrix da uva, em diferentes ambientes de cultivo.

1.5 Objetivos específicos

- Aplicar técnicas de aprendizagem de máquinas na redução da dimensionalidade em espectros de reflectância, identificando variáveis relevantes nas análises, como índices e comprimentos de onda específicos.
- Aplicar técnicas de aprendizagem de máquina para discriminar regiões produtoras e os tipos de variedades de videiras por meio do sensoriamento remoto hiperespectral.
- Verificar a relação dos parâmetros de clorofila (*Chl a*, *Chl b*, *Chl (a+b)*, *Chl*(*a/b)*) com espectros de reflectância em nível de folha em diferentes parcelas de Cabernet Sauvignon.
- Estudar a relação do grau Brix da uva, com dados de sensores remotos e proximais.

2 MATERIAL E MÉTODOS

2.1 Área de Estudo

Oito vinícolas localizadas no Rio Grande do Sul (Figura 1) foram selecionadas, e encontram-se distribuídos em um território de cerca de 500 km de largura, em diferentes tipos de rochas. As vinícolas são: a) Fazenda Almadén (W1) em Santana do Livramento; b) Adega Boscato em Nova Pádua, com dois vinhedos (W2 e W3, separados por dois quilômetros); c) Fazenda Chandon (W4) na Encruzilhada do Sul; d) Fazenda Luiz Argenta (W5) em Flores da Cunha; e) Adega Miolo em Bento Gonçalves (W6); f) Fazenda Miolo Seival (W7) em Candiota: g) Terra Sul (W8). Analisamos vinhas e suas respectivas folhas, das seguintes variedades de uvas: Cabernet Sauvignon (V1), Chardonnay (V2), Merlot (V3), Petit Verdot (V4), Pinot Grigio

(V5), Pinot Noir (V6), Riesling Italic (V7), Sauvignon Blanc (V8), Syrah (V9), Tannat (V10), Tempranillo (V11), and Viognier (V12).

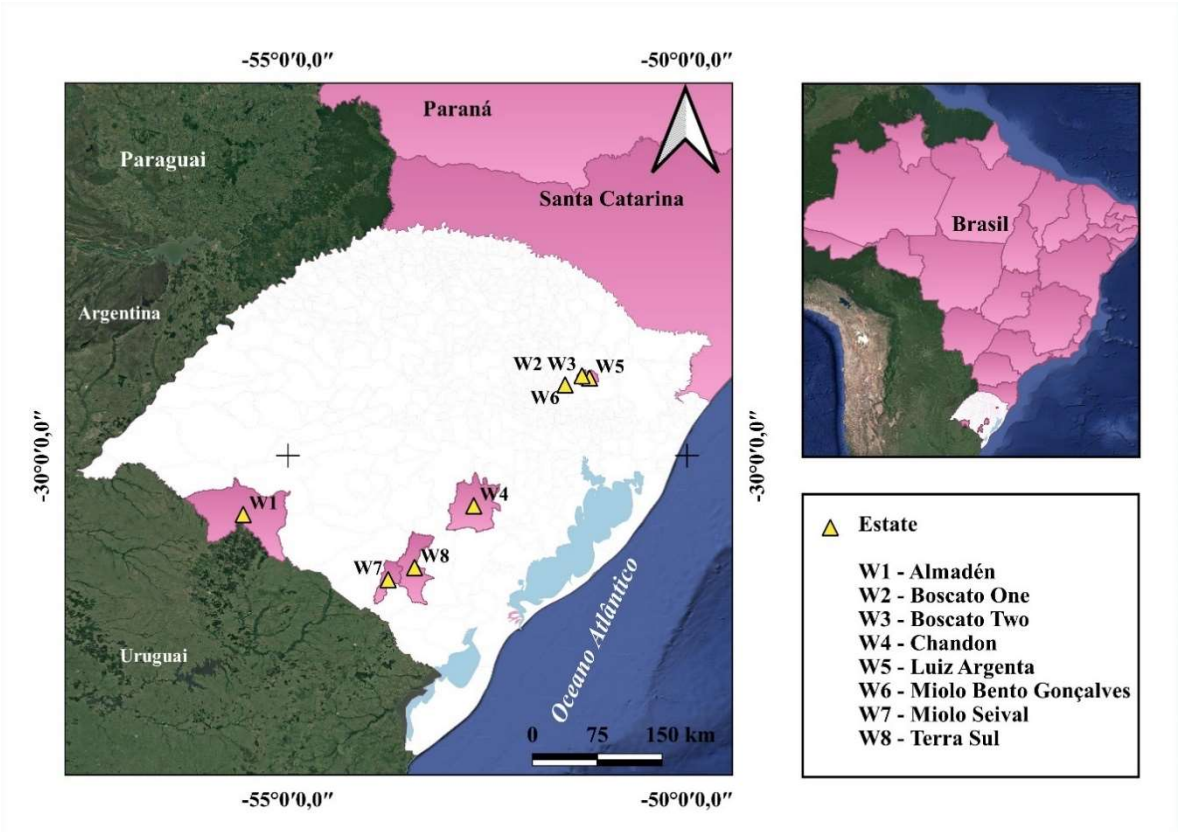

Figura 1. Mapa de Localização das vinícolas escolhidas como área de estudo neste trabalho.

2.2 Leituras de campo

Os trabalhos de campo foram realizados no período do verão do sul do Brasil, compreendendo o estágio fenológico da videira com maior densidade de folhas no dossel e vigor vegetativo. Optou-se por coletar os dados em dias ensolarados sem cobertura de nuvens, entre os horários das 10h até as 14 h, contemplando um padrão de exposição solar da videira. As dificuldades de locomoção entre parcelas e fileiras, foram solucionadas com a seleção das plantas centrais de cada parcela e as folhas do terço média da planta. As leituras espectrorradiométricas foram realizadas com o sensor proximal Field Spec 3 ASD®, onde foram acopladas ao sensor as folhas e medidas realizadas na parte adaxial da folha.

Em particular e no ponto de vista estratégico, as coletas de dados na vinícola Luiz Argenta foram realizadas de forma diferenciada das demais. Para contemplar um estudo de análise de escala local, em nível de parcela, foram coletados os parâmetros de clorofila, na parte adaxial da folha, com o aparelho ClorofiLOG®, um Clorofilômetro que mensura o conteúdo de clorofila na folha. Os parâmetros registrados são: Clorofila a(*Chl* a), Clorofila b(*Chl* b) e a Clorofila Total(*Chl* (a+b), razão das clorofilas (*Chl* (a/b). Logo após, iniciou-se as leituras espectrorradiométricas nas folhas isoladas, em seis parcelas cultivadas com a variedade Cabernet Sauvignon. As leituras em campo serão detalhadas no próximo tópico.

2.3 Tratamento dos dados hiperespectrais

As leituras de campo compreenderam a utilização de dois sensores: O espectrorradiômetro e o clorofilômetro para medidas tomadas na parte adaxial da folha da videira. O primeiro trata da reflectância hiperespectral, com alcance espectral de 350-2500nm, compreendendo 2151 comprimentos de onda. Os espectros foram manipulados em ambiente computacional Python, para modelagem dos dados no formato ASD®, tornando manipuláveis em formatos tabulares convencionais, e inseridos no aplicativo Excel e manipulados com as bibliotecas: *Pandas, Matplot, Numpy, Scikit-learn, seaborn* para análise exploratório de dados. Os espectros de reflectância hiperespectral foram carregadas no programa *ViewSpec*®, e transformadas para um arquivo no formato ASCII e posteriormente integradas as funções de bibliotecas em *python* para filtragem dos dados(*Savitzky-Golay Filter*) para retirar valores ruidosos dos espectros no range(350nm-2500). Uma função de ajuste de patamar entre as regiões do NIR e SWIR1 foi aplicada antes das análises. A queda de patamar foi ajustada para alguns espectros aplicando a função *Jump Correction*. E por último a normalização dos dados, por meio da Função *Normalize*.

O segundo levantamento foi utilizado o ClorofiLOG®, aparelho usado para medir os parâmetros de clorofila da folha, a partir do índice Falker®. As leituras foram copiadas do

aparelho e exportadas no formato de texto e posteriormente conRevivertidas para o formato de tabelas. Os dados foram tabulados em campo de forma organizada em cada ponto de coleta, e digitalizados para uma planilha no formato Excel®. As tomadas de medidas em campo foram organizadas de forma tabular e inseridas num banco de dados relacional (PostgresSQL©)

2.4 Processamento das imagens

As imagens adquiridas comtemplam do dia 16 de dezembro de 2017 até 28 de fevereiro de 2018. Essas datas estão próximas aos dias de campo realizado na vinícola Luiz Argenta. Os produtos baixados do site da empresa Planet's LandScope são nomeados como 3B, com correção atmosférica e geométrica, com valores dos seus pixels em reflectância de superfície. As imagens foram normalizadas pela técnica de Ponto Flutuante Invariante(PFI) desenvolvida em ambiente R, utilizando uma imagem como referência, a primeira data. O produto normalizado, foram coletadas as amostras de treinamento em cada ponto amostral. Uma área de abrangência foi realizada em torno de cada ponto coletada, totalizando uma área de 9m².

2.5 Modelos de Aprendizagem de máquina

O tratamento dos dados concluído, iniciou-se o processo de redução da dimensão dos dados, no qual foi selecionado os comprimentos de ondas e os índices de vegetação hiperespectrais. As técnicas utilizadas foram: *Band ratios*: Uma divisão espectral entre os espectros de reflectância médio para cada variedade e propriedade, com objetivo de identificar variações espectrais sutis na reflectância hiperespectral de cada folha isolada; e a Análise de Componentes Principais por Kernel – ACPK, uma técnica de análise fatorial para reduzir as dimensões em componentes principais, aglutinando o conjunto de informações importantes em poucas variáveis. O diferencial desse modelo e a função adicional, que utiliza o *Kernel* para estimar as componentes em outras dimensões, fugindo da linearização dos dados.

Primeiramente, foi aplicada à classificação dos dados para discriminar propriedades(localizadas sob diferentes regiões) e variedades de uva, espécies típicas para

cultivos de vinho fino. Os modelos de aprendizagem de máquina implementados foram LGBM(*Light Gradient Booster Machine*), RF(*Random Forest*), CDA(*Canonical Discriminant Analysis)*, SVM(*Support Vector Machine*). A avaliação das diferenças entre as medianas foi realizada pela aplicação do teste não paramétrico *KrusKal-Wallis*, considerando o nível de 5% de significância. Nas análises preditivas dos parâmetros de clorofila foram criados modelos a partir dos algoritmos PLSR(*Partial Least Square Regression*) e RFR*(Random Forest Regression*). E por final, foi desenvolvida uma análise utilizando as imagens do Planet Landscope e proximal, para estimar os dados de ºBrix da uva em diferentes parcelas de Cabernet Sauvignon. Logo mais, é descrito um esquema gráfico(Figura 2) descrevendo de forma resumida as etapas metodológicas da tese.

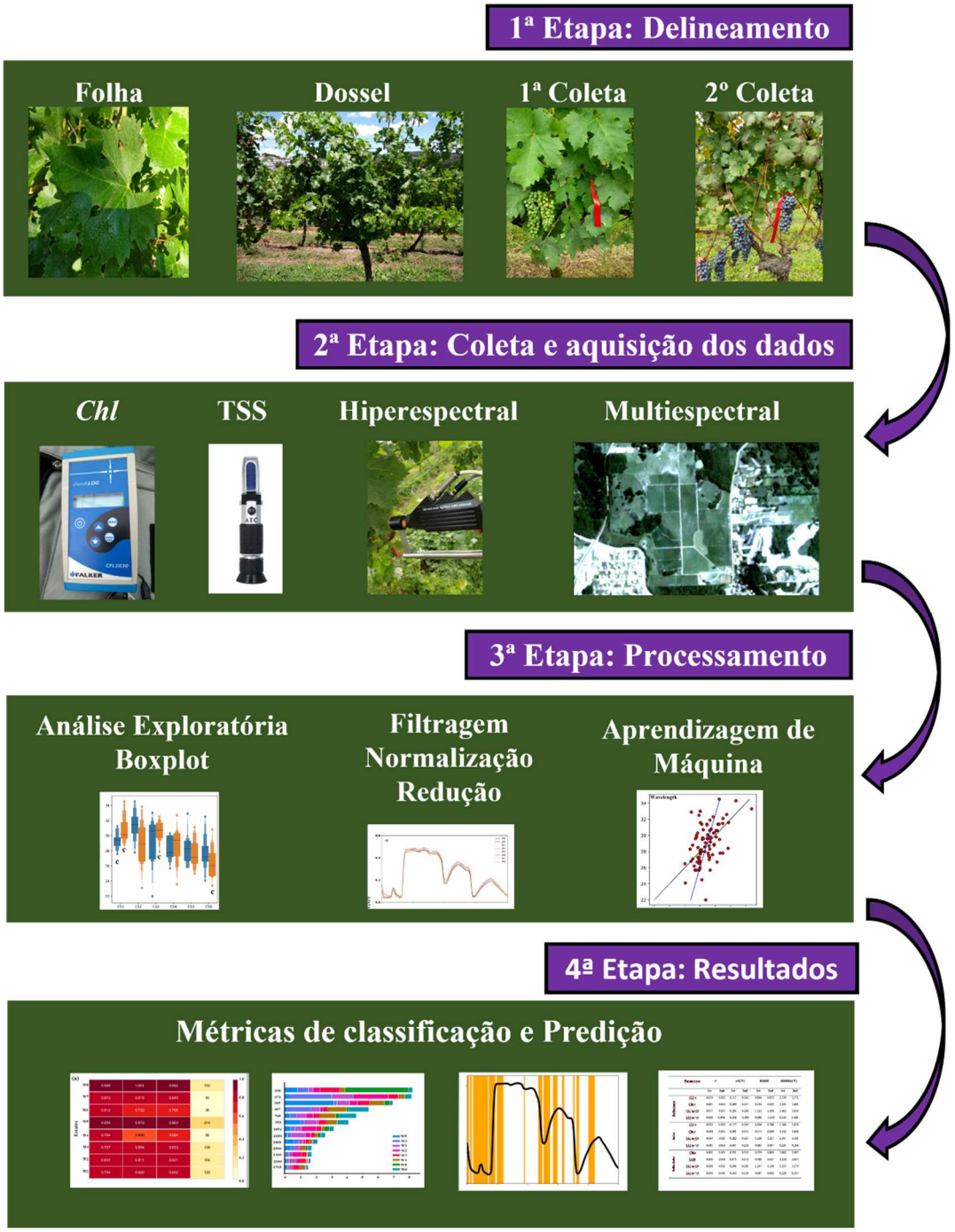

Figura 2. Esquema de rotinas desenvolvidas durante e execução do projeto de doutorado. Cada etapa do projeto é descrita de forma geral, para ilustrar cada processo de desenvolvimento do trabalho, iniciando com o delineamento das coletas até os parâmetros de classificação e predição dos dados.

3 RESULTADOS E DISCUSSÃO

Nesta etapa do trabalho serão apresentados os três artigos desenvolvidos durante à execução da metodologia proposta pela tese, intitulados:

Artigo 1: Aceito para publicação na revista Ciência Rural

Proximal hyperspectral analysis in grape leaves for region and variety identification

Artigo 2: Submetido para revista Ciência Rural

Hyperspectral data analysis for chlorophyll content derivation in Cabernet Sauvignon vines

Artigo 3: Submetido para revista Journal Applied Remote Rensing – JARS

Estimation of degree Brix in grapes by proximal hyperspectral sensing and nanosatellite imagery through the Random Forest Regression algorithm

3.1 ARTIGO 1: Proximal hyperspectral analysis in grape leaves for region and variety identification

Diniz Carvalho de Arruda[1] Jorge Ricardo Ducati[1] Adriane Brill Thum[2]

Tássia Fraga Belloli[1] Rosemary Hoff[3]

ABSTRACT

Reflectance measurements of plants of the same species can produce sets of data with differences between spectra, due to factors that can be external to the plant, like the environment where the plant grows, and to internal factors, for measurements of different varieties. This paper reports results of the analysis of radiometric measurements performed on leaves of vines of several grape varieties and on several sites. The objective of the research was, after the application of techniques of dimensionality reduction for the definition of the most relevant wavelengths, to evaluate four machine learning models applied to the observational sample aiming to discriminate classes of region and variety in vineyards. The tested machine learning classification models were Canonical Discrimination Analysis (CDA), Light Gradient Boosting Machine (LGBM), Random Forest (RF), and Support Vector Machine (SVM). From the results, we report that the LGBM model obtained better accuracy in spectral discrimination by region, with a value the 0.93, followed by the RF model. Regarding the discrimination between grape varieties, these two models also achieved better results, with accuracies of 0.88 and 0.89. The wavelengths more relevant for discrimination were at ultraviolet, followed by those at blue and green spectral regions. This work points toward the importance of defining the wavelengths more relevant to the characterization of the reflectance spectra of leaves of grape varieties and reveal the effective capability of discriminating vineyards by their region or grape variety, using machine learning models.

Keywords: vineyards, hyperspectral, spectroradiometer, machine learning.

RESUMO

Medições de refletância de plantas da mesma espécie podem produzir conjuntos de dados com diferenças entre os espectros, devido a fatores que podem ser externos à planta, como o ambiente onde a planta cresce, e fatores internos, para medições com diferentes variedades de plantas. Este artigo reporta resultados da análise de medições por espectroradiometia efetuadas em folhas de vinhas de diversas variedades e em diversas regiões. O objetivo desta pesquisa foi, após a aplicação de técnicas de redução de dimensionalidade para a definição dos comprimentos de onda mais relevantes, avaliar quatro modelos de aprendizado de máquina aplicados à amostra observacional visando discriminar classes de região e variedade. Os modelos de classificação de aprendizado de máquina testados foram Canonical Discrimination Analysis (CDA), Light Gradient Boosting Machine (LGBM), Random Forest (RF) e Support Vector Machine (SVM). A partir dos resultados, relatamos que o modelo LGBM obteve melhor acurácia na discriminação espectral por região, com valor de 0,93, seguido pelo modelo RF. Relativamente à discriminação entre castas, estes dois modelos também obtiveram melhores resultados, com acurácias de 0,88 e 0,89. Os comprimentos de onda mais importantes para as discriminações procuradas estiveram na região do ultravioleta, seguidos do azul e do verde. Este trabalho aponta para a importância de detectar os comprimentos de onda mais relevantes para a caracterização dos espectros de reflectância das folhas de diferentes variedades de vinhas, e revela a capacidade efetiva de discriminar vinhedos por suas regiões ou variedades, usando modelos de aprendizado de máquina.

Palavras-chave: Vinhedos, hiperespectral, aprendizagem de máquina

3.1.1 INTRODUCTION

The spectral response of vegetation expressed by its reflectance has been known to be a way to characterize different vegetal species, with applications in surveys and monitoring of forests, crops and other land uses (ZHANG, C. et al., 2014; MIRZAEI et al., 2019). Several

studies have applied techniques of remote sensing for data acquisition, including satellite or aerial imagery and/or field or laboratory spectroradiometer. In the first cases, the spectral resolution, in general, tends to be moderate, and only the main spectral features are acquired; even with this limitation, classifications with significant accuracies have been accomplished in studies on vineyards (KARAKIZI et al., 2016; MOGHIMI et al., 2020; SILVA & DUCATI, 2009) using conventional classification algorithms. In the latter cases, using a spectroradiometer extremely high spectral resolution can be attained, showing minute details of a spectrum, and allowing to detect subtle spectral features of vine leaves; these features express degrees or states of pigmentation, cell structure, and water content which, besides depending on intrinsic biological descriptors, can be influenced by environmental and geographical factors (CEROVIC et al., 2012; SMIT et al., 2016; THUM et al., 2020).

From this perspective, spectral data is valuable in studies focused on vine development in geographical contexts, since the high density of information carried by a high-resolution spectrum allows searching for differentiation between cultivars and from external influences caused by climate, soil, management, or other effects. Results from such studies are helpful to the characterization of viticultural regions aiming to distinguish themselves from other regions, contributing to the formation of a set of descriptors necessary to the attribution of a label of typicity of which AOC (Appellation d'Origine Controlée), IGT (Indicazione Geografica Tipica) or AVA (American Viticultural Area) are examples. Such characterizations, when coming from data of plant spectroscopy, have been achieved mainly through the use of conventional classification algorithms (SILVA & DUCATI, 2009; KARAKIZI et al., 2016), but few results have been reported of applications of Machine Learning models which, with present computational resources, can outperform already existent classification methods (ANGUITA et al., 2010).

This paper reports the results from spectroradiometric field measurements performed on vineyards located in southern Brazil, where we investigated their potential to discriminate vines by their locations or by variety. Here, the location factor is dominated by environmental constraints (soils, climate), while the variety factor tends to be dominated by biological (genetic characteristics) constraints. Both factors have significant impacts on plant metabolism and development (WHITE, 2009), influencing leaf structure and chemical composition and, therefore, its reflectance spectrum (THUM et al., 2020). Specifically, the objectives of this research were: a) To discriminate vineyards by region and variety from leaf reflectance data; b) To select a technique to reduce the number of wavelengths necessary for the first objective; c) To select, from a selected set of Machine Learning techniques, the ones with the best performances in the classification process.

3.1.2 MATERIALS AND METHODS

3.1.2.1 Study area

As study areas, eight vineyards were selected in Rio Grande do Sul, which is the southernmost state in Brazil. These vineyards are distributed over a territory of about 500 km wide, on terrains of different types of rocks, and belong to the following wineries: a) Almadén Estate (W1) in Santana do Livramento, in the Campanha Gaúcha wine region, with sandstone-based soils from the Guará Formation (WILDNER et al. 2008); b) Boscato Winery in Nova Pádua, with two vineyards (W2 and W3, two kilometers apart) on acidic volcanic rocks (rhyolite, rhyodacite and dacite) of the Palmas Formation (IBGE 2018, ROSSETI et al. 2017); c) Chandon Estate (W4) in Encruzilhada do Sul, on the gneiss of the Arroio dos Ratos Gneissic Complex (WILDNER et al. 2008); d) Luiz Argenta Estate (W5) in Flores da Cunha, over acidic volcanic rocks (rhyolite, rhyodacite and dacite) of the Palmas Formation (IBGE 2018, ROSSETI et al. 2017); e) Miolo Winery in Bento Gonçalves (W6) in the Serra Gaúcha wine region, with soil on acidic volcanic rocks (rhyolite, rhyodacite and dacite) of the Palmas

Formation (IBGE 2018, ROSSETI et al. 2017); f) Miolo Seival Estate (W7) in Candiota, in the wine-growing region of Campanha Gaúcha, whose soils are a transition between sandstone and claystone of the Rio Bonito and Palermo Formations (CAMOZZATO & LOPES, 2012); g) Terra Sul Winery (W8) in Pinheiro Machado, in the Serra do Sudeste wine region, with soils based on granitic rocks from the Pinheiro Machado Granitic-Gneissic Complex (WILDNER et al. 2008). From this description, it can be seen that the studied vineyards are over different soils, with varying amounts of sand, clay and organic matter. The balance of these soil components, meaning the variation in mineral content, play an important role in reflectance spectra, not only on the spectra of soils themselves (DEMATTÊ, 2002), but also on the spectra of vegetation growing on it (THUM et al. 2020), since many elements are important to plant metabolism; for example, CONRADIE (1981), SCHREINER et al (2006) and SCHREINER (2016) reported as elements like phosphorus potassium, calcium and magnesium move along vine tissues. It is known that different soils have different mineral availability to plant metabolism (WHITE, 2009), with an impact on leaf reflectance spectra (THUM et al. 2020). We note for the regions presently under study that iron availability (associated with clay content) changes greatly, possibly leading to significant changes on plant reflectance spectra. As additional information, we briefly discuss the reason of dividing Boscato Estate in two parts (W2 and W3). From a previous investigation of this winery (THUM et al., 2020), it was found that W2 (5.38 hectares) has elevations from 666 to 688m, and W3 (7.93 hectares) has elevations from 747 to 785m; in addition to the fact of W3 is at higher elevations, W3 displays steeper slopes. Furthermore, out of 21 measured agronomical parameters (data not presently shown), only 3 (P, Ca, Zn) had larger variability in W2; W2 is, therefore, much more homogeneous. Finally, measured soil profiles in W2 are deeper across that vineyard, what points for a likely reason of the larger variability of soil traits in W3, since shallower soils in a more rugged terrain would tend to put the surface in closer contact with deeper horizons and the bedrock, these two layers acting as

mineral suppliers. This condition of soil diversity in terrains seating on the same bedrock supplies an opportunity for assessing the limits of classification performances of the set of Machine Learning techniques to be presently tested. We also note that estates W1 and W7 are located at areas covered by the "Campanha Gaúcha" viticultural region; W2, W3 and W5 are in the "Altos Montes" viticultural region; W4 and W8 are at the "Serra do Sudeste" viticultural region; and W6 is at the "Vale dos Vinhedos" viticultural region. The distribution of these locations over the State's territory is shown in Figure 1.

As grape varieties or cultivars we selected twelve of those more commonly found in the chosen regions, which are: Cabernet Sauvignon (V1), Chardonnay (V2), Merlot (V3), Petit Verdot (V4), Pinot Grigio (V5), Pinot Noir (V6), Riesling Italic (V7) (also known as Welschriesling), Sauvignon Blanc (V8), Syrah (V9), Tannat (V10), Tempranillo (V11), and Viognier (V12). These twelve grape varieties are not present in all eight locations; for example, the Chandon Estate only has Pinot Noir, Chardonnay and Riesling Italic, and at Boscato only Cabernet Sauvignon and Merlot were measured. Detailed information on number of measurements is provided in Table 1. The climate in all regions is subtropical with well-defined seasons; however, the Serra Gaúcha region tends to have summers with higher humidity. We visited in total seventy-eight vine parcels.

3.1.2.2 Leaf reflectance acquisition

Field spectroscopic measurements were performed with a Malvern Panalytical Spectral Devices (ASD, Westborough, MA, USA) FieldSpec® 3 spectroradiometer, which has spectral sensitivity between 350nm and 2500nm, using the Leaf Clip sensor. Field trips were performed in December 2018 and January 2019, since these dates correspond to a period in the phenological cycle where grape leaves are already well-developed, in the stage of growth and ripening of berries represented on the BBCH scale in the sub-stages 81 to 83 (LORENZ et.al., 1995).

In each estate, we selected vine parcels with areas of about five hectares. At each parcel we chose rows centrally localized, at each row we selected four plants, and at each plant we measured four fully developed leaves at their adaxial sides. Calibration of the sensor, through optimization and measurement of the white reference plate of the Leaf Clip probe, was conducted before making the spectroradiometric readings. Every spectrum was recorded at one-nanometer intervals, resulting in 2151 reflectance values for the observed spectral domain (350 nm to 2500nm). The final sample had 3006 spectra corresponding to measurements of 1002 leaves (three spectra per leaf); however, the measurements used for the analyses were 2967 in total since 39 spectra were detected as being erroneous for several factors and were excluded.

3.1.2.3 Pre-processing of spectra

To mitigate the noise interference in the spectra, and to smooth the spectral breaks at the sensor's interfaces, we used the Savitzky-Golay filter and slice correction. The library packages used were SciPy, signal Filter, and Coefficients (VIRTANEN et al., 2020). Since high-resolution spectra tend to carry redundant information over neighboring wavelengths, a feature that tends to increase processing time of classification tasks with no sizable gains, the next step was to decrease the number of wavelengths by means of two spectral reduction techniques applied to the database, which were: Spectrum Ratio (SR) and Kernel Principal Component Analysis (KPCA).

3.1.2.4 Spectrum Ratio (SR)

The SR technique was applied after a normalization procedure was performed on each original spectrum. Since in each acquisition the sensor can receive a particular influx of energy, recorded levels of reflectance can vary from one spectrum to another; that is, each spectrum comes from the acquisition of a certain amount of energy across the observed wavelength domain, implying in a specific area under the spectral curve. The SR technique consists in the direct comparison of two spectra at the same scale, and so, original spectra were transformed

through a normalization procedure described elsewhere (PITHAN et al., 2021); we note that normalization is an operation that does not change the shape of any spectrum.

The Estates group had eight vineyards, so comparisons between them, by pairs, allowed twenty-eight combinations; for each estate, a mean spectrum was derived from all measurements, and this spectrum was divided by the mean spectrum of each other estate, an operation that, applied to all eight vineyards, resulted in twenty-eight "spectrum-ratios." The same procedure was followed for the Varieties group where, for twelve varieties, we obtained sixty-six possible "spectrum ratios". A typical "spectrum-ratio" has values around unity for all wavelengths, except at those wavelengths were spectral differences between classes (in Estates or in Varieties) exist. In this sense, the technique reveals where differences between classes exist, knowledge to be used in classification tasks.

The spectra were subjected to non-parametric correlations tests for the whole spectral domain. First, a correlational test, the Spearman rank correlation model, was used to evaluate collinearity between the 2151 wavelengths. The coefficient of determination (R^2) was used to adjust the correlations for each wavelength. Wavelengths having statistical significance expressed by a p-value < 0.05 were selected. Additionally, and to address the level of statistical significance, the Kruskal-Wallis H test was used to assess the real differences between the sample groups. Levels of statistical significance, α (0.05), were determined to verify the difference in statistical distributions of the sub-groups internal to each main group (Estates and Varieties).

3.1.2.5 Kernel Principal Component Analysis (KPCA)

KPCA, the second spectral dimension reduction technique, is a technique for transforming original data into components of uncorrelated variables, using Principal Component Analysis with extension Kernel in dimensionality reduction to create reliable

compositions, since the determination of decision limits between classes is performed in a non-linear way (Fauvel et al., 2009).

3.1.2.6 Hyperspectral Classification

The classification of reflectance spectra was performed from both input techniques, KPCA and SR. Four Machine Learning (ML) algorithms were used in processes, developed in Python language using the Scikit-Learn package and using the libraries Pandas and NumPy for the preparation of matrix and tables. The four ML algorithms selected for the spectral classification process were: a) Canonical Discriminant Analysis (CDA), which is a multivariate analysis algorithm with a procedure for grouping individuals from a previously defined group into exclusive classes of a group of independent variables (LARK, 1995); b) Random Forest (RF), a model tolerant of noisy data which evaluates correlations between variables using a random vector. The RF performance is high in setting spectral reflectance measurements, because of its low sensitivity to outliers (FLETCHER & REDDY, 2016; HONG et al., 2019); c) Support Vector Machine (SVM), a classifier that discriminates using separation hyper planes with support vectors, limiting the division area between the classes (MA & GUO, 2014); and d) Light Gradient Boosting Machine (LGBM), a gradient structure that uses learning algorithms on trees that grow vertically (FAN et al., 2019).

The training samples were selected at random from a data set with 70% (n = 2077) of reflectance spectra, with the remaining 30% (n = 890) being reserved for testing and validation of ML models. The quality of the validation procedure was evaluated by comparing some commonly used indicators of the performance of ML algorithms, such as Classification Accuracy, Area Under the ROC Curve (AUC), F1 Score, and Kappa, besides other parameters for validation metrics as Precision, Recall and Support. Finally, the wavelengths more relevant for the classifications were revealed through calculation of the Average Impact Magnitude parameter,

using values from the SHAP library which allow identification of the more notable features to the model, thus explaining the output of the machine learning model being studied.

3.1.3 RESULTS AND DISCUSSION

Average spectra for each Estate and each Variety classes are provided in Figure 2. As expected, all spectra display the usual features typical of healthy vegetation, with subtle differences between classes which will be further discussed in what follows.

Results from the correlational Spearman test by coefficients are shown in Figure 3, where in Figures 3a and 3b R^2 values are presented. Values of R^2 as high as 0.6 were observed for the spectral ranges corresponding to the UV (350 to 399nm), NIR (780nm), and SWIR (1100 to 2300nm) for both groups. In the figures, areas next to the main diagonal have strong associations between their wavelengths, while coefficients with lower R^2 values, the darkest colors, indicate the low collinearity between wavelengths. Figures 3c and 3d show the *p*-values, where it can be seen that the wavelengths located at the main diagonal or nearby present determination coefficients above 0.9 and p-values < 0.05, indicating statistical significance. After a correlational analysis has identified the spectral regions with low correlation (*p*-value < 0.05), fourteen wavelengths were selected as indicators of the most conspicuous spectral differences between the studied classes as revealed by the SR technique. These wavelengths were: 350nm; 358nm; 365nm; 467nm; 574nm; 705nm; 1350nm; 1410nm; 1420nm; 1723nm; 1850nm; 1894nm; 2306nm; and 2500nm.

Results from the non-parametric Kruskal-Wallis test for the fourteen wavelengths indicated significant differences ($p < 0.05$) at 365nm, 1350nm, 1420nm, 1850nm and 2306 nm at all Estates. The feasibility of spectral separability between classes within the Estates group has been previously reported, leading to the discrimination between vineyards located in different regions, a perception linked to the terroir concept expressing the soil-plant-climate-management relationship (CEMIN & DUCATI, 2011; THUM et al., 2020). In the Varieties

group, the wavelengths 350nm, 358nm, and 574nm are the more suited to variety separation, while at 2500nm little separation is achieved. These results, therefore, suggest that: a) variations either in region or in variety have a significant effect in the ultraviolet reflectance of vines (at 350nm, 358nm, and 365nm); b) concerning chlorophyll, these variations do not have a major effect on the 467nm band, and none at all at the 660nm band; c) a significant effect at near-infrared (NIR) bands was observed for region variation, and here it can be noted that in former studies a group of grape varieties was discriminated by hyperspectral sensors, pointing out the VIS and NIR spectral regions as crucial in the separability of vineyards (KARAKIZI et al., 2016; MIRZAEI et al., 2019); and d) the water absorption bands usually observed in vegetation (at 1450nm, 1950nm, and 2500nm) seem to have little importance on differentiation of vines induced by variation of region or variety.

The models' performance is presented in Table 2. The highest predictive accuracies for classification are those of the LGBM algorithm, with a maximum accuracy range of 0.99. For both the Estate and Variety groups, the best performances were attained by LGBM, followed by RF. For the dimensionality reduction, the best performance came from the SR technique but the KPCA method also yielded satisfactory results. Comparing KPCA and SR performances, the set of wavelengths extracted by SR showed an increase in performance from 0.91 to 0.93 (Estate) and 0.69 to 0.88 (Variety) using the LGBM algorithm and for RF accuracy raised from 0.74 to 0.92 (Estate) and from 0.45 to 0.89 (Variety). The CDA and SVM algorithms did not perform well by KPCA but showed significant improvements in their metrics for discrimination by SR.

The spectral separation between classes internal to the groups (Estates or Varieties) is shown in Figure 4, which displays the AUC values derived from the LGBM algorithm, the one with best performance, for both KPCA and SR. In this Figure it is possible to assess the separability between classes by inspecting the relations between true or false positives; the more

AUC values are near 1, the better the separation. Most AUC values were above 0.90, with the best fits to the discrimination being obtained by the SR method. For example, in Figure 4, using as input data the set generated by the SR method, for the class W6 the AUC value was 0.95, while using KPCA we had AUC = 0.90; at the Varieties Group, for V8 we had AUC = 0.99 from SR and AUC = 0.70 from KPCA. Therefore, significant separability for both groups was achieved using the LGBM model with both reduction methods, with some advantage to SR.

The classification metrics (Figures 5a and 5b) presents the performance of each class through wavelengths extraction by SR. In Figure 5a, the vineyards W4 and W6 obtained the smallest Recall (0.606 and 0.722) and F1-Score (0.684 and 0.765). With respect to separation between W2 and W3, which are 2km apart and on the same bedrock, inspection of Figure 5a reveals that classes W2 and W3 display similarity between True Positive and False Positive values, having AUC values near 1; therefore, these two classes show similar classification accuracies, being nevertheless separable, what can be explained by the fact that, even if being on the same bedrock, they have different soil profiles, with a possible influence on plant development. It can be noted that W2 and W3 belong to the same owner and have the same management, what excludes differentiation due to anthropogenic factors. Still focusing on Figure 5a, it can be seen that estates W1 and W7, both located at the Campanha Gaúcha viticultural region, are fairly separated, indicating non-negligible spectral differences; this fact, added to the one that W7 is on a transition of sandstone to clay, reinforces current perceptions that the presently established limits of this viticultural region are too wide, pointing to the future need of its division in more uniform territorial units. In Figure 5b, the result of classification between varieties indicates for V7 and V8 the smallest Recall (0.667 and 0.444) and F1-Score (0.800 and 0.615). The lowest precision was shown by V3, with a value of 0.647. Estates W5 and W8 and varieties V1, V2, V6, V10, V11, and V12 obtained the best performances, all of

them with values of F1-Score above 0.9. Furthermore, both groups obtained good discrimination accuracy, indicating the feasibility of spectral separability at leaf level.

Finally, the average Impact Magnitude of the wavelengths on the LGBM model using feature extraction by the SR method is shown in Figures 5c and 5d. The ultraviolet wavelengths (358nm, 574nm, and 365nm, in order of importance) presented a greater average impact magnitude for discrimination between Estates. The Variety classes displayed a similar average impact magnitude. The wavelengths in these spectral regions (green, blue, and ultraviolet) are important to detect changes in reflectance due to changes in pigment content (MERZLYAK et al., 1999), carotenoids(GITELSON et al., 2002), and anthocyanins(PROSHKIN et al., 2021) at leaf level.

Two more perceptions have to be noted. The spectral differences between classes, especially those revealed in the fourteen wavelengths described above, are subtle, as reported elsewhere (DELALIEUX et al., 2007; ETTABAA & SALEM, 2018); in fact, taking as reference the usual range of reflectance values (from zero to unity), the conspicuous differences revealed by the spectrum-ratio technique are of the order of 10^{-4} or even smaller. Their detection is due to the extreme signal-to-noise ratio of the measurements taken with the equipment presently employed, leading to the significant detection of faint spectral features. A lengthy discussion of this point can be found at former research reported by our group (PITHAN et al, 2020). Finally, the results presented here do not suggest a capability, from our data and analysis, to separate between red and white grape varieties (classes V1, V3, V4, V6, V9, V10 and V11 are red grapes); however, it was reported by SILVA & DUCATI (2009) that, using ASTER satellite data, these two greater classes can be discriminated. This is intriguing, since the spectral resolution of ASTER images is much coarser. A possible explanation may come from the classification algorithm used on the images, the maximum likelihood, which was not presently used.

From these results, it seems that purely environmental variations (bedrock, climate) are not decisive to differentiation within the Estates group, since, for example, the Estates on volcanic rocks (W2, W3, W5 and W6), all of them with a more humid climate, do not form a separate group. This suggests that more complex processes are involved in the construction of reflectance spectra of vines (or of vegetation in general) confronted to environmental changes.

3.1.4 CONCLUSION

In this research, we investigated the potential of field hyperspectral leaf reflectance measurements to differentiate grape varieties and grape production regions. Our results have demonstrated that such separability is indeed possible, with significant accuracies. Acquiring spectral information about the vines *in situ*, without removal of leaves for laboratory analysis represents a gain both in costs and in logistical preparations. Due to its extreme signal-to-noise ratio, allowing the detection of subtle spectral features, the hyperspectral proximal sensor data presently used was a crucial tool in the detailing of faint leaf traits, making possible to discriminate grapevine varieties and the influence of environmental aspects. In this sense, our results can contribute to the comprehension of terroir issues related to regional variations, as discussed by VAN LEEUWEN & SEGUIN (2006). In fact, focusing on the presently demonstrated capability of spectrally separating regions, even when the bedrock is similar (being the cases of estates W2, W3, W5, and W6, all on volcanic acidic rocks), we saw that the geological similarity was not a confounding factor; these classes were fairly separated, suggesting that additional discriminating factors, like climate, also play a role on plant development leading to specific spectral traits in leaf reflectance.

The wavelength extraction by the SR technique demonstrated advantages over the KPCA method when both were used for classification with the LGBM algorithm. This paper points towards the feasibility of the spectral discrimination of grapevines at leaf level, using a non-

destructive method, for identification of vine varieties and their region, with applications valuable to the producer, allowing building a spectral library of grape wines.

3.1.5 REFERENCES

AKBARZADEH, S. et al. Plant discrimination by Support Vector Machine classifier based on spectral reflectance. **Computers and Electronics in Agriculture**, 1 May. 2018. v. 148, p. 250–258. Available from: <https://doi.org/10.1016/j.compag.2018.03.026>. Accessed: Oct. 10, 2019. doi:10.1016/j.compag.2018.03.026.

ANGUITA, D. et al. Model selection for support vector machines: Advantages and disadvantages of the Machine Learning Theory. ***In*: Proceedings of the International Joint Conference on Neural Networks**. 2010 Available from: <https://ieeexplore.ieee.org/document/5596450/> Accessed: August 2010, Jan 8. doi: 10.1109/IJCNN.2010.5596450.

CEMIN, G.; DUCATI, J. R. Spectral Discrimination of Grape Varieties and a Search for Terroir Effects Using Remote Sensing. **Journal of Wine Research**, Mar. 2011. v. 22, n. 1, p. 57–78. Available from: <https://doi.org/10.1080/09571264.2011.550762>. Accessed: Oct. 15, 2018. doi: 10.1080/09571264.2011.550762.

CEROVIC, Z. G. et al. A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. **Physiologia Plantarum**, 6 Nov. 2012. v. 146, n. 3, p. 251–260. Available from: <https://doi.org/10.1111/j.1399-3054.2012.01639.x >. Accessed: Nov. 15, 2019. doi: 10.1111/j.1399-3054.2012.01639.x.

CAMOZZATO, E; LOPES, R. C. Carta Geológica Hulha Negra, Folha S H.22-Y-C-I. Estado do RS. Escala 1:100.000. 2002. CPRM, Porto Alegre. Available from: <https://rigeo.cprm.gov.br/handle/doc/19253> Accessed: Nov. 15, 2021.CONRADIE, W. J. Seasonal Uptake of Nutrients by Chenin Blanc in Sand Culture: I. Phosphorus, Potassium, Calcium and Magnesium. **South African Journal of Enology and Viticulture** 2: 7–13. 1981.

Available from: <https://doi.org/10.21548/2-1-2403.9176> . Accessed: Nov. 15, 2019. doi: 10.21548/2-1-2403.9176.

DANNER, M. et al. Spectral Sampling with the ASD FIELDSPEC 4 – Theory, Measurement, Problems, Interpretation. **EnMAP Field Guides Technical Report**, 2015. Available from: <https://doi.org/10.2312/enmap.2015.008>. Accessed: Dec. 02, 2021. doi: 10.2312/enmap.2015.008.

DELALIEUX, S., et al. Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications.**European Journal of Agronomy**, 2007(1), 130–143. Available from: <https://doi.org/10.1016/J.EJA.2007.02.005>.Accessed: Oct. 02, 2020. doi: 10.1016/J.EJA.2007.02.005.

DEMATTÊ, J. A. M. Characterization and discrimination of soils by their reflected electromagnetic energy. **Pesquisa Agropecuária Brasileira**, 2022, p,1445–1458. Available from: < https://doi.org/10.1590/S0100-204X2002001000013 >. Accessed: Oct. 02, 2020. doi: 10.1590/S0100-204X2002001000013.

IBGE – Instituto Brasileiro de Geografia e Estatística. Geologia. Available from: <hftp://geoftp.ibge.gov.br/informacoes_ambientais/geologia/levantamento_geologico/vetores/escala_250_mil/>. Accessed: Jan. 02, 2020.

ETTABAA, K.S, SALEM, M. B. Adaptive Progressive Band Selection for Dimensionality Reduction in Hyperspectral Images. **Journal of the Indian Society of Remote Sensing.** *2017. 46:2*, *46*(2), 157–167. Available from: <https://doi.org/10.1007/S12524-017-0691-9> Accessed: Oct. 02, 2020. doi: 10.1007/S12524-017-0691-9.

FAN, J. et al. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data.

Agricultural Water Management, Nov. 2019. v. 225, n. August, p. 105758. Available from: <https://doi.org/10.1155/2009/783194>. Accessed: Jan. 06, 2020. doi: 10.1155/2009/783194.

FAUVEL, M. et al. Kernel Principal Component Analysis for the Classification of Hyperspectral Remote Sensing Data over Urban Areas. **EURASIP Journal on Advances in Signal Processing**, 22 Dec. 2009. v. 2009, n. 1, p. 783194. Available from: <https://doi.org/10.1155/2009/783194>. Accessed: Jan. 06, 2020. doi: 10.1155/2009/783194.

FLETCHER, R. S.; REDDY, K. N. Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. **Computers and Electronics in Agriculture**, 1 Oct. 2016. v. 128, p. 199–206. Available from: <https://doi.org/10.1016/j.compag.2016.09.004>. Accessed: Jan. 06, 2020. doi: 10.1016/j.compag.2016.09.004.

GITELSON, A. et al. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy¶. **Photochemistry and Photobiology**, 2002. v. 75, n. 3, p. 272. Available from: < https://doi.org/10.1562/0031-8655(2002)075<0272:ACCIPL>2.0.CO;2>. Accessed: Jan. 25, 2018. doi: 10.1562/0031-8655(2002)075<0272:ACCIPL>2.0.CO;2.

HONG, Y. et al. Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest. **Science of the Total Environment**, Feb. 2019. v. 651, p. 1969–1982. Available from: <https://doi.org/10.1016/j.scitotenv.2018.09.391>.Accessed: Jan. 15, 2020. doi:10.1016/j.scitotenv.2018.09.391.

KARAKIZI, C.; OIKONOMOU, M.; KARANTZALOS, K. Vineyard detection and vine variety discrimination from very high-resolution satellite data. **Remote Sensing**, 2016. v. 8, n. 3, p. 1–25. Available from: < https://doi.org/10.3390/rs8030235 >. Accessed: Feb. 15, 2020. doi: 10.3390/rs8030235.

LARK, R. M. Components of accuracy of maps with special reference to discriminant analysis

on remote sensor data. **International Journal of Remote Sensing**, 20 May. 1995. v. 16, n. 8, p. 1461–1480. Available from: < https://doi.org/10.1080/01431169508954488 > Accessed: Feb. 15, 2020. doi: 10.1080/01431169508954488.

LORENZ, D. H. et al. Growth Stages of the Grapevine: Phenological growth stages of the grapevine (*Vitis vinifera* L. ssp. vinifera)—Codes and descriptions according to the extended BBCH scale. **Australian Journal of Grape and Wine Research**, 1 Jul. 1995. v. 1, n. 2, p. 100–103. Available from: < https://doi.org/10.1111/j.1755-0238.1995.tb00085.x > Accessed: Jan. 16, 2020. doi: 10.1111/j.1755-0238.1995.tb00085.x.

MA, Y.; GUO, G. **Support Vector Machines Applications**. Cham: Springer International Publishing, 2014. V. 9783319023. Available from: < https://doi.org/10.1007/978-3-319-023007 > . Accessed: Feb. 10, 2019. doi: 10.1007/978-3-319-02300-7.

MCCARTY, D. A. et al. Evaluation of Light Gradient Boosted Machine Learning Technique in Large Scale Land Use and Land Cover Classification. **Environments**, 3 Oct. 2020. v. 7, n. 10, p. 84. Available from: <https://doi.org/10.3390/environments7100084> Accessed: Feb. 02, 2021. doi: 10.3390/environments7100084.

MERZLYAK, M. N. et al. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. **Physiologia Plantarum**, May. 1999. v. 106, n. 1, p. 135–141. Available from: <https://doi.org/10.1034/j.1399-3054.1999.106119.x>. Accessed: Feb. 25, 2019. doi:10.1034/j.1399-3054.1999.106119.x

MIRZAEI, M. et al. Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. **International Journal of Applied Earth Observation and Geoinformation**, 2019. v. 80, n. January, p. 26–37. Available from: <https://doi.org/10.1016/j.jag.2019.04.002>. Accessed: Feb. 25, 2019. doi:10.1016/j.jag.2019.04.002.

MOGHIMI, A. et al. A novel machine learning approach to estimate grapevine leaf nitrogen

concentration using aerial multispectral imagery. **Remote Sensing**, 26 Oct. 2020. v. 12, n. 21, p. 1–21. Available from: <https://doi.org/10.3390/rs12213515 >. Accessed: Nov. 25, 2020. doi: 10.3390/rs12213515.

PITHAN, P. A. et al. Spectral characterization of fungal diseases downy mildew, powdery mildew, black-foot and Petri disease on Vitis vinifera leaves. **International Journal of Remote Sensing**, 3 Aug. 2021. v. 42, n. 15, p. 5680–5697. Available from: <https://doi.org/10.1080/01431161.2021.1929542>. Accessed: Aug. 25, 2021. doi: 10.1080/01431161.2021.1929542

PROSHKIN, Y. A. et al. Assessment of ultraviolet impact on main pigment content in purple basil (Ocimum basilicum L.) by the spectrometric method and hyperspectral images analysis. **Applied Sciences** (Switzerland), 2021. v. 11, n. 19. Available from: <https://doi.org/10.3390/app11198804 >. Accessed: Feb. 25, 2019. doi:10.3390/app11198804

RENZULLO, L. J.; BLANCHFIELD, A. L.; POWELL, K. S. A method of wavelength selection and spectral discrimination of hyperspectral reflectance spectrometry. **IEEE Transactions on Geoscience and Remote Sensing**, 2006. v. 44, n. 7, p. 1986–1994. Available from: <https://doi.org/10.1109/TGRS.2006.870441>. Accessed: Feb. 25, 2019. doi: 10.1109/TGRS.2006.870441.

ROSSETI, L.; et al. Lithostratigraphy and volcanology of the Serra Geral Group, Paraná-Etendeka Igneous Province in Southern Brazil: Towards a formal..., **J. Volcanol. Geotherm. Res**. 2017. Available from: <http://dx.doi.org/10.1016/j.jvolgeores.2017.05.008>. Accessed: Oct. 25, 2019. doi: 10.1016/j.jvolgeores.2017.05.008.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, 1 Jul. 1964. v. 36, n. 8, p. 1627–1639. . Available from: <https://doi.org/10.1016/j.dsp.2004.09.008>. Accessed: Sep. 05, 2019. doi: 10.1109/TGRS.2006.870441 doi:10.1016/j.dsp.2004.09.008.

SILVA, P. R.; DUCATI, J. R. Spectral features of vineyards in south Brazil from ASTER imaging. **International Journal of Remote Sensing**, 23 nov. 2009. v. 30, n. 23, p. 6085–6098. Available from: < https://doi.org/10.1080/01431160902810612> . Accessed: Sep. 05, 2017. doi: g/10.1080/01431160902810612.

SCHREINER, R. P. 2016. Nutrient Uptake and Distribution in Young Pinot Noir Grapevines over Two Seasons. **American Journal of Enology and Viticulture**. 2016. 67: 436–448. Available from: <Nutrient Uptake and Distribution in Young Pinot noir Grapevines over Two Seasons | American Journal of Enology and Viticulture (ajevonline.org) >. Accessed: Sep. 05, 2018. doi:10.5344/ajev.2016.16019.

SCHREINER, R. P., C. F. SCAGEL, and J. BAHAM. 2006. Nutrient Uptake and Distribution in a Mature 'Pinot Noir' Vineyard. **HortScience** 41: 336–345. Available from: <Nutrient Uptake and Distribution in a Mature `Pinot noir' Vineyard in: HortScience Volume 41 Issue 2 (2006) (ashs.org)> Accessed: Nov. 05, 2015. doi:10.21273/HORTSCI.41.2.336.

SMIT, J. L. et al. Vine signal extraction - an application of remote sensing in precision viticulture. **South African Journal of Enology and Viticulture**, 2010. v. 31, n. 2, p. 65–74. Available from: < https://doi.org/10.21548/31-2-1402 >. Accessed: Feb. 25, 2019. doi: 10.21548/31-2-1402.

THUM, A. B. et al. The influence of mineral content on spectral features of vine leaves. **International Journal of Remote Sensing**, 1 Dec. 2020. v. 41, n. 23, p. 9161–9179. Available from: <https://doi.org/10.1080/01431161.2020.1798547> . Accessed: Feb Dec. 5, 2020. doi: 10.1080/01431161.2020.1798547.

UPRETI, D. et al. Bayesian Calibration of the Aquacrop-OS Model for Durum Wheat by Assimilation of Canopy Cover Retrieved from VENµS Satellite Data. **Remote Sensing**, 18 Aug. 2020. v. 12, n. 16, p. 2666. Available from: < https://doi.org/10.3390/rs14061345 >. Accessed: Sep.15, 2020. doi: 10.3390/rs14061345.

VAN LEEUWEN, C.; SEGUIN, G. The concept of terroir in viticulture. 2007. Available from: <http://dx.doi.org/10.1080/09571260600633135>. Accessed: Oct.15, 2015. doi: 10.1080/09571260600633135.

VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, 3 Feb. 2020. v. 17, n. 3, p. 261–272. Available from: <https://doi.org/10.1038/s41592-019-0686-2>. Accessed: Feb. 25, 2019. doi: 10.1038/s41592-019-0686-2.

WHITE, R. E. 2009. Understanding Vineyard Soils. Oxford: Oxford University Press.

WILDNER, W. et al. Mapa Geológico do Estado do Rio Grande do Sul, Escala 1:750.000 [Geological Map of the State of Rio Grande do Sul, Scale 1:750.000]. 2008.Porto Alegre, Brazil: CPRM. Available from: <http://www.cprm.gov.br/publique/Geologia/Geologia-Basica/Cartografia-Geologica-Regional-624.html>. Accessed: Feb. 25, 2019

ZHANG, C. et al. Separating Mangrove Species and Conditions Using Laboratory Hyperspectral Data: A Case Study of a Degraded Mangrove Forest of the Mexican Pacific. **Remote Sensing**, 25 Nov. 2014. v. 6, n. 12, p. 11673–11688. Available from: <https://doi.org/10.3390/rs61211673>. Accessed: Dec. 02, 2021. doi: 10.0.3390/rs61211673.

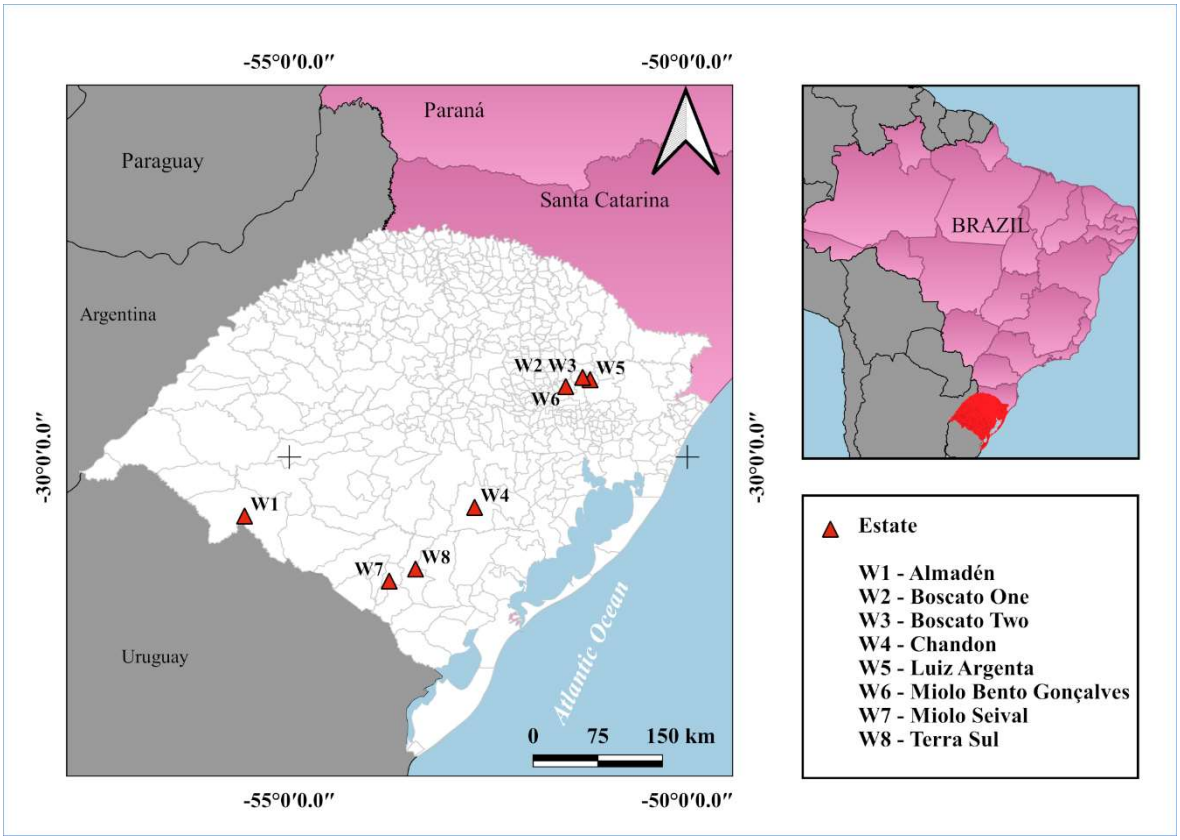

Figura 3. Study area location map.

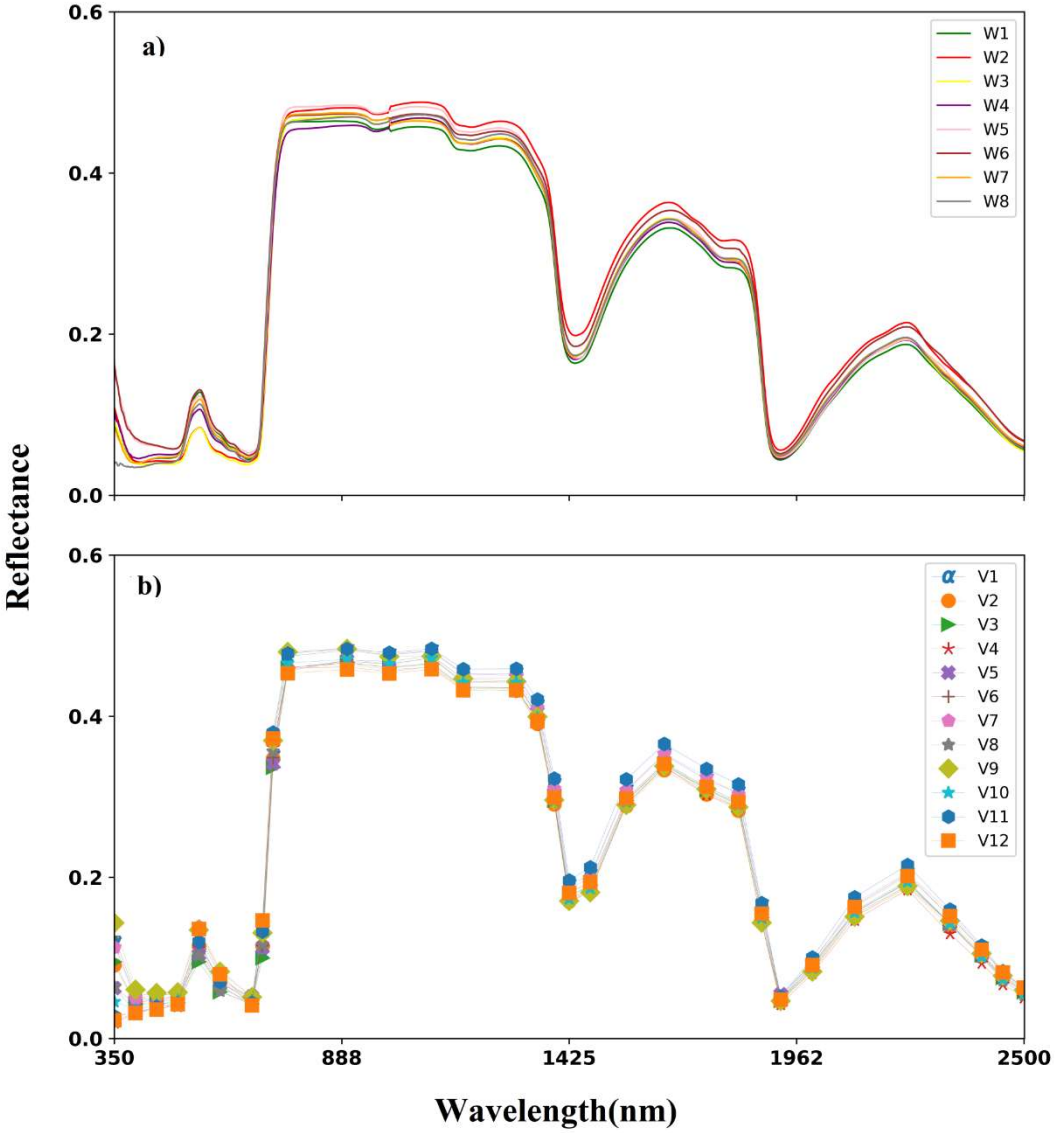

Figura 4. Reflectance spectra of field-measured vines. a) Estates; b) Varieties.

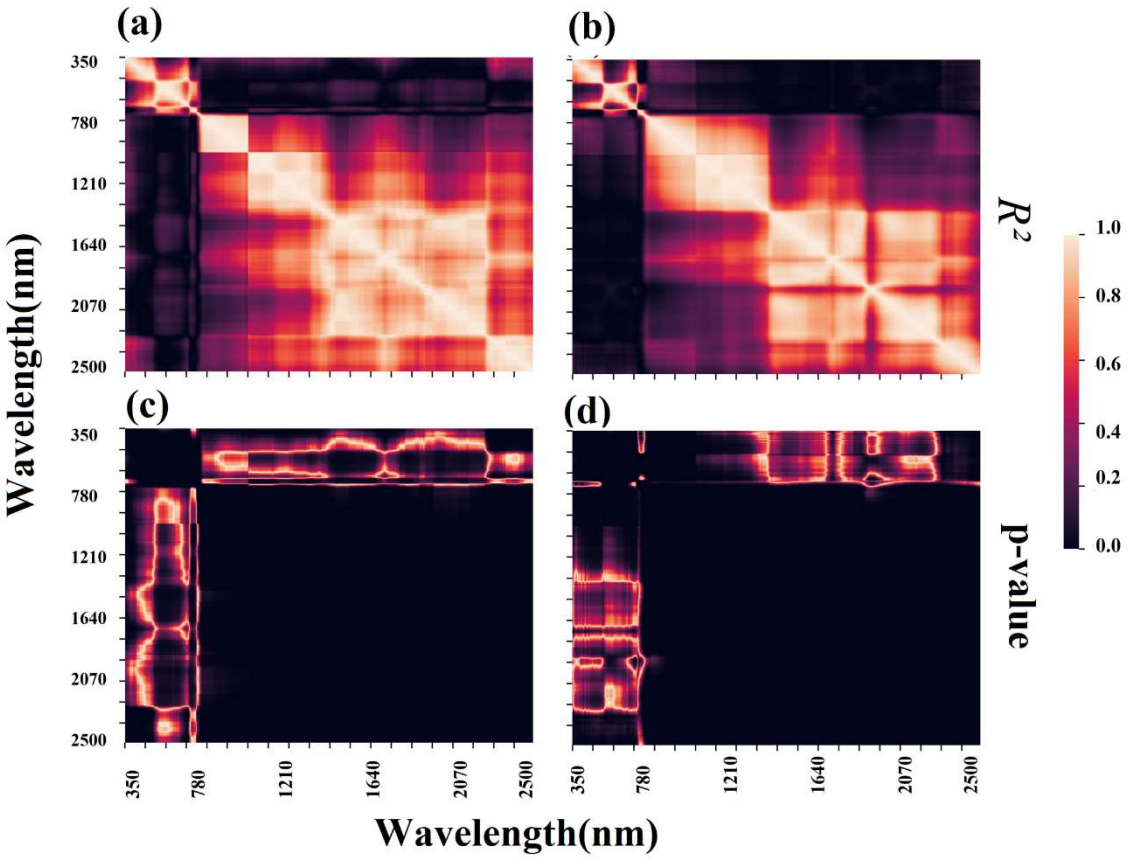

Figura 5. Coefficient of determination R^2 and p-value of the spectrum-ratios between the averages of each class. (a), (c), Estates; (b), (d), Varieties. The shaded scale shows values of the spectral regions with low collinearity.

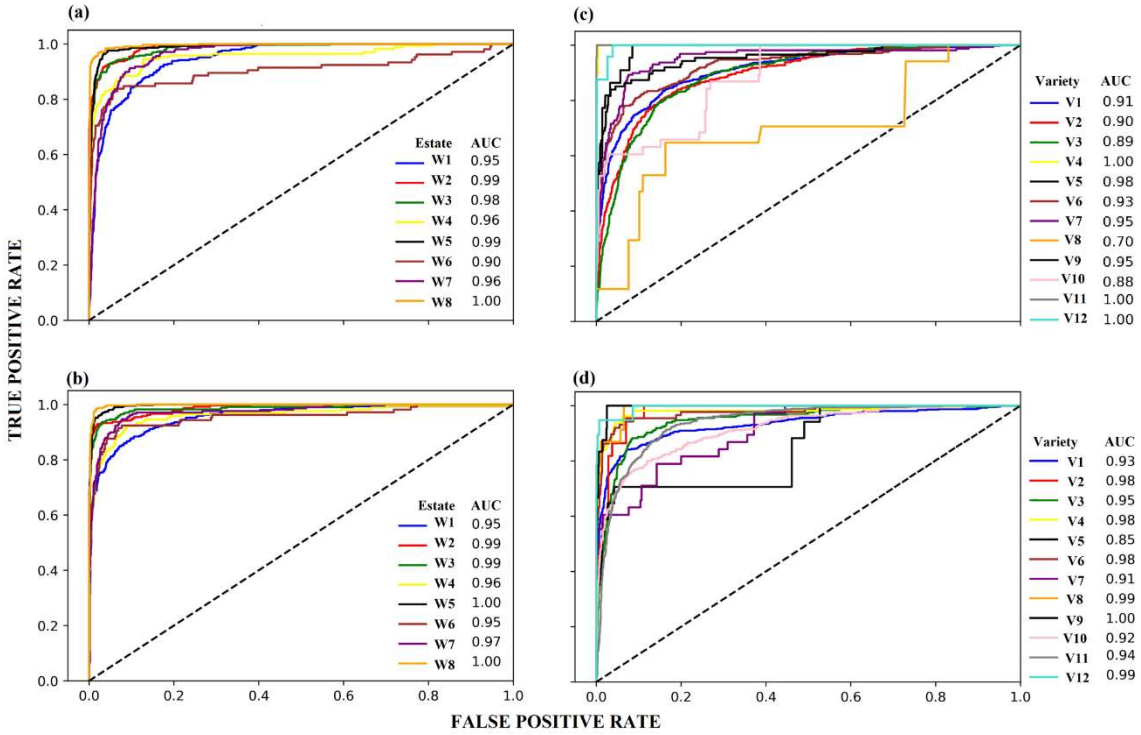

Figura 6. Area Under Curve (AUC) expressing the performance of the LGBM algorithm, using wavelengths selected by the KPCA method ((a) and (c)) and by the Spectrum Ratio method ((b) and (d)). Correspondences between Wn and Vn to their respective estates and varieties are given in Table 1.

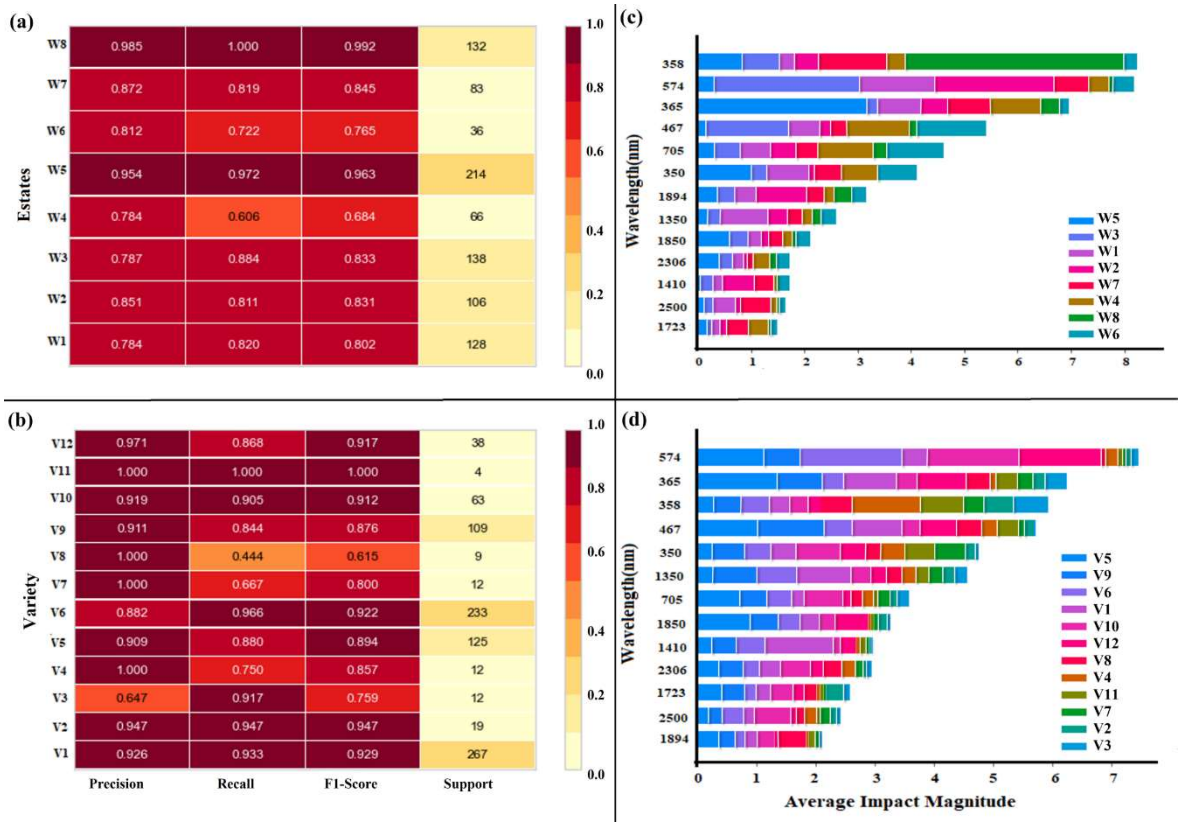

Figura 7. Validation Metrics (a) and (b) and Average Impact Magnitude (c) and (d) to evaluate the performance of the LGBM algorithm, using wavelengths selected by the Spectrum Ratio method. Correspondences between W*n* and V*n* to their respective estates and varieties are given in Table 1.

Tabela 1. Results obtained for spectral discrimination between the Leaf reflectance measured.

Variety/ Estate	W1-Almadén	W2- Boscato One	W3-Boscato Two	W4-Chandon	W5- Luiz Argenta	W6- Miolo BG	W7-Miolo Seival	W8 - Terra Sul
V1- Cabernet Sauvignon	0	237	143	0	223	159	0	105
V2- Chardonnay	132	0	0	93	60	0	100	30
V3-Merlot	144	96	338	0	60	0	97	80
V4-Petit Verdot	0	0	0	0	0	0	0	30
V5- Pinot Grigio	0	0	0	0	0	0	0	30
V6- Pinot Noir	0	0	0	96	121	0	100	46
V7-Riesling Italic	132	0	0	48	61	0	0	0
V8 - Sauvignon Blanc	0	0	0	0	0	0	0	20
V9 - Syrah	0	0	0	0	121	0	0	0
V10 - Tannat	0	0	0	0	0	0	0	56
V11 - Tempranillo	0	0	0	0	0	0	0	30
V12 - Viognier	0	0	0	0	0	0	0	30

Tabela 2. Results obtained for spectral discrimination between the Leaf reflectance measured.

Class	Reduction	Model	Accuracy	AUC	F1	Kappa
Estate	KPCA	LGBM	0.91	0.99	0.91	0.89
		RF	0.74	0.97	0.71	0.69
		CDA	0.50	0.71	0.50	0.42
		SVM	0.48	0.00	0.39	0.38
	SR	LGBM	0.93	0.99	0.93	0.92
		RF	0.92	0.99	0.92	0.90
		CDA	0.92	0.99	0.92	0.91
		SVM	0.61	0.00	0.55	0.54
Variety	KPCA	LGBM	0.69	0.92	0.67	0.60
		RF	0.45	0.82	0.36	0.24
		CDA	0.17	0.53	0.16	0.07
		SVM	0.31	0.00	0.24	0.12
	SR	LGBM	0.88	0.98	0.88	0.86
		RF	0.89	0.98	0.89	0.87
		CDA	0.67	0.91	0.66	0.59
		SVM	0.41	0.00	0.36	0.24

3.2 ARTIGO 2: Hyperspectral data analysis for chlorophyll content derivation in Cabernet Sauvignon vines

Diniz Carvalho de Arruda, Jorge Ricardo Ducati, Rosemary Hoff

ABSTRACT

Quality and yield of a vineyard are related to canopy biomass and leaf vigor, and proximal and remote sensing techniques have been used as alternatives to conventional methods to estimate these parameters, producing imagery and spectral information on leaf and canopy which can be used to those data retrieval. Knowledge on chlorophyll content is crucial to plant health assessments, and this information can be directly obtained through specific sensors. However, chlorophyll indices can also be extracted from reflectance spectra obtained for an ample range of applications. In this perspective, relations between chlorophyll indices obtained by direct measurements and derived from field radiometry were investigated, with the objective of assess the accuracy of predicted chlorophyll content by indirect measurements. The investigation was performed on Cabernet Sauvignon vines on a commercial vineyard, being based on direct chlorophyll surveys, vine leaf spectroradiometry and the derivation of Hyperspectral Vegetation Indices (HVIs), with data acquisition being performed on two stages of the vegetative cycle. Direct chlorophyll data was compared with predicted indices using two machine learning algorithms: Partial Least-Squares Regression (PLSR) and Random Forest Regressor (RFR), using data from reflectance spectra and derived HVIs. Results for estimates indicated that the higher coefficients of determination expressing the correlation between measurements and predictions were obtained for *Chl a* and *Chl(a/b)* modeled by the RFR algorithm, with R^2 values as high as 0.8 and Root Mean Squared Errors as low as 0.093. With respect to HVIs, the Plant Senescence Reflectance Index (PSRI) calculated for the second

acquisition run, corresponding to leaves reaching senescence was the one which produced the highest percentage of prediction explanations, as attested by the Gini index. This study can bring a significant contribution to the development of non-invasive techniques to vine monitoring, contributing to vineyard management by allowing fast, low-cost, real-time interventions by the producer.

Keywords: Hyperspectral, Vineyards, Partial Least-Squares Regression, Random Forest Regressor.

3.2.1 INTRODUCTION

Quality and yield of a vineyard are related to canopy biomass and leaf vigor. Along the vegetative cycle, leaf vigor is an important indicator of plant health status (Bergsträsser et al., 2015; Lacar et al., 2002). In this perspective, remote detection techniques have been applied in studies based on plant spectral patterns, and focused to analyze vegetative development, phenological dynamics, management practices and a diversity of stresses due to biotic and abiotic attacks (Junges et al., 2019; Loggenberg, 2018; Lv et al., 2018; Pithan et al., 2021; Thum et al., 2020; C. Zhang et al., 2013).

Chlorophyll levels are important indicators to the monitoring of nitrogen content in leaves, their determination being conventionally done by laboratory techniques applied on field-collected samples. As the characteristic plant green shades are due to the reflected light after interaction of illuminating radiation with leaf photosynthesizing pigments, chlorophyll amounts can be estimated by non-destructive methods (Fassnacht et al., 2015; Ordóñez et al., 2018; Steele et al., 2008a) like reflectance analysis at hyperspectral resolutions, allowing to map with improved performance the spectral properties of plant leaves at visible and near infrared wavelengths, studying color changes, hemispherical reflectance and subtle variations in leaf tissues (Meneses et al., 2019; Zhao et al., 2014). Chlorophyll levels are also sensitive to

water stress and to soil type (Mitra et al., 2018). Data from plant spectroscopy can be used to predictive models, providing estimations of plant physiological and morphological traits as an alternative to conventional methods. In terms of remote sensing, the use of hyperspectral sensors to the detection of variations in leaf pigmentation is an improvement compared to multispectral data, as more detailed information becomes available.

Ample use has been made of spectroscopy to estimate vine descriptors (Power et al., 2019). However, in hyperspectral data the large number of spectral bands tends to have a negative impact on the metrics expressing the performance of estimating models, a problem which is addressed by the use of dimensionality reduction techniques (Loggenberg et al., 2018; Saheb Ettabaa & ben Salem, 2017; L. Zhang et al., 2019) where machine learning algorithms perform a crucial role. Through machine learning models to data analysis, it is possible to regularize and reduce the number of wavelengths necessary to build structured spectral libraries. Models Partial Least Squared Regression (PLSR) and Random Forest Regression (RFR) are examples of robust algorithms to the characterization and analysis of spectral data, dimensionality reduction and parameters prediction using non-invasive methods (Cheng & Sun, 2017; El-Hendawy et al., 2019; Feng et al., 2017; Kawamura et al., 2017).

Furthermore, the arrival of new methods for data acquisition by *in situ* proximal remote sensing increased the potential to monitoring plant phenological dynamics during the growing cycle. Therefore, the objectives of this study were: a) to analyze the relations of chlorophyll parameters with plant hyperspectral spectroscopy, at leaf level, in commercial vineyards; b) to assess the performance of two machine learning models, PLSR and RFR, to attain the first objective; and c) to reveal the wavelengths more relevant to these tasks. It was expected that this study would bring a significant contribution to the development of non-invasive techniques to vine monitoring, contributing to vineyard management by allowing fast, low-cost, real-time interventions by the producer.

3.2.2 MATERIAL AND METHODS

3.2.2.1 Study area

As study area the Luiz Argenta Winery was chosen, due to its easy access and favorable topography. This estate is in a viticultural region called Vinhos dos Altos Montes (High Hills Wines), a geographical denomination ("Indicação de Procedência") located in north-east of Rio Grande do Sul State in south Brazil. Coordinates are 29º 01' 23.37" S and 51º11'02.23" O, being at a larger wine region called "Serra Gaúcha". The area with vines covers about 48 hectares with several *Vitis vinifera* grape varieties, with focus in the production of quality wines. All measurements were performed during the 2017/2018 season.

The grape variety Cabernet Sauvignon was chosen for the study, and specific parcels were selected considering ease of access, topography, uniformity, and availability of information on soils. Six vine plots were studied, and following the estate use they were called 4a, 4b, 16a, 16b, 19a, and 19b. Vines were planted in trellis driving system, on Paulsen 1103 rootstocks, distance between rows were 2.8m following east-west orientation (plots 4a, 4b, 16a, 16b) or north-south (19a, 19b), and distance between plants were 1.45m. These vineyards had conventional management with treatments on an approximate weekly basis. All plants used for the study were marked prior to the beginning of measurements.

To follow the evolution of spectral behavior of plants, measurements were performed in phases 81 and 83 of the vegetative cycle according to the BBCH scale (LORENZ et al., 1995). The first data acquisition was done on December 16, 2017, during the stage of the phenological cycle known as *véraison*, meaning the phase during which the berries acquire dark pigmentation. The second acquisition run took place on February 27, 2018, during the final ripening and harvest.

3.2.2.2 Data acquisition and treatment

3.2.2.2.1 Chlorophyll and radiometric *in situ* measurements

All measurements in this study took place during a four-hour interval between 10AM and 2PM, to ensure uniformity of plant conditions leading to uniformity in acquired data; reasons for this protocol are presented in more detail in Thum et al. (2020), added to the fact that earlier measurements tend to be done over humid leaves, affecting leaf spectral features. A sample of 24 plants was selected, meaning four plants per parcel, located in the two central rows of each vine plot.

Chlorophyll and radiometric data were acquired in succession, beginning with transmittance measurements at 635nm, 660nm and 880nm using a Falker CFL1030 (Porto Alegre, Brazil) chlorophyll meter (Schlichting et al., 2015), providing chlorophyll a, b and total chlorophyll content mediated by calibration using a white reference. These acquisitions were followed, for the same leaf, by spectroradiometric measurements using a Malvern Panalytical Spectral Devices (ASD, Westborough, MA, USA) FieldSpec® 3 spectroradiometer, which has spectral sensitivity between 350nm and 2500nm (Malvern Panalytical, 2020). A typical spectrum provides reflectance values between 0.0 and 1.0 at intervals of one nanometer, a calibration being made through measurements of a reference plate taken at regular time intervals. All measurements were made with an attached Leaf Clip probe, which carries an internal halogen light source and also an internal reference plate of Spectralon® (Labsphere, Inc., North Sutton, NH, USA). Every 15 minutes calibrations with white reference and optimization were performed, following the protocols described by Pithan et al. (2021) and Thum et al. (2020).

All reflectance field data was recorded in ASD format, and handled in computer environment in Python language, where proprietary codes were created to spectra treatment,

helped by public libraries. Exploratory data analysis used Panda's libraries, and the reflectance steps frequently found at 1000nm and at 1800nm were corrected using the Specdal library with the jump correction function; noisy lines were smoothed by applying the Savitzky-Golay filter.

3.2.2.2.2 Hyperspectral Vegetation Indices

Hyperspectral Vegetation Indices (HVIs), which are defined by the use of carefully selected wavelengths tuned to plant metabolical functions, may avoid the hyperspectral data redundancy problem through the selection of the more informative wavelengths, which are sensitive to plant characteristics such as cellular structure and biochemical and physiological processes. In this study we used 19 HVIs, calculated from the use of specific, discrete wavelengths: Anthocyanin Reflectance Index 1(ARI1) (Gitelson et al., 2001); Anthocyanin Reflectance Index 2 (ARI2) (Gitelson et al., 2001); Cellulose Absorption Index (CAI) (Nagler et al., 2003); Chlorophyll Absorption in Reflectance Index (CARI); Carotenoid Reflectance Index 1 (CRI1) (Gitelson et al., 2002); Carotenoid Reflectance Index 2 (CRI2) (Gitelson et al., 2002); Leaf Water Vegetation Index 2 (LWVI-2) (Galvão et al., 2005); Modified Chlorophyll Absorption in Reflectance Index (MCARI) (Yang et al., 2006) Normalized Difference Nitrogen Index (NDNI) (Serrano et al., 2002); Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974); Normalized Difference Water Index (NDWI) (Gao, 1996)); Photochemical Reflectance Index (PRI) (Peñuelas, Filella, et al., 1995); Pigment Specific Normalized Difference 1 (PSND1) (Blackburn, 1998); Pigment Specific Normalized Difference 2 (PSND2) (Blackburn, 1998); Plant Senescence Reflectance Index (PSRI) (Merzlyak et al., 1999); Plant Senescence Reflectance Index 2 (PSR2) (Merzlyak et al., 1999); Structure Insensitive Pigment Index (SIPI) (Peñuelas, Baret, et al., 1995); Vogelman Red Edge 1 (VOG1) (Vogelmann et al., 2007); and Water Index (WI) (Penuelas et al., 1997).

3.2.2.3 Modeling Process and Prediction Assessment

In this process we generated three data sets and applied the spectral dimension reduction process for each one. The three data sets were composed by: i) 2151 variables (only reflectance bands), ii) 19 HVIs, and iii) 2170 variables (reflectance bands + HVIs). The prediction response variable is represented by the chlorophyll parameters *Chl a*, *Chl b*, *Chl* (*a* + *b*), and *Chl* (*a/b*). We carried out the normalization process for each dataset with the Normalize function, to equalize the descriptors variation scale. The machine learning models used for the prediction analysis were the Partial Least Squares Regression (PLSR) and the Random Forest Regressor (RFR).

The PLSR is technically employed in the sensor calibration on, and spectral analysis associated with infrared spectroscopy and hyperspectral consensus. This is a linear model, easy to fit and presents low computational complexity (Cheng & Sun, 2017) and it is effective for selecting wavelengths employing score coefficients (MIRZAEI et al., 2019). Leaf traits and soil properties were measured non-invasively through PLSR factors, for selection of spectral variables and estimation of physical-chemical parameters (El-Hendawy et al., 2019; Thum et al., 2020; N. Zhang et al., 2017). The problems of high collinearities among wavelengths were solved by maximizing the covariance between measured and predicted (Viscarra Rossel et al., 2006). Moreover, many authors used this model for reducing the dimensionality of wavelengths both at leaf and canopy levels (Abbasi et al., 2020; Mirzaei et al., 2019).

The second algorithm, Random Forest Regressor, is a method based on predictions made in decision trees (estimators) randomly selected (Breiman, 2001). Empirically, it is possible to select the number of estimators, as tree depth, to develop the sample bagging, then extracting an Ouf of Bag (OOB) percentage (Palmer et al. 2007). To the final model an average of the results of the individual iterations was done. The selection of the wavelengths with greater impact was made from the importance of variables, calculated by the Gini index (Nembrini

et al., 2018), which is a way, in the RFR model, to address the magnitude of the relation between the wavelengths and the measured parameters.

Summing up the steps described above, the methodological approach used is this study was as follows: 1. In each one of the six vineyards, four plants were selected, and at each plant we selected a full-developed leaf, opposed to a grape cluster, under full solar illumination, located in the mean third of canopy, in a branch near the main vine trunk; being six parcels, a total of twenty-four leaves were selected. 2. Chlorophyll and radiometric measurements were performed at two phases of the vegetative cycle. These acquisitions produced, for each leaf and phase, a set of chlorophyll indices and a spectrum with 2151 reflectance values. 3. Using reflectance values for the required wavelengths, 19 different hyperspectral vegetation indices (HVI) were calculated for each leaf. 4. As one of our aims were to predict chlorophyll concentrations from radiometric data, three sets of input data were assembled to be applied by the two models to be tested (RFR and PLSR). These three data sets were composed by: i) 2151 variables (only reflectance bands), ii) 19 HVIs, and iii) 2170 variables (reflectance bands + HVIs).

5. Observed and predicted chlorophyll values were compared, and correlation accuracies were expressed, for each set of input data, by the following parameters: coefficient of determination (R^2), coefficient of determination with cross validation (R^2 (CV)), Root Mean Squared Error (RMSE), Root Mean Squared Error with cross validation (RMSE (CV)).

The model evaluation for prediction performance was made through the metrics R^2 and RMSE (*Root Mean Squared Error*), using Cross-Validation (CV) and dividing each input set in K-Fold 5. The Python library packages used were *pandas*, *Numpy*, *Scipy*, *Sk-learn* and *Matplotlib* (Pip · PyPI, 2022.)

3.2.3 RESULTS AND DISCUSSION

3.2.3.1 Chlorophyll Parameters

Variations in chlorophyll levels at the studied Cabernet Sauvignon parcels are shown in Figure 1. Among the six parcels, plot 16a presented a large variation in *Chl a* at the first acquisition. Parcels 4b and 16b showed large variations at the second acquisition run. It was noted that plot 16a, at the second acquisition run, presented a significant number of clusters with anomalies, as they were attained by a late-cycle fungus disease, *Glomerella cingulata*, which occurs at conditions of high temperatures and humidity, coupled with nitrogen excess. Other analyzed parameters were the ratios between chlorophylls. Parcels 16a (first acquisition), 16b(1^{st}), 19a(1^{st} and 2^{nd}) and 19b(1^{st}) presented medians of about 3.0, an indication of plants under high luminosity conditions (Lichtenthaler, 1987), what is confirmed by the open landscape at the site. Parcels 19a and 19b presented vines with smaller, open canopy, being noted that 19a is located has the highest elevation, with exposed, rocky soils and little vegetation between rows. The highest *Chl a* and *Chl b* concentrations were found at parcels 4a(2^{nd}), 4b(1^{st}) and 16a(1^{st}). In general, it was observed that as the vegetative cycle advanced and senescence approached a decrease in pigment variability and in the respective ratios took place.

3.2.3.2 Correlational Spectral Analysis

The correlograms for the reflectance spectra and the HVIs at both acquisition dates are presented at Figure 2. It was seen that graphs a) and c) suggest a decrease of the correlation between the more distant wavelengths, and a geometrical delineation between neighboring wavelengths, mainly at the NIR spectral range. Graphs b) and c) present positive and negative correlations between indices. Changes in leaf characteristics along the cycle lead to a decrease in water content, with a reduction in photosynthetic activity and changes in leaf colors, effects due to senescence which impact the spectrum (Boyer et al., 1988).

3.2.3.3 Prediction of Chlorophyll Parameters

The metrics which qualify the chlorophyll predictions from the PLSR model, for the three input sets and two acquisition dates, with a total of 24 results, are presented in Table 1. As an average, each sub-model had used from two to eight Principal Components (PC) as input data for the prediction. The estimate model for *Chl b* on the second date was the data set reduced by PCA with the largest number of wavelengths (892) in the analysis; on the other hand, the models for *Chl a* and *Chl (a/b)* had the smaller number of wavelengths used on the PCs, being eight and nine, respectively.

As the model's difficulty to predict chlorophyll levels increased, so increased the number of input variables; with the cycle's progress toward senescence, leaves have smaller photosynthetic activity, and leaf chlorophyll content gets smaller leading to color change to yellow-orange shades (Boyer et al., 1988; Jensen, 2006). It was noted that those models operating to chlorophyll parameters with larger number of input PCs presented a better performance as expressed by prediction metrics.

The estimate of *Chl b* from the Wavelengths data set was the one which obtained the largest R^2 (0.622) at the second acquisition run. Predictions using only the HVI data set presented lower performances, with $R^2 < 0.4$. However, using the whole input data set, Wavelengths + HVI, the PLSR model performed well for *Chl a*, *Chl b* and *Chl(a+b)*, with R^2 values above 0.60. Metrics expressed by R^2 with cross validation (R^2(CV)) were smaller than 0.462, suggesting insufficient entries to model validation. Finally, *Chl(a/b)* obtained the smallest RMSE at both dates, between 0.115 and 0.250, and similar values for RMSE (CV).

Going now to the Random Forest Regressor (RFR) model, R^2 values larger than 0.876 were obtained, regardless of the input set (Table 2). Parameter *Chl(a/b)* got RMSE values between 0.18 and 0.20 at the two measuring runs, these being the smaller errors among the measured parameters. Wavelengths selected by the model were concentrated at red and red-

edge spectral regions. Therefore, the RFR model presented higher predictive accuracy that the PLSR model. In general, the predictive metrics with cross validation presented values which were smaller than the training ones, with R^2(CV) around 0.113 and 0.387 and RMSE(CV) from 0.243 to 3.561.

For the PLSR coefficients the impact of each wavelength in the estimates of chlorophyll parameters are presented in Figure 3. Parameters *Chl* a and *Chl(a/b)* presented the lesser number of wavelengths in the spectral modeling, with the red and red-edge spectral regions presenting high sensitivity to chlorophyll content; in these spectral regions, reflectance depends linearly on leaf chlorophyll content (Steele et al., 2008b). For other parameters the model used data from several spectral regions, from the visible to SWIR2. A significant increase of wavelengths inserted at the Principal Components happened at the second acquisition run.

The importance of the Gini index to wavelengths and vegetation indices is expressed at Figure 4. The index is used to select the variables more relevant to prediction; presently, given the HVIs input data set, at first acquisition the most important indices were CAI, CARI and PSRI2, and for wavelengths the blue ones were more important. The PRI index was anomalous at both acquisitions.

3.2.4 CONCLUSION

The results presented in this investigation suggest that chlorophyll content can be predicted from hyperspectral data. Collinearity between wavelengths presented some stability in specific spectral regions, mainly later in the vegetative cycle, when the second data acquisition took place; at this stage, the observed indices tend to be more stable. Parameters Chl a and Chl(a/b) were in general the ones with less wavelengths used as input data sets in prediction models PLSR and RFR . The PSRI index was an important variable in almost all

calibrations done using the RFR model, while PLSR model included a larger variable number at second acquisition.

Forthcoming studies may focus on deeper investigations on effects of radiation on the physico-chemical structure of vine leaves, correlating spectroradiometric data with data from remote sensors at high resolutions, both in space and time. Data on physiological parameters bring more information on environmental effects on leaf characteristics, and extending the analysis presently reported to other grape varieties will help to a better understanding on the spectral behavior due to genetic factors.

3.2.5 REFERENCES

Abbasi, M., Verrelst, J., Mirzaei, M., Marofi, S., & Bakhtiari, H. R. R. (2020). Optimal spectral wavelengths for discriminating orchard species using multivariate statistical techniques. *Remote Sensing*, *12*(1). https://doi.org/10.3390/RS12010063

Bergsträsser, S., Fanourakis, D., Schmittgen, S., Cendrero-Mateo, M. P., Jansen, M., Scharr, H., & Rascher, U. (2015). HyperART: Non-invasive quantification of leaf traits using hyperspectral absorption-reflectance-transmittance imaging. *Plant Methods*, *11*(1), 1–17. https://doi.org/10.1186/s13007-015-0043-0

Blackburn, G. A. (1998). Quantifying Chlorophylls and Caroteniods at Leaf and Canopy Scales: An Evaluation of Some Hyperspectral Approaches. *Remote Sensing of Environment*, *66*(3), 273–285. https://doi.org/10.1016/S0034-4257(98)00059-5

Boyer, M., Miller, J., Belanger, M., Hare, E., & Wu, J. (1988). Senescence and spectral reflectance in leaves of northern pin oak (Quercus palustris Muenchh.). *Remote Sensing of Environment*, *25*(1), 71–87. https://doi.org/10.1016/0034-4257(88)90042-9

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cheng, J.-H., & Sun, D.-W. (2017). Partial Least Squares Regression (PLSR) Applied to NIR and HSI Spectral Data Modeling to Predict Chemical Properties of Fish Muscle. *Food Engineering Reviews*, *9*(1), 36–49. https://doi.org/10.1007/s12393-016-9147-1

El-Hendawy, S., Al-Suhaibani, N., Elsayed, S., Alotaibi, M., Hassan, W., & Schmidhalter, U. (2019). Performance of optimized hyperspectral reflectance indices and partial least squares regression for estimating the chlorophyll fluorescence and grain yield of wheat grown in simulated saline field conditions. *Plant Physiology and Biochemistry*, *144*(June), 300–311. https://doi.org/10.1016/j.plaphy.2019.10.006

Falker. (2009). *clorofiLOG CFL1030*. 1–6.

Fassnacht, F. E., Stenzel, S., & Gitelson, A. A. (2015). Non-destructive estimation of foliar carotenoid content of tree species using merged vegetation indices. *Journal of Plant Physiology*, *176*, 210–217. https://doi.org/10.1016/J.JPLPH.2014.11.003

Feng, Y., Cui, N., Gong, D., Zhang, Q., & Zhao, L. (2017). Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management*, *193*, 163–173. https://doi.org/10.1016/j.agwat.2017.08.003

Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves¶. *Photochemistry and Photobiology*, *74*(1), 38. https://doi.org/10.1562/0031-8655(2001)074<0038:opaneo>2.0.co;2

Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy¶. *Photochemistry and Photobiology*, *75*(3), 272. https://doi.org/10.1562/0031-8655(2002)075<0272:ACCIPL>2.0.CO;2

Jensen, J. R. (2006). *Remote Sesing of the Environment: An Earth Resource Perspective* (P. Hall, Ed.; 2ª).

Kawamura, K., Tsujimoto, Y., Rabenarivo, M., Asai, H., Andriamananjara, A., & Rakotoson, T. (2017). Vis-NIR Spectroscopy and PLS Regression with Waveband Selection for Estimating the Total C and N of Paddy Soils in Madagascar. *Remote Sensing*, *9*(10), 1081. https://doi.org/10.3390/rs9101081

Lacar, F. M., Lewis, M. M., & Grierson, I. T. (2002). Use of hyperspectral reflectance for discrimination between grape varieties. *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, *6*(C), 2878–2880. https://doi.org/10.1109/IGARSS.2001.978192

Lichtenthaler, H. K. (1987). [34] Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods in Enzymology*, *148*(C), 350–382. https://doi.org/10.1016/0076-6879(87)48036-1

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), 1–14. https://doi.org/10.3390/rs10020202

LORENZ, D. H., EICHHORN, K. W., BLEIHOLDER, H., KLOSE, R., MEIER, U., & WEBER, E. (1995). Growth Stages of the Grapevine: Phenological growth stages of the grapevine (Vitis vinifera L. ssp. vinifera)—Codes and descriptions according to the extended BBCH scale.

Australian Journal of Grape and Wine Research, *1*(2), 100–103. https://doi.org/10.1111/j.1755-0238.1995.tb00085.x

Meneses, P., Almeida, T., & Baptista, G. (2019). *Reflectância dos materiais terrestres: análise e interpretação*. Oficina e Textos.

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, *106*(1), 135–141. https://doi.org/10.1034/j.1399-3054.1999.106119.x

Mirzaei, M., Marofi, S., Abbasi, M., Solgi, E., Karimi, R., & Verrelst, J. (2019). Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. *International Journal of Applied Earth Observation and Geoinformation*, *80*(January), 26–37. https://doi.org/10.1016/j.jag.2019.04.002

Mitra, S., Irshad, M., Debnath, B., Lu, X., Li, M., Dash, C. K., Rizwan, H. M., Qiu, Z., & Qiu, D. (2018). Effect of vineyard soil variability on chlorophyll fluorescence, yield and quality of table grape as influenced by soil moisture, grown under double cropping system in protected condition. *PeerJ*, *2018*(9). https://doi.org/10.7717/PEERJ.5592

Nagler, P. L., Inoue, Y., Glenn, E. P., Russ, A. L., & Daughtry, C. S. T. (2003). Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes. *Remote Sensing of Environment*, *87*(2–3), 310–325. https://doi.org/10.1016/J.RSE.2003.06.001

Nembrini, S., Konig, I. R., & Wright, M. N. (2018). *Supplementary material for "The revival of the Gini importance?"*

Ordóñez, C., Oviedo de la Fuente, M., Roca-Pardiñas, J., & Rodríguez-Pérez, J. R. (2018). Determining optimum wavelengths for leaf water content estimation from reflectance: A

distance correlation approach. *Chemometrics and Intelligent Laboratory Systems*, *173*(November 2017), 41–50. https://doi.org/10.1016/j.chemolab.2017.12.001

Peñuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectânce. *Photosynthetica*, *2*, 221–230.

Peñuelas, J., Filella, I., & Gamon, J. A. (1995). Assessment of photosynthetic radiation-use efficiency with spectral reflectance. *New Phytologist*, *131*(3), 291–296. https://doi.org/10.1111/j.1469-8137.1995.tb03064.x

Penuelas, J., Pinol, J., Ogaya, R., & Filella, I. (1997). Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). *International Journal of Remote Sensing*, *18*(13), 2869–2875. https://doi.org/10.1080/014311697217396

pip · PyPI. (n.d.). Retrieved July 12, 2022, from https://pypi.org/project/pip/

Pithan, P. A., Ducati, J. R., Garrido, L. R., Arruda, D. C., Thum, A. B., & Hoff, R. (2021). Spectral characterization of fungal diseases downy mildew, powdery mildew, black-foot and Petri disease on Vitis vinifera leaves. *International Journal of Remote Sensing*, *42*(15), 5680–5697. https://doi.org/10.1080/01431161.2021.1929542

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*. https://doi.org/10.3390/ht8040021

Rouse, J. W. ;, Haas, R. H. ;, Schell, J. A. ;, & D.W., D. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the Third Earth Resources Technology Satellite- 1 Symposium*, 301–317.

Saheb Ettabaa, K., & ben Salem, M. (2017). Adaptive Progressive Band Selection for Dimensionality Reduction in Hyperspectral Images. *Journal of the Indian Society of Remote Sensing 2017 46:2*, *46*(2), 157–167. https://doi.org/10.1007/S12524-017-0691-9

Serrano, L., Peñuelas, J., & Ustin, S. L. (2002). Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sensing of Environment*, *81*(2–3), 355–364. https://doi.org/10.1016/S0034-4257(02)00011-1

Steele, M., Gitelson, A. A., & Rundquist, D. (2008a). Nondestructive estimation of leaf chlorophyll content in grapes. *American Journal of Enology and Viticulture*, *59*(3), 299–305.

Steele, M., Gitelson, A. A., & Rundquist, D. (2008b). Nondestructive estimation of leaf chlorophyll content in grapes. *American Journal of Enology and Viticulture*, *59*(3), 299–305.

Thum, A. B., Arruda, D. C., Ducati, J. R., Pithan, P. A., & Rolim, S. B. A. (2020). The influence of mineral content on spectral features of vine leaves. *International Journal of Remote Sensing*, *41*(23), 9161–9179. https://doi.org/10.1080/01431161.2020.1798547

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

Vogelmann, J. E., Rock, B. N., & Moss, D. M. (2007). Red edge spectral measurements from sugar maple leaves. *Http://Dx.Doi.Org/10.1080/01431169308953986*, *14*(8), 1563–1575. https://doi.org/10.1080/01431169308953986

Yang, X.-H., Huang, J.-F., Wang, F.-M., Wang, X.-Z., Qiu-Xiang, Y. I., & Yuan, W. (2006). A modified chlorophyll absorption continuum index for chlorophyll estimation *. *J Zhejiang Univ SCIENCE A*, *7*(12), 2002–2006. https://doi.org/10.1631/jzus.2006.A2002

Zhang, L., Su, H., & Shen, J. (2019). Hyperspectral dimensionality reduction based on multiscale superpixelwise kernel principal component analysis. *Remote Sensing*, *11*(10). https://doi.org/10.3390/rs11101219

Zhang, N., Liu, X., Jin, X., Li, C., Wu, X., Yang, S., Ning, J., & Yanne, P. (2017). Determination of total iron-reactive phenolics, anthocyanins and tannins in wine grapes of skins and seeds based on near-infrared hyperspectral imaging. *Food Chemistry*, *237*, 811–817. https://doi.org/10.1016/j.foodchem.2017.06.007

Zhao, J., Huang, L., Huang, W., Zhang, D., Yuan, L., Zhang, J., & Liang, D. (2014). Hyperspectral measurements of severity of stripe rust on individual wheat leaves. *European Journal of Plant Pathology*, *139*(2), 401–411. https://doi.org/10.1007/s10658-014-0397-6

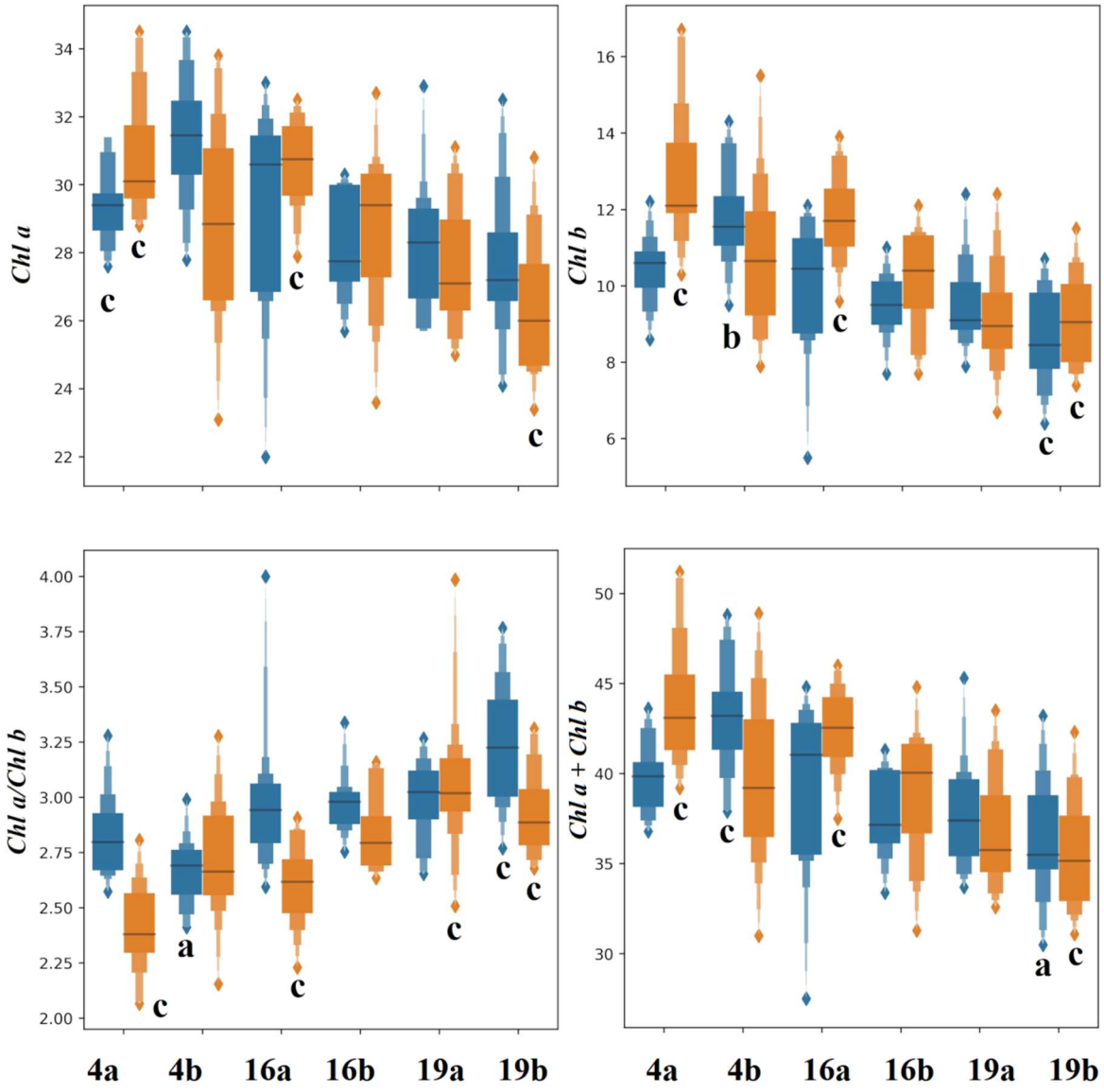

Figura 8. Main effects spatial variability among the six plots on the two collection days and the interactions among them (KrusKal-Wallis), with repeated measures ($P \leq 0.05$) for all leaf parameters studied: chlorophyll a concentration (Chl a), chlorophyll b (Chl b), total chlorophyll (Chl a + Chl b), chlorophyll ratio (Chl a / Chl b). Each letter represents the statistically significant differences between the parameters: a) differs from one plot; b) differs from two plots; c) differs from three plots or more.

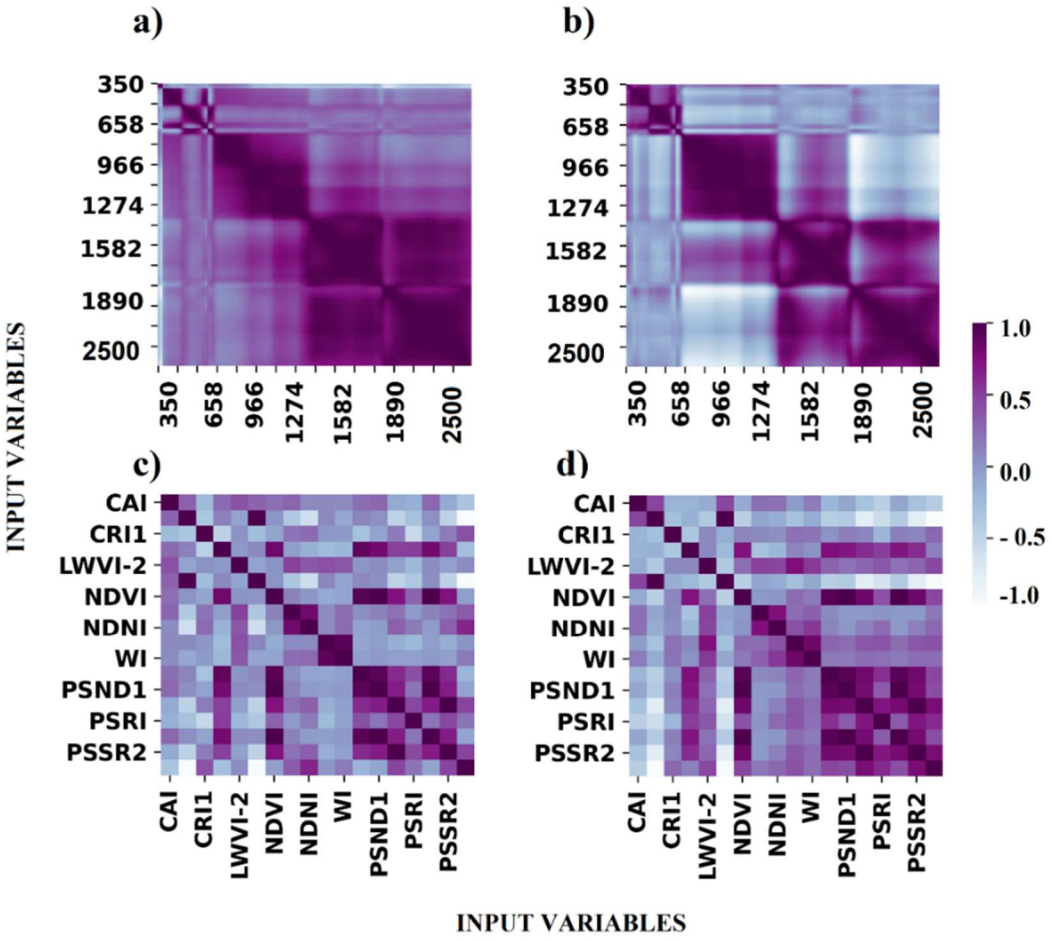

Figura 9. Spearman Correlation Coefficient Rank (r), applied to wavelengths (350nm - 2500nm) of hyperspectral leaf reflectance and to HVIs, for acquisition dates: Dec. 16, 2017 (a and c) and Feb. 27, 2018 (b and d).

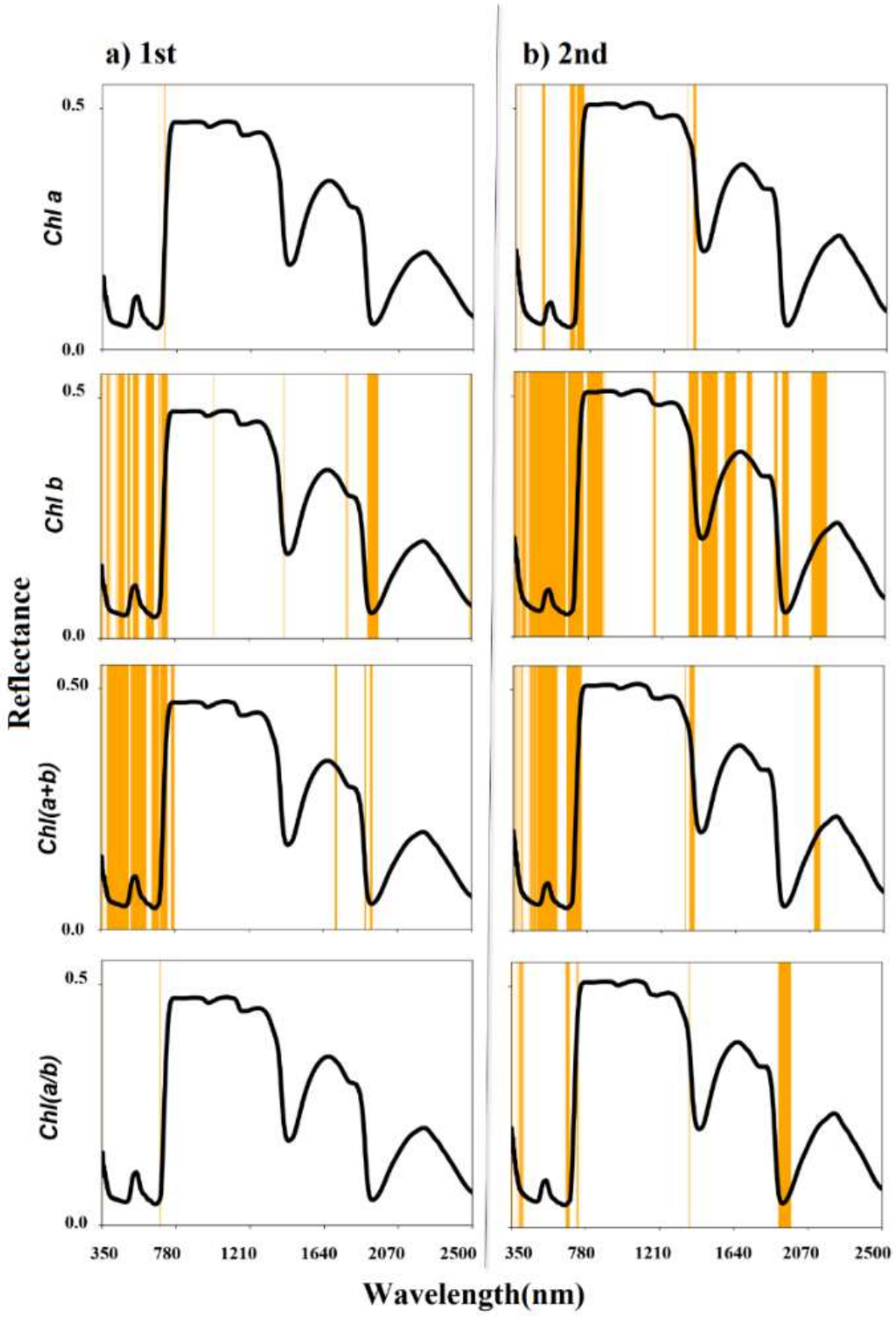

Figura 10. Selected wavelengths (highlighted in yellow) with the best coefficients of the PSLR components, at the observed spectral domain (350nm-2500nm), in each prediction; column a corresponds to the 1st acquisition date, and b to the 2nd.

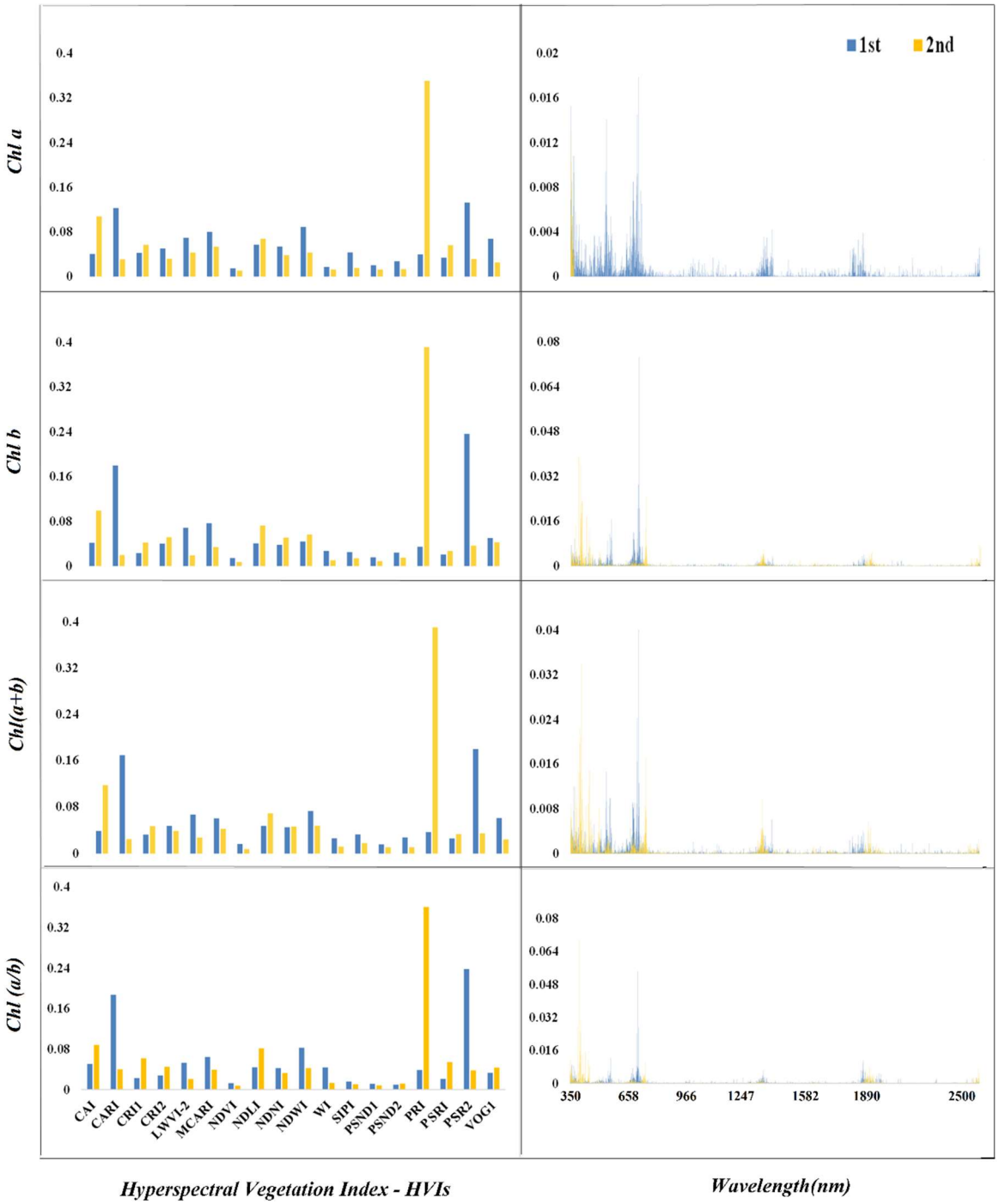

Figura 11. Gini Importance to the HVIs(a) and wavelengths(b) in modeling of each chlorophyll parameter measured in the field.

Tabela 3. Metrics used for the predictions carried out with the PLSR model for the three input databases and on the two acquisition dates. The parameters used in the analyses are: principal components number (PC), wavelengths numbers (VN), coefficient of determination (R^2), coefficient of determination with cross validation (R^2 (CV)), Root mean square error (RMSE), Root mean square error with cross validation (RMSE (CV)).

	Parameters	**PC**		**VN**		***R²***		***R²*(CV)**		**RMSE**		**RMSE(CV)**	
		1st	**2nd**	**1st**	**2nd**	**1st**	**2nd**	**1st**	**2nd**	**1st**	**2nd**	**1st**	**2nd**
Wavelengths	***Chl a***	3	5	9	141	0.294	0.551	0.214	0.415	1.995	1.748	2.104	1.994
	Chl b	8	8	332	892	0.585	0.622	0.359	0.432	1.025	1.254	1.274	1.537
	Chl (a+b)	3	6	391	360	0.380	0.579	0.296	0.462	3.073	2.952	3.274	3.336
	Chl (a/b)	7	4	8	157	0.576	0.528	0.471	0.391	0.183	0.214	0.205	0.243
HVIs	***Chl a***	2	5	6	7	0.286	0.468	0.228	0.336	2.006	1.902	2.086	2.125
	Chl b	3	4	5	5	0.404	0.411	0.317	0.274	1.228	1.565	1.315	1.738
	Chl (a+b)	3	5	4	6	0.345	0.446	0.278	0.311	3.157	3.385	3.316	3.775
	Chl (a/b)	3	4	6	9	0.365	0.352	0.204	0.115	0.225	0.250	0.251	0.293
Wavelenths+HVIs	***Chl a***	2	8	273	120	0.293	0.639	0.239	0.383	1.996	1.567	2.070	2.047
	Chl b	2	6	103	203	0.409	0.630	0.354	0.399	1.224	1.241	1.248	1.581
	Chl (a+b)	2	8	9	121	0.333	0.659	0.292	0.380	3.188	2.656	3.283	3.581
	Chl (a/b)	3	4	254	271	0.415	0.500	0.312	0.369	0.215	0.220	0.234	0.247

Tabela 4.. Predictions performance metrics for the Random Forest Regressor model for the three input databases on the two acquisition dates. The parameters used in the analyses **are** Coefficient of determination (r^2); Coefficient of determination with Cross validation (r^2 CV); Root mean square error (RMSE); Root mean square error with Cross validation(RMSE CV).

	Parameters	***R²***		***R²*(CV)**		**RMSE**		**RMSE(CV)**	
		1st	**2nd**	**1st**	**2nd**	**1st**	**2nd**	**1st**	**2nd**
Reflectance	***Chl a***	0.874	0.893	0.113	0.242	0.844	0.853	2.230	2.271
	Chl b	0.897	0.901	0.290	0.337	0.510	0.643	1.341	1.661
	Chl (a+b)	0.877	0.897	0.293	0.288	1.312	1.458	3.403	3.836
	Chl (a/ b)	0.899	0.956	0.258	0.293	0.090	1.250	0.243	1.100
Index	***Chl a***	0.876	0.909	0.157	0.365	0.836	0.786	2.180	2.078
	Chl b	0.911	0.905	0.391	0.355	0.475	0.629	1.242	1.638
	Chl (a+b)	0.894	0.915	0.282	0.387	1.269	1.327	3.307	3.561
	Chl (a/ b)	0.910	0.902	0.307	0.258	0.085	0.097	0.235	0.268
Reflec+Index	***Chla***	0.899	0.905	0.231	0.353	0.793	0.803	2.082	2.097
	Chlb	0.909	0.908	0.375	0.373	0.480	0.617	1.258	1.614
	Chl (a+b)	0.898	0.913	0.304	0.380	1.247	1.338	3.255	3.579
	Chl (a/ b)	0.906	0.911	0.342	0.339	0.087	0.093	0.228	0.253

3.3 ARTIGO 3: Estimation of degree Brix in grapes by proximal hyperspectral sensing and nanosatellite imagery through the Random Forest Regression algorithm

Diniz Carvalho de Arruda,[a] Jorge Ricardo Ducati,[a] Renata Pacheco Quevedo [b], Pâmela Aude Pithan,[a], Adriane Brill Thum.[c] Rosemary Hoff.[d]

a Universidade Federal do Rio Grande do Sul - UFRGS, Remote Sensing Center, Av.Bento Gonçalves 9500, Porto Alegre, Brazil, CEP 91501-970.

b Instituto Nacional de Pesquisas Espaciais - INPE, Av. dos Astronautas, 1758, São José dos Campos – SP, Brazil, CEP 12.227-010.

c Universidade do Vale do Rio dos Sinos - Unisinos , Av.Unisinos, São Leopoldo, Brazil, CEP 93022-000.

d Empresa Brasileira de Pesquisa Agropecuária – Embrapa , Av. Livramento 515, Bento Goncalves, Brazil, CEP 95701-008.

Abstract. The assessment of physiological parameters in vineyards can be done by direct measurements or by remote, indirect methods. The latter option frequently yields useful data and development of methods and techniques that make them possible is worthwhile. One of the parameters most looked for to define the quality status of a vineyard is the degree Brix of its grapes, a quantity usually determined by direct measurement. However, other ways may be possible, and presently Brix estimations in vineyards using as data sources field radiometry, localized Brix measurements and satellite imagery are reported. The investigation was developed in a commercial vineyard in south Brazil at two stages of the 2017/2018 vegetative cycle. Brix estimates were derived using a machine learning model, the Random Forest Regression (RFR) algorithm. Results produced coefficients of correlation between observed and predicted degrees Brix as high as 0.89. Analysis of an importance parameter, the Gini index, suggested that spectral data at ultraviolet, visible, and near-infrared wavelengths and the vegetation indices TGI and NDVI are the most important variables used for the predictive model. This methodology is potentially useful for the derivation of vineyard quality parameters at situations when specific vineyard conditions, as rugged terrain and large variations in soils, turn direct measurements a difficult task.

Keywords: degree Brix, hyperspectral data, Random Forest Regression.

*Diniz Carvalho de Arruda**,** E-mail: dinizcarvalho28@outlook.com

3.3.1 INTRODUCTION

Grape producers are in constant search for fruit excellency, and for this sake factors that have to be in close monitoring include soil fertility, vine susceptibility to pathogen attacks, and effects of humidity, temperature, and exposure to sunlight. These procedures are closely related to precision viticulture, and one of the techniques employed in such approaches are those of remote sensing, supporting the decision-making of producers with the help of remotely acquired data carrying information about the vineyard [1–4]. Studies using remote sensing technologies can

help check plots of land, spatially manage vine vigor characteristics, and qualify homogeneous production zones[5,6]. Medium and high-resolution images are used to delineate vineyards, and by spectral indices it is possible to detect stresses related to water status, diseases, and soil characteristics [3,7–9]

Moreover, remote sensing is also used at a proximal, near-plant scale, through the use of hyperspectral sensors, with the potential to detail the characteristics of vines in a non-destructive way[10–12] Hyperspectral indices were also used to check the grape composition in the field and laboratory [13,14]

Furthermore, spectroscopic modeling via proximal and satellite sensors supplies information on the progress of vineyard production throughout the harvest. Non-invasive methods for monitoring and controlling the nutritional and pathological state of the vineyard are essential. Soil macros/micronutrients, air and soil temperatures, solar exposure, rainfall, and management are relevant factors when fruit quality is looked for, and for this sake following vine development throughout the phenological cycle becomes indispensable.

One of the most searched parameters related to fruit quality is the berry sugar content, which is expressed by the Brix degree (ºBrix), or briefly Brix, a measure of the number of dissolved solids in a liquid via its specific gravity, used especially to measure dissolved sugar in fruits ([15]) ºBrix is commonly estimated at the vineyard by devices called refractometers, a direct measurement performed on berries, but alternative techniques may eventually prove as more practical, for example in occasions where ºBrix have to be known in remote vineyards, or in very large surfaces, especially if in this last case the vineyard extends over areas with rugged or inhomogeneous soils, where remote sensing techniques may be useful. Such indirect determination or prediction of needed parameters and its validation by comparison with direct measurements can be done by machine learning models like the Random Forest Regression

(RFR), a model tolerant to noisy data which evaluates correlations between variables using a random vector. The RFR performance is high in setting spectral reflectance measurements, because of its low sensitivity to outliers [16,17]. Hence, this study has the objective to estimate the grape degree Brix in a commercial vineyard by proximal sensing and nanosatellite imagery using the RFR machine learning model.

3.3.2 MATERIALS AND METHODS

3.3.2.1 Study area and sampling criteria

As study area the Luiz Argenta Winery was chosen, for reasons of access facility and favorable topography. It is important Brazilian producers of fine wines, located at the northeastern part of State of Rio Grande do Sul, in Flores da Cunha municipality. The total cultivated area is approximately 48 hectares, and six Cabernet Sauvignon plots were selected, this choice of vine grapes of the same variety being made to avoid confusion due to genetic differences from two or more different varieties with influences on observable properties.

Conduction system was espalier trellis, rootstocks were Paulsen 1103, viticultural treatments were conventional with weekly applications, row orientation was east-west solar exposure (plots 4a, 4b,16a, 16b, following the winery's naming use) and north-south (plots 19a, 19b, both at higher elevations with respect to other plots); row spacing was 2.8 meters, with 1.45 meters between plants at same row. The availability of soil sampling information, easy access to the vineyards, and collection time were considered in the selection process of these plots. At each plot four plants located at central rows were selected and marked.

The experiment was conducted in two stages of the grape ripening cycle. As data sources this investigation was based on three input data sets, considered to be fundamental to following the evolution of the spectral behavior of the leaves at distinct stages of sugar concentration in the grape, as follows:

3.3.2.2 Spectroradiometric

Field spectroradiometric measurements were performed on December 15, 2017 and February 27, 2018. The first date coincides with the veraison stage of the phenological cycle, and the second with the phenolic ripeness, being close to harvest time. The equipment used for the measurements was a spectroradiometer Field Spec® *3* ASD – Analytical Spectral Devices, which performs reflectance measurements at intervals of three to five angstroms, in the 350nm to 2500nm domain, including, therefore, the spectral ranges from the ultraviolet (UV) to the near and mid-infrared (NIR, SWIR1 and 2)[10].

The Leaf Clip probe accessory was used, and measurements were taken in four leaves per plant, in four points for each leaf, on the adaxial part of the selected leaves, opposite to the clusters under full light, in the middle third of the canopy, on branches closest to the main trunk, for four plants per plot. Calibration measurements were performed with a built-in Spectralon® reference plate.

Calculations were done to produce a single, average spectrum per plant. At the end of each observing run, considering four plants per vine plot and six plots, 24 reflectance spectra were available for analysis. On these spectra, corrections for filtering noisy data and reflective plateau shifts were performed using Savitzky-Golay[18] filters, Jump Correction [19] and normalize[20] function.

3.3.2.3 ºBrix measurements

A refractometer was used to measure the ºBrix of grapes, an expression of soluble solids content[21]. Measurements were made on February 8, 2018 at temperatures between 23ºC to 30ºC. Four grape clusters were chosen at lower parts of the selected plants, and at each cluster the degree Brix was measured at four berries at lower parts of the clusters. Averages were made for each plant and the resulting data set had twenty-four ºBrix values.

3.3.2.4 Satellite images

Two images from the Dove-R sensor aboard PlanetScope nanosatellites were acquired at dates December 16, 2017, and February 28, 2018. The product level 3B, Surface Reflectance (SR), provides images with radiometric and atmospheric correction, projected to plane coordinates, with spatial resolution of about three meters. There are four bands: blue (465nm-515nm), green (547nm – 583nm), red (650nm – 680nm), and NIR (845nm – 885 nm)[22].

Both images were treated to the radiometric normalization process, a process of reducing spectral uncertainties, influenced by atmospheric variations from one year to another. The code was developed in a computational environment with the R language, based on the normalization method. The radiometric normalization model was done through a PIF – Pseudo-Invariant Features technique, that is, using control points without radiometric variation, on a reference image with Normalized Difference Vegetation Index (NDVI)[23]. The 2017 image was used as a reference and the homologous points, with the least possible radiometric variation, band by band (José et al., 2017). A new linear transformation was applied to each band from a spectral sampling grid. The normalization method applied in this work was described by [25].

After the treatment of the images, a pixel sampling procedure was done for the creation of the grape Brix prediction models. The selection of pixel samples in the images happened with the creation of a buffer with an area of 9 m² around the midway point of the plants selected for this study.

3.3.2.4.1 Vegetation Index

Vegetation indices are biophysical parameters that provide important information about canopy structure and leaf traits of the vegetation[26,27] These variables were entered as input data into the estimation models and the following Vegetation Indices were calculated from the measured reflectance spectra: Normalized Difference Vegetation Index (NDVI)[23], Triangular

Greenness Index (TGI)[28] and Visible Atmospherically Resistant Index (VARI) [29]. These indices were added as input variables in the predictive models and analyzed on time with the field data.

3.3.2.5 Random Forest Regression

Machine learning models are widely used for classification processes [30] and estimation (Feng et al., 2017), and presently for prediction the non-parametric algorithm Random Forest Regressor (RFR) was used. The RF follows the principle of randomly using decision trees from a set of samples (Breiman, 2001). The main parameters of the model are estimators as tree depth, scores per bagging, and Out of Bagging (OOB), among others [20].

The selection of the input features by the RFR algorithm occurs by applying the importance criterion for the variables, Gini Index, a multivariate process to the variable that best divides the nodes randomly [30,33,34] Majority voting took place to select the best decision tree, in the classification process and validation utilizing OOB, a test set, consisting of 36% of the total samples (Immitzer et al., 2012). In tasks involving plant spectroscopy, OOB validations have supplied low error estimates in the visible and shortwave infrared spectral regions[16,30,35].

From the measured Brix values a table was compiled with the descriptive statistics for each of the six vine plots, including the maxima, average and minima values, besides coefficients of variation. The predicted Brix was estimated both from leaf hyperspectral reflectance spectra and from PlanetScope imagery, using the Random Forest Regression algorithm; hence, the data base formed by the average values for each plot was used for two input groups: ºBrix x Hyperspectral Reflectance and ºBrix x Surface Reflectance and Vegetation Indices. For RFR parameters *n_estimators* = [50,100,250,500] were used; these values express the maximal number of trees used at models to predict the Brix values from each sensor. Several estimators were tested to assess the performance stability of the model

predictions, expressed by the following metrics: Mean Square Error (MSE); Root Mean Square Error (RMSE); Coefficient of Determination (R^2); Adjusted R^2 (adj_R^2); and Out of Bagging (OOB). Processing was done in Python language environment, using libraries found in the open-source data manipulation and processing program Anaconda. The libraries used for exploratory, predictive, and visual analysis were Pandas, Numpy, MatplotLib, Seaborn, and Scikit- Learn.

After metrics analysis and comparison, the RFR models with 500 decision trees were chosen for spacialization using satellite data and the vegetation índices (NDVI, TGI and VARI) for the two acquisition dates, maps for predicted Brix being produced.

3.3.3 RESULTS AND DISCUSSION

Measurements of grape Brix (Figure 2) showed a coefficient of variation of about 5%-6%, except for plot 16a which presented the largest variation, around 9%. These results are compatible with other works estimating grape Brix values and variations close to summer harvest time (Pedro Júnior et al., 2014; Santos et al., 2011). The sugar content showed similar values for all plots, an expected behavior since a single variety was studied; however, the highest coefficient of variation was found for plot 16a in the second and last data reading, in which phytosanitary problems were found in some plants, bunches, and leaves with end-of-cycle diseases.

Peak ºBrix values were found at 19a plot; this parcel was the one with lesser density of plants and canopies, presenting smaller bunches, on stony soil with exposed surface. The higher variations in Brix at first acquisition date can be understood by the fact that at the beginning of the grape ripening cycle, in the berry color change, it is normal for the leaves to have gotten distinct chlorophyll values, since the plants are in the full phase of photosynthetic activity. [38].

The Cabernet Sauvignon grape variety is known for its long, late phenological cycle when compared to other varieties. However, the variations between plots can change the dynamics of each stage, each plot being days ahead or behind another plots with respect to the same stage. Therefore, and ideally, grape harvest does not occur simultaneously for all plots, grapes depending on climate-soil-plant interactions for definition of the best ripening.

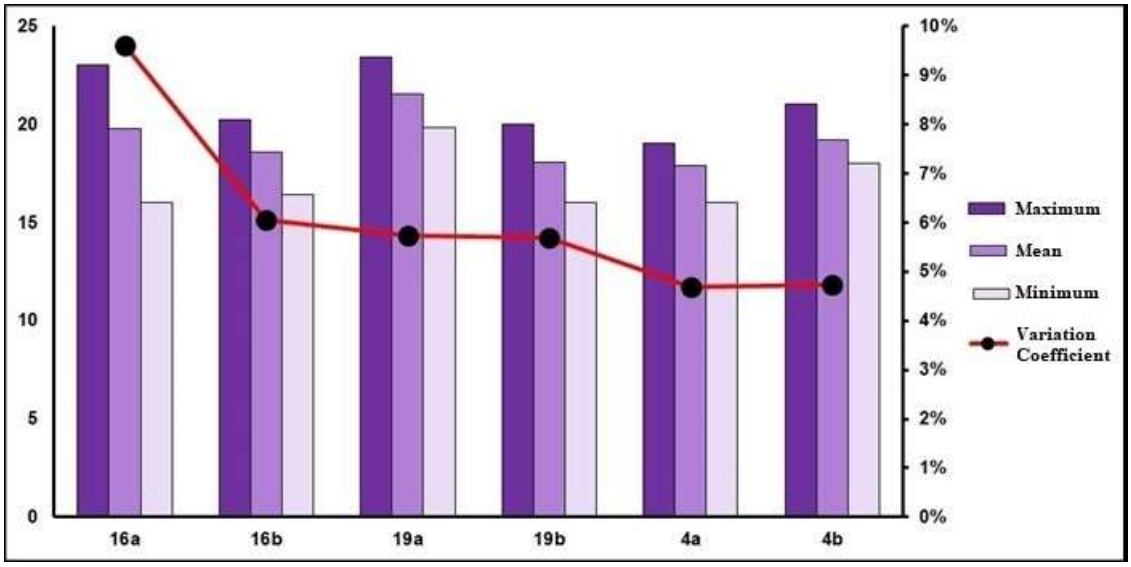

Figura 12. Descriptive statistics for grape Brix readings, refractometer data measured in situ on plants in the six studied Cabernet Sauvignon plots.

The prediction metrics of the models from both HR and SR input data sets are presented in Table 1, with number of estimators between 50 and 500. Predictions with the proximal sensor input data (HR) for the first date (1st) presented R^2 values above 0.86 independently of the number of estimators (1st), and for the second date (2nd) R^2 values were between 0.82 and 0.85. Still for HR, the OOB metric shows, with 100 estimators, a value of 0.03 and 0.19 for 2nd with 500 estimators, which were the best scores among all models for input from leaf hyperspectral reflectance. The estimation values are compatible with results reported by,[39], with R^2 between 0.83 and 0.92.

Keeping with Table 1, the results of the metrics for the input data from surface reflectance (SR) show that in both collection dates R^2 values were in the range between 0.84 and 0.87. The model with 500 estimators obtained the lowest OBB scores, with values of 0.01(1st) and 0.04 (2nd).

Tabela 5. Metrics for evaluating grape ºBrix estimates for hyperspectral data (Hyperspectral Reflectance, HR) and PlanetScope (Surface Reflectance and Indices, SR). Mean Square Error (MSE); Root Mean Square Error (RMSE); Coefficient of Determination (R^2); Adjusted R^2 (adj_R^2); Out of Bagging (OOB).

Parameters	Estimators	MSE		RMSE		R^2		adj_R^2		OOB	
		1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
HR	50	0.28	0.32	0.53	0.57	0.86	0.84	1.00	1.00	0.11	0.28
	100	0.27	0.36	0.52	0.60	0.87	0.82	1.00	1.00	0.03	0.20
	250	0.25	0.33	0.50	0.57	0.87	0.84	1.00	1.00	0.05	0.30
	500	0.26	0.30	0.51	0.55	0.87	0.85	1.00	1.00	0.08	0.19
SR	50	0.29	0.26	0.54	0.51	0.86	0.87	0.79	0.87	0.06	0.06
	100	0.21	0.25	0.46	0.50	0.89	0.88	0.85	0.82	0.14	0.05
	250	0.24	0.28	0.49	0.52	0.88	0.86	0.83	0.81	0.08	0.06
	500	0.27	0.26	0.52	0.51	0.86	0.87	0.81	0.82	0.01	0.04

Figure 3 shows the importance of the bands of each sensor in the estimation. For hyperspectral data (Figure 3a) the most important bands on the first date (1st) are concentrated in the spectral region of the ultraviolet, red, and near-infrared and some small areas in the SWIR2. In the second collection (2nd), the model selected wavelengths of the ultraviolet, blue, and SWIR1. For satellite data plus vegetation indices (Figure 3b) the Gini Index was used as indicator of importance, and indices VARI, NDVI, and TGI showed their importance to the spectral modeling. The NDVI showed relevance in both acquisition dates. TGI Index was only relevant in the first collection and B3 in the second collection. TGI is considered as a valuable indicator at more advanced phenological stages, at the onset of changes in leaf color and

decrease of photosynthetic[40,41] NDVI is a measure of the vegetative vigor of a vineyard, where high vigor is associated with low water stress, and medium vigor with greater leaf area; in general, it is considered that low vigor increases fruit quality parameters [42,43]. Spectral regions NIR and Red edge are sensitive to physiological changes in plants and to the stage of their development[38] .

Figure 5 presents the Brix estimate for the six vine plots. As the cycle progresses and grape ripeness approaches, it is possible to identify an increase in Brix (hatchings in dark violet shades). From the models it is suggested that rightness rate is high; even at coarser satellite resolutions it is possible to survey grape quality, as it is suggested by[44].

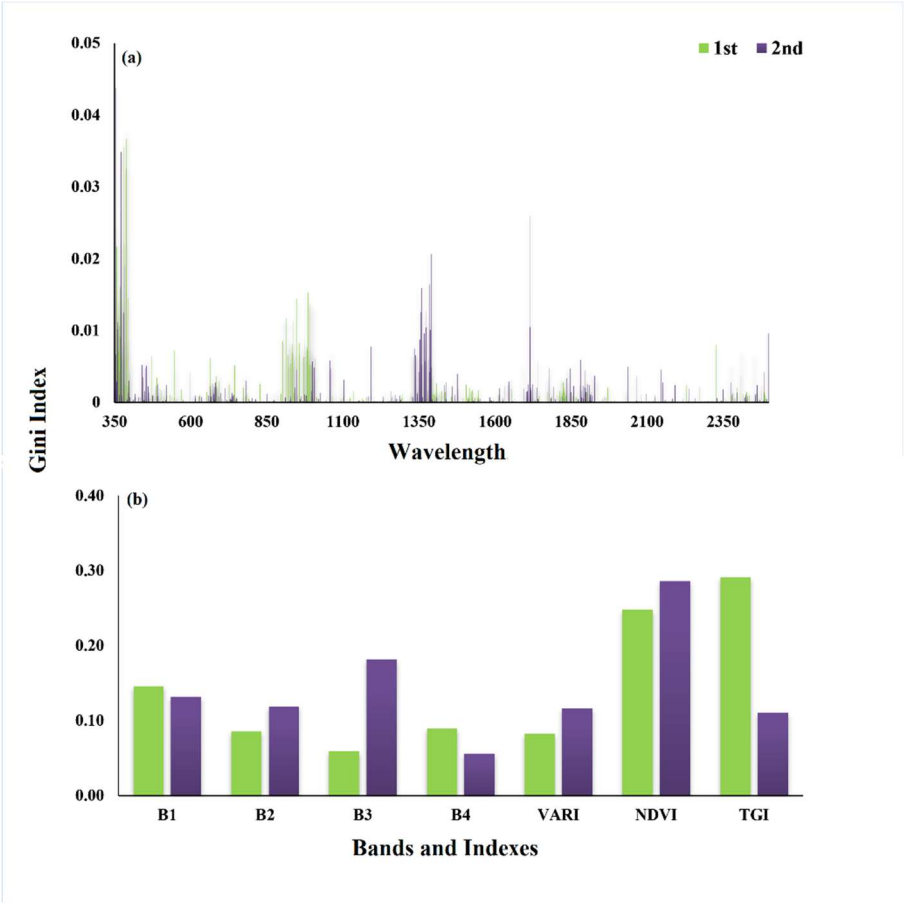

Figura 13.Gini index for feature selection in the models with 500 estimators: (a) Spectroradiometer; (b) PlanetScope.

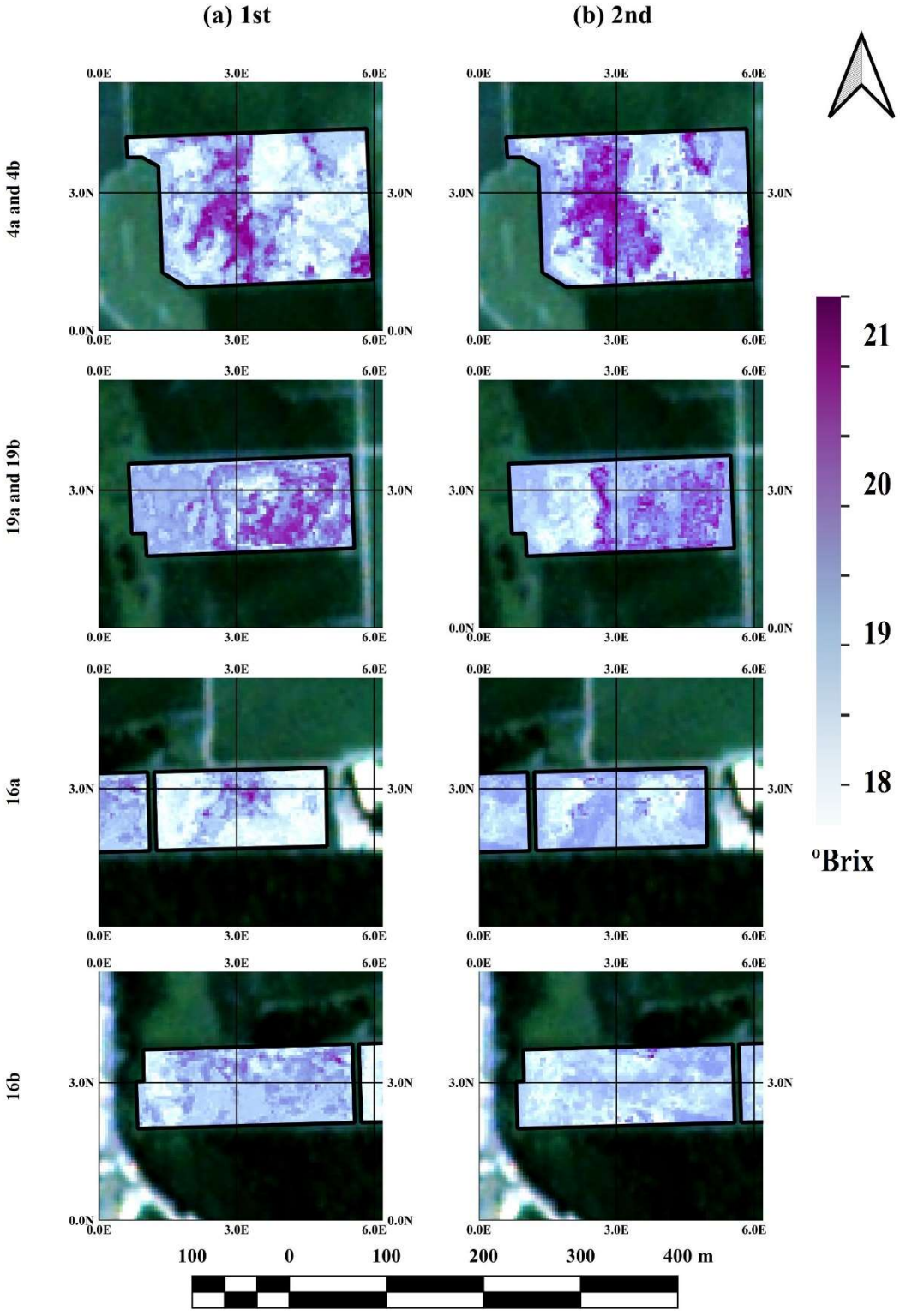

Figura 14.Degree Brix estimates to the six studied vineyards(4a, 4b, 19a, 19b, 16a and 16b). (a) December 2017; (b) February 2018.

4 CONCLUSION

The results presented here suggest that the derivations made using the Random Forest Regressor were significant, either based in proximal or satellite data; this perception is supported by analysis of the metrics evaluating the predictions for both dates. The measurements of ºBrix are linked to the vegetative state of the plant and the exchange processes between plant and its environment, involving an analysis more complex than only considering light reflection by leaves[42,43]. This is important to set up a better understanding between spatial factors and the development and monitoring of vine vegetative vigor.

The importance and utility of high-resolution sensors like Dove-R was highlighted, as well as that using nanosatellite and proximal derived products will help address the impacts of management, weather, soil, and climate on vineyard vegetative growth and consequently the quality of its grapes. For future studies, the development of a sampling grid with a higher quantity of plants, plots, and varieties will help to understand the identity of the spectral signatures of the vines and to seek new relationships with other grape quality and quantity parameters.

5 REFERENCES

C. Acevedo-Opazo et al., “The potential of high spatial resolution information to define within-vineyard zones related to vine water status,” Precis Agric **9**(5), 285–302 (2008) [https://doi.org/10.1007/s11119-008-9073-1].

2. A. Hosu et al., “The Influence of the Variety, Vineyard, and Vintage on the Romanian White Wines Quality,” J Anal Methods Chem **2016**, 1–33 (2016) [https://doi.org/10.1155/2016/4172187].

3. C. Poblete-Echeverría et al., “remote sensing Detection and Segmentation of Vine Canopy in Ultra-High Spatial Resolution RGB Imagery Obtained from Unmanned Aerial Vehicle (UAV): A Case Study in a Commercial Vineyard,” 268 (2017) [https://doi.org/10.3390/rs9030268].

4. C. V. Rombaldi et al., “Vineyard Yield and Grape Quality in Two Diferent Cultivation Systems,” Rev. Bras. Frutic., Jaboticabal-SP(1), 89–91 (2004).

5. J. R. Ducati, R. E. Sarate, and J. M. G. Fachel, “Application of remote sensing techniques to discriminate between conventional and organic vineyards in the Loire Valley, France,” OENO One **48**(3), 135 (2014) [https://doi.org/0.20870/oeno-one.2014.48.3.1574].

6. A. B. González-Fernández, T. Catanzarite Torres, and J. R. Rodriguez-Pérez, “Una Metodología Apoyada En Sig Para El Seguimiento Vitícola Y La Delimitación De Zonas Homogéneas De Vendimia En La D.O. Bierzo (León-España,” in Revista Internacional de Ciencia y Tercnología de la Información Geográfica (10), pp. 185–207 (2010).

7. A. Hall, J. Louis, and D. Lamb, “Characterising and mapping vineyard canopy using high-spatial- resolution aerial multispectral images,” Comput Geosci **29**(7), 813–822 (2003) [https://doi.org/10.1016/S0098-3004(03)00082-7].

8. A. Nasrallah et al., “A novel approach for mapping wheat areas using high resolution sentinel-2 images,” Sensors (Switzerland) **18**(7), MDPI AG (2018) [https://doi.org/10.3390/s18072089].

9. D. Helman et al., “Using time series of high-resolution planet satellite images to monitor grapevine stem water potential in commercial vineyards,” Remote Sens (Basel) **10**(10), 1615, MDPI AG (2018) [https://doi.org/10.3390/rs10101615].

10. M. Danner et al., *Spectral Sampling with the ASD FIELDSPEC 4 EnMAP Field Guides Technical Report* (2015) [https://doi.org/10.2312/enmap.2015.008].

11. J. Lu et al., “Field detection of anthracnose crown rot in strawberry using spectroscopy technology,” Comput Electron Agric **135**, 289–299, Elsevier B.V. (2017) [https://doi.org/10.1016/j.compag.2017.01.017].

12. E. Adam and O. Mutanga, “Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry,” ISPRS Journal of Photogrammetry and Remote Sensing **64**(6), 612–620 (2009) [https://doi.org/10.1016/j.isprsjprs.2009.04.004].

13. A. Power et al., “From the Laboratory to The Vineyard-Evolution of The Measurement of Grape Composition using NIR Spectroscopy towards High-Throughput Analysis” (2019) [https://doi.org/10.3390/ht8040021].

14. P. Martín et al., “Using hyperspectral remote sensing to map grape quality in ‘Tempranillo’ vineyards affected by iron deficiency chlorosis,” Vitis - Journal of Grapevine Research **46**(1), 7–14 (2007).

15. R. B. Boulton et al., “Principles and Practices of Winemaking,” Principles and Practices of Winemaking, Springer US (1999) [https://doi.org/10.1007/978-1-4757-6255-6].

16. R. S. Fletcher and K. N. Reddy, “Random Forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds,” Comput Electron Agric **128**, 199–206, Elsevier B.V. (2016) [https://doi.org/10.1016/j.compag.2016.09.004].

17. Y. Hong et al., “Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest,” Science of The Total Environment **651**, 1969–1982 (2019) [https://doi.org/10.1016/j.scitotenv.2018.09.391].

18. A. Savitzky and M. J. E. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” Anal Chem **36**(8), 1627–1639 (1964) [https://doi.org/10.1021/ac60214a047].

19. A. Hueni and A. Bialek, “Cause, Effect, and Correction of Field Spectroradiometer Interchannel Radiometric Steps,” IEEE J Sel Top Appl Earth Obs Remote Sens **10**(4), 1542–1551, Institute of Electrical and Electronics Engineers (2017) [https://doi.org/10.1109/JSTARS.2016.2625043].

20. P. Virtanen et al., “SciPy 1.0: fundamental algorithms for scientific computing in Python,” Nat Methods **17**(3), 261–272, Nature Publishing Group (2020) [https://doi.org/10.1038/s41592-019-0686-2].

21. M. A. Coêlho de Lima, “Colheita e pós-colheita,” Sistemas de Produção, August 2010, [http://www.cpatsa.embrapa.br:8080/sistema_producao/spuva/colheita.html#topo]

22. Planet, “Planet Imagery: Product Spesification,” Planet(January), 96 (2017).

23. J. W.; Rouse et al., “Monitoring vegetation systems in the Great Plains with ERTS,” Proceedings of the Third Earth Resources Technology Satellite- 1 Symposium, 301–317 (1974).

24. P. José et al., “Modelo Automático De Normalização Radiométrica De Série Multitemporal Landsat-5 Usando Pontos Pseudoinvariantes (Pif),” Revista Brasileira de Cartografia **69**(2), 241–251 (2017).

25. P. Gabriel and B. César, “Detecção de mudanças de uso e cobertur da terra por imagens de nanossatélites : estudo de caso do entorno da Aldeia Verdadeira (Anhetenguá) - Porto Alegre - RS” (2019).

26. A. Lausch et al., “Understanding Forest health with remote sensing-Part I-A review of spectral traits, processes and remote-sensing characteristics,” Remote Sens (Basel) **8**(12) (2016) [https://doi.org/10.3390/rs8121029].

27. M. H. Nunes, M. P. Davey, and D. A. Coomes, “On the challenges of using field spectroscopy to measure the impact of soil type on leaf traits,” Biogeosciences **14**(13), 3371–3385 (2017) [https://doi.org/10.5194/bg-14-3371-2017].

28. M. T. Yilmaz, E. R. Hunt, and T. J. Jackson, “Remote sensing of vegetation water content from equivalent water thickness using satellite imagery,” Remote Sens Environ **112**(5), 2514–2522 (2008) [https://doi.org/10.1016/j.rse.2007.11.014].

29. A. A. Gitelson et al., “Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction,” https://doi.org/10.1080/01431160110107806 **23**(13), 2537–2562, Taylor & Francis Group (2010) [https://doi.org/10.1080/01431160110107806].

30. B. H. Menze et al., “A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data,” BMC Bioinformatics **10**(1), 1–16, BioMed Central (2009) [https://doi.org/10.1186/1471-2105-10-213/TABLES/4].

31. Y. Feng et al., “Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling,” Agric Water Manag **193**, 163–173, Elsevier B.V. (2017) [https://doi.org/10.1016/j.agwat.2017.08.003].

32. L. Breiman, “Random forests,” Mach Learn **45**(1), 5–32 (2001) [doi:10.1023/A:1010933404324].

33. S. Nembrini, I. R. Konig, and M. N. Wright, “Supplementary material for ‘The revival of the Gini importance?’” (2018).

34. M. Immitzer, C. Atzberger, and T. Koukal, “Tree species classification with Random Forest using very high spatial resolution 8-band worldView-2 satellite data,” Remote Sens (Basel) **4**(9), 2661–2693 (2012) [https://doi.org/10.3390/rs4092661].

35. E. M. Adam et al., “Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and Hyperspectral data resampled to HYMAP,” Int J Remote Sens **33**(2), 552–569 (2012) [https://doi.org/10.1080/01431161.2010.543182].

36. M. J. Pedro Júnior et al., “Curva de maturação e estimativa do teor de sólidos solúveis e acidez total em função de graus-dia: uva IAC 138-22 ‘Máximo,’” Bragantia **73**(1), 81–85, Instituto Agronômico de Campinas (2014) [https://doi.org/10.1590/BRAG.2014.011].

37. A. O. Santos et al., “Parâmetros fitotécnicos e condições microclimáticas para videira vinífera conduzida sob dupla poda sequencial,” Revista Brasileira de Engenharia Agrícola e Ambiental **15**(12), 1251–1256, Departamento de Engenharia Agrícola - UFCG (2011) [https://doi.org/10.1590/S1415-43662011001200006].

38. C. Chea et al., “Optimal models under multiple resource types for Brix content prediction in sugarcane fields using machine learning,” Remote Sens Appl **26**, 100718, Elsevier (2022) [https://doi.org/10.1016/J.RSASE.2022.100718].

39. U. Zibrat et al., “Use of Remote sensing technology to assess grapevine quality,” 2019 IEEE International Workshop on Metrology for Agriculture and Forestry, MetroAgriFor 2019 - Proceedings, 260–263, Institute of Electrical and Electronics Engineers Inc. (2019) [https://doi.org/10.1109/METROAGRIFOR.2019.8909261].

40. E. R. Hunt Jr. et al., “Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index,” in Agronomy Journal **103**, pp. 1090–1099 (2011).

41. K. Starý et al., “Comparing RGB - based vegetation indices from UAV imageries to estimate hops canopy area,” Agronomy Research **18**(4), 2592–2601, Eesti Pollumajandusulikool (2020) [https://doi.org/10.15159/AR.20.169].

42. M. Ferrer et al., “Mapping vineyard vigor using airborne remote sensing: relations with yield, berry composition and sanitary status under humid climate conditions,” Precision Agriculture 2019 21:1 **21**(1), 178–197, Springer (2019) [https://doi.org/10.1007/S11119-019-09663-9].

43. I. Bonilla, F. M. de Toda, and J. A. Martínez-Casasnovas, “Vine vigor, yield and grape quality assessment by airborne remote sensing over three years: Analysis of unexpected relationships in cv. Tempranillo,” Spanish Journal of Agricultural Research **13**(2), e0903, Ministerio de Agricultura Pesca y Alimentacion (2015) [https://doi.org/10.5424/sjar/2015132-7809].

44. S. N. Fredes, L. Ruiz, and J. A. Recio, “Modeling °Brix and pH in Wine Grapes from Satellite Images in Colchagua Valley, Chile,” Agriculture 2021, Vol. 11, Page 697 **11**(8), 697, Multidisciplinary Digital Publishing Institute (2021) [https://doi.org/10.3390/AGRICULTURE11080697].

6 CONSIDERAÇÕES FINAIS

Os objetivos propostos nesta tese demonstraram um grande avanço no quesito reconhecimento espectral e espacial de vinhedos em diferentes ambientes de cultivos. A aplicação dos modelos de espectroscopia, para classificação e predição dos dados remotamente adquiridos e em campo, apresentou métricas de avaliação satisfatórias, o que garante uma performance aceitável dos modelos.

Este estudo demonstrou que a discriminação espectral de variedades e região, utilizando algoritmos de aprendizagem de máquina(CDA, LGBM, RF, SVM), são relevantes uma vez que ajudam na compreensão das transformações da superfície da folha em diferentes IGs, e a identidade espectral para cada variedades ao longo de um espectro hiperespectral. Ademais constatou-se regiões espectrais e bandas, que melhor explicam a separabilidade entre as vinícolas e variedades. Os modelos de redução da dimensão dos espectros de reflectância hiperespectral(*Band ration*, KPCA), foram cruciais no processo de diminuição do número de variáveis e agilidade na execução dos modelos de classificação.

A estimativa dos parâmetros de clorofila coletados nas seis parcelas da vinícola Luiz Argenta apresentou desempenho favorável a predição utilizando reflectância hiperespectral, com destaque para RFR, sua performance foi superior aos PLSR, qualificando como um modelo mais adequado para processos preditivos envolvendo sensores hiperespectrais. A regressão para os parâmetros de grau brix, tanto por sensor proximal como por nanossatélite, construíram valores preditivos similares aos mensurados com refratômetro, ou seja, um monitoramento(folhas e dossel), e dossel torna-se adequado no controle da qualidade da uva, pelo menos para apontamentos envolvendo a doçura da uva.

As folhas e dosséis vegetativos em vinhedos são unidades de análises de difícil padronização espectral, as vezes pela similaridade entre suas assinaturas(folhas medidas

isoladamente) ou/e influência da estrutura do vigor vegetativo e/ou pela mistura de outras classes de usos(dossel mensurado por satélite). Os parâmetros de pigmentos fotossintéticos e grau brixº da uva são vinculados aos processos externos, a iteração entre planta e à ambiência, isso é complexo de ser analisado pela reflexão da luz. No entanto, o vigor vegetativo da vinha está interligado aos fatores espaciais onde a planta é cultivada; identificar características específicas nas assinaturas espectrais ao longo das parcelas auxiliará na construção de zonas homogêneas e amplificar o número de folhas medidas e datas analisadas ajudará a entender a identidade das assinaturas espectrais das variedades de forma temporal (acompanhamento dos estágios do ciclo) e relacionar com outros parâmetros de qualidade e quantidade da uva.

7 REFERÊNCIAS

ABBASI, Mozhgan *et al.* Optimal spectral wavelengths for discriminating orchard species using multivariate statistical techniques. **Remote Sensing**, [*s. l.*], v. 12, n. 1, 2020. Disponível em: Acesso em: 3 mar. 2021.

ACEVEDO-OPAZO, C. *et al.* The potential of high spatial resolution information to define within-vineyard zones related to vine water status. **Precision Agriculture**, [*s. l.*], v. 9, n. 5, p. 285–302, 2008.

ADAM, E. M. *et al.* Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and Hyperspectral data resampled to HYMAP. **International Journal of Remote Sensing**, [*s. l.*], v. 33, n. 2, p. 552–569, 2012.

AKBARZADEH, Saman *et al.* Plant discrimination by Support Vector Machine classifier based on spectral reflectance. **Computers and Electronics in Agriculture**, [*s. l.*], v. 148, p. 250–258, 2018. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0168169917310268. Acesso em: 4 maio 2021.

BASHEER, Manar A.; EL KAFRAWY, Sameh B.; MEKAWY, Amal A. Identification of mangrove plant using hyperspectral remote sensing data along the red sea, Egypt. **Egyptian Journal of Aquatic Biology and Fisheries**, [*s. l.*], v. 23, n. 1, p. 27–36, 2019.

BERGSTRÄSSER, Sergej *et al.* HyperART: Non-invasive quantification of leaf traits using hyperspectral absorption-reflectance-transmittance imaging. **Plant Methods**, [*s. l.*], v. 11, n. 1, p. 1–17, 2015.

BONNARDOT, Valérie *et al.* Spatial variability of night temperatures at a fine scale over the stellenbosch wine district, SOUTH AFRICA. **Journal International des Sciences de la Vigne et du Vin**, [*s. l.*], v. 46, n. 1, p. 1–13, 2012.

BRAMLEY, R. G.V.; HAMILTON, R. P. Understanding variability in winegrape production systems 1. Within vineyard variation in yield over several vintages. **Australian Journal of Grape and Wine Research**, [*s. l.*], v. 10, n. 1, p. 32–45, 2004.

BRASIL. **LEI Nº 9.279**. [*S. l.*], 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9279.htm. Acesso em: 23 ago. 2022.

CABELLO-PASINI, Alejandro; MACÍAS-CARRANZA, Víctor. Optical properties of grapevine leaves: Reflectance, transmittance, absorptance and chlorophyll concentration. **Agrociencia**, [*s. l.*], v. 45, n. 8, p. 943–957, 2011. Disponível em: http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S1405-31952011000800007&lng=en&nrm=iso&tlng=en. Acesso em: 1 ago. 2021.

CEMIN, Gisele; DUCATI, Jorge Ricardo. Spectral Discrimination of Grape Varieties and a Search for Terroir Effects Using Remote Sensing. **Journal of Wine Research**, [*s. l.*], v. 22, n. 1, p. 57–78, 2011. Disponível em: http://www.tandfonline.com/doi/abs/10.1080/09571264.2011.550762.

DOGAN, Adnan *et al.* Grapevine leaf area measurements by using pixel values. **Comptes Rendus de L'Academie Bulgare des Sciences**, [*s. l.*], v. 71, n. 6, p. 772–779, 2018.

DUCATI, Jorge R.; SARATE, Rafael E.; FACHEL, Jandyra M. G. Application of remote sensing techniques to discriminate between conventional and organic vineyards in the Loire Valley, France. **OENO One**, [*s. l.*], v. 48, n. 3, p. 135, 2014. Disponível em: http://oeno-one.eu/article/view/1574.

ECKERT, Sandra; KNEUBÜHLER, Mathias. Application of HYPERION data to agricultural land classification and vegetation properties estimation in Switzerland. **Eckert, Sandra; Kneubühler, Mathias (2004). Application of HYPERION data to agricultural land classification and vegetation properties estimation in Switzerland. In: XXth ISPRS Congress, Istanbul, Turkey, 12 July 2004 - 23 July 2004, 866-872.**, [*s. l.*], n. Vol. 35, p. 866–872, 2004.

EMBRAPA. **A Viticultura no Brasil**. [*S. l.*], 2022. Disponível em: https://www.embrapa.br/cim-uva-e-vinho/a-viticultura-no-brasil. Acesso em: 23 ago. 2022.

FERREIRO-ARMÁN, M. *et al.* Vine variety discrimination with airborne imaging spectroscopy. **Remote Sensing and Modeling of Ecosystems for Sustainability IV**, [*s. l.*], v. 6679, p. 667909, 2007.

FORMAGGIO, Antonio Roberto; SANCHES, Ieda Del'Arco. **Sensoriamento Remoto na Agricultura** . São Paulo : Oficina Textos, 2017. v. 1ª

GALVÃO, Lênio Soares; FORMAGGIO, Antônio Roberto; TISOT, Daniela Arnold. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. **Remote Sensing of Environment**, [*s. l.*], v. 94, n. 4, p. 523–534, 2005. Disponível em: Acesso em: 19 dez. 2021.

GITELSON, Anatoly A. *et al.* Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy¶. **Photochemistry and Photobiology**, [*s. l.*], v. 75, n. 3, p. 272, 2002. Disponível em: http://doi.wiley.com/10.1562/00318655(2002)075%3C0272:ACCIPL%3E2.0.CO;2.Acesso em: 17 dez. 2021.

GONZÁLEZ-FERNÁNDEZ, A.B.; CATANZARITE TORRES, T.; RODRIGUEZ-PÉREZ, J.R. **Una Metodología Apoyada En Sig Para El Seguimiento Vitícola Y La Delimitación De Zonas Homogéneas De Vendimia En La D.O. Bierzo (León-España**. [*S. l.: s. n.*], 2010.

GONZATTI BOMBASSARO, Magno. ANÁLISE ESPECTRAL DE VARIEDADES DE Vitis vinifera A PARTIR DE DADOS RADIOMÉTRICOS DE DOSSEL EM VINHEDOS DE PINTO BANDEIRA, RS. [*s. l.*], 2016.

HALL, Andrew; LOUIS, John; LAMB, David. Characterising and mapping vineyard canopy using high-spatial- resolution aerial multispectral images. **Computers and Geosciences**, [*s. l.*], v. 29, n. 7, p. 813–822, 2003.

HELMAN, David *et al.* Using time series of high-resolution planet satellite images to monitor grapevine stem water potential in commercial vineyards. **Remote Sensing**, [*s. l.*], v. 10, n. 10, p. 1–22, 2018.

HENNESSY, Andrew; CLARKE, Kenneth; LEWIS, Megan. Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. **Remote Sensing**, [*s. l.*], v. 12, n. 1, p. 113, 2020. Disponível em: www.mdpi.com/journal/remotesensing.

HUNT JR., E Raymond *et al.* **Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index**. [*S. l.: s. n.*], 2011.

JOHNSON, L. F. *et al.* Mapping vineyard leaf area with multispectral satellite imagery. **Computers and Electronics in Agriculture**, [*s. l.*], v. 38, n. 1, p. 33–44, 2003.

JUNGES, Amanda Heemann *et al.* Leaf hyperspectral reflectance as a potential tool to detect diseases associated with vineyard decline. **Tropical Plant Pathology**, [*s. l.*], v. 45, n. 5, p. 522–533, 2020.

KARAKIZI, Christina; OIKONOMOU, Marios; KARANTZALOS, Konstantinos. Vineyard Detection and Vine Variety Discrimination from Very High-Resolution Satellite Data. **Remote Sensing**, [*s. l.*], v. 8, n. 3, p. 235, 2016. Disponível em: http://www.mdpi.com/2072-4292/8/3/235.

KNIPPER, Kyle R. *et al.* Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards. **Irrigation Science**, [*s. l.*], v. 37, n. 3, p. 431–449, 2019. Disponível em: http://dx.doi.org/10.1007/s00271-018-0591-y.

KOKALY, Raymond F. Investigating a Physical Basis for Spectroscopic Estimates of Leaf Nitrogen Concentration. **Remote Sensing of Environment**, [*s. l.*], v. 75, n. 2, p. 153–161, 2001. Disponível em: Acesso em: 13 set. 2022.

LACAR, F.M.; LEWIS, M.M.; GRIERSON, I.T. Use of hyperspectral reflectance for discrimination between grape varieties. *Em*: , 2002. **IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)**. [*S. l.*]: IEEE, 2002. p. 2878–2880. Disponível em: http://ieeexplore.ieee.org/document/978192/.

MARTÍN, P. *et al.* Using hyperspectral remote sensing to map grape quality in “Tempranillo” vineyards affected by iron deficiency chlorosis. **Vitis - Journal of Grapevine Research**, [*s. l.*], v. 46, n. 1, p. 7–14, 2007.

ORDÓÑEZ, Celestino *et al.* Determining optimum wavelengths for leaf water content estimation from reflectance: A distance correlation approach. **Chemometrics and Intelligent Laboratory Systems**, [*s. l.*], v. 173, n. November 2017, p. 41–50, 2018.

PEÑUELAS, Josep; FILELLA, Iolanda; GAMON, John A. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. **New Phytologist**, [*s. l.*], v. 131, n. 3, p. 291–296, 1995. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.1995.tb03064.x. Acesso em: 19 dez. 2021.

PITHAN, Pâmela A. *et al.* Spectral characterization of fungal diseases downy mildew, powdery mildew, black-foot and Petri disease on Vitis vinifera leaves. **International Journal of Remote Sensing**, [*s. l.*], v. 42, n. 15, p. 5680–5697, 2021. Disponível em: https://www.tandfonline.com/doi/full/10.1080/01431161.2021.1929542. Acesso em: 17 jun. 2021.

PLANETSCOPE - EARTH ONLINE. [*S. l.*], [*s. d.*]. Disponível em: https://earth.esa.int/eogateway/missions/planetscope. Acesso em: 14 set. 2022.

POBLETE-ECHEVERRÍA, Carlos *et al.* remote sensing Detection and Segmentation of Vine Canopy in Ultra-High Spatial Resolution RGB Imagery Obtained from Unmanned Aerial Vehicle (UAV): A Case Study in a Commercial Vineyard. [*s. l.*], v. 9, p. 268, 2017. Disponível em: www.mdpi.com/journal/remotesensing.

QIN, Jiang Lin *et al.* A non-linear model for measuring grapevine leaf thickness by means of red-edge/near-infrared spectral reflectance. **Acta Ecologica Sinica**, [*s. l.*], v. 30, n. 6, p. 297–303, 2010. Disponível em: http://dx.doi.org/10.1016/j.chnaes.2010.08.012.

RENZULLO, Luigi J.; BLANCHFIELD, Annette L.; POWELL, Kevin S. A method of wavelength selection and spectral discrimination of hyperspectral reflectance spectrometry. **IEEE Transactions on Geoscience and Remote Sensing**, [*s. l.*], v. 44, n. 7, p. 1986–1994, 2006.

REYNOLDS, ANDREW G.; BROWN, Ralph; Utilization of proximal sensing technology (Greenseeker) to map variability in Ontario vineyards. **19th International Symposium GiESCO**, [*s. l.*], n. July, p. 593–597, 2015.

RODRÍGUEZ-PÉREZ, José R. *et al.* Evaluation of hyperspectral reflectance indexes to detect grapevine water status in vineyards. **American Journal of Enology and Viticulture**, [*s. l.*], v. 58, n. 3, p. 302–317, 2007.

SCHOEDL, K. *et al.* An experimental design applied to vineyards for identifying spatially and temporally variable crop parameters. **Vitis - Journal of Grapevine Research**, [*s. l.*], v. 51, n. 2, p. 53–60, 2012.

SILVA, Patrícia Rodrigues da; DUCATI, Jorge Ricardo. Spectral features of vineyards in south Brazil from ASTER imaging. **International Journal of Remote Sensing**, [*s. l.*], v. 30, n. 23, p. 6085–6098, 2009. Disponível em: https://www.tandfonline.com/doi/full/10.1080/01431160902810612.

SLATON, Michèle R.; HUNT, E. Raymond; SMITH, William K. Estimating near-infrared leaf reflectance from leaf structural characteristics. **American Journal of Botany**, [*s. l.*], v. 88, n. 2, p. 278–284, 2001.

STEELE, Mark R. *et al.* Nondestructive estimation of anthocyanin content in grapevine leaves. **American Journal of Enology and Viticulture**, [*s. l.*], v. 60, n. 1, p. 87–92, 2009.

STEELE, Mark; GITELSON, Anatoly A.; RUNDQUIST, Donald. Nondestructive estimation of leaf chlorophyll content in grapes. **American Journal of Enology and Viticulture**, [*s. l.*], v. 59, n. 3, p. 299–305, 2008.

STREVER, A. E. A study of within-vineyard variability with conventional and remote sensing technology. **MSc(Agric) Viticulture thesis, Stellenbosch University.**, [*s. l.*], n. December, 2003.

TARANTINO, Eufemia; FIGORITO, Benedetto. Mapping Rural Areas with Widespread Plastic Covered Vineyards Using True Color Aerial Data. **Remote Sensing**, [*s. l.*], v. 4, n. 7, p. 1913–1928, 2012.

THUM, Adriane B. *et al.* The influence of mineral content on spectral features of vine leaves. **International Journal of Remote Sensing**, [*s. l.*], v. 41, n. 23, p. 9161–9179, 2020. Disponível em: https://www.tandfonline.com/doi/abs/10.1080/01431161.2020.1798547. Acesso em: 19 fev. 2021.

TIAN, Jinyan *et al.* Comparison of UAV and WorldView-2 imagery for mapping leaf area index of mangrove forest. **International Journal of Applied Earth Observation and Geoinformation**, [*s. l.*], v. 61, n. May, p. 22–31, 2017. Disponível em: http://dx.doi.org/10.1016/j.jag.2017.05.002.

TISOT, Daniela A. *et al.* Hyperion/EO-1 hyperspectral data efficacy for agricultural targets identification: A comparisonwith ETM+/Landsat-7 multispectral data. **Engenharia Agricola**, [*s. l.*], v. 27, n. 2, p. 511–519, 2007.

VANEGAS, Fernando *et al.* A novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data. **Sensors (Switzerland)**, [*s. l.*], v. 18, n. 1, p. 1–21, 2018.

VAUDOUR, Emmanuelle. Les terroirs viticoles. Définitions, caractérisation et protection. [*s. l.*], p. 312, 2003.

VAUDOUR, E.; CAREY, V. A.; GILLIOT, J. M. Digital zoning of South African viticultural terroirs using bootstrapped decision trees on morphometric data and multitemporal SPOT images. **Remote Sensing of Environment**, [*s. l.*], v. 114, n. 12, p. 2940–2950, 2010.

VILLACRÉS, Juan *et al.* Retrieval of vegetation indices related to leaf water content from a single index: A case study of eucalyptus globulus (labill.) and pinus radiata (d. don.). **Plants**, [*s. l.*], v. 10, n. 4, 2021.

ZHANG, Chunhua *et al.* Separating Mangrove Species and Conditions Using Laboratory Hyperspectral Data: A Case Study of a Degraded Mangrove Forest of the Mexican Pacific. **Remote Sensing**, [*s. l.*], v. 6, n. 12, p. 11673–11688, 2014. Disponível em: http://www.mdpi.com/2072-4292/6/12/11673.